

**ALTERNATIONS OF MICRORNAS, THE MICROBIOME, AND  
GUT-HOST INTERACTIONS IN GASTROINTESTINAL  
DISEASES**

---

**TUNG ON YAU**

A thesis submitted in partial fulfilment of the requirements of  
Nottingham Trent University for the degree of Doctor of Philosophy

**August 2022**

## ABSTRACT

Over the past few decades, an ageing population combined with a shift towards a Western lifestyle has predisposed many individuals towards inter-connected gastrointestinal (GI) diseases, including inflammatory bowel disease (IBD), colorectal cancer (CRC), gastric cancer (GC) and *Clostridioides difficile* (*C. difficile*) infection (CDI). anti-TNF- $\alpha$  treatment for IBD patients has a high unresponsive rate, by using bioinformatics approaches, I identified neutrophil chemotaxis may contribute to the treatment resistance and IL13RA2 is the best predictor to identify treatment unresponsive patients. On the other hand, in the intestinal tract, colonocytes consistently exfoliate and shed into the lumen, affecting gut microbiota composition. These molecular/microbial changes involved in disease pathogenesis can be detected in faeces. By using Taqman probe-based real-time polymerase chain reaction (RT-qPCR) assay, several non-coding microRNAs (such as miR-18a, miR-20a, miR-221 and miR-135b) and gut microbes (including *Fusobacterium nucleatum*, *Parvimonas micra*, *Gemella morbillorum*, *Peptostreptococcus anaerobius*, *Clostridium hathewayi* and *Lachnoclostrium sp.*) are highly expressed/enriched in faeces in CRC individuals compared to control subjects. The use of a faecal immunological test (FIT) in combination with these biomarkers may improve the non-invasive CRC screening accuracy. Furthermore, *Epstein-Barr* virus (EBV) is an oncogenic virus and EBV-driven GC accounts for roughly 10% of total GC cases. GC cells infected with EBV alter the molecular aspect at whole-genome, transcriptome, and epigenome levels. For instance, AKT2 activated by mutation in EBV<sub>positive</sub> GC cells affecting downstream MAPK and focal adhesion signalling pathways; AKT2 mutation associates with poor patient survival in EBV-positive GC. Furthermore, once patients have received GI treatments, it may suppress/interfere with the patients' immune system, disrupt the gut flora homeostasis and trigger CDI. Faecal microbiota transplantation (FMT) has been demonstrated as an effective and alternative treatment strategy for CDI patients. However, it is still in clinical trials due to safety concerns. My study revealed that serum miRNAs such as miR-23a-3p, miR-150-5p, miR-26b-5p and miR-28-5p could be used to monitor FMT treatment in CDI patients, and these markers inversely correlate with IL-12B, IL-18, FGF21 and TNFSRF9 at serum protein and mRNA levels, respectively. Furthermore, miR-23a and miR-150 showed cytoprotective effects against *C. difficile* Toxin B (TcdB).

## **COPYRIGHT STATEMENT**

The copyright in this work is held by the author. You may copy up to 5% of this work for private study or personal, non-commercial research. Any re-use of the information contained within this document should be fully referenced, quoting the author, title, university, degree level and pagination. Queries or requests for any other use, or if a more substantial copy is required, should be directed to the author.

## **ACKNOWLEDGEMENTS**

I would like to express my most profound appreciation to Prof Jun Yu and Prof Jessie Qiaoyi Liang from the Chinese University of Hong Kong; Prof Graham Ball from Anglia Ruskin University/Nottingham Trent University, Prof Christos Polytarchou, Dr Maria Hatzia Apostolou, Dr Benjamin Dickins and Prof Sergio Rutella from Nottingham Trent University and Prof Nicola Holden from Scotland's Rural College (SRUC)/James Hutton Institute, who employed me to participated in a variety of research projects.

I would like to thank Dr Benjamin Dickins from Nottingham Trent University and Prof Shan Gao from Nankai University for allowing me to participate in bioinformatics related research projects.

I especially thank Prof Jessie Qiaoyi Liang, Prof Graham Ball, Prof Sergio Rutella, Prof Tanya Monaghan, Dr Ceen-Ming Tang, Dr Shivadas Sivasubramaniam, Dr Cristina Montiel-Duarte, Dr David Boocock, Dr Kirsty Hunter and Dr Sammy Chung for their support of my long PhD degree application process, prolonged owing to a series of "administrative errors" made by the doctoral school.

Thank you to all the researchers, medical professionals, lecturers, and professors I met around the world, either face to face or even only on the Internet during my research journey, and all the data contributors who deposited their raw data and have thus contributed to our bioscience community.

Thank you too to all my family members and friends for encouraging and supporting me.

## TABLE OF CONTENTS

ABSTRACT .....	I
COPYRIGHT STATEMENT .....	II
ACKNOWLEDGEMENTS .....	III
TABLE OF CONTENTS .....	IV
LIST OF FIGURES .....	VI
LIST OF TABLES .....	VII
ABBREVIATIONS .....	VIII
LIST OF PEER-REVIEWED PUBLICATIONS .....	XII
<b>OVERVIEW OF COLORECTAL CANCER AND EPSTEIN-BARR VIRUS-ASSOCIATED GASTRIC CANCER</b> .....	XII
<b>INCREASE ANTI-TNF-ALPHA TREATMENT RESPONSIVENESS IN PATIENTS WITH INFLAMMATORY BOWEL DISEASE</b> .....	XIII
<b>STOOL-BASED MICRORNA AND GUT MICROBES AS NON-INVASIVE BIOMARKERS FOR COLORECTAL CANCER SCREENING</b> .....	XIII
<b>GUT-HOST INTERACTION</b> .....	XVI
<b>I. INTRODUCTION</b> .....	1
<b>II. COLORECTAL CANCER</b> .....	2
<b>i. SPORADIC COLORECTAL CANCER</b> .....	2
(1). <b>Classical adenoma-carcinoma sequence pathway</b> .....	3
<b>A. chromosomal instability</b> .....	3
<b>B. microsatellite instability</b> .....	5
<b>C. DNA hyper-methylation</b> .....	6
(2). <b>Serrated colorectal cancer pathway</b> .....	7
<b>ii. COLITIS-ASSOCIATED COLORECTAL CANCER</b> .....	10
(1). <b>Pathogenesis of colitis-driven colorectal cancer</b> .....	10
(2). <b>The role of tumour necrosis factor <math>\alpha</math> in colitis-associated colorectal cancer</b> .....	12
<b>iii. microRNA</b> .....	14
<b>i. CANONICAL PATHWAY OF microRNA BIOGENESIS</b> .....	14
<b>ii. NON-CANONICAL PATHWAYS OF microRNA BIOGENESIS</b> .....	15
<b>iii. NOMENCLATURE OF microRNA</b> .....	17
<b>iv. PROBE-BASED REAL-TIME POLYMERASE CHAIN REACTION</b> .....	18
<b>IV. MICROBIOTA</b> .....	21
<b>i. EPSTEIN-BARR VIRUS-INDUCED GASTRIC CANCER</b> .....	21
(1). <b>Identification of <i>Epstein-Barr</i> virus-associated gastric cancer</b> .....	21
<b>ii. CLOSTRIDIODES DIFFICILE INFECTION</b> .....	22
(1). <b><i>Clostridioides difficile</i> infection and microRNA dysregulation</b> .....	23
<b>iii. GUT MICROBES IN COLORECTAL CANCER</b> .....	24
(1). <b>Gut microbes affect host microRNAs in colorectal cancer carcinogenesis</b> .....	26
(2). <b>Gut flora influenced by host microRNA</b> .....	28
<b>V. RESEARCH METHODOLOGY</b> .....	29
<b>VI. RESULTS</b> .....	34
<b>i. INCREASE ANTI-TNF-<math>\alpha</math> TREATMENT RESPONSIVENESS IN PATIENTS WITH INFLAMMATORY BOWEL DISEASE</b> .....	34
(1). <b>Introduction</b> .....	34
(2). <b>Summary of the outcomes</b> .....	34
(3). <b>Discussions and Conclusions</b> .....	35
<b>ii. STOOL-BASED MICRORNA AS NON-INVASIVE BIOMARKERS FOR COLORECTAL CANCER SCREENING</b> .....	36
(1). <b>Introduction</b> .....	36
(2). <b>Summary of the outcomes</b> .....	37
(3). <b>Discussions and Conclusions</b> .....	39

A. microRNA-20a and miR-18a in microRNA-17-92 Cluster in Colorectal Cancer .....	40
B. microRNA-21 in Colorectal Cancer .....	43
C. microRNA-221 in Colorectal Cancer .....	44
iii. GUT MICROBES AS NON-INVASIVE BIOMARKERS FOR COLORECTAL CANCER SCREENING .....	45
(1). Introduction .....	45
(2). Summary of the outcomes .....	45
(3). Discussions and Conclusions .....	47
iv. GUT-HOST INTERACTION .....	48
(1). Introduction .....	48
(2). Summary of the outcomes .....	49
A. Host-microbe and Metabolic Interactions are Differentially Shaped by Geographic location and Body Weight .....	49
B. Faecal Microbial Transplantation is an effective treatment strategy for <i>Clostridioides difficile</i> infection .....	51
C. Genome-wide, Transcriptomic and Epigenomic level changes in Epstein-Barr virus-associated gastric cancer .....	52
(3). Discussions and Conclusions .....	53
VII. OVERALL DISCUSSION AND FUTURE DIRECTIONS .....	54
VIII. CONCLUSION .....	58
IX. REFERENCES .....	59
X. PUBLISHED PEER-REVIEWED ARTICLES .....	78
APPENDIX 1. PUBLISHED PEER-REVIEWED ARTICLES REVERENT TO THE THESIS .....	292
APPENDIX 2. CONFERENCE ABSTRACTS REVERENT TO THE THESIS TOPIC .....	344

## LIST OF FIGURES

<b>Figure 1.</b> Distribution of colorectal cancer subtypes.....	2
<b>Figure 2.</b> Classical sporadic colorectal cancer pathway.....	3
<b>Figure 3.</b> Canonical Wnt/ $\beta$ -catenin signalling pathway.....	4
<b>Figure 4.</b> Schematic diagram of DNA mismatch repair. ....	6
<b>Figure 5.</b> The molecular mechanism linking DNA methylation and inactive transcription. ....	7
<b>Figure 6.</b> Serrated colorectal cancer pathways.....	8
<b>Figure 7.</b> Distribution of inflammatory bowel disease subtypes.....	10
<b>Figure 8.</b> Colitis-associated colorectal cancer in classical pathways.....	11
<b>Figure 9.</b> Therapeutic pyramid for the step-up treatment of inflammatory bowel disease....	12
<b>Figure 10.</b> TNF $\alpha$ signalling pathway. ....	13
<b>Figure 11.</b> Canonical and non-canonical pathways of microRNA biogenesis.....	16
<b>Figure 12.</b> The hsa-miR-92 Family and its nomenclature. ....	18
<b>Figure 13.</b> microRNA detection and quantification by TaqMan <sup>®</sup> probe-based RT-qPCR.....	20
<b>Figure 14.</b> Pathogenicity loci of toxicogenic <i>Clostridioides difficile</i> .....	23
<b>Figure 15.</b> TLR4 signalling pathway. ....	27
<b>Figure 16.</b> The analysis workflow on anti-TNF- $\alpha$ treatment resistance IBD patients. ....	29
<b>Figure 17.</b> Flowchart diagram of faecal-based miRNAs study selections based on the inclusion and exclusion criteria.....	30
<b>Figure 18.</b> Schematic diagram of the overall study design on rural and urban areas of central India individuals.....	32
<b>Figure 19.</b> Schematic diagram of the study pipeline on four severe or fulminant <i>Clostridioides difficile</i> infection patients (three responders and one non-responder). ....	33
<b>Figure 20.</b> Neutrophils, endothelial cells, and B cells are significantly higher on the baseline anti-TNF- $\alpha$ treatment IBD nonresponders compared with responders patients. ....	35
<b>Figure 21.</b> Identification of upregulated microRNAs in colorectal cancer by qPCR profiling.....	37
<b>Figure 22.</b> The AUROC of the combination of miR-221, miR-18a and miR-135b in faecal-based miRNA screening. ....	38
<b>Figure 23.</b> Diagnostic accuracy in pooled stool-based miR-21 for colorectal cancer screening. ....	39
<b>Figure 24.</b> microRNA-17-92 cluster at chromosome 13q31.3. ....	40
<b>Figure 25.</b> Validating robust gene markers associated with colorectal cancer in quantitative PCR. ....	46
<b>Figure 26.</b> Comparison and combination of bacterial markers with faecal immunochemical test (FIT). ....	47
<b>Figure 27.</b> The microbiota is structurally distinct in participants from rural vs. urban areas. ....	50
<b>Figure 28.</b> Summary of findings on miRNA changes in patients with <i>Clostridium difficile</i> infection after receiving fecal microbial transplants. ....	52
<b>Figure 29.</b> Summary of the study outcomes in <i>Epstein-Barr</i> virus associated gastric cancer .....	53

## LIST OF TABLES

<b>Table 1.</b> Clinical features of common serrated polyps.....	7
<b>Table 2.</b> Comparison between TaqMan® probe-based and SYBR® green for Real-Time quantitative polymerase chain reaction. ....	19
<b>Table 3.</b> The major reported gut microbes in colorectal cancer.....	25
<b>Table 4.</b> MicroRNA-microbiota interaction in mice models. ....	25
<b>Table 5.</b> Anti-tumour necrosis factor (TNF) therapeutic agents for moderate-to-severe IBD patients. ....	34
<b>Table 6.</b> microRNA differentially expressed in tumours compared with adjacent normal tissues. ....	38
<b>Table 7.</b> The reported microRNA-20a studies with the corresponding direct target genes in gastrointestinal diseases.....	41
<b>Table 8.</b> Reported microRNA-18a studies with the corresponding direct target genes in gastrointestinal diseases.....	42
<b>Table 9.</b> The reported microRNA-21 studies with the corresponding direct target genes in gastrointestinal diseases.....	44
<b>Table 10.</b> Reported microRNA-221 studies with the corresponding direct target genes in gastrointestinal diseases.....	45
<b>Table 11.</b> Statistically significant threshold difference in between responders and non-responder on four severe or fulminant <i>Clostridioides difficile</i> infection patients (three responders and one non-responder) in the taxonomy analysis (average across all timepoints). ....	51



## ABBREVIATIONS

### Gene names

Gene Abbreviation	Full Name
ABCB1	ATP Binding Cassette Subfamily B Member 1
ACAP2	ArfGAP With Coiled-Coil, Ankyrin Repeat And PH Domains 2
ACTB	Actin Beta
AGO	Argonate
AKT	Protein kinase B
APC	adenomatous polyposis coli
ASK1	Apoptosis signal-regulating kinase 1
ATG7	Autophagy Related 7
ATM	ATM serine/threonine kinase
BAX	BCL2 associated X protein
BCL2	B-cell lymphoma 2
BID	BH3 interacting-domain death agonist
BNIP2	BCL2 Interacting Protein 2
BRAF	B-Raf proto-oncogene serine/threonine kinase
c-Myc	MYC proto-oncogene
CAC1	CACUL1, CDK2 associated cullin domain 1
CACNA1G	Calcium Voltage-Gated Channel Subunit Alpha1 G
CACNA2D2	Calcium Voltage-Gated Channel Auxiliary Subunit Alpha2delta 2
CCND1	Cyclin D1
CASP5	Caspase 5
CD36	Cluster of differentiation 36
CDC42	GTPase cell division control protein 42
CDKN1C	Cyclin-Dependent Kinase Inhibitor 1C
CDKN2A	p16, p16 <sup>INK4a</sup> , cyclin-dependent kinase inhibitor 2A
CRABP1	Cellular Retinoic Acid Binding Protein 1
COL2A1	Collagen, type II, alpha 1
COX-2	Prostaglandin-endoperoxide synthase 2
CXCL8	Chemokine (C-X-C motif) ligand 8
CXCR1	C-X-C Motif Chemokine Receptor 1
DCC	DCC Netrin 1 Receptor
DDR1	CD167a, Discoidin domain receptor family, member 1
DGCR8	DiGeorge Syndrome Critical Region 8
DUSP8	Dual Specificity Phosphatase 8
<i>E-cadherin</i>	Epithelial cadherin
E2F4	E2F Transcription Factor 4
<i>EBER</i>	Epstein-Barr virus-encoded small RNA
EBNA-1	Epstein-Barr nuclear antigen 1
<i>EpCAM</i>	Epithelial cell adhesion molecule
ERK	Extracellular signal-regulated kinase
FADD	Fas-associated protein with death domain
FENDRR	FOXF1 Adjacent Non-Coding Developmental Regulatory RNA
FGF21	Fibroblast growth factor 21
FOXJ2	Forkhead Box J2
GDH	Glutamate dehydrogenase
GLUT4	Glucose transporter type 4
<i>HNRNPA1</i>	Heterogeneous nuclear ribonucleoprotein A1
<i>HPGD</i>	Hydroxyprostaglandin dehydrogenase 15-(NAD)
ID4	Inhibitor Of DNA Binding 4, HLH Protein
IFN- $\gamma$	Interferon gamma
IGF2	Insulin-like growth factor 2
IGF2R	Insulin-like growth factor 2 receptor
IGFBP7	Insulin Like Growth Factor Binding Protein 7

<b>Gene Abbreviation</b>	<b>Full Name</b>
<b>IKK</b>	IκB kinase
<b>IL-1β</b>	Interleukin 1 beta
<b>IL-6</b>	Interleukin 6
<b>IL-8</b>	Interleukin 8
<b>IL12</b>	Interleukin 12
<b>IL-12B</b>	Interleukin 12 beta
<b>IL-17</b>	Interleukin 17
<b>IL17a</b>	Interleukin 17 alpha
<b>IL-18</b>	Interleukin 18
<b>IL-22</b>	Interleukin 22
<b>IRAK1</b>	Interleukin-1 receptor-associated kinase 1
<b>IRAK4</b>	Interleukin-1 receptor-associated kinase 4
<b>IRF2</b>	Interferon regulatory factor 2
<b>IRF3</b>	Interferon regulatory factor 3
<b>KRAS</b>	KRAS Proto-Oncogene, GTPase
<b>JNK</b>	c-Jun N-terminal kinase
<b>JNK1</b>	Jun N-terminal Kinase 1
<b>Ki67</b>	Marker Of Proliferation Ki-67
<b>MAPK</b>	Mitogen-activated protein kinase
<b>MICA</b>	MHC Class I Polypeptide-Related Sequence A
<b>MCP-1</b>	Monocyte chemoattractant protein-1
<b>MD2</b>	LY96, Lymphocyte Antigen 96
<b>MEK</b>	Mitogen-Activated Protein Kinase Kinase;
<b>MGMT</b>	O-6-Methylguanine-DNA Methyltransferase
<b>MLH1</b>	MutL homolog 1
<b>MMP9</b>	matrix metalloproteinase 9
<b>MSH2</b>	MutL homolog 2
<b>MSH3</b>	MutL homolog 3
<b>MSH6</b>	MutL homolog 6
<b>MYC</b>	MYC Proto-Oncogene
<b>MyD88</b>	myeloid differentiation primary response
<b>NCOA4</b>	Nuclear Receptor Coactivator 4
<b>NEDD9</b>	Neural Precursor Cell Expressed, Developmentally Down-Regulated 9
<b>NEUROG1</b>	Neurogenin 1
<b>NF-κB</b>	Nuclear factor-kappa B
<b>p50</b>	Nuclear factor NF-kappa-B p105 subunit
<b>p53</b>	Tumour protein P53
<b>p65</b>	Transcription factor p65
<b>PDCD4</b>	Programmed cell death 4
<b>PDLIM2</b>	PDZ And LIM Domain 2
<b>PMS2</b>	PMS1 homolog 2
<b>PPARγ</b>	Peroxisome proliferator-activated receptor gamma
<b>PTEN</b>	Phosphatase And Tensin Homolog
<b>RIP1</b>	Receptor-interacting serine/threonine-protein kinase 1
<b>RIPK1</b>	Receptor-interacting serine/threonine-protein kinase 1
<b>RISC</b>	RNA-induced silencing complex
<b>ROCK1</b>	Rho Associated Coiled-Coil Containing Protein Kinase 1
<b>RNF43</b>	Ring Finger Protein 43
<b>RUNX3</b>	Runt-related transcription factor 3
<b>SEC23A</b>	SEC23 Homolog A
<b>SMAD2</b>	SMAD Family Member 2
<b>SMAD4</b>	SMAD Family Member 4
<b>SOCS1</b>	Suppressor Of Cytokine Signaling 1
<b>SPRY1</b>	Sprouty 1
<b>SPRY2</b>	Sprouty 2

<b>Gene Abbreviation</b>	<b>Full Name</b>
<b>STAT3</b>	Signal transducer and activator of transcription 3
<b>TAB</b>	transforming growth factor- $\beta$ -activated kinase 1-binding protein
<b>TACE</b>	TNF-converting enzyme
<b>TAK1</b>	Transforming growth factor $\beta$ -activated kinase 1
<b>TBK1</b>	TANK-binding kinase 1
<b>TGF-<math>\beta</math></b>	Transforming growth factor beta
<b>TGFBR</b>	Transforming growth factor- $\beta$ receptor
<b>TGFBR2</b>	Transforming Growth Factor Beta Receptor 2
<b>TIAM1</b>	TIAM Rac1 Associated GEF 1
<b>TIM</b>	TRAF-interacting
<b>TIMP2</b>	TIMP Metallopeptidase Inhibitor 2
<b>TIRAP</b>	Toll/interleukin-1 receptor domain-containing adapter protein
<b>TLR4</b>	Toll-like receptor 4
<b>TNF<math>\alpha</math></b>	Tumor Necrosis Factor- $\alpha$
<b>TNFR1</b>	tumour necrosis factor receptor 1
<b>TNFR2</b>	tumour necrosis factor receptor 2
<b>TNFRSF9</b>	TNF Receptor Superfamily Member 9
<b>TP53</b>	Tumour protein 53
<b>TP53INP1</b>	Tumour protein 53-induced nuclear protein 1
<b>TRADD</b>	TNFRSF1A Associated Via Death Domain
<b>TRAF2</b>	TNF receptor-associated factor 2
<b>TRAF3</b>	TNF receptor-associated factor 3
<b>TRAF6</b>	TNF receptor-associated factor 6
<b>TRAM</b>	Translocation Associated Membrane Protein
<b>TRBP</b>	Trans-activation-responsive RNA binding protein
<b>TRIF</b>	TIR-domain-containing adapter-inducing interferon- $\beta$
<b>Ubc13</b>	Ubiquitin-conjugating enzyme E2 13
<b>Uev1A</b>	Ubiquitin-conjugating enzyme E2 variant 1A
<b>ULK1</b>	Urothelial Cancer Associated 1
<b>XPO-5</b>	Exportin 5

## Others

<b>Abbreviation</b>	<b>Full Name</b>
<b>3'UTRs</b>	3'-untranslated regions
<b>m<sup>7</sup>G</b>	7-methylguanosine
<b>AGO</b>	Argonaute
<b>AP-1</b>	Activator protein 1
<b>AOM</b>	Azoxymethane
<b>BMI</b>	Body mass index
<b><i>C. difficile</i></b>	<i>Clostridioides difficile</i>
<b>CAC</b>	Colitis-associated colorectal cancer
<b>CCTA</b>	Cell cytotoxicity neutralisation assay
<b>CD</b>	Crohn's disease
<b>CDI</b>	<i>Clostridioides difficile</i> infection
<b><i>C. hathewayi</i></b>	<i>Clostridium hathewayi</i>
<b>CIMP</b>	CpG island methylator phenotype
<b>CIN</b>	Chromosomal instability
<b>CpGs</b>	Cytosine-phosphate-guanine dinucleotides
<b>CRC</b>	Colorectal cancer
<b>COVID-19</b>	Coronavirus disease 2019
<b>DCA</b>	Deoxycholic acid
<b>DNMTs</b>	DNA methyltransferase(s)
<b>EBV</b>	Epstein-Barr virus
<b>EBVaGC</b>	EBV-associated gastric cancer

<b>Abbreviation</b>	<b>Full Name</b>
<b>EIA</b>	Enzyme immunoassay
<b>ELISA</b>	Enzyme-linked immunosorbent assay
<b>FAP</b>	Familial adenomatous polyposis
<b>FFPE</b>	Formaldehyde and embedded in paraffin
<b>FIT</b>	Faecal immunological test
<b>FMT</b>	Fecal Microbiota Transplantation
<b>Fn</b>	<i>Fusobacterium nucleatum</i>
<b>GC</b>	Gastric cancer
<b>GCHP</b>	Goblet cell hyperplastic polyp
<b>GI</b>	Gastrointestinal
<b>GSP</b>	Glycated serum protein
<b><i>H. pylori</i></b>	<i>Helicobacter pylori</i>
<b>HCC</b>	hepatocellular carcinoma
<b>HNPCC</b>	Hereditary nonpolyposis colorectal cancer
<b>HP</b>	Hyperplastic polyp
<b>HPS</b>	Hamartomatous polyposis syndromes
<b>IBD</b>	Inflammatory bowel disease
<b>IBDU</b>	IBD unclassified
<b>IBS</b>	Irritable bowel syndrome
<b>lncRNA</b>	Long non-coding RNA
<b>ISH</b>	<i>In situ</i> hybridisation
<b>LBP</b>	Lipopolysaccharide-binding protein
<b>LOH</b>	loss of heterozygosity
<b>LPS</b>	Lipopolysaccharides
<b>miRNA</b>	MicroRNAs
<b>MMR</b>	Mismatch repair
<b>MSI</b>	Microsatellite instability
<b>MSS</b>	Microsatellite stable
<b>MVHP</b>	Microvesicular hyperplastic polyp
<b>NAAT</b>	nucleic acid amplification test
<b>NAFLD</b>	non-alcoholic fatty liver disease
<b>NASH</b>	Non-alcoholic steatohepatitis
<b>PaLoc</b>	Pathogenicity locus
<b>PCNA</b>	proliferating cell nuclear antigen
<b>P. micra</b>	Parvimonas micra
<b>PMCID</b>	PubMed Central Identifier
<b>RISC</b>	RNA-induced silencing complex
<b>SARS-COV-2</b>	Severe acute respiratory syndrome coronavirus 2
<b>SCFAs</b>	Short-chain fatty acids
<b>SFCDI</b>	severe or fulminant <i>Clostridioides difficile</i> infection
<b>SNP</b>	Single-nucleotide polymorphism
<b>SSL</b>	Sessile serrated lesion
<b>SUMOylation</b>	Small ubiquitin-like modification
<b>TC</b>	Toxigenic culture
<b>TSA</b>	Traditional serrated adenomas
<b>UC</b>	Ulcerative colitis
<b>WHO</b>	World Health Organisation

## LIST OF PEER-REVIEWED PUBLICATIONS

### OVERVIEW OF COLORECTAL CANCER AND EPSTEIN-BARR VIRUS-ASSOCIATED GASTRIC CANCER

Order (Page)	Title	Authors	Journal name	IF	CS	No. of Cit.	Contribution towards publication	Access domains / Identifier
1 (79)	Non-Invasive Colorectal Cancer Screening: An Overview	Melanie Tepus, <b>Tung On Yau</b>	<i>Gastrointestinal Tumors</i> . 2020;3–4(7):62–73.	-	-	27	<ul style="list-style-type: none"> <li>• Corresponding author (80%)</li> <li>• Design the writing strategy</li> <li>• Wrote and revised the manuscript</li> </ul>	Journal URL: <a href="http://karger.com/Article/FullText/507701">karger.com/Article/FullText/507701</a> DOI: <a href="https://doi.org/10.1002/jgh3.12153">10.1002/jgh3.12153</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/32903904/">32903904</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC7445682/">PMC7445682</a>
2 (92)	Precision Treatment in Colorectal Cancer: Now and the Future	<b>Tung On Yau</b>	<i>JGH Open: An Open Access Journal of Gastroenterology and Hepatology</i> . 2019; 3(5):361–69.	-	1.9	30	<ul style="list-style-type: none"> <li>• First author (100%)</li> <li>• Design the searching strategy</li> <li>• Wrote and revised the manuscript</li> </ul>	Journal URL: <a href="https://onlinelibrary.wiley.com/doi/full/10.1002/jgh3.12153">onlinelibrary.wiley.com/doi/full/10.1002/jgh3.12153</a> DOI: <a href="https://doi.org/10.1002/jgh3.12153">10.1002/jgh3.12153</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/31633039/">31633039</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC6788378/">PMC6788378</a>
3 (102)	The Role of microRNAs in Development of Colitis-Associated Colorectal Cancer	Marco Bocchetti, Maria Grazia Ferraro, Filippo Ricciardiello, Alessandro Ottaiano, Amalia Luce, Alessia Maria Cossu, Marianna Scrima, Wing-Yan Leung, Marianna Abate, Paola Stiuso, Michele Caraglia, Silvia Zappavigna#, <b>Tung On Yau</b> #	<i>International Journal of Molecular Sciences</i> . 2021; 22(8): 3967.	6.208	6.9	11	<ul style="list-style-type: none"> <li>• Co-corresponding author (65%)</li> <li>• Conceptualisation</li> <li>• Visualisation</li> <li>• Wrote the original draft revised and edited the manuscript</li> </ul>	Journal URL: <a href="https://mdpi.com/1422-0067/22/8/3967">mdpi.com/1422-0067/22/8/3967</a> DOI: <a href="https://doi.org/10.3390/ijms22083967">10.3390/ijms22083967</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/33921348/">33921348</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC8068787/">PMC8068787</a>
4 (119)	Epigenetic Dysregulation in Epstein-Barr Virus-Associated Gastric Carcinoma: Disease and Treatments	<b>Tung On Yau</b> , Ceen-Ming Tang, Jun Yu	<i>World Journal of Gastroenterology</i> . 2014; 20(21): 6448.	5.374	8.1	61	<ul style="list-style-type: none"> <li>• First author (80%)</li> <li>• Conceptualisation</li> <li>• Visualisation</li> <li>• Wrote the manuscript</li> </ul>	Journal URL: <a href="http://wjnet.com/1007-9327/full/v20/i21/6448.htm">wjnet.com/1007-9327/full/v20/i21/6448.htm</a> DOI: <a href="https://doi.org/10.3748/wjg.v20.i21.6448">10.3748/wjg.v20.i21.6448</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/24914366/">24914366</a>

IF, Clarivate impact factor (2021); CS, CiteScore (2021); Cit., citation, based on Google Scholar record dated July 11, 2022

## INCREASE ANTI-TNF-ALPHA TREATMENT RESPONSIVENESS IN PATIENTS WITH INFLAMMATORY BOWEL DISEASE

Order (Page)	Title	Authors	Journal name	IF	CS	No. of Cit.	Contribution towards publication	Access domains / Identifier
5 (129)	Hyperactive neutrophil chemotaxis contributes to anti-tumor necrosis factor- $\alpha$ treatment resistance in inflammatory bowel disease	<b>Tung On Yau</b> , Jayakumar Vadakekolathu, Gemma Ann Foulds, Guodong Du, Benjamin Dickins, Christos Polytarchou, Sergio Rutella	Journal of gastroenterology and hepatology	4.369	6.0	0	<ul style="list-style-type: none"> <li>• First author (90%)</li> <li>• Design the searching strategy</li> <li>• Data collection, integration, analysis and visualisation</li> <li>• Wrote and revised the manuscript</li> </ul>	Journal URL: <a href="https://onlinelibrary.wiley.com/doi/10.1111/jgh.15764">onlinelibrary.wiley.com/doi/10.1111/jgh.15764</a> DOI: <a href="https://doi.org/10.1111/jgh.15764">10.1111/jgh.15764</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/34931384/">34931384</a>

## STOOL-BASED MICRORNA AND GUT MICROBES AS NON-INVASIVE BIOMARKERS FOR COLORECTAL CANCER SCREENING

Order (Page)	Title	Authors	Journal name	IF	CS	No. of Cit.	Contribution towards publication	Access domains / Identifier
6 (141)	Identification of MicroRNA-135b in Stool as a Potential Noninvasive Biomarker for Colorectal Cancer and Adenoma	Chung Wah Wu, Siew Chien Ng, Yujuan Dong, Linwei Tian, Simon Siu Man Ng, Wing Wa Leung, Wai Tak Law, <b>Tung On Yau</b> , Francis Ka Leung Chan, Joseph Jao Yiu Sung, Jun Yu.	<i>Clinical Cancer Research</i> . 2014;20(11):299-3002.	13.801	19.8	136	<ul style="list-style-type: none"> <li>• Co-author (25%)</li> <li>• Clinical sample/data collection</li> <li>• Experimental design</li> <li>• Performed experiments</li> </ul>	Journal URL: <a href="https://clincancerres.aacrjournals.org/content/20/11/2994">clincancerres.aacrjournals.org/content/20/11/2994</a> DOI: <a href="https://doi.org/10.1158/1078-0432.CCR-13-1750">10.1158/1078-0432.CCR-13-1750</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/24691020/">24691020</a>
7 (152)	MicroRNA-221 and MicroRNA-18a Identification in Stool as Potential Biomarkers for the Non-Invasive Diagnosis of Colorectal Carcinoma	<b>Tung On Yau</b> , Chung Wah Wu, Yujuan Dong, Ceen-Ming Tang, Simon Siu Man Ng, Francis Ka Leung Chan, Joseph Jao Yiu Sung, Jun Yu.	<i>British Journal of Cancer</i> . 2014;111(9):176-71.	9.089	12.4	96	<ul style="list-style-type: none"> <li>• First author (75%)</li> <li>• Clinical sample collection</li> <li>• Study design</li> <li>• Conduct experiments</li> <li>• Statistical analysis</li> <li>• Wrote and revised the manuscript</li> </ul>	Journal URL: <a href="https://www.nature.com/articles/bjc2014484">nature.com/articles/bjc2014484</a> DOI: <a href="https://doi.org/10.1038/bjc.2014.484">10.1038/bjc.2014.484</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/25233396/">25233396</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC4453736/">PMC4453736</a>

IF, Clarivate impact factor (2021); CS, CiteScore (2021); Cit., citation, based on Google Scholar record dated July 11, 2022

**STOOL-BASED MICRORNA AND GUT MICROBES AS NON-INVASIVE BIOMARKERS FOR COLORECTAL CANCER SCREENING - continued**

Order (Page)	Title	Authors	Journal name	IF	CS	No of Cit.	Contribution towards publication	Access domains / Identifier
8 (160)	MicroRNA-20a in Human Faeces as a Non-Invasive Biomarker for Colorectal Cancer	<b>Tung On Yau</b> , Chung Wah Wu, Ceen-Ming Tang, Yingxuan Chen, Jingyuan Fang, Yujuan Dong, Qiaoyi Liang, Simon Siu Man Ng, Francis Ka Leung Chan, Joseph Jao Yiu Sung, Jun Yu.	<i>Oncotarget</i> . <b>2016</b> ;7(2):1559–68.	5.168	9.8	52	<ul style="list-style-type: none"> <li>• First Author (80%)</li> <li>• Clinical samples collection</li> <li>• Study design</li> <li>• Performed experiments</li> <li>• Statistical analysis</li> <li>• Wrote and revised the manuscript</li> </ul>	Journal URL: <a href="https://oncotarget.com/article/6403/">oncotarget.com/article/6403/</a> DOI: <a href="https://doi.org/10.18632/oncotarget.6403">10.18632/oncotarget.6403</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/26621842/">26621842</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC4811480/">PMC4811480</a>
9 (171)	Faecal MicroRNAs as a Non-Invasive Tool in the Diagnosis of Colonic Adenomas and Colorectal Cancer: A Meta-Analysis	<b>Tung On Yau</b> , Ceen-Ming Tang, Elinor K. Harriss, Benjamin Dickins, Christos Polytaichou.	<i>Scientific Reports</i> . <b>2019</b> ;9(1):9491	4.996	6.9	37	<ul style="list-style-type: none"> <li>• First author (80%)</li> <li>• Experimental design</li> <li>• Data Collection</li> <li>• Carry out the computational experiments</li> </ul>	Journal URL: <a href="https://www.nature.com/articles/s41598-019-45570-9">nature.com/articles/s41598-019-45570-9</a> DOI: <a href="https://doi.org/10.1038/s41598-019-45570-9">10.1038/s41598-019-45570-9</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/31263200/">31263200</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC6603164/">PMC6603164</a>

IF, Clarivate impact factor (2021); CS, CiteScore (2021); Cit., citation, based on Google Scholar record dated July 11, 2022

**STOOL-BASED MICRORNA AND GUT MICROBES AS NON-INVASIVE BIOMARKERS FOR COLORECTAL CANCER SCREENING - continued**

Order (Page)	Title	Authors	Journal name	IF	CS	No of Cit.	Contribution towards publication	Access domains / Identifier
10 (185)	Metagenomic Analysis of Faecal Microbiome as a Tool towards Targeted Non-Invasive Biomarkers for Colorectal Cancer	Jun Yu*, Qiang Feng*, Sunny Hei Wong*, Dongya Zhang*, Qiao yi Liang*, Youwen Qin, Longqing Tang, Hui Zhao, Jan Stenvang, Yanli Li, Xiaokai Wang, Xiaoqiang Xu, Ning Chen, William Ka Kei Wu, Jumana Al-Aama, Hans Jørgen Nielsen, Pia Kilerich, Benjamin Anderschou Holbech Jensen, <b>Tung On Yau</b> , Zhou Lan, Huijue Jia, Junhua Li, Liang Xiao, Thomas Yuen Tung Lam, Siew Chien Ng, Alfred Sze-Lok Cheng, Vincent Wai-Sun Wong, Francis Ka Leung Chan, Xun Xu, Huanming Yang, Lise Madsen, Christian Datz, Herbert Tilg, Jian Wang, Nils Brünner, Karsten Kristiansen, Manimozhiyan Arumugam#, Joseph Jao-Yiu Sung#, Jun Wang#	<i>Gut</i> . 2017;66(1):70–78.	31.840	40.1	640	<ul style="list-style-type: none"> <li>• Co-author (10%)</li> <li>• Clinical samples/data collection</li> <li>• Performed validation experiments</li> <li>• Data analysis</li> </ul>	Journal URL: <a href="http://gut.bmj.com/content/66/1/70">gut.bmj.com/content/66/1/70</a> DOI: <a href="https://doi.org/10.1136/gutjnl-2015-309800">10.1136/gutjnl-2015-309800</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/26408641/">26408641</a>
11 (196)	A Novel <i>Faecal Lachnospirillum</i> Marker for the Non-Invasive Diagnosis of Colorectal Adenoma and Cancer	Jessie Qiaoyi Liang, Tong Li, Geicho Nakatsu, Ying-Xuan Chen, <b>Tung On Yau</b> , Eagle Chu, Sunny Wong, Chun Ho Szeto, Siew C. Ng, Francis KL Chan, Jing-Yuan Fang, Joseph JY Sung, Jun Yu.	<i>Gut</i> . 2020;69(7):1248-1257.	31.840	40.1	83	<ul style="list-style-type: none"> <li>• Co-author (40%)</li> <li>• Experimental design</li> <li>• Clinical sample management</li> <li>• Carry out the experiments</li> </ul>	Journal URL: <a href="http://gut.bmj.com/content/69/7/1248">gut.bmj.com/content/69/7/1248</a> DOI: <a href="https://doi.org/10.1136/gutjnl-2019-318532">10.1136/gutjnl-2019-318532</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/31776231/">31776231</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC7306980/">PMC7306980</a>

IF, Clarivative impact factor (2021); CS, CiteScore (2021); Cit., citation, based on Google Scholar record dated July 11, 2022



## GUT-HOST INTERACTION

Order (Page)	Title	Authors	Journal name	IF	CS	No. of Cit.	Contribution towards publication	Access domains / Identifier
12 (208)	Multiomics Profiling Reveals Signatures of Dysmetabolism in Urban Populations in Central India	Tanya M. Monaghan, Rima N. Biswas, Rupam R. Nashine, Samidha S. Joshi, Benjamin H. Mullish, Anna M. Seekatz, Jesus Miguens Blanco, Julie A. K. McDonald, Julian R. Marchesi, <b>Tung On Yau</b> , Niki Christodoulou, Maria Hatziapostolou, Maja Pucic-Bakovic, Frano Vuckovic, Filip Klicek, Gordan Lauc, Ning Xue, Tania Dottorini, Shrikant Ambalkar, Ashish Satav, Christos Polyarchou, Animesh Acharjee, Rajpal Singh Kashyap	<i>Microorganisms</i> . 2021;9(7):1485.	4.926	4.1	0	<ul style="list-style-type: none"> <li>Co-author (15%)</li> <li>Clinical sample collection/management</li> <li>Experimental design</li> <li>Performed experiments</li> </ul>	Journal URL: <a href="https://mdpi.com/2076-2607/9/7/1485">mdpi.com/2076-2607/9/7/1485</a> DOI: <a href="https://doi.org/10.3390/microorganisms9071485">10.3390/microorganisms9071485</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/34361920/">34361920</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC8307859/">PMC8307859</a>
13 (232)	Fecal microbiota transplantation for recurrent <i>Clostridioides difficile</i> infection associates with functional alterations in circulating microRNAs	Tanya M Monaghan <sup>#</sup> , Anna M Seekatz*, Nicholas O Markham*, <b>Tung On Yau*</b> , Maria Hatziapostolou*, Tahseen Jilani, Niki Christodoulou, Brandi Roach, Eleni Birli, Odette Pomenya, Thomas Louie, D Borden Lacy, Peter Kim, Christine Lee, Dina Kao <sup>#</sup> , Christos Polyarchou <sup>#</sup>	<i>Gastroenterology</i> . 2021; 161(1):255-270.e4.	33.883	33.0	11	<ul style="list-style-type: none"> <li>Co-contributed author (40%)</li> <li>Conduct experiments</li> <li>Data analysis</li> <li>Clinical sample/data collection</li> </ul>	Journal URL: <a href="https://gastrojournal.org/article/S0016-5085(21)00577-1/fulltext">gastrojournal.org/article/S0016-5085(21)00577-1/fulltext</a> ScienceDirect: <a href="https://sciencedirect.com/science/article/abs/pii/S0016508521005771">sciencedirect.com/science/article/abs/pii/S0016508521005771</a> DOI: <a href="https://doi.org/10.1053/j.gastro.2021.03.050">10.1053/j.gastro.2021.03.050</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/33844988/">33844988</a>

IF, Clarivative impact factor (2021); CS, CiteScore (2021); Cit., citation, based on Google Scholar record dated July 11, 2022

## GUT-HOST INTERACTION - continued

Order (Page)	Title	Authors	Journal name	IF	CS	No. of Cit.	Contribution towards publication	Access domains / Identifier
14 (258)	A multi-factorial observational study on sequential fecal microbiota transplant in patients with medically refractory <i>Clostridioides difficile</i> infection	Tanya M Monaghan <sup>**</sup> , Niharika A Duggal <sup>*</sup> , Elisa Rosati <sup>*</sup> , Ruth Griffin, Jamie Hughes, Brandi Roach, David Y Yang, Christopher Wang, Karen Wong, Lynora Saxinger, Maja Pučić-Baković, Frano Vučković, Filip Klicek, Gordan Lauc, Paddy Tighe, Benjamin H Mullish, Jesus Miguens Blanco, Julie AK McDonald, Julian R Marchesi, Ning Xue, Tania Dottorini, Animesh Acharjee, Andre Franke, Yingrui Li, Gane Ka-Shu Wong, Christos Polytarchou, <b>Tung On Yau</b> , Niki Christodoulou, Maria Hatzia Apostolou, Minkun Wang, Lindsey A Russell <sup>#</sup> , Dina H Kao <sup>#</sup>	<i>Cells</i> <b>2021</b> , 10(11), 3234	7.666	6.7	2	<ul style="list-style-type: none"> <li>• Co-author (15%)</li> <li>• Conduct experiments</li> <li>• Data analysis</li> <li>• Clinical sample/data collection</li> </ul>	Journal URL: <a href="https://www.mdpi.com/2073-4409/10/11/3234">https://www.mdpi.com/2073-4409/10/11/3234</a> DOI: <a href="https://doi.org/10.3390/cells10113234">10.3390/cells10113234</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/34831456/">34831456</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC8624539/">PMC8624539</a>
15 (277)	Integrative Identification of Epstein-Barr Virus-Associated Mutations and Epigenetic Alterations in Gastric Cancer	Qiaoyi Liang <sup>*</sup> , Xiaotian Yao <sup>*</sup> , Senwei Tang, Jingwan Zhang, <b>Tung On Yau</b> , Xiaoxing Li, Ceen-Ming Tang, Wei Kang, Raymond Wm M. Lung, Jing Woei Li, Ting Fung Chan, Rui Xing, Youyong Lu, Kwok Wai Lo, Nathalie Wong, Ka Fai To, Chang Yu, Francis KL. Chan, Joseph JY Sung, Jun Yu	<i>Gastroenterology</i> . <b>2014</b> ;147(6):1350-62.e4.	33.883	33.0	110	<ul style="list-style-type: none"> <li>• Co-author (25%)</li> <li>• Experimental design</li> <li>• EBV and DNA methylated genes screening in GC cell and clinical samples</li> <li>• Clinical sample/data collection</li> </ul>	Journal URL: <a href="https://gastrojournal.org/article/S0016-5085(14)01075-0/">gastrojournal.org/article/S0016-5085(14)01075-0/</a> ScienceDirect: <a href="https://www.sciencedirect.com/science/article/pii/S016508514010750">sciencedirect.com/science/article/pii/S016508514010750</a> DOI: <a href="https://doi.org/10.1053/j.gastro.2014.08.036">10.1053/j.gastro.2014.08.036</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/25173755/">25173755</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC4047330/">PMC4047330</a>

IF, Clarivate impact factor (2021); CS, CiteScore (2021); Cit., citation, based on Google Scholar record dated July 11, 2022

## I. INTRODUCTION

Over the past few decades, an ageing population combined with dramatic modifications in lifestyle and obesity has predisposed many individuals towards interconnected gastrointestinal (GI) diseases, such as non-alcoholic fatty liver disease (NAFLD), inflammatory bowel disease (IBD), colorectal cancer (CRC), gastric cancer (GC) and *Clostridioides difficile* (*C. diff*) infection (CDI). It characterises one of the major global health issues, causing around 8 million deaths per year worldwide (Kim et al. 2014). In Europe, GI cancer is the leading cause of cancer death, with the most and the second-most common cancer in men and women, respectively (Farthing et al. 2014). GC and CRC are the major malignancy in GI cancer and account for the second and fourth most cancer-related deaths worldwide.

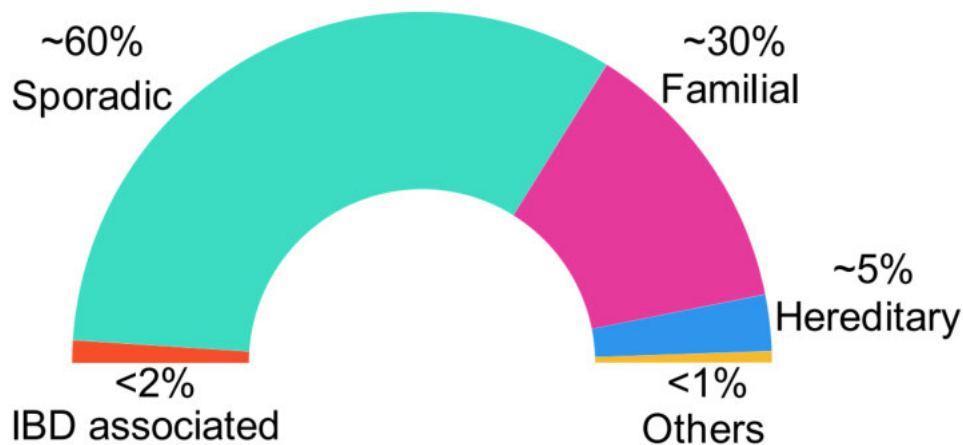
Benefiting from bowel cancer screening programmes (**Publication 1, page 79**) (Tepus, Yau 2020) and advanced therapeutic strategies (**Publication 2, page 92**) (Yau 2019), the population-based mortality for CRC has been falling for the past half-century, especially in Central, Western and Northern European countries, and particularly in men (Bray et al. 2018). However, the use of low-cost haemoglobin-based CRC screening such as the faecal immunological test (FIT) is limited by haemoglobin degradation and intermittent bleeding patterns in large intestines, so that a large number of individuals cannot be identified until the late stages, leading to poor treatment responses and prognosis (Tepus, Yau 2020). Thus, finding reliable non-invasive biomarkers is still ongoing; and the understanding of CRC pathogenesis on microRNA (miRNA) and gut microbes is critical in helping with biomarker discovery.

On the other hand, due to poor disease management of long-standing chronic colonic inflammation, a certain percentage of CRC cases are induced by IBD. One of the critical issues is the unresponsive treatment of anti-tumour necrosis factor  $\alpha$  (TNF $\alpha$ ) agents, which has a rate of up to 30% of initial unresponsive treatment and up to 50% of diminishing response over time (Ben-Horin et al. 2014). **(1)** Thus, finding biomarkers to predict treatment outcome and the mechanisms of unresponsive anti-TNF $\alpha$  are required. Here, microRNA is involved in the progression of colitis-associated CRC (CAC) (**Publication 3, page 102**) (Bocchetti et al. 2021) and the use of **(2)** faecal-based miRNA and **(3)** gut microbes could be useful for non-invasive CRC screening. Besides, **(4)** study of gut-host interactions is critically important for gaining new insight into gastrointestinal diseases. IBD patients also have a greater risk of CDI, and IBD patients with CDI have disproportionately higher morbidity and mortality (Balram et al. 2019). Faecal microbiota transplantation (FMT) is an effective alternative treatment strategy

for CDI, however, the molecular mechanisms of how the host responds to FMT remain unclear. Moreover, changes in stomach microbiota composition also increase the incidence of *Epstein-Barr* virus (EBV)-induced GC, which leads to significantly increased healthcare costs (**publication 4, page 119**) (Yau, Tang, et al. 2014), unravelling the differences between EBV infected and non-infected GCs through the use of whole genome, transcriptome and epigenome sequence analysis may improve our understanding of the disease, leading to precision medicine.

## II. COLORECTAL CANCER

Approximately 70-90% of total CRC cases are sporadic, thus about 10-30% of CRC cases are familial (Monahan et al. 2020). Around 5% of CRC subjects are linked with hereditary cancer predisposition, the majority (1-3% of total CRC cases) being Lynch syndrome (also called hereditary nonpolyposis CRC, HNPCC); hamartomatous polyposis syndromes (HPS) and familial adenomatous polyposis (FAP) are relatively low, with < 0.1% and < 1% of total CRC cases, respectively (Nguyen, Duong 2018; Monahan et al. 2020). Long-term IBD is also the main risk factor of CRC oncogenesis (Nadeem et al. 2020), with an estimated proportion of fewer than 2% of all CRC cases (Eluri et al. 2017) (**Figure 1**).



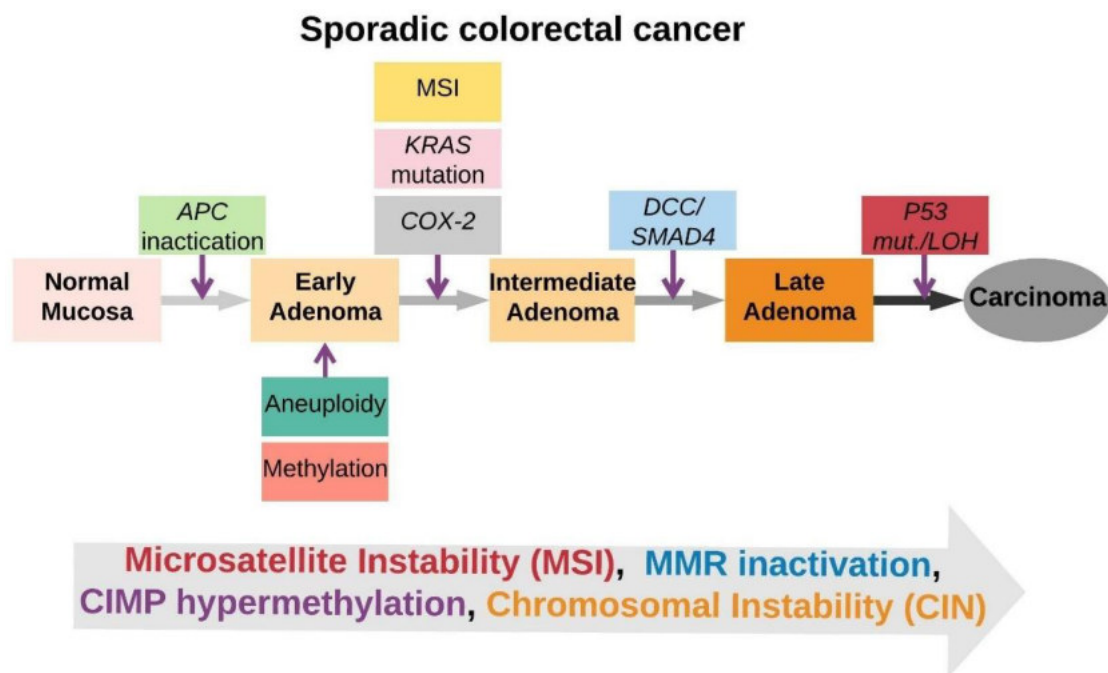
**Figure 1. Distribution of colorectal cancer subtypes.**

### i. SPORADIC COLORECTAL CANCER

The pathogenesis of CRC follows a stepwise progression from aberrant crypt formation in colonic mucosa to benign adenomas and malignant adenocarcinomas. Without early intervention, it gradually evolves into invasive and metastatic tumours (Toyota, Suzuki 2010; You, Jones 2012; Migliore et al. 2011). There are two major distinct clinicopathologic routes in sporadic CRCs based on genetic classifications: the classical adenoma-carcinoma pathway and serrated pathways (De Palma et al. 2019; Jass 2001).

### (1). Classical adenoma-carcinoma sequence pathway

CRC induced by adenoma-carcinoma sequence is histologically homogeneous and begins with tubulovillous or tubular adenoma precursor lesions (Nguyen, Duong 2018). The histological alterations are the consequence of molecular dysregulation (Armaghany et al. 2012). Broadly, the inactivation of adenomatous polyposis coli (*APC*) is the initial event in adenoma CRC formation, followed by *KRAS* mutation at the early neoplasm development. Then, the deletion and/or inactivation of the SMAD family member (*SMAD2* and *SMAD4*) on chromosome arm 18q and tumour-suppressor gene *TP53* on chromosome arm 17p occur during the transition to malignancy (**Figure 2**) (Diep et al. 2006; Jasmine et al. 2012; Zarzour et al. 2015). This proposed adenoma-carcinoma sequence pathway is a combination of the three tumorigenesis mechanisms of CRC, including (A) chromosomal instability (CIN); (B) microsatellite instability (MSI); and (C) DNA hyper-methylation (Gupta et al. 2018; Pino, Chung 2010). Any of the mechanisms may also involve in other types of CRC pathogenesis.

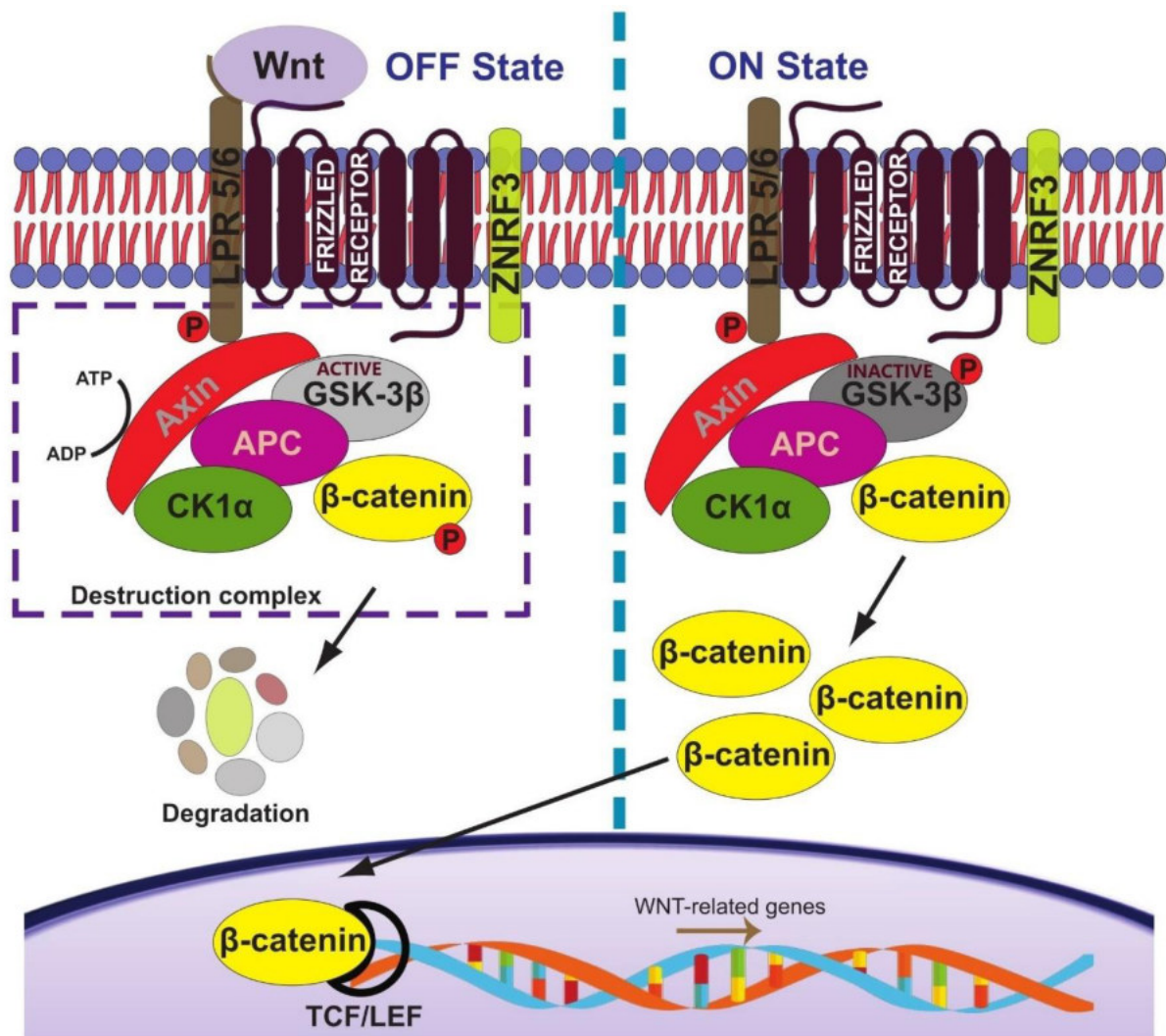


**Figure 2. Classical sporadic colorectal cancer pathway.** APC, Adenomatous Polyposis Coli; MSI, microsatellite instability; KRAS, KRAS Proto-Oncogene, GTPase; COX-2, Prostaglandin-endoperoxide synthase 2; DCC, Deleted in Colorectal Cancer; SMAD4, SMAD Family Member 4; p53, Tumour protein P53; LOH, loss of heterozygosity. Adopted and modified from (Cerrito, Grassilli 2021).

#### A. chromosomal instability

CIN is the most prevalent form of genomic instability that comprises deletion and/or duplication of either whole or parts of chromosomes. Roughly, 85% of sporadic CRCs are CIN and commonly found in the distal colon (Gupta et al. 2018; Pino, Chung 2010). CIN is usually associated with tumour suppressor genes inactivation, such as *APC*, *TP53* and/or *KRAS* mutation (Geigl et al. 2008; Pino, Chung 2010). Allelic deletion(s) and/or somatic mutation(s)

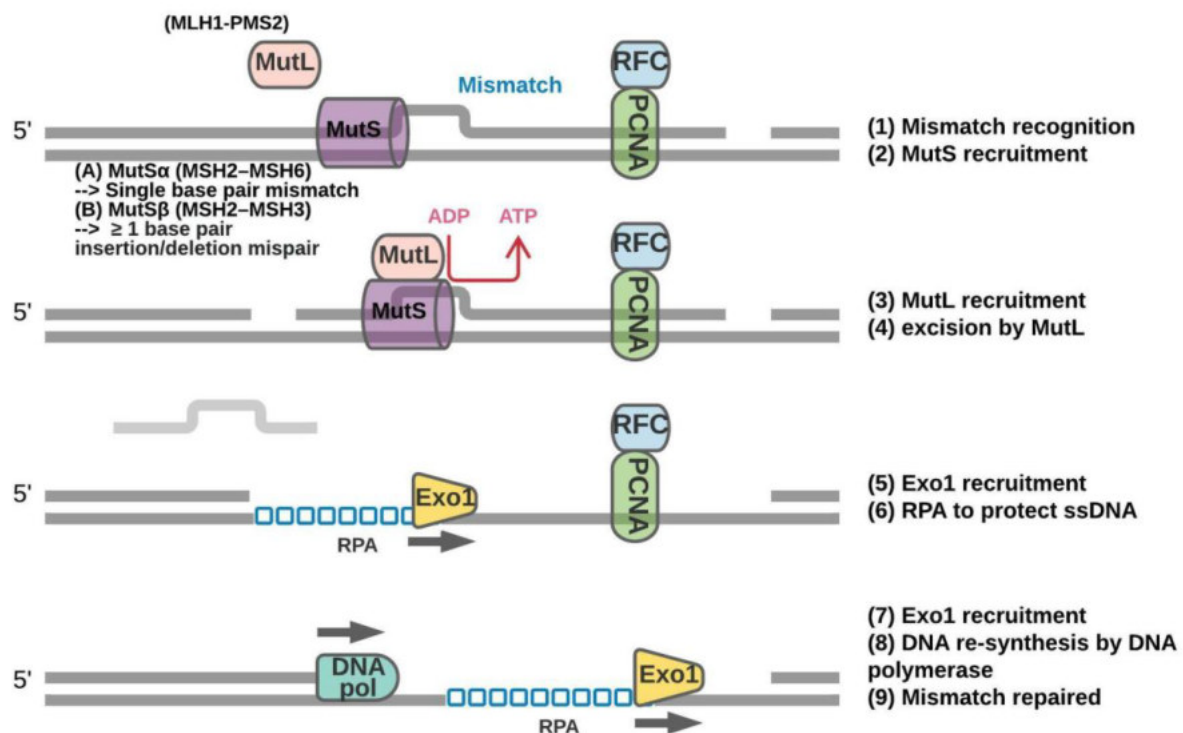
of APC are a critical factor for sporadic CRC (Kapitanović et al. 2004; Kinzler, Vogelstein 1996). Silencing of APC triggers canonical Wnt signalling cascade from the accumulation and stabilisation of  $\beta$ -catenin in the cytoplasm and is further translocated into the nucleus with T-cell factor/lymphoid enhancer factor (Tcf/Lef) complexes (**Figure 3**) (Cadigan, Waterman 2012; Pai et al. 2017). Then, it activates the transcription of *CCND1*, *AXIN2* and *MYC* (Shang et al. 2017; MacDonald et al. 2009; Coppedè et al. 2014), and promotes cell proliferation, invasion, migration and cancer metastasis (Stanczak et al. 2011).



**Figure 3. Canonical Wnt/ $\beta$ -catenin signalling pathway. OFF state:** The presence of WNT binds to a frizzled receptor and its co-receptor LRP5/6 activates  $\beta$ -catenin phosphorylation and interacts with the destruction complex (CK1 $\alpha$ , GSK-3 $\beta$ , Axin and APC), and subsequent degradation. **ON state:** The absence of WNT protein activates glycogen synthase kinase-3 $\beta$  (GSK-3 $\beta$ ). Here, the tumour suppressor of ZNRF3 induces degradation of WNT receptor. It prevents  $\beta$ -catenin phosphorylation/degradation and thus accumulates  $\beta$ -catenin in the cytoplasm and further translocated into the nucleus with Tcf/Lef complexes, subsequently inducing cell proliferation. P, phosphate; APC, adenomatous polyposis coli; GSK-3 $\beta$ , glycogen synthase kinase 3 $\beta$ ; LRP, lipoprotein receptor-related protein; Tcf/Lef, T-cell factor/lymphoid enhancer factor, CK, Casein kinase. Adopted and modified from (Silva-García et al. 2019; Zhang, Wang 2020).

## **B. microsatellite instability**

MSI consists of an accumulation of deletions and/or insertions of short nucleotide repeats (microsatellite) in DNA mismatch repair (MMR)-related genes without affecting chromosomal integrity (Boland, Goel 2010; Gupta et al. 2018). The DNA MMR system is there to identify and repair mistaken misincorporation, deletion and/or insertion of DNA base(s) during replication repair, and recombination (**Figure 4**). MSI occurs in about 15% of CRC cases; and over 80% of sporadic CRC with MSI are mainly associated with *MLH1* inactivation – one of the key regulating genes in MMR – via CpG island methylator phenotype (CIMP)-driven methylation (Weisenberger et al. 2006; Popat et al. 2005; Boland, Goel 2010). MSI can be classified and determined as stable (MSS), low (MSI<sup>Low</sup>) and high (MSI<sup>High</sup>) by using PCR based on the Bethesda panel assay (BAT25, BAT26, D2S123, D5S346, and D17S250) (Losso et al. 2012; Hegde et al. 2014). The MSI status can also be identified from whole-genome sequencing data using some bioinformatics tools (Hempelmann et al. 2015; Salipante et al. 2014; Niu et al. 2014). MSI<sup>High</sup> CRC often has *BRAF* mutations (Cancer Genome Atlas Network 2012) and has high frequent frameshift mutations on the genes that contain nucleotide repeats in exon coding regions, including *MSH6*, *MSH3*, *CASP5*, *E2F4*, *TGFBR2* and *IGF2R* (Souza et al. 1996; Saeki et al. 2001; Yoshitaka et al. 1996; Schwartz et al. 1999).

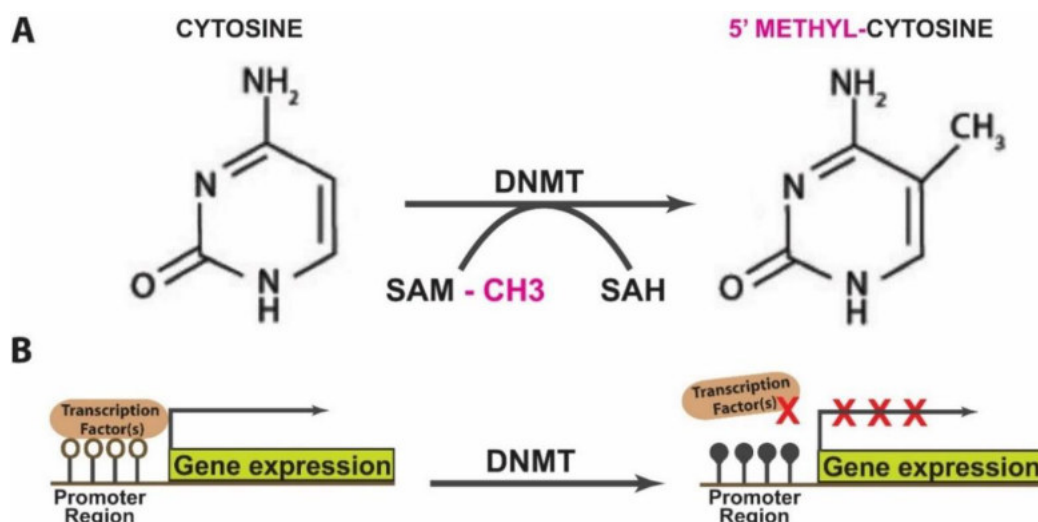


**Figure 4. Schematic diagram of DNA mismatch repair.** Recognition of a mismatch starts with MutS (MutS $\alpha$ : MSH2–MSH6 for single base-pair mismatch or MutS $\beta$ : MSH2–MSH3 for over 1 base pair of insertion/deletion mispair) and MutL (MutL $\alpha$ : MLH1–PMS2) forming a ternary complex. The DNA-protein and protein-protein interactions need ATP/ADP cofactor to bind MutL and MutS (MutS $\alpha$  or MutS $\beta$ ). PCNA and RFC are DNA replication factors that target the excision step to the newly synthesised strand. Then, Exo1 indicates the excision of one side of the DNA duplex having mismatched nucleotide(s) and coated with RPA on the single-strand gap. Resynthesis by the replicative DNA polymerase repairs the integrity of the duplex. MSH, mutS homolog; Exo1, excision by exonuclease 1; PCNA, proliferating cell nuclear antigen; RFC, replication factor C; DNA pol, DNA polymerase; ATP, adenosine triphosphate; ADP, adenosine diphosphate; RPA, replication protein A. Adopted and modified from (Pećina-Šlaus et al. 2020).

### C. DNA hyper-methylation

DNA methylation is an epigenetics process by which a methyl (CH<sub>3</sub>) group is added to a carbon-5 (C5) position of cytosine to form a 5-methyl-cytosine. This process precedes a CpG site by DNA methyltransferase(s) (DNMTs) without changing the DNA sequence (**Figure 5A**). Hyper-methylation of CpG islands in gene promoter regions presents a conformational change of chromatin, avoiding RNA polymerase and/or other regulatory protein(s) from accessing the promoter region, and thus suppressing gene transcription (**Figure 5B**) (Jin et al. 2011; Moore et al. 2013). The degree of DNA methylation can be utilised for CRC classification based on the CpG island methylator phenotype (CIMP), showing diverse molecular features, precursor lesions and histology (Issa 2004; Ang et al. 2010; Shen et al. 2007). The degree of CIMP can be sub-classified into high (CIMP<sup>High</sup>) low (CIMP<sup>Low</sup>) and negative (CIMP<sup>Negative</sup>) (Drew et al. 2017; Ogino et al. 2006).





**Figure 5. The molecular mechanism linking DNA methylation and inactive transcription.** (A) Cytosine converts to 5'-methyl-cytosine via DNA methyltransferase (DNMT). (B) The CpG islands methylated sites at promoter regions preventing the binding transcription factors via methyl CpG-binding protein, suppressing the down-stream transcription. DNMT, DNA methyltransferase; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine.

## (2). Serrated colorectal cancer pathway

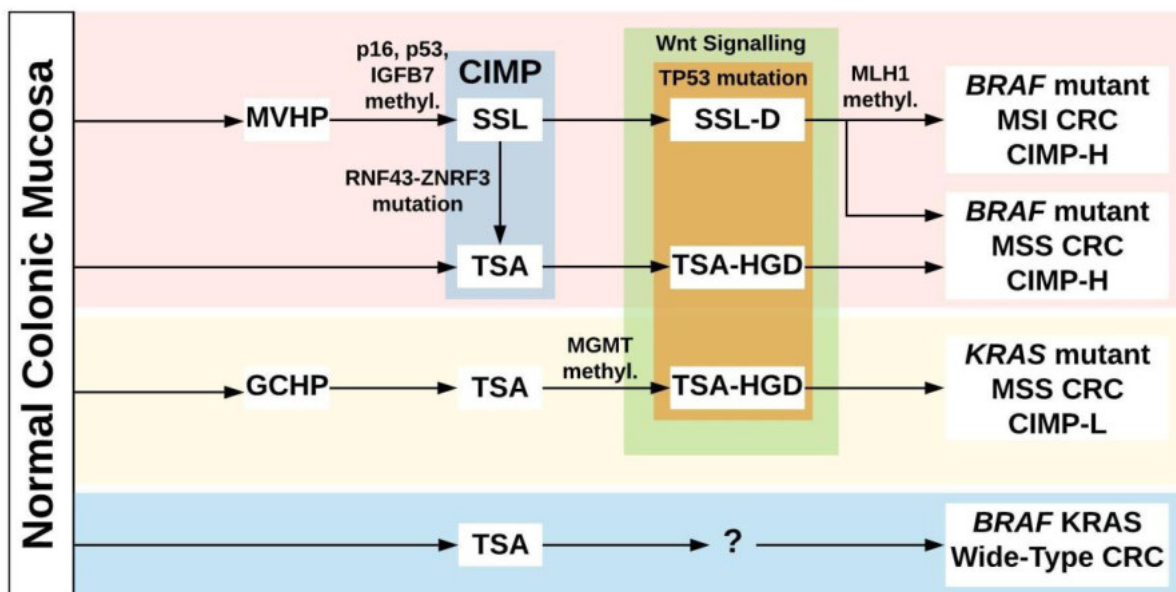
Based on the 5<sup>th</sup> edition of the World Health Organisation (WHO) classification of tumours of the digestive system, serrated colorectal lesions are classified into traditional serrated adenomas (TSAs), sessile serrated lesions (SSLs), hyperplastic polyps (HPs) and unclassified serrated adenomas (Nagtegaal et al. 2020). Two variants of HPs are microvesicular (MVHP) and goblet cell (GCHP) (Kim, Kang 2020). The prevalence of serrated class polyps is 20-40% in average-risk individuals; approximately 75% of serrated polyps detected are HPs (Crockett, Nagtegaal 2019). SSLs are the most common premalignant serrated subtype and are found in up to 15% of average-risk patients (**Table 1**). Approximately 70-80% of TSAs are located in the distal colon (Gui et al. 2020; Chetty 2016; Bettington et al. 2015). Based on clinical histopathology, TSA has three variants, including filiform, flat and mucin-rich (Childs et al. 2014; Bettington et al. 2015), and are often admixed with HP or SSL (Tsai et al. 2014; Kim et al. 2013; Bettington et al. 2015).

**Table 1.** Clinical features of common serrated polyps.

Histological Classification	CRC Frequency (%)	Serrated polyps Frequency	Location	Size	Malignant Progression
Hyperplastic Polyps (HPs)*	20%-30%	~75%	70~80% Distal	<5mm	No
Sessile Serrated Lesions <sup>^</sup> (SSLs)	5%-15%	~25%	75%-90% Proximal	5mm-7mm	Yes
Traditional serrated polyp (TSAs) <sup>#</sup>	<1%	1%-7%	Mostly Distal	Usually > 5mm	Yes

\*Hyperplastic polyp includes microvesicular hyperplastic polyp (MVHP) and goblet cell hyperplastic polyp (GCHP). <sup>^</sup>Previously version called sessile serrated adenomas/polyps. <sup>#</sup>Three variants of TSA are filiform, flat and mucin-rich. Adopted from (Crockett, Nagtegaal 2019).

In most serrated polyps, the first critical step is the acquisition of a gene mutation that regulates the mitogen-activated protein kinase (MAPK) pathway and DNA methylation (Pino, Chung 2010). The activation of MAPK signalling induces cell proliferation from normal epithelium cells to serrated aberrant hyperplastic crypt foci, which are considered at the earliest histological lesions (Rosenberg et al. 2007). TSA has been considered a genetically heterogeneous polyp and mainly recognised as a precursor lesion for MSS or MSI<sup>Low</sup> (Leggett, Whitehall 2010; Saito et al. 2015), and can be CIMP<sup>High</sup> or CIMP<sup>Low</sup> (Rex et al. 2012; Bettington et al. 2013; Tsai, Cheng, et al. 2016). It has been hypothesised that GCHP is the precursor to TSA, although no clear evidence has been put forward (Yang et al. 2004). Pooled data showed that *KRAS* or *BRAF* mutations are present in the majority of TSAs, with an incidence of approximately 32%, and 56%, respectively (Gui et al. 2020). Based on the molecular features, TSA could broadly be divided into *KRAS*- or *BRAF*-driven pathways, while *KRAS* and *BRAF* wide type TSAs are still unknown (McCarthy et al. 2019) (Figure 6).



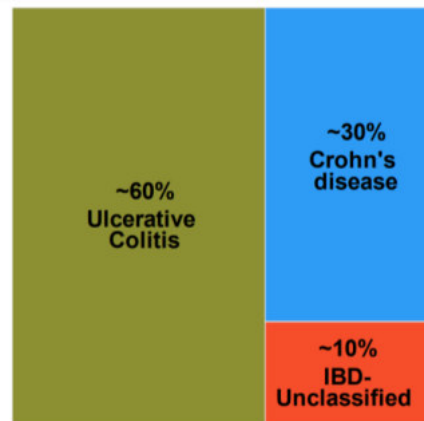
**Figure 6. Serrated colorectal cancer pathways.** Serrated colorectal cancer (CRC) can be broadly divided by BRAF and KRAS molecular pathways. In the sessile serrated pathway, normal mucosa transformation begins with BRAF mutations, potentially from MVHP, followed by p16 and IGFBP7 promoter hyper-methylation toward serrated adenocarcinoma, mainly through MLH1 epigenetic alterations. In contrast, the traditional serrated pathway (TSA) involves *KRAS* or *BRAF* mutations in normal colon cells, following *MGMT* methylation towards high-grade dysplasia and serrated adenocarcinoma (SAC). RNF43-ZNRF3 mutation may be the key regulator by transforming from SSL to TSA. CIMP, CpG island methylator phenotype; SSL, Sessile Serrated Lesions; KRAS, KRAS proto-oncogene GTPase; BRAF, B-Raf proto-oncogene serine/threonine kinase; APC, adenomatous polyposis; TP53, tumour protein 53; TGFBR, transforming growth factor-β receptor; BAX, BCL2 associated X apoptosis regulator; IGF2R, insulin-like growth factor 2 receptor Adopted and modified from (McCarthy et al. 2019).

*BRAF* mutation (especially V600E mutation sites) accounts for approximately 10% of CRC cases (Kambara et al. 2004). It is relatively less in conventional adenomas/Lynch syndrome CRC but common in SSLs/TSAs CRC tumours (Kakar et al. 2008; Spring et al. 2006; Leggett, Whitehall 2010). Most of the *BRAF*-mutated TSA-induced tumours are serrated with MVHP or SSL precursor at proximal colon (Kim et al. 2010; Lash et al. 2010; Sekine et al. 2020). It has been hypothesised that *MLH1* hyper-methylation (especially 93G>A polymorphism) leads the tumour progression from *BRAF*-mutated SSL, SSL with dysplasia (SSL-D) to MSI<sup>High</sup> CRC sequence (Nourbakhsh, Minoo 2020). CIMP<sup>High</sup> with *MLH1* hyper-methylation is also one of the critical factors from SSL to SSL-D or even serrated CRC transformation (Nourbakhsh, Minoo 2020). Due to the high mutation rate and oncogenic burden, SSL-D is speedily transformed into CRC (Amemori et al. 2019). Around three-quarters of SSL-D cases are loss of *MLH1* gene expression (Fennell et al. 2018); and *BRAF*-mutated SSL without *MLH1* methylation could be transformed into TSA and become MSS (Yan et al. 2017; Nourbakhsh, Minoo 2020). *BRAF* mutation also leads to a high frequency of *RNF43* mutations, up-regulation of p16<sup>INK4a</sup> and induces *IGFBP7* secretion at aberrant crypt foci (Hashimoto et al. 2019; Sekine et al. 2017; Tsai, Liau, et al. 2016). Moreover, *BRAF* mutation might promote global CpG island methylation by increasing the promoter binding of the transcriptional repressor gene *MAFG* (Fang et al. 2014; Fang et al. 2016). A *BRAF* mutation-induced oncogenesis in mouse colon-derived organoid focuses on ageing-related hyper-methylation, which suggests that *BRAF* mutation is not a prerequisite for CIMP development in SSLs (Tao et al. 2019).

*KRAS* is a membrane-bound GTP/GDP-binding protein (Ogino et al. 2009) and its gene mutations are common in codon 12 (G12D, G12V) and codon 13 (G13D). *KRAS* mutation appears in the early stages of CRC tumorigenesis (Fearon 2011; Fernández-Medarde, Santos 2011; Kosmidou et al. 2014) and impairs intrinsic GTPase activity from active GDP-binding stage to GTP-binding stage, resulting in *KRAS* accumulation and downstream proliferative signalling pathways (Schubbert et al. 2007). *KRAS* mutation with CIMP<sup>Low</sup> and MSS TSAs is frequently located in the distal colon (descending and sigmoid colon) and rectum. The epigenetic silencing of *MGMT* instead of *MLH1* induces the mutations' accumulation and may lead to TSA formation (Huang et al. 2011; Whitehall et al. 2001). *KRAS*-mutated TSA has a higher frequency with *PTPRK-RSPO3* gene fusions (RSPO fusions), resulting in down-regulation of *RNF43* and R-spondin over-expression (de Lau et al. 2014; McCarthy et al. 2019).

## ii. COLITIS-ASSOCIATED COLORECTAL CANCER

Inflammatory bowel disease (IBD) is a group of idiopathic and relapsing-remitting chronic inflammatory disorders defined by a wide range of criteria, based on clinical, endoscopic, laboratory, radiological and histological outcomes. Crohn's disease (CD) and ulcerative colitis (UC) are the two major subtypes of IBD, accounting for approximately 30% and 60% respectively, and up to 10% of IBD patients are grouped as IBD unclassified (IBDU) due to lacking specific UC or CD features (Thurgate et al. 2019) (**Figure 7**).



**Figure 7. Distribution of inflammatory bowel disease subtypes.**

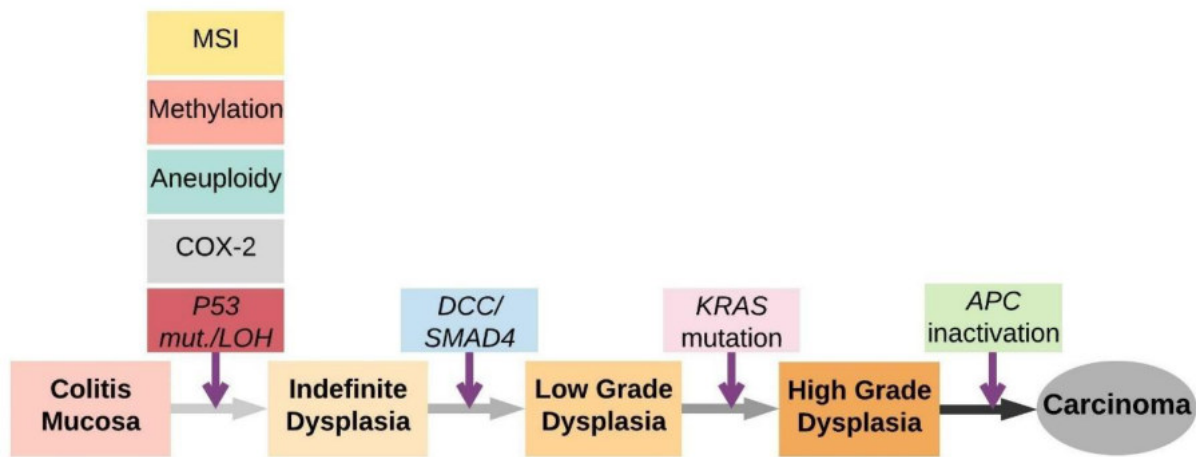
Chronic inflammation appeared in one-fifth of all types of cancers, and CRC has the highest correlation between tumorigenesis and long-standing inflammation (De Marzo et al. 2007). This long-standing chronic inflammation is characterised by a susceptible genetic background, underlying immunological deregulation and dysbiosis, inducing intestinal barrier injury, permeability and destruction of tight junctions, leading to intestinal mucosa damage (Turner 2006; Abraham, Cho 2009). The degree of colonic inflammation and the duration of the disease is related to the development of neoplasia, with approximately 300-500% higher risk of CRC (Choi et al. 2019; Samadder et al. 2019). Studies reported that UC-associated CRC has a more unfavourable survival rate than sporadic CRC, accounting for one-sixth of UC-related deaths (Jensen et al. 2006; Watanabe et al. 2011; Jess et al. 2006).

### (1). Pathogenesis of colitis-driven colorectal cancer

In terms of clinical histopathology, CAC tissues are more often likely to have a background of chronic inflammation, a substantial portion of mucinous, and a higher number of signet ring cells and multifocal dysplasia (Kulaylat, Dayton 2010; Ullman, Itzkowitz 2011). More frequently, CAC develops initially from non-polypoid and flat dysplasia in the area of abnormal epithelial hyperplasia and progresses to invasive adenocarcinoma (Xie, Itzkowitz 2008; Jensen et al. 2006; Watanabe et al. 2011). At the molecular level, CIN and MSI incidences in CAC are similar to sporadic CRC, while the occurrence time and frequency differ

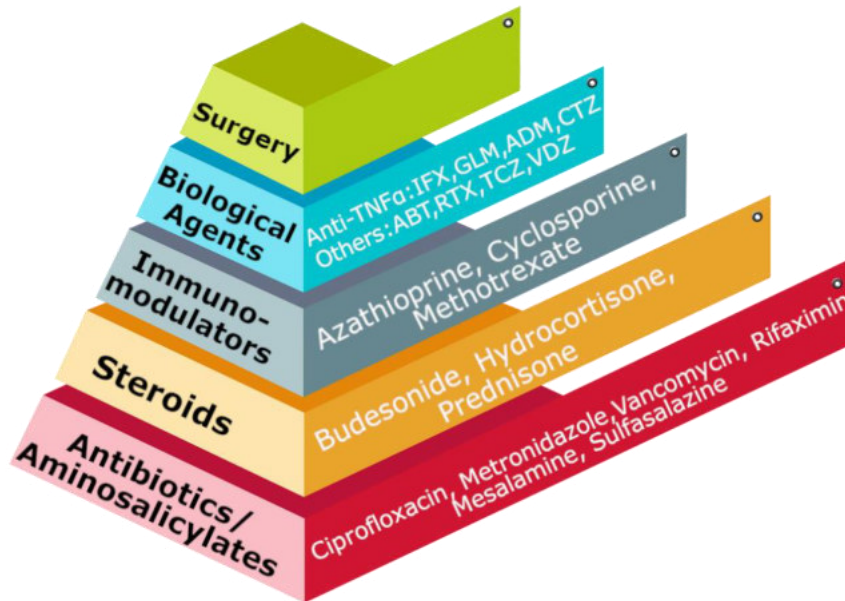
(Lakatos, Lakatos 2008; Triantafillidis et al. 2009; Ullman, Itzkowitz 2011). CAC has a lower *APC* dysfunction rate and appears at the late stage of carcinogenesis; the loss of *TP53* function occurs at the early stage of carcinogenesis compared to sporadic (Dyson, Rutter 2012; Pilley et al. 2021; Bieging et al. 2014) (**Figure 8**). Approximately 50% - 85% of CACs have *TP53* mutations and it seems more likely for neoplasia initiation in IBD patients (Du et al. 2017).

### Colitis-associated colorectal cancer



**Figure 8. Colitis-associated colorectal cancer in classical pathways.** MSI, microsatellite instability; COX-2, Prostaglandin-endoperoxide synthase 2; p53, Tumour protein P53; LOH, loss of heterozygosity; DCC, Deleted in Colorectal Cancer; SMAD4, SMAD Family Member 4; KRAS, KRAS Proto-Oncogene, GTPase; APC, Adenomatous Polyposis Coli. Adopted and modified from (Kobayashi et al. 2017).

Once individuals have been diagnosed with IBD, drugs treatment may be given depending on the disease type, severity, responsiveness to treatment and potential side effects, in order to obtain remission and mucosal healing, thereby lowering the need for surgery, disease progression and potential CRC development. Different classes of approved pharmacological drugs have diverse anti-inflammatory mechanisms of action and the potential side effects for IBD patients (McQuaid 2018) (**Figure 9**).



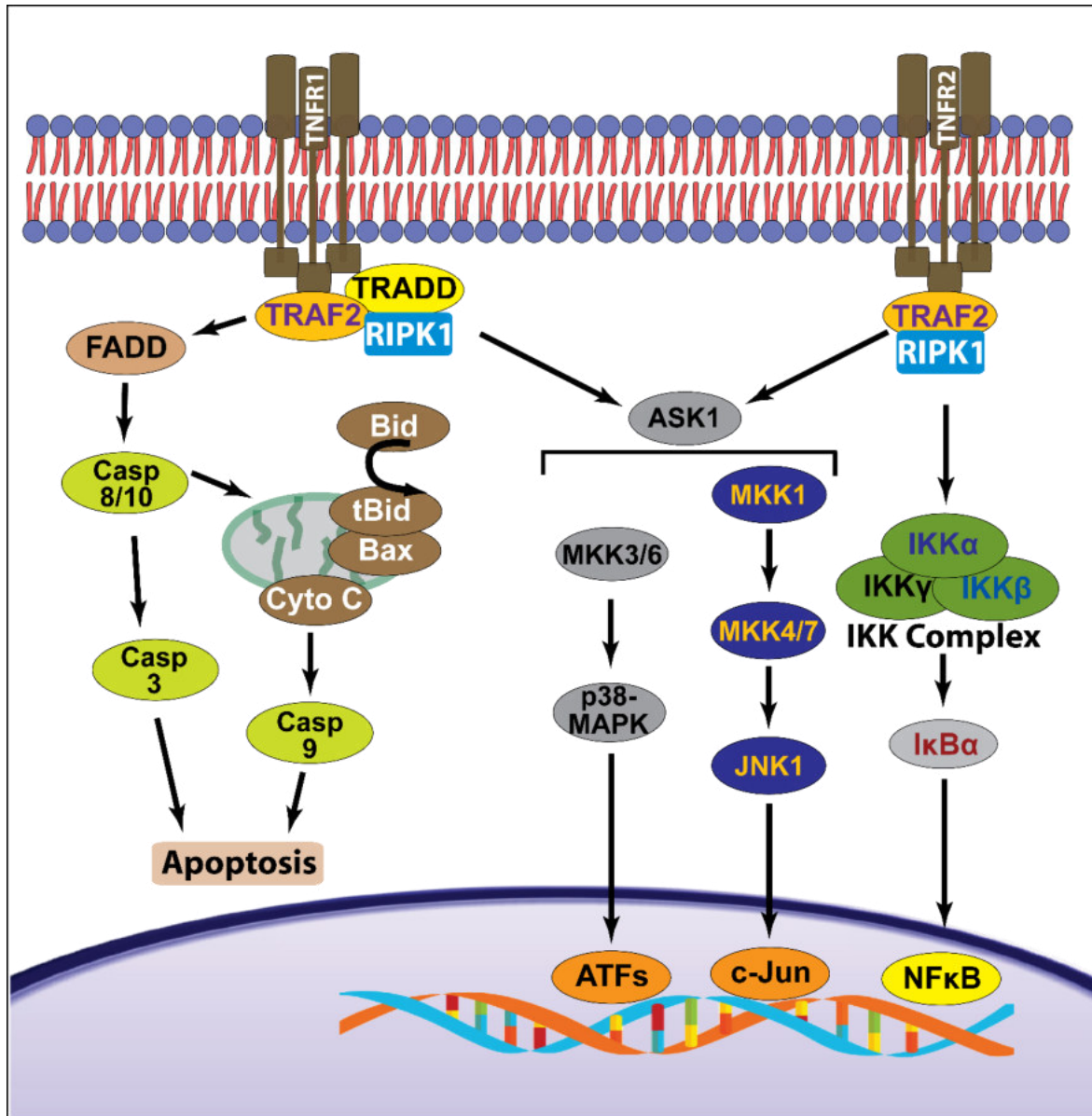
**Figure 9. Therapeutic pyramid for the step-up treatment of inflammatory bowel disease.** Step-up approach: from mild to stronger and more toxic therapies. ABT, Abatacept; RTX, Rituximab; TCZ, Tocilizumab; VDZ, vedolizumab; IFX, Infliximab; ADM, Adalimumab; CTZ, Certolizumab Pegol; GLM, Golimumab.

## (2). The role of tumour necrosis factor $\alpha$ in colitis-associated colorectal cancer

Cytokines are small proteins (< 40 kDa) secreted in response to infection and inflammation, influencing disease progression including IBD, CAC and CRC (Kany et al. 2019). It is believed that IL6, IL1 $\beta$  and TNF $\alpha$  are the key cytokines in IBD (Ahluwalia et al. 2018). Among them, therapeutic agents targeting TNF $\alpha$  have altered the treatment strategy in IBD patients with moderate-to-severe disease (D’Haens, van Deventer 2021). TNF $\alpha$  is induced by mononuclear cells through activation of cellular receptors and pattern recognition receptors, such as toll-like receptor 4 (TLR4) (Sansone, Medzhitov 2009).

The biologically active homotrimer TNF $\alpha$  originally derived from the precursor transmembrane TNF $\alpha$  (tmTNF- $\alpha$ , 26 kDa), and is proteolysed by TNF-converting enzyme (TACE) to form soluble TNF $\alpha$  (17 kDa) (Kalliolias, Ivashkiv 2016). The canonical TNF $\alpha$  pathway is activity-mediated via binding to the TNF receptors I and II (TNFR1 and TNFR2), and further activates MAPKs and NF- $\kappa$ B pathways, inducing cell proliferation, differentiation and up-regulation of several pro-inflammatory cytokines (McDaniel et al. 2016; Yeung et al. 2018; Ruiz et al. 2021). To be specific, TNF $\alpha$  binds to TNFR1, inducing the recruitment of death domain (DD) incorporated with TNFR-associated DD (TRADD) protein (Hsu et al. 1995) and, further, the recruitment of fas-associated protein with death domain (FADD), activating caspase-8, -10 and -3, (Hsu et al. 1996). TNF $\alpha$  can also trigger mitochondria signalling, release cytochrome C and Bax, thereby activating caspases, leading to apoptosis (Vringer, Tait 2019). TNFR2 contains a TRAF-interacting (TIM) motif domain and enables binding

TRAF2 via the activation of TNF $\alpha$ . This pathway also leads to association with multiple signalling pathways, including ERK (extracellular signal-related kinase), JNK (c-Jun N-terminal kinase), pI3K (phosphoinositide 3-kinase) MAPK, NF- $\kappa$ B and AKT, and share a similar path to TNFR1 (Figure 10) (Wajant, Siegmund 2019).



**Figure 10. TNF $\alpha$  signalling pathway.** TNF $\alpha$  binds to the membrane receptors TNFR1 or TNFR2, leading to downstream apoptosis and inflammatory signalling. The activation of TNFR1 induces the formation of a death-inducing signalling complex such as TRADD, and further activates FADD and caspase, eventually inducing apoptosis. Pro-inflammatory signalling pathways can be activated via TNFR1 or TNFR2, then, through the adaptor protein TRAF2 to activate RIPK1, ASK1 or MEKs, resulting in the activation of MAPK and canonical NF- $\kappa$ B activation. TNFR1, tumour necrosis factor receptor 1; TNFR2, tumour necrosis factor receptor 2; TRADD, tumour necrosis factor receptor type 1-associated DEATH domain protein; FADD, Fas-associated protein with death domain; TRAF2, TNF receptor-associated factor 2; RIPK1, Receptor-interacting serine/threonine-protein kinase 1; ASK1, Apoptosis signal-regulating kinase 1; MEK, Mitogen-Activated Protein Kinase Kinase; NF- $\kappa$ B, (nuclear factor-kappa B); IKK, I $\kappa$ B kinase; BID, BH3 interacting-domain death agonist; BAX, Bcl-2-associated X protein; Cyto C, Cytochrome c; Casp, Caspases; JNK1, Jun N-terminal Kinase 1. Adopted from (Wajant, Siegmund 2019).

### **III. microRNA**

MicroRNAs (miRNA) belong to a class of highly-conserved short single-stranded non-coding RNA segments (18-24 nucleotides), which regulate gene expression at the transcriptional level after their nuclear hybridisation and reimportation (Place et al. 2008). miRNAs are commonly located within introns of host genes and potentially have a co-regulation of transcription which has an independent transcriptional start site and regulatory elements. It can also be controlled by a different promoter and/or other regulatory sequences. Thus, miRNA expression does not necessarily correlate with the levels of their host gene (Bartel 2004).

#### **Biogenesis of miRNA has both canonical and non-canonical pathways**

##### **i. CANONICAL PATHWAY OF microRNA BIOGENESIS**

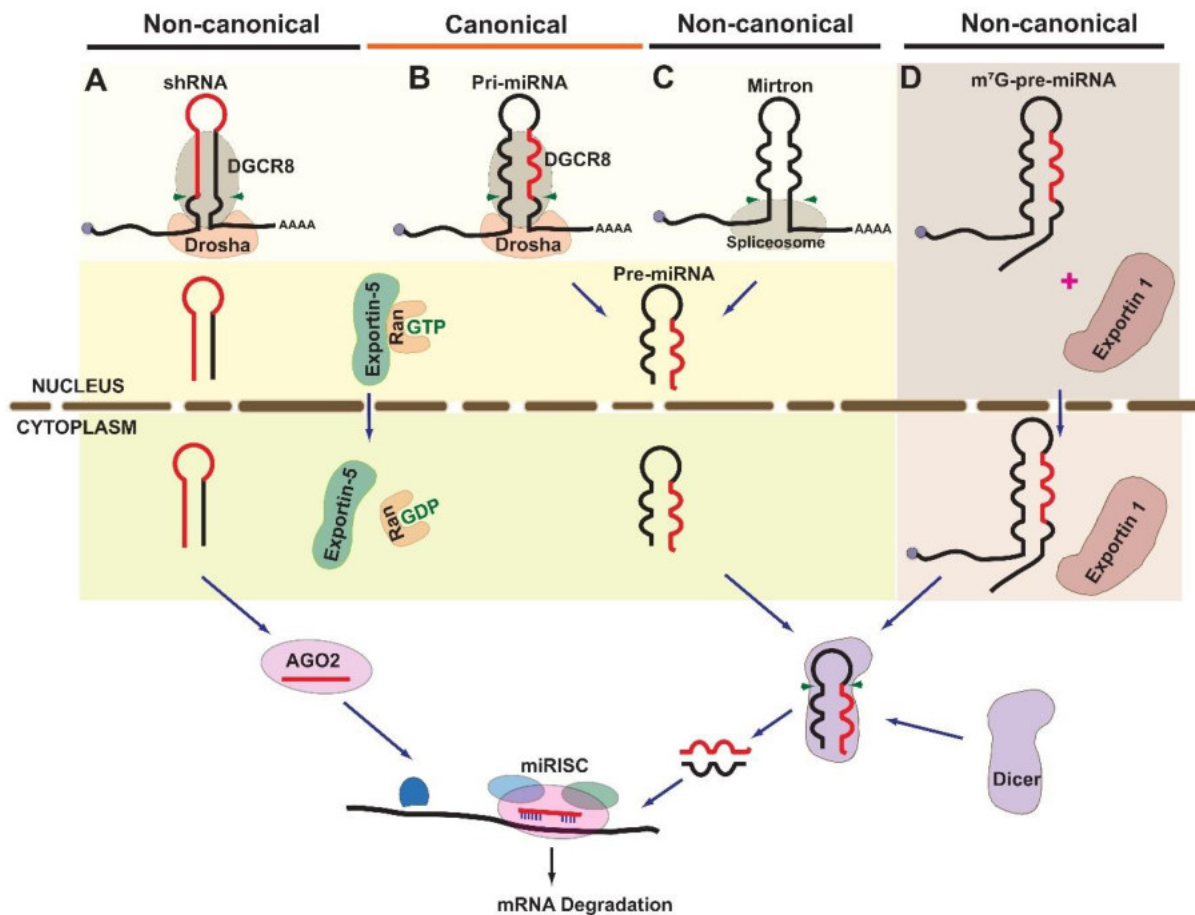
The canonical miRNA pathway is the dominant approach for miRNA biogenesis (**Figure 11B**). A mature miRNA is first transcribed by RNA polymerase II (POL II) as a primary miRNA (pri-miRNA) (>1 kb). It contains a 5'-cap with a poly-A tail (Lee et al. 2004) and can be recognised by a DiGeorge Syndrome Critical Region 8 (DGCR8) - an RNase III enzyme Drosha complex in the nucleus (Denli et al. 2004; Han 2004). The Drosha cleaves the pri-miRNA into precursor miRNA (pre-miRNA, an approximately 70 nucleotides stem-loop precursor) by recognising the N6-methyladenylated GGAC and other motif(s) such as CNNC and UG motifs (Lee et al. 2003). The pri-miRNA to pre-miRNA process needs a ~35 nucleotides hairpin stem or a 3–23 nucleotides apical loop (Roden et al. 2017; Adams 2017). Then, the pre-miRNA is exported to the cytoplasm via the small Ran GTPase and exportin 5 (XPO-5) complex (Bohnsack et al. 2004; Yi et al. 2003) and is therefore cleaved by the endonuclease DICER/trans-activation-responsive RNA binding protein (TRBP) into a duplex miRNA (Hutvagner et al. 2001; Bartel 2018). This duplex is bound by the Argonaute (AGO) in an RNA-induced silencing complex (RISC), which eliminates one strand of the duplex to conserve the guide strand corresponding to the mature miRNA (Kobayashi, Tomari 2016). This mature miRNA at the 5' region (corresponding 2 to 7 nucleotides) enables recognition and binding of the mRNA target transcript(s), mostly in their 3'-untranslated regions (3'UTRs). If the pairing is perfect, AGO produces an endonucleolytic cleavage leading to mRNA degradation; if mismatch(es) are present, the mRNA could first be deadenylated and then degraded or translational repression may occur (Jonas, Izaurralde 2015; Pasquinelli 2012). Rarely, miRNA could be localised in the open reading frame (ORF) or 5'UTR of mRNA (Bartel 2009; Helwak et al. 2013), or up-regulate the expression of target genes (Mortensen et al. 2011; Oldenburg et al. 2017; Vasudevan et al. 2007). In certain circumstances, miRNAs can



also control the translation activation and even regulate transcription rates (Vasudevan 2012), as they can shuttle between different subcellular compartments (Makarova et al. 2016).

## ii. NON-CANONICAL PATHWAYS OF microRNA BIOGENESIS

Non-canonical miRNA pathways are composed mainly of proteins involved in the canonical pathway by bypassing one or more steps, while Dicer is still required. It is broadly considered as (1) Drosha/DGCR8-independent and (2) Dicer-independent pathways (Abdelfattah et al. 2014). miRNA is processed by Drosha via Dicer-independent pathway typically from the endogenous short transcripts, such as short-hairpin RNA (shRNA) and a short length of pre-miRNA product (Yang et al. 2010). For instance, the maturation process of miR-451 requires AGO2 within the cytoplasm, due to the relatively short length of the pre-miR-451 stem-loop structure to be a Dicer substrate (Cheloufi et al. 2010; Yang et al. 2010) (**Figure 11A**). On the Drosha/DGCR8-independent pathway (**Figure 11C**), mirtron- (a class of miRNAs produced by spliced or debranched mRNA) derived pri-miRNA has a stable hairpin with a shorter stem-loop, which is suitable for Dicer cleavage and enables bypassing the microprocessor (Berezikov et al. 2007; Westholm, Lai 2011). Another example is 7-methylguanosine ( $m^7G$ )-capped pre-miRNA (**Figure 11D**), which is transferred to the cytoplasm via exportin 1 without Drosha cleavage. Noted that the end miRNA product is mainly from the 3p strand, this is due to the  $m^7G$  cap preventing 5p strand loading into an AGO (Xie et al. 2013).



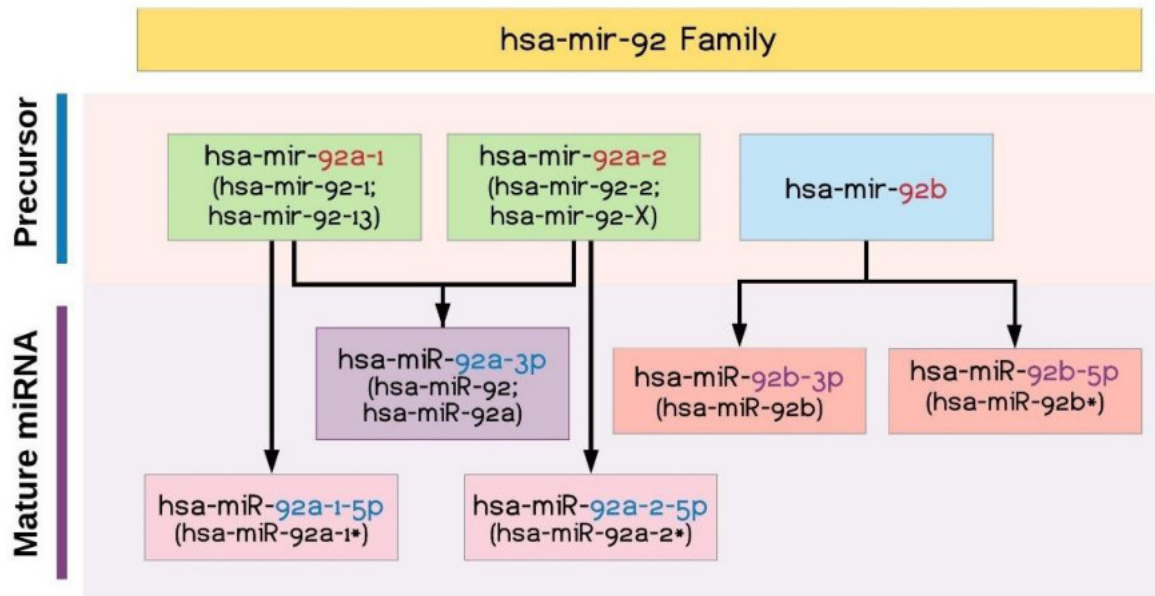
**Figure 11. Canonical and non-canonical pathways of microRNA biogenesis.** (B) The canonical miRNA biogenesis starts with RNA polymerase II-driven transcription and produces the primary miRNA (pri-miRNA) in the nucleus, with the sequence ends polyadenylated and capped. Then, the DGCR8 and Drosha microprocessor complex cleaves the pri-miRNA into pre-miRNA – a short hairpin-like structure with 60-90 nucleotides – and delivers it to the cytoplasm by Ran/exportin-5 in a GTP-dependent approach. The pre-miRNA will further be modified by Dicer – an RNase III nuclease – to form a double-stranded duplex miRNA. The duplex miRNA will be bound by the Argonaute proteins (AGO2) in an RNA-induced silencing complex (RISC) for a strand selection process. Within the RISC, the mature single-strand miRNA at the 5' region (corresponding 2 to 8 nucleotides) enables recognition and binding of the complementary mRNA target transcript(s), mostly in their 3'-untranslated regions (3'UTRs). In non-canonical pathways, (A) small hairpin RNA (shRNA) is initially cleaved by the DGCR8 and Drosha microprocessor complex and exported to the cytoplasm via Exportin5/RanGTP, and further processed via AGO2 directly without Dicer; (C) Mirtrons and (D) 7-methylguanine capped (m7G)-pre-miRNA require Exportin5/RanGTP and Exportin1 respectively, to export the nucleus; both require Dicer to complete the maturation in the cytoplasm. Moreover, mirtrons require spliceosome sizing from a pre-miRNA. All pathways ultimately lead to a functional miRISC complex like canonical miRNA biogenesis. This interaction then triggers the repression of translation and degradation of the target mRNA. Adopted and modified from (O'Brien et al. 2018).

### iii. NOMENCLATURE OF microRNA

Based on the miRNA naming system, a typical miRNA name is composed of a minimum of two components (e.g., miR-221 or miR-21) and applied in only one specific species study in a research article. For studies involving more than one species, a three-letter specific prefix is required to separate the origin of miRNAs. For instance, "hsa" refers to human (*Homo sapiens*), "rno" refers to common rat (*Rattus norvegicus*) and "ebv" refers to Epstein-Barr virus (e.g. hsa-miR-221, rno-miR-221 and ebv-miR-BART3) (AMBROS 2003; Meyers et al. 2008; Liang et al. 2014).

The *majuscule* "miR-" refers to the mature form of the miRNA and *minuscule* "mir-" refers to the corresponding pre- or pri-miRNA stem-loop form. A suffix number often suggests the order of naming, so that miR-21 is likely discovered before miR-221. A *minuscule* letter after the number is used to distinguish among multiple members of the same family (e.g., miR-20a and miR-20b) and normally starts from the letter "a". 3p or 5p names after a mature miRNA sequence indicate a position from the pre-miRNA hairpin cleaved by the RNase III enzyme Dicer: the -3p strand is positioned in the reverse (3'-5') position while the -5p located in the forward (5'-3') position (e.g., miR-21-5p from the 5' arm of the precursor and miR-21-3p from the 3' arm of the precursor). If two diverse loci produce identical mature products, an additional number will be given after the name. For instance, hsa-mir-92a-1 (from chromosome 13) and hsa-mir-92a-2 (from chromosome X) form the same final -3p miRNA product (hsa-miR-92a-3p) while each locus can generate two separate -5p (hsa-miR-92a-1-5p and hsa-miR-92a-2-5p) end mature products (**Figure 12**) (Kozomara et al. 2019; Kozomara, Griffiths-Jones 2014; Kozomara, Griffiths-Jones 2011; Griffiths-Jones et al. 2007; Griffiths-Jones 2006; Griffiths-Jones 2004).

Additionally, a miRNA with \* (star symbol, an abandoned identification system) indicates the mature miRNA from the same hairpin precursor, but less predominant from the opposite arm (e.g., miR-21\* is the former name of miR-21-3p). Noted that the name of a miRNA may be changed or abandoned (e.g., hsa-mir-923, false discovery due to the fragment originally from 28S ribosomal RNA (rRNA)) (miRBase 2020). The latest miRNA identifiers can be found from the miRbase database (<http://www.mirbase.org/>) for published (and dead entry) miRNA (Kozomara et al. 2019; Kozomara, Griffiths-Jones 2014; Kozomara, Griffiths-Jones 2011; Griffiths-Jones et al. 2007; Griffiths-Jones 2006; Griffiths-Jones 2004).



**Figure 12. The hsa-miR-92 Family and its nomenclature.** Based on the miRbase database dated 10-03-2021. Brackets indicate the previous miRNA identifier(s).

#### iv. PROBE-BASED REAL-TIME POLYMERASE CHAIN REACTION

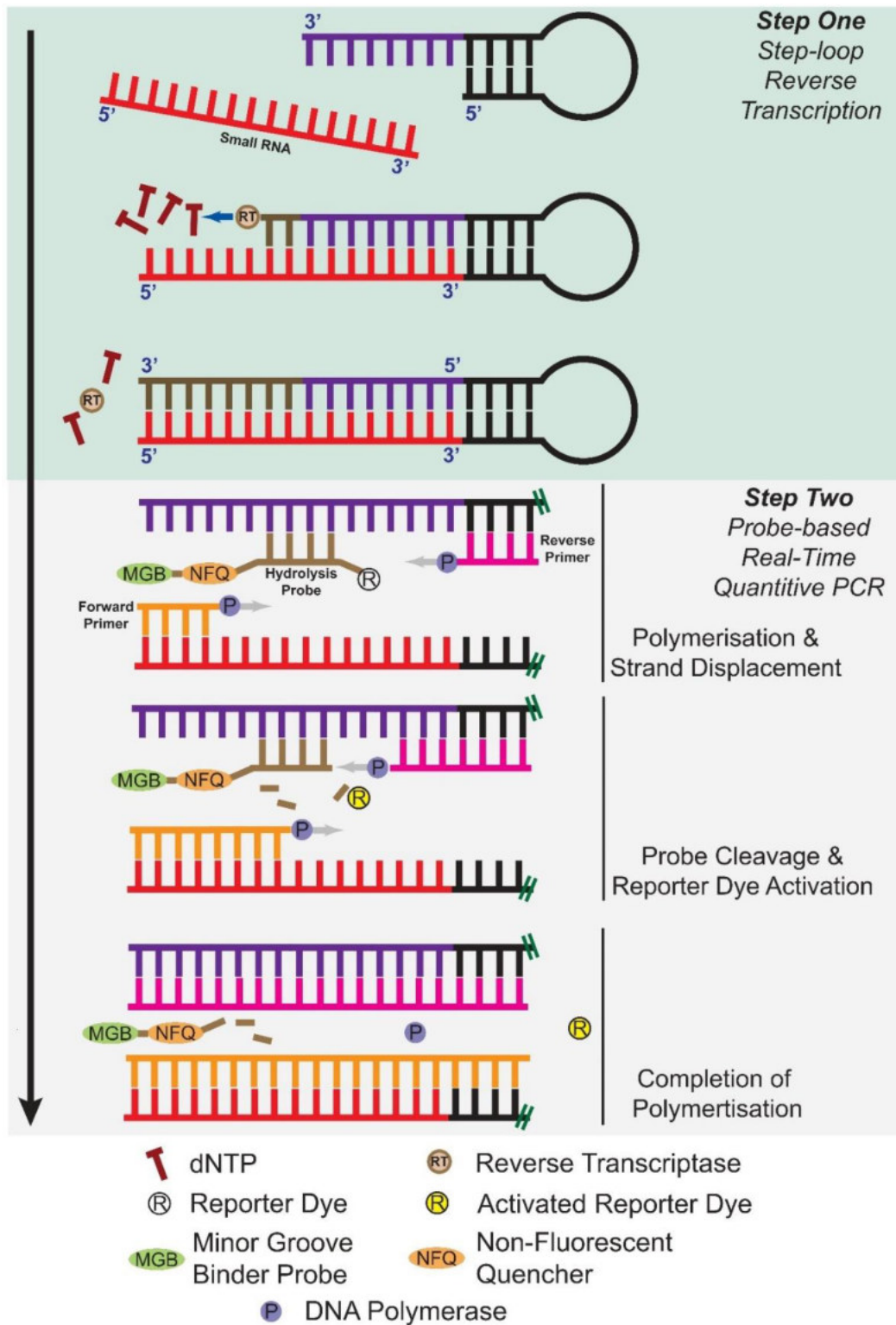
SYBR® Green is the most commonly used reagent for Real-Time quantitative polymerase chain reaction (qPCR). In traditional real-time qPCR, it only requires a pair of 18~22 based-pair primers targeting the specific target sequence of interest. However, SYBR Green is able to bind to any double-stranded DNA, including non-specific amplification products. Thus, using a specific complement sequence probe with a pair of primers can minimise the chance of RT-qPCR non-specific amplification product(s). This approach can also increase the detection specificity and prefer to use it in a complex environment, such as DNA/RNA samples from faeces. Noted that the probe-based RT-qPCR approach has been extensively used in SARS-COV-2 screening during the COVID-19 pandemic (Bustin, Nolan 2020; Vogels et al. 2020). A hydrolysis fluorescent-labelling probe enhances the detection of specificity and adjusts the melting temperature ( $T_m$ ). A probe conjugate with a minor groove binder (MGB) can minimise the length of its oligonucleotide to approach the target sequence as well as stabilising the quantification of PCR products via a reporter dye (**Table 2**) (Chen 2005; Afonina et al. 2002). The use of probe-based qPCR can be utilised in multiple targets in one single reaction (Liang et al. 2019). It is based on the use of different probes by selecting different wavelength reporters and the corresponding relative intensities. Probe design, fluorescent-label selection, and the number of RT-qPCR targets in one single reaction are varied and rely on the qPCR instrument itself. FAM is the most common qPCR probe while other types, such as VIC™ and NFQ™ are available (Mark D. et al. 2005).

**Table 2. Comparison between TaqMan<sup>®</sup> probe-based and SYBR<sup>®</sup> Green for Real-Time quantitative polymerase chain reaction.**

TaqMan <sup>®</sup> Probe-based	SYBR <sup>®</sup> Green
<b>Chemistry</b>	
Fluorogenic probe(s) in detection of target sequence(s)	Bind to double-stranded DNA
<b>Advantage / Disadvantage</b>	
High specificity	High sensitivity
Low signal to background	High signal to background
Require fluorescent-labelled probe(s)	No fluorescent-labelled probe
Relatively High cost	Relatively low cost
Available for multiplexing	Only one single target per reaction
Accurate quantitation	May have non-specific amplification(s)

TaqMan<sup>®</sup> is a registered trademark of Roche Molecular Systems, Inc.; SYBR<sup>®</sup> is a registered trademark of Molecular Probes, Inc.

Compared to a typical protein-coding gene expression qPCR assay, small RNAs, including mature miRNAs, are too short to accommodate for “traditional” PCR amplification. In order to overcome these issues, using a one-strand cDNA synthesis with a highly stable miRNA-specific stem-loop structure primer can lengthen the original target small RNA based on the 6 base-pair binding regions and forms the 3' end RT primer/mature miRNA chimaera for the specific mature miRNA (**Figure 13, Step one**). These extended cDNA products also optimised for the *T<sub>m</sub>* for a standard TaqMan<sup>®</sup> probe-based Real-Time PCR (**Figure 13, Step two**). To apply this stem-loop primer structure for SYBR<sup>®</sup> Green qPCR, an 11 base pair (bp) extended binding region was designed to bring down the cost per single reaction (Tong et al. 2015).



**Figure 13. microRNA detection and quantification by TaqMan<sup>®</sup> probe-based RT-qPCR. (Step 1)** miRNA templates are reverse transcribed using a stem-loop specific primer. **(Step 2)** The reverse transcribed product is amplified using the miRNA specific primers and the hydrolysable probes. The process of hydrolysis displaces the probes and Taq polymerase, resulting in the fluorophore's separation and quencher. Accumulation of reporter dye (the fluorescence signal intensity) can determine the amplification efficiency based on a cut-off cycle. microRNA, miRNA.

## IV. MICROBIOTA

There are trillions of microbes colonised in the human intestinal flora, of which there are about 100–1,000 bacterial species that interact closely with host cells, affecting the immune and metabolome in the gastrointestinal tract, and shaping the homeostasis of the intestine (Qin et al. 2010). Dysbiosis of gut microbiota has been linked to various GI diseases, including IBD, CRC and *Clostridioides difficile* (also formerly called *Clostridium difficile*, *C. difficile*) infection (CDI) (Lee et al. 2016). In the stomach, the presence of *Helicobacter pylori* (*H. pylori*) and/or Epstein-Barr virus (EBV) have been reported in association with the induction of GC (Yau, Tang, et al. 2014). Recently, an increasing number of studies indicated that miRNAs from the host, or microbes such as EBV, play an important role in influencing the stomach microenvironment to maintain intestinal homeostasis and/or disease(s) progression.

### i. EPSTEIN-BARR VIRUS-INDUCED GASTRIC CANCER

*Helicobacter pylori* (*H. pylori*) and EBV are the oncogenic microbes that induce GC; among them, EBV-associated GC (EBVaGC) accounts for nearly 10% of GC cases (Tavakoli et al. 2020). Compared to the other GCs, EBVaGC is recognised for its unique clinicopathological characteristics (Chen et al. 2015). EBVaGC is a latency I EBV infection, in which *EBER*, *EBNA1* and *LMPs* viral genes are expressed in the host cells, affecting the host cells' molecular mechanisms, including their epigenetics. The mechanism of EBV infection and the epigenetic changes in EBVaGC have been previously discussed (**publication 4, page 119**) (Yau, Tang, et al. 2014). For example, EBV latent membrane protein 2A (LMP2A) induces DNMT1 transcription via the phosphorylation of STAT3 (Hino et al. 2009). Virally-encoded miRNAs such as ebv-miR-BART1-5p, ebv-miR-BART3, and ebv-miR-BART6 target host DICER1 to regulate host miRNA expression (Wang et al. 2017). Host cell DNA methylation can be activated to suppress tumour suppressor gene expression by EBV infection in EBVaGC (Wang et al. 2017).

#### (1). Identification of Epstein–Barr virus-associated gastric cancer

The detection of EBV genes in host GC cells was first described by using PCR (Burke et al. 1990) while *in situ* hybridisation (ISH) targeting EBV-encoded small RNA (EBER-ISH) is considered the gold standard for EBVaGC identification (Fukayama et al. 2020). This identification can also be applied to genome sequencing (Camargo et al. 2016). Serology-based EBV detection, such as enzyme-linked immunosorbent assay (ELISA) and immunofluorescence against one or more viral antigens such as nuclear EBNA, capsid VCA, and early EA-D/EA-R antigens, are available; however, it is less likely for GC patients (Koh et al. 2019).

PCR amplification for EBV genes usually targets BamHI-W and/or Epstein–Barr nuclear antigen 1 (EBNA-1) (Shukla et al. 2011). In general, the outcome from PCR-based detection is commonly higher in EBVaGC (Rymbai et al. 2015; Shukla et al. 2011); and the prevalence of positive cases is about 80% compared with EBER-ISH (Shukla et al. 2012; Shukla et al. 2011). Recently, using droplet digital PCR (ddPCR), it has become possible for the EBV-DNA load from GC patients to be calculated according to the copy number of BamH1-W fragments (Shuto et al. 2019). However, due to heterogeneity and a high false-positive rate, PCR methods cannot determine the source of virus fragments. This is firstly because of viral DNA potentially from EBV-induced inflammation caused by infiltrating lymphocytes, instead of tumour cells; secondly, nine out of ten people have a latent stage of EBV infection from lymphocytes (Kim et al. 2009). These drawbacks also applied in serology-based EBV detection. The gold standard of EBER-ISH requires detection of both EBER-1 and EBER-2 viral genes from resected GC specimens fixed by formaldehyde and embedded in paraffin (FFPE), as positive cases have high copies of EBV ( $10^{6-7}$  per cell) in the nucleus (Tokunaga et al. 1993; Lee et al. 2009). However, there are shortcomings, including a complex experimental process, the expense and time-consumption. If poor fixation appeared, nucleic acids in the tissue may be diffused and/or denatured, reducing the binding efficiency and potentially appearing as false positive/negative results (Chen et al. 2015).

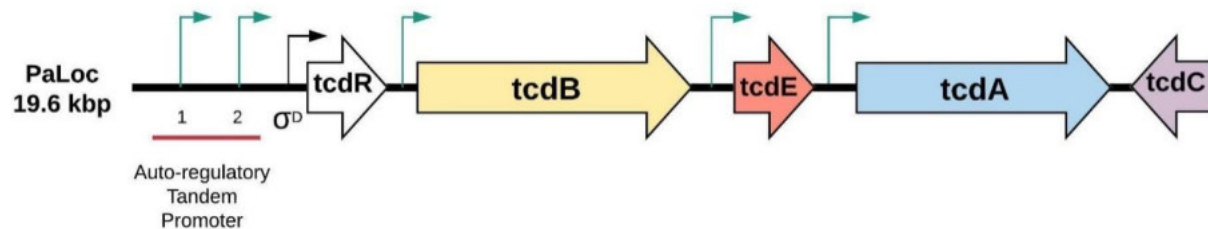
## ii. *CLOSTRIDIoidES DIFFICILE* INFECTION

*C. difficile* is an anaerobic gram-positive spore-forming bacillus and typically harmless in a balanced gut flora microenvironment. Once patients have been diagnosed with digestive diseases, gastrointestinal treatments, such as gastric-acid suppressing agents (e.g. H<sub>2</sub>-receptor antagonists and proton pump inhibitors), broad-spectrum antibiotics, chemotherapy and/or GI tract surgery may be required to overcome their diseases (Janarthanan et al. 2012; Roughead et al. 2016). Such treatments could suppress or interfere with the patients' immune system and disrupt gut flora homeostasis, creating a suitable micro-ecosystem for overgrowth of *C. difficile* (Rineh et al. 2014).

The life cycle of *C. difficile* includes spore formation, germination, and growth. At the spore stage, when *C. difficile* is in the dormant phase, it can resist oxygen, heat, and many other environmental insults, including ethanol-based disinfectants (Wilcox 2003). Once stable, spores germinate rapidly and produce two major toxins – toxin A and toxin B encode by *TcdA* and *TcdB* respectively, and locate at the 19.6 kbp long pathogenicity locus (PaLoc) region (**Figure 14**). The other coding genes (*tcdC*, *tcdE* and *tcdR*) are also located at PaLoc. Toxin A and toxin B trigger cytosol translocation of target host cells and inactive small GTP-binding



proteins (such as CDC42, Rho and Rac) through monoglucosylation, leading to actin condensation, disintegration of cytoskeleton, cell rounding and apoptosis (Voth, Ballard 2005). *tcdR* is an RNA polymerase sigma factor that initiates translation of *tcdA* and *tcdB*, activated by its two tandem promoters (Mani, Dupuy 2001; Mani et al. 2002).



**Figure 14. Pathogenicity loci of toxicogenic *Clostridioides difficile*.** The 19.6kbp long pathogenicity loci (PaLoc) in toxicogenic *Clostridioides difficile*, including *tcdC*, *tcdA*, *tcdE*, *tcdB*, and *tcdR* coding genes. Adopted and modified from (Isidro et al. 2017).

### (1). *Clostridioides difficile* infection and microRNA dysregulation

Study of the relationship between CDI and host miRNA is limited. To the best of my knowledge, the first reported study indicated the C57BL/6J wild-type mice with CDI induced mmu-miR-146b, mmu-miR-1940, and mmu-miR-1298 expressions, and up-regulated the pro-inflammatory cytokine expressions, such as MCP-1, IL-6, IL-17 and IL-1 $\beta$  in colonic tissues (Viladomiu et al. 2012). The miR-146b potential target *NCOA4*, *CD36* and *GLUT4* mRNA expression levels were down-regulated. *in silico* simulation predicts up-regulation of mmu-miR-146b and *IL-17* down-regulated *NCOA4* and *PPAR $\gamma$*  in mice after CDI. Null mice silencing *PPAR $\gamma$*  in T cells following CDI presents severe colonic disease activity, inflammatory lesions and inflammatory cytokine expressions (Viladomiu et al. 2012). A higher level of faecal hsa-miR-1246 was found in human CDI patients, compared to the control group in a small-scale study (Verdier et al. 2020).

Faecal microbiota transplantation (FMT) has been proven in the treatment of recurrent CDI (Drekonja et al. 2015; Hensley-McBain et al. 2016). It is intended to restore colonic microbiota through the introduction of "healthy" bacteria *via* colonoscopy, enema or oral capsules that contain powder-form substance. However, the safety concerns of FMT, such as volunteers who found SARS-COV-2 in their faeces, could be an obstacle to extending the application as a regular treatment strategy (Y. Xu et al. 2020; Alang, Kelly 2015). To monitor treatment conditions, detecting a panel of miRNA markers in circulations could help physicians make a clinical decision (Tanya M. Monaghan, Seekatz, et al. 2021). A study shows 71 circulating miRNAs were identified at 4 and 12 weeks following FMT from 126 sera from 42 patients at the screening. qRT-PCR and 3'UTR reporter assays validating the top miRNA

candidates showed that hsa-miR-23a-3p, hsa-miR-150-5p, hsa-miR-26b-5p and miR-28-5p expression levels target and inversely correlate with the sera protein and mRNA levels of IL-12B, IL-18, FGF21 and TNFRSF9 respectively. In a mouse model of relapsing-CDI, RT-qPCR analyses of caecal and sera RNA extracts showed inhibition of these miRNAs, and the inhibitory effect can be recovered by FMT. An additional study showed TcdB mediated the suppressive effects of CDI on miRs in human colonoids and mouse colon models. miR-23a and miR-150 demonstrated cytoprotective effects against TcdB (Tanya M. Monaghan, Seekatz, et al. 2021).

### iii. GUT MICROBES IN COLORECTAL CANCER

Gut microbiota imbalance (also called dysbiosis) and the corresponding microbial by-product changes have been recognised as playing crucial roles in both CRC and CAC pathogenesis. These alterations further affect the form of bacterial biofilm and the interaction of host immunity and inflammation (Keku et al. 2015; Flemer et al. 2017). The disruption of bacterial biofilm may lose the front line of defence protecting the host epithelial cells, inducing intestinal inflammation from direct bacteria attacks and genotoxic metabolites derived from the gut bacteria (Wu, Wu 2012; Belkaid, Hand 2014). It subsequently damages the mucous membrane, promotes colonic inflammation and neoplastic lesions (Wu, Wu 2012; Belkaid, Hand 2014). For instance, a number of *in vivo* studies demonstrated that faeces from human CRC patients transplanted into AOM-induced CRC mice and *Apc*<sup>min/+</sup> mice models further promote colon neoplasia, compared to the corresponding controls (Wong et al. 2017; Li et al. 2019). The composition of the various microbes has been investigated and several overgrowth bacteria have been identified (**Table 3**), these pathogenic microorganisms induced intestinal inflammation through several immune pathways, including endocytosis via invasion, toxin, secretion and/or adherence (Sansone, Medzhitov 2009). In addition, there is an increasing evidence suggesting that gut microbes can inter-react with host miRNAs and demonstrated in several mice models (**Table 4**).

**Table 3.** The major reported gut microbes in colorectal cancer.

Genus/Species	Change	Sample Type	Reference
<i>Anaerotruncus</i>	Up	Faeces	(N. Wu et al. 2013; Li et al. 2019)
<i>Actinomyces odontolyticus</i>	Up	Faeces	(Kasai et al. 2016; Mizutani et al. 2020; Yachida et al. 2019)
(Enterotoxigenic) <i>Bacteroides fragilis</i>	Up	Faeces/Mucosa	(Zhou et al. 2016; Hwang et al. 2018; Sobhani et al. 2011; Wei et al. 2016; Kasai et al. 2016)
<i>Bifidobacterium</i>	Down	Mucosa	(Chen et al. 2012; Masanobu et al. 2011; Fanning et al. 2012)
<i>Blautia</i>	Down	Mucosa	(Yuan et al. 2018; Ocvirk et al. 2020; Chen et al. 2012)
<i>Clostridioides difficile</i>	Up	Faeces	(Zheng et al. 2017; Fukugaiti et al. 2015)
<i>Campylobacter</i>	Up	Faeces	(Warren et al. 2013; N. Wu et al. 2013)
<i>Collinsella</i>	Up	Faeces	(Sheng et al. 2019; N. Wu et al. 2013)
<i>Enterococcus</i>	Up	Faeces	(Balamurugan et al. 2008; Zorron Cheng Tao Pu et al. 2020; Gaines et al. 2020)
<i>Enterococcaceae</i>	Up	Faeces	(Wang, Huycke 2007; N. Wu et al. 2013)
<i>Erysipelotrichaceae</i>	Up	Faeces	(Turnbaugh et al. 2009; N. Wu et al. 2013)
<i>Faecalibacterium</i>	Down	Faeces/Mucosa	(N. Wu et al. 2013; Chen et al. 2012)
<i>Fusobacterium</i>	Up	Faeces/Mucosa	(Castellari et al. 2012; Chen et al. 2012; N. Wu et al. 2013)
<i>Mogibacterium</i>	Up	Mucosa	(Hale et al. 2017; Chen et al. 2012)
<i>Parvimonas micra</i>	Up	Faeces	(Löwenmark et al. 2020; J. Xu et al. 2020; Flemer et al. 2018; Zhao et al. 2022)
<i>Peptostreptococcus</i>	Up	Faeces/Mucosa	(N. Wu et al. 2013; Chen et al. 2012)
<i>Porphyromonas</i>	Up	Mucosa	(Chen et al. 2012; Zorron Cheng Tao Pu et al. 2020)
<i>Roseburia</i>	Down	Faeces	(N. Wu et al. 2013; Chen et al. 2012)
<i>Prevotella</i>	Up	Stool/Mucosa	(Flemer et al. 2017; Kasai et al. 2016; Sobhani et al. 2011)
<i>Streptococcus sanguinis</i>	Up	Faeces	(Nijjer, Dubrey 2010; Fass et al. 1995; Rawla et al. 2017; Chen et al. 2017)

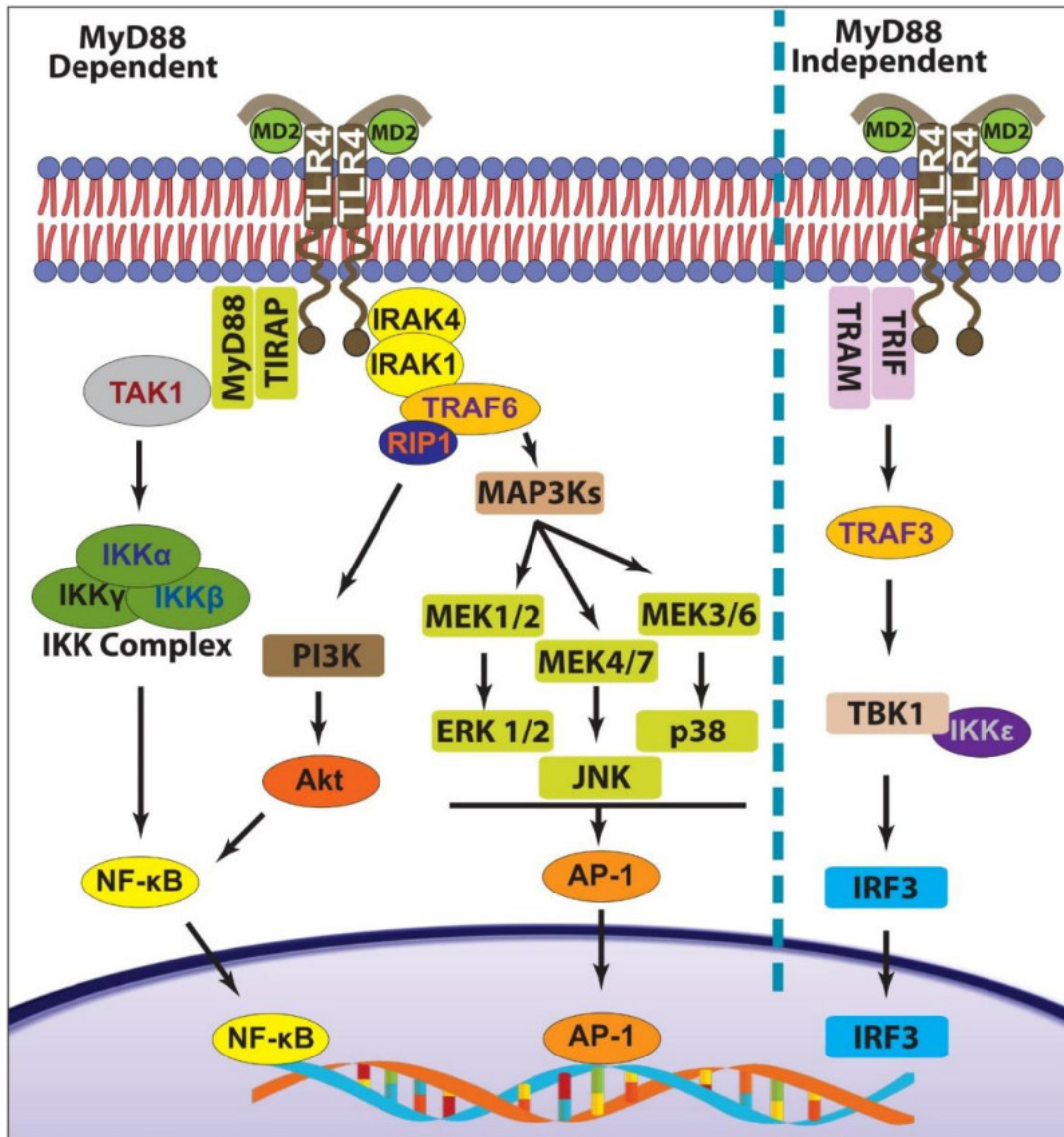
**Table 4.** MicroRNA-microbiota interaction in mice models.

Bacteria	Model	Target	Reference
<i>E. coli</i> (Nissle 1917)	• C57BL/6J w/ DSS	miR-223, miR-155, miR-150, miR-143, miR-375	(Rodríguez-Nogales, Algeri, Garrido-Mesa, Vezza, Maria P. Utrilla, et al. 2018)
<i>E. coli</i>	• C57BL/6J • Germ-free • Dicer1 <sup>ΔIEC</sup>	miR-1226-5p	(Liu et al. 2016)
<i>Saccharomyces boulardii</i>	• C57BL/6J w/ DSS or probiotics	miR-155, miR-223	(Rodríguez-Nogales, Algeri, Garrido-Mesa, Vezza, M. Pilar Utrilla, et al. 2018)
<i>L. fermentum</i>	• C57BL/6J w/ DSS or probiotics	miR-150, miR-155, miR-223, miR-143	(Rodríguez-Nogales et al. 2017)
<i>L. salivarius</i>	• C57BL/6J w/ DSS or probiotics	miR-155, miR-223	(Rodríguez-Nogales et al. 2017)
<i>Listeria monocytogenes</i>	• Germfree • <i>Listeria</i> infected C57BL/6J	miR-143, miR-148a miR-200b, miR-200c	(Archambaud et al. 2013)
<i>F. nucleatum</i>	• BALB/C nude • C57BL miR21a <sup>-/-</sup> • C57BL APC <sup>min/+</sup> (w/., w/out AOM/DSS)	miR-21	(Yongzhi Yang et al. 2017)
	• C57BL/6J • Germ-free • Dicer1 <sup>ΔIEC</sup>	miR-515-5p	(Liu et al. 2016)
<i>B. bifidum</i>	• BALB/c w/ AOM, <i>L. acidophilus</i> or <i>B. bifidum</i>	miR-135b, miR-155 miR-26b, miR-18a	(Heydari et al. 2019)
<i>L. acidophilus</i>	• BALB/c w/ AOM, <i>L. acidophilus</i> or <i>B. bifidum</i> • DSS-induced colitis	miR-135b, miR-155 miR-26b, miR-18a, miR-107, miR-146, miR-223	(Heydari et al. 2019)
<i>B. longum</i>	• Swiss CD-1 w/ AOM/DSS	miR-145, miR-15a miR-146a	(Fahmy et al. 2019)
<i>L. rhamnosus</i>	• C57BL/6 • IL-22 knockout • Germ-free	gma-miR396e mdo-miR-7267-3p	(Teng et al. 2018)

DSS, dextran sulphate sodium; AOM, azoxymethane

### (1). Gut microbes affect host microRNAs in colorectal cancer carcinogenesis

*F. nucleatum* was shown to increase CRC cell proliferation and tumour growth in mice by up-regulating miR-21 in the tumour cells by activating the TLR4–MyD88 signalling cascade (Yongzhi Yang et al. 2017). The loss of miR-21 reduces susceptibility in the experimental murine colitis model (Johnston et al. 2018). TLR4 (with the recruitment of MD2 accessory proteins) can be activated by lipopolysaccharides (LPS), bacterial metabolites or derivatives from Gram-negative bacteria (Liu et al. 2007; Wu et al. 2019). The activation of TLR4 dimerises and recruits downstream adaptor molecules MyD88/TIRAP (MyD88 dependent) or TRIF/TRAM (MyD88 independent) to mount an inflammatory response. The initiation of MyD88 dependent pathway activates IRAK4, IRAK1, TRAF6 and TAK1, forming an IKK complex and promoting NF- $\kappa$ B expression. Next, NF- $\kappa$ B translocates to the nucleus and furthers the transcription of proinflammatory cytokine-related genes (such as TNF $\alpha$ , IL-6, IL-1 $\alpha/\beta$  and IL-18) (Walsh et al. 2015). The stimulation of NF- $\kappa$ B can be archived via PI3K-AKT signalling (Walsh et al. 2015). On the other hand, MyD88 independent pathway relies on TRIF, activates TRAF3 and IRF3 through TBK1, inducing transcription of type I interferons (IFNs) and IFN-inducible genes (**Figure 15**). This LPS-TLR4 signalling pathway promotes TNF $\alpha$  production in intestinal immune cells, which further act on epithelial cells via TNFR1 (Ruder et al. 2019). The activation of TNF $\alpha$  induce immune response and pro-inflammatory cytokine and chemokine productions, which can also be found other colonic diseases including IBD (Bocchetti et al. 2021) and *C. difficile* infection (Tanya M. Monaghan, Seekatz, et al. 2021). Studies indicate that rs11536898 single-nucleotide polymorphism (SNP) in *TLR4* is significantly associated with CRC compared to healthy individuals (Slattery et al. 2012); and *TLR4* polymorphism (Thr399Ile and/or Asp299Gly) is significantly associated with CRC and *KRAS* gene mutations (Messaritakis et al. 2018).



**Figure 15. TLR4 signalling pathway.** TLR4, Toll-like receptor 4; MD2, lymphocyte antigen 96; MAPKs, mitogen-activated protein kinases; IRF3, interferon regulatory factor 3; TIRAP, Toll/interleukin-1 receptor domain-containing adapter protein; MyD88, myeloid differentiation primary response; TRAM, Translocation Associated Membrane Protein; TRIF, TIR-domain-containing adapter-inducing interferon- $\beta$ ; IRAK1, IL-1 receptor-associated kinase 1; IRAK4, IL-1 receptor-associated kinase 4; RIP1, receptor-interacting serine/threonine-protein kinase 1; TRAF 3, TNF receptor-associated factor 3; TRAF 6, TNF receptor-associated factor 6; TAK1, Transforming growth factor  $\beta$ -activated kinase1; IKK, I $\kappa$ B kinase; TBK1, TANK-binding Kinase 1, MEK, mitogen-activated protein kinase; JNK, c-Jun N-terminal kinase; ERK, extracellular signal-regulated kinase; AP-1, activator protein 1; p38, p38 mitogen-activated protein kinases; AKT, Protein kinase B; NF- $\kappa$ B, Nuclear factor-kappa B.

Colibactin is a genotoxic compound that encodes from *E. coli* and harbours pks genomic island (pks<sup>+</sup> *E. coli*) (Nougayrede 2006; Chagneau et al. 2019). Colibactin can stimulate host cells' inflammation and alkylate DNA on adenine residues to induce double-strand breaks, causing chromosomal instability and mutations, and promoting cancer development (Wilson et al. 2019; Xue et al. 2019; Nougayrede 2006). Colibactin in CRC cells can also induce c-Myc expression, resulting in the up-regulation of miR-20a-5p, and suppress

*SENP1* gene expression. The lower expression of *SENP1* negatively regulates cell senescence via p53 small ubiquitin-like modification (SUMOylation); drives a senescence-associated secretory phenotype (SASP); and secretes the growth factors that stimulate tumour growth (Dalmasso et al. 2014; Tjalsma et al. 2012; Cougnoux et al. 2014).

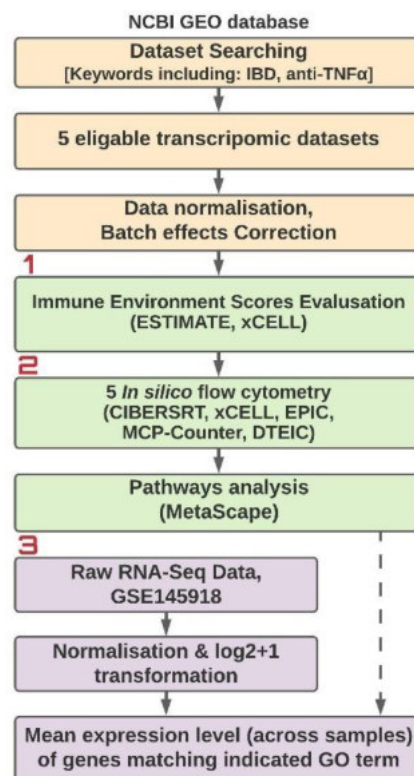
In addition to *F. nucleatum* and *pks<sup>+</sup> E. coli*, enterotoxigenic *B. fragilis* (*ETBF*) that encode a 20 kDa metalloprotease called – *B. fragilis* toxin (BFT) (DeStefano Shields et al. 2016; Vetizou et al. 2015) have been linked to diarrheal disorders, IBD and CRC pathogenesis (Chung et al. 2018; Zamani et al. 2020; Purcell et al. 2017). BFT-treated cells cause persistent cellular proliferation, target the epithelial cell tight junctions and enhance barrier permeability (Wu et al. 2006). BFT also induces IL-8 secretion, triggers E-cadherin cleavage and c-myc translation, initiates the nuclear localisation of  $\beta$ -catenin and actuates NF- $\kappa$ B signalling (Hwang et al. 2013; Wu et al. 2007; Sears 2009; Jeon et al. 2019). The expression of  $\beta$ -catenin can return to its original level relatively late after stimulation (Jeon et al. 2019). The toxin can also induce *B. fragilis*-associated lncRNA1 (*BFAL1*) expression via the Ras homolog - the MTORC1 binding target of rapamycin (RHEB/mTOR) signalling pathway. *BFAL1* modulates *RHEB* expression by competitively sponging miR-200a-3p and miR-155-5p (Bao et al. 2019).

## **(2). Gut flora influenced by host microRNA**

It is reported that intestinal microbiota negatively regulates host miR-107 in macrophages and dendritic cells in NF- $\kappa$ B- and MyD88-dependent manner to maintain intestinal homeostasis (Xue et al. 2014). A study reported that hsa-miR-515-5p targets 16S/23S rRNAs of *F. nucleatum*, while hsa-miRNA-1226-5p modifies *yegH* gene expression of *E. coli* (Liu et al. 2016). The authors first used a prediction tool for miRNA candidates and cultured the bacteria separately with the target miRNA mimics and observed the promotion of bacteria growth. Then, knockout of DICER at the intestinal epithelial cells in experiential mice resulted in a significant reduction in the total miRNA level in faeces, transformed the gut microenvironment and exacerbated colitis. Faeces transplanted from the wild-type mice into the Dicer knockout mice ameliorated damage and inflammation from colitis (Liu et al. 2016). Here, the delivery of host-derived miRNAs may be transferred via extracellular vesicles, affecting the gut microbiome composition, and regulating bacterial growth. For instance, the gain of function p53 mutant CRC cells selectively abandons exosomal transcriptome with hsa-miR-1246-5p (Cooks et al. 2018). Uptake of exosomal hsa-miR-1246-5p by tumour-associated macrophages stimulates the reprogramming into an anti-inflammatory condition, increases activity of TGF- $\beta$  and supports cancer growth (Cooks et al. 2018).

## V. RESEARCH METHODOLOGY

To improve IBD management and prevent neoplasia development in colon and rectum, identifying molecular pathways and finding reliable diagnostic biomarkers for patient response to anti-TNF- $\alpha$  treatment are necessary. In the beginning, a total of 449 freely available Affymetrix transcriptomic data from five eligible datasets were systemically retrieved from Gene Expression Omnibus (GEO) database, including control, baseline, and after primary anti-TNF- $\alpha$  therapy in IBD patients or placebo. Then, two immune microenvironments algorithms and five *in silico* flow cytometry programs were used to evaluate immune cell populations and Metascape – a pathway analysis tool for gene enrichment analysis. The outcome from the pathway analysis was validated on neutrophils isolated from choriodecidua cells with LPS-induced inflammation treated with or without adalimumab (Yau et al. 2022) (**Figure 16**). The AUC curve analysis was applied to all the genes to identify the best prediction biomarkers (**Publication 5, page 129**) (Yau et al. 2022).

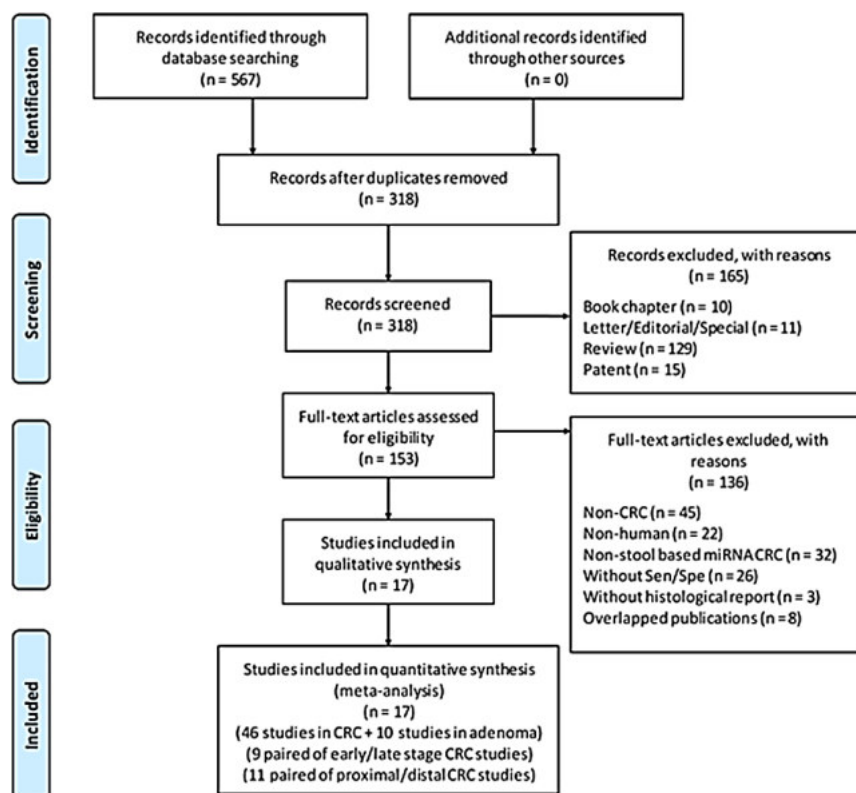


**Figure 16. The analysis workflow on anti-TNF- $\alpha$  treatment resistance IBD patients.**

Other than improving disease management for IBD patients, CRC development from neoplasia can be prevented by using non-invasive stool-based screening approaches such as FIT. However, there are a number of limitations in FIT and thus new non-invasive biomarkers to improve the screening accuracy are required. In the study, miRNA and gut microbes have the potential to archive the goal. Initially, we focus on faecal-based miRNA. Briefly, reverse

transcription for miRNA array carries out using Human Pools A and B v2.1 Kit (Applied Biosystems) – a megaplex primer pools in the 5 paired CRC-adjunct normal tissues to identify high express miRNAs (Wu et al. 2014). Then, some of the target miRNAs will further be validated in 40 paired CRC-para normal tissues and CRC cell lines. Next, total RNA from the faecal samples (n = 595, containing 198 CRCs, 199 adenomas and 198 healthy subjects) were isolated from the TRIzol-chloroform mixture using the miRNeasy Mini Kit (Qiagen) and qRT-PCR of individual miRNA was performed using TaqMan miRNA Reverse Transcription Kit and TaqMan Human MiRNA Assay (Applied Biosystems) (**Publication 6-8, from page 141**)(Yau, Wu, et al. 2014; Yau et al. 2016).

To further investigate the use of faecal-based miRNA potential for CRC screening, the Cochrane Handbook for Diagnostic Test Accuracy Review protocol was followed to perform the meta-analysis to evaluate the utility of faecal miRNAs as a non-invasive tool in CRC screening. A systematic literature search in five databases (PubMed, Ovid Embase, The Cochrane Library, Scopus and Web of Science) identified 17 research articles including 6475, 783 and 5569 faecal-based miRNA tests in CRC, adenoma patients and healthy individuals, respectively (**Figure 17**)(**Publication 9, page 171**)(Yau et al. 2019).

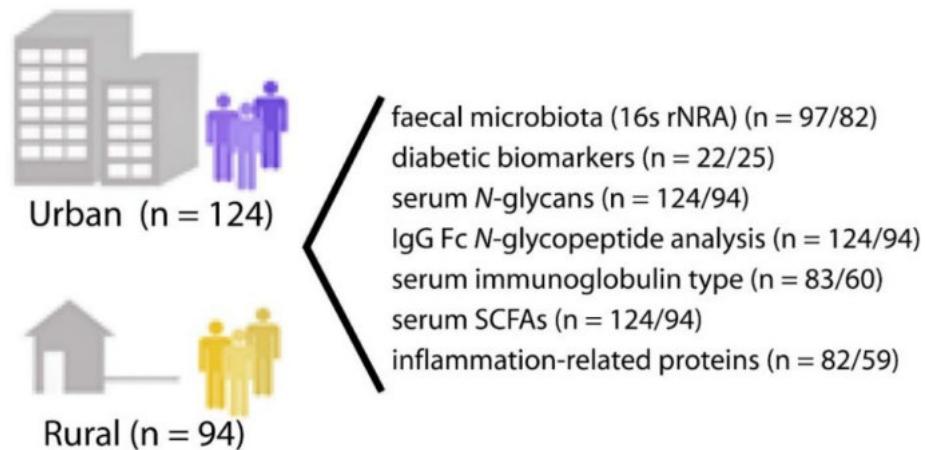


**Figure 17. Flowchart diagram of faecal-based miRNAs study selections based on the inclusion and exclusion criteria.** Adopted from (Yau et al. 2019).



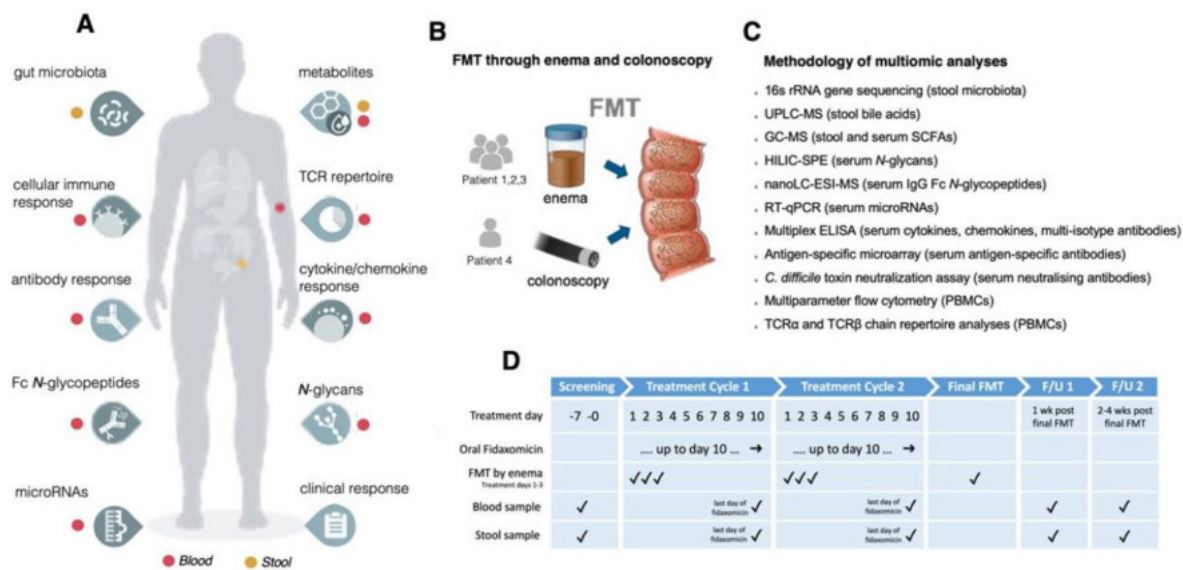
In addition to faecal-based miRNA, gut dysbiosis in CRC has been investigated and some of them have been reported as potential biomarkers for non-invasive CRC screening. To search for the candidate targets, at first, the metagenome-wide association studies on faecal samples from 74 patients with CRC and 54 controls from the East Asian cohort, and validated the early outcomes in 16 patients and 24 controls from Denmark. Additional confirmation on the selected biomarkers for the two reported European cohorts. Finally, we applied targeted qPCR assays to evaluate the diagnostic potential of selected biomarkers in an independent Chinese cohort of 47 patients and 109 controls and identified 20 bacterial gene marker candidates that may serve as non-invasive biomarkers for CRC (**Publication 10, page 185**)(J. Yu et al. 2017). These findings further confirmed that faecal bacteria may be able to serve as non-invasive targets for CRC screening by using target qPCR. These biomarkers, by combination with *F. nucleatum*, undefined 'm7', *Bacteroides clarus* (*B. clarus*) and *Clostridium hathewayi* (*C. hathewayi*) could improve the CRC diagnostic performance but the diagnostic performance for adenoma is not satisfactory (Liang et al. 2017). Thus, a new *Lachnoclostridium* gene marker (labelled as 'm3') was identified and further evaluated the utility for the diagnosis of colorectal adenoma. The diagnostic accuracy of m3, compared with and/or in combination with FIT and other bacterial gene markers, was tested in 1012 subjects (274 CRC, 353 adenoma and 385 controls) from two separated East Asian cohorts (**Publication 11, page 196**)(Liang et al. 2020).

Furthermore, to demonstrate and validate gut microbiome varies to factors including, but not limited geographic locations, lifestyles, diets, disease types and statuses widely between people, an in-depth phenotypic prospective study was conducted based on 218 adults from rural (n = 94) and urban (n = 124) areas of central India by using multi-omics analysis to determine the relationship between disease risk, circulating biomarkers and microbial taxa (**Figure 18**). High-throughput experiments, including multiplex assays for serum diabetic proteins, cytokines, chemokines and polyisotypic antibodies, quantification of serum short-chain fatty acids by gas chromatography-mass spectrometry (GC-MS) and analysis of faecal microbiota by 16s ribosomal RNA gene amplicon sequencing were used. Sera were also analysed for *N*-glycans and immunoglobulin G Fc *N*-glycopeptides (**Publication 12, page 208**)(Tanya M. Monaghan, Biswas, et al. 2021).



**Figure 18. Schematic diagram of the overall study design on rural and urban areas of central India.** n; number of urban/rural samples. Adopted from (Tanya M. Monaghan, Biswas, et al. 2021).

*C. difficile* is a type of bacteria that can cause colitis, it is sometimes life-threatening. Infections from *C. difficile* often start with taking antibiotics and FMT is an unconventional treatment strategy for CDI patients. Although the growing evidence to support the use of FMT in CDI, the multifactorial mechanisms that underpin the efficacy of FMT remain unclear. Thus, a deep phenomics study on four adults (three responders and one non-responder) receiving sequential FMT for severe or fulminant CDI (SFCDI), in which we performed a longitudinal, integrative analysis of multiple host factors and intestinal microbiome changes. Faecal samples were profiled in a variety of assays for changes in gut microbiota and metabolites, and blood samples for alterations in targeted epigenomic (including miRNAs), metabonomic, glycomic, immune proteomic, immunophenotyping, immune functional assays, and T-cell receptor (TCR) repertoires, respectively (**Figure 19**)(**Publication 13, page 232**)(Tanya M. Monaghan, Duggal, et al. 2021).



**Figure 19. Schematic diagram of the study workflow on severe or fulminant *Clostridioides difficile* infection patients on three faecal microbiota transplantation (FMT) responders and one FMT non-responder.** (A) Multidimensional, longitudinal assays applied in patients receiving sequential FMT either by colonoscopy or enema with severe or fulminant *Clostridioides difficile* infection; (B) FMT Delivery route for each participant; (C) Methodologies applied; (D) Treatment timelines and sampling strategy. Adopted from (Tanya M Monaghan, Duggal, et al. 2021).

Based on the investigative study, next, an in-depth study concentrates on the molecular mechanisms of miRNA alterations underlying successful FMT for recurrent CDI (rCDI) patients. To be more specific, sera samples from 2 prospective multicentre randomised controlled trials were used for miRNA profiling on the Nanostring nCounter platform and revealed dysregulation circulating miRNAs 4 and 12 weeks after FMT compared with pre-treatment screening, of which the top miRNAs were validated in the discovery cohort by means of RT-qPCR. In a murine model of rCDI, RT-qPCR analyses of sera and caecal RNA extracts demonstrated suppression of these miRs, an effect reversed by FMT (**Publication 14, page 258**)(Tanya M. Monaghan, Seekatz, et al. 2021).

In addition to colorectum, a relatively stable microbiota microenvironment is also important in stomach. *Helicobacter pylori* and EBV are the key oncogenic bacteria, GC patients present stomach dysbiosis and form a genotoxic microbial community by reduced microbial diversity and the abundance of *Helicobacter*, with over-population of new bacterial genera (Ferreira et al. 2018). Thus, we worked on whole-genome, transcriptome, and epigenome sequence analyses of a gastric adenocarcinoma cell line (AGS cells) before and after EBV infection. Gastric tumour samples with (n = 34) or without (n = 100) EBV infection were also applied to look for the alteration (**Publication 15, page 277**)(Liang et al. 2014).

## VI. RESULTS

### i. INCREASE ANTI-TNF- $\alpha$ TREATMENT RESPONSIVENESS IN PATIENTS WITH INFLAMMATORY BOWEL DISEASE

#### (1). Introduction

Currently, anti-TNF $\alpha$  compounds such as full monoclonal IgG1 antibodies (infliximab and adalimumab), pegylated anti-TNF $\alpha$  F[ab']<sub>2</sub> fragment (certolizumab), and IgG1 $\kappa$  monoclonal antibody - derived from immunising genetically engineered mice with human TNF $\alpha$  (golimumab) have been approved for IBD patients (**Table 5**) (Berns, Hommes 2016; Levin et al. 2016; Moroi et al. 2013). Although the retrospective study indicated that TNF $\alpha$  blockers lower the risk of CRC in IBD patients (Alkhayyat et al. 2020), approximately 30% of patients do not respond to anti-TNF induction therapy (primary non-response), and up to 50% of the patients lose response to treatment over time, after initially experiencing clinical improvement (secondary loss of response). More importantly, anti-TNF treatment could induce the risk of serious infection (Gerriets et al. 2020; Poullenot et al. 2016). Thus, improving the anti-TNF $\alpha$  treatment responsiveness in IBD is needed to improve the disease management.

**Table 5.** Anti-tumour necrosis factor (TNF) therapeutic agents for moderate-to-severe IBD patients.

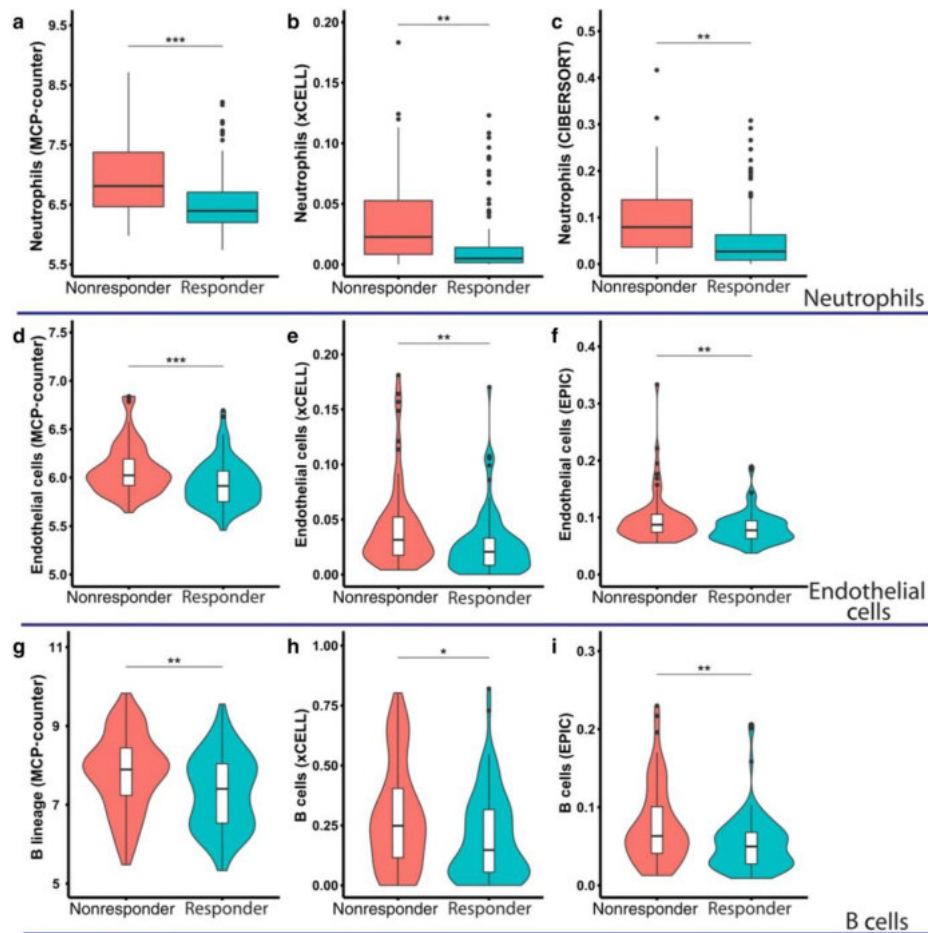
Anti-TNF $\alpha$	Abbv.	Structure	Administration	Adverse effects (Unique)	Indication
<b>Adalimumab</b>	ADM	Human monoclonal antibody	• Subcutaneous	• Injection site reactions	<b>CD:</b> induction (include IFX failure, maintenance) <b>UC:</b> induction, maintenance
<b>Certolizumab Pegol</b>	CTZ	Recombinant antigen binding Fab fragment conjugated to polyethylene glycol	• Subcutaneous	• Injection site reactions	<b>CD:</b> maintenance
<b>Golimumab</b>	GLM	Human monoclonal antibody	• Subcutaneous	• Injection site reactions	<b>UC:</b> induction (include corticosteroid-dependent), maintenance
<b>Infliximab</b>	IFX	75% human monoclonal antibody + 25% chimeric mouse	• Intravenous • Subcutaneous	• Infusion reactions • Serum sickness • Risk of heart failure	<b>CD:</b> induction, maintenance <b>UC:</b> induction, maintenance, corticosteroid elimination

Abbv., Abbreviation; CD, Crohn's disease; UC, ulcerative colitis.

#### (2). Summary of the outcomes

To find out the reason behind the treatment irresponsiveness, optimise the disease management and minimise the chance to develop colonic neoplasms in IBD patients, in the beginning, a total of 449 publicly available transcriptomic data from five eligible datasets were systemically retrieved, including control, baseline, and after primary anti-TNF- $\alpha$  therapy in IBD patients or placebo. Then, based on the outcomes from *in silico* flow cytometry algorithms analysis, neutrophil, endothelial cell, and B-cell populations were higher in baseline nonresponders (**Figure 20**) and revealed that neutrophil chemotaxis pathways – identified from Metascape – may contribute to the treatment resistance. Genes identified within the

neutrophil chemotaxis enrichment pathways were validated in LPS-induced inflammation on neutrophils in a previously published RNA-sequencing dataset from the anti-TNF- $\alpha$ -treated animal model. The AUC analysis was applied to all the genes to identify the best prediction biomarkers. Interleukin 13 receptor subunit alpha 2 (*IL13RA2*) is the best predictor (AUC: 80.7%, 95% confidence interval: 73.8%–87.5%), with a specificity of 84.93% and sensitivity of 68.13%, and significantly higher in nonresponders compared with responders ( $P < 0.0001$ ) (Publication 5, page 129)(Yau et al. 2022).



**Figure 20. Neutrophils, endothelial cells, and B cells are significantly higher on the baseline anti-TNF- $\alpha$  treatment IBD nonresponders compared with responders patients.** Immune cell population evaluated in five *in silico* flow cytometry, and (a–c) neutrophils, (d–f) endothelial cells, and (g–i) B cells can be recognized in three out of five algorithms. B-cell, endothelial cell, and neutrophil populations are higher on baseline anti-TNF- $\alpha$  nonresponders compared with responders. P-value determined by Mann–Whitney U-test. Asterisks denote statistically significant differences (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , and \*\*\*\* $P < 0.0001$ ). Adopted from (Yau et al. 2022).

### (3). Discussions and Conclusions

In a typical inflammatory response, immune cells such as T lymphocytes, dendritic cells, macrophages, and natural killer cells release pro-inflammatory cytokine TNF- $\alpha$ , leading to the activation of neutrophils and endothelial cells (Balamayooran et al. 2010). The activation

of endothelial cells in colonic mucosa enhances vascular permeability, induces the recruitment of immune cells, and thus activates chemotaxis. At the same time, the activation of neutrophils follows the tethering, rolling, crawling, and transmigration process from the blood vessel into the inflamed colonic tissues (Wéra et al. 2016). When neutrophils engulf invasive gut microbiome, they release granule proteins and chromatin to form neutrophil extracellular traps and secrete antimicrobial peptides to mediate extracellular killing of microbial pathogens (Schmidt et al. 2011). Nevertheless, hyperactive neutrophils trigger an unrestrained activity of the positive feedback amplification loops, leading to endothelial cells and the surrounding tissues damage, inducing resolution delay (IL-6, TNF- $\alpha$ , and IFN- $\gamma$ ) and chemokines (IL-8, CCL3, and CCL4), which further the recruitment of neutrophils, monocytes, and macrophages to the inflamed sites (Mortaz et al. 2018).

The use of anti-TNF- $\alpha$  blockers significantly suppresses the infiltration of neutrophil and B-cell population in the inflamed mucosa, and suppresses pro-inflammatory mediators, such as calprotectin (S100A8/A9), IL-8, IL-6, and TNF- $\alpha$  production,(C. Zhang et al. 2018; Timmermans et al. 2016) and matched with our finding only in responders. However, the unwanted immunogenicity has a high level of B cells due to the presence of anti-drug antibodies (Vaisman-Mentesh et al. 2019), including anti-TNF- $\alpha$  monotonical antibodies (De Groot, Scott 2007) which is matched with our findings. In addition, the study also identified that IL13RA2 is stand-alone in the volcano plot with the highest fold change and the lowest  $p$ -value, suggesting IL13RA2 is a potential biomarker to predict anti-TNF- $\alpha$  treatment response.

In summary, nonresponders presented higher populations of neutrophils, endothelial cells, and B cells compared to responders at baseline level. IL13RA2 is a potential biomarker to predict anti-TNF- $\alpha$  treatment response. These findings may help us to further investigate the drug resistance in anti-TNF- $\alpha$  monotonical antibodies against inflammation.

## **ii. STOOL-BASED MICRORNA AS NON-INVASIVE BIOMARKERS FOR COLORECTAL CANCER SCREENING**

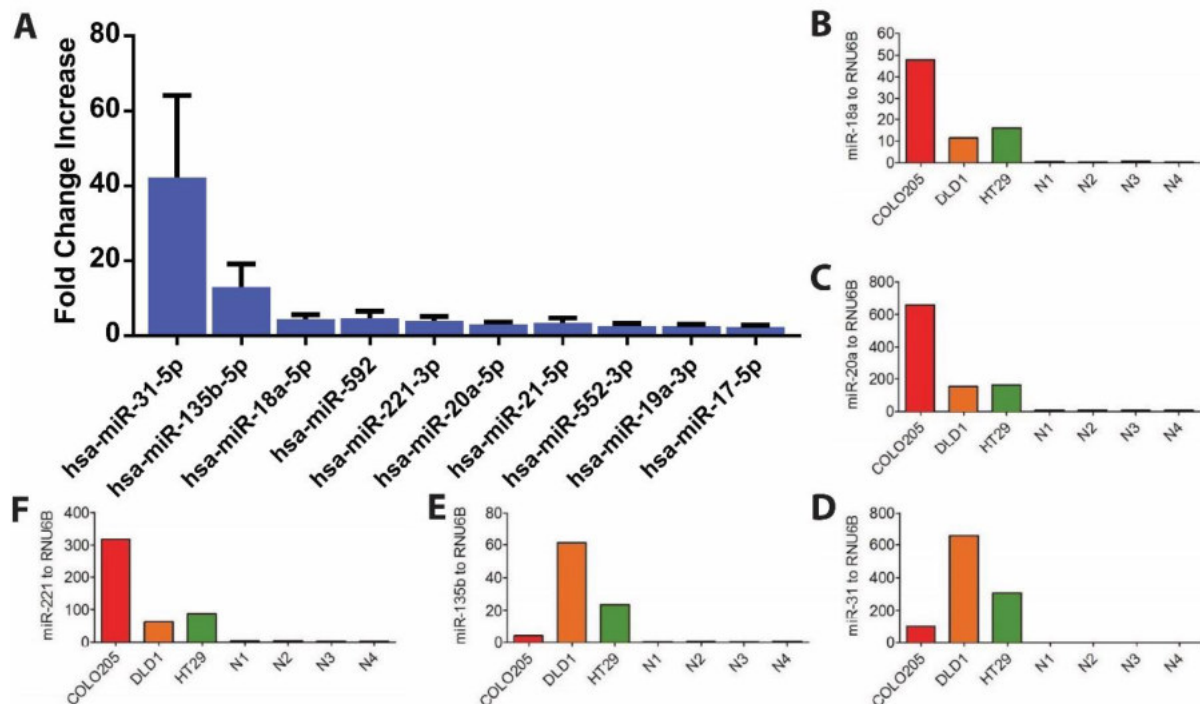
### **(1). Introduction**

Implementation of CRC screening programs in communities allows early detection of colonic neoplasm(s) to lower the treatment need, morbidity, and mortality (Zauber et al. 2012). However, the use of low-cost haemoglobin-based CRC screening such as FIT is limited by haemoglobin degradation and intermittent bleeding patterns in large intestines, so that there is still a large number of individuals cannot be identified until the late stages, leading to poor treatment responses and prognosis (**Publication 2, page 92**)(Tepus, Yau 2020). The search

for reliable non-invasive biomarkers is therefore still ongoing, and faecal-based miRNAs and gut microbes are crucial tools in helping to discover new biomarkers. The expression of several miRNAs differs significantly between normal colonic tissues and CRC, including IBD and colitis-induced CRC (**Publication 3, page 102**)(Bocchetti et al. 2021). These abnormal colonocytes consistently exfoliate and shed into the lumen. miRNAs are highly stable and detectable within samples throughout a 72-hour incubation period at room temperature, due to protection from ribonuclease degradation by exosomes (Hunter et al. 2008; Mitchell et al. 2008). Thus, miRNA levels can be detected in faecal specimens.

## (2). Summary of the outcomes

To begin with, a total of 10 RNA samples from 5 paired CRC-adjacent normal tissues were applied for miRNA profiling, and identified 10 highly expressed miRNAs (**Figure 21A, replotted the figure based on the supplementary data from Publication 6, page 141**) (Wu et al. 2014). Next, five of the target miRNAs were validated in 40 pairs of CRC tissues (**Table 6, data reorganised from Publication 6-8, from page 141**) (Wu et al. 2014; Yau, Wu, et al. 2014; Yau et al. 2016) and CRC cell lines (**Figure 21B-F, previously unpublished data**).



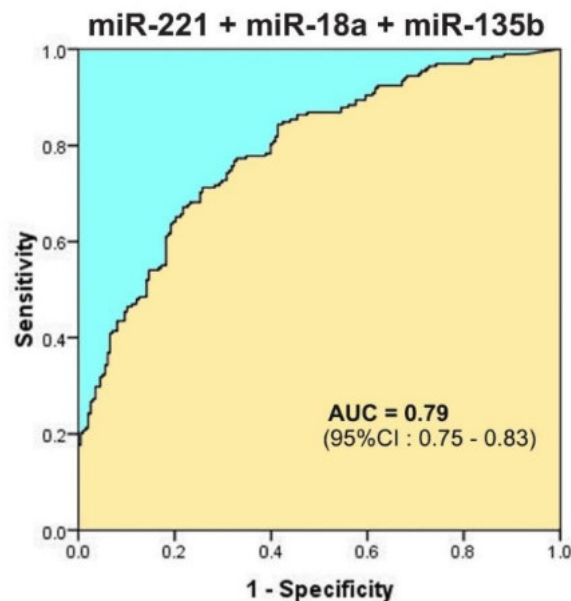
**Figure 21. Identification of upregulated microRNAs in colorectal cancer by qPCR profiling.** (A) 667 microRNAs in five paired colorectal cancer (CRC) patients were evaluated. Fold change below 0.8 on each analyte in each paired CRC sample was eliminated. Adapted from (Wu et al. 2014). Levels of (B) miR-18a-5p, (C) miR-20a-5p, (D) miR-31-5p, (E) miR-135b-5p, (F) miR-221-5p in CRC epithelial cell lines Colo-205, DLD-1, HT-29, and 4 colon biopsies from patients of normal colonoscopy findings. Levels of microRNA are normalised to internal control RNU6B.

**Table 6. microRNA differentially expressed in tumours compared with adjacent normal tissues.** Adopted from (Wu et al. 2014; Yau, Wu, et al. 2014; Yau et al. 2016).

miRNA	Chromosomal location	% of samples with elevated expression in tumour	Fold change (Interquartile range)	P-value*
miR-135b	1q32.1	92.5% (37/40)	13.73 (5.532-45.66)	< 0.0001
miR-31	9p21.3	87.5% (35/40)	10.45 (2.443-99.31)	< 0.0001
miR-221	Xp11.3	77.5% (31/40)	1.96 (1.025-3.319)	< 0.0001
miR-18a	13q31.3	77.5% (31/40)	2.65 (1.078-6.828)	0.0003
miR-20a	13q31.3	70.0% (28/40)	2.063 (0.910-5.418)	0.0065
Combined	-	97.5% (39/40)	-	-

\*P values were analysed by Wilcoxon matched-pairs test.

Then, a total of 595 faeces samples from healthy subjects (n = 198), polyps (n = 199) and CRC patients with different cancer stagings (n = 198) were extracted for miRNA-specific reverse transcription and Taqman probe-based real-time polymerase chain reaction (RT-qPCR) for the specific miRNAs. Our experimental findings confirmed that miR-18a, miR-20a, miR-221 and miR-135b are potential non-invasive biomarkers (**Publication 6-8, from page 141**) (Wu et al. 2014; Yau, Wu, et al. 2014; Yau et al. 2016). The combination of miR-221, miR-18a and miR-135b, for example, is the best faecal-based miRNAs combination in our study and showed an improvement in screening accuracy for CRC patients (AUC: 0.79, 95% CI 0.75 – 0.83)(**Publication 7, page 152**)(Yau, Wu, et al. 2014).

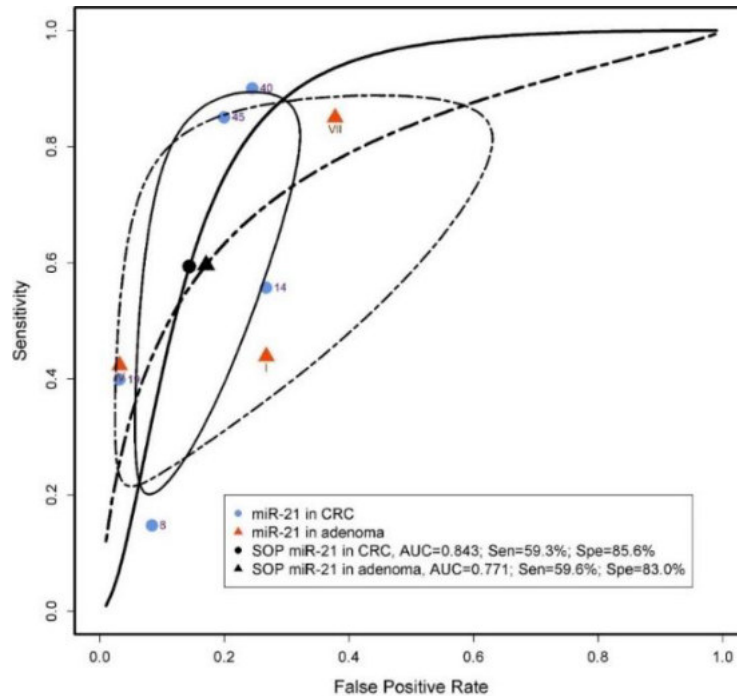


**Figure 22. The combination of miR-221, miR-18a and miR-135b has an AUC of 0.79.** Adopted from (Yau, Wu, et al. 2014).

Afterwards, a meta-analysis approach was applied to assess different faecal-based miRNA studies for CRC and colonic adenoma screening from the published research articles in any language up to November 17, 2017. The study identified 21 miRNAs and estimated that miR-21 may be the most reliable miRNA marker. miR-21 has an AUC of 0.84 with a sensitivity



of 59.3% and a specificity of 85.6% for CRC patients based on the 5 studies; and has an AUC of 0.78 with a sensitivity of 59.6% and a specificity of 83.0% for colonic adenoma according to the 3 published outcomes (**Figure 23**)(**Publication 9, page 171**)(Yau et al. 2019).



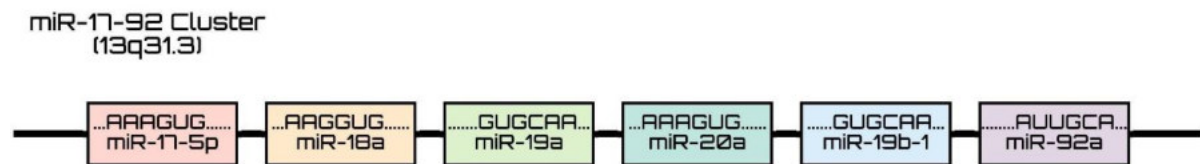
**Figure 23. Diagnostic accuracy in pooled stool-based miR-21 for colorectal cancer screening.** SROC for pooled miR-21 in the detection of CRC (n = 5) and colonic adenoma (n = 3). The number next to the dot/triangle corresponding to the study ID in Table 1 (Blue dots: CRC) or Table 2 (Red triangles: colonic adenoma) in the publication (Yau et al. 2019). Sen, sensitivity; Spe, specificity; SOP, summary operating point. The circular regions (95% confidence contour) contain likely combinations of the mean value of sensitivity and specificity. Adopted from (Yau et al. 2019).

### (3). Discussions and Conclusions

Unlike FIT, which is currently used for CRC screening, faecal-based miRNA tests do not require troublesome drug and dietary restrictions. Therefore, the uptake of faecal-based miRNA tests may be higher than that of FIT. As a result, quantitation of miRNA biomarkers in human faeces by qRT-PCR is a promising non-invasive approach for screening CRC patients. We have investigated the expression profile of 667 mRNAs, and reported that miR-18a and miR-20a from microRNA-17-92 cluster, and miR-221 as candidate biomarkers (Wu et al. 2014; Yau, Wu, et al. 2014; Yau et al. 2016). Also, the meta-analysis summarised 21 miRNAs from 17 different publications and miR-21 could be the best miRNA biomarker for CRC screening (Yau et al. 2019). In conclusion, faecal-based microRNA could be a useful target for CRC, in combination with other biological markers along with FIT could improve the detection accuracy. The biological functions of these miRNAs were also discussed below.

### A. microRNA-20a and miR-18a in microRNA-17-92 Cluster in Colorectal Cancer

miR-17-92 is one of the most extensively investigated miRNA clusters and located in chromosome 13q31.3. It comprises six mature miRNAs, including miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1 and miR-92a-1 (**Figure 24**)(Ventura et al. 2008). Each of the individual miRNAs in the miR-17-92 cluster expresses varies in cancers compared to normal controls (G. Yu et al. 2012; Diosdado et al. 2009; Humphreys et al. 2014).



**Figure 24. microRNA-17-92 cluster at chromosome 13q31.3.**

miR-20a is an oncomiR and has been suggested as a diagnostic biomarker in GI cancers (**Table 7**)(Moody et al. 2019; Stojanovic et al. 2019). In CRC, over-expression of miR-20a promotes cell cancer proliferation, migration, and invasion, and inversely correlates with *Smad4* expression. The abolition of *Smad4* induces epithelial-mesenchymal transition (EMT), mediated by miR-20a (Zhang et al. 2014) and confirmed by dual-luciferase reporter assays (Cheng et al. 2016). The presence of miR-20a modulates the expression of tissue inhibitor of metalloproteinases-2 (*TIMP2*) and matrix metalloproteinase 9 (*MMP9*) to promote EMT; aids the detachment of CRC cells from the tissue parenchyma; and goes into systemic circulation during cancer metastasis (Xu et al. 2015). miR-20a is also involved in the cell cycle by regulating *FOXJ2* to control cell cycle arrest at G1 stage (Qiang et al. 2020), and *Myc/p21 (CDKN1A)* expression via TGF- $\beta$  by preventing its delay of G1/S transition (Sokolova et al. 2015). Silencing of miR-20a in SW480 CRC cell lines increases the anti-tumour effect of TRAIL through the caspase-8 dependent pathway (Huang et al. 2017), and inhibits translocation of truncated BID (tBID) to induce the mitochondrial pathway of apoptosis (Li et al. 1998; Orzechowska et al. 2015). By controlling *MICA* and *CXCL8*, miR-20a respectively modulates NK cells and influences tumour latency from IBD (Tang et al. 2019; Signs et al. 2018).

**Table 7.** The reported microRNA-20a studies with the corresponding direct target genes in gastrointestinal diseases.

Target Gene	Gene Name	Main Biological roles	GI Disease [Reference]
<b>Apoptosis</b>			
<i>BID</i>	BH3 Interacting Domain Death Agonist	Induce mitochondrial pathway of apoptosis	<b>CRC</b> (Huang et al. 2017)
<i>SENP1</i>	SUMO Specific Peptidase 1	Proliferation, apoptosis, invasion, migration & chemo-resistance	<b>CRC</b> (Huang et al. 2017; Xu et al. 2015)
<i>PDCD4</i>	programmed cell death factor 4	Cell proliferation, migration & invasion; promote cell apoptosis & 5-FU resistance	<b>CRC</b> (Jiang et al. 2020)
<b>Cell Cycle</b>			
<i>CDKN1A</i>	Cyclin Dependent Kinase Inhibitor 1A (p21)	Interfere colonic epithelium homeostasis by disrupting the regulation of Myc/p21 by TGF- $\beta$	<b>CRC</b> (Sokolova et al. 2015)
<i>FOXJ2</i>	Forkhead Box J2	Proliferation, migration, invasion& cell cycle arrest at G1 stage	<b>CRC</b> (Qiang et al. 2020)
<b>Drug Resistance</b>			
<i>ASK1</i>	Apoptosis signal-regulating kinase 1	Enhance sensitivity to cisplatin	<b>CRC</b> (L. Zhang et al. 2018)
<i>BNIP2</i>	BCL2 Interacting Protein 2	5-FU, oxaliplatin & teniposide resistance	<b>CRC</b> (Chai et al. 2011)
<i>HAND2-AS1</i>	heart and neural crest derivatives expressed 2-antisense RNA 1 - <i>lncRNA</i>	5-FU resistance, cell proliferation, migration & invasion & promoted cell apoptosis	<b>CRC</b> (Jiang et al. 2020)
<b>Extracellular Growth Factors</b>			
<i>SMAD4</i>	Mothers against decapentaplegic homolog 4	EMT signalling	<b>CRC</b> (Cheng et al. 2016; Zhang et al. 2014; Xu et al. 2015)
<b>G-protein coupled receptor</b>			
<i>GABBR1</i>	Gamma-Aminobutyric Acid Type B Receptor Subunit 1	Promote proliferation & invasion	<b>CRC</b> (Longqiu et al. 2016)
<b>Immune Response</b>			
<i>MICA</i>	Major Histocompatibility Complex (MHC) class I-related chain genes A	Regulate the sensitivity of CRC cells to NK cells	<b>CRC</b> (Tang et al. 2019)
<i>CXCL8</i>	Chemokine (C-X-C motif) ligand 8	Influence tumour latency from IBD	<b>IBD induced CRC</b> (Signs et al. 2018)

GI, Gastrointestinal, CRC, colorectal cancer; IBD, inflammatory bowel disease; GC, gastric cancer; HCC, Hepatocellular carcinoma; 5-FU, Fluorouracil; EMT, Epithelial-mesenchymal transition, TRAIL, TNF-related apoptosis-inducing ligand; lncRNA, long non-coding RNA.

miR-18a has a dual-functional role in either promoting or inhibiting oncogenesis in different human cancers and presents as an oncomiR in CRC (**Table 8**). The up-regulation of miR-18a was shown to inhibit the malignant progression of CRC by inhibiting GTPase cell division control protein 42 (*CDC42*), a mediator of PI3K pathway (Humphreys et al. 2014). The suppression of miR-18a restores *CDC42* mRNA expression and cell growth (Humphreys et al. 2014). The activation of *CDC42* promotes adhesion and invasion of CRC cells (Gao et al. 2013), and the malignant progression through silencing of *ID4* and *CACNA2D2* (Sakamori et al. 2014; Gómez Del Pulgar et al. 2008). On the other hand, the long non-coding RNA (lncRNA) *UCA1* was identified and up-regulates in CRC, promoting cell proliferation and tumorigenicity (Ni et al. 2015). *UCA1* can be a "sponge" to reduce the regulatory effect on miR-18a and thus increase *CDC42* expression to enhance the sensitivity to oncolytic vaccinia virus cell-to-cell spread by activating filopodia formation (Horita et al. 2019). Furthermore, the presence of lncRNA *FENDRR* can also act as a "sponge" to restrain the aggressiveness of CRC cells

through the regulation of miR-18a-5p/ING4 axis (Yin et al. 2019). Additionally, studies indicated that the expression of miR-18a attenuates DNA damage repair via suppressing *ATM* and induces CRC cell apoptosis via the autophagolysosomal degradation of *HNRNPA1* (C.-W. Wu et al. 2013; Fujiya et al. 2014). miR-18a can be a treatment target by encapsulation in grapefruit-derived nano-vectors (GNV) to induce M1 macrophage interferon gamma (IFN- $\gamma$ ) by targeting interferon regulatory factor 2 (*IRF2*) to induce IL12, and to activate NKT cells and NK cells suppressing cell metastasis to the liver (Teng et al. 2016). In colitis-associated colorectal cancer (CAC), up-regulation of miR-18a inhibits *PIAS3* expression and activates NF- $\kappa$ B and STAT3 in the colon of both CAC and CRC patients (Ma et al. 2018). This *PIAS3*-mediated autoregulatory feedback loop (*PIAS3*/NF- $\kappa$ B or *STAT3*/miR-18a) is verified in an AOM-DSS-induced mice model. Modulation of the feedback loops through suppression of miR-18a or over-expression of *PIAS3* significantly repressed cell proliferation in a mouse CRC xenograft model. By contrast, up-regulation of *PIAS3* by intra-colonic administration of *PIAS3* or anti-miR-18a lentivirus in AOM-DSS-induced mice reduces both tumour sizes and numbers (Ma et al. 2018).

**Table 8.** Reported microRNA-18a studies with the corresponding direct target genes in gastrointestinal diseases.

Target Gene	Gene Name	Main Biological roles	GI Disease [Reference]
<b>Apoptosis</b>			
<i>HNRNPA1</i>	heterogeneous nuclear ribonucleoprotein A1	Induces apoptosis via autophagolysosomal degradation	<b>CRC</b> (Fujiya et al. 2014)
<i>IRF2</i>	Interferon regulatory factor 2	Promote apoptosis, inhibit cell proliferation & migration; Induction of macrophage IFN $\gamma$ & activates NK & NK T cells in CRC	<b>CRC</b> (Teng et al. 2016); <b>GC</b> (Chen et al. 2016); <b>HCC</b> (Yongyu et al. 2018)
<b>Cell Cycle</b>			
<i>ATM</i>	Ataxia telangiectasia mutated	Attenuate DNA damage repair	<b>CRC</b> (C.-W. Wu et al. 2013)
<i>CDC42</i>	Cell division control protein 42	Acts as a GTPase in the PI3K/AKT pathway	<b>CRC</b> (Humphreys et al. 2014)
<i>UCA1</i>	Urothelial Cancer Associated 1 (lncRNA)	Sponging miR-18a, promoting Cdc42 activation, regulate oncolytic vaccinia virus cell-to-cell spread & filopodia formation	<b>CRC</b> (Horita et al. 2019)
<b>Transcription Factor</b>			
<i>TBPL1</i>	TATA-Box Binding Protein Like 1	Regulate tumour proliferation & invasion	<b>CRC</b> (Liu et al. 2015)
<i>PIAS3</i>	Protein inhibitor of activated STAT3	Inhibit STAT3 activity, downstream of the Wnt/ $\beta$ -catenin pathway	<b>Colitis-associated CRC</b> (Ma et al. 2018); <b>GC</b> (W. Wu et al. 2013)
<i>FENRR</i>	FOXF1 Adjacent Non-Coding Developmental Regulatory RNA (lncRNA)	Exert an inhibitory role by interacting with miR-18a-5p & increase ING4 expression	<b>CRC</b> (Yin et al. 2019)

CRC, colorectal cancer; GC, gastric cancer; HCC, Hepatocellular carcinoma; 5-FU, Fluorouracil; EMT, Epithelial-mesenchymal transition; NK, natural killer; lncRNA, long non-coding RNA.

## B. microRNA-21 in Colorectal Cancer

miR-21 is one of the most highly-expressed miRNAs in CRC, IBD and colitis-associated CRC and could also be utilised to distinguish these diseases for patient diagnosis and prognosis (Svrcek et al. 2013; Shi et al. 2013; Ando et al. 2016). The up-regulation of miR-21 in CRC correlates to cell migration, invasion and proliferation, as well as promoting miR-21-mediated transformation in somatic cells (Iliopoulos et al. 2010; Shi et al. 2016). The expression of miR-21 regulates a plethora of genes in GI cancer (**Table 9**), including programmed cell death 4 (*PDCD4*) (Asangani et al. 2008; Peacock et al. 2014; Lu et al. 2008), Sprouty 1 and 2 (*SPRY1* and *SPRY2*) (Sayed et al. 2008; Thum et al. 2008; Feng et al. 2012), *E-cadherin* (Kang et al. 2015), and *SEC23A* (Li et al. 2016). miR-21 is also involved in cancer stem cells (CSCs) stemness and cell cycle. For instance, FOLFOX (a combination of 5-FU, leucovorin and oxaliplatin)-resistant HT-29 and HCT-116 CSCs show suppression of *PDCD4* and has an up to 7-fold increase of pre or mature miR-21 (Yingjie Yu et al. 2012). The stable over-expression of miR-21 in the HCT-116 cell line down-regulates *PDCD4* and *TGF $\beta$ R2* by binding to 3'UTR sequences, suppresses c-Myc and Cyclin D1, and increases stemness via the Wnt/ $\beta$ -catenin signalling pathway (Yingjie Yu et al. 2012). The presence of circular RNA - circRNA-ACAP2 acts as an miRNA sponge to regulate *TIAM1* expression by removing the inhibitory effect of miR-21-5p, affecting cell proliferation, migration and invasion (He et al. 2018). An integrative combination meta-analysis and bioinformatics analysis indicated that *ABCB1*, *HPGD*, *BCL2*, *TIAM1* and *PDCD4* are the most common targets of miR-21 in CRC patients and are potentially biomarkers for diagnosis and prognosis (Saheb Sharif-Askari et al. 2020).

**Table 9.** The reported microRNA-21 studies with the corresponding direct target genes in gastrointestinal diseases.

Target Gene	Gene Name	Main Biological roles	GI Disease [Reference]
<b>Apoptosis</b>			
<i>CDH1</i>	Cadherin 1 ( <i>E-cadherin</i> )	An independent predictor of early tumour relapse	<b>CRC</b> (Kang et al. 2015)
<i>ROHB</i>	Ras homolog gene family, member B	Proliferation, invasion & apoptosis	<b>CRC</b> (Liu et al. 2011)
<i>PDCD4</i>	Programmed cell death 4	ROS promotes gastric carcinogenesis inflammation	<b>CRC</b> (Asangani et al. 2008; Peacock et al. 2014; Lu et al. 2008)
<b>Cell Cycle</b>			
<i>CDC25A</i>	Cell Division Cycle 25A	Negatively regulates G(1)-S transition, involve in DNA damage-induced G(2)-M checkpoint	<b>CRC</b> (Wang et al. 2009)
<i>TGFBR2</i>	Transforming growth factor, beta receptor 2	Induce stemness	<b>CRC</b> (Y. Yu et al. 2012)
<i>PTEN</i>	Phosphatase and tensin homolog	Regulate cell cycle, induces resistance of interferon- $\alpha$ /5-FU & metformin	<b>CRC</b> (Feng et al. 2012)
<b>Cytoplasmic vesicle</b>			
<i>SEC23A</i>	Sec23 Homolog A, COPII Coat Complex Component	Proliferation, migration, & invasion	<b>CRC</b> (Li et al. 2016)
<b>Growth Factor</b>			
<i>SPRY2</i>	Sprouty RTK Signalling Antagonist 2	Enhance the cytotoxic effect of 5-FU & metformin	<b>CRC</b> (Feng et al. 2012)
<b>Immune Response</b>			
<i>BCL2</i>	B-cell lymphoma 2	Increases beta cell death	<b>CRC</b> (Sims et al. 2017)
<i>TIAM1</i>	T-cell lymphoma invasion and metastasis 1	CircRNA-ACAP2/miR-21-5p/TIAM1 feedback circuit affects proliferation, migration & invasion	<b>CRC</b> (Cottonham et al. 2010; He et al. 2018)
<b>Metabolism</b>			
<i>HPGD</i>	15-Hydroxyprostaglandin Dehydrogenase	COX-2-dependent mechanism	<b>CRC</b> (Monteleone et al. 2019)
<i>DUSP8</i>	Dual Specificity Phosphatase 8	Repress cell growth & metastasis; alter AKT & ERK signalling	<b>CRC</b> (Ding et al. 2018)

CRC, colorectal cancer; NASH, non-alcoholic steatohepatitis, NAFLD, non-alcoholic fatty liver disease; GC, gastric cancer; HCC, Hepatocellular carcinoma; NK, natural killer; 5-FU, Fluorouracil.

### C.microRNA-221 in Colorectal Cancer

miR-221 is also one of the highly-expressed miRNAs in CRC tissues compared to the para-tumour tissues (Di Martino et al. 2016). The level of miR-221 positively correlates to local invasion, advanced tumour node metastasis and shorter patient survival. The depletion of miR-221 diminishes colony formation, cell migration, invasion and proliferation in both *in vivo* and *in vitro* CRC models (**Table 10**)(Tao et al. 2014; Liu et al. 2014; Qin, Luo 2014; Liao et al. 2018; Liu et al. 2018). miR-221 could directly bind to reversion-inducing-cysteine-rich protein with Kazal motifs (*RECK*) 3'UTR to promote CRC cell invasion and metastasis (Qin, Luo 2014). Silencing of miR-221 enables induction of apoptosis and autophagy via Cyclin-Dependent Kinase Inhibitor 1C (*CDKN1C*) and tumour protein 53-induced nuclear protein 1 (*TP53INP1*) respectively, and suppresses the S-phase and overexpresses the G<sub>0</sub>/G<sub>1</sub> population (Sun et al. 2011; Liao et al. 2018). A dose-dependent X-radiation affects the expression of miR-221 in Caco2 CRC cells and suppresses the *PTEN* expression. Inhibition of miR-221 sensitises radiosensitivity (Xue et al. 2013). In addition, the use of microcystin-LR promotes DLD-1 CRC cell migration through the miR-221/*PTEN*, as well as *STAT3* signalling with the accumulation of  $\beta$ -catenin in nuclei (Ren et al. 2019).

miR-221 and miR-222 are encoded from a tandem gene cluster located on chromosome Xp11.3, and thus co-expression of both miRNAs has been well studied. A three-dimensional (3D) cell culture study indicated *KRAS* mutation induces the expression of miR-221/222 and suppresses PTEN protein expression (Tsunoda et al. 2011). Over-expression of miR-221/222 by using miRNA mimics activates *NF-κB* and *STAT3* and binds to the 3' UTR of *PDLIM2* – a nuclear ubiquitin E3 ligase for RelA and *STAT3*. Injection of lentivirus-expressing miR-221/222 sponges suppresses CRC tumour formations in AOM/DSS-induced CRC mice, with fewer Ki67-positive cells compared with the control group (Ebert et al. 2007; Liu et al. 2014).

**Table 10.** Reported microRNA-221 studies with the corresponding direct target genes in gastrointestinal diseases.

Target Gene	Gene Name	Main Biological roles	GI Disease [Reference]
<b>Apoptosis</b>			
<i>GAS5</i>	lncRNA	Induces G0/G1 arrest & apoptosis	<b>CRC</b> (Yang Yang et al. 2017; Liu et al. 2018)
<b>Autophagy</b>			
<i>TP53INP1</i>	Tumour protein p53-inducible nuclear protein 1	Inhibit autophagy	<b>CRC</b> (Liao et al. 2018)
<b>Cell Cycle</b>			
<i>PTEN</i>	Phosphatase and Tensin Homolog	Cell proliferation, invasion & sphere formation to increase radio-sensitivity; down-regulate by microcystin-LR, radiation & Oroxin B	<b>CRC</b> (Xue et al. 2013; Ren et al. 2019; Tsunoda et al. 2011)
<b>Immune Response</b>			
<i>PDLIM2</i>	PDZ And LIM Domain 2	Degrade <i>RelA</i> & <i>STAT3</i>	<b>CRC &amp; IBD</b> (Liu et al. 2014)
<b>Metalloendopeptidase Activity</b>			
<i>RECK</i>	Reversion Inducing Cysteine Rich Protein with Kazal Motifs	Promote cancer cell invasion & metastasis	<b>CRC</b> (Qin, Luo 2014)

HBV; Hepatitis B virus, CRC, colorectal cancer; GC, gastric cancer; HCC, Hepatocellular carcinoma; 5-FU, Fluorouracil; EMT, Epithelial-mesenchymal transition; NK, natural killer.

### iii. GUT MICROBES AS NON-INVASIVE BIOMARKERS FOR COLORECTAL CANCER SCREENING

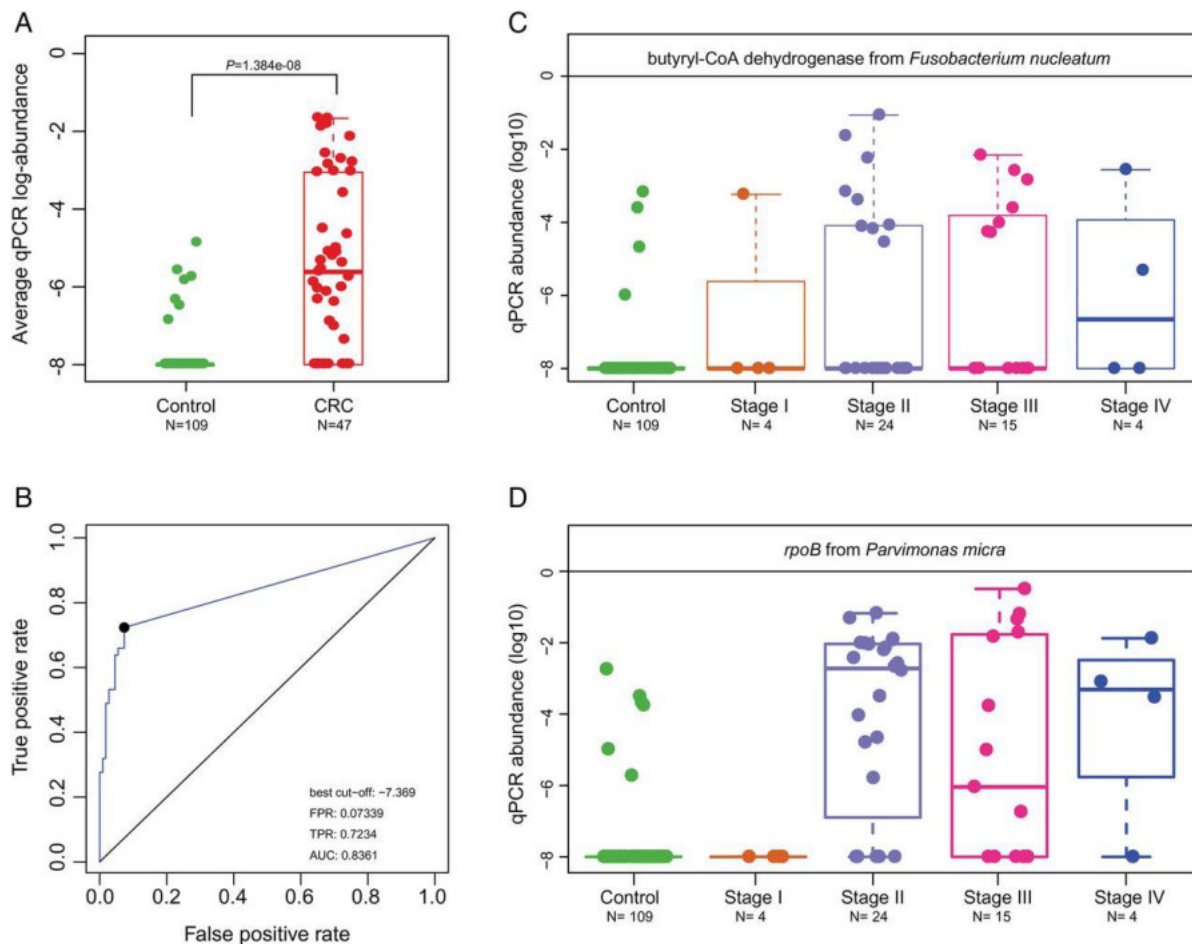
#### (1). Introduction

Besides faecal-based miRNAs, gut microbiota dysbiosis and the corresponding microbial composition changes have been recognised in a variety of diseases including the pathogenesis of CRC. These patterns could be used as microbial diagnostic biomarkers for potential CRC screening tool development.

#### (2). Summary of the outcomes

At the beginning of the study, our early shotgun metagenomics analysis from CRC patients firstly identified 20 bacteria abundant markers in CRC patients compared to controls, these are including *Fusobacterium nucleatum* (*F. nucleatum*), *Parvimonas micra* (*P. micra*,

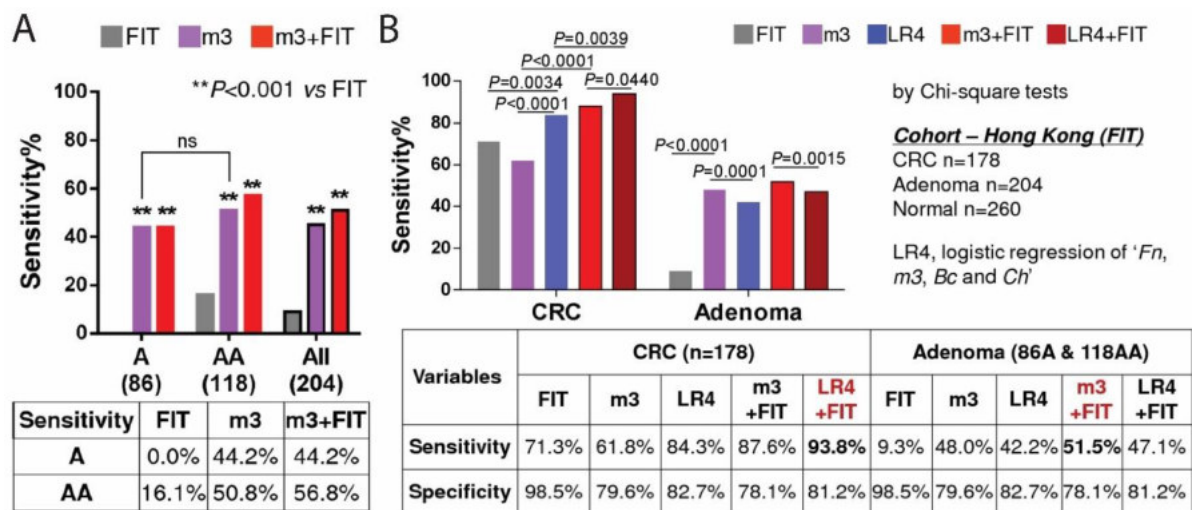
formerly *Peptostreptococcus micros*), *Gemella morbillorum*, *Peptostreptococcus anaerobius* and *Solobacterium moorei*. The selected markers were trans-continental validated in the Danish cohort. Further validation of the four gene markers in published metagenomes CRC cohorts from the French and Austrian have the AUC of 0.72 and 0.77, respectively. qPCR abundance of two gene markers (combined butyryl-CoA dehydrogenase from *F. nucleatum* and RNA polymerase subunit  $\beta$ , *rpoB*, from *P. micra*) separates CRC microbiomes from controls in an independent cohort consisting of 47 cases and 109 healthy controls, with the odds ratio of 23 and AUC of 0.84 (**Figure 25**)(**Publication 10, page 185**)(J. Yu et al. 2017).



**Figure 25. Validating robust gene markers associated with colorectal cancer in quantitative PCR (qPCR).** The abundance of two gene markers (butyryl-CoA dehydrogenase from *Fusarium nucleatum* and RNA polymerase subunit beta, *rpoB*, from *Parvimonas micra*) were further evaluated in 47 CRC cases and 109 healthy controls. (**A**) The combined log<sub>10</sub> abundance of the two genes enable to distinguish the CRC patients from the control group, and (**B**) classification of the CRC microbiomes with an area under the receiver operating characteristic curve of 0.84. (**C-D**) The two markers presented relatively high abundance in stage II and III CRC compared to the control and stage I. Zero abundance is plotted on a log<sub>10</sub> scale as -8. FPR, false positive rate; TPR, true positive rate AUC; area under the receiver operating curve. Adopted from (J. Yu et al. 2017).



Then, further additional metagenomic analysis identified *Clostridium hathewayi* (*C. hathewayi*) *F. nucleatum*, and “m3” from a *Lachnoclostridium sp.* significantly enriched in adenoma. Faecal m3 and *F. nucleatum* were significantly increased from normal to adenoma to CRC (linear trend by one-way ANOVA:  $p < 0.0001$ , in group I ( $n = 698$ ) and group II ( $n = 313$ ) cohorts). *F. nucleatum* showed better in detecting CRC (AUC: *F. nucleatum* = 0.862 vs m3 = 0.741,  $p < 0.0001$ ), while 'm3' may performed better than *F. nucleatum* in categorising adenomas in the control group (AUC: m3 = 0.675 vs *F. nucleatum* = 0.620,  $p = 0.09$ ) (Liang et al. 2020). By setting a specificity threshold of 78.5%, *F. nucleatum* and m3 had sensitivities of 77.8% and 62.1% for CRC and 33.8% and 48.3% for adenoma, respectively. In a subgroup Hong Kong cohort comparison with FIT ( $n = 642$ ), m3 performed better in detecting adenomas and advanced adenomas, with sensitivities of 44.2% and 50.8% at 79.6% specificity, compared to 0% and 16.1% for FIT, at 98.5% specificity. Combining m3 with FIT increased the sensitivity of advanced adenomas to 56.8% (**Figure 26A**). Finally, the combination of m3 with *F. nucleatum*, *C. hathewayi*, *B. clarus* and FIT performed best in the diagnosis of CRC with a specificity of 81.2% and a sensitivity of 93.8% (**Figure 26B**)(**Publication 11, page 196**) (Liang et al. 2020).



**Figure 26. Comparison and combination of bacterial markers with faecal immunochemical test (FIT).** (A) Comparison of the sensitivities of FIT, m3 and their combination in detecting non-advanced and advanced adenomas. (B) Comparison of sensitivity and specificity of FIT, m3, the combination of four makers (LR4: m3 + *F. nucleatum* + *C. hathewayi* + *B. clarus*) and combination of bacterial markers with FIT in a subgroup from Hong Kong. LR4 combined with FIT performed best for colorectal cancer (CRC) detection, while m3 combined with FIT performed best for detecting adenoma. All comparison of sensitivities was conducted by  $X^2$  tests. A, non-advanced adenoma; AA, advanced adenoma. Adopted from (Liang et al. 2020).

### (3). Discussions and Conclusions

In addition to faecal miRNA, other investigative targets, including gut microbes are one of the potential markers that could be used to improve CRC screening accuracy. Here, we

showed several faecal bacterial markers that can be used to distinguish between healthy subjects and CRC patients and m3 could be able for adenoma. In the meantime, a new panel of faecal bacterial markers (m3 + *F. nucleatum* + *C. hathewayi* + *B. clarus*) enhanced diagnostic power for CRC compared to the microbial markers reported in our studies in the past. The carcinogenicity of some overgrown bacteria identified in our studies, such as *P. micra* and *F. nucleatum* has been studied (**Publication 11, page 196**)(Liang et al. 2020).

*P. micra* is a Gram-positive anaerobic cocci species known as a commensal bacterium in the human oral cavity (García Carretero et al. 2016). The overgrowth *P. micra* has been studied in several CRC cohorts in both faeces and tissues and has been reported as a potential non-invasive screening biomarker (Löwenmark et al. 2020; J. Xu et al. 2020; Flemer et al. 2018). Functional studies showed that *P. micra* induces cell proliferation in CRC cell lines and germ-free mice; it also increases Ki-67<sub>positive</sub> cells and PCNA protein expression, and pro-inflammatory cytokines (TNF $\alpha$ , IL17a, IL6 and CXCR1) in the germ-free mice model (Zhao et al. 2022). Flow cytometry analyses identified Th2 and Th17 immune cell populations are higher, while Th1 cells were reduced in the lamina propria in *P. micra* mice. In the *Apc*<sup>min/+</sup> mice model, *P. micra*-driven CRC is significantly higher in tumour burden and tumour load compared to mice gavage with non-pathogenic *E. coli* or non-bacterial control (Zhao et al. 2022). *F. nucleatum* can be found in approximately 70% of CRC patients (J. Yu et al. 2017). It adheres and invades colonic epithelial cells, binds to E-cadherin via FadA to activate  $\beta$ -catenin signalling and regulates the inflammatory and oncogenic responses (Rubinstein et al. 2013). A higher abundance of *F. nucleatum* induces tumour multiplicity through the recruitment of tumour-infiltrating myeloid cells to generate a pro-inflammatory microenvironment (Kostic et al. 2012). Virulence factors such as FadA and Fap2 proteins from *F. nucleatum*, *Bacteroides fragilis* (*B. fragilis*)-produced toxins and the bacteria wall extracts, have also been identified and play as influential modulators in the evolution of normal colonic epithelial cells to tumour cells (Gholizadeh et al. 2017; Boleij et al. 2015; Mima et al. 2016; Abed et al. 2016).

In addition to the known and/or culturable microbes, there are a number of “unculturable” microbes and other identified gut bacteria that may need to be further investigated for their functions and the correlations in the role (s) of CRC and/or other colonic diseases.

#### **iv. GUT-HOST INTERACTION**

##### **(1). Introduction**

Whilst communicable diseases caused by infectious microbes continue, non-communicable diseases (NCDs) have become a major cause of morbidity and mortality in

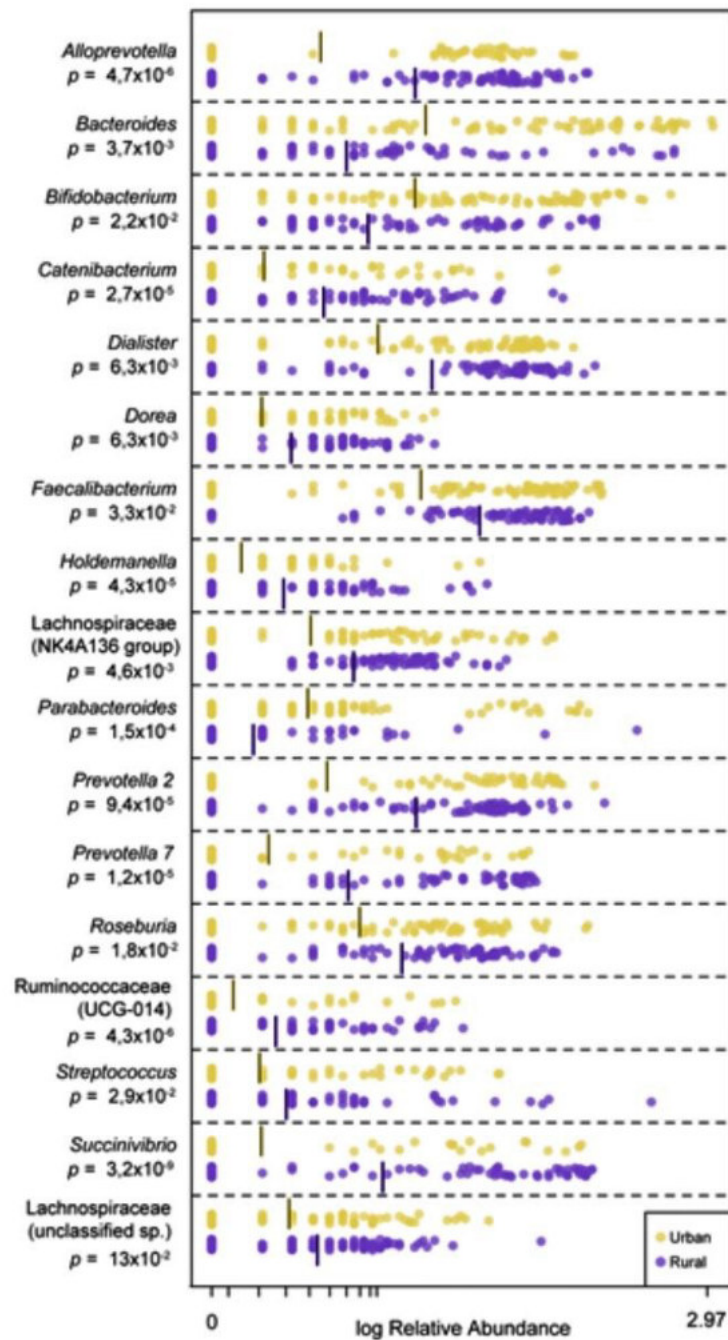
many countries (India State-Level Disease Burden Initiative Collaborators 2017; Arokiasamy 2018; Mohan et al. 2019). Perturbation of host-microbiome interactions may be a key mechanism by which lifestyle-related risk factors, for example, alcohol consumption, tobacco use, and physical inactivity influence metabolism caused by metainflammation (Furman et al. 2019; Lumeng, Saltiel 2011). Metainflammation contributes to the development of many NCDs, such as diabetes, which has increased rapidly in India over the past 25 years, rising to 65 million prevalent cases in 2016 from 26 million in 1990 (India State-Level Disease Burden Initiative Diabetes Collaborators 2018). Therefore, there is clearly an urgent need to identify relevant metabolic disorder traits to predict the risk of metabolic disorders and their associated diseases (Anjana et al. 2017).

Furthermore, once patients have been diagnosed with digestive diseases, gastrointestinal treatments may be required (Janarthanan et al. 2012; Roughead et al. 2016) and thus could suppress or interfere with the patients' immune system and disrupt gut flora homeostasis, creating a suitable micro-ecosystem for overgrowth of *C. difficile* (Rineh et al. 2014). In addition to antibiotics, FMT is an alternative treatment strategy for CDI patients with increasing evidence to support the use of FMT in SFCDI. However, the multifactorial mechanisms that underpin the efficacy of FMT are not fully understood. Other than CDI in colorectum, EBV infection in stomach should not be negated and thus functional studies are required to have a better understanding of the mechanisms including the epigenetic changes of the oncogenic viral infection (**Publication 4, page 119**)(Yau, Tang, et al. 2014).

## **(2). Summary of the outcomes**

### **A. Host-microbe and Metabolic Interactions are Differentially Shaped by Geographic location and Body Weight**

In comparison between countryside and municipal areas of central India, the study revealed metabolic and host-microbe interactions are differentially shaped by body weight and geographical location, multiple hallmarks of dysmetabolism were identified in urbanites and young overweight adults, the majority of whom did not have a known diagnosis of diabetes (**Publication 12, page 208**)(Tanya M. Monaghan, Biswas, et al. 2021). To be more specific, the faecal taxonomic composition profile revealed a number of overrepresented genera belonging to the *Firmicutes* phylum in the rural population, including *Faecalibacterium*, *Roseburia*, *Ruminococcaceae* and unclassified *Lachnospiraceae* by using Linear discriminant analysis Effect Size (LEfSe). *Prevotella* and *Alloprevotella* are dominant in the rural microbiota while *Bacteroides* and *Parabacteroides* are overrepresented in the urban microbiota (**Figure 27**).



**Figure 27. The microbiota is structurally distinct in participants from rural vs. urban areas.** Log-transformed relative abundance of significantly differential genera between participants from rural or urban areas, as determined by Linear discriminant analysis Effect Size (LEfSe). Adopted from (Tanya M. Monaghan, Biswas, et al. 2021).

The specific variation in serum short-chain fatty acids (SCFAs) - 2-hydroxybutyrate levels were positively correlated with IgG4 levels in the rural group ( $p < 0.05$ ), and IgG4 was strongly positively associated with *Campylobacter*, *Gemella*, *Leptotrichia*, *Neisseria*, *Porphyromonas* and *Streptococcus* ( $p < 0.0001$ ). Tetrasialylated and tetragalactosylated serum glycans (pathogenic complex glycans) were positively associated with serum caproate in the urban residents as well as *Holdemania* and *Klebsiella* in the rural residents, showing a

potentially diabetogenic role for the serum metabolite and genera. Also, glycated serum protein (GSP) levels were assessed (n = 135) and significantly higher in pre-obese (BMI 25–29.9) and overweight (BMI 23–24.9) individuals compared to normal (BMI 18.5–22.9), and in urban participants compared to rural subjects ( $p < 0.001$ ). Across all the subjects, high GSP levels were associated with significantly lower circulating IgG2, IgM, caproate, and valerate levels, and lower relative abundance of *Roseburia* and *Dorea* ( $p < 0.05$ ) (**Publication 13, page 232**)(Tanya M. Monaghan, Biswas, et al. 2021).

## B. Faecal Microbial Transplantation is an effective treatment strategy for *Clostridioides difficile* infection

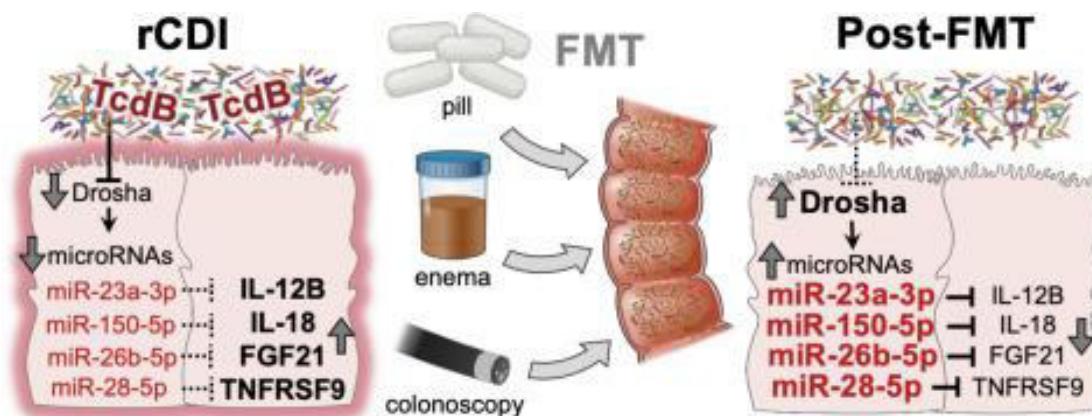
FMT is highly effective for recurrent CDI and there is growing evidence to support FMT for SFCDI. In this early exploratory longitudinal analysis study using a small sample size (3 responders and 1 non-responder for sequential FMT), a total of 562 characteristics were used for analysis, 78 of which were significantly different at all time points ( $p < 0.05$ ). This is including *Acidaminococcaceae*, *Phascolarctobacterium*, *Enterobacteriaceae*, *Pseudocitrobacter*, *Enterococcaceae* and *Enterococcus*, which are higher in the FMT non-responder compared to the responders (**Table 11**)(Tanya M Monaghan, Duggal, et al. 2021). This multi-omics study highlights initial novel features of dynamic phenotypic changes in FMT non-responders under different expectations, which prompted an in-depth study focusing on the molecular and immunological mechanisms of miRNA alterations in successful FMT in patients with recurrent CDI.

**Table 11.** Statistically significant threshold difference between responders and non-responder on four severe or fulminant *Clostridioides difficile* infection patients (three responders and one non-responder) in the taxonomy analysis (average across all time points). Adopted from (Tanya M Monaghan, Duggal, et al. 2021).

Taxa	Classification	P-value	log2 Fold Change (Succ/Fail)	Responders (Mean Value)	Non-Responder (Value)
<i>Acidaminococcaceae</i>	Family	0.0212	-3.6351	0.1625	2.0193
<i>Phascolarctobacterium</i>	Genus	0.0212	-3.6351	0.1625	2.0193
<i>Enterobacteriaceae_unclassified</i>	Genus	0.0013	-2.2113	1.0058	4.6577
<i>Pseudocitrobacter</i>	Genus	0.0080	-2.2079	0.7009	3.2383
<i>Enterococcaceae</i>	Family	0.0035	-1.8467	1.4509	5.2185
<i>Enterococcus</i>	Genus	0.0035	-1.8467	1.4509	5.2185

For the in-depth study, sera from two prospective multicentre randomised controlled trials (NCT02254811 and NCT01398969) for miRNA profiling on the Nanostring nCounter platform uncovered 64 upregulated circulating miRNAs at 4 and 12 weeks after FMT compared with screening (pre-treatment phase), of which the top 6 miRNAs were validated in the discovery cohort by means of RT-qPCR. In a murine model of rCDI, RT-qPCR analyses of cecal and sera RNA extracts proved repression of these miRs, an effect reversed by FMT. In mouse colons and human colonoids, *C. difficile* TcdB mediated the suppressive effects of

CDI on miRNAs. CDI dysregulated DROSHA, but not Ago2 or Dicer1, and its effect can be reversed by FMT. Correlation analyses, qPCR, and 3'UTR reporter assays revealed that miR-26b, miR-28, miR-23a and miR-150, target directly the 3'UTRs of FGF21, TNFRSF9, IL12B and IL18, respectively. miR-23a and miR-150 demonstrated cytoprotective effects against TcdB on colonic epithelial cells. These results provide novel and provocative evidence that modulation of the gut microbiome via FMT induces alterations in circulating and intestinal tissue miRs and identify new potential targets for therapeutic intervention in rCDI (**Figure 28**) (**Publication 14, page 258**)(Tanya M. Monaghan, Seekatz, et al. 2021).

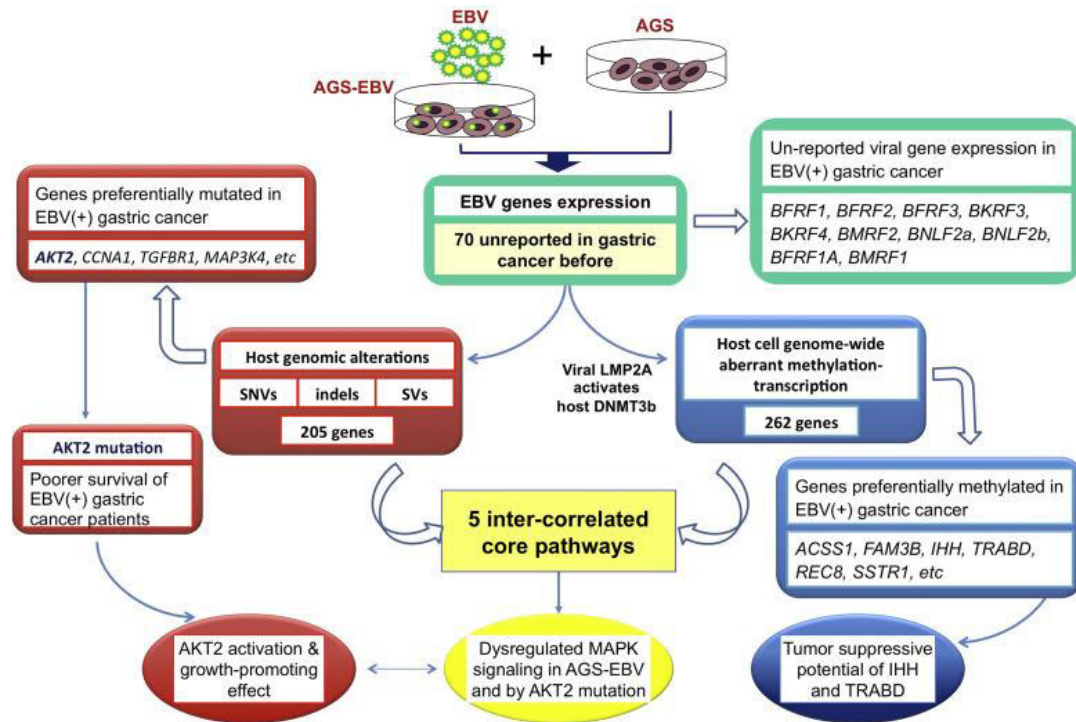


**Figure 28.** Summary of findings on miRNA changes in patients with *Clostridium difficile* infection after receiving faecal microbial transplants. rCDI; recurrent *Clostridioides difficile*, FMT, Faecal microbiota transplantation. Adopted from (Tanya M. Monaghan, Seekatz, et al. 2021).

### C. Genome-wide, Transcriptomic and Epigenomic level changes in Epstein-Barr virus-associated gastric cancer

In the EBV-driven GC study, a number of sequencing approaches including whole-genome, transcriptome, and epigenome sequencing were applied on EBV infected vs control AGS GC cell line. Integrated epigenome and transcriptome analyses identified 216 genes transcriptionally down-regulated by EBV-associated hypermethylation; methylation of ACSS1, FAM3B, IHH, and TRABD increased significantly in EBVaGC. Overexpression of Indian hedgehog (IHH) and TraB domain containing (TRABD) increased proliferation and colony formation of GC cells, whereas knockdown of these genes reduced these activities. Genomic analyses spotted AKT2 from the 44 mutation genes as it has 2 nonsynonymous point mutations in AGS–EBV cells and is linked with poorer patient survival in EBVaGC. AKT2 and phosphate-AKT2 also have a higher expression in EBVaGC cell lines and up-regulate the down-stream MAPK signalling of ERK and activator protein 1 (AP-1), eventually inducing GC cell proliferation. AKT2 mutation associates with poor patient survival in EBVaGC. AKT2 is a

putative oncoprotein that participates in important cancer pathways; our study uncovered EBVaGC, mainly involved in focal adhesion and MAPK signalling (Liang et al. 2014). In addition to the well-documented EBV viral gene markers, such as *EBNA1*, *EBER1* and *EBER2* found in EBVaGC patients, we reported 10 different EBV genes expressed in EBVaGC patients (**Figure 29**)(**Publication 15, page 277**)(Liang et al. 2014).



**Figure 29. Summary of the study outcomes in Epstein-Barr virus associated gastric cancer.** EBV gene expression profile, EBV<sub>positive</sub> host genomic and epigenomic alterations were identified in a cell model and validated further in primary EBV(+) gastric cancers. Adopted from (Liang et al. 2014).

### (3). Discussions and Conclusions

There is substantial evidence that the gut microbiome is associated with interrelated physiological parameters of different gastrointestinal diseases such as CDI and EBVaGC, as well as metabolic diseases including diabetes. Currently, a relatively few deep phenomic studies have been undertaken and thus a better understanding of the biological processes associated with healthy individuals and those potentially at risk of diabetes.

By using the Central-Indian study cohort, the serum *N*-glycan profiles showed a more complex glyco-phenotype in urban population, which may be associated with a higher risk of developing T2DM and poorer blood glucose regulation (Rudman et al. 2019). The findings of SCFAs in our study were previously observed in the non-industrial microbiome, with more SCFAs present in human faeces and greater diversity of genes engaged in complex carbohydrate metabolism compared to the industrial microbiome (Ou et al. 2013; Obregon-Tito et al. 2015). Multi-hallmarks of dysmetabolism were identified in urbanites and young

overweight adults from a higher burden of diabetic-related proteins and a complex glyco-phenotype in the circulation, the majority of whom did not have a known diagnosis of diabetes. It is important to identify people at high risk of diabetes as early intervention may delay or even prevent overt diabetes. By unravelling the immunometabolic interactions between gut microbiome and host, personalised therapeutic approaches, including prebiotics, probiotics, postbiotics and synbiotics could be utilised and investigated to prevent or even treat cardiometabolic-related diseases (Vallianou et al. 2019; Barengolts 2016).

In the context of rCDI studies, our findings demonstrate the synergistic regulation of miRNA expression by FMT, confirming several significant alterations in circulating miRNAs following successful FMT treatment in two independent cohorts of rCDI patients. The miRNA signature was further validated in rCDI mouse models and human colonoids. miRNA processing in colonic epithelial cells was directly altered by *C. difficile* toxin and may be affected by *C. difficile*-associated dysbiosis. Conditional knockdown of the miRNA processing enzyme Dicer in murine intestinal epithelial cells has been shown to regulate the intestinal microbiota and exacerbate colitis. From this observation, we found that FMT-regulated miRNAs can control cellular properties and target TNFRSF9, FGF21, IL12B and IL18, all of which are integral to pathways associated with cancer, inflammation and autoimmunity, and that CDI-induced colitis can suppress circulating miRNAs, which can be restored by FMT. These findings may be further explored to investigate the potential of rCDI as a therapeutic target.

Regarding the study on EBVaGC, gene expressions in EBVaGC cell was shown by transcriptome analysis and confirmed in EBVaGC primary gastric tumours. Whole-genome sequencing showed EBV-associated host mutations in genes including AKT2, CCNA1, MAP3K4, and TGFBR1. AKT2 mutations are associated with reduced survival times of patients with EBVaGC. Epigenome analysis uncovered hypermethylation of genes including ACSS1, IHH, FAM3B, and TRABD through the infection. Five core pathways were shown to be dysregulated by EBV-associated host genomic and epigenomic aberrations in GC. These findings provide a systematic view of EBV-associated host genomic and epigenomic abnormalities and signalling networks that may govern the pathogenesis of EBVaGC.

## VII. OVERALL DISCUSSION AND FUTURE DIRECTIONS

Over the past century, an ageing population coupled with a shift to a Western lifestyle has left many people vulnerable to interrelated digestive disorders and there is an urgent need to understand the pathogenesis and mechanisms of disease in order to improve diagnostic



accuracy and treatment outcomes. First, by integrating publicly available datasets with bioinformatics analysis, I identified hyperactive neutrophil chemotaxis could be the major factor for anti-TNF- $\alpha$  treatment resistance in IBD patients. Then, utilising faecal miRNA and gut microbes, several upregulated miRNA and over-abundance of gut microbes in CRC were identified. In addition, studies regarding FMT confirmed the alternation of host-miRNA expression and relief of the infection symptoms in rCDI patients; EBV<sub>positive</sub> GC has clear molecular changes at whole-genome, transcriptome, and epigenome levels.

Currently, in non-invasive CRC screening studies, our faecal-based miRNA and gut microbe biomarker discoveries focus on the general population in the prevention of cancer development, however, the strategy for other diseases and/or other conditions in clinical applications should not be neglected, such as chemotherapy. For instance, miR-20a dysregulation has been reported in several chemo-resistance studies, including CRC. miR-20a directly targets apoptosis signal-regulating kinase 1 (*ASK1*) in cisplatin resistance CRC cells (L. Zhang et al. 2018) and BCL2 Interacting Protein 2 (*BNIP2*) in 5-Fluorouracil (5-FU), oxaliplatin and teniposide resistance CRC cells (Chai et al. 2011). Long non-coding RNA (lncRNA) *HAND2-AS1* acts as a miR-20a sponge to control the expression of miR-20a target genes in 5-FU chemotherapy-resistance CRC cells (Jiang et al. 2020). In addition, a recent study indicated that the enrichment of *F. nucleatum* regulates the autophagy signalling pathway through *ATG7* and *ULK1* to increase chemo-resistance to CRC (T. Yu et al. 2017). *ATG7* and *ULK1* can be targeted by hsa-miR-4802 and hsa-miR-18a-3p, respectively, through initiating TLR4–MyD88 activation and thus avoiding undergoing chemotherapy-caused apoptosis (T. Yu et al. 2017). Hence, our findings on miR-18a and *F. nucleatum* could also be further expanded to investigate the prediction potential for chemotherapeutic treatment response in CRC patients, and additional studies are thus required.

More recently, loop-mediated isothermal amplification (LAMP) also offers extensive research, and detection opportunities for target-specific amplification under isothermal conditions, especially samples from a complicated environment such as soil and faeces. LAMP typically requires recognising 6-8 various regions by using 4-6 primers to target a specific DNA or cDNA for an amplification reaction. This requires a DNA polymerase with strand-displacing ability to begin synthesis and two specially designed primers to form 'loop' structures that facilitate subsequent rounds of amplification by extension on the loop and additional annealing of the primers. The endpoint LAMP DNA products can exceed 20kb and are formed by many repeats of short target sequences (typically between 80-250bp) linked to single-stranded loop regions. LAMP could be applied to real-time fluorescence assays using intercalators or probes, and both agarose gel and lateral flow assays are directly compatible

with LAMP reactions (Panno et al. 2020). The use of LAMP may present better accuracy compared to probe-based qPCR techniques, however, this may not apply well to faecal-based miRNAs due to the relatively short sequences.

In addition to increasing non-invasive CRC screening and the treatment accuracy for IBD patients to prevent colonic polyps development, studies on diet-microRNAs-microbiome interactions may also be able for us to investigate disease prevention strategies including CRC. For instance, high-fat diets have been considered a risk factor for CRC. Studies have shown that high-fat diets stimulate the production of hepatic primary bile acids and transfer them into the colon, which is then metabolised by gut microbes into genotoxic bile acids such as deoxycholic acid (DCA) (Ajouz et al. 2014). An increased level of 7 $\alpha$ -dehydroxylating intestinal bacterium has also been reported in high-fat diet studies (Ocvirk, O’Keefe 2017). DCA was shown to promote CRC through down-regulation of miR-199a-5p (Kong et al. 2012). The expression of miR-199a-5p in CRC cells may lead to the inhibition of tumour-cell growth and metastasis, and regulate EMT signalling by targeting *ROCK1* and *DDR1* (Zhu et al. 2018; Hu et al. 2014), restore therapeutic sensitivity through *CAC1* (Kong et al. 2012), and control cell apoptosis through HIF-1 $\alpha$  and VEGF (Ye et al. 2015).

On the other hand, SCFAs such as acetate, propionate, and butyrate are the end metabolite products of dietary fibres via saccharolytic fermentation in the gut. *Faecalibacterium prausnitzii* is one of the abundant bacterial species in the gut producing butyrate. Butyrate is a source of colonocytes and has a potential protective role against colonic diseases, such as IBD and CRC (Martín et al. 2017), by reducing mucosa inflammation through upregulating PPAR $\gamma$  (Schwab et al. 2007) and inhibiting NF- $\kappa$ B transcription factor activation (Inan et al. 2000) and IFN- $\gamma$  (Klampfer et al. 2003). Butyrate is also a histone deacetylase (HDAC) inhibitor that regulates *PTEN*, *BCL2L11* and *CDKN1A* at post-transcriptional level, presenting anti-tumour effects (Wawruszak et al. 2019; Hu et al. 2011). A number of dysregulated miRNAs in butyrate-treated cells have been reported. For instance, in CRC cells, butyrate can inhibit miR-92a in a cMyc-dependent manner, allowing p57 – a CDK inhibitor – to exert control over cell cycle progression (Hu et al. 2015); suppression of miR-106b by butyrate promotes the expression of tumour suppressor p21 (p21 acts on proliferating cell nuclear antigen (PCNA) to inhibit DNA replication and intermediate cell cycle arrest triggered by p53 (Huang et al. 2000; Schlörmann et al. 2015; Hu et al. 2011)); miR-203 was identified by butyrate in several CRC cell lines and down-regulate *NEDD9* – a scaffolding protein belonging to the Cas family – resulting in reduced cancer cell migration and invasion (Han et al. 2016; Shagisultanova et al. 2015).

In addition, diets themselves could regulate intestinal microbes and affect homeostasis. For instance, miRNAs are abundant in raw milk, which is resistant to RNase digestion; stable in normal storage and acidic conditions and freeze-thaw cycles (Kosaka et al. 2010; Weber et al. 2010; Izumi et al. 2012); and ~40% of miRNAs remain after pasteurisation and homogenisation processes (Howard et al. 2015). Milk miRNAs can tolerate the upper digestive tract and are attained by absorption in the intestine via exosome (Munagala et al. 2016). Also, plant miR-159 is rich in broccoli and stable under storage, processing and cooking conditions (Chin et al. 2016; Philip et al. 2015). A DSS-induced mouse colitis model showed that ginger exosome-like nanoparticle (GELN)-carried mdo-miR-7267-3p can target *Lactobacillus rhamnosus* monooxygenase coding gene *ycnE*; boost up the expression of indole-3-carboxaldehyde – a ligand for the aryl hydrocarbon receptor; and stimulate host IL-22 production, eventually relieving colitis symptoms through improving barrier function (Teng et al. 2018). By using *Lactobacillus acidophilus* and *Bifidobacterium bifidum* as probiotics in AOM-induced CRC mouse model, the researchers identified miR-135b, miR-155 and *KRAS* expression are increased; while miR-18a, miR-26b, APC, PU.1, and PTEN expressions are reduced (Heydari et al. 2019).

Currently, various studies have been focused on therapies that control gut microbiota composition; and antibiotics treating gut-related diseases is one of the most remarkable achievements. For instance, both vancomycin and metronidazole are commonly used in CDI and fidaxomicin may apply for recurrence or relapse CDI; vancomycin is also frequently used for IBD flares associated with *C. difficile* colitis. Rifaximin, ciprofloxacin and metronidazole can be applied in IBD, especially Crohn's disease (Ledder 2019; Czepiel et al. 2019). However, the extensive use of antibiotics has led to the global challenge of antibiotic resistance; thus, alternative treatment strategies, such as pro-/pre-biotics, bacteriophage, FMT, as well as miRNA, are under investigation. Bacteriophage (also called phages) therapy relies on the use of naturally-occurring and/or bio-engineered phages to infect and lyse target bacteria at the affected infection site(s). Its therapeutic efficiency relies on a particular setting and remains at an exploratory stage, due to its complex microbial communities and ethical challenges (Anomaly 2020). Phages may also translocate across the intestinal epithelium and subsequently circulate within the blood (Górski et al. 2006). The so-called “cocktail phage therapy” was first applied in the United Kingdom for a 15-year-old patient with cystic fibrosis with drug-resistant *Mycobacterium abscessus* infection (Dedrick et al. 2019); and the US FDA recently approved a phase one bacteriophage clinical trial for *Pseudomonas aeruginosa* infections (Voelker 2019).

On the other hand, gut dysbiosis could be changed by taking prebiotics – specialised dietary fibres that stimulate “friendly” bacteria growth, or probiotics – living “good” bacteria in food. In prebiotics, for example, the intake of lactose will break down the lactose in milk (-containing) products to alleviate the symptoms of lactose intolerance and maintain the balance of gut microbiota (Dastar et al. 2016). Lactose itself may contribute to the growth of *Bifidobacterium* and other lactic acid bacteria (Romero-Velarde et al. 2019); consuming milk in patients with lactose intolerance may decrease *Megamonas* and increase *Bifidobacterium*, *Anaerostipe* and *Blautia* (Li et al. 2018). On the other hand, there is some evidence that probiotics may ease some symptoms of irritable bowel syndrome (IBS) (Cristofori et al. 2021) and prevent diarrhoea when taking antibiotics (Stavropoulou, Bezirtzoglou 2020); while developing probiotic strains for treatment/prevention purposes requires intensive studies to ensure the possible treatment effect and safety. Moreover, there is increasing evidence that miRNAs regulate gut homeostasis; modulating gut microbiota via miRNAs is an unexplored therapeutic strategy. Studies to understand host-miRNA-microbiota communication are essential for the development of miRNA-based therapies. For example, a recent study identified a synthetic miRNA that can specifically modulate the microbiome and ameliorate an inflammatory autoimmune disease (Liu et al. 2019). Other hypotheses include using milk or plant as delivering vectors to transfer miRNAs that target specific gut bacteria (Rome 2019; Munagala et al. 2016).

## VIII. CONCLUSION

In conclusion, inter-connected GI diseases are the major global public health issues affecting almost the entire human population. The complex interactions among host epithelial cells, immunity, epigenetics and gut microbiota have been suggested as major factors in the maintenance of a healthy GI environment. The molecular changes, such as miRNA from disease(s) or infection, can be found in blood/faeces and modulate the distribution and composition of the gut flora microenvironment. These changes can potentially be a non-invasive tool to improve the detection accuracy of the current use of FIT for CRC screening or for recurrent CDI patients.

## IX. REFERENCES

- Abdelfattah, A.M., Park, C., Choi, M.Y., 2014. Update on non-canonical microRNAs. *Biomolecular Concepts*, 5(4), pp.275–287. 10.1515/bmc-2014-0012.
- Abed, J. et al., 2016. Fap2 Mediates *Fusobacterium nucleatum* Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. *Cell host & microbe*, 20(2), pp.215–25. 10.1016/j.chom.2016.07.006.
- Abraham, C., Cho, J.H., 2009. Inflammatory bowel disease. *The New England journal of medicine*, 361(21), pp.2066–78. 10.1056/NEJMra0804647.
- Adams, L., 2017. Pri-miRNA processing: structure is key. *Nature Reviews Genetics*, 18(3), pp.145–145. 10.1038/nrg.2017.6.
- Afonina, I.A. et al., 2002. Minor Groove Binder-Conjugated DNA Probes for Quantitative DNA Detection by Hybridization-Triggered Fluorescence. *BioTechniques*, 32(4), pp.940–949. 10.2144/02324pf01.
- Ahluwalia, B. et al., 2018. Immunopathogenesis of inflammatory bowel disease and mechanisms of biological therapies. *Scandinavian journal of gastroenterology*, 53(4), pp.379–389. 10.1080/00365521.2018.1447597.
- Ajouz, H., Mukherji, D., Shamseddine, A., 2014. Secondary bile acids: an underrecognized cause of colon cancer. *World Journal of Surgical Oncology*, 12(1), p.164. 10.1186/1477-7819-12-164.
- Alang, N., Kelly, C.R., 2015. Weight gain after fecal microbiota transplantation. *Open forum infectious diseases*, 2(1), p.ofv004. 10.1093/ofid/ofv004.
- Alkhayat, M. et al., 2020. Lower Rates of Colorectal Cancer in Patients With Inflammatory Bowel Disease Using Anti-TNF Therapy. *Inflammatory bowel diseases*. 10.1093/ibd/izaa252.
- AMBROS, V., 2003. A uniform system for microRNA annotation. *RNA*, 9(3), pp.277–279. 10.1261/rna.2183803.
- Amemori, S. et al., 2019. Sessile serrated adenoma/polyp showed rapid malignant transformation in the final 13 months. *Digestive endoscopy : official journal of the Japan Gastroenterological Endoscopy Society*. 10.1111/den.13572.
- Ando, Y. et al., 2016. Downregulation of MicroRNA-21 in Colonic CD3+ T Cells in UC Remission. *Inflammatory bowel diseases*, 22(12), pp.2788–2793. 10.1097/MIB.0000000000000969.
- Ang, P.W. et al., 2010. Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features. *BMC cancer*, 10, p.227. 10.1186/1471-2407-10-227.
- Anjana, R.M. et al., 2017. Prevalence of diabetes and prediabetes in 15 states of India: results from the ICMR-INDIAB population-based cross-sectional study. *The lancet. Diabetes & endocrinology*, 5(8), pp.585–596. 10.1016/S2213-8587(17)30174-2.
- Anomaly, J., 2020. The Future of Phage: Ethical Challenges of Using Phage Therapy to Treat Bacterial Infections. *Public health ethics*, 13(1), pp.82–88. 10.1093/phe/phaa003.
- Archambaud, C. et al., 2013. The intestinal microbiota interferes with the microRNA response upon oral *Listeria* infection. *mBio*, 4(6), pp.e00707-13. 10.1128/mBio.00707-13.
- Armaghany, T. et al., 2012. Genetic alterations in colorectal cancer. *Gastrointestinal cancer research : GCR*, 5(1), pp.19–27.
- Arokiasamy, P., 2018. India's escalating burden of non-communicable diseases. *The Lancet. Global health*, 6(12), pp.e1262–e1263. 10.1016/S2214-109X(18)30448-0.
- Asangani, I.A. et al., 2008. MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdc4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene*, 27(15), pp.2128–36. 10.1038/sj.onc.1210856.
- Balamayoran, G. et al., 2010. Mechanisms of neutrophil accumulation in the lungs against bacteria. *American journal of respiratory cell and molecular biology*, 43(1), pp.5–16. 10.1165/rcmb.2009-0047TR.
- Balamurugan, R. et al., 2008. Real-time polymerase chain reaction quantification of specific butyrate-producing bacteria, *Desulfovibrio* and *Enterococcus faecalis* in the feces of patients with colorectal cancer. *Journal of gastroenterology and hepatology*, 23(8 Pt 1), pp.1298–303. 10.1111/j.1440-1746.2008.05490.x.
- Balram, B. et al., 2019. Risk Factors Associated with *Clostridium difficile* Infection in Inflammatory Bowel Disease: A Systematic Review and Meta-Analysis. *Journal of Crohn's & colitis*, 13(1), pp.27–38. 10.1093/ecco-jcc/jjy143.
- Bao, Y. et al., 2019. Long noncoding RNA BFAL1 mediates enterotoxigenic *Bacteroides fragilis*-related carcinogenesis in colorectal cancer via the RHEB/mTOR pathway. *Cell Death & Disease*, 10(9), p.675. 10.1038/s41419-019-1925-2.
- Barengolts, E., 2016. Gut Microbiota, Prebiotics, Probiotics, and Synbiotics in Management of Obesity

- and Prediabetes: Review of Randomized Controlled Trials. *Endocrine Practice*, 22(10), pp.1224–1234. 10.4158/EP151157.RA.
- Bartel, D.P., 2018. Metazoan MicroRNAs. *Cell*, 173(1), pp.20–51. 10.1016/j.cell.2018.03.006.
- Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2), pp.281–97. 10.1016/s0092-8674(04)00045-5.
- Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2), pp.215–33. 10.1016/j.cell.2009.01.002.
- Belkaid, Y., Hand, T.W., 2014. Role of the Microbiota in Immunity and Inflammation. *Cell*, 157(1), pp.121–141. 10.1016/j.cell.2014.03.011.
- Ben-Horin, S., Kopylov, U., Chowers, Y., 2014. Optimizing anti-TNF treatments in inflammatory bowel disease. *Autoimmunity Reviews*, 13(1), pp.24–30. 10.1016/j.autrev.2013.06.002.
- Berezikov, E. et al., 2007. Mammalian Mirtron Genes. *Molecular Cell*, 28(2), pp.328–336. 10.1016/j.molcel.2007.09.028.
- Berns, M., Hommes, D.W., 2016. Anti-TNF- $\alpha$  therapies for the treatment of Crohn's disease: the past, present and future. *Expert Opinion on Investigational Drugs*, 25(2), pp.129–143. 10.1517/13543784.2016.1126247.
- Bettington, M. et al., 2013. The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology*, 62(3), pp.367–86. 10.1111/his.12055.
- Bettington, M.L. et al., 2015. A clinicopathological and molecular analysis of 200 traditional serrated adenomas. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 28(3), pp.414–27. 10.1038/modpathol.2014.122.
- Biegging, K.T., Mello, S.S., Attardi, L.D., 2014. Unravelling mechanisms of p53-mediated tumour suppression. [online]. *Nature reviews. Cancer*, 14(5), pp.359–70. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24739573>.
- Bocchetti, M. et al., 2021. The Role of microRNAs in Development of Colitis-Associated Colorectal Cancer. *International journal of molecular sciences*, 22(8). 10.3390/ijms22083967.
- Bohnsack, M.T., Czaplinski, K., Gorlich, D., 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA (New York, N.Y.)*, 10(2), pp.185–91.
- Boland, C.R., Goel, A., 2010. Microsatellite instability in colorectal cancer. *Gastroenterology*, 138(6), pp.2073-2087.e3. 10.1053/j.gastro.2009.12.064.
- Boleij, A. et al., 2015. The Bacteroides fragilis Toxin Gene Is Prevalent in the Colon Mucosa of Colorectal Cancer Patients. *Clinical Infectious Diseases*, 60(2), pp.208–215. 10.1093/cid/ciu787.
- Bray, F. et al., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6), pp.394–424. 10.3322/caac.21492.
- Burke, A.P. et al., 1990. Lymphoepithelial carcinoma of the stomach with Epstein-Barr virus demonstrated by polymerase chain reaction. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 3(3), pp.377–80.
- Bustin, S.A., Nolan, T., 2020. RT-qPCR Testing of SARS-CoV-2: A Primer. *International journal of molecular sciences*, 21(8). 10.3390/ijms21083004.
- Cadigan, K.M., Waterman, M.L., 2012. TCF/LEFs and Wnt signaling in the nucleus. *Cold Spring Harbor perspectives in biology*, 4(11). 10.1101/cshperspect.a007906.
- Camargo, M.C. et al., 2016. Validation and calibration of next-generation sequencing to identify Epstein-Barr virus-positive gastric cancer in The Cancer Genome Atlas. *Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association*, 19(2), pp.676–681. 10.1007/s10120-015-0508-x.
- Cancer Genome Atlas Network, 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407), pp.330–7. 10.1038/nature11252.
- Castellari, M. et al., 2012. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Research*, 22(2), pp.299–306. 10.1101/gr.126516.111.
- Cerrito, M.G., Grassilli, E., 2021. Identifying Novel Actionable Targets in Colon Cancer. *Biomedicines*, 9(5), p.579. 10.3390/biomedicines9050579.
- Chagneau, C. V. et al., 2019. The Polyamine Spermidine Modulates the Production of the Bacterial Genotoxin Colibactin D'Orazio, S. E. F., ed. *mSphere*, 4(5). 10.1128/mSphere.00414-19.
- Chai, H. et al., 2011. MiR-20a targets BNIP2 and contributes chemotherapeutic resistance in colorectal adenocarcinoma SW480 and SW620 cell lines. *Acta biochimica et biophysica Sinica*, 43(3), pp.217–25. 10.1093/abbs/gmq125.
- Cheloufi, S. et al., 2010. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature*, 465(7298), pp.584–589. 10.1038/nature09092.
- Chen, C., 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids*

- Research*, 33(20), pp.e179–e179. 10.1093/nar/gni178.
- Chen, F.-L. et al., 2017. Streptococcus sanguis bacteremia complicating endocarditis associated with colonic adenocarcinoma. *Journal of Microbiology, Immunology and Infection*, 50(3), pp.399–400. 10.1016/j.jmii.2016.11.002.
- Chen, W. et al., 2012. Human Intestinal Lumen and Mucosa-Associated Microbiota in Patients with Colorectal Cancer Moschetta, A., ed. *PLoS ONE*, 7(6), p.e39743. 10.1371/journal.pone.0039743.
- Chen, X.-Z. et al., 2015. Epstein-Barr virus infection and gastric cancer: a systematic review. *Medicine*, 94(20), p.e792. 10.1097/MD.0000000000000792.
- Chen, Y.-J. et al., 2016. MicroRNA-18a modulates P53 expression by targeting IRF2 in gastric cancer patients. *Journal of gastroenterology and hepatology*, 31(1), pp.155–63. 10.1111/jgh.13041.
- Cheng, D. et al., 2016. MicroRNA-20a-5p promotes colorectal cancer invasion and metastasis by downregulating Smad4. *Oncotarget*, 7(29), pp.45199–45213. 10.18632/oncotarget.9900.
- Chetty, R., 2016. Traditional serrated adenoma (TSA): morphological questions, queries and quandaries. *Journal of clinical pathology*, 69(1), pp.6–11. 10.1136/jclinpath-2015-203452.
- Childs, B.G. et al., 2014. Senescence and apoptosis: dueling or complementary cell fates? *EMBO reports*, 15(11), pp.1139–53. 10.15252/embr.201439245.
- Chin, A.R. et al., 2016. Cross-kingdom inhibition of breast cancer growth by plant miR159. *Cell research*, 26(2), pp.217–28. 10.1038/cr.2016.13.
- Choi, C.-H.R. et al., 2019. Cumulative burden of inflammation predicts colorectal neoplasia risk in ulcerative colitis: a large single-centre study. *Gut*, 68(3), pp.414–422. 10.1136/gutjnl-2017-314190.
- Chung, L. et al., 2018. Bacteroides fragilis Toxin Coordinates a Pro-carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. *Cell Host & Microbe*, 23(2), pp.203-214.e5. 10.1016/j.chom.2018.01.007.
- Cooks, T. et al., 2018. Mutant p53 cancers reprogram macrophages to tumor supporting macrophages via exosomal miR-1246. *Nature Communications*, 9(1), p.771. 10.1038/s41467-018-03224-w.
- Coppedè, F. et al., 2014. Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer. *World journal of gastroenterology*, 20(4), pp.943–56. 10.3748/wjg.v20.i4.943.
- Cottonham, C.L., Kaneko, S., Xu, L., 2010. MiR-21 and miR-31 Converge on TIAM1 to Regulate Migration and Invasion of Colon Carcinoma Cells. *Journal of Biological Chemistry*, 285(46), pp.35293–35302. 10.1074/jbc.M110.160069.
- Cougnoux, A. et al., 2014. Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut*, 63(12), pp.1932–1942. 10.1136/gutjnl-2013-305257.
- Cristofori, F. et al., 2021. Anti-Inflammatory and Immunomodulatory Effects of Probiotics in Gut Inflammation: A Door to the Body. *Frontiers in immunology*, 12, p.578386. 10.3389/fimmu.2021.578386.
- Crockett, S.D., Nagtegaal, I.D., 2019. Terminology, Molecular Features, Epidemiology, and Management of Serrated Colorectal Neoplasia. *Gastroenterology*, 157(4), pp.949-966.e4. 10.1053/j.gastro.2019.06.041.
- Czepiel, J. et al., 2019. Clostridium difficile infection: review. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*, 38(7), pp.1211–1221. 10.1007/s10096-019-03539-6.
- D’Haens, G.R., van Deventer, S., 2021. 25 years of anti-TNF treatment for inflammatory bowel disease: lessons from the past and a look to the future. *Gut*. 10.1136/gutjnl-2019-320022.
- Dalmaso, G. et al., 2014. The bacterial genotoxin colibactin promotes colon tumor growth by modifying the tumor microenvironment. *Gut Microbes*, 5(5), pp.675–80. 10.4161/19490976.2014.969989.
- Dastar, B. et al., 2016. Effect of calcium with and without probiotic, lactose, or both on organ and body weights, immune response and caecal microbiota in moulted laying hens. *Journal of animal physiology and animal nutrition*, 100(2), pp.243–50. 10.1111/jpn.12358.
- Dedrick, R.M. et al., 2019. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. *Nature medicine*, 25(5), pp.730–733. 10.1038/s41591-019-0437-z.
- Denli, A.M. et al., 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014), pp.231–235. 10.1038/nature03049.
- DeStefano Shields, C.E. et al., 2016. Reduction of Murine Colon Tumorigenesis Driven by

- Enterotoxigenic *Bacteroides fragilis* Using Cefoxitin Treatment. *Journal of Infectious Diseases*, 214(1), pp.122–129. 10.1093/infdis/jiw069.
- Diep, C.B. et al., 2006. The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. *Genes, chromosomes & cancer*, 45(1), pp.31–41. 10.1002/gcc.20261.
- Ding, T. et al., 2018. Antisense Oligonucleotides against miR-21 Inhibit the Growth and Metastasis of Colorectal Carcinoma via the DUSP8 Pathway. *Molecular Therapy - Nucleic Acids*, 13, pp.244–255. 10.1016/j.omtn.2018.09.004.
- Diosdado, B. et al., 2009. MiR-17-92 cluster is associated with 13q gain and c-myc expression during colorectal adenoma to adenocarcinoma progression. *British journal of cancer*, 101(4), pp.707–14. 10.1038/sj.bjc.6605037.
- Drekonja, D. et al., 2015. Fecal Microbiota Transplantation for Clostridium difficile Infection: A Systematic Review. *Annals of internal medicine*, 162(9), pp.630–8. 10.7326/M14-2693.
- Drew, D.A. et al., 2017. A Prospective Study of Smoking and Risk of Synchronous Colorectal Cancers. *The American journal of gastroenterology*, 112(3), pp.493–501. 10.1038/ajg.2016.589.
- Du, L. et al., 2017. KRAS and TP53 mutations in inflammatory bowel disease-associated colorectal cancer: a meta-analysis. *Oncotarget*, 8(13), pp.22175–22186. 10.18632/oncotarget.14549.
- Dyson, J.K., Rutter, M.D., 2012. Colorectal cancer in inflammatory bowel disease: what is the real magnitude of the risk? *World journal of gastroenterology*, 18(29), pp.3839–48. 10.3748/wjg.v18.i29.3839.
- Ebert, M.S., Neilson, J.R., Sharp, P.A., 2007. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature methods*, 4(9), pp.721–6. 10.1038/nmeth1079.
- Eluri, S. et al., 2017. Nearly a Third of High-Grade Dysplasia and Colorectal Cancer Is Undetected in Patients with Inflammatory Bowel Disease. *Digestive Diseases and Sciences*, 62(12), pp.3586–3593. 10.1007/s10620-017-4652-5.
- Fahmy, C.A. et al., 2019. Bifidobacterium longum Suppresses Murine Colorectal Cancer through the Modulation of oncomiRs and Tumor Suppressor miRNAs. *Nutrition and Cancer*, 71(4), pp.688–700. 10.1080/01635581.2019.1577984.
- Fang, M. et al., 2016. Common BRAF(V600E)-directed pathway mediates widespread epigenetic silencing in colorectal cancer and melanoma. *Proceedings of the National Academy of Sciences of the United States of America*, 113(5), pp.1250–5. 10.1073/pnas.1525619113.
- Fang, M. et al., 2014. The BRAF oncoprotein functions through the transcriptional repressor MAFG to mediate the CpG Island Methylator phenotype. *Molecular cell*, 55(6), pp.904–915. 10.1016/j.molcel.2014.08.010.
- Fanning, S. et al., 2012. Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. *Proceedings of the National Academy of Sciences*, 109(6), pp.2108–2113. 10.1073/pnas.1115621109.
- Farthing, M. et al., 2014. Survey of digestive health across Europe: Final report. Part 1: The burden of gastrointestinal diseases and the organisation and delivery of gastroenterology services across Europe. *United European gastroenterology journal*, 2(6), pp.539–43. 10.1177/2050640614554154.
- Fass, R., Alim, A., Kaunitz, J.D., 1995. Adenocarcinoma of the colon presenting as Streptococcus sanguis bacteremia. *The American journal of gastroenterology*, 90(8), pp.1343–5.
- Fearon, E.R., 2011. Molecular genetics of colorectal cancer. *Annual review of pathology*, 6, pp.479–507. 10.1146/annurev-pathol-011110-130235.
- Feng, Y.-H. et al., 2012. MicroRNA-21-mediated regulation of Sprouty2 protein expression enhances the cytotoxic effect of 5-fluorouracil and metformin in colon cancer cells. *International journal of molecular medicine*, 29(5), pp.920–6. 10.3892/ijmm.2012.910.
- Fennell, L.J. et al., 2018. MLH1–93 G/a polymorphism is associated with MLH1 promoter methylation and protein loss in dysplastic sessile serrated adenomas with BRAFV600E mutation. *BMC Cancer*, 18(1), p.35. 10.1186/s12885-017-3946-5.
- Fernández-Medarde, A., Santos, E., 2011. Ras in cancer and developmental diseases. *Genes & cancer*, 2(3), pp.344–58. 10.1177/1947601911411084.
- Ferreira, R.M. et al., 2018. Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota. *Gut*, 67(2), pp.226–236. 10.1136/gutjnl-2017-314205.
- Flemer, B. et al., 2018. The oral microbiota in colorectal cancer is distinctive and predictive. *Gut*, 67(8), pp.1454–1463. 10.1136/gutjnl-2017-314814.
- Flemer, B. et al., 2017. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut*, 66(4), pp.633–643. 10.1136/gutjnl-2015-309595.
- Fujiya, M. et al., 2014. MicroRNA-18a induces apoptosis in colon cancer cells via the



- autophagolysosomal degradation of oncogenic heterogeneous nuclear ribonucleoprotein A1. *Oncogene*, 33(40), pp.4847–56. 10.1038/onc.2013.429.
- Fukayama, M. et al., 2020. Thirty years of Epstein-Barr virus-associated gastric carcinoma. *Virchows Archiv : an international journal of pathology*, 476(3), pp.353–365. 10.1007/s00428-019-02724-4.
- Fukugaiti, M.H. et al., 2015. High occurrence of *Fusobacterium nucleatum* and *Clostridium difficile* in the intestinal microbiota of colorectal carcinoma patients. *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology]*, 46(4), pp.1135–40. 10.1590/S1517-838246420140665.
- Furman, D. et al., 2019. Chronic inflammation in the etiology of disease across the life span. *Nature medicine*, 25(12), pp.1822–1832. 10.1038/s41591-019-0675-0.
- Gaines, S. et al., 2020. Western Diet Promotes Intestinal Colonization by Collagenolytic Microbes and Promotes Tumor Formation After Colorectal Surgery. *Gastroenterology*, 158(4), pp.958-970.e2. 10.1053/j.gastro.2019.10.020.
- Gao, L., Bai, L., Nan, Q. zhen, 2013. Activation of Rho GTPase Cdc42 promotes adhesion and invasion in colorectal cancer cells. *Medical science monitor basic research*, 19, pp.201–7. 10.12659/MSMBR.883983.
- García Carretero, R. et al., 2016. Bacteraemia due to *Parvimonas micra*, a commensal pathogen, in a patient with an oesophageal tumour. *BMJ case reports*, 2016. 10.1136/bcr-2016-217740.
- Geigl, J.B. et al., 2008. Defining 'chromosomal instability'. *Trends in genetics : TIG*, 24(2), pp.64–9. 10.1016/j.tig.2007.11.006.
- Gerriets, V. et al., 2020. *Tumor Necrosis Factor (TNF) Inhibitors*.
- Gholizadeh, P., Eslami, H., Kafil, H.S., 2017. Carcinogenesis mechanisms of *Fusobacterium nucleatum*. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, 89, pp.918–925. 10.1016/j.biopha.2017.02.102.
- Gómez Del Pulgar, T. et al., 2008. Cdc42 is highly expressed in colorectal adenocarcinoma and downregulates ID4 through an epigenetic mechanism. *International journal of oncology*, 33(1), pp.185–93.
- Górski, A. et al., 2006. Bacteriophage translocation. *FEMS immunology and medical microbiology*, 46(3), pp.313–9. 10.1111/j.1574-695X.2006.00044.x.
- Griffiths-Jones, S., 2006. MiRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, 34(90001), pp.D140–D144. 10.1093/nar/gkj112.
- Griffiths-Jones, S. et al., 2007. MiRBase: tools for microRNA genomics. *Nucleic Acids Research*, 36(Database), pp.D154–D158. 10.1093/nar/gkm952.
- Griffiths-Jones, S., 2004. The microRNA Registry. *Nucleic Acids Research*, 32(90001), pp.109D – 111. 10.1093/nar/gkh023.
- De Groot, A.S., Scott, D.W., 2007. Immunogenicity of protein therapeutics. *Trends in immunology*, 28(11), pp.482–90. 10.1016/j.it.2007.07.011.
- Gui, H., Husson, M.A., Mannan, R., 2020. Correlations of morphology and molecular alterations in traditional serrated adenoma. *World Journal of Gastrointestinal Pathophysiology*, 11(4), pp.78–83. 10.4291/wjgp.v11.i4.78.
- Gupta, R., Sinha, S., Paul, R.N., 2018. The impact of microsatellite stability status in colorectal cancer. *Current problems in cancer*, 42(6), pp.548–559. 10.1016/j.currproblcancer.2018.06.010.
- Hale, V.L. et al., 2017. Shifts in the Fecal Microbiota Associated with Adenomatous Polyps. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 26(1), pp.85–94. 10.1158/1055-9965.EPI-16-0337.
- Han, J., 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development*, 18(24), pp.3016–3027. 10.1101/gad.1262504.
- Han, R. et al., 2016. Sodium Butyrate Upregulates miR-203 Expression to Exert Anti-Proliferation Effect on Colorectal Cancer Cells. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*, 39(5), pp.1919–1929. 10.1159/000447889.
- Hashimoto, T. et al., 2019. Acquisition of WNT Pathway Gene Alterations Coincides With the Transition From Precursor Polyps to Traditional Serrated Adenomas. *The American journal of surgical pathology*, 43(1), pp.132–139. 10.1097/PAS.0000000000001149.
- He, J.-H. et al., 2018. The CircRNA-ACAP2/Hsa-miR-21-5p/ Tiam1 Regulatory Feedback Circuit Affects the Proliferation, Migration, and Invasion of Colon Cancer SW480 Cells. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*, 49(4), pp.1539–1550. 10.1159/000493457.
- Hegde, M. et al., 2014. ACMG technical standards and guidelines for genetic testing for inherited

- colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). *Genetics in medicine : official journal of the American College of Medical Genetics*, 16(1), pp.101–16. 10.1038/gim.2013.166.
- Helwak, A. et al., 2013. Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell*, 153(3), pp.654–665. 10.1016/j.cell.2013.03.043.
- Hempelmann, J.A. et al., 2015. MSIplus for Integrated Colorectal Cancer Molecular Testing by Next-Generation Sequencing. *The Journal of molecular diagnostics : JMD*, 17(6), pp.705–14. 10.1016/j.jmoldx.2015.05.008.
- Hensley-McBain, T. et al., 2016. Effects of Fecal Microbial Transplantation on Microbiome and Immunity in Simian Immunodeficiency Virus-Infected Macaques. *Journal of virology*, 90(10), pp.4981–4989. 10.1128/JVI.00099-16.
- Heydari, Z. et al., 2019. Effects of Lactobacillus acidophilus and Bifidobacterium bifidum Probiotics on the Expression of MicroRNAs 135b, 26b, 18a and 155, and Their Involving Genes in Mice Colon Cancer. *Probiotics and antimicrobial proteins*, 11(4), pp.1155–1162. 10.1007/s12602-018-9478-8.
- Hino, R. et al., 2009. Activation of DNA methyltransferase 1 by EBV latent membrane protein 2A leads to promoter hypermethylation of PTEN gene in gastric carcinoma. *Cancer research*, 69(7), pp.2766–74. 10.1158/0008-5472.CAN-08-3070.
- Horita, K. et al., 2019. Long noncoding RNA UCA1 enhances sensitivity to oncolytic vaccinia virus by sponging miR-18a/miR-182 and modulating the Cdc42/filopodia axis in colorectal cancer. *Biochemical and biophysical research communications*, 516(3), pp.831–838. 10.1016/j.bbrc.2019.06.125.
- Howard, K.M. et al., 2015. Loss of miRNAs during processing and storage of cow's (*Bos taurus*) milk. *Journal of agricultural and food chemistry*, 63(2), pp.588–92. 10.1021/jf505526w.
- Hsu, H. et al., 1996. TRADD-TRAF2 and TRADD-FADD interactions define two distinct TNF receptor 1 signal transduction pathways. *Cell*, 84(2), pp.299–308. 10.1016/s0092-8674(00)80984-8.
- Hsu, H., Xiong, J., Goeddel, D. V., 1995. The TNF receptor 1-associated protein TRADD signals cell death and NF-kappa B activation. *Cell*, 81(4), pp.495–504. 10.1016/0092-8674(95)90070-5.
- Hu, S. et al., 2015. Butyrate inhibits pro-proliferative miR-92a by diminishing c-Myc-induced miR-17-92a cluster transcription in human colon cancer cells. *Molecular cancer*, 14, p.180. 10.1186/s12943-015-0450-x.
- Hu, S. et al., 2011. The microbe-derived short chain fatty acid butyrate targets miRNA-dependent p21 gene expression in human colon cancer. Navarro, A., ed. *PLoS ONE*, 6(1), p.e16221. 10.1371/journal.pone.0016221.
- Hu, Y. et al., 2014. MiR-199a-5p Loss Up-Regulated DDR1 Aggravated Colorectal Cancer by Activating Epithelial-to-Mesenchymal Transition Related Signaling. *Digestive Diseases and Sciences*, 59(9), pp.2163–2172. 10.1007/s10620-014-3136-0.
- Huang, C.S. et al., 2011. The clinical significance of serrated polyps. *The American journal of gastroenterology*, 106(2), pp.229–40; quiz 241. 10.1038/ajg.2010.429.
- Huang, G. et al., 2017. MiR-20a-directed regulation of BID is associated with the TRAIL sensitivity in colorectal cancer. *Oncology reports*, 37(1), pp.571–578. 10.3892/or.2016.5278.
- Huang, L. et al., 2000. Activation of the p21WAF1/CIP1 promoter independent of p53 by the histone deacetylase inhibitor suberoylanilide hydroxamic acid (SAHA) through the Sp1 sites. *Oncogene*, 19(50), pp.5712–9. 10.1038/sj.onc.1203963.
- Humphreys, K.J., McKinnon, R.A., Michael, M.Z., 2014. MiR-18a Inhibits CDC42 and Plays a Tumour Suppressor Role in Colorectal Cancer Cells Yue, J., ed. *PLoS one*, 9(11), p.e112288. 10.1371/journal.pone.0112288.
- Hunter, M.P. et al., 2008. Detection of microRNA expression in human peripheral blood microvesicles. *PLoS one*, 3(11), p.e3694. 10.1371/journal.pone.0003694.
- Hutvagner, G. et al., 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science (New York, N.Y.)*, 293(5531), pp.834–8. 10.1126/science.1062961.
- Hwang, S. et al., 2013. Bacteroides fragilis Toxin Induces IL-8 Secretion in HT29/C1 Cells through Disruption of E-cadherin Junctions. *Immune Network*, 13(5), p.213. 10.4110/in.2013.13.5.213.
- Hwang, S. et al., 2018. Immunofluorescence Microscopic Evaluation of Tight Junctional Proteins during Enterotoxigenic Bacteroides fragilis (ETBF) Infection in Mice. *Biomedical Science Letters*, 24(3), pp.275–279. 10.15616/BSL.2018.24.3.275.
- Iliopoulos, D. et al., 2010. STAT3 Activation of miR-21 and miR-181b-1 via PTEN and CYLD Are Part of the Epigenetic Switch Linking Inflammation to Cancer. *Molecular Cell*, 39(4), pp.493–506. 10.1016/j.molcel.2010.07.023.

- Inan, M.S. et al., 2000. The luminal short-chain fatty acid butyrate modulates NF-kappaB activity in a human colonic epithelial cell line. *Gastroenterology*, 118(4), pp.724–34. 10.1016/s0016-5085(00)70142-9.
- India State-Level Disease Burden Initiative Collaborators, 2017. Nations within a nation: variations in epidemiological transition across the states of India, 1990-2016 in the Global Burden of Disease Study. *Lancet (London, England)*, 390(10111), pp.2437–2460. 10.1016/S0140-6736(17)32804-0.
- India State-Level Disease Burden Initiative Diabetes Collaborators, 2018. The increasing burden of diabetes and variations among the states of India: the Global Burden of Disease Study 1990-2016. *The Lancet. Global health*, 6(12), pp.e1352–e1362. 10.1016/S2214-109X(18)30387-5.
- Issa, J.-P., 2004. CpG island methylator phenotype in cancer. *Nature reviews. Cancer*, 4(12), pp.988–93. 10.1038/nrc1507.
- Izumi, H. et al., 2012. Bovine milk contains microRNA and messenger RNA that are stable under degradative conditions. *Journal of dairy science*, 95(9), pp.4831–4841. 10.3168/jds.2012-5489.
- Janarthanan, S. et al., 2012. Clostridium difficile-Associated Diarrhea and Proton Pump Inhibitor Therapy: A Meta-Analysis. *The American Journal of Gastroenterology*, 107(7), pp.1001–1010. 10.1038/ajg.2012.179.
- Jasmine, F. et al., 2012. A genome-wide study of cytogenetic changes in colorectal cancer using SNP microarrays: opportunities for future personalized treatment. *PloS one*, 7(2), p.e31968. 10.1371/journal.pone.0031968.
- Jass, J.R., 2001. Serrated route to colorectal cancer: back street or super highway? *The Journal of pathology*, 193(3), pp.283–5. 10.1002/1096-9896(200103)193:3<283::AID-PATH799>3.0.CO;2-9.
- Jensen, A.B. et al., 2006. Survival After Colorectal Cancer in Patients with Ulcerative Colitis: A Nationwide Population-Based Danish Study. *The American Journal of Gastroenterology*, 101(6), pp.1283–1287. 10.1111/j.1572-0241.2006.00520.x.
- Jeon, J.I., Ko, S.H., Kim, J.M., 2019. Intestinal Epithelial Cells Exposed to Bacteroides fragilis Enterotoxin Regulate NF-κB Activation and Inflammatory Responses through β-Catenin Expression Torres, V. J., ed. *Infection and Immunity*, 87(11). 10.1128/IAI.00312-19.
- Jess, T. et al., 2006. Risk of Intestinal Cancer in Inflammatory Bowel Disease: A Population-Based Study From Olmsted County, Minnesota. *Gastroenterology*, 130(4), pp.1039–1046. 10.1053/j.gastro.2005.12.037.
- Jiang, Z. et al., 2020. LncRNA HAND2-AS1 inhibits 5-fluorouracil resistance by modulating miR-20a/PDCD4 axis in colorectal cancer. *Cellular signalling*, 66, p.109483. 10.1016/j.cellsig.2019.109483.
- Jin, B., Li, Y., Robertson, K.D., 2011. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes & cancer*, 2(6), pp.607–17. 10.1177/1947601910393957.
- Johnston, D.G.W. et al., 2018. Loss of microRNA-21 Influences the Gut Microbiota Causing Reduced Susceptibility in a Murine Model of Colitis. *Journal of Crohn's & colitis*. 10.1093/ecco-jcc/jjy038.
- Jonas, S., Izaurrealde, E., 2015. Towards a molecular understanding of microRNA-mediated gene silencing. *Nature Reviews Genetics*, 16(7), pp.421–433. 10.1038/nrg3965.
- Kakar, S. et al., 2008. CpG island methylation is frequently present in tubulovillous and villous adenomas and correlates with size, site, and villous component. *Human Pathology*, 39(1), pp.30–36. 10.1016/j.humpath.2007.06.002.
- Kalliolias, G.D., Ivashkiv, L.B., 2016. TNF biology, pathogenic mechanisms and emerging therapeutic strategies. *Nature reviews. Rheumatology*, 12(1), pp.49–62. 10.1038/nrrheum.2015.169.
- Kambara, T. et al., 2004. BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. *Gut*, 53(8), pp.1137–44. 10.1136/gut.2003.037671.
- Kang, W.K. et al., 2015. Stromal expression of miR-21 in T3-4a colorectal cancer is an independent predictor of early tumor relapse. *BMC gastroenterology*, 15, p.2. 10.1186/s12876-015-0227-0.
- Kany, S., Vollrath, J.T., Relja, B., 2019. Cytokines in Inflammatory Disease. *International journal of molecular sciences*, 20(23). 10.3390/ijms20236008.
- Kapitanović, S. et al., 2004. APC gene loss of heterozygosity, mutations, E1317Q, and I1307K germ-line variants in sporadic colon cancer in Croatia. *Experimental and molecular pathology*, 77(3), pp.193–200. 10.1016/j.yexmp.2004.06.001.
- Kasai, C. et al., 2016. Comparison of human gut microbiota in control subjects and patients with colorectal carcinoma in adenoma: Terminal restriction fragment length polymorphism and next-generation sequencing analyses. *Oncology reports*, 35(1), pp.325–33. 10.3892/or.2015.4398.
- Keku, T.O. et al., 2015. The gastrointestinal microbiota and colorectal cancer. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 308(5), pp.G351–G363.

- 10.1152/ajpgi.00360.2012.
- Kim, H.P., Crockett, S.D., Shaheen, N.J., 2014. The Burden of Gastrointestinal and Liver Disease Around the World. In: *GI Epidemiology*. Chichester, UK: John Wiley & Sons, Ltd, 2014, pp. 1–13. 10.1002/9781118727072.ch1.
- Kim, J.H., Kang, G.H., 2020. Evolving pathologic concepts of serrated lesions of the colorectum. *Journal of pathology and translational medicine*, 54(4), pp.276–289. 10.4132/jptm.2020.04.15.
- Kim, K.-M. et al., 2010. KRAS mutations in traditional serrated adenomas from Korea herald an aggressive phenotype. *The American journal of surgical pathology*, 34(5), pp.667–75. 10.1097/PAS.0b013e3181d40cb2.
- Kim, M.-J. et al., 2013. Traditional serrated adenoma of the colorectum: clinicopathologic implications and endoscopic findings of the precursor lesions. *American journal of clinical pathology*, 140(6), pp.898–911. 10.1309/AJCPDJC9VC5KTYUS.
- Kim, Y. et al., 2009. Epstein-Barr virus antibody level and gastric cancer risk in Korea: a nested case-control study. *British journal of cancer*, 101(3), pp.526–9. 10.1038/sj.bjc.6605146.
- Kinzler, K.W., Vogelstein, B., 1996. Lessons from hereditary colorectal cancer. *Cell*, 87(2), pp.159–70. 10.1016/s0092-8674(00)81333-1.
- Klampfer, L. et al., 2003. Inhibition of interferon gamma signaling by the short chain fatty acid butyrate. *Molecular cancer research : MCR*, 1(11), pp.855–62.
- Kobayashi, H., Tomari, Y., 2016. RISC assembly: Coordination between small RNAs and Argonaute proteins. *Biochimica et biophysica acta*, 1859(1), pp.71–81. 10.1016/j.bbagr.2015.08.007.
- Kobayashi, K. et al., 2017. P53 Expression as a Diagnostic Biomarker in Ulcerative Colitis-Associated Cancer. *International Journal of Molecular Sciences*, 18(6), p.1284. 10.3390/ijms18061284.
- Koh, J. et al., 2019. Development and Validation of an Easy-to-Implement, Practical Algorithm for the Identification of Molecular Subtypes of Gastric Cancer: Prognostic and Therapeutic Implications. *The oncologist*, 24(12), pp.e1321–e1330. 10.1634/theoncologist.2019-0058.
- Kong, Y. et al., 2012. The deoxycholic acid targets miRNA-dependent CAC1 gene expression in multidrug resistance of human colorectal cancer. *The International Journal of Biochemistry & Cell Biology*, 44(12), pp.2321–2332. 10.1016/j.biocel.2012.08.006.
- Kosaka, N. et al., 2010. MicroRNA as a new immune-regulatory agent in breast milk. *Silence*, 1(1), p.7. 10.1186/1758-907X-1-7.
- Kosmidou, V. et al., 2014. Tumor heterogeneity revealed by KRAS, BRAF, and PIK3CA pyrosequencing: KRAS and PIK3CA intratumor mutation profile differences and their therapeutic implications. *Human mutation*, 35(3), pp.329–40. 10.1002/humu.22496.
- Kostic, A.D. et al., 2012. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome research*, 22(2), pp.292–8. 10.1101/gr.126573.111.
- Kozomara, A., Birgaoanu, M., Griffiths-Jones, S., 2019. MiRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(D1), pp.D155–D162. 10.1093/nar/gky1141.
- Kozomara, A., Griffiths-Jones, S., 2014. MiRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, 42(D1), pp.D68–D73. 10.1093/nar/gkt1181.
- Kozomara, A., Griffiths-Jones, S., 2011. MiRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, 39(Database), pp.D152–D157. 10.1093/nar/gkq1027.
- Kulaylat, M.N., Dayton, M.T., 2010. Ulcerative colitis and cancer. *Journal of surgical oncology*, 101(8), pp.706–12. 10.1002/jso.21505.
- Lakatos, P.-L., Lakatos, L., 2008. Risk for colorectal cancer in ulcerative colitis: changes, causes and management strategies. *World journal of gastroenterology*, 14(25), pp.3937–47.
- Lash, R.H., Genta, R.M., Schuler, C.M., 2010. Sessile serrated adenomas: prevalence of dysplasia and carcinoma in 2139 patients. *Journal of Clinical Pathology*, 63(8), pp.681–686. 10.1136/jcp.2010.075507.
- de Lau, W. et al., 2014. The R-spondin/Lgr5/Rnf43 module: regulator of Wnt signal strength. *Genes & development*, 28(4), pp.305–16. 10.1101/gad.235473.113.
- Ledder, O., 2019. Antibiotics in inflammatory bowel diseases: do we know what we're doing? *Translational pediatrics*, 8(1), pp.42–55. 10.21037/tp.2018.11.02.
- Lee, C.H. et al., 2016. Frozen vs Fresh Fecal Microbiota Transplantation and Clinical Resolution of Diarrhea in Patients With Recurrent Clostridium difficile Infection: A Randomized Clinical Trial. *JAMA*, 315(2), pp.142–9. 10.1001/jama.2015.18098.
- Lee, J.-H. et al., 2009. Clinicopathological and molecular characteristics of Epstein-Barr virus-associated gastric carcinoma: a meta-analysis. *Journal of gastroenterology and hepatology*, 24(3), pp.354–65. 10.1111/j.1440-1746.2009.05775.x.
- Lee, Y. et al., 2004. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal*, 23(20), pp.4051–4060. 10.1038/sj.emboj.7600385.

- Lee, Y. et al., 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425(6956), pp.415–419. 10.1038/nature01957.
- Leggett, B., Whitehall, V., 2010. Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology*, 138(6), pp.2088–100. 10.1053/j.gastro.2009.12.066.
- Levin, A.D., Wildenberg, M.E., van den Brink, G.R., 2016. Mechanism of Action of Anti-TNF Therapy in Inflammatory Bowel Disease. *Journal of Crohn's and Colitis*, 10(8), pp.989–997. 10.1093/ecco-jcc/jjw053.
- Li, Chenli et al., 2016. MicroRNA-21 promotes proliferation, migration, and invasion of colorectal cancer, and tumor growth associated with down-regulation of sec23a expression. *BMC Cancer*, 16(1), p.605. 10.1186/s12885-016-2628-z.
- Li, H. et al., 1998. Cleavage of BID by Caspase 8 Mediates the Mitochondrial Damage in the Fas Pathway of Apoptosis. *Cell*, 94(4), pp.491–501. 10.1016/S0092-8674(00)81590-1.
- Li, L. et al., 2019. Gut microbiota from colorectal cancer patients enhances the progression of intestinal adenoma in Apcmin/+ mice. *EBioMedicine*, 48, pp.301–315. 10.1016/j.ebiom.2019.09.021.
- Li, X. et al., 2018. Effects of Whole Milk Supplementation on Gut Microbiota and Cardiometabolic Biomarkers in Subjects with and without Lactose Malabsorption. *Nutrients*, 10(10). 10.3390/nu10101403.
- Liang, J.Q. et al., 2020. A novel faecal Lachnoclostridium marker for the non-invasive diagnosis of colorectal adenoma and cancer. *Gut*, 69(7), pp.1248–1257. 10.1136/gutjnl-2019-318532.
- Liang, J.Q. et al., 2019. A Novel Microrna Panel for Non-Invasive Diagnosis and Prognosis of Colorectal Cancer. *Gastroenterology*, 156(6), p.S-496. 10.1016/S0016-5085(19)38106-5.
- Liang, Q. et al., 2017. Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer. *Clinical Cancer Research*, 23(8), pp.2061–2070. 10.1158/1078-0432.CCR-16-1599.
- Liang, Q. et al., 2014. Integrative Identification of Epstein–Barr Virus–Associated Mutations and Epigenetic Alterations in Gastric Cancer. *Gastroenterology*, 147(6), pp.1350-1362.e4. 10.1053/j.gastro.2014.08.036.
- Liao, D. et al., 2018. MiR-221 inhibits autophagy and targets TP53INP1 in colorectal cancer cells. *Experimental and therapeutic medicine*, 15(2), pp.1712–1717. 10.3892/etm.2017.5522.
- Liu, G. et al., 2015. Tumor suppressor microRNA-18a regulates tumor proliferation and invasion by targeting TBPL1 in colorectal cancer cells. *Molecular medicine reports*, 12(5), pp.7643–8. 10.3892/mmr.2015.4335.
- Liu, H., Redline, R.W., Han, Y.W., 2007. Fusobacterium nucleatum Induces Fetal Death in Mice via Stimulation of TLR4-Mediated Placental Inflammatory Response. *The Journal of Immunology*, 179(4), pp.2501–2508. 10.4049/jimmunol.179.4.2501.
- Liu, L. et al., 2018. Prognostic and predictive value of long non-coding RNA GAS5 and mircoRNA-221 in colorectal cancer and their effects on colorectal cancer cell proliferation, migration and invasion. *Cancer Biomarkers*, 22(2), pp.283–299. 10.3233/CBM-171011.
- Liu, M. et al., 2011. MiR-21 targets the tumor suppressor RhoB and regulates proliferation, invasion and apoptosis in colorectal cancer cells. *FEBS letters*, 585(19), pp.2998–3005. 10.1016/j.febslet.2011.08.014.
- Liu, S. et al., 2014. A microRNA 221- and 222-mediated feedback loop maintains constitutive activation of NFκB and STAT3 in colorectal cancer cells. *Gastroenterology*, 147(4), pp.847-859.e11. 10.1053/j.gastro.2014.06.006.
- Liu, S. et al., 2019. Oral Administration of miR-30d from Feces of MS Patients Suppresses MS-like Symptoms in Mice by Expanding Akkermansia muciniphila. *Cell host & microbe*, 26(6), pp.779-794.e8. 10.1016/j.chom.2019.10.008.
- Liu, S. et al., 2016. The Host Shapes the Gut Microbiota via Fecal MicroRNA. *Cell Host & Microbe*, 19(1), pp.32–43. 10.1016/j.chom.2015.12.005.
- Longqiu, Y. et al., 2016. A miRNAs panel promotes the proliferation and invasion of colorectal cancer cells by targeting GABBR1. *Cancer medicine*, 5(8), pp.2022–31. 10.1002/cam4.760.
- Losso, G.M. et al., 2012. Instabilidade de microssatélite - MSI nos marcadores (BAT26, BAT25, D2s123, D5S346, D17S250) no câncer de reto. *ABCD. Arquivos Brasileiros de Cirurgia Digestiva (São Paulo)*, 25(4), pp.240–244. 10.1590/S0102-67202012000400006.
- Löwenmark, T. et al., 2020. Parvimonas micra as a putative non-invasive faecal biomarker for colorectal cancer. *Scientific reports*, 10(1), p.15250. 10.1038/s41598-020-72132-1.
- Lu, Z. et al., 2008. MicroRNA-21 promotes cell transformation by targeting the programmed cell death 4 gene. *Oncogene*, 27(31), pp.4373–4379. 10.1038/onc.2008.72.
- Lumeng, C.N., Saltiel, A.R., 2011. Inflammatory links between obesity and metabolic disease. *The*

- Journal of clinical investigation*, 121(6), pp.2111–7. 10.1172/JCI57132.
- Ma, J. et al., 2018. PIAS3-mediated feedback loops promote chronic colitis-associated malignant transformation. *Theranostics*, 8(11), pp.3022–3037. 10.7150/thno.23046.
- MacDonald, B.T., Tamai, K., He, X., 2009. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Developmental cell*, 17(1), pp.9–26. 10.1016/j.devcel.2009.06.016.
- Makarova, J.A. et al., 2016. Intracellular and extracellular microRNA: An update on localization and biological role. *Progress in histochemistry and cytochemistry*, 51(3–4), pp.33–49. 10.1016/j.proghi.2016.06.001.
- Mani, N. et al., 2002. Environmental response and autoregulation of *Clostridium difficile* TxeR, a sigma factor for toxin gene expression. *Journal of bacteriology*, 184(21), pp.5971–8. 10.1128/jb.184.21.5971-5978.2002.
- Mani, N., Dupuy, B., 2001. Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proceedings of the National Academy of Sciences of the United States of America*, 98(10), pp.5844–9. 10.1073/pnas.101126598.
- Mark D., T. et al., 2005. Quantitation of DNA for Forensic DNA Typing by qPCR (quantitative PCR): Singleplex and Multiple Modes for Nuclear and Mitochondrial Genomes, and the Y Chromosome.
- Martín, R. et al., 2017. Functional Characterization of Novel *Faecalibacterium prausnitzii* Strains Isolated from Healthy Volunteers: A Step Forward in the Use of *F. prausnitzii* as a Next-Generation Probiotic. *Frontiers in Microbiology*, 8, p.1226. 10.3389/fmicb.2017.01226.
- Di Martino, M.T. et al., 2016. Mir-221/222 are promising targets for innovative anticancer therapy. *Expert opinion on therapeutic targets*, 20(9), pp.1099–108. 10.1517/14728222.2016.1164693.
- De Marzo, A.M. et al., 2007. Inflammation in prostate carcinogenesis. *Nature reviews. Cancer*, 7(4), pp.256–69. 10.1038/nrc2090.
- Masanobu, F. et al., 2011. Probiotic treatment reduced preneoplastic lesions through the downregulation of toll like receptor 4 in a chemo-induced carcinogenic model. *Journal of Clinical Biochemistry and Nutrition*, 49(1), pp.57–61. 10.3164/jcbn.10-114.
- McCarthy, A.J., Serra, S., Chetty, R., 2019. Traditional serrated adenoma: an overview of pathology and emphasis on molecular pathogenesis. *BMJ open gastroenterology*, 6(1), p.e000317. 10.1136/bmjgast-2019-000317.
- McDaniel, D.K. et al., 2016. Emerging Roles for Noncanonical NF- $\kappa$ B Signaling in the Modulation of Inflammatory Bowel Disease Pathobiology. *Inflammatory bowel diseases*, 22(9), pp.2265–79. 10.1097/MIB.0000000000000858.
- McQuaid, K.R., 2018. Drugs Used in the Treatment of Gastrointestinal Diseases. In: Frassetto, D., Henriquez, R. D., eds. *Basic & Clinical Pharmacology*. McGraw-Hill Education, 2018.
- Messaritakis, I. et al., 2018. Evaluation of the detection of Toll-like receptors (TLRs) in cancer development and progression in patients with colorectal cancer De Re, V., ed. *PLOS ONE*, 13(6), p.e0197327. 10.1371/journal.pone.0197327.
- Meyers, B.C. et al., 2008. Criteria for Annotation of Plant MicroRNAs. *The Plant Cell*, 20(12), pp.3186–3190. 10.1105/tpc.108.064311.
- Migliore, L. et al., 2011. Genetics, cytogenetics, and epigenetics of colorectal cancer. *Journal of biomedicine & biotechnology*, 2011, p.792362. 10.1155/2011/792362.
- Mima, K. et al., 2016. *Fusobacterium nucleatum* in Colorectal Carcinoma Tissue According to Tumor Location. *Clinical and translational gastroenterology*, 7(11), p.e200. 10.1038/ctg.2016.53.
- miRBase, 2020. *Stem-loop sequence hsa-mir-923*.
- Mitchell, P.S. et al., 2008. Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America*, 105(30), pp.10513–8. 10.1073/pnas.0804549105.
- Mizutani, S., Yamada, T., Yachida, S., 2020. Significance of the gut microbiome in multistep colorectal carcinogenesis. *Cancer Science*, 111(3), pp.766–773. 10.1111/cas.14298.
- Mohan, P., Mohan, S.B., Dutta, M., 2019. Communicable or noncommunicable diseases? Building strong primary health care systems to address double burden of disease in India. *Journal of family medicine and primary care*, 8(2), pp.326–329. 10.4103/jfmpc.jfmpc\_67\_19.
- Monaghan, Tanya M., Duggal, N.A., et al., 2021. A Multi-Factorial Observational Study on Sequential Fecal Microbiota Transplant in Patients with Medically Refractory *Clostridioides difficile* Infection. *Cells*, 10(11), p.3234. 10.3390/cells10113234.
- Monaghan, Tanya M., Duggal, N.A., et al., 2021. A Multi-Factorial Observational Study on Sequential Fecal Microbiota Transplant in Patients with Medically Refractory *Clostridioides difficile* Infection. *Cells*, 10(11). 10.3390/cells10113234.
- Monaghan, Tanya M., Seekatz, A.M., et al., 2021. Fecal Microbiota Transplantation for Recurrent

- Clostridioides difficile Infection Associates With Functional Alterations in Circulating microRNAs. *Gastroenterology*, 161(1), pp.255-270.e4. 10.1053/j.gastro.2021.03.050.
- Monaghan, Tanya M., Biswas, R.N., et al., 2021. Multiomics Profiling Reveals Signatures of Dysmetabolism in Urban Populations in Central India. *Microorganisms*, 9(7), p.1485. 10.3390/microorganisms9071485.
- Monahan, K.J. et al., 2020. Guidelines for the management of hereditary colorectal cancer from the British Society of Gastroenterology (BSG)/Association of Coloproctology of Great Britain and Ireland (ACPGBI)/United Kingdom Cancer Genetics Group (UKCGG). *Gut*, 69(3), pp.411–444. 10.1136/gutjnl-2019-319915.
- Monteleone, N.J. et al., 2019. MiR-21-mediated regulation of 15-hydroxyprostaglandin dehydrogenase in colon cancer. *Scientific Reports*, 9(1), p.5405. 10.1038/s41598-019-41862-2.
- Moody, L. et al., 2019. The Efficacy of miR-20a as a Diagnostic and Prognostic Biomarker for Colorectal Cancer: A Systematic Review and Meta-Analysis. *Cancers*, 11(8). 10.3390/cancers11081111.
- Moore, L.D., Le, T., Fan, G., 2013. DNA methylation and its basic function. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 38(1), pp.23–38. 10.1038/npp.2012.112.
- Moroi, R. et al., 2013. FCGR3A-158 polymorphism influences the biological response to infliximab in Crohn's disease through affecting the ADCC activity. *Immunogenetics*, 65(4), pp.265–271. 10.1007/s00251-013-0679-8.
- Mortaz, E. et al., 2018. Update on Neutrophil Function in Severe Inflammation. *Frontiers in Immunology*, 9. 10.3389/fimmu.2018.02171.
- Mortensen, R.D. et al., 2011. Posttranscriptional activation of gene expression in *Xenopus laevis* oocytes by microRNA-protein complexes (microRNPs). *Proceedings of the National Academy of Sciences of the United States of America*, 108(20), pp.8281–6. 10.1073/pnas.1105401108.
- Munagala, R. et al., 2016. Bovine milk-derived exosomes for drug delivery. *Cancer letters*, 371(1), pp.48–61. 10.1016/j.canlet.2015.10.020.
- Nadeem, M.S. et al., 2020. Risk of colorectal cancer in inflammatory bowel diseases. *Seminars in cancer biology*, 64, pp.51–60. 10.1016/j.semcancer.2019.05.001.
- Nagtegaal, I.D. et al., 2020. The 2019 WHO classification of tumours of the digestive system. *Histopathology*, 76(2), pp.182–188. 10.1111/his.13975.
- Nguyen, H.T., Duong, H.-Q., 2018. The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy. *Oncology letters*, 16(1), pp.9–18. 10.3892/ol.2018.8679.
- NI, B. et al., 2015. Increased urothelial cancer associated 1 is associated with tumor proliferation and metastasis and predicts poor prognosis in colorectal cancer. *International Journal of Oncology*, 47(4), pp.1329–1338. 10.3892/ijo.2015.3109.
- Nijjer, S., Dubrey, S.W., 2010. Streptococcus sanguis endocarditis associated with colonic carcinoma. *Case Reports*, 2010(feb19 1), pp.bcr0920092311–bcr0920092311. 10.1136/bcr.09.2009.2311.
- Niu, B. et al., 2014. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics (Oxford, England)*, 30(7), pp.1015–6. 10.1093/bioinformatics/btt755.
- Nougayrede, J.-P., 2006. Escherichia coli Induces DNA Double-Strand Breaks in Eukaryotic Cells. *Science*, 313(5788), pp.848–851. 10.1126/science.1127059.
- Nourbakhsh, M., Minoo, P., 2020. Annexin A10 and HES-1 Immunohistochemistry in Right-sided Traditional Serrated Adenomas Suggests an Origin From Sessile Serrated Adenoma. *Applied Immunohistochemistry & Molecular Morphology*, 28(4), pp.296–302. 10.1097/PAI.0000000000000740.
- Obregon-Tito, A.J. et al., 2015. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nature communications*, 6, p.6505. 10.1038/ncomms7505.
- Ocvirk, S. et al., 2020. A prospective cohort analysis of gut microbial co-metabolism in Alaska Native and rural African people at high and low risk of colorectal cancer. *The American journal of clinical nutrition*, 111(2), pp.406–419. 10.1093/ajcn/nqz301.
- Ocvirk, S., O'Keefe, S.J., 2017. Influence of Bile Acids on Colorectal Cancer Risk: Potential Mechanisms Mediated by Diet - Gut Microbiota Interactions. *Current nutrition reports*, 6(4), pp.315–322. 10.1007/s13668-017-0219-5.
- Ogino, S. et al., 2009. CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut*, 58(1), pp.90–6. 10.1136/gut.2008.155473.
- Ogino, S. et al., 2006. Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. *The Journal of molecular diagnostics : JMD*, 8(2), pp.209–17. 10.2353/jmoldx.2006.050135.

- Oldenburg, A. et al., 2017. A lipodystrophy-causing lamin A mutant alters conformation and epigenetic regulation of the anti-adipogenic MIR335 locus. *The Journal of Cell Biology*, 216(9), p.jcb.201701043. 10.1083/jcb.201701043.
- Orzechowska, E.J. et al., 2015. Synergy of BID with doxorubicin in the killing of cancer cells. *Oncology reports*, 33(5), pp.2143–50. 10.3892/or.2015.3841.
- Ou, J. et al., 2013. Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. *The American journal of clinical nutrition*, 98(1), pp.111–20. 10.3945/ajcn.112.056689.
- Pai, S.G. et al., 2017. Wnt/beta-catenin pathway: modulating anticancer immune response. *Journal of hematology & oncology*, 10(1), p.101. 10.1186/s13045-017-0471-6.
- De Palma, F.D.E. et al., 2019. The Molecular Hallmarks of the Serrated Pathway in Colorectal Cancer. *Cancers*, 11(7). 10.3390/cancers11071017.
- Panno, S. et al., 2020. Loop Mediated Isothermal Amplification: Principles and Applications in Plant Virology. *Plants (Basel, Switzerland)*, 9(4). 10.3390/plants9040461.
- Pasquinelli, A.E., 2012. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nature reviews. Genetics*, 13(4), pp.271–82. 10.1038/nrg3162.
- Peacock, O. et al., 2014. Inflammation and MiR-21 pathways functionally interact to downregulate PDCD4 in colorectal cancer. *PLoS one*, 9(10), p.e110267. 10.1371/journal.pone.0110267.
- Pećina-Šlaus, N. et al., 2020. Mismatch Repair Pathway, Genome Stability and Cancer. *Frontiers in Molecular Biosciences*, 7. 10.3389/fmolb.2020.00122.
- Philip, A., Ferro, V.A., Tate, R.J., 2015. Determination of the potential bioavailability of plant microRNAs using a simulated human digestion process. *Molecular nutrition & food research*, 59(10), pp.1962–72. 10.1002/mnfr.201500137.
- Pilley, S., Rodriguez, T.A., Vousden, K.H., 2021. Mutant p53 in cell-cell interactions. [online]. *Genes & development*, 35(7–8), pp.433–448. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/33861719>.
- Pino, M.S., Chung, D.C., 2010. The chromosomal instability pathway in colon cancer. *Gastroenterology*, 138(6), pp.2059–72. 10.1053/j.gastro.2009.12.065.
- Place, R.F. et al., 2008. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 105(5), pp.1608–13. 10.1073/pnas.0707594105.
- Popat, S., Hubner, R., Houlston, R.S., 2005. Systematic review of microsatellite instability and colorectal cancer prognosis. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 23(3), pp.609–18. 10.1200/JCO.2005.01.086.
- Poullenet, F. et al., 2016. Risk of Incident Cancer in Inflammatory Bowel Disease Patients Starting Anti-TNF Therapy While Having Recent Malignancy. *Inflammatory Bowel Diseases*, 22(6), pp.1362–1369. 10.1097/MIB.0000000000000741.
- Purcell, R. V. et al., 2017. Colonization with enterotoxigenic *Bacteroides fragilis* is associated with early-stage colorectal neoplasia Brim, H., ed. *PLOS ONE*, 12(2), p.e0171602. 10.1371/journal.pone.0171602.
- Qiang, Y. et al., 2020. MiR-20a/Foxj2 Axis Mediates Growth and Metastasis of Colorectal Cancer Cells as Identified by Integrated Analysis. *Medical Science Monitor*, 26. 10.12659/MSM.923559.
- Qin, J. et al., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), pp.59–65. 10.1038/nature08821.
- Qin, J., Luo, M., 2014. MicroRNA-221 promotes colorectal cancer cell invasion and metastasis by targeting RECK. *FEBS letters*, 588(1), pp.99–104. 10.1016/j.febslet.2013.11.014.
- Rawla, P. et al., 2017. Colon Carcinoma Presenting as Streptococcus anginosus Bacteremia and Liver Abscess. *Gastroenterology research*, 10(6), pp.376–379. 10.14740/gr884w.
- Ren, Y. et al., 2019. Microcystin-LR promotes migration via the cooperation between microRNA-221/PTEN and STAT3 signal pathway in colon cancer cell line DLD-1. *Ecotoxicology and environmental safety*, 167, pp.107–113. 10.1016/j.ecoenv.2018.09.065.
- Rex, D.K. et al., 2012. Serrated lesions of the colorectum: review and recommendations from an expert panel. *The American journal of gastroenterology*, 107(9), pp.1315–29; quiz 1314, 1330. 10.1038/ajg.2012.161.
- Rineh, A. et al., 2014. Clostridium difficile infection: molecular pathogenesis and novel therapeutics. *Expert review of anti-infective therapy*, 12(1), pp.131–50. 10.1586/14787210.2014.866515.
- Roden, C. et al., 2017. Novel determinants of mammalian primary microRNA processing revealed by systematic evaluation of hairpin-containing transcripts and human genetic variation. *Genome research*, 27(3), pp.374–384. 10.1101/gr.208900.116.
- Rodríguez-Nogales, A. et al., 2017. Differential intestinal anti-inflammatory effects of Lactobacillus



- fermentum and *Lactobacillus salivarius* in DSS mouse colitis: impact on microRNAs expression and microbiota composition. *Molecular Nutrition & Food Research*, 61(11), p.1700144. 10.1002/mnfr.201700144.
- Rodríguez-Nogales, A., Algieri, F., Garrido-Mesa, J., Vezza, T., Utrilla, M. Pilar, et al., 2018. Intestinal anti-inflammatory effect of the probiotic *Saccharomyces boulardii* in DSS-induced colitis in mice: Impact on microRNAs expression and gut microbiota composition. *The Journal of Nutritional Biochemistry*, 61, pp.129–139. 10.1016/j.jnutbio.2018.08.005.
- Rodríguez-Nogales, A., Algieri, F., Garrido-Mesa, J., Vezza, T., Utrilla, Maria P., et al., 2018. The Administration of *Escherichia coli* Nissle 1917 Ameliorates Development of DSS-Induced Colitis in Mice. *Frontiers in Pharmacology*, 9. 10.3389/fphar.2018.00468.
- Rome, S., 2019. Biological properties of plant-derived extracellular vesicles. *Food & function*, 10(2), pp.529–538. 10.1039/c8fo02295j.
- Romero-Velarde, E. et al., 2019. The Importance of Lactose in the Human Diet: Outcomes of a Mexican Consensus Meeting. *Nutrients*, 11(11). 10.3390/nu11112737.
- Rosenberg, D.W. et al., 2007. Mutations in BRAF and KRAS differentially distinguish serrated versus non-serrated hyperplastic aberrant crypt foci in humans. *Cancer research*, 67(8), pp.3551–4. 10.1158/0008-5472.CAN-07-0343.
- Roughead, E.E. et al., 2016. Proton pump inhibitors and risk of *Clostridium difficile* infection: a multi-country study using sequence symmetry analysis. *Expert Opinion on Drug Safety*, 15(12), pp.1589–1595. 10.1080/14740338.2016.1238071.
- Rubinstein, M.R. et al., 2013. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/ $\beta$ -catenin signaling via its FadA adhesin. *Cell Host & Microbe*, 14(2), pp.195–206. 10.1016/j.chom.2013.07.012.
- Ruder, B., Atreya, R., Becker, C., 2019. Tumour Necrosis Factor Alpha in Intestinal Homeostasis and Gut Related Diseases. *International journal of molecular sciences*, 20(8). 10.3390/ijms20081887.
- Rudman, N., Gornik, O., Lauc, G., 2019. Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes. *FEBS letters*, 593(13), pp.1598–1615. 10.1002/1873-3468.13495.
- Ruiz, A. et al., 2021. Transmembrane TNF and Its Receptors TNFR1 and TNFR2 in Mycobacterial Infections. *International journal of molecular sciences*, 22(11). 10.3390/ijms22115461.
- Rymbai, M.L. et al., 2015. Frequency of Epstein–Barr virus infection as detected by messenger RNA for EBNA 1 in histologically proven gastric adenocarcinoma in patients presenting to a tertiary care center in South India. *Indian journal of medical microbiology*, 33(3), pp.369–73. 10.4103/0255-0857.158556.
- Saeki, H. et al., 2001. Genetic alterations in the human Tcf-4 gene in Japanese patients with sporadic gastrointestinal cancers with microsatellite instability. *Oncology*, 61(2), pp.156–61. 10.1159/000055367.
- Saheb Sharif-Askari, N. et al., 2020. Integrative systematic review meta-analysis and bioinformatics identifies MicroRNA-21 and its target genes as biomarkers for colorectal adenocarcinoma. *International journal of surgery (London, England)*, 73, pp.113–122. 10.1016/j.ijsu.2019.11.017.
- Saito, S., Tajiri, H., Ikegami, M., 2015. Serrated polyps of the colon and rectum: Endoscopic features including image enhanced endoscopy. *World journal of gastrointestinal endoscopy*, 7(9), pp.860–71. 10.4253/wjge.v7.i9.860.
- Sakamori, R. et al., 2014. CDC42 inhibition suppresses progression of incipient intestinal tumors. *Cancer research*, 74(19), pp.5480–92. 10.1158/0008-5472.CAN-14-0267.
- Salipante, S.J. et al., 2014. Microsatellite instability detection by next generation sequencing. *Clinical chemistry*, 60(9), pp.1192–9. 10.1373/clinchem.2014.223677.
- Samadder, N.J. et al., 2019. Family History Associates With Increased Risk of Colorectal Cancer in Patients With Inflammatory Bowel Diseases. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*, 17(9), pp.1807-1813.e1. 10.1016/j.cgh.2018.09.038.
- Sansonetti, P.J., Medzhitov, R., 2009. Learning tolerance while fighting ignorance. *Cell*, 138(3), pp.416–20. 10.1016/j.cell.2009.07.024.
- Sayed, D. et al., 2008. MicroRNA-21 Targets Sprouty2 and Promotes Cellular Outgrowths Chernoff, J., ed. *Molecular Biology of the Cell*, 19(8), pp.3272–3282. 10.1091/mbc.e08-02-0159.
- Schlörmann, W. et al., 2015. Influence of miRNA-106b and miRNA-135a on butyrate-regulated expression of p21 and Cyclin D2 in human colon adenoma cells. *Genes & nutrition*, 10(6), p.50. 10.1007/s12263-015-0500-4.
- Schmidt, E.P. et al., 2011. On, Around, and Through: Neutrophil-Endothelial Interactions in Innate Immunity. *Physiology*, 26(5), pp.334–347. 10.1152/physiol.00011.2011.

- Schubbert, S., Shannon, K., Bollag, G., 2007. Hyperactive Ras in developmental disorders and cancer. *Nature reviews. Cancer*, 7(4), pp.295–308. 10.1038/nrc2109.
- Schwab, M. et al., 2007. Involvement of different nuclear hormone receptors in butyrate-mediated inhibition of inducible NF kappa B signalling. *Molecular immunology*, 44(15), pp.3625–32. 10.1016/j.molimm.2007.04.010.
- Schwartz, S. et al., 1999. Frameshift mutations at mononucleotide repeats in caspase-5 and other target genes in endometrial and gastrointestinal cancer of the microsatellite mutator phenotype. *Cancer research*, 59(12), pp.2995–3002.
- Sears, C.L., 2009. Enterotoxigenic *Bacteroides fragilis*: a Rogue among Symbiotes. *Clinical Microbiology Reviews*, 22(2), pp.349 LP – 369. 10.1128/CMR.00053-08.
- Sekine, S. et al., 2020. Clinicopathological and molecular correlations in traditional serrated adenoma. *Journal of gastroenterology*, 55(4), pp.418–427. 10.1007/s00535-020-01673-z.
- Sekine, S. et al., 2017. Comprehensive characterization of RSPO fusions in colorectal traditional serrated adenomas. *Histopathology*, 71(4), pp.601–609. 10.1111/his.13265.
- Shagisultanova, E. et al., 2015. Preclinical and clinical studies of the NEDD9 scaffold protein in cancer and other diseases. *Gene*, 567(1), pp.1–11. 10.1016/j.gene.2015.04.086.
- Shang, S., Hua, F., Hu, Z.-W., 2017. The regulation of  $\beta$ -catenin activity and function in cancer: therapeutic opportunities. *Oncotarget*, 8(20), pp.33972–33989. 10.18632/oncotarget.15687.
- Shen, L. et al., 2007. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 104(47), pp.18654–9. 10.1073/pnas.0704652104.
- Sheng, Q. et al., 2019. Characteristics of fecal gut microbiota in patients with colorectal cancer at different stages and different sites. *Oncology letters*, 18(5), pp.4834–4844. 10.3892/ol.2019.10841.
- Shi, C. et al., 2013. MicroRNA-21 Knockout Improve the Survival Rate in DSS Induced Fatal Colitis through Protecting against Inflammation and Tissue Injury Heimesaat, M. M., ed. *PLoS ONE*, 8(6), p.e66814. 10.1371/journal.pone.0066814.
- Shi, C. et al., 2016. Novel evidence for an oncogenic role of microRNA-21 in colitis-associated colorectal cancer. *Gut*, 65(9), pp.1470–1481. 10.1136/gutjnl-2014-308455.
- Shukla, S.K. et al., 2011. Epstein-Barr virus DNA load and its association with *Helicobacter pylori* infection in gastroduodenal diseases. *The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases*, 15(6), pp.583–90.
- Shukla, S.K. et al., 2012. Expression profile of latent and lytic transcripts of epstein-barr virus in patients with gastroduodenal diseases: a study from northern India. *Journal of medical virology*, 84(8), pp.1289–97. 10.1002/jmv.23322.
- Shuto, T. et al., 2019. Establishment of a Screening Method for Epstein-Barr Virus-Associated Gastric Carcinoma by Droplet Digital PCR. *Microorganisms*, 7(12). 10.3390/microorganisms7120628.
- Signs, S.A. et al., 2018. Stromal miR-20a controls paracrine CXCL8 secretion in colitis and colon cancer. *Oncotarget*, 9(16), pp.13048–13059. 10.18632/oncotarget.24495.
- Silva-García, O., Valdez-Alarcón, J.J., Baizabal-Aguirre, V.M., 2019. Wnt/ $\beta$ -Catenin Signaling as a Molecular Target by Pathogenic Bacteria. *Frontiers in Immunology*, 10. 10.3389/fimmu.2019.02135.
- Sims, E.K. et al., 2017. MicroRNA 21 targets BCL2 mRNA to increase apoptosis in rat and human beta cells. *Diabetologia*, 60(6), pp.1057–1065. 10.1007/s00125-017-4237-z.
- Slattery, M.L. et al., 2012. Toll-like receptor genes and their association with colon and rectal cancer development and prognosis. *International Journal of Cancer*, 130(12), pp.2974–2980. 10.1002/ijc.26314.
- Sobhani, I. et al., 2011. Microbial dysbiosis in colorectal cancer (CRC) patients. *PloS one*, 6(1), p.e16393. 10.1371/journal.pone.0016393.
- Sokolova, V. et al., 2015. The Effects of miR-20a on p21: Two Mechanisms Blocking Growth Arrest in TGF- $\beta$ -Responsive Colon Carcinoma. *Journal of cellular physiology*, 230(12), pp.3105–14. 10.1002/jcp.25051.
- Souza, R.F. et al., 1996. Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours. *Nature genetics*, 14(3), pp.255–7. 10.1038/ng1196-255.
- Spring, K.J. et al., 2006. High Prevalence of Sessile Serrated Adenomas With BRAF Mutations: A Prospective Study of Patients Undergoing Colonoscopy. *Gastroenterology*, 131(5), pp.1400–1407. 10.1053/j.gastro.2006.08.038.
- Stanczak, A. et al., 2011. Prognostic significance of Wnt-1,  $\beta$ -catenin and E-cadherin expression in advanced colorectal carcinoma. *Pathology oncology research : POR*, 17(4), pp.955–63. 10.1007/s12253-011-9409-4.

- Stavropoulou, E., Bezirtzoglou, E., 2020. Probiotics in Medicine: A Long Debate. *Frontiers in immunology*, 11, p.2192. 10.3389/fimmu.2020.02192.
- Stojanovic, J. et al., 2019. MicroRNAs expression profiles as diagnostic biomarkers of gastric cancer: a systematic literature review. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals*, 24(2), pp.110–119. 10.1080/1354750X.2018.1539765.
- Sun, K. et al., 2011. MicroRNA-221 inhibits CDKN1C/p57 expression in human colorectal carcinoma. *Acta pharmacologica Sinica*, 32(3), pp.375–84. 10.1038/aps.2010.206.
- Svrcek, M. et al., 2013. Overexpression of microRNAs-155 and 21 targeting mismatch repair proteins in inflammatory bowel diseases. *Carcinogenesis*, 34(4), pp.828–34. 10.1093/carcin/bgs408.
- Tang, S. et al., 2019. MiR-20a regulates sensitivity of colorectal cancer cells to NK cells by targeting MICA. *Bioscience reports*, 39(7). 10.1042/BSR20180695.
- Tao, K. et al., 2014. Prognostic value of miR-221-3p, miR-342-3p and miR-491-5p expression in colon cancer. *American journal of translational research*, 6(4), pp.391–401.
- Tao, Y. et al., 2019. Aging-like Spontaneous Epigenetic Silencing Facilitates Wnt Activation, Stemness, and BrafV600E-Induced Tumorigenesis. *Cancer cell*, 35(2), pp.315-328.e6. 10.1016/j.ccell.2019.01.005.
- Tavakoli, A. et al., 2020. Association between Epstein-Barr virus infection and gastric cancer: a systematic review and meta-analysis. *BMC cancer*, 20(1), p.493. 10.1186/s12885-020-07013-x.
- Teng, Y. et al., 2016. Grapefruit-derived nanovectors deliver miR-18a for treatment of liver metastasis of colon cancer by induction of M1 macrophages. *Oncotarget*, 7(18), pp.25683–97. 10.18632/oncotarget.8361.
- Teng, Y. et al., 2018. Plant-Derived Exosomal MicroRNAs Shape the Gut Microbiota. *Cell Host & Microbe*, 24(5), pp.637-652.e8. 10.1016/j.chom.2018.10.001.
- Tepus, M., Yau, T.O., 2020. Non-Invasive Colorectal Cancer Screening: An Overview. *Gastrointestinal Tumors*, 3–4(7), pp.62–73. 10.1159/000507701.
- Thum, T. et al., 2008. MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature*, 456(7224), pp.980–4. 10.1038/nature07511.
- Thurgate, L.E. et al., 2019. An Overview of Inflammatory Bowel Disease Unclassified in Children. *Inflammatory intestinal diseases*, 4(3), pp.97–103. 10.1159/000501519.
- Timmermans, W.M.C. et al., 2016. B-Cell Dysregulation in Crohn's Disease Is Partially Restored with Infliximab Therapy Richard, Y., ed. *PLOS ONE*, 11(7), p.e0160103. 10.1371/journal.pone.0160103.
- Tjalsma, H. et al., 2012. A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects. *Nature Reviews Microbiology*, 10(8), pp.575–582. 10.1038/nrmicro2819.
- Tokunaga, M. et al., 1993. Epstein-Barr virus in gastric carcinoma. *The American journal of pathology*, 143(5), pp.1250–4.
- Tong, L. et al., 2015. Improved RT-PCR Assay to Quantitate the Pri-, Pre-, and Mature microRNAs with Higher Efficiency and Accuracy. *Molecular Biotechnology*, 57(10), pp.939–946. 10.1007/s12033-015-9885-y.
- Toyota, M., Suzuki, H., 2010. Epigenetic drivers of genetic alterations. *Advances in genetics*, 70, pp.309–23. 10.1016/B978-0-12-380866-0.60011-3.
- Triantafyllidis, J.K., Nasioulas, G., Kosmidis, P.A., 2009. Colorectal cancer and inflammatory bowel disease: epidemiology, risk factors, mechanisms of carcinogenesis and prevention strategies. *Anticancer research*, 29(7), pp.2727–37.
- Tsai, J.-H., Liao, J.-Y., et al., 2016. RNF43 Is an Early and Specific Mutated Gene in the Serrated Pathway, With Increased Frequency in Traditional Serrated Adenoma and Its Associated Malignancy. *The American journal of surgical pathology*, 40(10), pp.1352–9. 10.1097/PAS.0000000000000664.
- Tsai, J.-H. et al., 2014. Traditional serrated adenoma has two pathways of neoplastic progression that are distinct from the sessile serrated pathway of colorectal carcinogenesis. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 27(10), pp.1375–85. 10.1038/modpathol.2014.35.
- Tsai, J.-H., Cheng, C.-H., et al., 2016. Traditional serrated adenoma with BRAF mutation is associated with synchronous/metachronous BRAF -mutated serrated lesions. *Histopathology*, 68(6), pp.810–818. 10.1111/his.12814.
- Tsunoda, T. et al., 2011. Oncogenic KRAS regulates miR-200c and miR-221/222 in a 3D-specific manner in colorectal cancer cells. *Anticancer research*, 31(7), pp.2453–9.
- Turnbaugh, P.J. et al., 2009. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Science Translational Medicine*, 1(6), pp.6ra14-6ra14. 10.1126/scitranslmed.3000322.

- Turner, J.R., 2006. Molecular Basis of Epithelial Barrier Regulation. *The American Journal of Pathology*, 169(6), pp.1901–1909. 10.2353/ajpath.2006.060681.
- Ullman, T.A., Itzkowitz, S.H., 2011. Intestinal inflammation and cancer. *Gastroenterology*, 140(6), pp.1807–16. 10.1053/j.gastro.2011.01.057.
- Vaisman-Mentesh, A. et al., 2019. Molecular Landscape of Anti-Drug Antibodies Reveals the Mechanism of the Immune Response Following Treatment With TNF $\alpha$  Antagonists. *Frontiers in immunology*, 10, p.2921. 10.3389/fimmu.2019.02921.
- Vallianou, N. et al., 2019. Understanding the Role of the Gut Microbiome and Microbial Metabolites in Obesity and Obesity-Associated Metabolic Disorders: Current Evidence and Perspectives. *Current obesity reports*, 8(3), pp.317–332. 10.1007/s13679-019-00352-2.
- Vasudevan, S., 2012. Posttranscriptional Upregulation by MicroRNAs. *Wiley Interdisciplinary Reviews: RNA*, 3(3), pp.311–330. 10.1002/wrna.121.
- Vasudevan, S., Tong, Y., Steitz, J.A., 2007. Switching from repression to activation: microRNAs can up-regulate translation. *Science (New York, N.Y.)*, 318(5858), pp.1931–4. 10.1126/science.1149460.
- Ventura, A. et al., 2008. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell*, 132(5), pp.875–86. 10.1016/j.cell.2008.02.019.
- Verdier, J. et al., 2020. Faecal Micro-RNAs in Inflammatory Bowel Diseases. *Journal of Crohn's & colitis*, 14(1), pp.110–117. 10.1093/ecco-jcc/jjz120.
- Vetizou, M. et al., 2015. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science*, 350(6264), pp.1079–1084. 10.1126/science.aad1329.
- Viladomiu, M. et al., 2012. Modeling the role of peroxisome proliferator-activated receptor  $\gamma$  and microRNA-146 in mucosal immune responses to *Clostridium difficile*. Heimesaat, M. M., ed. *PloS one*, 7(10), p.e47525. 10.1371/journal.pone.0047525.
- Voelker, R., 2019. FDA Approves Bacteriophage Trial. *JAMA*, 321(7), p.638. 10.1001/jama.2019.0510.
- Vogels, C.B.F. et al., 2020. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer-probe sets. *Nature microbiology*, 5(10), pp.1299–1305. 10.1038/s41564-020-0761-6.
- Voth, D.E., Ballard, J.D., 2005. *Clostridium difficile* toxins: mechanism of action and role in disease. *Clinical microbiology reviews*, 18(2), pp.247–63. 10.1128/CMR.18.2.247-263.2005.
- Vringer, E., Tait, S.W.G., 2019. Mitochondria and Inflammation: Cell Death Heats Up. *Frontiers in cell and developmental biology*, 7, p.100. 10.3389/fcell.2019.00100.
- Wajant, H., Siegmund, D., 2019. TNFR1 and TNFR2 in the Control of the Life and Death Balance of Macrophages. *Frontiers in cell and developmental biology*, 7, p.91. 10.3389/fcell.2019.00091.
- Walsh, M.C., Lee, J., Choi, Y., 2015. Tumor necrosis factor receptor- associated factor 6 (TRAF6) regulation of development, function, and homeostasis of the immune system. *Immunological reviews*, 266(1), pp.72–92. 10.1111/imr.12302.
- Wang, P. et al., 2009. MicroRNA-21 negatively regulates Cdc25A and cell cycle progression in colon cancer cells. *Cancer research*, 69(20), pp.8157–65. 10.1158/0008-5472.CAN-09-1996.
- Wang, X., Huycke, M.M., 2007. Extracellular Superoxide Production by *Enterococcus faecalis* Promotes Chromosomal Instability in Mammalian Cells. *Gastroenterology*, 132(2), pp.551–561. 10.1053/j.gastro.2006.11.040.
- Wang, Y. et al., 2017. BART miRNAs: an unimaginable force in the development of nasopharyngeal carcinoma. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP)*, 26(2), pp.144–150. 10.1097/CEJ.0000000000000221.
- Warren, R.L. et al., 2013. Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome*, 1(1), p.16. 10.1186/2049-2618-1-16.
- Watanabe, T. et al., 2011. Ulcerative colitis-associated colorectal cancer shows a poorer survival than sporadic colorectal cancer: A nationwide Japanese study. *Inflammatory Bowel Diseases*, 17(3), pp.802–808. 10.1002/ibd.21365.
- Wawruszak, A. et al., 2019. Histone Deacetylase Inhibitors and Phenotypical Transformation of Cancer Cells. *Cancers*, 11(2), p.148. 10.3390/cancers11020148.
- Weber, J.A. et al., 2010. The microRNA spectrum in 12 body fluids. *Clinical chemistry*, 56(11), pp.1733–41. 10.1373/clinchem.2010.147405.
- Wei, Z. et al., 2016. Could gut microbiota serve as prognostic biomarker associated with colorectal cancer patients' survival? A pilot study on relevant mechanism. *Oncotarget*, 7(29), pp.46158–46172. 10.18632/oncotarget.10064.
- Weisenberger, D.J. et al., 2006. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nature genetics*,

- 38(7), pp.787–93. 10.1038/ng1834.
- Wéra, O., Lancellotti, P., Oury, C., 2016. The Dual Role of Neutrophils in Inflammatory Bowel Diseases. *Journal of Clinical Medicine*, 5(12), p.118. 10.3390/jcm5120118.
- Westholm, J.O., Lai, E.C., 2011. Mirtrons: microRNA biogenesis via splicing. *Biochimie*, 93(11), pp.1897–1904. 10.1016/j.biochi.2011.06.017.
- Whitehall, V.L. et al., 2001. Methylation of O-6-methylguanine DNA methyltransferase characterizes a subset of colorectal cancer with low-level DNA microsatellite instability. *Cancer research*, 61(3), pp.827–30.
- Wilcox, M.H., 2003. Gastrointestinal disorders and the critically ill. Clostridium difficile infection and pseudomembranous colitis. *Best practice & research. Clinical gastroenterology*, 17(3), pp.475–93.
- Wilson, M.R. et al., 2019. The human gut bacterial genotoxin colibactin alkylates DNA. *Science*, 363(6428), p.eaar7785. 10.1126/science.aar7785.
- Wong, S.H. et al., 2017. Gavage of Fecal Samples From Patients With Colorectal Cancer Promotes Intestinal Carcinogenesis in Germ-Free and Conventional Mice. *Gastroenterology*, 153(6), pp.1621-1633.e6. 10.1053/j.gastro.2017.08.022.
- Wu, C.-W. et al., 2013. MicroRNA-18a attenuates DNA damage repair through suppressing the expression of ataxia telangiectasia mutated in colorectal cancer. *PloS one*, 8(2), p.e57036. 10.1371/journal.pone.0057036.
- Wu, C.W. et al., 2014. Identification of microRNA-135b in Stool as a Potential Noninvasive Biomarker for Colorectal Cancer and Adenoma. *Clinical Cancer Research*, 20(11), pp.2994–3002. 10.1158/1078-0432.CCR-13-1750.
- Wu, H.-J., Wu, E., 2012. The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes*, 3(1), pp.4–14. 10.4161/gmic.19320.
- Wu, J., Li, Q., Fu, X., 2019. Fusobacterium nucleatum Contributes to the Carcinogenesis of Colorectal Cancer by Inducing Inflammation and Suppressing Host Immunity. *Translational Oncology*, 12(6), pp.846–851. 10.1016/j.tranon.2019.03.003.
- Wu, N. et al., 2013. Dysbiosis Signature of Fecal Microbiota in Colorectal Cancer Patients. *Microbial Ecology*, 66(2), pp.462–470. 10.1007/s00248-013-0245-9.
- Wu, S. et al., 2007. Bacteroides fragilis toxin stimulates intestinal epithelial cell shedding and -secretase-dependent E-cadherin cleavage. *Journal of Cell Science*, 120(11), pp.1944–1952. 10.1242/jcs.03455.
- Wu, S. et al., 2006. The Bacteroides fragilis Toxin Binds to a Specific Intestinal Epithelial Cell Receptor. *Infection and Immunity*, 74(9), pp.5382–5390. 10.1128/IAI.00060-06.
- Wu, W. et al., 2013. MicroRNA-18a modulates STAT3 activity through negative regulation of PIAS3 during gastric adenocarcinogenesis. *British journal of cancer*, 108(3), pp.653–61. 10.1038/bjc.2012.587.
- Xie, J., Itzkowitz, S.H., 2008. Cancer in inflammatory bowel disease. *World journal of gastroenterology*, 14(3), pp.378–89.
- Xie, M. et al., 2013. Mammalian 5'-Capped MicroRNA Precursors that Generate a Single MicroRNA. *Cell*, 155(7), pp.1568–1580. 10.1016/j.cell.2013.11.027.
- Xu, J. et al., 2020. Alteration of the abundance of Parvimonas micra in the gut along the adenoma-carcinoma sequence. *Oncology letters*, 20(4), p.106. 10.3892/ol.2020.11967.
- Xu, T. et al., 2015. MicroRNA-20a enhances the epithelial-to-mesenchymal transition of colorectal cancer cells by modulating matrix metalloproteinases. *Experimental and therapeutic medicine*, 10(2), pp.683–688. 10.3892/etm.2015.2538.
- Xu, Y. et al., 2020. Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. *Nature Medicine*, 26(4), pp.502–505. 10.1038/s41591-020-0817-4.
- Xue, M. et al., 2019. Structure elucidation of colibactin and its DNA cross-links. *Science*, 365(6457), p.eaax2685. 10.1126/science.aax2685.
- Xue, Q. et al., 2013. Anti-miRNA-221 sensitizes human colorectal carcinoma cells to radiation by upregulating PTEN. *World journal of gastroenterology*, 19(48), pp.9307–17. 10.3748/wjg.v19.i48.9307.
- Xue, X. et al., 2014. Downregulation of microRNA-107 in intestinal CD11c(+) myeloid cells in response to microbiota and proinflammatory cytokines increases IL-23p19 expression. *European journal of immunology*, 44(3), pp.673–82. 10.1002/eji.201343717.
- Yachida, S. et al., 2019. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature medicine*, 25(6), pp.968–976. 10.1038/s41591-019-0458-7.

- Yan, H.H.N. et al., 2017. RNF43 germline and somatic mutation in serrated neoplasia pathway and its association with BRAF mutation. *Gut*, 66(9), pp.1645–1656. 10.1136/gutjnl-2016-311849.
- Yang, J.-S. et al., 2010. Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proceedings of the National Academy of Sciences*, 107(34), pp.15163–15168. 10.1073/pnas.1006432107.
- Yang, S. et al., 2004. BRAF and KRAS Mutations in hyperplastic polyps and serrated adenomas of the colorectum: relationship to histology and CpG island methylation status. *The American journal of surgical pathology*, 28(11), pp.1452–9. 10.1097/01.pas.0000141404.56839.6a.
- Yang, Yongzhi et al., 2017. Fusobacterium nucleatum Increases Proliferation of Colorectal Cancer Cells and Tumor Development in Mice by Activating Toll-Like Receptor 4 Signaling to Nuclear Factor- $\kappa$ B, and Up-regulating Expression of MicroRNA-21. *Gastroenterology*, 152(4), pp.851-866.e24. 10.1053/j.gastro.2016.11.018.
- Yang, Yang et al., 2017. Long non-coding RNA GAS5 inhibits cell proliferation, induces G0/G1 arrest and apoptosis, and functions as a prognostic marker in colorectal cancer. *Oncology Letters*, 13(5), pp.3151–3158. 10.3892/ol.2017.5841.
- Yau, T.O. et al., 2019. Faecal microRNAs as a non-invasive tool in the diagnosis of colonic adenomas and colorectal cancer: A meta-analysis. *Scientific reports*, 9(1), p.9491. 10.1038/s41598-019-45570-9.
- Yau, T.O. et al., 2022. Hyperactive neutrophil chemotaxis contributes to anti-tumor necrosis factor- $\alpha$  treatment resistance in inflammatory bowel disease. *Journal of Gastroenterology and Hepatology*, 37(3), pp.531–541. 10.1111/jgh.15764.
- Yau, T.O. et al., 2016. MicroRNA-20a in human faeces as a non-invasive biomarker for colorectal cancer. *Oncotarget*, 7(2), pp.1559–68. 10.18632/oncotarget.6403.
- Yau, T.O., Wu, C.W., et al., 2014. MicroRNA-221 and microRNA-18a identification in stool as potential biomarkers for the non-invasive diagnosis of colorectal carcinoma. *British Journal of Cancer*, 111(9), pp.1765–1771. 10.1038/bjc.2014.484.
- Yau, T.O., 2019. Precision treatment in colorectal cancer: Now and the future. *JGH open : an open access journal of gastroenterology and hepatology*, 3(5), pp.361–369. 10.1002/jgh3.12153.
- Yau, T.O., Tang, C.-M., Yu, J., 2014. Epigenetic dysregulation in Epstein-Barr virus-associated gastric carcinoma: Disease and treatments. *World Journal of Gastroenterology*, 20(21), p.6448. 10.3748/wjg.v20.i21.6448.
- Ye, H. et al., 2015. A critical role of mir-199a in the cell biological behaviors of colorectal cancer. *Diagnostic pathology*, 10, p.65. 10.1186/s13000-015-0260-x.
- Yeung, Y.T. et al., 2018. Signaling Pathways in Inflammation and Anti-inflammatory Therapies. *Current pharmaceutical design*, 24(14), pp.1449–1484. 10.2174/1381612824666180327165604.
- Yi, R. et al., 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & development*, 17(24), pp.3011–6. 10.1101/gad.1158803.
- Yin, S.L. et al., 2019. Long non-coding RNA FENDRR restrains the aggressiveness of CRC via regulating miR-18a-5p/ING4 axis. *Journal of cellular biochemistry*. 10.1002/jcb.29555.
- Yongyu, Z. et al., 2018. MicroRNA-18a targets IRF2 and CBX7 to promote cell proliferation in hepatocellular carcinoma. *Oncology research*. 10.3727/096504018X15165493852990.
- Yoshitaka, T. et al., 1996. Mutations of E2F-4 trinucleotide repeats in colorectal cancer with microsatellite instability. *Biochemical and biophysical research communications*, 227(2), pp.553–7. 10.1006/bbrc.1996.1544.
- You, J.S., Jones, P.A., 2012. Cancer genetics and epigenetics: two sides of the same coin? *Cancer cell*, 22(1), pp.9–20. 10.1016/j.ccr.2012.06.008.
- Yu, G. et al., 2012. Prognostic values of the miR-17-92 cluster and its paralogs in colon cancer. *Journal of surgical oncology*, 106(3), pp.232–7. 10.1002/jso.22138.
- Yu, J. et al., 2017. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut*, 66(1), pp.70–78. 10.1136/gutjnl-2015-309800.
- Yu, T. et al., 2017. Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell*, 170(3), pp.548-563.e16. 10.1016/j.cell.2017.07.008.
- Yu, Yingjie et al., 2012. MicroRNA-21 induces stemness by downregulating transforming growth factor beta receptor 2 (TGF R2) in colon cancer cells. *Carcinogenesis*, 33(1), pp.68–76. 10.1093/carcin/bgr246.
- Yu, Y. et al., 2012. MicroRNA-21 induces stemness by downregulating transforming growth factor beta receptor 2 (TGF R2) in colon cancer cells. *Carcinogenesis*, 33(1), pp.68–76. 10.1093/carcin/bgr246.
- Yuan, L. et al., 2018. The influence of gut microbiota dysbiosis to the efficacy of 5-Fluorouracil treatment on colorectal cancer. *Biomedicine & Pharmacotherapy*, 108, pp.184–193.

- 10.1016/j.biopha.2018.08.165.
- Zamani, S. et al., 2020. Enterotoxigenic *Bacteroides fragilis*: A Possible Etiological Candidate for Bacterially-Induced Colorectal Precancerous and Cancerous Lesions. *Frontiers in Cellular and Infection Microbiology*, 9. 10.3389/fcimb.2019.00449.
- Zarzour, P. et al., 2015. Single nucleotide polymorphism array profiling identifies distinct chromosomal aberration patterns across colorectal adenomas and carcinomas. *Genes, chromosomes & cancer*, 54(5), pp.303–14. 10.1002/gcc.22243.
- Zauber, A.G. et al., 2012. Colonoscopic Polypectomy and Long-Term Prevention of Colorectal-Cancer Deaths. *New England Journal of Medicine*, 366(8), pp.687–696. 10.1056/NEJMoa1100370.
- Zhang, C. et al., 2018. Anti-TNF-  $\alpha$  Therapy Suppresses Proinflammatory Activities of Mucosal Neutrophils in Inflammatory Bowel Disease. *Mediators of Inflammation*, 2018, pp.1–12. 10.1155/2018/3021863.
- Zhang, G.-J.J. et al., 2014. MiR-20a is an independent prognostic factor in colorectal cancer and is involved in cell metastasis. *Molecular medicine reports*, 10(1), pp.283–91. 10.3892/mmr.2014.2144.
- Zhang, L. et al., 2018. Knockdown of MiR-20a Enhances Sensitivity of Colorectal Cancer Cells to Cisplatin by Increasing ASK1 Expression. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*, 47(4), pp.1432–1441. 10.1159/000490834.
- Zhang, Y., Wang, X., 2020. Targeting the Wnt/ $\beta$ -catenin signaling pathway in cancer. *Journal of Hematology & Oncology*, 13(1), p.165. 10.1186/s13045-020-00990-3.
- Zhao, L. et al., 2022. *Parvimonas micra* promotes colorectal tumorigenesis and is associated with prognosis of colorectal cancer patients [online]. *Oncogene*. Available at: <https://www.nature.com/articles/s41388-022-02395-7>.
- Zheng, Y. et al., 2017. *Clostridium difficile* colonization in preoperative colorectal cancer patients. *Oncotarget*, 8(7), pp.11877–11886. 10.18632/oncotarget.14424.
- Zhou, Youlian et al., 2016. Association of oncogenic bacteria with colorectal cancer in South China. *Oncotarget*, 7(49), pp.80794–80802. 10.18632/oncotarget.13094.
- Zhu, Q.D. et al., 2018. MiR-199a-5p Inhibits the Growth and Metastasis of Colorectal Cancer Cells by Targeting ROCK1. *Technology in Cancer Research & Treatment*, 17, p.153303461877550. 10.1177/1533034618775509.
- Zorron Cheng Tao Pu, L. et al., 2020. Microbiota profile is different for early and invasive colorectal cancer and is consistent throughout the colon. *Journal of gastroenterology and hepatology*, 35(3), pp.433–437. 10.1111/jgh.14868.

# **X. PUBLISHED PEER-REVIEWED ARTICLES**



# Non-Invasive Colorectal Cancer Screening: An Overview

Melanie Tepus, **Tung On Yau**

*Gastrointestinal Tumors*. 2020;3–4(7):62–73.

Journal URL: [karger.com/Article/FullText/507701](https://karger.com/Article/FullText/507701)

DOI: [10.1002/igh3.12153](https://doi.org/10.1002/igh3.12153)

PMID: [32903904](https://pubmed.ncbi.nlm.nih.gov/32903904/)

PMCID: [PMC7445682](https://pubmed.ncbi.nlm.nih.gov/PMC7445682/)

Review Article

# Non-Invasive Colorectal Cancer Screening: An Overview

Melanie Tepus Tung On Yau

John van Geest Cancer Research Centre, School of Science and Technology,  
Nottingham Trent University, Nottingham, UK

## Keywords

Colorectal cancer screening · Tumour biomarkers · Diagnostic biomarkers · microRNA · Gut microbiota

## Abstract

**Background:** Colorectal cancer (CRC) follows a protracted stepwise progression, from benign adenomas to malignant adenocarcinomas. If detected early, 90% of deaths are preventable. However, CRC is asymptomatic in its early-stage and arises sporadically within the population. Therefore, CRC screening is a public health priority. **Summary:** Faecal immunochemical test (FIT) is gradually replacing guaiac faecal occult blood test and is now the most commonly used screening tool for CRC screening program globally. However, FIT is still limited by the haemoglobin degradation and the intermittent bleeding patterns, so that one in four CRC cases are still diagnosed in a late stage, leading to poor prognosis. A multi-target stool DNA test (Cologuard, a combination of *NDRG4* and *BMP3* DNA methylation, *KRAS* mutations, and haemoglobin) and a plasma *SEPT9* DNA methylation test (Epi proColon) are non-invasive tools also approved by the US FDA, but those screening approaches are not cost-effective, and the detection accuracies remain unsatisfactory. In addition to the approved tests, faecal-/blood-based microRNA and CRC-related gut microbiome screening markers are under development, with work ongoing to find the best combination of molecular biomarkers which maximise the screening sensitivity and specificity. **Key Message:** Maximising the detection accuracy with a cost-effective approach for non-invasive CRC screening is urgently needed to further reduce the incidence of CRC and associated mortality rates.

© 2020 The Author(s)

Published by S. Karger AG, Basel

Tung On Yau  
John van Geest Cancer Research Centre, School of Science and Technology  
Nottingham Trent University, Clifton Lane  
Nottingham NG11 8NS (UK)  
tungon@gmail.com

## Introduction

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related death worldwide, with over 1.8 million new cases and causing approximately 900,000 deaths in 2018 [1]. The incidence rate varies among countries, with a rate about 3-times higher in developed versus developing countries, while the mortality rate has less variation [1]. The improvement in cancer treatment and the introduction of CRC screening programs have further reduced mortality arising from CRC in developed countries [2–4]. However, the pathogenesis of CRC follows a stepwise progression from benign adenomas to malignant adenocarcinomas, often over a course of 10 years. It is often asymptomatic in its early stages and remains undiagnosed until late stages, where prognosis becomes unfavourable [2]. If detected early, up to 90% of deaths can be prevented [5]. As a result, a well-planned public health policy with the development of effective and non-invasive biomarkers could overcome the problem.

## Colorectal Cancer Screening Program

Implementation of CRC screening programs in communities allows early detection of colonic neoplasm(s) to lower the treatment need, morbidity, and mortality [6]. However, CRC screening programs in different countries differ in their approach [7]. These programs can be broadly divided by structured opportunistic and population-based organised (pilot) screening programs (Table 1) [4]. Population-based organised programs have been introduced into the United Kingdom (UK), Croatia, and Hong Kong, with the governments providing a well-organised systematic process of inviting a specific group of individuals for testing. By contrast, structured opportunistic screening programs are implemented on an ad hoc basis, usually through fee-for-service reimbursement of physicians, such as the United States (US) (Table 2).

### *The United Kingdom*

The UK is a typical example of a population-based organised screening program, where the National Health Service (NHS) has been providing a free-for-charge nationwide Bowel Cancer Screening Program (BCSP) for UK residents since 2006. The BCSP was originally intended for the population between 60 to 69 years of age and recently extended the age range to between age 50 to 74 for their biannual tests. To increase detection accuracy, the screening guidelines have shifted from guaiac faecal occult blood test (gFOBT) to faecal immunochemical test (FIT) since April 2018. Moreover, the NHS also offers a one-off flexible sigmoidoscopy at the age of 55 [8].

### *Croatia*

The early CRC screening program in Croatia was established in 2007 following recommendations by the European Council in 2003. The program provides a non-invasive gFOBT for the population aged 50 to 74 every 2 years. Positive cases of gFOBT may further be referred for a colonoscopy to confirm the finding [9]. However, the participation rate was below 20% for 5 years (2007 to 2011), the lowest rate in the European Union [4].

### *Hong Kong*

The Hong Kong government has a CRC screening program for citizens in the age range of 50 to 74, which is considered an average risk age group. Eligible citizens should receive a FIT every 2 years in this screening program. The guideline from the Department of Health in Hong

**Table 1.** Global status of structured and organised colorectal cancer screening by continent in 2018

Continent	Population-based organised	Population-based organised pilot	Structured opportunistic
Europe	<ol style="list-style-type: none"> <li>1. Belgium</li> <li>2. Croatia</li> <li>3. Czech Republic</li> <li>4. Denmark</li> <li>5. Estonia</li> <li>6. France</li> <li>7. Ireland</li> <li>8. Italy</li> <li>9. Lithuania</li> <li>10. Luxembourg</li> <li>11. Malta</li> <li>12. Montenegro</li> <li>13. The Netherlands</li> <li>14. Norway</li> <li>15. Poland</li> <li>16. Slovenia</li> <li>17. Spain</li> <li>18. United Kingdom</li> </ol>	<ol style="list-style-type: none"> <li>1. Austria</li> <li>2. Cyprus</li> <li>3. Georgia</li> <li>4. Hungary</li> <li>5. Portugal</li> <li>6. Serbia</li> <li>7. Sweden</li> <li>8. Switzerland</li> </ol>	<ol style="list-style-type: none"> <li>1. Austria</li> <li>2. Germany</li> <li>3. Greece</li> <li>4. Latvia</li> </ol>
North and Latin America	<ol style="list-style-type: none"> <li>1. Canada</li> <li>2. Uruguay</li> </ol>	<ol style="list-style-type: none"> <li>1. Argentina</li> <li>2. Brazil</li> <li>3. Chile</li> </ol>	<ol style="list-style-type: none"> <li>1. USA</li> <li>2. Colombia</li> </ol>
Africa	–	–	<ol style="list-style-type: none"> <li>1. Morocco</li> </ol>
Central, West, South Asia	<ol style="list-style-type: none"> <li>1. Israel</li> <li>2. UAE</li> </ol>	<ol style="list-style-type: none"> <li>1. Bahrain</li> <li>2. Kuwait</li> <li>3. Kazakhstan</li> <li>4. Lebanon</li> <li>5. Qatar</li> <li>6. Saudi Arabia</li> </ol>	<ol style="list-style-type: none"> <li>1. Iran</li> </ol>
Far East Asia and Oceania	<ol style="list-style-type: none"> <li>1. Taiwan</li> <li>2. Korea</li> <li>3. Hong Kong</li> <li>4. Singapore</li> <li>5. Australia</li> <li>6. New Zealand</li> </ol>	<ol style="list-style-type: none"> <li>1. PR China</li> <li>2. Thailand</li> </ol>	<ol style="list-style-type: none"> <li>1. Japan</li> <li>2. Malaysia</li> </ol>

Kong also recommended self-funded invasive screening, such as sigmoidoscopy every 5 years or colonoscopy every 10 years [10]. A similar screening program can also be found in Macau and Taiwan [11, 12].

#### *The United States*

In the US, the CRC screening program is largely opportunistic, and the guidelines are relying on both government institutions as well as national independent bodies, such as the US Preventive Services Task Force and the American Cancer Society (ACS). These organisations provide their professional guidelines for the choice of CRC screening tests according to the latest prevention and evidence-based medicine [4]. Currently, CRC screening is indicated for the patients aged 50 to 75, although the 2018 ACS guideline

**Table 2.** Colorectal cancer screening programs in the selected countries

	United States	United Kingdom	Hong Kong/ Macau/Taiwan	Croatia
Screening program	Opportunistic	Population-based	Population-based	Population-based
Age	45 to 75 76 to 85: consult	60 to 75	50 to 75	50 to 74
Non-invasive	Annual gFOBT Annual FIT Triennial mt-sDNA	Biennial FIT	Biennial FIT	Biennial gFOBT
Invasive	CT colonography (every 5 years) Flexible sigmoidoscopy (every 5 years) Colonoscopy (every 10 years)	One-off flexible sigmoidoscopy for age 55	Sigmoidoscopy (every 5 years)* Colonoscopy (every 10 years)*	Colonoscopy (when gFOBT positive)

gFOBT, guaiac faecal occult blood test; FIT, faecal immunochemical test; mt-sDNA, multi-target faecal-based DNA screening test.  
 \* Fee-of-charge service.

**Table 3.** Clinically available non-invasive CRC screening tools

Screening tool	Sample	Detection target	Specificity, % (95% CI) [Ref.]	Sensitivity, % (95% CI) [Ref.]	Advanced adenoma sensitivity, % (95% CI) [Ref.]	Cost, USD [Ref.]
mSEPT9 (Epi proColon)	Serum	<i>SEPT9</i> DNA methylation	92 (89–94) [70]	71 (67–75) [70]	11.2 (7.2–15.7) [71]	273–445 [71]
mt-sDNA (Cologuard)	Faeces	<i>NDRG4</i> and <i>BMP3</i> DNA methylation, <i>KRAS</i> mutations and haemoglobin	89.8 (88.9–90.7) [72]	92.3 (83–97.5) [72]	42.4 (38.9–46) [72]	492.72 [73]
FIT	Faeces	Haemoglobin	90* [74]	78* [74]	39* [74]	20–21.65 [73, 75]
gFOBT	Faeces	Haemoglobin	90.0 (84.2–93.8) [76]	62.6 (34.9–83.9) [76]	–	3.31–5 [71, 75]

\* No 95% CI reported from the meta-analysis. 95% CI, 95% confidence interval; gFOBT, guaiac faecal occult blood test; FIT, faecal immunochemical test; mt-sDNA, multi-target stool DNA test; mSEPT9, plasma *SEPT9* DNA methylation test.

recommended that screening should begin at the age of 45 and does not recommend CRC screening for anyone over 85. People in the age range between 76 and 85 should consult their medical providers. The ACS guideline also recommended a regular faecal-based non-invasive examination, such as FIT (every year) and mt-sDNA (every 3 years) [13]. gFOBT is no longer recommended due to the high false-positive rate as well as the dietary and pharmaceutical restrictions [14, 15]. The ACS guideline also proposed visual invasive examinations, such as CT colonography (every 5 years), flexible sigmoidoscopy (every 5 years) or colonoscopy (every 10 years) [16]. Although opportunistic screening is effective in reducing CRC-related mortality in the US [17], access to CRC screening is not equal [4]. Residents who are in poverty, uninsured, or underinsured are less likely to undergo regular CRC screening [18, 19].

## Approved Colorectal Cancer Screening Tools

### *Faecal Occult Blood Test and Faecal Immunochemical Test*

Currently, the most common and low-cost non-invasive faecal tests for CRC screening are gFOBT and FIT (Table 3). Both gFOBT and FIT enable detection of a tiny amount of blood by targeting haemoglobin [20]. Individuals with a positive gFOBT or FIT result may receive a gold-standard, invasive colonoscopy to confirm the results and/or removal of polyp(s). A meta-analysis of four randomised controlled trials revealed that annual or biennial gFOBT screening caused roughly a 16% reduction in CRC-related mortality with no significant effect on CRC incidence [21]. However, gFOBT is limited by its relatively poor sensitivity in advanced colonic adenoma and also requires repeat screening and dietary restrictions [22]. Thus, it is gradually being replaced by FIT [23, 24]. FIT has a relatively better detection accuracy and can be quantified, providing a tailored screening approach by optimising the cut-off level [25, 26]. A low cut-off reduces the specificity and requires more follow-up with colonoscopy, but increases the sensitivity to identify more individuals with precancerous polyp(s) [23, 27]. Further optimisation for FIT to improve detection accuracy is still ongoing, including the formulation of a FIT buffer for haemoglobin stabilisation [28, 29], the best haemoglobin detection concentration for automated FIT systems, as well as a single-sample (1-FIT) and two-sample (2-FIT) faecal sample protocol per one specimen [30]. Although improvement in FIT detection is still ongoing, a more accurate non-invasive test is urgently needed. At present, the Food and Drug Administration (FDA) in the US approved two other commercially available CRC tests in clinical use, including the multi-target stool DNA (mt-sDNA) test (Cologuard) and plasma *SEPT9* DNA methylation test (Epi proColon) (Table 3).

### *Multi-Target Stool DNA Test (Cologuard)*

Since colonocytes consistently exfoliate and shed into the lumen of the gastrointestinal tract, molecular alterations in faeces, such as DNA methylations, have been widely investigated [31]. The mt-sDNA screening test (also called Cologuard) is an FDA-approved non-invasive CRC screening tool in 2014, developed by EXACT Sciences Corporation (NASDAQ: EXAS) and Mayo Clinic [32]. The test is Clinical Laboratory Improvement Amendments (CLIA) certified and accredited by the College of American Pathologists [33]. It is designed to detect faeces-based DNA biomarkers with occult haemoglobin. The initial development utilised a pre-commercial 23-marker assay, with subsequent findings that there were 3 broadly informative markers for colorectal neoplasia [34]. Based on the findings, the preliminary version of mt-sDNA utilised *NDRG4*, *BMP3*, *VIM*, and *TFP12* genes as DNA methylation targets, with mutant *KRAS* and faecal haemoglobin. At the threshold of 90% specificity from the 293 healthy controls, the sensitivities for CRC ( $n = 252$ ) and adenomas  $\geq 1$  cm ( $n = 133$ ) were 85 and 54%, respectively [35]. The size of the tumour correlated to the detection sensitivity, increasing from 54% in 1 cm to 92% in  $\geq 4$  cm adenomas [35]. Afterwards, the commercially available, FDA-approved mt-sDNA version 2.0 by the “DeeP-C” study utilised *NDRG4* and *BMP3* DNA as methylation markers with *KRAS* mutations plus FIT. The “DeeP-C” prospective study recruited almost ten thousand participants in an average-risk community asymptomatic of CRC [36]. The sensitivity for CRC and pre-cancerous lesions was 92.3 and 42.4%, respectively, and presented a higher sensitivity compared to only one FIT kit at one cut-off (CRC: 73.8%,  $p = 0.002$ ; pre-cancerous lesions: 23.8%,  $p < 0.001$ ) [36].

Following the approval, further clinical trials were continued. In the Alaska native cohort ( $n = 661$ ) [37], the test presented a sensitivity of 49% for advanced colorectal neoplasms ( $n = 92$ ) versus 28% for FIT ( $p < 0.001$ ). The specificity of mt-sDNA was 93%, which is 3% lower than FIT ( $p = 0.034$ ) in the subjects in whom no adenomas were detected [37]. Later,

the mt-sDNA test was applied in frozen samples ( $n = 1,047$ ) from the Netherlands prospective COCOS study for further FIT comparison in advanced colorectal neoplasia ( $n = 102$ ). The mt-sDNA had a sensitivity of 49% and specificity of 89%, showing better accuracy than FIT, which had a sensitivity of 25% and specificity of 96% [38, 39]. An additional clinical trial at the Netherlands cancer institute is still ongoing [40]. It should be taken into account that a positive test result with no findings on colonoscopy may be due to other causes as *NDRG4* and *BMP3* methylation can be found from other gastrointestinal diseases such as gastric and pancreatic cancers, although it is rare [41–43].

#### *Plasma SEPT9 DNA Methylation Test (Epi proColon)*

In addition to faeces, DNA methylation can also be determined from blood. The *SEPT9* methylation detection in plasma has been evaluated in multiple studies. Epigenomics AG (ECX: FRA) in Germany first implemented the *SEPT9* methylation biomarker in Europe in 2008 [44]. Two years later, the commercialised Epi proColon qPCR kit version 1.0 was launched in Europe and later upgraded to version 2.0 [45]. The Chinese Food and Drug Administration and the US FDA approved the Epi proColon kit in 2015 and 2016, respectively. In the prospective “PRESEPT” study, the methylated *SEPT9* assay demonstrated a sensitivity of 48% for CRC (from stages I–IV, 35, 63, 46, and 77%, respectively) with a specificity of 92%; however, merely 11% of advanced adenomas were identified [46]. The commercially available kit provides 2 different algorithms, the 2/3 algorithm test has a relatively high true negative rate, while the sensitivity is higher in the 1/3 algorithm [47]. A meta-analysis study published in 2017 including 25 research articles found that the *SEPT9* assay is only superior to the FIT in the symptomatic population [48]. Due to its relatively poor sensitivity, the US Preventive Services Task Force and the ACS currently do not include the Epi proColon test in their CRC screening guidelines [49].

### Screening Tools under Development

One of the biggest challenges in early cancer diagnosis and/or prognosis is the lack of reliable biomarkers, leading to several screening tools, such as PreGen-Plus™ (*KRAS*, *APC*, and *p53* mutations) [50], ColoSure™ (*VIM* methylation), and COLVERA™ (*BCAT1* and *IKZF1* methylation) [51], being withdrawn from the market [50, 52]. Therefore, developing a low-invasive biomarker that can be easily performed with a clear clinical outcome is necessary. Apart from DNA methylation tests in both faeces and blood samples, other molecular biomarkers are being developed for CRC screening, such as circulating tumour DNA (ctDNA) [53], tumour-derived circulating cell (CTC) [54], circular RNA (circRNA) [55], PIWI-interacting RNA (piRNA) [56], microRNA (miRNA) [57–60], and gut microbes (Table 4). Studies relating to miRNA and gut bacteria will further be discussed.

#### *microRNAs Detection in Blood and Faeces*

miRNAs are a class of conserved endogenous, non-coding RNAs with approximately 18–24 nucleotides and play an important role in post-transcriptional regulation of protein-coding gene expression(s) through binding primarily to the 3′-untranslated region of the target mRNA(s), resulting in mRNA degradation and/or translational repression [61]. Thus, aberrant miRNA expression leads to disease progression and thus can be useful as diagnostic and/or prognostic predictors to human diseases. Until now, numerous research articles have reported that both blood- and faecal-based miRNAs can be utilised as biomarkers for CRC screening. Among them, miR-21 and miR-92a are the highly reported miRNAs for CRC screening [62].

**Table 4.** Selected developing molecular biomarkers for colorectal cancer screening

Types	Sample	Molecular biomarker(s)	AUC (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Ref.	
DNA methylation	Plasma	<i>APC, MGMT, RASSF2A, WIF1</i>	0.927	86.5 (81.7–90.8)	92.1 (88.2–95.0)	[81]	
	Plasma	<i>BCAT1, IKZF1</i>	NA	66	94	[82]	
	Serum	<i>SDC2</i>	NA	87	95.2	[83]	
	Plasma cfDNA	<i>LINE-1</i>	0.81	65.8	90	[84]	
	Faeces		<i>NDRG4</i>	T: 0.77	T: 61 (43–79)	T: 93 (90–97)	[85]
				V: NA	V: 53 (39–67)	V: 100 (86–100)	
Faeces		<i>TFPI2</i>	NA	76 (60–88)	93 (77–99)	[86]	
Circulating tumour DNA	Serum cfDNA	<i>ALU115</i>	0.8458	69.23	99.09	[87]	
		<i>ALU247/115</i>	0.8551	73.08	97.27		
Circular RNA (circRNA)	Plasma	<i>91H, PVT1, MEG3</i>	0.877	82.76	78.57	[88]	
	Serum	<i>LOC285194, RP11-462C24.1, Nbla12061</i>	0.79 (0.71–0.86)	68.33	86.89	[89]	
	Whole-blood	<i>NEAT1_v1</i>	0.73 (0.60–0.83)	56.7	83.3	[90]	
	Whole-blood	<i>NEAT1_v2</i>	0.85 (0.73–0.93)	86.6	83.3	[90]	
PIWI-interacting RNA (piRNA)	Serum	<i>piR-5937</i>	T: 0.8060	71.8	72.5	[56]	
			V: 0.7673	73.6	65.3		
		<i>piR-28876</i>	T: 0.8065	75.3	70.0	[56]	
			V: 0.7074	66.0	65.3		
microRNA (miRNA)	Plasma	<i>miR-92a</i>	0.885	89	70	[91]	
	Serum	<i>miR-210</i>	0.82	74.6	73.5	[92]	
	Plasma	<i>miR-24</i>	8.84 (0.79–0.89)	78.4	83.9	[93]	
	Faeces	<i>miR-221</i>	0.73 (0.68–0.78)	62 (55–68)	74 (67–80)	[60]	
	Faeces	<i>miR-20a</i>	0.73 (0.68–0.78)	55 (47–62)	82 (76–87)	[57]	
	Faeces	<i>miR-135b</i>	0.79	78 (69–85)	68 (58–77)	[59]	
	Faeces	<i>miR-92a, miR-21, miR-135b, miR-145, miR-133a</i>	0.849	81	80	[94]	
	Saliva	<i>miR-21</i>	NA	97	91	[95]	
Exosomal microRNA	Plasma	<i>miR-27a</i>	0.87 (0.77–0.96)	81.82	90.91	[96]	
	Plasma	<i>miR-130a</i>	0.82(0.73–0.90)	69.32	100		
Tumour-derived circulating cell	Whole-blood	Circulating endothelial cell clusters	0.92 (0.84–1.00)	NA	NA	[54]	
Gut microbes	Faeces	<i>F. nucleatum, Parvimonas micra</i>	0.84	NA	NA	[76]	
	Faeces	<i>F. nucleatum, Clostridium hathewayi, Lachnoclostridium sp., Bacteroides clarus, and FIT</i>	NA	93.8	81.2	[80]	

95% CI, 95% confidence interval; NA, not available; T, training; V, validation; cfDNA, cell-free DNA.



In a meta-analysis of the blood-based miRNA study, the overall sensitivity and specificity of blood-based miRNAs for CRC is 76% (95% CI, 72–80%) and 76% (95% CI, 72–80%), respectively [63]. The most predictive miRNA was miR-92a, ranging from a sensitivity of 65.5% to 89%, and from a specificity of 70% to 82.5%, with the area under the receiver-operating characteristics (AUC) between 0.786 and 0.890 [63]. However, the major shortcoming of using blood-based miRNA for CRC screening is the detection specificity. This is because miRNAs might arise from other cancer(s) [64, 65], depression [66], and virus infection(s) [67, 68]. Therefore, faecal-based miRNA detection may be an alternative option [57, 59, 60, 69]. It has been demonstrated that miRNAs are highly stable short sequences which remain detectable within samples throughout a 72-hour incubation period due to protection from ribonuclease degradation by exosomes [70, 71]. A meta-analysis showed that miR-21 is the most reliable miRNA [72]. However, as faeces contain abundant amounts of proteins and DNA from gut microbes, the purity of RNA samples from faeces may determine the result outcomes. As a result, faecal-miRNA detection in combination with FIT is a reliable approach to enhance the detection accuracy. Previous studies indicated that the combination of miR-21 and miR-92a with FIT had a specificity of 96.8% and sensitivity of 78.4%, while FIT alone only had a specificity of 98.4% and sensitivity of 66.7% [73].

#### *Faecal-Based Microbe Detection*

The gut flora habitat has a vast amount of microbes and plays an important role in maintaining our health. Changes in microbiome composition have been linked to multiple diseases including cancer. The study of biomarkers from the gut microbiome has a particular focus on CRC, where clinical use is already on the horizon [74]. It is known that dysbacteriosis alters metabolic activities and induces inflammatory stimuli to the gastrointestinal tract, and eventually induces mutations to colonic cells, thus contributing to the development of the CRC. Among multiple microbial taxonomic markers, intensive research showed that *Fusobacterium nucleatum* is enriched in tumour neoplasms as well as in faeces from CRC patients [75, 76]. *F. nucleatum* belongs to the class of asaccharolytic bacteria. Enrichment of *F. nucleatum* in the gut does not only recruit tumour-infiltrating immune cells and induces a pro-inflammatory microenvironment, but also contributes to CRC tumorigenesis through its strong adhesive abilities and invasive effects on epithelial cells [77]. Furthermore, *F. nucleatum* survives and divides in the hypoxic tumour microenvironment, contributing to the cell proliferation and angiogenesis. A meta-analysis indicated that the use of *F. nucleatum* as a biomarker for CRC screening has a sensitivity of 71% (95% CI, 61–79%) and a specificity of 76% (95% CI, 66–84%), with the AUC of 0.80 (95% CI, 0.76–0.83); the sensitivity and specificity for advanced colorectal neoplasia is 36% (95% CI, 27–46%) and 73% (95% CI, 65–79%), respectively, with the AUC of 0.60 (95% CI, 0.56–0.65) [78]. The use of *F. nucleatum* together with FIT may improve the detection accuracy for advanced colorectal neoplasia [79]. A most recent report indicated that the combination of *F. nucleatum*, *Clostridium hathewayi*, *Lachnoclostridium* sp., *Bacteroides clarus*, and FIT presented a high detection accuracy, with a specificity of 81.2% and sensitivity of 93.8% [80].

#### **Conclusion**

CRC is the third most aggressive cancer worldwide with a high mortality rate due to the lack of robust biomarkers. Current CRC screening programs are mostly only available to those above age 50 or 55, despite the latest guidelines recommending that screening should begin from the age of 45. Although colonoscopy is the gold standard for CRC, it being labour-intensive and invasive means that it cannot be applied for everyone. Thus, there is a great

need for a cost-effective and non-invasive CRC screening test to improve the screening accuracy and acceptability. The use of circulating and/or faecal-based miRNAs, as well as gut bacteria, could be the next generation CRC screening biomarkers.

### Disclosure Statement

The authors have no conflicts of interest to declare.

### Funding Sources

Not available.

### Author Contributions

Conceptualisation and writing of the main text and tables, T.O.Y.; revision and proofreading of the manuscript, M.T. All authors have read and agreed to the published version of the manuscript.

### References

- 1 Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018 Nov;68(6):394–424.
- 2 Yau TO. Precision treatment in colorectal cancer: now and the future. *JGH Open*. 2019 Feb;3(5):361–9.
- 3 Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017 Apr;66(4):683–91.
- 4 Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJ, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. *Gut*. 2015 Oct;64(10):1637–49.
- 5 Simon K. Colorectal cancer development and advances in screening. *Clin Interv Aging*. 2016 Jul;11:967–76.
- 6 Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med*. 2012 Feb;366(8):687–96.
- 7 Young GP, Rabeneck L, Winawer SJ. The Global Paradigm Shift in Screening for Colorectal Cancer. *Gastroenterology*. 2019 Mar;156(4):843–851.e2.
- 8 Koo S, Neilson LJ, Von Wagner C, Rees CJ. The NHS Bowel Cancer Screening Program: current perspectives on strategies for improvement. *Risk Manag Healthc Policy*. 2017 Dec;10:177–87.
- 9 Navarro M, Nicolas A, Ferrandez A, Lanás A. Colorectal cancer population screening programs worldwide in 2016: an update. *World J Gastroenterol*. 2017 May;23(20):3632–42.
- 10 Lam TH, Wong KH, Chan KK, Chan MC, Chao DV, Cheung AN, et al.; Cancer Expert Working Group on Cancer Prevention and Screening. Recommendations on prevention and screening for colorectal cancer in Hong Kong. *Hong Kong Med J*. 2018 Oct;24(5):521–6.
- 11 Wang YW, Chen HH, Wu MS, Chiu HM; Taiwanese Nationwide Colorectal Cancer Screening Program. Current status and future challenge of population-based organized colorectal cancer screening: lesson from the first decade of Taiwanese program. *J Formos Med Assoc*. 2018 May;117(5):358–64.
- 12 Colorectal Cancer Screening Program [Internet]. Serviços Saúde do Gov da Região Adm Espec Macau. [cited 2019 Jun 30]. Available from: <https://www.ssm.gov.mo/apps1/coloncancer/en.aspx>
- 13 Wolf AM, Fontham ET, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA Cancer J Clin*. 2018 Jul;68(4):250–81.
- 14 Robertson DJ, Lee JK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, et al. Recommendations on Fecal Immunochemical Testing to Screen for Colorectal Neoplasia: A Consensus Statement by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology*. 2017 Apr;152(5):1217–1237.e3.
- 15 Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling JW Jr, García FA, et al.; US Preventive Services Task Force. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2016 Jun;315(23):2564–75.
- 16 American Cancer Society updates its colorectal cancer screening guideline: new recommendation is to start screening at age 45 years. *Cancer*. 2018 Sep;124(18):3631–2.
- 17 Sur D, Colceriu M, Sur G, Floca E, Dascal L, Irimie A. Colorectal cancer: evolution of screening strategies. *Med Pharm Reports*. 2019 Jan;92(1):21–4. <https://doi.org/10.15386/cjmed-1104>.

- 18 Welch HG, Robertson DJ. Colorectal Cancer on the Decline – Why Screening Can't Explain It All. *N Engl J Med*. 2016 Apr;374(17):1605–7.
- 19 de Moor JS, Cohen RA, Shapiro JA, Nadel MR, Sabatino SA, Robin Yabroff K, et al. Colorectal cancer screening in the United States: trends from 2008 to 2015 and variation by health insurance coverage. *Prev Med*. 2018 Jul;112:199–206.
- 20 Young GP, Symonds EL, Allison JE, Cole SR, Fraser CG, Halloran SP, et al. Advances in Fecal Occult Blood Tests: the FIT revolution. *Dig Dis Sci*. 2015 Mar;60(3):609–22.
- 21 Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. *Am J Gastroenterol*. 2008 Jun;103(6):1541–9.
- 22 Brenner H, Hoffmeister M, Birkner B, Stock C. Diagnostic performance of guaiac-based fecal occult blood test in routine screening: state-wide analysis from Bavaria, Germany. *Am J Gastroenterol*. 2014 Mar;109(3):427–35.
- 23 Hol L, Wilschut JA, van Ballegooijen M, van Vuuren AJ, van der Valk H, Reijerink JC, et al. Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. *Br J Cancer*. 2009 Apr;100(7):1103–10.
- 24 van Rossum LG, van Rijn AF, Laheij RJ, van Oijen MG, Fockens P, van Krieken HH, et al. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. *Gastroenterology*. 2008 Jul;135(1):82–90.
- 25 Wilschut JA, Hol L, Dekker E, Jansen JB, Van Leerdam ME, Lansdorp-Vogelaar I, et al. Cost-effectiveness analysis of a quantitative immunochemical test for colorectal cancer screening. *Gastroenterology*. 2011 Nov;141(5):1648–55.e1.
- 26 Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, Habbema JD, Kuipers EJ. Effect of rising chemotherapy costs on the cost savings of colorectal cancer screening. *J Natl Cancer Inst*. 2009 Oct;101(20):1412–22.
- 27 Kim DH, Pickhardt PJ, Taylor AJ. Characteristics of advanced adenomas detected at CT colonographic screening: implications for appropriate polyp size thresholds for polypectomy versus surveillance. *AJR Am J Roentgenol*. 2007 Apr;188(4):940–4.
- 28 Catomeris P, Baxter NN, Boss SC, Paszat LF, Rabeneck L, Randell E, et al. Effect of Temperature and Time on Fecal Hemoglobin Stability in 5 Fecal Immunochemical Test Methods and One Guaiac Method. *Arch Pathol Lab Med*. 2018 Jan;142(1):75–82.
- 29 Grazzini G, Ventura L, Rubeca T, Rapi S, Cellai F, Di Dia PP, et al. Impact of a new sampling buffer on faecal haemoglobin stability in a colorectal cancer screening programme by the faecal immunochemical test. *Eur J Cancer Prev*. 2017 Jul;26(4):285–91.
- 30 Liles EG, Perrin N, Rosales AG, Smith DH, Feldstein AC, Mosen DM, et al. Performance of a quantitative fecal immunochemical test for detecting advanced colorectal neoplasia: a prospective cohort study. *BMC Cancer*. 2018 May;18(1):509.
- 31 Li B, Gan A, Chen X, Wang X, He W, Zhang X, et al. Diagnostic Performance of DNA Hypermethylation Markers in Peripheral Blood for the Detection of Colorectal Cancer: A Meta-Analysis and Systematic Review. *PLoS One*. 2016 May;11(5):e0155095.
- 32 Ahlquist DA. Multi-target stool DNA test: a new high bar for noninvasive screening. *Dig Dis Sci*. 2015 Mar;60(3):623–33.
- 33 Prince M, Lester L, Chiniwala R, Berger B. Multitarget stool DNA tests increases colorectal cancer screening among previously noncompliant Medicare patients. *World J Gastroenterol*. 2017 Jan;23(3):464–71.
- 34 Ahlquist DA, Sargent DJ, Loprinzi CL, Levin TR, Rex DK, Ahnen DJ, et al. Stool DNA and occult blood testing for screen detection of colorectal neoplasia. *Ann Intern Med*. 2008 Oct;149(7):441–50.
- 35 Ahlquist DA, Zou H, Domanico M, Mahoney DW, Yab TC, Taylor WR, et al. Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. *Gastroenterology*. 2012 Feb;142(2):248–56.
- 36 Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med*. 2014 Jul;371(2):187–8.
- 37 Redwood DG, Asay ED, Blake ID, Sacco PE, Christensen CM, Sacco FD, et al. Stool DNA Testing for Screening Detection of Colorectal Neoplasia in Alaska Native People. *Mayo Clin Proc*. 2016 Jan;91(1):61–70.
- 38 Dublin Pathology. Dublin Pathology 2015. 8th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, 23–25 June 2015. *J Pathol*. 2015 Sep;237 Suppl 1:S1–52.
- 39 Berger BM, Levin B, Hilsden RJ. Multitarget stool DNA for colorectal cancer screening: A review and commentary on the United States Preventive Services Draft Guidelines. *World J Gastrointest Oncol*. 2016 May;8(5):450–8.
- 40 van Lanschot MC, Carvalho B, Coupé VM, van Engeland M, Dekker E, Meijer GA. Molecular stool testing as an alternative for surveillance colonoscopy: a cross-sectional cohort study. *BMC Cancer*. 2017 Feb;17(1):116.
- 41 Majumder S, Raimondo M, Taylor WR, Yab TC, Berger CK, Dukek BA, et al. Methylated DNA in Pancreatic Juice Distinguishes Patients With Pancreatic Cancer From Controls. *Clin Gastroenterol Hepatol*. 2020 Mar;18(3):676–683.e3.
- 42 Chen X, Yang Y, Liu J, Li B, Xu Y, Li C, et al. NDRG4 hypermethylation is a potential biomarker for diagnosis and prognosis of gastric cancer in Chinese population. *Oncotarget*. 2017 Jan;8(5):8105–19.
- 43 Yu C, Hao X, Zhang S, Hu W, Li J, Sun J, et al. Characterization of the prognostic values of the NDRG family in gastric cancer. *Therap Adv Gastroenterol*. 2019 Jul;12:1756284819858507.
- 44 Payne SR. From discovery to the clinic: the novel DNA methylation biomarker (m)SEPT9 for the detection of colorectal cancer in blood. *Epigenomics*. 2010 Aug;2(4):575–85.

- 45 Issa IA, Nouredine M. Colorectal cancer screening: an updated review of the available options. *World J Gastroenterol*. 2017 Jul;23(28):5086–96.
- 46 Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, et al.; PRESEPT Clinical Study Steering Committee, Investigators and Study Team. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut*. 2014 Feb;63(2):317–25.
- 47 Bretthauer M, Kaminski MF, Løberg M, Zauber AG, Regula J, Kuipers EJ, et al.; Nordic-European Initiative on Colorectal Cancer (NordICC) Study Group. Population-Based Colonoscopy Screening for Colorectal Cancer: A Randomized Clinical Trial. *JAMA Intern Med*. 2016 Jul;176(7):894–902.
- 48 Song L, Jia J, Peng X, Xiao W, Li Y. The performance of the SEPT9 gene methylation assay and a comparison with other CRC screening tests: A meta-analysis. *Sci Rep*. 2017 Jun;7(1):3032.
- 49 Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling JW Jr, García FA, et al.; US Preventive Services Task Force. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2016 Jun;315(23):2564–75.
- 50 Hamzehzadeh L, Yousefi M, Ghaffari S-H. Colorectal Cancer Screening: A Comprehensive Review to Recent Non-Invasive Methods. *Int J Hematol Stem Cell Res*. 2017 Jul;11(3):250–61.
- 51 Symonds EL, Pedersen SK, Murray D, Byrne SE, Hollington P, Rabbitt P, et al. Performance comparison of the methylated BCAT1 / IKZF1 ctDNA test (COLVERA) with the CEA assay for detection of recurrent colorectal cancer. *J Clin Oncol*. 2019 May;37(15\_suppl):3589–3589.
- 52 Locke WJ, Guanzon D, Ma C, Liew YJ, Duesing KR, Fung KY, et al. DNA Methylation Cancer Biomarkers: translation to the Clinic. *Front Genet*. 2019 Nov;10:1150.
- 53 Reinert T, Henriksen TV, Christensen E, Sharma S, Salari R, Sethi H, et al. Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer. *JAMA Oncol*. 2019 May;5(8):1124.
- 54 Cima I, Kong SL, Sengupta D, Tan IB, Phyo WM, Lee D, et al. Tumor-derived circulating endothelial cell clusters in colorectal cancer. *Sci Transl Med*. 2016 Jun;8(345):345ra89–345ra89.
- 55 Pan B, Qin J, Liu X, He B, Wang X, Pan Y, et al. Identification of Serum Exosomal hsa-circ-0004771 as a Novel Diagnostic Biomarker of Colorectal Cancer. *Front Genet*. 2019 Nov;10:1096.
- 56 Vychytilova-Falteskova P, Stitkovcova K, Radova L, Sachlova M, Kosarova Z, Slaba K, et al. Circulating PIWI-Interacting RNAs piR-5937 and piR-28876 Are Promising Diagnostic Biomarkers of Colon Cancer. *Cancer Epidemiol Biomarkers Prev*. 2018 Sep;27(9):1019–28.
- 57 Yau TO, Wu CW, Tang CM, Chen Y, Fang J, Dong Y, et al. MicroRNA-20a in human faeces as a non-invasive biomarker for colorectal cancer. *Oncotarget*. 2016 Jan;7(2):1559–68.
- 58 Yu G, Tang JQ, Tian ML, Li H, Wang X, Wu T, et al. Prognostic values of the miR-17-92 cluster and its paralogs in colon cancer. *J Surg Oncol*. 2012 Sep;106(3):232–7.
- 59 Wu CW, Ng SC, Dong Y, Tian L, Ng SS, Leung WW, et al. Identification of microRNA-135b in stool as a potential noninvasive biomarker for colorectal cancer and adenoma. *Clin Cancer Res*. 2014 Jun;20(11):2994–3002.
- 60 Yau TO, Wu CW, Dong Y, Tang CM, Ng SS, Chan FK, et al. microRNA-221 and microRNA-18a identification in stool as potential biomarkers for the non-invasive diagnosis of colorectal carcinoma. *Br J Cancer*. 2014 Oct;111(9):1765–71.
- 61 Liz J, Esteller M. lncRNAs and microRNAs with a role in cancer development. *Biochim Biophys Acta*. 2016 Jan;1859(1):169–76.
- 62 Toiyama Y, Okugawa Y, Fleshman J, Richard Boland C, Goel A. MicroRNAs as potential liquid biopsy biomarkers in colorectal cancer: A systematic review. *Biochim Biophys Acta Rev Cancer*. 2018 Dec;1870(2):274–82.
- 63 Carter JV, Galbraith NJ, Yang D, Burton JF, Walker SP, Galandiuk S. Blood-based microRNAs as biomarkers for the diagnosis of colorectal cancer: a systematic review and meta-analysis. *Br J Cancer*. 2017 Mar;116(6):762–74.
- 64 Lamichhane SR, Thachil T, De Ieso P, Gee H, Moss SA, Milic N. Prognostic Role of MicroRNAs in Human Non-Small-Cell Lung Cancer: A Systematic Review and Meta-Analysis. *Dis Markers*. 2018 Oct;2018:8309015.
- 65 Ma C, Nguyen HP, Luwor RB, Stylli SS, Gogos A, Paradiso L, et al. A comprehensive meta-analysis of circulation miRNAs in glioma as potential diagnostic biomarker. *PLoS One*. 2018 Feb;13(2):e0189452.
- 66 Liu S, Zhang F, Wang X, Shugart YY, Zhao Y, Li X, et al. Diagnostic value of blood-derived microRNAs for schizophrenia: results of a meta-analysis and validation. *Sci Rep*. 2017 Nov;7(1):15328.
- 67 Lai FW, Stephenson KB, Mahony J, Lichty BD. Human coronavirus OC43 nucleocapsid protein binds microRNA 9 and potentiates NF-κB activation. *J Virol*. 2014 Jan;88(1):54–65.
- 68 Fiorino S, Bacchi-Reggiani ML, Visani M, Acquaviva G, Fornelli A, Masetti M, et al. MicroRNAs as possible biomarkers for diagnosis and prognosis of hepatitis B- and C-related-hepatocellular-carcinoma. *World J Gastroenterol*. 2016 Apr;22(15):3907–36.
- 69 Wu CW, Ng SS, Dong YJ, Ng SC, Leung WW, Lee CW, et al. Detection of miR-92a and miR-21 in stool samples as potential screening biomarkers for colorectal cancer and polyps. *Gut*. 2012 May;61(5):739–45.
- 70 Hunter MP, Ismail N, Zhang X, Aguda BD, Lee EJ, Yu L, et al. Detection of microRNA expression in human peripheral blood microvesicles. *PLoS One*. 2008;3(11):e3694.
- 71 Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci USA*. 2008 Jul;105(30):10513–8.
- 72 Yau TO, Tang CM, Harriss EK, Dickins B, Polytaichou C. Faecal microRNAs as a non-invasive tool in the diagnosis of colonic adenomas and colorectal cancer: A meta-analysis. *Sci Rep*. 2019 Jul;9(1):9491.

- 73 Kanaoka S, Kuriyama S, Iwaizumi M, Yamada T, Sugimoto M, Osawa S, et al. Potential Usefulness of Fecal Immunochemical Test Plus Fecal MicroRNA Assay As a Marker for Colorectal Cancer Screening. *Gastroenterology*. 2013 May;144(5):S-599–600.
- 74 Gagnière J, Raisch J, Veziat J, Barnich N, Bonnet R, Buc E, et al. Gut microbiota imbalance and colorectal cancer. *World J Gastroenterol*. 2016 Jan;22(2):501–18.
- 75 Kwong TN, Wang X, Nakatsu G, Chow TC, Tipoe T, Dai RZ, et al. Association between Bacteremia from Specific Microbes and Subsequent Diagnosis of Colorectal Cancer. *Gastroenterology*. 2018 Aug;155(2):383–390.e8.
- 76 Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut*. 2017 Jan;66(1):70–8.
- 77 Wu J, Li Q, Fu X. *Fusobacterium nucleatum* Contributes to the Carcinogenesis of Colorectal Cancer by Inducing Inflammation and Suppressing Host Immunity. *Transl Oncol*. 2019 Jun;12(6):846–51.
- 78 Zhang X, Zhu X, Cao Y, Fang JY, Hong J, Chen H. Fecal *Fusobacterium nucleatum* for the diagnosis of colorectal tumor: A systematic review and meta-analysis. *Cancer Med*. 2019 Feb;8(2):480–91.
- 79 Wong SH, Kwong TN, Chow TC, Luk AK, Dai RZ, Nakatsu G, et al. Quantitation of faecal *Fusobacterium* improves faecal immunochemical test in detecting advanced colorectal neoplasia. *Gut*. 2017 Aug;66(8):1441–8.
- 80 Liang JQ, Li T, Nakatsu G, Chen Y-X, Yau TO, Chu E, et al. A novel faecal *Lachnospirillum* marker for the non-invasive diagnosis of colorectal adenoma and cancer. *Gut*. 2019 Nov;gutjnl-2019-318532.
- 81 Lee BB, Lee EJ, Jung EH, Chun HK, Chang DK, Song SY, et al. Aberrant methylation of APC, MGMT, RASSF2A, and Wif-1 genes in plasma as a biomarker for early detection of colorectal cancer. *Clin Cancer Res*. 2009 Oct;15(19):6185–91.
- 82 Pedersen SK, Symonds EL, Baker RT, Murray DH, McEvoy A, Van Doorn SC, et al. Evaluation of an assay for methylated BCAT1 and IKZF1 in plasma for detection of colorectal neoplasia. *BMC Cancer*. 2015 Oct;15(1):654.
- 83 Oh T, Kim N, Moon Y, Kim MS, Hoehn BD, Park CH, et al. Genome-wide identification and validation of a novel methylation biomarker, SDC2, for blood-based detection of colorectal cancer. *J Mol Diagn*. 2013 Jul;15(4):498–507.
- 84 Nagai Y, Sunami E, Yamamoto Y, Hata K, Okada S, Muroto K, et al. LINE-1 hypomethylation status of circulating cell-free DNA in plasma as a biomarker for colorectal cancer. *Oncotarget*. 2017 Feb;8(7):11906–16.
- 85 Melotte V, Lentjes MH, van den Bosch SM, Hellebrekers DM, de Hoon JP, Wouters KA, et al. N-Myc downstream-regulated gene 4 (NDRG4): a candidate tumor suppressor gene and potential biomarker for colorectal cancer. *J Natl Cancer Inst*. 2009 Jul;101(13):916–27.
- 86 Glöckner SC, Dhir M, Yi JM, McGarvey KE, Van Neste L, Louwagie J, et al. Methylation of TFPI2 in stool DNA: a potential novel biomarker for the detection of colorectal cancer. *Cancer Res*. 2009 Jun;69(11):4691–9.
- 87 Hao TB, Shi W, Shen XJ, Qi J, Wu XH, Wu Y, et al. Circulating cell-free DNA in serum as a biomarker for diagnosis and prognostic prediction of colorectal cancer. *Br J Cancer*. 2014 Oct;111(8):1482–9.
- 88 Liu H, Ye D, Chen A, Tan D, Zhang W, Jiang W, et al. A pilot study of new promising non-coding RNA diagnostic biomarkers for early-stage colorectal cancers. *Clin Chem Lab Med*. 2019 Jun;57(7):1073–83.
- 89 Wang C, Yu J, Han Y, Li L, Li J, Li T, et al. Long non-coding RNAs LOC285194, RP11-462C24.1 and Nbla12061 in serum provide a new approach for distinguishing patients with colorectal cancer from healthy controls. *Oncotarget*. 2016 Oct;7(43):70769–78.
- 90 Wu Y, Yang L, Zhao J, Li C, Nie J, Liu F, et al. Nuclear-enriched abundant transcript 1 as a diagnostic and prognostic biomarker in colorectal cancer. *Mol Cancer*. 2015 Nov;14(1):191.
- 91 Ng EK, Chong WW, Jin H, Lam EK, Shin VY, Yu J, et al. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. *Gut*. 2009 Oct;58(10):1375–81.
- 92 Wang W, Qu A, Liu W, Liu Y, Zheng G, Du L, et al. Circulating miR-210 as a diagnostic and prognostic biomarker for colorectal cancer. *Eur J Cancer Care (Engl)*. 2017 Jul;26(4):e12448.
- 93 Fang Z, Tang J, Bai Y, Lin H, You H, Jin H, et al. Plasma levels of microRNA-24, microRNA-320a, and microRNA-423-5p are potential biomarkers for colorectal carcinoma. *J Exp Clin Cancer Res*. 2015 Aug;34(1):86.
- 94 Liang JQ, Yau TO, Yau TC, Szeto CH, Zhao FS, Chan FK, et al. A Novel MicroRNA Panel for Non-Invasive Diagnosis and Prognosis of Colorectal Cancer. *Gastroenterology*. 2019 May;156(6):S-496.
- 95 Sazanov AA, Kiselyova EV, Zakharenko AA, Romanov MN, Zaraysky MI. Plasma and saliva miR-21 expression in colorectal cancer patients. *J Appl Genet*. 2017 May;58(2):231–7.
- 96 Liu X, Pan B, Sun L, Chen X, Zeng K, Hu X, et al. Circulating Exosomal miR-27a and miR-130a Act as Novel Diagnostic and Prognostic Biomarkers of Colorectal Cancer. *Cancer Epidemiol Biomarkers Prev*. 2018 Jul;27(7):746–54.

# Precision treatment in colorectal cancer: Now and the future

Tung On Yau

*JGH Open: An Open Access Journal of Gastroenterology and Hepatology.* **2019**; 3(5):361–69.

Journal URL: [onlinelibrary.wiley.com/doi/full/10.1002/jgh3.12153](https://onlinelibrary.wiley.com/doi/full/10.1002/jgh3.12153)

DOI: [10.1002/jgh3.12153](https://doi.org/10.1002/jgh3.12153)

PMID: [31633039](https://pubmed.ncbi.nlm.nih.gov/31633039/)

PMCID: [PMC6788378](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC6788378/)

REVIEW ARTICLE

# Precision treatment in colorectal cancer: Now and the future

Tung On Yau 

John van Geest Cancer Research Centre, School of Science and Technology, Nottingham Trent University, Nottingham, UK

**Key words**

chemotherapy, colorectal cancer, epidermal growth factor receptor, personalized medicine, precision treatment, vascular endothelial growth factor.

Accepted for publication 14 January 2019.

**Correspondence**

Tung On Yau, John van Geest Cancer Research Centre, School of Science and Technology, Nottingham Trent University, Nottingham, NG11 8NS, UK.

Email: tungon@gmail.com; payton.yau@ntu.ac.uk

**Declaration of conflict of interest:** The authors declare no conflict of interest.

**Abstract**

Until recently, a one-drug-fits-all model was applied to every patient diagnosed with the same condition. But not every condition is the same, and this has led to many cases of ineffective treatment. Pharmacogenetics is increasingly used to stratify patients for precision medicine treatments, for instance, the UGT1A1\*28 polymorphism as a dosage indicator for the use of irinotecan as well as epidermal growth factor receptor (EGFR) immunohistochemistry and KRAS Proto-Oncogene (*KRAS*) exon 2 mutation tests for determining the likelihood of treatment response to cetuximab or panitumumab treatment in metastatic colorectal cancer (CRC). The other molecular subtypes, such as *KRAS* exon 3/4, B-Raf Proto-Oncogene, *NRAF*, *PIK3CA*, and *PETN*, were also reported as potential new pharmacogenetic targets for the current and the newly discovered anticancer drugs. In addition to next-generation sequencing (NGS), primary tumor cells for *in vivo* and *in vitro* drug screening, imaging biomarker 3'-Deoxy-3'-18F-fluorothymidine positron emission tomography, and circulating tumor DNA (ctDNA) detection methods are being developed and may represent the future direction of precision medicine. This review will discuss the current environment of precision medicine, including clinically approved targeted therapies, the latest potential therapeutic agents, and the ongoing pharmacogenetic trials for CRC patients.

**INTRODUCTION**

Colorectal cancer (CRC) is the second most common cancer and cause of death across European countries. In 2012, approximately 447 000 Europeans were diagnosed, and 215 000 died from the disease.<sup>1</sup> Over the past few decades, patients with CRC were treated homogeneously and provided with the same “standard” care. In addition to the standard colorectal surgery, the recommendation of standard drug treatment based on the tumor staging has successfully improved the treatment efficacy for CRC patients in both overall survival (OS) and disease-free survival (DFS).<sup>2</sup> However, not every patient’s condition is the same, and decisions on treatment options made by relying solely on CRC staging is simplistic. This has likely led to many cases of ineffective treatment, adverse drug reactions, and multiple side effects.

Precision cancer treatment could be one of the possible ways to tackle this problem. Precision medicine, also known as personalized medicine, goes beyond a conventional one-drug-fits-all model to match therapy by using particular environmental, lifestyle, cancer staging, and biological characteristics to identify which approach will be most effective for a particular individual. This thereby increases his or her likelihood of response to treatment and reduces the number of adverse drug effects.<sup>3</sup>

Currently, there are several drugs that have been approved for CRC treatment, and a variety of pharmacogenetic tests involving biomarkers have been accepted to aid the patient

selection process (Table 1). The aim of this review is to discuss the current state of precision drug treatments, including clinically approved chemotherapy drugs, molecularly targeted therapies such as anti-VEGF (vascular endothelial growth factor) and anti-EGFR (epidermal growth factor receptor) treatments, and the latest ongoing clinical trials for CRC patients.

**Precision treatment and implications for early-stage CRC**

There are several methods for staging CRC, including the tumour, node, and metastases (TNM) system, Dukes classification, and Astler-Coller classification. Using the most common TNM staging system, CRC can be broadly subdivided into five phases (Table 2).<sup>4</sup> This staging system is important because it forms the basis for decisions regarding treatment options for CRC. For example, patients with stage I CRC normally receive colonoscopic polypectomy, endoscopic mucosal resection, or endoscopic submucosal dissection as their main form of treatment, whereas those with more advanced stages require surgical resection with or without (neo)adjuvant chemotherapy.<sup>5</sup>

More recent research has, however, suggested that a subset of patients with stage I CRC have lymph node metastasis (LNM) and requires additional surgery.<sup>6</sup> Unfortunately, current best practice lacks relevant risk assessment tools, and there is no clear definition of LNM for patients classified with T1 histopathology. This results in several patients being under- or overtreated,

**Table 1** Clinically approved drugs and its approved pharmacogenetic targets in colorectal cancer patients

Class of agent	Name	Biological target	Detection target <sup>†</sup>	U.S. FDA-approved testing kit for CRC (detection method)
Cytotoxic chemotherapy	5-FU	TS	DYPD	—
	Irinotecan	TOP1	UGT1A1*28	—
	Oxaliplatin	—	—	—
	Raltitrexed <sup>‡</sup>	TS	—	—
	Lonsurf (trifluridine/tipiracil)	TS	—	—
VEGF	Bevacizumab	VEGF-A	—	—
	Ziv-aflibercept	VEGF-A	—	—
	Ramucirumab	VEGFR-2	—	—
	Regorafenib	Series of protein kinases <sup>§</sup>	—	—
EGFR	Cetuximab	EGFR	1. EGFR	1. DAKO EGFR PharmDx Kit (IHC)
	Panitumumab	EGFR	2. KRAS exon 2 3 & 4	2. cobas <sup>®</sup> KRAS Test (qPCR) 3. theascreen KRAS Test (qPCR)

<sup>†</sup>U.S. FDA-approved pharmacogenomic biomarkers on drug labeling.

<sup>‡</sup>NICE UK-approved drug.

<sup>§</sup>Regorafenib targeted proteins are VEGF receptors 1–3, TIE2, KIT, RET, *RAF1*, *BRAF* V600E, PDGFR, and FGFR.

DYPD, Dihydropyrimidine Dehydrogenase [NADP(+)]; EGFR, epidermal growth factor receptor; IHC, immunohistochemistry; qPCR, quantitative reverse transcription polymerase chain reaction; TOP1, Topoisomerase 1; TS, thymidylate synthase; VEGF, vascular endothelial growth factor.

causing unnecessary treatment side effects and excess morbidity.<sup>7</sup> The use of biomarkers may aid in the further subclassification of this set of patients. One study has shown that EZR is a potential biomarker for LNM and that this may guide decisions about the need for further surgery.<sup>8</sup> A panel of five biomarkers—BMI, ETV6, H3F3B, RPS10, and VEGFA—was also shown to outperform clinicopathological prognostic factors for node-negative CRC.<sup>9</sup>

Current best practice recommends adjuvant chemotherapy for patients with stage II CRC and high-risk clinicopathological features, but there is also no consensus on how to define the high-risk characteristics.<sup>10</sup> Several molecular assays, such as ColoPrint and Oncotype DX, offer additional means for analyzing patients' risk of recurrence. In the Prospective Analysis of Risk Stratification by ColoPrint (PARSC) study (NCT00903565), relapse rates in stage II CRC were evaluated, and it was demonstrated that ColoPrint may improve the prognostic accuracy beyond the clinical variables and microsatellite instability (MSI) status.<sup>11</sup> The Oncotype DX Colon Cancer assay, which utilizes quantitative polymerase chain reaction (qPCR) to measure 12 biomarkers (seven cancer-related—BGN, C-MYC, FAP, GADD45B, INHBA, Ki-67, MYBL2; 5 reference genes—ATP5E, GPX1, PGK1, UBB, VDAC2), produces a score from 0 to 100, which represents the predicted recurrence risk to inform decisions regarding adjuvant chemotherapy for CRC patients.<sup>12,13</sup> It has been shown to predict recurrence risk more accurately than when using T-stage and mismatch repair status alone (NCT01479894).<sup>13</sup> Studies have also shown that other biomarkers, such as a lack of *CDX2* expression, may offer further insight into the subgroup of patients with high-risk stage II CRC who benefit from receiving adjuvant chemotherapy (5-year DFS: 91% vs 56%;  $P = 0.006$ ).<sup>14</sup>

## Chemotherapy drugs for precision treatment

Cytotoxic agents such as 5-fluorouracil (5-FU), irinotecan, and oxaliplatin are commonly used as chemotherapy agents for CRC treatment. However, a proportion of CRC patients does not respond to this chemotherapy regimen and/or suffer from severe drug toxicities. 5-FU is a widely used thymidylate synthase (TS) inhibitor that acts as an antimetabolite to block the pyrimidine thymidine synthesis required for DNA replication.<sup>15</sup> In the early years, studies demonstrated that high-frequency microsatellite instability (MSI-H), due to loss of DNA mismatch repair function, is correlated with poor response to 5-FU-based treatment compared to CRC patients with stable microsatellites.<sup>16,17</sup> Controversially, negative results were also reported by the other researchers.<sup>18</sup> The latest systematic review with meta-analysis summarized fourteen 5-FU-based trials and concluded that MSI status has a limited effect on both DFS and OS and is therefore not valuable in guiding 5-FU-based treatment selection.<sup>19</sup> Dihydropyrimidine dehydrogenase ([NADP<sup>+</sup>], DYPD)—a pyrimidine catabolic enzyme that metabolizes thymine (T) and uracil (U) nucleotides—was later discovered and enables the identification of the 3% of CRC patients who cannot sufficiently metabolize 5-FU. Patients with DYPD deficiency could experience severe 5-FU-related toxicities.<sup>20</sup> Further research found that the DYPD variants DPYD\*2A (relative risk: 2.9,  $P < 0.0001$ ), c.1679 T > G (relative risk: 4.4,  $P < 0.0001$ ), c.1236G > A/HapB3 (relative risk: 1.6,  $P < 0.0001$ ), and c.2846A > T (relative risk: 3.0,  $P < 0.0001$ ) are clinically relevant as predictors of fluoropyrimidine-associated intolerance.<sup>21</sup> A prospective trial proved that DPYD\*2A-guided 5-FU dosing has significantly reduced the incidence of severe toxicity in DPYD\*2A carriers,



**Table 2** TNM staging system of colorectal cancer (AJCC 8th edition)

Stage	T (primary tumour)	N (regional lymph nodes)	M (distant metastasis)
0	Tis	N0	M0
I	T1–T2	N0	M0
IIA	T3	N0	M0
IIB	T4a	N0	M0
IIC	T4b	N0	M0
IIIA	T1–T2	N1 or N1c	M0
	T1	N2a	M0
IIIB	T3–T4a	N1 or N1c	M0
	T2–T3	N2a	M0
	T1–T2	N2b	M0
IIIC	T4a	N2a	M0
	T3–T4a	N2b	M0
	T4b	N1–N2	M0
IVA	Any T	Any N	M1a
IVB	Any T	Any N	M1b
IVC	Any T	Any N	M1c

*Primary tumour (T):* Tx, primary tumour of unknown; T0, no evidence of primary tumour; Tis, carcinoma *in situ*; T1, tumour invades submucosa; T2, tumour invades muscularis propria; T3, tumour invades through the muscularis propria into the peri colorectal tissues; T4a, tumour invades through the visceral peritoneum; T4b: tumour directly invades or adheres to other adjacent organs or structures.

*Regional lymph nodes (N):* Nx, lymph nodes cannot be assessed; N0, no lymph node metastases; N1, 1–3 lymph node involvement; N1a, 1 lymph node; N1b, 2–3 lymph nodes; N1c, non-nodal tumour deposits without identified lymph node metastases; N2, 4 or more lymph node involvement; N2a: 4–6 lymph nodes; N2b: 7 or more lymph nodes.

*Distant metastasis (M):* Mx, distant metastasis cannot be assessed; M0, no distant metastasis by imaging; M1, distant metastasis; M1a, metastasis to one organ or site without peritoneal metastasis; M1b, metastasis to two or more organs or sites without peritoneal metastasis; M1c, peritoneal involvement regardless of other organ involvement.

from 73 to 28% ( $P < 0.001$ ).<sup>22</sup> Although DPYD pretreatment screening has been proven to improve drug safety for DPYD\*2A carriers by the Food and Drug Administration (FDA) in the United States, the current European Society for Medical Oncology (ESMO) guidelines do not “routinely recommend” upfront genotyping of DPYD\*2A before the administration of 5-FU in metastatic CRC (mCRC) patients.<sup>23</sup> This recommendation is now being reviewed.<sup>24</sup>

Irinotecan is a topoisomerase 1 (TOP1) inhibitor that has a specific pharmacodiagnostic test.<sup>25</sup> Clinical studies demonstrated that the inhibition of TOP1 by irinotecan blocks the DNA ligation process during the cell cycle. However, CRC patients with uridine diphosphate glucuronosyltransferase 1A1 (UGT1A1) deficiency cannot sufficiently excrete the active metabolite SN-38, which primarily undergoes glucuronidation in their livers.<sup>26</sup> As a result, a high dose of irinotecan in UGT1A1-deficient CRC patients is associated with severe adverse drug responses such as neutropenia and diarrhea.<sup>27</sup> This has been confirmed by other studies and verified by a meta-analysis.<sup>28</sup> Therefore, the U.S. FDA has recommended a dose reduction of irinotecan for patients with homozygous UGT1A1\*28 based on A(TA-6)TAA

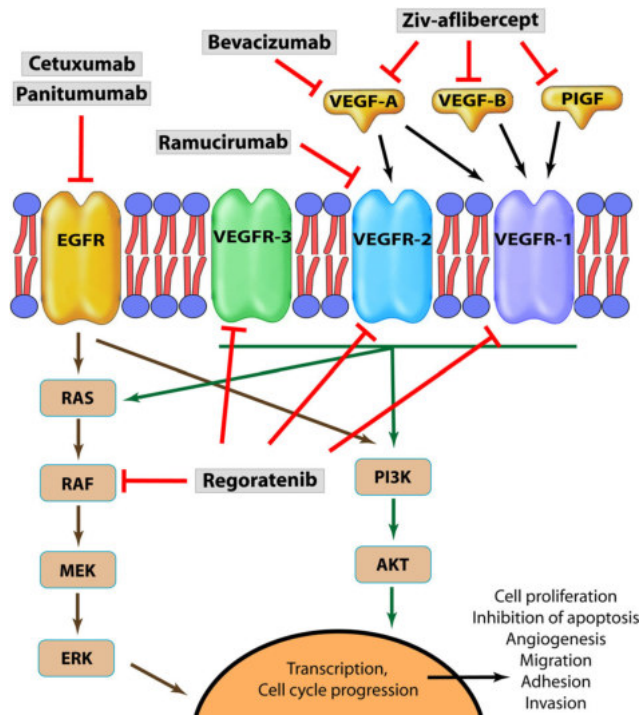
and A(TA-7)TAA genotyping.<sup>29</sup> Clinical trials focusing on the other UGT1A1 gene polymorphisms, such as UGT1A1\*1 (ClinicalTrials.gov Identifier: NCT01639326 and NCT02138617) and UGT1A1\*6 (NCT02497157), are still ongoing.

Similar to 5-FU and irinotecan, oxaliplatin is another common antineoplastic agent to which there are varying levels of chemo resistance in CRC patients.<sup>30</sup> The treatment efficacy of this platinum-based regimen can be modulated by excision repair cross-complementing group 1 (ERCC1)—one of the ERCC1-XPF enzyme complexes that play a crucial role in the nucleotide excision and repair (NER) pathway for DNA recombination and DNA repair.<sup>31</sup> In particular, ERCC1-C118T (T/T or T/C) polymorphism<sup>32</sup> or a lower expression of ERCC1<sup>33</sup> has been reported as being associated with unfavorable prognosis in patients undergoing treatment with oxaliplatin. It has therefore been proposed as a surrogate biomarker for oxaliplatin resistance. However, clinical trials have not demonstrated the predictive ability of ERCC1 in oxaliplatin-based treatment.<sup>34</sup> Thus, EMSO has not recommended ERCC1 testing prior to the use of oxaliplatin in routine practice.<sup>23</sup>

More recently, a new cytotoxic drug, lonsurf, was approved by the U.S. FDA, National Institute of Health and Care Excellence (NICE) in England, and the European Medicines Agency (EMA) for refractory mCRC patients. Lonsurf is a combination of trifluridine (thymidine-based nucleoside analogue) and tipiracil (a potent thymidine phosphorylase inhibitor) that suppresses cancer cell proliferation by interfering with DNA synthesis.<sup>35</sup> Based on the RECOURSE group’s phase III randomized trial, which included nearly 800 participants from three different geographical areas, lonsurf results in a 1.8-month improvement in median OS compared with the placebo group.<sup>36</sup> Methods for optimizing lonsurf treatment are currently under investigation, including the development of a CRC xenograft experimental model that predicts treatment outcome;<sup>37</sup> the use of 3′-Deoxy-3′-18F-fluorothymidine positron emission tomography ([<sup>18</sup>F]FLT-PET) as a noninvasive radio-traceable substitute for thymidine; and using the MSI status as an indicator for the use of lonsurf in combination with nivolumab, a PD-1 inhibitor, in refractory mCRC patients (NCT02860546).

## EGFR therapies

EGFR is a transmembrane tyrosine kinase receptor that regulates the serine/threonine-specific protein kinase (AKT), JNK, and mitogen-activated protein kinase (MAPK)/ERK signaling pathways responsible for DNA synthesis, cell proliferation, apoptosis, and motility (Fig. 1). Overexpression of EGFR is associated with tumor progression in various cancer types, including CRC.<sup>38</sup> Blocking the EGFR by using monoclonal antibodies such as cetuximab or panitumumab<sup>39,40</sup> with a chemotherapy formula combination with 5-FU, leucovorin plus oxaliplatin (FOLFOX) or a chemotherapy formula combination with 5-FU, leucovorin plus irinotecan (FOLFIRI) results in a better treatment response in mCRC patients.<sup>41,42</sup> Those treatments can be tailored using one of the FDA-approved pharmacogenetic tools that measure a patient’s EGFR expression level<sup>43</sup> or detect KRAS Proto-Oncogene (*KRAS*) exon 2 (codon 12/13) mutations<sup>44</sup> (Table 1). However, the effectiveness of these pharmacogenetic tests in detecting and improving treatment response is uncertain. For



**Figure 1** The approved EGFR and VEGF targeted drugs and its receptors in colorectal cancer. AKT, protein kinase B; EGFR, epidermal growth factor receptor; ERK, extracellular-regulated kinase; MEK, mitogen-activated protein/extracellular signal-regulated kinase; PI3K, phosphatidylinositol 3-kinase; PlGF, placental growth factor; RAF, rapidly accelerated fibrosarcoma; RAS, retrovirus-associated DNA sequences; VEGFR, vascular endothelial growth factor receptor

example, many pathologists have expressed concern about the EGFR detection criteria in the PharmDx™ immunohistochemistry (IHC) test.<sup>45,46</sup> Many clinicians also doubt the benefits of anti-EGFR treatment in EGFR-positive CRC patients.<sup>45,47</sup> The alternative option of *KRAS* exon 2 mutation screening is also problematic because testing is limited to one *KRAS* exon region, and studies have shown that CRC patients with other *KRAS* mutations will still benefit from anti-EGFR treatment.<sup>48</sup> In fact, up to 35% of *KRAS* exon 2 wild-type<sup>49</sup> and approximately 25% of EGFR-negative patients responded to EGFR inhibitor treatments.<sup>50</sup> Therefore, other *RAS* signaling biomarkers, such as *KRAS* exon 3 (codons 59/61) and 4 (codons 117/146), as well as *NRAS* proto-oncogene (*NRAS*) exon 2 (codon 12/13), 3 (codons 59/61) and 4 (codons 117/146) mutations, are being investigated for further pharmacodiagnostic development.<sup>51–53</sup>

In a retrospective analysis of the CRYSTAL study, authors assessed the status of other *RAS* mutations (*KRAS* exons 3 and 4; *NRAS* exons 2, 3 and 4). Of the 367 *RAS* wild-type CRC patients, treatment with FOLFIRI plus cetuximab was better than FOLFIRI alone in both PFS (11.4 vs 8.4 months, HR: 0.56  $P < 0.001$ ) and OS (28.4 vs 20.2 months, HR: 0.69  $P = 0.0024$ ). There was no difference in the other *RAS* mutant populations ( $n = 63$ ).<sup>54</sup> Similar results were also reported in another phase III trial for a second-line therapy based on *RAS* mutation status (*KRAS* exons 3, 4; *NRAS* exons 2, 3, 4; and

*BRAF* exon 15). The use of FOLFIRI with or without panitumumab in the wild-type *RAS* population improved survival in mCRC patients (PFS: 6.4 vs 4.6 months, HR: 0.70,  $P = 0.007$ ) compared with the *KRAS* exon 2 wild-type individuals (PFS: 5.9 vs 3.9 months, HR: 0.73,  $P = 0.004$ ).<sup>55</sup> Based on the published results of the *RAS* mutation combination analysis, the ESMO,<sup>56</sup> European Society of Pathology (ESP), and Association of Clinical Pathologists Molecular Pathology and Diagnostics Group in the United Kingdom recommended the *KRAS*/*NRAS* mutation test for mCRC patients.<sup>57</sup>

In addition to the *RAS* mutation, other potential biomarkers have been uncovered and may help in the selection of CRC patients suitable for anti-EGFR treatment. These biomarkers include *PIK3CA*, PTEN, Human Epidermal Growth Factor Receptor 2 (HER2), HER3, and the EGFR ligands EREG and AREG.<sup>53,58–61</sup> Although these biomarkers are not yet available for clinical use, the combination of multiple biomarkers may have a stronger predictive power than using one alone.<sup>58</sup> Further prospective studies are needed to substantiate predictive biomarker combinations for EGFR-targeted treatment (Table 3).

## VEGF receptor therapies

The VEGF receptor is a transmembrane protein containing a split tyrosine–kinase domain at the intracellular level and seven immunoglobulin-like domains at extracellular levels for angiogenesis and vasculogenesis.<sup>62</sup> Overexpression of VEGF results in tumor progression and metastasis as well as lower patient survival rates.<sup>63,64</sup> Today, three approved biological agents targeting VEGF are available for CRC patients. Ramucirumab targets the VEGF-A receptor activation by modulating VEGFR-2; ziv-aflibercept inhibits placental growth factor (PlGF), VEGF-A, and VEGF-B by using its IgG1 Fc-VEGFR; and bevacizumab blocks VEGF-A to cause ligand sequestering (Fig. 1).<sup>65</sup> Interestingly, the use of FOLFIRI in combination with ziv-aflibercept (VELOUR trial),<sup>66</sup> bevacizumab (ML18147 trial),<sup>67</sup> or ramucirumab (PRAISE trial)<sup>68</sup> in mCRC patients presented similar treatment benefits in median OS (1.4, 1.4, and 1.6 months) and PFS (2.2, 1.6 and 1.2 months). All three antiangiogenic regimens also present with similar types of adverse drug events (e.g. proteinuria, hemorrhage, and hypertension).<sup>69</sup> However, the differences in tolerability and the study design in those clinical trials vary.<sup>70</sup>

Although no obvious difference was found between the approved VEGF-targeted treatments, they also do not directly replace each other due to the different VEGF subtype targets (Fig. 1) and the treatment effectiveness in patient-derived xenograft mouse models.<sup>71</sup> Hence, an ongoing PERMAD phase II trial (NCT02331927) is investigating potential cytokine and/or angiogenic factor(s) as biomarker(s) for a treatment shift from bevacizumab to ziv-aflibercept to increase the treatment effectiveness and limit drug resistance. Furthermore, studies also found that the continuous administration of bevacizumab leads to better OS<sup>67,72</sup> as planned treatment breaks or discontinuation in antiangiogenic therapy could lead to rapid tumor regrowth.<sup>73,74</sup> To monitor the tumor growth and treatment response, the CIRCUS research team is prospectively evaluating circulating VEGFR-2 levels as a predictor of the continuation of bevacizumab treatment in mCRC patients (NCT02623621). Several

**Table 3** Ongoing clinical trials for molecular biomarkers in approved CRC drugs

Drug	Biomarker	ClinicalTrials.gov identifier:
Bevacizumab + chemotherapy	VEGFR-2	NCT02623621
Bevacizumab/cetuximab + FOLFIRI	<i>BRAF</i> & <i>PIK3K</i> in <i>RAS</i> wild-type mCRC	NCT01640444
Bevacizumab, cetuximab + irinotecan	<i>KRAS</i> wild-type, Irinotecan refractory	NCT02292758
Cetuximab + FOLFIRI/mFOLFOX6	ERCC1	NCT01703390
Cetuximab or panitumumab	EGFR domain III region	NCT01726309
Panitumumab + FOLFIRI	<i>RAS</i> & <i>BRAF</i> wild-type mCRC	NCT02508077
Regorafenib	[ <sup>18</sup> F] FLT-PET	NCT02175095
Regorafenib	<i>RAS</i> -mutant advanced CRC	NCT02619435
Ziv-aflibercept	Cytokines & angiogenic factors	NCT02331927

[<sup>18</sup>F] FLT-PET, 3'-deoxy-3'-<sup>18</sup>F-fluorothymidine positron emission tomography; mCRC, metastatic colorectal cancer.

potential new biomarkers have also been reported for VEGF inhibitors, including *KRAS* (codons 12 and 13),<sup>75</sup> VEGF(165)b: VEGF(total) expression ratio,<sup>76</sup> VEGF-D,<sup>77</sup> miR-126,<sup>78</sup> EGFL7,<sup>79</sup> Ang-2,<sup>80</sup> NRP-1,<sup>81</sup> IL-8,<sup>82</sup> and G12 V and G12A *KRAS* mutations.<sup>83</sup> However, prospective studies are necessary to verify the results.

In addition to the VEGF single-targeting agents, regorafenib is a dual-targeted VEGFR2-TIE2 tyrosine kinase inhibitor that suppresses a set of protein kinases involved in oncogenesis (B-Raf Proto-Oncogene [*BRAF*], *RAF1*, *RET* and *KIT*) and angiogenesis (tyrosine receptor kinase-2 [*TIE2*], VEGFR 1–3, fibroblast growth factor receptor [*FGFR*] and platelet-derived growth factor receptor [*PDGFR*]).<sup>84</sup> mCRC patients who received regorafenib treatment demonstrated a statistically significant improvement in survival rate when compared with placebo in the CORRECT (OS: 6.4 vs 5.0 months, HR = 0.77, *P* = 0.0052; PFS: 1.9 vs 1.7 months, HR = 0.49, *P* < 0.0001)<sup>85</sup> and CONCUR (OS: 8.8 vs 6.3 months, HR = 0.55, *P* = 0.0016) trials.<sup>86</sup> Several clinical studies on regorafenib are ongoing to find suitable biomarkers to stratify CRC patients.<sup>85,87</sup> This includes identifying *RAS* subtypes (NCT02619435), as well as using imaging biomarkers such as [<sup>18</sup>F] FLT-PET (NCT02175095) (Table 3). Several clinical trials investigating biomarkers for regorafenib in mCRC patients who failed one prior anticancer treatment are ongoing (NCT01949194, NCT01996969, and NCT02402036).

## The development of new molecular targeted therapy in CRC

The development of new molecular targeted therapy in CRC and investigations into their use in combination are ongoing. For instance, selumetinib, a MEK1 and MEK2 inhibitor,<sup>88</sup> in combination with afatinib, an approved EGFR inhibitor for non-small cell lung carcinoma,<sup>88</sup> is currently being tested in an early-stage randomized clinical trial for *KRAS* mutant and *PIK3CA* wild-type CRC patients (NCT02450656) (Table 4). Dual anti-EGFR and anti-VEGF treatments for CRC are also being studied. For example, the use of cetuximab plus regorafenib inhibited AKT and MAPK signaling pathways in *BRAF*-mutated, *KRAS*-mutated, and cetuximab-resistant CRC cell lines and presented a synergistic apoptotic as well as antiproliferative effect in an *in vivo* model.<sup>89</sup> This combination was proven and well tolerated in the phase I clinical trial, and the antitumor effect may greatly benefit MSI-H CRC patients.<sup>90</sup> The next phase of the trial may be conducted in the near future.

More recently, monoclonal antibodies against programmed cell death-1 (PD-1) receptor or its ligand PD-L1 have shown promising results in several types of cancers. PD-1 is an immune checkpoint protein expressed on the surface of T-cells and plays a key role in promoting self-tolerance by suppressing T-cell cytokine production. PD-L1 is frequently upregulated in tumor cells

**Table 4** Ongoing clinical trials for new CRC drugs and their respective biomarkers

Target molecule	Drug name	Biomarker	Trial phase	ClinicalTrials.gov identifier
AKT	Trametinib	<i>BRAF</i> mutant	I/II	NCT01902173
	GSK2141795	<i>BRAF</i> mutant	I/II	NCT01902173
<i>BRAF</i>	Dabrafenib	<i>BRAF</i> mutant	I/II	NCT01902173
cMET	Tivantinib	<i>KRAS</i> wild-type	II	NCT01892527
	PF-02341066	<i>RAS</i> mutant & over-active MET	I/II	NCT02510001
Glutaminase	CB-839	Fluoropyrimidine Resistant & <i>PIK3CA</i> mutant	I/II	NCT02861300
HER2	Ado-Trastuzumab Emtansine	HER2	I/II	NCT02465060
PD-1	Pembrolizumab	<i>KRAS</i> , <i>BRAF</i> & <i>NRAS</i> wild-type	II	NCT02318901
	Nivolumab	MSI status	III	NCT01876511, NCT02563002
MEK	Selumetinib	<i>KRAS</i> mutant & <i>PIK3CA</i> wild-type	II	NCT02860546, NCT03104439
	PD-0325901	<i>RAS</i> mutant & over-active MET	I/II	NCT02510001
Tyrosine Kinase	Entrectinib	<i>NTRK1/2/3</i> , <i>ROS1</i> , & <i>ALK</i> gene fusion	II	NCT02568267
PI3K	BKM120	<i>RAS</i> wild-type	I/II	NCT01304602, NCT01591421

*BRAF*, B-Raf proto-oncogene; *HER2*, human epidermal growth factor receptor 2; *NRAS*, *NRAS* proto-oncogene.

and deactivates antitumor activity in cytotoxic T-cells.<sup>91,92</sup> Research has shown that CRC with MSI highly expresses immune checkpoint molecules, including PD-L1.<sup>93</sup> Thus, in a phase II clinical trial, pembrolizumab, a U.S. FDA-approved PD-1 targeted therapy, was utilized to treat both MSI-H and microsatellite-stable (MSS) CRC patients. The response rate and the 12-week PFS to pembrolizumab in MSI-H mCRC patients (n = 10) were 40 and 78% compared to 0 and 11% in MSS mCRC patients (n = 18), respectively.<sup>94</sup> Combination treatment with pembrolizumab and itacitinib, a JAK1 inhibitor, is also under investigation for use in any patient with MSI instability (NCT02646748). In addition to the clinical trials stratifying treatment based on MSI status (NCT01876511 and NCT02563002), treatment for different molecular subtypes such as pembrolizumab plus trastuzumab treatment for mCRC patients with *KRAS*, *BRAF*, and *NRAS* wild-types (NCT02318901) are under investigation (Table 3).

## Future directions and conclusions

Patients with the “same” cancer often respond differently to treatment—this challenge has baffled medical oncologists for decades. Pharmacodiagnostic testing is now becoming an essential tool for selecting the right medication for the right patient. Since the U.S. FDA approved next-generation sequencing (NGS) devices for clinical diagnosis in November 2013,<sup>95</sup> the use of NGS has become a popular tool for the investigation of diseases. For example, NGS was used by Hagemann *et al.* in patients with non-small cell lung cancer to match 11% of their patients with a targeted therapy.<sup>96</sup> NGS can also be applied to noninvasively detect circulating tumor DNA (ctDNA) in CRC patients for real-time monitoring of the disease, facilitating early identification of disease progression.<sup>97</sup> For instance, *KRAS* mutant alleles can be detected in blood plasma from the acquired tumor-resistant patients 10 months before cancer progression is otherwise detected.<sup>98</sup> This is because genetic aberrations coding treatment resistance accumulate during tumor progression and are released from tumor cells into the blood circulation.<sup>99</sup> Another effective method to improve treatment selection was demonstrated by Pauli *et al.*,<sup>100</sup> where tumor tissue collected from a patient was subjected to four separate experiments: (i) NGS for molecular subtype analysis, (ii) primary cell culture, (iii) patient-derived xenograft (PDX) models, and (iv) patient-derived tumor organoids. This cutting-edge screening strategy facilitated precision treatment, but the process itself is costly and therefore may not currently be affordable to the wider public.

In conclusion, the aim of precision medicine is to develop a tailored treatment for each individual and his or her unique condition to maximize potential treatment response and minimize adverse drug reactions. The stratification of patients through the use of biomarkers is thus key. As the use of newer therapeutic agents connected with specific genetic sup-type(s) will increase, ultimately increasing patients' quality of life and life expectancy.

## REFERENCES

- 1 Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J *et al.* Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur. J. Cancer.* 2013; **49**: 1374–403.
- 2 Wilkinson NW, Yothers G, Lopa S, Costantino JP, Petrelli NJ, Wolmark N. Long-term survival results of surgery alone versus surgery plus 5-fluorouracil and leucovorin for stage II and stage III colon cancer: pooled analysis of NSABP C-01 through C-05. A baseline from which to compare modern adjuvant trials. *Ann. Surg. Oncol.* 2010; **17**: 959–66.
- 3 Verma M. Personalized medicine and cancer. *J. Pers. Med.* 2012; **2**: 1–14.
- 4 Amin MB, Greene FL, Edge SB *et al.* The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA Cancer J. Clin.* 2017; **67**: 93–9.
- 5 McQuade RM, Stojanovska V, Bornstein JC, Nurgali K. Colorectal cancer chemotherapy: the evolution of treatment and new approaches. *Curr. Med. Chem.* 2017; **24**: 1537–57.
- 6 Sohn DK, Chang HJ, Park JW *et al.* Histopathological risk factors for lymph node metastasis in submucosal invasive colorectal carcinoma of pedunculated or semipedunculated type. *J. Clin. Pathol.* 2007; **60**: 912–5.
- 7 Glynne-Jones R, Wyrwicz L, Tiret E *et al.* Rectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 2017; **28**(Suppl. 4): iv22–40.
- 8 Mori K, Toiyama Y, Otake K *et al.* Successful identification of a predictive biomarker for lymph node metastasis in colorectal cancer using a proteomic approach. *Oncotarget.* 2017; **8**: 106935–47.
- 9 Molloy MP, Engel A. Precision medicine beyond medical oncology: using molecular analysis to guide treatments of colorectal neoplasia. *Expert Rev. Gastroenterol. Hepatol.* 2018; **12**: 1179–81.
- 10 Kannarkatt J, Joseph J, Kurniali PC, Al-Janadi A, Hrinchenko B. Adjuvant chemotherapy for stage II colon cancer: a clinical dilemma. *J. Oncol. Pract.* 2017; **13**: 233–41.
- 11 Kopetz S, Taberero J, Rosenberg R *et al.* Genomic classifier ColoPrint predicts recurrence in stage II colorectal cancer patients more accurately than clinical factors. *Oncologist.* 2015; **20**: 127–33.
- 12 You YN, Rustin RB, Sullivan JD. Oncotype DX<sup>®</sup> colon cancer assay for prediction of recurrence risk in patients with stage II and III colon cancer: a review of the evidence. *Surg. Oncol.* 2015; **24**: 61–6.
- 13 Bailey H, Turner M, Stoppler MC, Chao C. The 12-gene oncotype DX colon recurrence score (RS) test: experience with >20,000 stage 2 patients (pts). *J. Clin. Oncol.* 2018; **36**(4 Suppl.): 618–18.
- 14 Dalerba P, Sahoo D, Paik S *et al.* CDX2 as a prognostic biomarker in stage II and stage III colon cancer. *N. Engl. J. Med.* 2016; **374**: 211–22.
- 15 Longley DB, Harkin DP, Johnston PG. 5-Fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer.* 2003; **3**: 330–8.
- 16 Bertagnolli MM, Niedzwiecki D, Compton CC *et al.* Microsatellite instability predicts improved response to adjuvant therapy with irinotecan, fluorouracil, and leucovorin in stage III colon cancer: Cancer and Leukemia Group B Protocol 89803. *J. Clin. Oncol.* 2009; **27**: 1814–21.
- 17 Des Guetz G, Schischmanoff O, Nicolas P, Perret G-Y, Morere J-F, Uzzan B. Does microsatellite instability predict the efficacy of adjuvant chemotherapy in colorectal cancer? A systematic review with meta-analysis. *Eur. J. Cancer.* 2009; **45**: 1890–6.
- 18 Sargent DJ, Shi Q, Yothers G *et al.* Prognostic impact of deficient mismatch repair (dMMR) in 7,803 stage II/III colon cancer (CC) patients (pts): a pooled individual pt data analysis of 17 adjuvant trials in the ACCENT database. *J. Clin. Oncol.* 2014; **32** (15 suppl): 3507.
- 19 Webber EM, Kauffman TL, O'Connor E, Goddard KA. Systematic review of the predictive effect of MSI status in colorectal cancer patients undergoing 5FU-based chemotherapy. *BMC Cancer.* 2015; **15**: 156.
- 20 Sanoff HK, McLeod HL. Predictive factors for response and toxicity in chemotherapy: pharmacogenomics. *Semin. Colon Rectal. Surg.* 2008; **19**: 226–30.

- 21 Meulendijks D, Henricks LM, Sonke GS *et al.* Clinical relevance of DPYD variants c.1679T>G, c.1236G>A/HapB3, and c.1601G>A as predictors of severe fluoropyrimidine-associated toxicity: a systematic review and meta-analysis of individual patient data. *Lancet Oncol.* 2015; **16**: 1639–50.
- 22 Deenen MJ, Meulendijks D, Cats A *et al.* Upfront genotyping of DPYD\*2A to individualize fluoropyrimidine therapy: a safety and cost analysis. *J. Clin. Oncol.* 2016; **34**: 227–34.
- 23 Van Cutsem E, Cervantes A, Adam R *et al.* ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann. Oncol.* 2016; **27**: 1386–422.
- 24 Deenen MJ, Meulendijks D. Recommendation on testing for dihydropyrimidine dehydrogenase deficiency in the ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann. Oncol.* 2017; **28**: 184.
- 25 Ratain MJ. From bedside to bench to bedside to clinical practice: an odyssey with irinotecan. *Clin. Cancer Res.* 2006; **12**: 1658–60.
- 26 Iyer L, King CD, Whittington PF *et al.* Genetic predisposition to the metabolism of irinotecan (CPT-11). Role of uridine diphosphate glucuronosyltransferase isof orm 1A1 in the glucuronidation of its active metabolite (SN-38) in human liver microsomes. *J. Clin. Invest.* 1998; **101**: 847–54.
- 27 Innocenti F. Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *J. Clin. Oncol.* 2004; **22**: 1382–8.
- 28 Campbell JM, Stephenson MD, Bateman E, Peters MDJ, Keefe DM, Bowen JM. Irinotecan-induced toxicity pharmacogenetics: an umbrella review of systematic reviews and meta-analyses. *Pharmacogenomics J.* 2017; **17**: 21–8.
- 29 CAMPTOSAR. *CAMPTOSAR-irinotecan hydrochloride injection, solution*, 2016.
- 30 Meyerhardt JA, Mayer RJ. Systemic therapy for colorectal cancer. *New Engl. J. Med.* 2005; **352**: 476–87.
- 31 Martin LP, Hamilton TC, Schilder RJ. Platinum resistance: the role of DNA repair pathways. *Clin. Cancer Res.* 2008; **14**: 1291–5.
- 32 Shahnam A, Ridha Z, Wiese MD, Kichenadasse G, Sorich MJ. Pharmacogenetic and ethnicity influence on oxaliplatin therapy for colorectal cancer: a meta-analysis. *Pharmacogenomics.* 2016; **17**: 1725–32.
- 33 Kassem AB, Salem SE, Abdelrahim ME *et al.* ERCC1 and ERCC2 as predictive biomarkers to oxaliplatin-based chemotherapy in colorectal cancer patients from Egypt. *Exp. Mol. Pathol.* 2017; **102**: 78–85.
- 34 Lenz H-J, Lee F-C, Yau L *et al.* MAVERICC, a phase 2 study of mFOLFOX6-bevacizumab (BV) vs FOLFIRI-BV with biomarker stratification as first-line (1L) chemotherapy (CT) in patients (pts) with metastatic colorectal cancer (mCRC). *J. Clin. Oncol.* 2016; **34** (4 Suppl.): 493.
- 35 Burness CB, Duggan ST. Trifluridine/tipiracil: a review in metastatic colorectal cancer. *Drugs.* 2016; **76**: 1393–402.
- 36 Mayer RJ, Van Cutsem E, Falcone A *et al.* Randomized trial of TAS-102 for refractory metastatic colorectal cancer. *N. Engl. J. Med.* 2015; **372**: 1909–19.
- 37 Kim S-Y, Jung JH, Lee HJ *et al.* [18F]fluorothymidine PET informs the synergistic efficacy of capecitabine and trifluridine/tipiracil in colon cancer. *Cancer Res.* 2017; **77**: 7120–30.
- 38 Oda K, Matsuoka Y, Funahashi A, Kitano H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.* 2005; **1**: E1–17.
- 39 Tveit KM, Guren T, Glimelius B *et al.* Phase III trial of cetuximab with continuous or intermittent fluorouracil, leucovorin, and oxaliplatin (Nordic FLOX) versus FLOX alone in first-line treatment of metastatic colorectal cancer: the NORDIC-VII study. *J. Clin. Oncol.* 2012; **30**: 1755–62.
- 40 Peeters M, Price TJ, Cervantes A *et al.* Randomized phase III study of panitumumab with fluorouracil, leucovorin, and irinotecan (FOLFIRI) compared with FOLFIRI alone as second-line treatment in patients with metastatic colorectal cancer. *J. Clin. Oncol.* 2010; **28**: 4706–13.
- 41 Shitara K, Yokota T, Takahari D *et al.* Cetuximab plus FOLFOX for patients with metastatic colorectal cancer with poor performance status and/or severe tumor-related complications. *Case Rep. Oncol.* 2010; **3**: 282–6.
- 42 Peeters M, Price TJ, Cervantes A *et al.* Final results from a randomized phase 3 study of FOLFIRI panitumumab for second-line treatment of metastatic colorectal cancer. *Ann. Oncol.* 2014; **25**: 107–16.
- 43 Abd El All HS, Mishriky AM, Mohamed FA. Epidermal growth factor receptor in colorectal carcinoma: correlation with clinicopathological prognostic factors. *Color Dis.* 2008; **10**: 170–78. 070817144921004.
- 44 U.S. Food and Drug Administration. *List of Cleared or Approved Companion Diagnostic Devices (In Vitro and Imaging Tools)* 2015.
- 45 Buckley AF, Kakar S. Comparison of the Dako EGFR pharmDx Kit and Zymed EGFR antibody for assessment of EGFR status in colorectal adenocarcinoma. *Appl. Immunohistochem. Mol. Morphol.* 2007; **15**: 305–9.
- 46 Shiogama K, Wongsiri T, Mizutani Y, Inada K, Tsutsumi Y. High-sensitivity epidermal growth factor receptor immunostaining for colorectal carcinomas, compared with EGFR PharmDx™: a study of diagnostic accuracy. *Int. J. Clin. Exp. Pathol.* 2013; **6**: 24–30.
- 47 Hecht JR, Mitchell E, Neubauer MA *et al.* Lack of correlation between epidermal growth factor receptor status and response to panitumumab monotherapy in metastatic colorectal cancer. *Clin. Cancer Res.* 2010; **16**: 2205–13.
- 48 Tejpar S, Celik I, Schlichting M, Sartorius U, Bokemeyer C, Van Cutsem E. Association of KRAS G13D tumor mutations with outcome in patients with metastatic colorectal cancer treated with first-line chemotherapy with or without cetuximab. *J. Clin. Oncol.* 2012; **30**: 3570–7.
- 49 Allegra CJ, Jessup JM, Somerfield MR *et al.* American Society of Clinical Oncology Provisional Clinical Opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J. Clin. Oncol.* 2009; **27**: 2091–6.
- 50 Chung KY. Cetuximab shows activity in colorectal cancer patients with tumors that do not express the epidermal growth factor receptor by immunohistochemistry. *J. Clin. Oncol.* 2005; **23**: 1803–10.
- 51 Ciardiello F, Lenz H-J, Kohne C-H *et al.* Effect of KRAS and NRAS mutational status on first-line treatment with FOLFIRI plus cetuximab in patients with metastatic colorectal cancer (mCRC): new results from the CRYSTAL trial. *J. Clin. Oncol.* 2014; **32** (3 Suppl.): LBA443.
- 52 Peeters M, Oliner K, Price T *et al.* Updated analysis of KRAS/NRAS and BRAF mutations in study 20050181 of panitumumab (pmab) plus FOLFIRI for second-line treatment (tx) of metastatic colorectal cancer (mCRC). *Ann. Oncol.* 2014; **25**(Suppl. 2): ii5–5.
- 53 Pentheroudakis G, Kotoula V, De Roock W *et al.* Biomarkers of benefit from cetuximab-based therapy in metastatic colorectal cancer: interaction of EGFR ligand expression with RAS/RAF, PIK3CA genotypes. *BMC Cancer.* 2013; **13**: 49.
- 54 Van Cutsem E, Lenz H-J, Kohne C-H *et al.* Fluorouracil, leucovorin, and irinotecan plus cetuximab treatment and RAS mutations in colorectal cancer. *J. Clin. Oncol.* 2015; **33**: 692–700.
- 55 Peeters M, Oliner KS, Price TJ *et al.* Analysis of KRAS/NRAS mutations in a phase III study of panitumumab with FOLFIRI compared with FOLFIRI alone as second-line treatment for metastatic colorectal cancer. *Clin. Cancer Res.* 2015; **21**: 5469–79.
- 56 Van Cutsem E, Cervantes A, Nordlinger B, Arnold D. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 2014; **25**(Suppl. 3): iii1–9.

- 57 Wong NA, Gonzalez D, Salto-Tellez M *et al.* RAS testing of colorectal carcinoma—a guidance document from the Association of Clinical Pathologists Molecular Pathology and Diagnostics Group. *J. Clin. Pathol.* 2014; **67**: 751–7.
- 58 Yang Z-Y, Wu X-Y, Huang Y-F *et al.* Promising biomarkers for predicting the outcomes of patients with KRAS wild-type metastatic colorectal cancer treated with anti-epidermal growth factor receptor monoclonal antibodies: a systematic review with meta-analysis. *Int. J. Cancer.* 2013; **133**: 1914–25.
- 59 De Roock W, Claes B, Bernasconi D *et al.* Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet Oncol.* 2010; **11**: 753–62.
- 60 Baker JB, Dutta D, Watson D *et al.* Tumour gene expression predicts response to cetuximab in patients with KRAS wild-type metastatic colorectal cancer. *Br. J. Cancer.* 2011; **104**: 488–95.
- 61 Martin V, Landi L, Molinari F *et al.* HER2 gene copy number status may influence clinical efficacy to anti-EGFR monoclonal antibodies in metastatic colorectal cancer patients. *Br. J. Cancer.* 2013; **108**: 668–75.
- 62 Palmer BF, Clegg DJ. Oxygen sensing and metabolic homeostasis. *Mol. Cell Endocrinol.* 2014; **397**: 51–8.
- 63 Hashim A, Al-Janabi A, Mahdi L, Al-Toriahi K, Yasseen A. Vascular endothelial growth factor ( VEGF ) receptor expression correlates with histologic grade and stage of colorectal cancer. *Libyan J. Med.* 2010; **5**: 5059.
- 64 Martins SF, Garcia EA, Luz MAM, Pardal F, Rodrigues M, Filho AL. Clinicopathological correlation and prognostic significance of VEGF-A, VEGF-C, VEGFR-2 and VEGFR-3 expression in colorectal cancer. *Cancer Genomics Proteomics.* 2013; **10**: 55–67.
- 65 Ferrara N, Adamis AP. Ten years of anti-vascular endothelial growth factor therapy. *Nat. Rev. Drug Discov.* 2016; **15**: 385–403.
- 66 Van Cutsem E, Tabernero J, Lakomy R *et al.* Addition of aflibercept to fluorouracil, leucovorin, and irinotecan improves survival in a phase III randomized trial in patients with metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. *J. Clin. Oncol.* 2012; **30**: 3499–506.
- 67 Bennouna J, Sastre J, Arnold D *et al.* Continuation of bevacizumab after first progression in metastatic colorectal cancer (ML18147): a randomised phase 3 trial. *Lancet Oncol.* 2013; **14**: 29–37.
- 68 Tabernero J, Yoshino T, Cohn AL *et al.* RAISE Study Investigators. Ramucirumab versus placebo in combination with second-line FOLFIRI in patients with metastatic colorectal carcinoma that progressed during or after first-line therapy with bevacizumab, oxaliplatin, and a fluoropyrimidine (RAISE): a randomised, double-blind. *Lancet Oncol.* 2015; **16**: 499–508.
- 69 Goel G, Sun W. Ramucirumab, another anti-angiogenic agent for metastatic colorectal cancer in second-line setting—its impact on clinical practice. *J. Hematol. Oncol.* 2015; **8**: 92.
- 70 Diaz-Serrano A, Riesco-Martinez MC, Garcia-Carbonero R. The safety and efficacy of ramucirumab for the treatment of metastatic colorectal cancer. *Expert Rev. Anticancer Ther.* 2016; **16**: 585–95.
- 71 Chiron M, Bagley RG, Pollard J *et al.* Differential antitumor activity of aflibercept and bevacizumab in patient-derived xenograft models of colorectal cancer. *Mol. Cancer Ther.* 2014; **13**: 1636–44.
- 72 Grothey A, Sugrue MM, Purdie DM *et al.* Bevacizumab beyond first progression is associated with prolonged overall survival in metastatic colorectal cancer: results from a large observational Cohort study (BRiTE). *J. Clin. Oncol.* 2008; **26**: 5326–34.
- 73 Griffioen AW, Mans LA, de Graaf AMA *et al.* Rapid angiogenesis onset after discontinuation of sunitinib treatment of renal cell carcinoma patients. *Clin. Cancer Res.* 2012; **18**: 3961–71.
- 74 Desar I, Mulder S, Stillebroer A *et al.* The reverse side of the victory: flare up of symptoms after discontinuation of sunitinib or sorafenib in renal cell cancer patients. A report of three cases. *Acta Oncol. (Madr).* 2009; **48**: 927–31.
- 75 Kubicka S, Greil R, Andre T *et al.* Bevacizumab plus chemotherapy continued beyond first progression in patients with metastatic colorectal cancer previously treated with bevacizumab plus chemotherapy: ML18147 study KRAS subgroup findings. *Ann. Oncol.* 2013; **24**: 2342–9.
- 76 Bates DO, Catalano PJ, Symonds KE *et al.* Association between VEGF splice isoforms and progression-free survival in metastatic colorectal cancer patients treated with bevacizumab. *Clin. Cancer Res.* 2012; **18**: 6384–91.
- 77 Weickhardt AJ, Williams DS, Lee CK *et al.* Vascular endothelial growth factor D expression is a potential biomarker of bevacizumab benefit in colorectal cancer. *Br. J. Cancer.* 2015; **113**: 37–45.
- 78 Hansen TF, Carlsen AL, Heegaard NHH, Sørensen FB, Jakobsen A. Changes in circulating microRNA-126 during treatment with chemotherapy and bevacizumab predicts treatment response in patients with metastatic colorectal cancer. *Br. J. Cancer.* 2015; **112**: 624–9.
- 79 Hansen TF, Nielsen BS, Sorensen FB, Johnsson A, Jakobsen A. Epidermal growth factor-like domain 7 predicts response to first-line chemotherapy and bevacizumab in patients with metastatic colorectal cancer. *Mol. Cancer Ther.* 2014; **13**: 2238–45.
- 80 Goede V, Coutelle O, Neuneier J *et al.* Identification of serum angiopoietin-2 as a biomarker for clinical outcome of colorectal cancer patients treated with bevacizumab-containing therapy. *Br. J. Cancer.* 2010; **103**: 1407–14.
- 81 Benson A, Krivoshik AK, Van Sant C, Gyuris J, Feng B. Abstract A24: Neupilin 1 (NRP1) as a potential biomarker for tivozanib + mFOLFOX6 versus bevacizumab + mFOLFOX6 in metastatic colorectal cancer (mCRC): post-hoc biomarker analysis of BATON-CRC phase 2 trial. *Mol. Cancer Ther.* 2015; **14**(12 Suppl. 1): A24–4.
- 82 Lambrechts D, Thienpont B, Thuillier V *et al.* Evaluation of efficacy and safety markers in a phase II study of metastatic colorectal cancer treated with aflibercept in the first-line setting. *Br. J. Cancer.* 2015; **113**: 1027–34.
- 83 Fiala O, Buchler T, Mohelnikova-Duchonova B *et al.* G12V and G12A KRAS mutations are associated with poor outcome in patients with metastatic colorectal cancer treated with bevacizumab. *Tumor Biol.* 2016; **37**: 6823–30.
- 84 Wilhelm SM, Dumas J, Adnane L *et al.* Regorafenib (BAY 73-4506): a new oral multikinase inhibitor of angiogenic, stromal and oncogenic receptor tyrosine kinases with potent preclinical antitumor activity. *Int. J. Cancer.* 2011; **129**: 245–55.
- 85 Grothey A, Van Cutsem E, Sobrero A *et al.* Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet.* 2013; **381**: 303–12.
- 86 Li J, Qin S, Xu R *et al.* Regorafenib plus best supportive care versus placebo plus best supportive care in Asian patients with previously treated metastatic colorectal cancer (CONCUR): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* 2015; **16**: 619–29.
- 87 Goldstein DA, Ahmad BB, Chen Q *et al.* Cost-effectiveness analysis of regorafenib for metastatic colorectal cancer. *J. Clin. Oncol.* 2015; **33**: 3727–32.
- 88 Yeh TC, Marsh V, Bernat BA *et al.* Biological characterization of ARRY-142886 (AZD6244), a potent, highly selective mitogen-activated protein kinase kinase 1/2 inhibitor. *Clin. Cancer Res.* 2007; **13**: 1576–83.
- 89 Napolitano S, Martini G, Rinaldi B *et al.* Primary and acquired resistance of colorectal cancer to anti-EGFR monoclonal antibody can be overcome by combined treatment of regorafenib with cetuximab. *Clin. Cancer Res.* 2015; **21**: 2975–83.
- 90 Subbiah V, Khawaja MR, Hong DS *et al.* First-in-human trial of multikinase VEGF inhibitor regorafenib and anti-EGFR antibody

- cetuximab in advanced cancer patients. *JCI Insight*. 2017; **2**: e90380.
- 91 Liu J, Yuan Y, Chen W *et al*. Immune-checkpoint proteins VISTA and PD-1 nonredundantly regulate murine T-cell responses. *Proc. Natl. Acad. Sci*. 2015; **112**: 6682–7.
- 92 Francisco LM, Sage PT, Sharpe AH. The PD-1 pathway in tolerance and autoimmunity. *Immunol. Rev*. 2010; **236**: 219–42.
- 93 Kim JH, Park HE, Cho N-Y, Lee HS, Kang GH. Characterisation of PD-L1-positive subsets of microsatellite-unstable colorectal cancers. *Br. J. Cancer*. 2016; **115**: 490–6.
- 94 Le DT, Uram JN, Wang H *et al*. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med*. 2015; **372**: 2509–20.
- 95 Sheridan C. Milestone approval lifts Illumina's NGS from research into clinic. *Nat. Biotechnol*. 2014; **32**: 111–2.
- 96 Hagemann IS, Devarakonda S, Lockwood CM *et al*. Clinical next-generation sequencing in patients with non-small cell lung cancer. *Cancer*. 2015; **121**: 631–9.
- 97 Diaz LA, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. *J. Clin. Oncol*. 2014; **32**: 579–86.
- 98 Misale S, Yaeger R, Hobor S *et al*. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature*. 2012; **486**: 532–6.
- 99 Murtaza M, Dawson S-J, Tsui DWY *et al*. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*. 2013; **497**: 108–12.
- 100 Pauli C, Hopkins BD, Prandi D *et al*. Personalized in vitro and in vivo cancer models to guide precision medicine. *Cancer Discov*. 2017; **7**: 462–77.

# The Role of microRNAs in Development of Colitis-Associated Colorectal Cancer

Marco Bocchetti, Maria Grazia Ferraro, Filippo Ricciardiello, Alessandro Ottaiano, Amalia Luce, Alessia Maria Cossu, Marianna Scrima, Wing-Yan Leung, Marianna Abate, Paola Stiuso, Michele Caraglia, Silvia Zappavigna<sup>#</sup>, **Tung On Yau<sup>#</sup>**

*International Journal of Molecular Sciences*. **2021**; 22(8): 3967.

Journal URL: [mdpi.com/1422-0067/22/8/3967](https://mdpi.com/1422-0067/22/8/3967)

DOI: [10.3390/ijms22083967](https://doi.org/10.3390/ijms22083967)

PMID: [33921348](https://pubmed.ncbi.nlm.nih.gov/33921348/)

PMCID: [PMC8068787](https://pubmed.ncbi.nlm.nih.gov/pmc/articles/PMC8068787/)



1<sup>st</sup> July, 2021

To Whom It May Concern,

**Statement of Authorship**

**Title of publication:** The Role of microRNAs in Development of Colitis-Associated Colorectal Cancer

**Journal:** International Journal of Molecular Sciences

**Publication date:** 12<sup>th</sup> April, 2021

**Issue:** Volume 22, Issue 8, 3967

**PubMed ID:** 33921348

**Authors:** Marco Bocchetti, Maria Grazia Ferraro, Filippo Ricciardiello, Alessandro Ottiano, Amalia Luce, Alessia Maria Cossu, Marianna Scrima, Wing-Yan Leung, Marianna Abate, Paola Stiuso, Michele Caraglia, Silvia Zappavigna\* and Tung On Yau\*

\* co-corresponding authors

I hereby confirm that Mr. Tung On YAU is one of the corresponding authors in the above publication. He was the key contributor to the concept of the topic, design searching strategy and responsible for screening literatures. He was also contributed to the schematic diagrams visualisation, writing of the manuscript, and revised the article.

Yours sincerely,



Marco Bocchetti, MSc  
PhD Candidate  
Department of Precision Medicine,  
University of Campania  
“L. Vanvitelli” via L. De Crecchio 7  
80138 Naples, Italy  
Email: marco.bocchetti@unicampania.it



Silvia Zappavigna, PhD  
Professor  
Department of Precision Medicine,  
University of Campania  
“L. Vanvitelli” via L. De Crecchio 7  
80138 Naples, Italy  
Email: silvia.zappavigna@unicampania.it



Review

# The Role of microRNAs in Development of Colitis-Associated Colorectal Cancer

Marco Bocchetti <sup>1,2</sup> , Maria Grazia Ferraro <sup>3</sup> , Filippo Ricciardiello <sup>4</sup>, Alessandro Ottaiano <sup>5</sup> , Amalia Luce <sup>1,6</sup> , Alessia Maria Cossu <sup>1,2</sup>, Marianna Scrima <sup>2</sup>, Wing-Yan Leung <sup>7</sup>, Marianna Abate <sup>1</sup>, Paola Stiuso <sup>1</sup> , Michele Caraglia <sup>1,2</sup> , Silvia Zappavigna <sup>1,\*</sup> and Tung On Yau <sup>8,\*</sup>

- <sup>1</sup> Department of Precision Medicine, University of Campania “Luigi Vanvitelli”, 80131 Naples, Italy; marco.bocchetti@unicampania.it (M.B.); amalia.luce@unicampania.it (A.L.); alessiamaria.cossu@biogem.it (A.M.C.); marianna.abate@unicampania.it (M.A.); paola.stiuso@unicampania.it (P.S.); michele.caraglia@unicampania.it (M.C.)
- <sup>2</sup> Biogem Scarl, Molecular Oncology and Precision Medicine Laboratory, via Camporeale, 83031 Ariano Irpino, Italy; marianna.scrima@biogem.it
- <sup>3</sup> Department of Pharmacy, School of Medicine and Surgery, University of Naples “Federico II”, via D. Montesano 49, 80131 Naples, Italy; mariagrazia.ferraro@unina.it
- <sup>4</sup> Ear, Nose, and Throat Unit, AORN “Antonio Cardarelli”, 80131 Naples, Italy; filipporicciardiello@virgilio.it
- <sup>5</sup> SSD-Innovative Therapies for Abdominal Metastases, Istituto Nazionale Tumori di Napoli, IRCCS “G. Pascale”, via M. Semmola, 80131 Naples, Italy; a.ottaiano@istitutotumori.na.it
- <sup>6</sup> School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK
- <sup>7</sup> Division of Haematology, Department of Medicine, The University of Hong Kong, Hong Kong, China; thomas83@hku.hk
- <sup>8</sup> John van Geest Cancer Research Centre, School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK
- \* Correspondence: silvia.zappavigna@unicampania.it (S.Z.); payton.yau@ntu.ac.uk (T.O.Y.)



**Citation:** Bocchetti, M.; Ferraro, M.G.; Ricciardiello, F.; Ottaiano, A.; Luce, A.; Cossu, A.M.; Scrima, M.; Leung, W.-Y.; Abate, M.; Stiuso, P.; et al. The Role of microRNAs in Development of Colitis-Associated Colorectal Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 3967. <https://doi.org/10.3390/ijms22083967>

Academic Editor: Daniela Taverna

Received: 10 March 2021

Accepted: 8 April 2021

Published: 12 April 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Colorectal cancer (CRC) is the third most deadly cancer worldwide, and inflammatory bowel disease (IBD) is one of the critical factors in CRC carcinogenesis. IBD is responsible for an unphysiological and sustained chronic inflammation environment favoring the transformation. MicroRNAs (miRNAs) belong to a class of highly conserved short single-stranded segments (18–25 nucleotides) non-coding RNA and have been extensively discussed in both CRC and IBD. However, the role of miRNAs in the development of colitis-associated CRC (CAC) is less clear. The aim of this review is to summarize the major upregulated (miR-18a, miR-19a, miR-21, miR-31, miR-155 and miR-214) and downregulated (miR-124, miR-193a-3p and miR-139-5p) miRNAs in CAC, and their roles in genes’ expression modulation in chronic colonic-inflammation-induced carcinogenesis, including programmed cell-death pathways. These miRNAs dysregulation could be applied for early CAC diagnosis, to predict therapy efficacy and for precision treatment.

**Keywords:** colorectal cancer; colitis-associated colorectal cancer; inflammatory bowel disease; microRNA; biomarkers

## 1. Introduction

Inflammatory bowel disease (IBD) is a group of idiopathic and relapsing-remitting chronic inflammatory disorders comprising the two major subtypes: Crohn’s disease (CD) and ulcerative colitis (UC). IBDs are characterized by a susceptible genetic background, underlying immunological deregulation and intestinal microbiome dysbiosis leading to intestinal mucosa damage [1]. It is well recognized that the long-standing chronic inflammation in intestinal mucosa induces intestinal barrier injury: resulting in increased permeability and destruction of the tight junctions [2], and eventually colorectal cancer (CRC) onset [3,4]. The degree of colonic inflammation together with the disorder duration is correlated with the development of colonic neoplasia [5,6]. The influence of UC on CRC risk is approximately 2%, 8% and 18% after one, two and three decades of the disease,

respectively [7]. The cumulative risk for CRC in CD is approximately 3% after 10 years, 6% after 20 years and 8% after 30 years of the disease duration [8]. Several studies have reported that UC-associated CRC has an unfavorable survival compared to sporadic CRC and is responsible for one-sixth of UC-related deaths [9–11].

Colitis-associated carcinogenesis is a multi-stage process starting with chronic inflammation and affected by environmental, genetic and immunologic factors, as well [12,13], eventually presenting differences when compared with sporadic CRC. In addition, drug trials in prodromal phases of IBD and/or colitis-associated CRC (CAC) appeared not completely adequate. Animal-based colitis models could improve the science capability to address exact research questions, potentially giving better disease prevention, control, and supervision. The most common CAC model is the use of dextran sulfate sodium (DSS) plus azoxymethane (AOM), other viable options are summarized in Table 1. It is important to take into account that mouse strain may play a significant role in the severity and variability of the disease model and response to the same treatment.

**Table 1.** The most commonly used animal models in colitis and colitis-associated colorectal tumorigenesis.

Type	Method	Prevalent of Response	Limitations
Colitis Models			
Chemically Induced	DSS	Epithelial damage	Does not require T and/or B cell responses [14], high severity variability [15]
	TNBS/DNBS	Epithelial damage, Immune-driven	Aetiopathogenesis not clear [16]
	Oxazolone	Epithelial damage, Immune-driven	International administration required [15]
Spontaneous Mutation	SAMP1/Yit	Immune-driven	Affect small intestine only [17], low breeding rate [15]
	C3H/HeJBir	Immune-driven	Greatly influenced by caging conditions [15]
Adoptive T Cell Transfer	CD4 <sup>+</sup> CD45RB <sup>hi</sup>	Immune-driven	Lack of a full overview of colitis development [18], expensive
Genetically Engineered	IL-10 <sup>-/-</sup>	Immune-driven	Lack of focal granulomatous inflammation and Transmural inflammation [17]
Colitis-Associated Colorectal Tumorigenesis Models			
Chemically Induced	DSS	Epithelial damage	Low cancer incidents [19].
	AOM/DSS	Epithelial damage	The most common CAC model [20]
Genetically Engineered	IL-10 <sup>-/-</sup>	Immune-driven	~60% of cancer Incidence [21]

AOM, Azoxymethane; DSS, dextran sulfate sodium; TNBS, 2,4,6-trinitrobenzene sulfonic acid; DNBS, dinitrobenzene sulfonic acid.

During the cancer development, sporadic CRC (or spontaneous, unrelated to the genetics of family and CRC history) typically present a stepwise “normal mucosa-adenoma-dysplasia-carcinoma” sequence, while CAC arises as an “inflamed mucosa-dysplasia-carcinoma” sequence. Moreover, there are also unique histological and genetic alterations [22]. In clinical histopathology, CAC tissues often have a background of chronic inflammation, a higher number of signet ring cells, and a substantial portion of mucinous. Frequently, it invokes a cascade within the abnormal epithelial proliferative region, progressing to invasive adenocarcinoma from flat and non-polypoid dysplasia [9,10,12]. The difference between sporadic and CAC can also be found at the molecular level [23]. This mainly involves pro-inflammatory signaling pathways and immune responses, promoting tumorigenesis by inducing the production of inflammatory mediators, induce the expression of the anti-apoptotic genes, and stimulating cells proliferation and angiogenesis.

These processes can be regulated by microRNA (miRNA). miRNA belongs to a class of highly conserved short single-stranded segments (18–25 nucleotides) non-coding RNA, which post-transcriptionally regulate protein expression inducing messenger RNA (mRNA) degradation and/or inhibit translation of target genes binding to the 3'-untranslated regions

(3'-UTR) to regulate gene expression [24,25]. During the occurrence of IBD, miRNAs play important roles either inhibiting or enhancing immune and inflammation signals by regulating the expression of the positive or negative components of immune signaling pathways associated with IBD and CRC progression. Several miRNAs studies are focusing on the CAC (Table 2); hence, the aim of this review is to discuss the major finding on this topic, and the roles of miRNAs in the progression of IBD and the CAC development.

**Table 2.** Major microRNA studies in colitis-associated colorectal cancer.

miRNA	Target Gene(s)	Function	Reference(s)
		Upregulation	
miR-18a	<i>PIAS3</i>	Proliferation, cell apoptosis	[26]
miR-19a	<i>TNFAIP3</i>	Activate NF- $\kappa$ B signaling	[27]
miR-21	<i>PDCD4, PTEN</i>	Invasion, intravasation, metastasis, apoptosis	[28–30]
miR-26b	<i>CCNDBP1</i>	Tumorigenesis and development of digestive diseases	[31]
miR-31	<i>HIF1, WDR5, IL13RA1</i>	Activate RAS signaling, stimulating tumorigenesis and correlates with serrated CRC	[32–36]
miR-146b	<i>TRAF6, IRAK1</i>		[34]
miR-155	<i>IL13RA1</i>	Negative feedback loop controlling IL-1 $\beta$	[34,37,38]
miR-181b-1	<i>CYLD</i>	Cellular transformation	[30,34]
miR-214	<i>PDLIM2, PTEN</i>	Malignant transformation	[39]
miR-221	<i>PDLIM2</i>		[34]
miR-223	<i>RASA1</i>	Cell proliferation	[40]
miR-301a	<i>BTG1</i>	Promote intestinal inflammation	[41]
		Downregulation	
miR-34a	<i>IL6/EMT/EGR1</i>	Suppresses migration and invasion	[34,42]
miR-124	<i>STAT3/ROCK1</i>	Inhibits neoplastic transformation	[43,44]
miR-139-5p	<i>IGF-1R</i>	Maintain intestinal homeostasis	[45,46]
miR-185-3p	<i>MLCK</i>	Regulate via lncRNA <i>CCAT1</i>	[47]
miR-193a-3p	<i>SLC15A1</i>	Suppress NF- $\kappa$ B signaling	[48,49]

CRC, colorectal cancer; lncRNA, long non-coding RNA.

## 2. MicroRNAs Overexpression Induces Colitis-Associated Colorectal Carcinogenesis

### 2.1. MiR-17-92 Cluster

Both miR-19a and miR-18a belong to miRNA17-92 cluster. MiR-19a has been proven to be an oncomiR, which regulate cell proliferation, differentiation, apoptosis and angiogenesis during the cancer development. In CRC, miR-19a enhances cells invasion, progression and lymph node metastasis by mediating the inhibition of Transglutaminase-2 (*TG-2*) [50], T-cell intracellular antigen 1 (*TIA1*) [51] and the inflammatory cytokine tumor necrosis factor  $\alpha$  (*TNF $\alpha$* ) [52]. Overexpression of miR-19a induces epithelial–mesenchymal transition (EMT) signaling in CRC cells, confirmed by N-cadherin, Vimentin, and Fibronectin levels [52]. In DSS-induced colitis mice treated with miR-19a mimic, colon tumor numbers, sizes and tumor loads are higher compared to the control group; pro-inflammatory cytokines, including IL-1 $\beta$ , IL-6, IL-17a, IFN- $\gamma$  and TNF- $\alpha$  are also upregulated [27,53]. MiR-19a mimic was also administered in the AOM/DSS-induced CRC mice and induced pro-inflammatory cytokines (IL-6, TNF- $\alpha$ , IL-1 $\beta$  and IL-17a), tumor proliferation marker (Ki-67) and NF- $\kappa$ B signaling markers (p-P65 and COX-2) via targeting TNF- $\alpha$ -induced protein 3 (*TNFAIP3*) [27]. The stimulation of TNF- $\alpha$  induces miR-19a expression in CAC and its overexpression activates NF- $\kappa$ B signaling and increases TNFAIP3. The regulatory effects of miR-19a on TNFAIP3 and NF- $\kappa$ B were also found in clinical tissue samples [27].

MiR-18a belongs to miRNA17-92 cluster, as well. Upregulation of miR-18a downregulates Protein Inhibitor Of Activated STAT 3 (*PIAS3*) expression and activates NF- $\kappa$ B and STAT3 in both CAC/CRC patients and AOM/DSS-induced mice. To be more specific, in vitro studies demonstrated that PIAS3 significantly repressed the activation of NF- $\kappa$ B and STAT3, while the activation of NF- $\kappa$ B and STAT3 transcriptionally regulate miR-18a expression level. The PIAS3/NF- $\kappa$ B and STAT3/miR-18a autoregulatory feedback loops

are involved in cell proliferation regulation [26,54]. PIAS3 overexpression or miR-18a knockdown significantly inhibited cell proliferation in the mouse CRC xenograft model. Intracolonic administration of PIAS3 lentivirus or anti-miR-18a lentivirus in AOM/DSS-induced mice led to dramatically reduced tumor sizes/numbers, whereas knockdown of PIAS3 in CAC mice significantly promoted tumor growth [26]. Higher expression of miR-18a can be detected in feces from CRC patients [55].

### 2.2. MiR-21

MiR-21 is one of the most overexpressed and well-studied miRNA in both cancers and inflammatory-related diseases. miR-21 expression patterns can be distinguished in between IBD, CRC, CAC and normal controls and appear related to CRC patients' survival [56–59]. The upregulation of miR-21 in cancer correlates to cell migration, invasion and proliferation, and promotes miR-21-mediated transformation in somatic cells [29,30]. In IBD, the deletion of miR-21 in C57BL/6 mice results in the exacerbation in both T-cells transfer and TNBS-induced colitis models, CD4<sup>+</sup>CD45R<sup>high</sup> T-cells from miR-21<sup>-/-</sup> mice were disposed to Th1 polarization [60]. MiR-21 knockout mice which received AOM/DSS presented a reduction of neoplasms size and numbers and induced inflammatory and carcinogenic cytokines such as IL-6, IL-17A, IL-21, and IL-23 [29]. This process further reduces BCL2 and STAT3 activation, attenuated cancer cells proliferation, simultaneously increase E-cadherin and decrease  $\beta$ -catenin, SOX9 and Ki-67 expressions [29]. Moreover, miR-21 expression in IBD significantly upregulates CD3<sup>+</sup> T-cells and negatively correlates to PDCD4 expression in UC remission patients [59]. The abovementioned miR-21 and PDCD4 correlation can be found in CAC, as well, increasing the apoptosis, and subsequently activating NF- $\kappa$ B [29]. Using a non-transformed mammary epithelial cell MCF-10A overexpressing v-Src, Iliopoulos et al. [30] indicated that the transient activation of v-Src is sufficient to induce transformation. The activation of STAT3 via v-Src enhances the transcription of *MIR21*, leading to increase NF- $\kappa$ B and IL-6 production and inhibit PTEN.

### 2.3. MiR-31-5p

MiR-31 has both oncogenic and suppressive roles in different types of cancers. The phenotype caused by aberrant miR-31 expression seems to be strongly dependent on the endogenous expression levels. In CRC, high level of miR-31 correlates with serrated CRC [61], *KRAS* [62] and *BRAF* [61,63] mutations. MiR-31-5p activates *RAS* signaling pathway via inhibition of *RASA1* translation, increasing CRC cell growth, stimulating tumorigenesis [64]. The expression of *EZH2* reported as a prognostic biomarker candidate for anti-EGFR treatment [65] correlates with miR-31 serrated pathway [66,67]. Thereby, in addition to *RAS* signaling related genes [68], miR-31-5p could potentially be an additional predictor for precision anti-EGFR therapy [61–63,69]. The transcription of *MIR31* can also be activated by NF- $\kappa$ B and STAT3 confirmed by using LoVo CRC cells and organoids derived from mouse colon cells in response to TNF and IL-6 [36]. Moreover, miR-31 negatively correlated with *HIF1AN* expression in CRC tissue samples and cell lines compared with the corresponding adjacent normal tissue [70], and directly regulate *HIF1AN* expression in CRC confirmed by luciferase reporter assay. The presents of *HIF1AN* inhibits hypoxia-inducible factor 1 $\alpha$  (HIF1 $\alpha$ ), and downregulation of *HIF1AN* promotes tumor angiogenesis, cell invasion and proliferation.

The induced expression of miR-31 can be identified in both UC and CD patients [35,37]. Inflamed UC mucosae showed decreased *IL13RA1* mRNA and protein expressions compared to healthy controls. MiR-31 mimics transfection in HT-29 colon cancer cells reduces both *IL13RA1* protein and mRNA expression, blocks pSTAT6, *SOCS1* and *CCL26* expression. The use of miR-31 mimic also downregulate *IL13RA1* in ex vivo human inflamed UC biopsies [37].

MiR-31 expression also targets and inversely correlates IL-25, regulates IL-12/23-mediated Th1/Th17 inflammatory responses during the chronic inflammation process in TNBS-induced and IL-10 knockout colitis mice models [35]. IL-25 is a regulatory cytokine

that has a key role in mucosal immune tolerance during inflammation response. *MIR31*-knockout DSS and TNBS-treated mice established severe colitis, induced immune responses with higher inflammatory cytokine receptors (IL7R and IL17RA) and signaling proteins (GP130) compared to the control group [36]. *IL7R*, *IL17RA* and *IL16ST* are the targets of miR-31-5p confirmed by 3'UTR-Luciferase reporter assays [36]. MiR-31 also regulates Hippo and WNT signaling pathways to promote epithelial regeneration [36]. The expression of miR-31 has been found dysregulated in IBD-related neoplasia compared to adjacent normal tissue in human [32,49] and the AOM/DSS-induced CRC mice models [33]. Mice with colon epithelium-specific deletion of miR-31 were utilized and presented a severe CAC compared to the wild-type mice. WD Repeat Domain 5 (WDR5) [33] and HIF1 [32] could be the targets of miR-31 for CAC development [71].

#### 2.4. MiR-155

MiR-155 is a multi-functional miRNA with inflammation-related and oncogenic roles. High-level of miR-155 can be detected in CRC, promoting cells proliferation, invasion, migration, and closely related to tumor location, TNM staging, metastasis and prognosis [72–75]. MiR-155 reported as a part of a negative feedback loop controlling IL-1 $\beta$  and inflammatory cytokines production during LPS-mediated dendritic cells (DC) activation. This process is directly targeting TAB2, a signaling transduction adaptor of the TLR/IL-1 biochemical cascade in response to microbial stimuli [76]. miR-155 is also characterized as a macrophage response factor and affects several immune-related mediators (TLR-3, IFN- $\beta$  and TNF- $\alpha$ ) [77]. It increases IL-8 throughout the inflammatory process [78,79], modulates the inflammatory phenotype of intestinal fibroblasts and myofibroblasts via NF- $\kappa$ B [80]. A recent study showed that miR-155 mediates intestinal barrier dysfunction in DSS-induced mice colitis through HIF1 $\alpha$ /TFF-3 axis [81]. The knock-down of miR-155 protects the experimental colitis mice by decreasing IFN $\gamma$ , TNF $\alpha$ , IL-6, IL-12 and IL-17 production, reducing Th1 response and suppressing the T-cells activation by DCs [82].

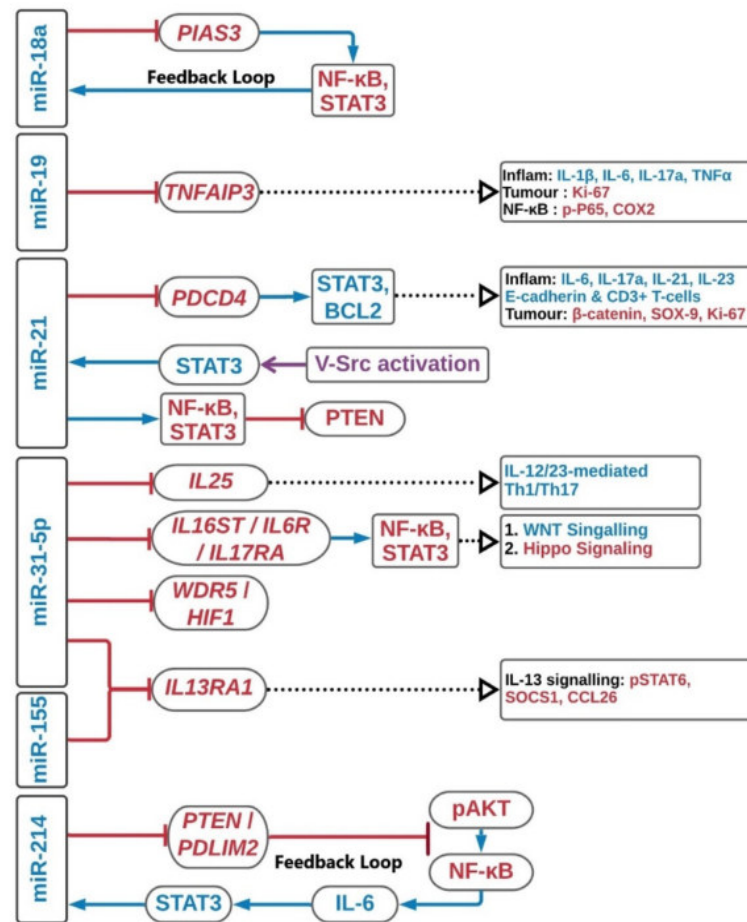
MiR-155 is significantly overexpressed in affected colonic mucosa and neoplastic tissues from IBD patients compared with non-IBD controls and can be detected from the distant non-neoplastic mucosa [57]. It is also highly expressed in the tumor region from patients with CAC compared to the non-tumor colon tissues. The deletion of miR-155 in AOM/DSS-induced CRC mouse models produced a higher grade of epithelial dysplasia, number of polyps and symptom severity scores; the models also have an unfavorable survival rate compared with the control group. The enhanced tumorigenic response in miR-155 knockout mice is associated with increased neutrophils and decreased macrophages activity and to the activation of the TGF $\beta$ /SMAD signaling activity [38]. MiR-155 also in have the same effect as miR-31 [37]. The use of miR-155 mimics in both HT-29 CRC cells and in ex vivo human inflamed UC biopsies reduced mRNA *IL13RA1* expression, reduced IL-13-dependent pSTAT6 via Janus kinases (JAK) and *CCL26* and *SOCS1* expression [37].

#### 2.5. MiRNA-214

MiR-214 has been reported to be a tumor-suppressor miRNA in CRC, which reduce tumor cell growth, migration, invasion [83], and modulates autophagy [84]. Downregulation of miR-214 might happen due to promoter hypermethylation [85] and correlate with liver [86–88], lymphatic [85], and lung metastasis in CRC patients [87]. For instance, several studies have indicated that the increased level of FGFR1 may contribute to increasing CRC liver metastasis regulated by miR-214 [86,88]. The expression of miR-214 may serve as a potential marker to predict CRC patient survival [86].

In UC-associated CRC, it was reported that miR-214 was highly expressed in UC cancer compared with CD cancer and its predicted targeting action on PTEN altered the p53 signaling pathway and caused UC-induced CRC [89]. To further investigated this phenomenon, Polyarchou et al. revealed that high level of miR-214 was detected in active UC or CAC patients' tissues, and correlated with the disease progression [39]. This correlation cannot be found in CD, IBS or healthy controls. To be more specific, miR-214 targets

PDLIM2 and PTEN in UC, and suppresses NF- $\kappa$ B expression and AKT phosphorylation. Then, NF- $\kappa$ B modulates IL-6 expression which regulates STAT3 activity. STAT3-driven transcriptional activation of miR-214 triggers the positive feedback loop circuit, and the circuit is weakened in the inactive state UC. In long-standing UC, overexpression of miR-214 and hyper-activation of the inflammatory feedback loop circuit further increases the colitis-induced CRC development [39]. This feedback circuit can be blocked by a miR-214 chemical inhibitor, which reduced the size and the number of tumors and the colitis severity in several in vivo and in vitro models. The upregulated miRNAs interactions described above are summarized in Figure 1.



**Figure 1.** Upregulation of miRNAs in colitis-associated colorectal cancer. The schematic diagram summarizes the major pathways in the upregulation of miRNAs. Red, blue and purple in the diagram represent inhibition, regulation and enhancer, respectively. Inflammation, inflammation.

### 3. MicroRNAs as Suppressors of Colitis-Associated Colorectal Carcinogenesis

#### 3.1. MiR-124

MiR-124 has been reported to inhibit cell proliferation and metastasis, induce apoptosis and oxidative stress, and correlate with favorable overall CRC patient survival [90–92]. The tumor suppression mediated by miR-124 may contribute to the maintenance of the Warburg effect [91,93], exhibiting a metabolic phenotype characterized by increased glycolysis, regardless of oxygen availability [94]. This effect is partly achieved via controlling the alternative splicing of pyruvate kinase muscle (PKM) isoforms expressions - PKM1 and PKM2 in feedback loops [95]. The expression of miR-124 targets polypyrimidine tract-binding protein 1 (PTB1) to activate the PKM1 and suppress the PKM2, which further downregulates c-Myc, E2F1 as well as STAT3 [91,96,97]. MiR-124 also activates the mitochondrial apoptosis pathway through PKM1 to facilitate HNF4 $\alpha$  binding to the miR-124

promoter region [98] and activate oxidative stress HIF1 $\alpha$  via PKM2 [91,99,100]. Thus, miR-124, PTB1, PKM1 and PKM2 constitute a feedback cascade and regulate cancer cells growth in human CRC [101]. MiR-124 also directly targets DDX6, which first regulates c-Myc, and further regulate the PTB1 levels [93].

In IBD, IL-6/STAT3 signaling has been reported as an important regulator in colon inflammation [102,103] in UC development and CAC progression [104] and can be modulated by miR-124. In particular, the expression of miR-124 was found inversely correlate with STAT3 in both DSS-induced and IL-10 knockout colitis mice models [43] and CRC patients, as well [105]. Downregulation of miR-124 in UC active paediatric patients resulted in upregulation of STAT3 and modulation of its related downstream targets—VEGF, BCL2, BCLXL and MMP9 [43]. Suppression of miR-124 in tissues from UC active paediatric patients was attributed to hypermethylation of STAT3 promoter region. This hypermethylation can also be found in CRC [106,107] and restored by using 5-AZA—a DNA methyltransferases inhibitor [43]. Applying nicotine as a treatment agent could ameliorate UC symptoms through upregulation of miR-124 expression by blocking STAT3 activation in DSS-induced colitis mice and epithelial cells. Therefore, targeting miR-124 and STAT3 by using nicotine may present a novel approach for treating UC [44].

### 3.2. MiR-139-5p

The role of miR-139-5p has been well studied in CRC; it is related to apoptosis, cell cycle arrest, cellular migration and invasion by targeting various coding genes [108–110]. Briefly, lower expression of miR-139-5p in CRC induces apoptosis, concomitantly with upregulation of apoptosis-related genes such as caspase-3, caspase-7, caspase-8 and PARP [108]; increases p27<sup>Kip1</sup> and p21<sup>Cip1/Waf1</sup>, and also increases the G0/G1 phase regulators which cause cell cycle arrest [108]; and suppresses cell migration and invasiveness through IGF-IR/MEK/ERK signaling pathway by targeting IGF-IR, MMP-2, MMP-7 and MMP-9 [108,110]. Moreover, miR-139-5p expression can be regulated by NOTCH1, the knock-down of NOTCH1 phenocopied the inhibitory effect of miR-139-5p on CRC metastasis [108,109], inhibited EMT and enhanced the chemotherapeutic sensitivity of CRC by downregulating BCL2 [111,112]. Moreover, Bian et al. [113] reported that LINC00152—a long non-coding RNA enables to regulate NOTCH1 expression via sponging miR-139-5p, which may control CRC progression and development.

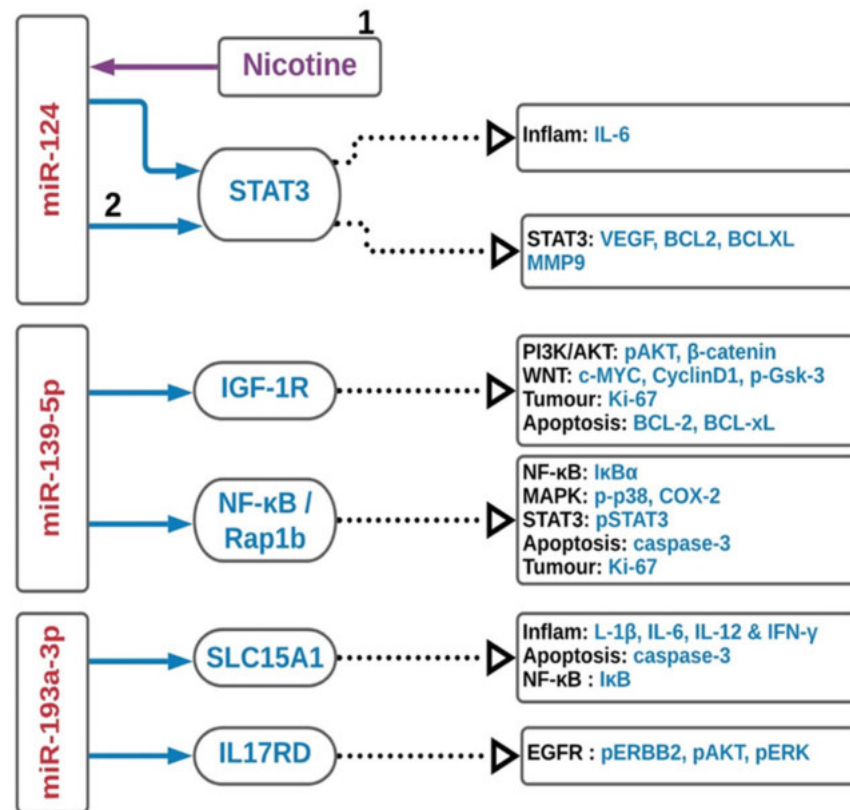
In miR-139-5p knockout mice models, worse clinical symptoms were observed for the DSS-induced colitis compared with the wild-type mice [45,46]. The enhanced formation of intestinal neoplasia was found for the AOM/DSS-induced CRC model. The miR-139-5p knockout mice enhanced the colon inflammation and tumor development via targeting both NF- $\kappa$ B and Rap1b, affect NF- $\kappa$ B, MAPK and STAT3 signaling activities [45]. The suppression of miR-139-5p was also involving Wnt signaling, associated with cell proliferation and differentiation and promoting  $\beta$ -catenin nuclear accumulation [46]. Overexpression of miR-139-5p in CRC cell lines inhibited PI3K/AKT/Wnt signaling pathways through IGF-1R [46]. These results pointed out that miR-139-5p act as a protective agent against both colitis and CRC, maintaining intestinal homeostasis.

### 3.3. MiR-193a-3p

Low expression of miR-193a-3p in CRC is associated with an unfavorable prognosis [114,115]: This miRNA acts as a tumor suppressor miRNA targeting *KRAS* and correlates with *BRAF*-mutated CRC [116]. The expression of miR-193-3p is significantly downregulated in UC active, UC-induced neoplasia and UC-driven CRC compared to adjacent normal tissues [48,49]. Studies revealed that miR-193-3p expression inversely correlates with *SLC15A1*, which is a positive TLR4-mediated inflammatory response regulator, and subsequently suppress NF- $\kappa$ B signaling. The use of miR-193a-3p mimic in CRC tissue significantly relief the colitis symptoms in DSS-induced colitis mice model. Overexpression of *SLC15A1* by using *SLC15A1* 3'-UTR mutant lentiviral vector neutralized the anti-inflammatory effect of miR-193a-3p [48]. IL17RD in UC-associated CRC directly



targets and inversely correlates with miR-193a-3p. MiR-193a-3p transfection reduced cells proliferation acting on EGFR signaling by targeting IL17RD; moreover, it regulates pAKT, pERK and pERBB2 expression [49]. The downregulated miRNAs interactions described above are summarized in Figure 2.



**Figure 2.** Downregulation of miRNAs in colitis-associated colorectal cancer. The schematic diagram summarizes the major pathways in the downregulation of miRNAs. Red, blue and purple in the diagram represent inhibition, regulation and enhancer, respectively. Inflam, inflammation.

#### 4. MicroRNAs as a Treatment Tool for Inflammatory Bowel Disease

The advantage of miRNA-based therapy is to differentially modulate target genes at post-transcriptional levels in multi-pathological pathways and appeared to be more flexible over siRNAs, negatively regulating target genes [117]. MiRNA-based therapies comprise two fundamental strategies: miRNA mimicry and antagonism [118]. Using miRNA mimics to restore miRNA expression have been extensively applied in miRNA research, including miR19a [27], miR-193a-3p [48], miR-31 [37] and miR-155 [37] studies mentioned above. Mimic particles can also be encapsulated into oxidized konjac glucomannan (OKGM) microspheres [36]. For example, the administration of peptosome-MIR31 via enema into the large intestines of mice with DSS-induced colitis demonstrated a reduction of inflammatory response, gain in body weight and increased colon length, promoting epithelial cell proliferation [36]. On the other hand, it is possible to reduce miRNA overexpression aberrantly acting on its target genes by the use of miRNA antagonists [119]; for example, Lu et al. [120] reported that the antagomir miR-155 alleviated DSS-induced colitis in mice by targeting *SHIP-1* and noted that individual mRNA may be regulated by more than one miRNA, and each miRNA may regulate numerous mRNAs [121]. Moreover, the existence of long non-coding RNAs giving another layer of complexity, forming a complex biological regulation, leading to unexpected outcomes [122]. Thus, miRNA delivery is a vital challenge. Since miRNAs might act on different mRNA targets, with different abundance from tissue to tissue, possibly leading to undesired effects, the targeted delivery strategy is crucial for two reasons: first to prioritize target tissue and avoid unspecific and random

tissue distribution and also to stabilize and preserve miRNAs structure and properties in physiological fluids [123].

### 5. MicroRNAs as Predictive Inflammatory Bowel Disease Biomarker(s)

The IBD treatment goal is to obtain remission and recovery of the altered mucosa, in order to avoid or sizing surgical intervention. It is important to mention that the latest therapies target inflammatory processes, for example, pro-inflammatory cytokine TNF inhibitors. However, approximately 30% of the patients are not responding to the treatment from the beginning (primary non-responders), while 50% of the primary responders start to lose the initial therapeutic benefit over time (secondary non-responders) [124–126]. Morilla et al. [127] reported that the neural-network-developed algorithms utilized a total of nine miRNAs with five clinical features in IBD patients associated with anti-TNF monoclonal antibody therapy treatment response. Moreover, fecal-based miRNA screening has been investigated to find the pattern to detect colon diseases, including CRC [55,56,128–130] and IBD [131,132], in the early phases with non-invasive procedures. For example, stool miR-16, miR-21, miR-223 and miR-1246 expression resulted in being upregulated in active CD and UC patients compared to healthy controls and circulating miRNAs detection may further prevent the chance of *Clostridioides difficile* infection [133] in IBD patients. Thus, miRNAs are the potential biomarkers to further investigate for an early non-invasive IBD screening [134].

### 6. Conclusions

MiRNAs play a variety of biological roles in cancer and (chronic) inflammation development, which induces an unphysiological condition favoring the malignant transformation. MiRNAs nowadays are becoming more and more important because of their role in biochemical pathways regulation, and some of them are showing promising therapeutic effects. There are a wide variety of genes finely regulated by miRNAs; the strategy, as we mentioned, could be utilized to restore their physiological levels to maintain the homeostasis. It is also important to take into account that miRNAs may have different binding sites on different targets, and their deregulation is not easy to control, which is also facing the similar challenge to the conventional pharmacological drugs development. To build a real translational therapeutic approach, it is important to study the physiological levels, first in silico and then in vitro and in vivo to compare with healthy and cancerous tissues. The next step will be to clarify and assess the putative mechanism of action and the corresponding target genes involved in miRNA regulation and the effectiveness of the pathology. Moreover, miRNAs may be exploited as biomarkers; their differential expression is important for early diagnosis and even to predict CRC conventional chemotherapy response. We hope this manuscript will help future speculation and discussion on CRC diagnosis and precision treatments, with the involvement of those promising small non-coding RNAs.

**Author Contributions:** Conceptualization, T.O.Y.; Writing—original draft, M.B., M.G.F., W.-Y.L., T.O.Y.; Writing—review & editing, M.B., M.G.F., W.-Y.L., M.C., S.Z., T.O.Y.; Visualization, M.B., W.-Y.L., T.O.Y.; Supervision, M.B., M.G.F., F.R., A.O., A.L., A.M.C., M.S., W.-Y.L., M.A., P.S., M.C., S.Z., T.O.Y.; Funding Acquisition, M.C., S.Z.. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by AIRC (IG 2017, code 20711).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

BTG1	BTG Anti-Proliferation Factor 1
CCNDBP1	Cyclin D1 Binding Protein 1
CYLD	CYLD Lysine 63 Deubiquitinase
EGR1	Early growth response protein 1
EMT	Epithelial–mesenchymal transition
HIF1	Hypoxia-inducible factor 1
IGF-1R	Insulin-like growth factor 1
IL13RA1	Interleukin 13 Receptor Subunit Alpha 1
IL6	Interleukin-6
IRAK1	Interleukin-1 receptor-associated kinase 1
MLCK	Myosin light chain kinase
PDCD4	Programmed Cell Death 4
PDLIM2	PDZ And LIM Domain 2
PIAS3	Protein Inhibitor Of Activated STAT 3
PTEN	Phosphatase and tensin homolog
RASA1	RAS p21 protein activator 1
ROCK1	Rho Associated Coiled-Coil Containing Protein Kinase 1
SLC15A1	Solute Carrier Family 15 Member 1
STAT3	Signal transducer and activator of transcription 3
TNFAIP3	TNF alpha induced protein 3
TRAF6	TNF Receptor Associated Factor 6
WDR5	WD Repeat Domain 5

## References

1. Abraham, C.; Cho, J.H. Inflammatory bowel disease. *N. Engl. J. Med.* **2009**, *361*, 2066–2078. [[CrossRef](#)] [[PubMed](#)]
2. Turner, J.R. Molecular Basis of Epithelial Barrier Regulation. *Am. J. Pathol.* **2006**, *169*, 1901–1909. [[CrossRef](#)] [[PubMed](#)]
3. Bernstein, C.N.; Blanchard, J.F.; Kliever, E.; Wajda, A. Cancer risk in patients with inflammatory bowel disease: A population-based study. *Cancer* **2001**, *91*, 854–862. [[CrossRef](#)]
4. Jess, T.; Rungoe, C.; Peyrin-Biroulet, L. Risk of colorectal cancer in patients with ulcerative colitis: A meta-analysis of population-based cohort studies. *Clin. Gastroenterol. Hepatol.* **2012**, *10*, 639–645. [[CrossRef](#)] [[PubMed](#)]
5. Rubin, D.T.; Huo, D.; Kinnucan, J.A.; Sedrak, M.S.; McCullom, N.E.; Bunnag, A.P.; Raun-Royer, E.P.; Cohen, R.D.; Hanauer, S.B.; Hart, J.; et al. Inflammation Is an Independent Risk Factor for Colonic Neoplasia in Patients with Ulcerative Colitis: A Case—Control Study. *Clin. Gastroenterol. Hepatol.* **2013**, *11*, 1601–1608. [[CrossRef](#)]
6. Nieminen, U.; Jussila, A.; Nordling, S.; Mustonen, H.; Färkkilä, M.A. Inflammation and disease duration have a cumulative effect on the risk of dysplasia and carcinoma in IBD: A case-control observational study based on registry data. *Int. J. Cancer* **2014**, *134*, 189–196. [[CrossRef](#)]
7. Eaden, J.A. The risk of colorectal cancer in ulcerative colitis: A meta-analysis. *Gut* **2001**, *48*, 526–535. [[CrossRef](#)]
8. Canavan, C.; Abrams, K.R.; Mayberry, J. Meta-analysis: Colorectal and small bowel cancer risk in patients with Crohn’s disease. *Aliment. Pharmacol. Ther.* **2006**, *23*, 1097–1104. [[CrossRef](#)]
9. Jensen, A.B.; Larsen, M.; Gislum, M.; Skriver, M.V.; Jepsen, P.; Norgaard, B.; Sorensen, H.T. Survival after Colorectal Cancer in Patients with Ulcerative Colitis: A Nationwide Population-Based Danish Study. *Am. J. Gastroenterol.* **2006**, *101*, 1283–1287. [[CrossRef](#)]
10. Watanabe, T.; Konishi, T.; Kishimoto, J.; Kotake, K.; Muto, T.; Sugihara, K. Ulcerative colitis-associated colorectal cancer shows a poorer survival than sporadic colorectal cancer: A nationwide Japanese study. *Inflamm. Bowel Dis.* **2011**, *17*, 802–808. [[CrossRef](#)]
11. Jess, T.; Loftus, E.V.; Velayos, F.S.; Harmsen, W.S.; Zinsmeister, A.R.; Smyrk, T.C.; Schleck, C.D.; Tremaine, W.J.; Melton, L.J.; Munkholm, P.; et al. Risk of Intestinal Cancer in Inflammatory Bowel Disease: A Population-Based Study from Olmsted County, Minnesota. *Gastroenterology* **2006**, *130*, 1039–1046. [[CrossRef](#)] [[PubMed](#)]
12. Xie, J.; Itzkowitz, S.H. Cancer in inflammatory bowel disease. *World J. Gastroenterol.* **2008**, *14*, 378–389. [[CrossRef](#)] [[PubMed](#)]
13. Rutter, M.; Saunders, B.; Wilkinson, K.; Rumbles, S.; Schofield, G.; Kamm, M.; Williams, C.; Price, A.; Talbot, I.; Forbes, A. Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis. *Gastroenterology* **2004**, *126*, 451–459. [[CrossRef](#)]
14. Oh, S.Y.; Cho, K.-A.; Kang, J.L.; Kim, K.H.; Woo, S.-Y. Comparison of experimental mouse models of inflammatory bowel disease. *Int. J. Mol. Med.* **2014**, *33*, 333–340. [[CrossRef](#)]
15. Morgan, S.J.; Elangbam, C.S. Gastrointestinal Injury Models. In *Drug Discovery Toxicology: From Target Assessment to Translational Biomarkers*; John Wiley & Sons: Hoboken, NJ, USA, 2016; pp. 273–278.
16. Antoniou, E.; Margonis, G.A.; Angelou, A.; Pikouli, A.; Argiri, P.; Karavokyros, I.; Papalois, A.; Pikoulis, E. The TNBS-induced colitis animal model: An overview. *Ann. Med. Surg.* **2016**, *11*, 9–15. [[CrossRef](#)]

17. Pizarro, T.T.; Arseneau, K.O.; Cominelli, F. Lessons from genetically engineered animal models XI. Novel mouse models to study pathogenic mechanisms of Crohn's disease. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2000**, *278*, G665–G669. [[CrossRef](#)]
18. Byrne, F.R.; Morony, S.; Warmington, K.; Geng, Z.; Brown, H.L.; Flores, S.A.; Fiorino, M.; Yin, S.L.; Hill, D.; Porkess, V.; et al. CD4+CD45RB<sup>hi</sup> T cell transfer induced colitis in mice is accompanied by osteopenia which is treatable with recombinant human osteoprotegerin. *Gut* **2005**, *54*, 78–86. [[CrossRef](#)]
19. Seamons, A.; Treuting, P.M.; Brabb, T.; Maggio-Price, L. Characterization of Dextran Sodium Sulfate-Induced Inflammation and Colonic Tumorigenesis in Smad3<sup>-/-</sup> Mice with Dysregulated TGFβ. *PLoS ONE* **2013**, *8*, e79182. [[CrossRef](#)]
20. Lee, H.-N.; Yum, H.-W.; Surh, Y.-J. The Azoxymethane Plus Dextran Sulfate Sodium-Induced Mouse Colon Cancer Model for the Study of Dietary Chemoprevention of Inflammation-Associated Carcinogenesis. In *Cancer Prevention*; Humana Press: New York, NY, USA, 2014; pp. 155–172.
21. Berg, D.J.; Davidson, N.; Kühn, R.; Müller, W.; Menon, S.; Holland, G.; Thompson-Snipes, L.; Leach, M.W.; Rennick, D. Enterocolitis and colon cancer in interleukin-10-deficient mice are associated with aberrant cytokine production and CD4(+) TH1-like responses. *J. Clin. Investig.* **1996**, *98*, 1010–1020. [[CrossRef](#)]
22. Feagins, L.A.; Souza, R.F.; Spechler, S.J. Carcinogenesis in IBD: Potential targets for the prevention of colorectal cancer. *Nat. Rev. Gastroenterol. Hepatol.* **2009**, *6*, 297–305. [[CrossRef](#)]
23. Robles, A.I.; Traverso, G.; Zhang, M.; Roberts, N.J.; Khan, M.A.; Joseph, C.; Lauwers, G.Y.; Selaru, F.M.; Popoli, M.; Pittman, M.E.; et al. Whole-Exome Sequencing Analyses of Inflammatory Bowel Disease–Associated Colorectal Cancers. *Gastroenterology* **2016**, *150*, 931–943. [[CrossRef](#)]
24. O'Brien, J.; Hayder, H.; Zayed, Y.; Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front. Endocrinol.* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]
25. Tong, A.W.; Nemunaitis, J. Modulation of miRNA activity in human cancer: A new paradigm for cancer gene therapy? *Cancer Gene Ther.* **2008**, *15*, 341–355. [[CrossRef](#)] [[PubMed](#)]
26. Ma, J.; Yang, Y.; Fu, Y.; Guo, F.; Zhang, X.; Xiao, S.; Zhu, W.; Huang, Z.; Zhang, J.; Chen, J. PIAS3-mediated feedback loops promote chronic colitis-associated malignant transformation. *Theranostics* **2018**, *8*, 3022–3037. [[CrossRef](#)]
27. Wang, T.; Xu, X.; Xu, Q.; Ren, J.; Shen, S.; Fan, C.; Hou, Y. miR-19a promotes colitis-associated colorectal cancer by regulating tumor necrosis factor alpha-induced protein 3-NF-κB feedback loops. *Oncogene* **2017**, *36*, 3240–3251. [[CrossRef](#)]
28. Schetter, A.J.; Leung, S.Y.; Sohn, J.J.; Zanetti, K.A.; Bowman, E.D.; Yanaihara, N.; Yuen, S.T.; Chan, T.L.; Kwong, D.L.W.; Au, G.K.H.; et al. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA* **2008**, *299*, 425–436. [[CrossRef](#)]
29. Shi, C.; Yang, Y.; Xia, Y.; Okugawa, Y.; Yang, J.; Liang, Y.; Chen, H.; Zhang, P.; Wang, F.; Han, H.; et al. Novel evidence for an oncogenic role of microRNA-21 in colitis-associated colorectal cancer. *Gut* **2016**, *65*, 1470–1481. [[CrossRef](#)]
30. Iliopoulos, D.; Jaeger, S.A.; Hirsch, H.A.; Bulyk, M.L.; Struhl, K. STAT3 Activation of miR-21 and miR-181b-1 via PTEN and CYLD Are Part of the Epigenetic Switch Linking Inflammation to Cancer. *Mol. Cell* **2010**, *39*, 493–506. [[CrossRef](#)]
31. Benderska, N.; Dittrich, A.-L.; Knaup, S.; Rau, T.T.; Neufert, C.; Wach, S.; Fahlbusch, F.B.; Rauh, M.; Wirtz, R.M.; Agaimy, A.; et al. miRNA-26b Overexpression in Ulcerative Colitis-associated Carcinogenesis. *Inflamm. Bowel Dis.* **2015**, *21*, 2039–2051. [[CrossRef](#)]
32. Olaru, A.V.; Selaru, F.M.; Mori, Y.; Vazquez, C.; David, S.; Paun, B.; Cheng, Y.; Jin, Z.; Yang, J.; Agarwal, R.; et al. Dynamic changes in the expression of MicroRNA-31 during inflammatory bowel disease-associated neoplastic transformation. *Inflamm. Bowel Dis.* **2011**, *17*, 221–231. [[CrossRef](#)]
33. Liu, Z.; Bai, J.; Zhang, L.; Lou, F.; Ke, F.; Cai, W.; Wang, H. Conditional knockout of microRNA-31 promotes the development of colitis associated cancer. *Biochem. Biophys. Res. Commun.* **2017**, *490*, 62–68. [[CrossRef](#)]
34. El-Daly, S.M.; Omara, E.A.; Hussein, J.; Youness, E.R.; El-Khayat, Z. Differential expression of miRNAs regulating NF-κB and STAT3 crosstalk during colitis-associated tumorigenesis. *Mol. Cell. Probes* **2019**, *47*, 101442. [[CrossRef](#)]
35. Shi, T.; Xie, Y.; Fu, Y.; Zhou, Q.; Ma, Z.; Ma, J.; Huang, Z.; Zhang, J.; Chen, J. The signaling axis of microRNA-31/interleukin-25 regulates Th1/Th17-mediated inflammation response in colitis. *Mucosal Immunol.* **2017**, *10*, 983–995. [[CrossRef](#)]
36. Tian, Y.; Xu, J.; Li, Y.; Zhao, R.; Du, S.; Lv, C.; Wu, W.; Liu, R.; Sheng, X.; Song, Y.; et al. MicroRNA-31 Reduces Inflammatory Signaling and Promotes Regeneration in Colon Epithelium, and Delivery of Mimics in Microspheres Reduces Colitis in Mice. *Gastroenterology* **2019**, *156*, 2281–2296.e6. [[CrossRef](#)]
37. Gwiggner, M.; Martinez-Nunez, R.T.; Whiteoak, S.R.; Bondanese, V.P.; Claridge, A.; Collins, J.E.; Cummings, J.R.F.; Sanchez-Elsner, T. MicroRNA-31 and MicroRNA-155 Are Overexpressed in Ulcerative Colitis and Regulate IL-13 Signaling by Targeting Interleukin 13 Receptor α-1. *Genes* **2018**, *9*, 85. [[CrossRef](#)]
38. Velázquez, K.T.; Enos, R.T.; McClellan, J.L.; Cranford, T.L.; Chatzistamou, I.; Singh, U.P.; Nagarkatti, M.; Nagarkatti, P.S.; Fan, D.; Murphy, E.A. MicroRNA-155 deletion promotes tumorigenesis in the azoxymethane-dextran sulfate sodium model of colon cancer. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2016**, *310*, G347–G358. [[CrossRef](#)]
39. Polyarchou, C.; Hommes, D.W.; Palumbo, T.; Hatziapostolou, M.; Koutsoumpa, M.; Koukos, G.; van der Meulen-de Jong, A.E.; Oikonomopoulos, A.; van Deen, W.K.; Vorvis, C.; et al. MicroRNA214 Is Associated With Progression of Ulcerative Colitis, and Inhibition Reduces Development of Colitis and Colitis-Associated Cancer in Mice. *Gastroenterology* **2015**, *149*, 981–992.e11. [[CrossRef](#)]

40. Sun, D.; Wang, C.; Long, S.; Ma, Y.; Guo, Y.; Huang, Z.; Chen, X.; Zhang, C.; Chen, J.; Zhang, J. C/EBP- $\beta$ -activated microRNA-223 promotes tumour growth through targeting RASA1 in human colorectal cancer. *Br. J. Cancer* **2015**, *112*, 1491–1500. [[CrossRef](#)] [[PubMed](#)]
41. He, C.; Yu, T.; Shi, Y.; Ma, C.; Yang, W.; Fang, L.; Sun, M.; Wu, W.; Xiao, F.; Guo, F.; et al. MicroRNA 301A Promotes Intestinal Inflammation and Colitis-Associated Cancer Development by Inhibiting BTG1. *Gastroenterology* **2017**, *152*, 1434–1448. [[CrossRef](#)] [[PubMed](#)]
42. Zhu, W.; Long, J.; Yin, Y.; Guo, H.; Jiang, E.; Li, Y.; He, Q.; Zeng, C.; Sun, Y. MicroRNA-34a suppresses the invasion and migration of colorectal cancer cells by enhancing EGR1 and inhibiting vimentin. *Exp. Ther. Med.* **2019**. [[CrossRef](#)]
43. Koukos, G.; Polytaichou, C.; Kaplan, J.L.; Morley-Fletcher, A.; Gras-Miralles, B.; Kokkotou, E.; Baril-Dore, M.; Pothoulakis, C.; Winter, H.S.; Iliopoulos, D. MicroRNA-124 Regulates STAT3 Expression and Is Down-regulated in Colon Tissues of Pediatric Patients With Ulcerative Colitis. *Gastroenterology* **2013**, *145*, 842–852. [[CrossRef](#)] [[PubMed](#)]
44. Qin, Z.; Wan, J.-J.; Sun, Y.; Wu, T.; Wang, P.-Y.; Du, P.; Su, D.-F.; Yang, Y.; Liu, X. Nicotine protects against DSS colitis through regulating microRNA-124 and STAT3. *J. Mol. Med.* **2017**, *95*, 221–233. [[CrossRef](#)] [[PubMed](#)]
45. Zou, F.; Mao, R.; Yang, L.; Lin, S.; Lei, K.; Zheng, Y.; Ding, Y.; Zhang, P.; Cai, G.; Liang, X.; et al. Targeted deletion of miR-139-5p activates MAPK, NF- $\kappa$ B and STAT3 signaling and promotes intestinal inflammation and colorectal cancer. *FEBS J.* **2016**, *283*, 1438–1452. [[CrossRef](#)]
46. Mao, R.; Zou, F.; Yang, L.; Lin, S.; Li, Y.; Ma, M.; Yin, P.; Liang, X.; Liu, J. The loss of MiR-139-5p promotes colitis-associated tumorigenesis by mediating PI3K/AKT/Wnt signaling. *Int. J. Biochem. Cell Biol.* **2015**, *69*, 153–161. [[CrossRef](#)] [[PubMed](#)]
47. Ma, D.; Cao, Y.; Wang, Z.; He, J.; Chen, H.; Xiong, H.; Ren, L.; Shen, C.; Zhang, X.; Yan, Y.; et al. CCAT1 lncRNA Promotes Inflammatory Bowel Disease Malignancy by Destroying Intestinal Barrier via Downregulating miR-185-3p. *Inflamm. Bowel Dis.* **2019**, *25*, 862–874. [[CrossRef](#)] [[PubMed](#)]
48. Dai, X.; Chen, X.; Chen, Q.; Shi, L.; Liang, H.; Zhou, Z.; Liu, Q.; Pang, W.; Hou, D.; Wang, C.; et al. MicroRNA-193a-3p Reduces Intestinal Inflammation in Response to Microbiota via Down-regulation of Colonic PepT1. *J. Biol. Chem.* **2015**, *290*, 16099–16115. [[CrossRef](#)] [[PubMed](#)]
49. Pekow, J.; Meckel, K.; Dougherty, U.; Huang, Y.; Chen, X.; Almoghrabi, A.; Mustafi, R.; Ayaloglu-Butun, F.; Deng, Z.; Haider, H.I.; et al. miR-193a-3p is a Key Tumor Suppressor in Ulcerative Colitis-Associated Colon Cancer and Promotes Carcinogenesis through Upregulation of IL17RD. *Clin. Cancer Res.* **2017**, *23*, 5281–5291. [[CrossRef](#)]
50. Cellura, D.; Pickard, K.; Quarantino, S.; Parker, H.; Strefford, J.C.; Thomas, G.J.; Mitter, R.; Mirnezami, A.H.; Peake, N.J. miR-19-Mediated Inhibition of Transglutaminase-2 Leads to Enhanced Invasion and Metastasis in Colorectal Cancer. *Mol. Cancer Res.* **2015**, *13*, 1095–1105. [[CrossRef](#)]
51. Liu, Y.; Liu, R.; Yang, F.; Cheng, R.; Chen, X.; Cui, S.; Gu, Y.; Sun, W.; You, C.; Liu, Z.; et al. miR-19a promotes colorectal cancer proliferation and migration by targeting TIA1. *Mol. Cancer* **2017**, *16*, 53. [[CrossRef](#)]
52. Huang, L.; Wang, X.; Wen, C.; Yang, X.; Song, M.; Chen, J.; Wang, C.; Zhang, B.; Wang, L.; Iwamoto, A.; et al. Hsa-miR-19a is associated with lymph metastasis and mediates the TNF- $\alpha$  induced epithelial-to-mesenchymal transition in colorectal cancer. *Sci. Rep.* **2015**, *5*, 13350. [[CrossRef](#)]
53. Chen, B.; She, S.; Li, D.; Liu, Z.; Yang, X.; Zeng, Z.; Liu, F. Role of miR-19a targeting TNF- $\alpha$  in mediating ulcerative colitis. *Scand. J. Gastroenterol.* **2013**, *48*, 815–824. [[CrossRef](#)] [[PubMed](#)]
54. Zhang, L.; Li, J.; Wang, Q.; Meng, G.; Lv, X.; Zhou, H.; Li, W.; Zhang, J. The relationship between microRNAs and the STAT3-related signaling pathway in cancer. *Tumor Biol.* **2017**, *39*, 101042831771986. [[CrossRef](#)]
55. Yau, T.O.; Wu, C.W.; Dong, Y.; Tang, C.-M.; Ng, S.S.M.; Chan, F.K.L.; Sung, J.J.Y.; Yu, J. microRNA-221 and microRNA-18a identification in stool as potential biomarkers for the non-invasive diagnosis of colorectal carcinoma. *Br. J. Cancer* **2014**, *111*, 1765–1771. [[CrossRef](#)] [[PubMed](#)]
56. Yau, T.O.; Tang, C.-M.; Harriss, E.K.; Dickins, B.; Polytaichou, C. Faecal microRNAs as a non-invasive tool in the diagnosis of colonic adenomas and colorectal cancer: A meta-analysis. *Sci. Rep.* **2019**, *9*, 9491. [[CrossRef](#)] [[PubMed](#)]
57. Svrcek, M.; El-Murr, N.; Wanherdrick, K.; Dumont, S.; Beaugerie, L.; Cosnes, J.; Colombel, J.-F.; Turet, E.; Fléjou, J.-F.; Lesuffleur, T.; et al. Overexpression of microRNAs-155 and 21 targeting mismatch repair proteins in inflammatory bowel diseases. *Carcinogenesis* **2013**, *34*, 828–834. [[CrossRef](#)]
58. Shi, C.; Liang, Y.; Yang, J.; Xia, Y.; Chen, H.; Han, H.; Yang, Y.; Wu, W.; Gao, R.; Qin, H. MicroRNA-21 Knockout Improve the Survival Rate in DSS Induced Fatal Colitis through Protecting against Inflammation and Tissue Injury. *PLoS ONE* **2013**, *8*, e66814. [[CrossRef](#)]
59. Ando, Y.; Mazzurana, L.; Forkel, M.; Okazaki, K.; Aoi, M.; Schmidt, P.T.; Mjösberg, J.; Bresso, F. Downregulation of MicroRNA-21 in Colonic CD3+ T Cells in UC Remission. *Inflamm. Bowel Dis.* **2016**, *22*, 2788–2793. [[CrossRef](#)]
60. Wu, F.; Dong, F.; Arendovich, N.; Zhang, J.; Huang, Y.; Kwon, J.H. Divergent influence of microRNA-21 deletion on murine colitis phenotypes. *Inflamm. Bowel Dis.* **2014**, *20*, 1972–1985. [[CrossRef](#)]
61. Nosh, K.; Igarashi, H.; Nojima, M.; Ito, M.; Maruyama, R.; Yoshii, S.; Naito, T.; Sukawa, Y.; Mikami, M.; Sumioka, W.; et al. Association of microRNA-31 with BRAF mutation, colorectal cancer survival and serrated pathway. *Carcinogenesis* **2014**, *35*, 776–783. [[CrossRef](#)]

62. Kiss, I.; Mlcochova, J.; Bortlicek, Z.; Poprach, A.; Drabek, J.; Vychytilova-Faltejskova, P.; Svoboda, M.; Buchler, T.; Batko, S.; Ryska, A.; et al. Efficacy and Toxicity of Panitumumab After Progression on Cetuximab and Predictive Value of MiR-31-5p in Metastatic Wild-type KRAS Colorectal Cancer Patients. *Anticancer Res.* **2016**, *36*, 4955–4960. [[CrossRef](#)]
63. Ito, M.; Mitsushashi, K.; Igarashi, H.; Noshio, K.; Naito, T.; Yoshii, S.; Takahashi, H.; Fujita, M.; Sukawa, Y.; Yamamoto, E.; et al. MicroRNA-31 expression in relation to BRAF mutation, CpG island methylation and colorectal continuum in serrated lesions. *Int. J. Cancer* **2014**, *135*, 2507–2515. [[CrossRef](#)]
64. Sun, D.; Yu, F.; Ma, Y.; Zhao, R.; Chen, X.; Zhu, J.; Zhang, C.-Y.; Chen, J.; Zhang, J. MicroRNA-31 Activates the RAS Pathway and Functions as an Oncogenic MicroRNA in Human Colorectal Cancer by Repressing RAS p21 GTPase Activating Protein 1 (RAS A1). *J. Biol. Chem.* **2013**, *288*, 9508–9518. [[CrossRef](#)] [[PubMed](#)]
65. Yamamoto, I.; Noshio, K.; Kanno, S.; Igarashi, H.; Kurihara, H.; Ishigami, K.; Ishiguro, K.; Mitsushashi, K.; Maruyama, R.; Koide, H.; et al. EZH2 expression is a prognostic biomarker in patients with colorectal cancer treated with anti-EGFR therapeutics. *Oncotarget* **2017**, *8*, 17810. [[CrossRef](#)]
66. Aoki, H.; Noshio, K.; Igarashi, H.; Ito, M.; Mitsushashi, K.; Naito, T.; Yamamoto, E.; Tanuma, T.; Nomura, M.; Maguchi, H.; et al. MicroRNA-31 expression in colorectal serrated pathway progression. *World J. Gastroenterol.* **2014**, *20*, 12346–12349. [[CrossRef](#)]
67. Kurihara, H.; Maruyama, R.; Ishiguro, K.; Kanno, S.; Yamamoto, I.; Ishigami, K.; Mitsushashi, K.; Igarashi, H.; Ito, M.; Tanuma, T.; et al. The relationship between EZH2 expression and microRNA-31 in colorectal cancer and the role in evolution of the serrated pathway. *Oncotarget* **2016**, *7*. [[CrossRef](#)] [[PubMed](#)]
68. Yau, T.O. Precision treatment in colorectal cancer: Now and the future. *JGH Open J. Gastroenterol. Hepatol.* **2019**, *3*, 361–369. [[CrossRef](#)]
69. Igarashi, H.; Kurihara, H.; Mitsushashi, K.; Ito, M.; Okuda, H.; Kanno, S.; Naito, T.; Yoshii, S.; Takahashi, H.; Kusumi, T.; et al. Association of MicroRNA-31-5p with Clinical Efficacy of Anti-EGFR Therapy in Patients with Metastatic Colorectal Cancer. *Ann. Surg. Oncol.* **2015**, *22*, 2640–2648. [[CrossRef](#)]
70. Chen, T.; Yao, L.-Q.; Shi, Q.; Ren, Z.; Ye, L.-C.; Xu, J.-M.; Zhou, P.-H.; Zhong, Y.-S. MicroRNA-31 contributes to colorectal cancer development by targeting factor inhibiting HIF-1 $\alpha$  (FIH-1). *Cancer Biol. Ther.* **2014**, *15*, 516–523. [[CrossRef](#)]
71. Necela, B.M.; Carr, J.M.; Asmann, Y.W.; Thompson, E.A. Differential expression of microRNAs in tumors from chronically inflamed or genetic (APC(Min/+)) models of colon cancer. *PLoS ONE* **2011**, *6*, e18501. [[CrossRef](#)] [[PubMed](#)]
72. Qu, Y.-L.; Wang, H.-F.; Sun, Z.-Q.; Tang, Y.; Han, X.-N.; Yu, X.-B.; Liu, K. Up-regulated miR-155-5p promotes cell proliferation, invasion and metastasis in colorectal carcinoma. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 6988–6994. [[PubMed](#)]
73. Khoshinani, H.M.; Afshar, S.; Pashaki, A.S.; Mahdavinezhad, A.; Nikzad, S.; Najafi, R.; Amini, R.; Gholami, M.H.; Khoshghadam, A.; Saidijam, M. Involvement of miR-155/FOXO3a and miR-222/P TEN in acquired radioresistance of colorectal cancer cell line. *Jpn. J. Radiol.* **2017**, *35*, 664–672. [[CrossRef](#)] [[PubMed](#)]
74. Al-Haidari, A.A.; Syk, I.; Thorlacius, H. MiR-155-5p positively regulates CCL17-induced colon cancer cell migration by targeting RhoA. *Oncotarget* **2017**, *8*, 14887–14896. [[CrossRef](#)]
75. Bakirtzi, K.; Hatziapostolou, M.; Karagiannides, I.; Polyarchou, C.; Jaeger, S.; Iliopoulos, D.; Pothoulakis, C. Neurotensin signaling activates microRNAs-21 and -155 and Akt, promotes tumor growth in mice, and is increased in human colon tumors. *Gastroenterology* **2011**, *141*, 1749–1761.e1. [[CrossRef](#)] [[PubMed](#)]
76. Ceppi, M.; Pereira, P.M.; Dunand-Sauthier, I.; Barras, E.; Reith, W.; Santos, M.A.; Pierre, P. MicroRNA-155 modulates the interleukin-1 signaling pathway in activated human monocyte-derived dendritic cells. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 2735–2740. [[CrossRef](#)] [[PubMed](#)]
77. O'Connell, R.M.; Taganov, K.D.; Boldin, M.P.; Cheng, G.; Baltimore, D. MicroRNA-155 is induced during the macrophage inflammatory response. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1604–1609. [[CrossRef](#)]
78. Min, M.; Peng, L.; Yang, Y.; Guo, M.; Wang, W.; Sun, G. MicroRNA-155 Is Involved in the Pathogenesis of Ulcerative Colitis by Targeting FOXO3a. *Inflamm. Bowel Dis.* **2014**, *20*, 652–659. [[CrossRef](#)]
79. Cremer, T.J.; Ravneberg, D.H.; Clay, C.D.; Piper-Hunter, M.G.; Marsh, C.B.; Elton, T.S.; Gunn, J.S.; Amer, A.; Kanneganti, T.-D.; Schlesinger, L.S.; et al. MiR-155 Induction by *F. novicida* but Not the Virulent *F. tularensis* Results in SHIP Down-Regulation and Enhanced Pro-Inflammatory Cytokine Response. *PLoS ONE* **2009**, *4*, e8508. [[CrossRef](#)]
80. Pathak, S.; Grillo, A.R.; Scarpa, M.; Brun, P.; D'Incà, R.; Nai, L.; Banerjee, A.; Cavallo, D.; Barzon, L.; Palù, G.; et al. MiR-155 modulates the inflammatory phenotype of intestinal myofibroblasts by targeting SOCS1 in ulcerative colitis. *Exp. Mol. Med.* **2015**, *47*, e164. [[CrossRef](#)]
81. Liu, Y.; Zhu, F.; Li, H.; Fan, H.; Wu, H.; Dong, Y.; Chu, S.; Tan, C.; Wang, Q.; He, H.; et al. MiR-155 contributes to intestinal barrier dysfunction in DSS-induced mice colitis via targeting HIF-1 $\alpha$ /TFF-3 axis. *Aging* **2020**, *12*, 14966–14977. [[CrossRef](#)]
82. Singh, U.P.; Murphy, A.E.; Enos, R.T.; Shamran, H.A.; Singh, N.P.; Guan, H.; Hegde, V.L.; Fan, D.; Price, R.L.; Taub, D.D.; et al. miR-155 deficiency protects mice from experimental colitis by reducing T helper type 1/type 17 responses. *Immunology* **2014**, *143*, 478–489. [[CrossRef](#)]
83. Chandrasekaran, K.S.; Sathyanarayanan, A.; Karunakaran, D. MicroRNA-214 suppresses growth, migration and invasion through a novel target, high mobility group AT-hook 1, in human cervical and colorectal cancer cells. *Br. J. Cancer* **2016**, *115*, 741–751. [[CrossRef](#)] [[PubMed](#)]
84. Hu, J.L.; He, G.Y.; Lan, X.L.; Zeng, Z.C.; Guan, J.; Ding, Y.; Qian, X.L.; Liao, W.T.; Ding, Y.Q.; Liang, L. Inhibition of ATG12-mediated autophagy by miR-214 enhances radiosensitivity in colorectal cancer. *Oncogenesis* **2018**, *7*, 16. [[CrossRef](#)]

85. He, G.Y.; Hu, J.L.; Zhou, L.; Zhu, X.H.; Xin, S.N.; Zhang, D.; Lu, G.F.; Liao, W.T.; Ding, Y.Q.; Liang, L. The FOXD3/miR-214/MED19 axis suppresses tumour growth and metastasis in human colorectal cancer. *Br. J. Cancer* **2016**, *115*, 1367–1378. [[CrossRef](#)]
86. Chen, D.-L.; Wang, Z.; Zeng, Z.-L.; Wu, W.-J.; Zhang, D.-S.; Luo, H.-Y.; Wang, F.-H.; Qiu, M.-Z.; Wang, D.; Ren, C.; et al. Identification of microRNA-214 as a negative regulator of colorectal cancer liver metastasis by way of regulation of fibroblast growth factor receptor 1 expression. *Hepatology* **2014**, *60*, 598–609. [[CrossRef](#)] [[PubMed](#)]
87. Cristóbal, I.; Caramés, C.; Madoz-Gúrpide, J.; Rojo, F.; Aguilera, O.; García-Foncillas, J. Downregulation of miR-214 is specific of liver metastasis in colorectal cancer and could play a role determining the metastatic niche. *Int. J. Colorectal Dis.* **2014**, *29*, 885. [[CrossRef](#)] [[PubMed](#)]
88. Wang, J.; Li, J.; Wang, X.; Zheng, C.; Ma, W. Downregulation of microRNA-214 and overexpression of FGFR-1 contribute to hepatocellular carcinoma metastasis. *Biochem. Biophys. Res. Commun.* **2013**, *439*, 47–53. [[CrossRef](#)] [[PubMed](#)]
89. Kanaan, Z.; Rai, S.N.; Eichenberger, M.R.; Barnes, C.; Dworkin, A.M.; Weller, C.; Cohen, E.; Roberts, H.; Keskey, B.; Petras, R.E.; et al. Differential MicroRNA expression tracks neoplastic progression in inflammatory bowel disease-associated colorectal cancer. *Hum. Mutat.* **2012**, *33*, 551–560. [[CrossRef](#)]
90. Jinushi, T.; Shibayama, Y.; Kinoshita, I.; Oizumi, S.; Jinushi, M.; Aota, T.; Takahashi, T.; Horita, S.; Dosaka-Akita, H.; Iseki, K. Low expression levels of microRNA-124-5p correlated with poor prognosis in colorectal cancer via targeting of SMC4. *Cancer Med.* **2014**, *3*, 1544–1552. [[CrossRef](#)]
91. Taniguchi, K.; Sugito, N.; Kumazaki, M.; Shinohara, H.; Yamada, N.; Nakagawa, Y.; Ito, Y.; Otsuki, Y.; Uno, B.; Uchiyama, K.; et al. MicroRNA-124 inhibits cancer cell growth through PTB1/PKM1/PKM2 feedback cascade in colorectal cancer. *Cancer Lett.* **2015**, *363*, 17–27. [[CrossRef](#)]
92. Sun, Y.; Zhao, X.; Zhou, Y.; Hu, Y. miR-124, miR-137 and miR-340 regulate colorectal cancer growth via inhibition of the Warburg effect. *Oncol. Rep.* **2012**, *28*, 1346–1352. [[CrossRef](#)]
93. Taniguchi, K.; Sugito, N.; Kumazaki, M.; Shinohara, H.; Yamada, N.; Matsushashi, N.; Futamura, M.; Ito, Y.; Otsuki, Y.; Yoshida, K.; et al. Positive feedback of DDX6/c-Myc/PTB1 regulated by miR-124 contributes to maintenance of the Warburg effect in colon cancer cells. *Biochim. Biophys. Acta* **2015**, *1852*, 1971–1980. [[CrossRef](#)] [[PubMed](#)]
94. Warburg, O. On the Origin of Cancer Cells. *Science* **1956**, *123*, 309–314. [[CrossRef](#)]
95. Chen, M.; Zhang, J.; Manley, J.L. Turning on a Fuel Switch of Cancer: hnRNP Proteins Regulate Alternative Splicing of Pyruvate Kinase mRNA. *Cancer Res.* **2010**, *70*, 8977–8980. [[CrossRef](#)]
96. Christofk, H.R.; Vander Heiden, M.G.; Harris, M.H.; Ramanathan, A.; Gerszten, R.E.; Wei, R.; Fleming, M.D.; Schreiber, S.L.; Cantley, L.C. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **2008**, *452*, 230–233. [[CrossRef](#)]
97. Demaria, M.; Poli, V. PKM2, STAT3 and HIF-1 $\alpha$ . *JAK-STAT* **2012**, *1*, 194–196. [[CrossRef](#)]
98. Sun, Y.; Zhao, X.; Luo, M.; Zhou, Y.; Ren, W.; Wu, K.; Li, X.; Shen, J.; Hu, Y. The Pro-Apoptotic Role of the Regulatory Feedback Loop between miR-124 and PKM1/HNF4 $\alpha$  in Colorectal Cancer Cells. *Int. J. Mol. Sci.* **2014**, *15*, 4318–4332. [[CrossRef](#)]
99. Demaria, M.; Giorgi, C.; Lebedzinska, M.; Esposito, G.; D’Angeli, L.; Bartoli, A.; Gough, D.J.; Turkson, J.; Levy, D.E.; Watson, C.J.; et al. A STAT3-mediated metabolic switch is involved in tumour transformation and STAT3 addiction. *Aging* **2010**, *2*, 823–842. [[CrossRef](#)]
100. Marzec, M.; Liu, X.; Wong, W.; Yang, Y.; Pasha, T.; Kantekure, K.; Zhang, P.; Woetmann, A.; Cheng, M.; Odum, N.; et al. Oncogenic kinase NPM/ALK induces expression of HIF1 $\alpha$  mRNA. *Oncogene* **2011**, *30*, 1372–1378. [[CrossRef](#)] [[PubMed](#)]
101. Zhang, H.-F.; Lai, R. STAT3 in Cancer—Friend or Foe? *Cancers* **2014**, *6*, 1408–1440. [[CrossRef](#)] [[PubMed](#)]
102. Sugimoto, K. Role of STAT3 in inflammatory bowel disease. *World J. Gastroenterol.* **2008**, *14*, 5110–5114. [[CrossRef](#)]
103. Musso, A.; Dentelli, P.; Carlino, A.; Chiusa, L.; Repici, A.; Sturm, A.; Fiocchi, C.; Rizzetto, M.; Pegoraro, L.; Sategna-Guidetti, C.; et al. Signal transducers and activators of transcription 3 signaling pathway: An essential mediator of inflammatory bowel disease and other forms of intestinal inflammation. *Inflamm. Bowel Dis.* **2005**, *11*, 91–98. [[CrossRef](#)]
104. Atreya, R.; Neurath, M.F. Involvement of IL-6 in the pathogenesis of inflammatory bowel disease and colon cancer. *Clin. Rev. Allergy Immunol.* **2005**, *28*, 187–196. [[CrossRef](#)]
105. Zhang, J.; Lu, Y.; Yue, X.; Li, H.; Luo, X.; Wang, Y.; Wang, K.; Wan, J. MiR-124 suppresses growth of human colorectal cancer by inhibiting STAT3. *PLoS ONE* **2013**, *8*, e70300. [[CrossRef](#)] [[PubMed](#)]
106. Wang, X.; Ji, P.; Zhang, Y.; LaComb, J.F.; Tian, X.; Li, E.; Williams, J.L. Aberrant DNA Methylation: Implications in Racial Health Disparity. *PLoS ONE* **2016**, *11*, e0153125. [[CrossRef](#)]
107. Harada, T.; Yamamoto, E.; Yamano, H.; Nojima, M.; Maruyama, R.; Kumegawa, K.; Ashida, M.; Yoshikawa, K.; Kimura, T.; Harada, E.; et al. Analysis of DNA methylation in bowel lavage fluid for detection of colorectal cancer. *Cancer Prev. Res.* **2014**, *7*, 1002–1010. [[CrossRef](#)] [[PubMed](#)]
108. Zhang, L.; Dong, Y.; Zhu, N.; Tsoi, H.; Zhao, Z.; Wu, C.W.; Wang, K.; Zheng, S.; Ng, S.S.; Chan, F.K.; et al. microRNA-139-5p exerts tumor suppressor function by targeting NOTCH1 in colorectal cancer. *Mol. Cancer* **2014**, *13*, 124. [[CrossRef](#)] [[PubMed](#)]
109. Song, M.; Yin, Y.; Zhang, J.; Zhang, B.; Bian, Z.; Quan, C.; Zhou, L.; Hu, Y.; Wang, Q.; Ni, S.; et al. MiR-139-5p inhibits migration and invasion of colorectal cancer by downregulating AMFR and NOTCH1. *Protein Cell* **2014**, *5*, 851–861. [[CrossRef](#)]
110. Shen, K.; Liang, Q.; Xu, K.; Cui, D.; Jiang, L.; Yin, P.; Lu, Y.; Li, Q.; Liu, J. MiR-139 inhibits invasion and metastasis of colorectal cancer by targeting the type I insulin-like growth factor receptor. *Biochem. Pharmacol.* **2012**, *84*, 320–330. [[CrossRef](#)]

111. Liu, H.; Yin, Y.; Hu, Y.; Feng, Y.; Bian, Z.; Yao, S.; Li, M.; You, Q.; Huang, Z. miR-139-5p sensitizes colorectal cancer cells to 5-fluorouracil by targeting NOTCH-1. *Pathol.-Res. Pract.* **2016**, *212*, 643–649. [[CrossRef](#)]
112. Li, Q.; Liang, X.; Wang, Y.; Meng, X.; Xu, Y.; Cai, S.; Wang, Z.; Liu, J.; Cai, G. miR-139-5p Inhibits the Epithelial-Mesenchymal Transition and Enhances the Chemotherapeutic Sensitivity of Colorectal Cancer Cells by Downregulating BCL2. *Sci. Rep.* **2016**, *6*, 27157. [[CrossRef](#)]
113. Bian, Z.; Zhang, J.; Li, M.; Feng, Y.; Yao, S.; Song, M.; Qi, X.; Fei, B.; Yin, Y.; Hua, D.; et al. Long non-coding RNA LINC00152 promotes cell proliferation, metastasis, and confers 5-FU resistance in colorectal cancer by inhibiting miR-139-5p. *Oncogenesis* **2017**, *6*, 395. [[CrossRef](#)]
114. Mamoori, A.; Wahab, R.; Islam, F.; Lee, K.; Vider, J.; Lu, C.-T.; Gopalan, V.; Lam, A.K. Clinical and biological significance of miR-193a-3p targeted KRAS in colorectal cancer pathogenesis. *Hum. Pathol.* **2018**, *71*, 145–156. [[CrossRef](#)]
115. Lin, M.; Duan, B.; Hu, J.; Yu, H.; Sheng, H.; Gao, H.; Huang, J. Decreased expression of miR-193a-3p is associated with poor prognosis in colorectal cancer. *Oncol. Lett.* **2017**, *14*, 1061–1067. [[CrossRef](#)] [[PubMed](#)]
116. Takahashi, H.; Takahashi, M.; Ohnuma, S.; Unno, M.; Yoshino, Y.; Ouchi, K.; Takahashi, S.; Yamada, Y.; Shimodaira, H.; Ishioka, C. microRNA-193a-3p is specifically down-regulated and acts as a tumor suppressor in BRAF-mutated colorectal cancer. *BMC Cancer* **2017**, *17*, 723. [[CrossRef](#)] [[PubMed](#)]
117. Ling, H.; Fabbri, M.; Calin, G.A. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat. Rev. Drug Discov.* **2013**, *12*, 847–865. [[CrossRef](#)] [[PubMed](#)]
118. Baumann, V.; Winkler, J. miRNA-based therapies: Strategies and delivery platforms for oligonucleotide and non-oligonucleotide agents. *Future Med. Chem.* **2014**, *6*, 1967–1984. [[CrossRef](#)]
119. Chapman, C.G.; Pekow, J. The emerging role of miRNAs in inflammatory bowel disease: A review. *Therap. Adv. Gastroenterol.* **2015**, *8*, 4–22. [[CrossRef](#)] [[PubMed](#)]
120. Lu, Z.-J.; Wu, J.-J.; Jiang, W.-L.; Xiao, J.-H.; Tao, K.-Z.; Ma, L.; Zheng, P.; Wan, R.; Wang, X.-P. MicroRNA-155 promotes the pathogenesis of experimental colitis by repressing SHIP-1 expression. *World J. Gastroenterol.* **2017**, *23*, 976–985. [[CrossRef](#)]
121. Selbach, M.; Schwanhäusser, B.; Thierfelder, N.; Fang, Z.; Khanin, R.; Rajewsky, N. Widespread changes in protein synthesis induced by microRNAs. *Nature* **2008**, *455*, 58–63. [[CrossRef](#)]
122. López-Urrutia, E.; Bustamante Montes, L.P.; Ladrón de Guevara Cervantes, D.; Pérez-Plasencia, C.; Campos-Parra, A.D. Crosstalk Between Long Non-coding RNAs, Micro-RNAs and mRNAs: Deciphering Molecular Mechanisms of Master Regulators in Cancer. *Front. Oncol.* **2019**, *9*. [[CrossRef](#)] [[PubMed](#)]
123. Chakraborty, C.; Wen, Z.-H.; Agoramoorthy, G.; Lin, C.-S. Therapeutic microRNA Delivery Strategies with Special Emphasis on Cancer Therapy and Tumorigenesis: Current Trends and Future Challenges. *Curr. Drug Metab.* **2016**, *17*, 469–477. [[CrossRef](#)]
124. Tenore, G.C.; Ritieni, A.; Campiglia, P.; Stiuso, P.; Di Maro, S.; Sommella, E.; Pepe, G.; D’Urso, E.; Novellino, E. Antioxidant peptides from “Mozzarella di Bufala Campana DOP” after simulated gastrointestinal digestion: In vitro intestinal protection, bioavailability, and anti-haemolytic capacity. *J. Funct. Foods* **2015**, *15*, 365–375. [[CrossRef](#)]
125. Nielsen, O.H.; Ainsworth, M.A. Tumor necrosis factor inhibitors for inflammatory bowel disease. *N. Engl. J. Med.* **2013**, *369*, 754–762. [[CrossRef](#)]
126. Ben-Horin, S.; Chowers, Y. Review article: Loss of response to anti-TNF treatments in Crohn’s disease. *Aliment. Pharmacol. Ther.* **2011**, *33*, 987–995. [[CrossRef](#)] [[PubMed](#)]
127. Morilla, I.; Uzzan, M.; Laharie, D.; Cazals-Hatem, D.; Denost, Q.; Daniel, F.; Belleanne, G.; Bouhnik, Y.; Wainrib, G.; Panis, Y.; et al. Colonic MicroRNA Profiles, Identified by a Deep Learning Algorithm, That Predict Responses to Therapy of Patients With Acute Severe Ulcerative Colitis. *Clin. Gastroenterol. Hepatol.* **2019**, *17*, 905–913. [[CrossRef](#)] [[PubMed](#)]
128. Wu, C.W.; Ng, S.C.; Dong, Y.; Tian, L.; Ng, S.S.M.; Leung, W.W.; Law, W.T.; Yau, T.O.; Chan, F.K.L.; Sung, J.J.Y.; et al. Identification of microRNA-135b in Stool as a Potential Noninvasive Biomarker for Colorectal Cancer and Adenoma. *Clin. Cancer Res.* **2014**, *20*, 2994–3002. [[CrossRef](#)] [[PubMed](#)]
129. Yau, T.O.; Wu, C.W.; Tang, C.-M.; Chen, Y.; Fang, J.; Dong, Y.; Liang, Q.; Ng, S.S.M.; Chan, F.K.L.; Sung, J.J.Y.; et al. microRNA-20a in human faeces as a non-invasive biomarker for colorectal cancer. *Oncotarget* **2016**, *7*, 1559–1568. [[CrossRef](#)] [[PubMed](#)]
130. Tepus, M.; Yau, T.O. Non-Invasive Colorectal Cancer Screening: An Overview. *Gastrointest. Tumors* **2020**, *3–4*, 62–73. [[CrossRef](#)]
131. Verdier, J.; Breunig, I.R.; Ohse, M.C.; Roubrocks, S.; Kleinfeld, S.; Roy, S.; Streetz, K.; Trautwein, C.; Roderburg, C.; Sellge, G. Faecal Micro-RNAs in Inflammatory Bowel Diseases. *J. Crohns. Colitis* **2020**, *14*, 110–117. [[CrossRef](#)]
132. Ji, Y.; Li, X.; Zhu, Y.; Li, N.; Zhang, N.; Niu, M. Faecal microRNA as a biomarker of the activity and prognosis of inflammatory bowel diseases. *Biochem. Biophys. Res. Commun.* **2018**, *503*, 2443–2450. [[CrossRef](#)]
133. Monaghan, T.M.; Seekatz, A.M.; Markham, N.O.; Yau, T.O.; Hatzia Apostolou, M.; Jilani, T.; Christodoulou, N.; Roach, B.; Birli, E.; Pomenya, O.; et al. Fecal microbiota transplantation for recurrent *Clostridioides difficile* infection associates with functional alterations in circulating microRNAs. *Gastroenterology* **2021**. [[CrossRef](#)]
134. Schönauen, K.; Le, N.; von Arnim, U.; Schulz, C.; Malfertheiner, P.; Link, A. Circulating and Fecal microRNAs as Biomarkers for Inflammatory Bowel Diseases. *Inflamm. Bowel Dis.* **2018**, *24*, 1547–1557. [[CrossRef](#)] [[PubMed](#)]



# Epigenetic dysregulation in Epstein-Barr virus-associated gastric carcinoma: Disease and treatments

Tung On Yau, Ceen-Ming Tang, Jun Yu

*World Journal of Gastroenterology*. 2014;20(21):6448.

Journal URL: [wjnet.com/1007-9327/full/v20/i21/6448.htm](http://wjnet.com/1007-9327/full/v20/i21/6448.htm)

DOI: [10.3748/wjg.v20.i21.6448](https://doi.org/10.3748/wjg.v20.i21.6448)

PMID: [24914366](https://pubmed.ncbi.nlm.nih.gov/24914366/)

WJG 20<sup>th</sup> Anniversary Special Issues (8): Gastric cancer**Epigenetic dysregulation in Epstein-Barr virus-associated gastric carcinoma: Disease and treatments**

Tung On Yau, Ceen-Ming Tang, Jun Yu

Tung On Yau, Ceen-Ming Tang, Jun Yu, Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, LKS Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China

Ceen-Ming Tang, Department of Pharmacology, University of Oxford, Oxford OX1 3QT, United Kingdom

Author contributions: Yau TO wrote this paper; Tang CM revised the article; Yu J supervised the work.

Supported by Research Grants of National Basic Research Program of China (973 Program, 2010CB529305); and Innovation and Technology Support Programme, Hong Kong (ITS/214/12)

Correspondence to: Jun Yu, MD, PhD, Professor, Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, LKS Institute of Health Sciences, The Chinese University of Hong Kong, Room 707A, Li Ka Shing Medical Sciences Building, Prince of Wales Hospital, Shatin, Hong Kong, China. [junyu@cuhk.edu.hk](mailto:junyu@cuhk.edu.hk)  
Telephone: +852-37636099 Fax: +852-21445330

Received: November 1, 2013 Revised: February 16, 2014

Accepted: March 12, 2014

Published online: June 7, 2014

© 2014 Baishideng Publishing Group Inc. All rights reserved.

**Key words:** Epstein-Barr virus; Gastric carcinoma; Epigenetic dysregulation; Aberrant DNA methylation; Epigenetic therapies

**Core tip:** Epstein-Barr virus (EBV)-associated gastric carcinoma (EBVaGC) comprises nearly 10% of gastric carcinoma cases worldwide. In the present review, we critically discuss the role of EBV in gastric carcinogenesis, summarising the role of viral proteins and microRNAs with respect to aberrant methylation in EBVaGC. Given the role of epigenetic dysregulation in tumorigenesis, epigenetic modifiers may represent a novel therapeutic strategy.

Yau TO, Tang CM, Yu J. Epigenetic dysregulation in Epstein-Barr virus-associated gastric carcinoma: Disease and treatments. *World J Gastroenterol* 2014; 20(21): 6448-6456 Available from: <http://www.wjgnet.com/1007-9327/full/v20/i21/6448.htm>  
DOI: <http://dx.doi.org/10.3748/wjg.v20.i21.6448>

**Abstract**

Epstein-Barr virus (EBV)-associated gastric carcinoma (EBVaGC) comprises nearly 10% of gastric carcinoma cases worldwide. Recently, it was recognised to have unique clinicopathologic characteristics, including male predominance, lower rates of lymph node involvement, and better prognosis. EBVaGC is further characterised by abnormal hypermethylation of tumour suppressor gene promoter regions, causing down-regulation of their expression. In the present review, we critically discuss the role of EBV in gastric carcinogenesis, summarising the role of viral proteins and microRNAs with respect to aberrant methylation in EBVaGC. Given the role of epigenetic dysregulation in tumorigenesis, epigenetic modifiers may represent a novel therapeutic strategy.

**INTRODUCTION**

Epstein-Barr virus (EBV) infection is ubiquitous, and is accepted as a causative microorganism for various malignancies including nasopharyngeal carcinoma (NPC), Burkitt's lymphoma, and gastric carcinoma (GC). EBV-associated GC (EBVaGC) accounts for approximately 10% of cases worldwide<sup>[1,2]</sup>, and is characterised by unique clinicopathologic features including a relatively favourable prognosis (Table 1)<sup>[1-4]</sup>. In recent years, the molecular mechanisms underlying EBV-related carcinogenesis have become increasingly understood. EBV may contribute to tumorigenesis through the expression of viral proteins and microRNAs (miRNAs). Previous studies have also reported that promoter methylation was observed more

**Table 1 Clinical and pathological features of Epstein-Barr virus-associated gastric carcinoma**

Clinical and pathological features	
Age	Younger <sup>1</sup>
Gender	Male predominance
Associations	Smoking
Prevalence	10% of gastric carcinoma cases
Location	Gastric body/cardia Remnant stomach
Clinical	Multiple carcinomas <sup>1</sup> Thickening of gastric wall Ulcerated (saucer-like) neoplasm Lower rate of lymph node involvement <sup>1</sup>
Histology	Lymphoepithelioma-like Lymphocytic infiltration in various degrees Atrophic gastritis Lace pattern within the mucosa Moderate to poorly differentiated adenocarcinoma
Prognosis	Longer survival <sup>1</sup>

<sup>1</sup>Items are controversial and subject to on-going research.

frequently in EBVaGC. Hence another method by which EBV contributes to gastric carcinogenesis is through aberrant DNA methylation and histone modification. Thus EBVaGC is characterised by distinct variations on genomic, epigenomic, and transcriptomic levels<sup>[5]</sup>. Here, we review the mechanism by which EBV infection causes aberrant methylation, transformation, cancer development, and its associated therapeutic implications.

### MECHANISM OF EBV INFECTION

EBV may infect host gastric epithelial cells directly or indirectly (Figure 1). In direct infection, the viral envelope glycoprotein BMRF-2 interacts with cellular  $\beta 1$  integrins. Subsequently, viral protein gH/gL attaches to cellular  $\alpha v\beta 6/8$  integrins, and triggers fusion of the viral envelope with the epithelial cell membrane<sup>[6]</sup>. EBV preferentially infects B lymphocytes, which then mediates subsequent infection to epithelial cells<sup>[7]</sup>. In B cell invasion, EBV envelope glycoproteins gp350/220 bind to B cell receptors CD21 and/or CD35<sup>[8,9]</sup>. Simultaneously, viral glycoprotein gp42 interacts with Human Leukocyte Antigen (HLA) class II molecules on the B cell membrane to trigger the core fusion complex, enabling EBV entry into the B cell (Figure 2)<sup>[8,10]</sup>. Through direct cell-to-cell contact, EBV-infected B cells may subsequently infect epithelial cells<sup>[11]</sup>. The exact mechanism of epithelial cell invasion is unclear, but involves CD21-mediated co-capping of EBV and integrins on B cells, as well as conjugate formation between EBV-infected B cells and epithelial cells *via* the capped adhesion molecules<sup>[11]</sup>. Once EBV enters epithelial cells, the viral capsid dissolves and the viral genome is transported to the cell nucleus.

### LATENCY, VIRAL PROTEINS, AND CARCINOGENESIS

Following infection, EBV typically persists in a latent

**Table 2 Latent gene expression patterns in Epstein-Barr virus infected malignancies**

Genes	Latency I a	Latency I b	Latency II	Latency III
<i>EBNA1</i>	+	+	+	+
<i>EBNA2</i>	-	-	-	+
<i>EBNA3a</i>	-	-	-	+
<i>EBNA3b</i>	-	-	-	+
<i>EBNA3c</i>	-	-	-	+
<i>EBNA-LP</i>	-	-	+	+
<i>LMP1</i>	-	-	+	+
<i>LMP2A</i>	-	+	+	+
<i>LMP2B</i>	-	-	+	+
<i>EBER1</i>	+	+	+	+
<i>EBER2</i>	+	+	+	+
<i>BARTs</i>	+	+	+	+
Disease	EBVaGC, Burkitt's lymphoma	EBVaGC	NPC, Hodgkin's lymphoma, NK/T-cell lymphoma	AIDS- associated B-cell lymphomas, Pyothorax- associated lymphoma

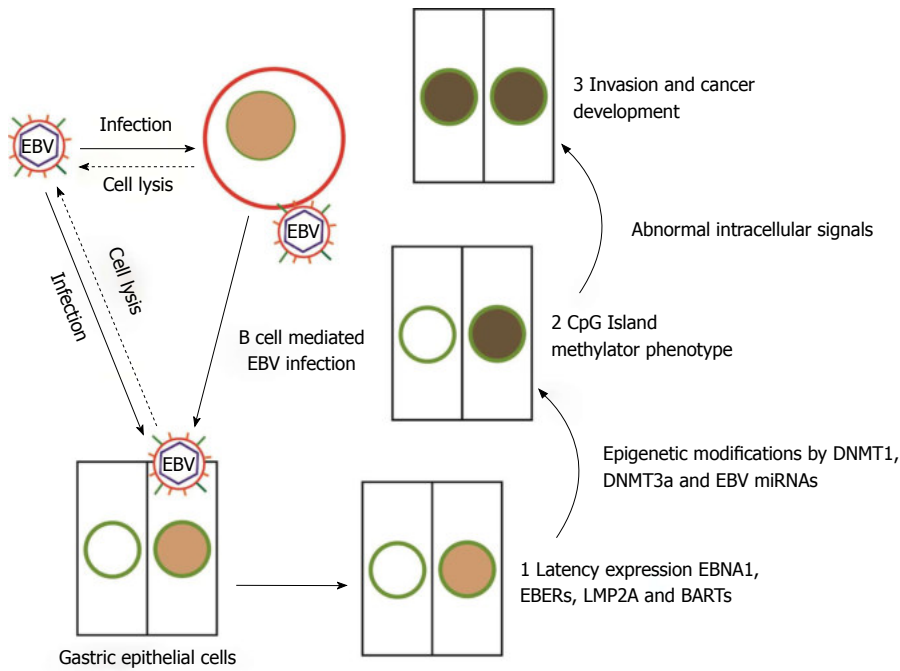
EBVaGC: Epstein-Barr virus-associated gastric carcinoma; NPC: Nasopharyngeal carcinoma; AIDS: Acquired-immunodeficiency-syndrome.

stage. During latency, the viral genome is largely silenced by host-driven methylation of CpG island motifs. Based on the subset of viral genes which are expressed, tumours may be classified into four types; latency I a, I b, II, and III (Table 2). EBVaGC belongs to latency type I, where the viral genes EBV nuclear antigen 1 (EBNA1), EBV-encoded small RNA (EBER1/2), BamHI-A rightward transcripts (BARTs), and latent membrane protein 2A (LMP2A) may be expressed<sup>[12,13]</sup>. Notably, the expression of latency genes is associated with malignancy. For example, EBER1 up-regulates the expression of insulin-growth factor-1, an autocrine growth factor which accelerates cell proliferation in EBVaGC<sup>[14]</sup>.

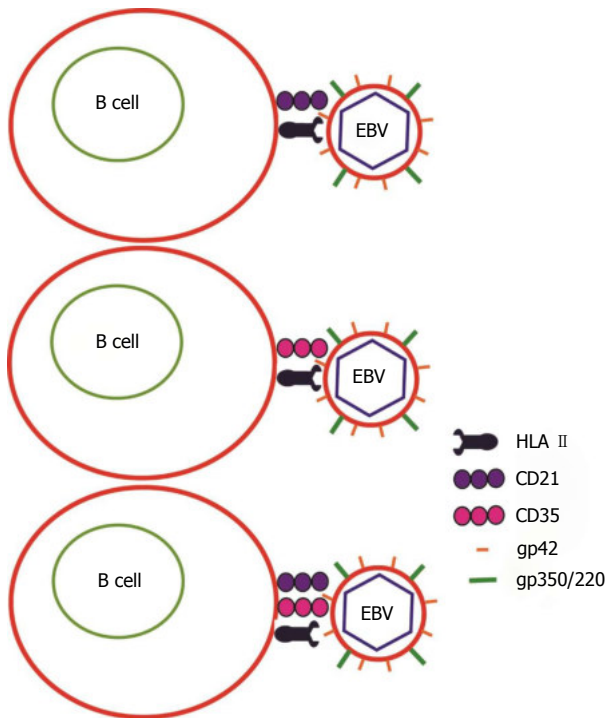
Half of all EBVaGCs also express LMP2A. LMP2A plays a critical role in the oncogenic processes in EBVaGC, and thus EBV latency patterns should be further subdivided into Ia or Ib based on the absence or presence of LMP2A<sup>[12,15]</sup>. LMP2A not only inhibits apoptosis through up-regulation of the cellular survivin gene *via* the NF- $\kappa$ B pathway<sup>[16]</sup>, but induces expression of phosphorylated signal transducer and activator of transcription 3 (pSTAT3), which causes up-regulation of DNA methyltransferase DNMT1<sup>[16]</sup> and DNMT3B<sup>[17]</sup> in EBV-infected GC cells. DNA methyltransferases play important roles in controlling DNA methylation. The subsequent overdrive of CpG methylation and silencing of tumour suppressor genes such as PTEN, p16, and p73 leads to the transformation of EBV-infected cells. Hence epigenetic dysregulation plays an important role in gastric carcinogenesis.

### EBVaGC AND EPIGENETIC ALTERATIONS

Epigenetics refer to functionally relevant and heritable changes in gene expression that occur without alteration of the underlying DNA sequence. The two primary mechanisms which may produce this change are DNA



**Figure 1 Epstein-Barr virus infects host gastric epithelial cells through direct and indirect mechanisms.** Epstein-Barr virus (EBV) preferentially infects B lymphocytes, which subsequently infects gastric epithelial cells through direct cell-to-cell contact. EBV infection causes expression of latency 1a and/or 1b proteins in gastric epithelial cells. EBV infection also up-regulates genes including *EBNA1*, *EBER*, *LMP2A*, and *BART*, altering expression of DNMTs and miRNAs. Collectively, the abnormal intracellular signals lead to carcinogenesis and tumour development.



**Figure 2 Mechanism of Epstein-Barr virus infection in B lymphocytes.** Epstein-Barr virus (EBV) envelope glycoproteins gp350/220 bind to B lymphocyte receptors CD21 and/or CD35. Simultaneously, viral glycoprotein gp42 interacts with HLA II on the B cell membrane to trigger the core fusion complex, enabling EBV entry into the B cell.

methylation, and histone modification. According to the epigenetic progenitor model, tumour-progenitor genes promote the polyclonal epigenetic disruption of stem

cells as a first step in the development of cancer<sup>[18]</sup>. This epigenetic plasticity causes genomic instability, and collectively drives tumour progression<sup>[19]</sup>.

The CpG island methylator phenotype was first observed in EBVaGC in 1999<sup>[20]</sup>. EBV infection was shown to induce extensive methylation and repression of tumour suppressor genes over 18 wk in MKN7, a low methylation GC cell line<sup>[21]</sup>. Subsequent studies confirmed that EBVaGC has higher rates of aberrant DNA methylation than EBV non-associated GC (EBVnGC)<sup>[21,22]</sup>. Nevertheless, the mechanisms by which EBV induces aberrant DNA methylation and histone modification remain poorly understood.

### EBV AND microRNA

Viral encoded miRNAs play a pivotal role in alterations to DNA methylation status in host cells. The expression of EBV miRNAs vary under different latency programs (Table 3)<sup>[23]</sup>. For example, miR-BART1-5p, 6, and 17-5p suppresses LMP1 expression<sup>[24]</sup>, whilst miR-BART-22 regulates expression of LMP2A<sup>[25]</sup>. EBV miRNAs further repress cellular proteins, including PUMA, DICER1, and BIM. EBV infection may also affect host cell miRNA expression. Specifically, miR-200a and miR-200b are down-regulated in EBVaGC compared to EBVnGC and adjacent mucosa. This down-regulation may be mediated by viral proteins such as BRAF0, EBV, and LMP2A, as well as by aberrant DNA methylation following EBV infection<sup>[26]</sup>. More recently, miRNA sequencing studies have revealed that EBV-infection mediates down-regulation of tumour suppressor miRNAs including the Let-7 family.

**Table 3 Epstein-Barr virus-driven miRNAs and their target genes**

Gene name	Gene targets in EBV	Gene targets in host cell
miR-BHRF1-1	-	<i>GUF1</i> <sup>[58]</sup> , <i>SCRN1</i> <sup>[58]</sup>
miR-BART1-5p	<i>LMP1</i> <sup>[24]</sup>	<i>CLEC2D</i> <sup>[58,59]</sup> , <i>LY75</i> <sup>[58,59]</sup> , <i>SP100</i> <sup>[58,59]</sup> , <i>DICER1</i> <sup>[58,59]</sup> , <i>MICB</i> <sup>[58,59]</sup>
miR-BART1-3	-	<i>CXCL11</i> <sup>[60]</sup>
miR-BART2-5p	<i>BALF5</i> <sup>[61]</sup>	<i>MICB</i> <sup>[62]</sup>
miR-BART3	-	<i>DICER1</i> <sup>[58]</sup> , <i>MICB</i> <sup>[58]</sup>
miR-BART3-3p	-	<i>IPO7</i> <sup>[63]</sup>
miR-BART5	<i>LMP1</i> <sup>[59]</sup>	<i>PUMA</i> <sup>[66]</sup>
miR-BART6	<i>LMP1</i> <sup>[24]</sup>	<i>DICER1</i> <sup>[65]</sup>
miR-BART10	<i>BHRF1</i> <sup>[59]</sup>	-
miR-BART13	-	<i>CAPRN2</i> <sup>[59]</sup>
miR-BART16	<i>LMP1</i> <sup>[24]</sup>	<i>TOMM22</i> <sup>[63]</sup>
miR-BART17-p	<i>LMP1</i> <sup>[24]</sup>	-
miR-BART19	<i>LMP1</i> <sup>[59]</sup>	-
miR-BART22	<i>LMP2A</i> <sup>[25]</sup>	-
miR-BARTs	-	<i>BIM</i> <sup>[66]</sup>

EBV: Epstein-Barr virus.

Further research is required to elucidate their role in tumorigenesis<sup>[27]</sup>.

### ABERRANT DNA METHYLATION IN EBVaGC

Currently, GC is subdivided into three subtypes based on CpG-island methylator phenotype (CIMP). Defined as high (CIMP-H), low (CIMP-L), or none (CIMP-N), the classification is based on the number of methylated loci ( $\geq 4$ , 1-3, and 0 respectively) in the promoter regions of five genes (LOX, HRASLS, FLNC, HAND1, and THBD)<sup>[28]</sup>. It was previously shown that promoter methylation of cancer-related genes was seen more frequently in EBVaGC than EBVnGC. EBVaGC is thus classified as CIMP-H<sup>[29]</sup>.

In a genome-wide study comparing promoter methylation between EBV-infected and EBV non-infected GC cell lines, hundreds of genes involved in cancer pathways such as cell adhesion molecules, wnt signalling pathway, and mitogen-activated protein kinase signalling were observed to be hypermethylated following EBV infection<sup>[17]</sup>. Further investigation through epigenomic and transcriptomic sequencing revealed that 216 genes were down-regulated by promoter hypermethylation. Significantly, hypermethylation of tumour suppressor genes, including p14, p15, p16, APC, E-cadherin, and PTEN were noted in EBVaGC, but not EBVnGC<sup>[30,31]</sup>. All studies unanimously agreed that p16 was significantly more hypermethylated in EBVaGC<sup>[29-35]</sup>. P16 is a tumour suppressor gene which acts in the G1 phase of the cell cycle to phosphorylate the retinoblastoma gene product (pRb). Loss of p16 leads to uncontrolled cell growth<sup>[36]</sup>, and is thus commonly found in tumours<sup>[37,38]</sup>.

Another important cellular abnormality in EBVaGC is its resistance to apoptosis. The frequency of apoptosis is significantly lower in EBVaGC than in EBVnGC<sup>[38]</sup>.

**Table 4 Hypermethylated genes verified in Epstein-Barr virus-associated gastric carcinoma tissue**

Function	Hypermethylated genes
Apoptosis	<i>DAPK</i> <sup>[30]</sup> , <i>BNIP3</i> <sup>[29]</sup> , <i>FAM3B</i> <sup>[5]</sup> , <i>HRK</i> <sup>[29]</sup> , <i>IL15RA</i> <sup>[17]</sup> , <i>MINT31</i> <sup>[32]</sup> , <i>p16</i> <sup>[29-35]</sup> , <i>p73</i> <sup>[30,32-34]</sup> , <i>PTEN</i> <sup>[31,67]</sup> , <i>RASS-FIA</i> <sup>[31]</sup>
Cell adhesion	<i>EPHB6</i> <sup>[17]</sup> , <i>FLNC</i> <sup>[30]</sup> , <i>FSD1</i> <sup>[34]</sup> , <i>REC8</i> <sup>[17]</sup> , <i>CSPG2</i> <sup>[29]</sup>
Cell-cell interactions	<i>MDGA2</i> <sup>[17]</sup> , <i>THBS1</i> <sup>[31]</sup>
Cell cycle regulation	<i>APC</i> <sup>[31]</sup> , <i>p15</i> <sup>[30]</sup> , <i>p16</i> <sup>[29-35]</sup> , <i>p57</i> <sup>[29]</sup> , <i>p73</i> <sup>[30,32-34]</sup>
Cell invasion	<i>E-Cadherin</i> <sup>[30,68,69]</sup>
Cell migration	<i>EPHB6</i> <sup>[17]</sup>
Cell proliferation	<i>E-Cadherin</i> <sup>[30,68,69]</sup> , <i>HRASLS</i> <sup>[30]</sup> , <i>IL15RA</i> <sup>[17]</sup> , <i>MINT31</i> <sup>[32]</sup> , <i>NKX3.1</i> <sup>[34]</sup> , <i>RUNX3</i> <sup>[32]</sup> , <i>TIMP2</i> <sup>[21]</sup> , <i>TIMP3</i> <sup>[30]</sup>
Cell signalling	<i>14-3-3Sigma</i> <sup>[31]</sup> , <i>CSPG2</i> <sup>[29]</sup> , <i>MINT1</i> <sup>[31]</sup> , <i>MINT2</i> <sup>[31,32]</sup> , <i>PLXND1</i> <sup>[21]</sup>
Differentiation	<i>HAND1</i> <sup>[30]</sup>
Dna repair	<i>hMLH1</i> <sup>[32,43,53]</sup> , <i>MGMT</i> <sup>[31]</sup>
Exocytosis	<i>SCRN1</i> <sup>[34]</sup>
Metastasis	<i>E-Cadherin</i> <sup>[30,68,69]</sup> , <i>LOX</i> <sup>[30]</sup>
Other	<i>BCL7A</i> <sup>[34]</sup> , <i>BLU</i> <sup>[34]</sup> , <i>CHFR</i> <sup>[29]</sup> , <i>CXXC4</i> <sup>[21]</sup> , <i>GSTP1</i> <sup>[30,31,40]</sup> , <i>HLTF</i> <sup>[29]</sup> , <i>HOXA10</i> <sup>[70]</sup> , <i>IHH</i> <sup>[5]</sup> , <i>MARK1</i> <sup>[34]</sup> , <i>MINT25</i> <sup>[31]</sup> , <i>PAX5-β</i> <sup>[29]</sup> , <i>SCARF2</i> <sup>[17]</sup> , <i>SSTR1</i> <sup>[17,39]</sup> , <i>THBD</i> <sup>[30]</sup> , <i>WNT5A</i> <sup>[71]</sup>

It may be caused by hypermethylation of SSTR1 and GSTP1; both genes are frequently hypermethylated in NPC and GC infected EBV tissues, and regulate cell migration, proliferation, and apoptosis<sup>[30,31,39-42]</sup>. Notably, EBV infection also up-regulates expression of FAM3B and IHH<sup>[5]</sup>. FAM3B is associated with invasion<sup>[43]</sup>, and Indian Hedgehog (IHH) with increased metastatic potential through angiogenesis and Snail protein expression, as well as a decrease in e-cadherin and tight junctions<sup>[44,45]</sup>. Table 4 shows a comprehensive list of hypermethylated genes and their role in carcinogenesis. Hence aberrant DNA methylation plays an important role in gastric carcinogenesis.

### IMPLICATIONS FOR TREATMENT

Current treatment guidelines from the National Institute for Health and Clinical Excellence (NICE) for the management of GC depends on the stage of disease. Broadly, the mainstay for cure is surgical excision with clearance of adjacent lymph nodes. Radiotherapy, and chemotherapeutic agents including cisplatin, docetaxel, epirubicin, and 5-fluorouracil (5-FU) may be used as adjuvants or in the palliative setting. Notably, no differentiation is made between the distinct subtypes of GC in the treatment guidelines.

Research has established that EBVaGC represents a distinct entity of GC, characterised not only by unique genomic aberrations, but also by clinicopathologic features such as less lymph node involvement, and significantly better prognosis<sup>[2]</sup>. Naturally, there are associated therapeutic implications, as evidenced by resistance to docetaxel and 5-FU in EBV-positive, but not EBVnGC cell lines<sup>[46,47]</sup>. The chemoresistance is mediated by EBV-lytic gene expression, which induces expression of Bcl-2

**Table 5 Selection of epigenetic therapeutics in cancer chemotherapy**

Target	Drug name	Status	Ref.	
DNA methylation	Azacitine (5-Aza-CR)	Approved	[72]	
	Decitabine (5-Aza-CdR)	Approved	[72]	
	Hydralazine	Phase II/III	[73]	
	Epigallocatechin-3-gallate (EGCG)	Phase II	[74-76]	
	5-Fluoro-deoxycytidine (FdCyd/FDAC)	Phase I/II	[77]	
	5-fluoro-2'-deoxycytidine (FCdR)	Phase I/II	[78]	
	Procainamide	Phase I	[79]	
	Procaine	Phase I	[80]	
	Psammaplin A	Phase 0	[81,82]	
	RG108	Phase 0	[83-86]	
	Zebularine	Phase 0	[87-89]	
	Histone deacetylases	Vorinostat	Approved	[90]
		Romidepsin	Approved	[90]
Panobinostat		Phase II	[91]	
SEN196		Phase II	[92]	
Phenyl butyrate		Phase I/II	[93-95]	
Valporic acid		Phase I	[93-95]	
Compound 6j (R = -C <sub>4</sub> H <sub>9</sub> )		Phase 0	[96]	

and survivin whilst simultaneously suppressing p21 to inhibit apoptosis<sup>[48]</sup>. In support of this hypothesis, silencing of EBV-lytic gene LMP1 through specific small interfering RNA (siRNA) enhanced chemosensitivity of cancer cells to bleomycin and cisplatin<sup>[49]</sup>. Since epigenetic dysregulation is implicated in the expression of EBV-lytic genes and consequent tumour progression, we believe that epigenetic processes are a rational therapeutic target in EBVaGC.

Crucially, aberrant DNA methylation in cancer is reversible. Thus the enzymes which regulate epigenetic modifications are attractive targets for pharmacological intervention. Current epigenetic therapies may be classified into histone acetyltransferase (HAT), histone deacetylase (HDAC), and DNA methyltransferase (DNMT) inhibitors (Table 5). Broadly, they facilitate demethylation and re-expression of epigenetically silenced tumour suppressor genes, lowering the apoptotic threshold to sensitise tumour cells to chemotherapy and radiotherapy. Consequently, there has been an emphasis on investigating the clinical value of epigenetic therapies in combination with conventional cytotoxic agents and radiation.

It was previously reported that the combination of irradiation and 5-aza-CdR significantly decreased growth activity compared with irradiation alone in OCUM-2M, OCUM-12, and MKN-45 GC cell lines ( $P < 0.05$ ). The cell cycle arrest and increased apoptotic rate may be partly mediated by enhanced expression of p53, RASSF1, and DAPK gene families by 5-aza-CdR<sup>[50]</sup>. The use of epigenetic therapies in conjunction with targeted therapies such as gefitinib in lung cancer, imatinib in chronic myeloid leukemia, and trastuzumab in breast cancer cell lines and *in vivo* tumour models also had synergistic effects on the induction of apoptosis<sup>[51,52]</sup>. More recently,

epigenetic modifiers and ZEB1 inhibitors have been used to induce lytic transformation of EBV-infected gastric cancer cells. Expressed only in the lytic form of infection, virally encoded kinases convert ganciclovir into its active form, potentiating its cytotoxic effects<sup>[53,54]</sup>. Hence epigenetic modifiers may be a useful therapeutic strategy in EBVaGC.

However, several problems must be considered. Firstly, methylation is reversible, so re-methylation and re-silencing after cessation of drug therapy may occur<sup>[55]</sup>. Moreover, there have been numerous concerns raised regarding the systemic effects of non-specific gene activation in non-cancerous cells by epigenetic therapies. Conflicting evidence exists in the literature regarding the effect of epigenetic therapies on normal cells. Some studies have demonstrated that 5-Aza and decitabine increases mutation frequency, causes chromosomal rearrangements, and decreases fertility in mice. Conversely, no increase in chromosomal integrity was observed following administration of low dose 5-aza-CdR in patients with myelodysplastic syndrome<sup>[56]</sup>. Additionally, treatment of 41 leukemia patients with 5-aza-CdR showed only mild effects on global genomic de-methylation, as measured by changes in Alu methylation<sup>[57]</sup>. Few adverse effects were observed, and original methylation levels were regained within two weeks after therapy. No development of secondary malignancies were recorded. Consequently, further studies are needed to investigate the long term effects of epigenetic therapies.

## CONCLUSION

EBVaGC is a unique type of GC. The characteristic global hypermethylation of the promoter region in tumour-suppressor genes may be due to overexpression of DNMTs by viral latent proteins, miRNAs, and various epigenomic changes. However, the precise role of EBV in the multifactorial etiology of GC is still not fully understood. Further studies are needed to elucidate the intricate relationship between EBV infection, environmental factors, genetic backgrounds, and aberrant DNA methylation in GC. A better understanding of the role of EBV in gastric carcinogenesis will enable discovery of novel therapeutic targets and strategies.

## REFERENCES

- 1 **Fukayama M**, Hino R, Uozaki H. Epstein-Barr virus and gastric carcinoma: virus-host interactions leading to carcinoma. *Cancer Sci* 2008; **99**: 1726-1733 [PMID: 18616681 DOI: 10.1111/j.1349-7006.2008.00888.x]
- 2 **van Beek J**, zur Hausen A, Klein Kranenbarg E, van de Velde CJ, Middeldorp JM, van den Brule AJ, Meijer CJ, Bloemena E. EBV-positive gastric adenocarcinomas: a distinct clinicopathologic entity with a low frequency of lymph node involvement. *J Clin Oncol* 2004; **22**: 664-670 [PMID: 14966089 DOI: 10.1200/JCO.2004.08.061]
- 3 **Camargo MC**, Kim WH, Chiaravalli AM, Kim KM, Corvalan AH, Matsuo K, Yu J, Sung JJ, Herrera-Goepfert R, Meneses-Gonzalez F, Kijima Y, Natsugoe S, Liao LM, Lissowska J, Kim S, Hu N, Gonzalez CA, Yatabe Y, Koriyama C, Hewitt

- SM, Akiba S, Gulley ML, Taylor PR, Rabkin CS. Improved survival of gastric cancer with tumour Epstein-Barr virus positivity: an international pooled analysis. *Gut* 2014; **63**: 236-243 [PMID: 23580779 DOI: 10.1136/gutjnl-2013-304531]
- 4 **Camargo MC**, Koriyama C, Matsuo K, Kim WH, Herrera-Goepfert R, Liao LM, Yu J, Carrasquilla G, Sung JJ, Alvarado-Cabrero I, Lissowska J, Meneses-Gonzalez F, Yatabe Y, Ding T, Hu N, Taylor PR, Morgan DR, Gulley ML, Torres J, Akiba S, Rabkin CS. Case-case comparison of smoking and alcohol risk associations with Epstein-Barr virus-positive gastric cancer. *Int J Cancer* 2014; **134**: 948-953 [PMID: 23904115 DOI: 10.1002/ijc.28402]
  - 5 **Liang Q**, Yao X, Tang S, Yau TO, Zhao J, Sung JJ, Yu J. Integrative Identification of EBV-Associated Variations At Genomic, Epigenomic and Transcriptomic Levels in Gastric Cancer. *Gastroenterology* 2013; **144**: S525 [DOI: 10.1016/S0016-5085(13)63525-8]
  - 6 **Tugizov SM**, Berline JW, Palefsky JM. Epstein-Barr virus infection of polarized tongue and nasopharyngeal epithelial cells. *Nat Med* 2003; **9**: 307-314 [PMID: 12592401 DOI: 10.1038/nm830]
  - 7 **Sixbey JW**, Davis DS, Young LS, Hutt-Fletcher L, Tedder TF, Rickinson AB. Human Epithelial Cell Expression of an Epstein-barr Virus Receptor. *J Gen Virol* 1987; **68**: 805-811 [DOI: 10.1099/0022-1317-68-3-805]
  - 8 **Tanner J**, Weis J, Fearon D, Whang Y, Kieff E. Epstein-Barr virus gp350/220 binding to the B lymphocyte C3d receptor mediates adsorption, capping, and endocytosis. *Cell* 1987; **50**: 203-213 [PMID: 3036369]
  - 9 **Ogembo JG**, Kannan L, Ghiran I, Nicholson-Weller A, Finberg RW, Tsokos GC, Fingerhuth JD. Human complement receptor type 1/CD35 is an Epstein-Barr Virus receptor. *Cell Rep* 2013; **3**: 371-385 [PMID: 23416052 DOI: 10.1016/j.celrep.2013.01.023]
  - 10 **Tanner J**, Whang Y, Sample J, Sears A, Kieff E. Soluble gp350/220 and deletion mutant glycoproteins block Epstein-Barr virus adsorption to lymphocytes. *J Virol* 1988; **62**: 4452-4464 [PMID: 2460635]
  - 11 **Shannon-Lowe C**, Rowe M. Epstein-Barr virus infection of polarized epithelial cells via the basolateral surface by memory B cell-mediated transfer infection. *PLoS Pathog* 2011; **7**: e1001338 [PMID: 21573183 DOI: 10.1371/journal.ppat.1001338]
  - 12 **Sugiura M**, Imai S, Tokunaga M, Koizumi S, Uchizawa M, Okamoto K, Osato T. Transcriptional analysis of Epstein-Barr virus gene expression in EBV-positive gastric carcinoma: unique viral latency in the tumour cells. *Br J Cancer* 1996; **74**: 625-631 [PMID: 8761381]
  - 13 **Niedobitek G**, Agathangelou A, Herbst H, Whitehead L, Wright DH, Young LS. Epstein-Barr virus (EBV) infection in infectious mononucleosis: virus latency, replication and phenotype of EBV-infected cells. *J Pathol* 1997; **182**: 151-159 [PMID: 9274524]
  - 14 **Iwakiri D**, Eizuru Y, Tokunaga M, Takada K. Autocrine growth of Epstein-Barr virus-positive gastric carcinoma cells mediated by an Epstein-Barr virus-encoded small RNA. *Cancer Res* 2003; **63**: 7062-7067 [PMID: 14612496]
  - 15 **Luo B**, Wang Y, Wang XF, Liang H, Yan LP, Huang BH, Zhao P. Expression of Epstein-Barr virus genes in EBV-associated gastric carcinomas. *World J Gastroenterol* 2005; **11**: 629-633 [PMID: 15655811]
  - 16 **Hino R**, Uozaki H, Inoue Y, Shintani Y, Ushiku T, Sakatani T, Takada K, Fukayama M. Survival advantage of EBV-associated gastric carcinoma: survivin up-regulation by viral latent membrane protein 2A. *Cancer Res* 2008; **68**: 1427-1435 [PMID: 18316606 DOI: 10.1158/0008-5472.CAN-07-3027]
  - 17 **Zhao J**, Liang Q, Cheung KF, Kang W, Lung RW, Tong JH, To KF, Sung JJ, Yu J. Genome-wide identification of Epstein-Barr virus-driven promoter methylation profiles of human genes in gastric cancer cells. *Cancer* 2013; **119**: 304-312 [PMID: 22833454 DOI: 10.1002/cncr.27724]
  - 18 **Feinberg AP**, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet* 2006; **7**: 21-33 [PMID: 16369569 DOI: 10.1038/nrg1748]
  - 19 **Pfeifer GP**, Tang M, Denissenko MF. Mutation hotspots and DNA methylation. *Curr Top Microbiol Immunol* 2000; **249**: 1-19 [PMID: 10802935]
  - 20 **Toyota M**, Ahuja N, Suzuki H, Itoh F, Ohe-Toyota M, Imai K, Baylin SB, Issa JP. Aberrant methylation in gastric cancer associated with the CpG island methylator phenotype. *Cancer Res* 1999; **59**: 5438-5442 [PMID: 10554013]
  - 21 **Matsusaka K**, Kaneda A, Nagae G, Ushiku T, Kikuchi Y, Hino R, Uozaki H, Seto Y, Takada K, Aburatani H, Fukayama M. Classification of Epstein-Barr virus-positive gastric cancers by definition of DNA methylation epigenotypes. *Cancer Res* 2011; **71**: 7187-7197 [PMID: 21990320 DOI: 10.1158/0008-5472.CAN-11-1349]
  - 22 **Kaneda A**, Matsusaka K, Aburatani H, Fukayama M. Epstein-Barr virus infection as an epigenetic driver of tumorigenesis. *Cancer Res* 2012; **72**: 3445-3450 [PMID: 22761333 DOI: 10.1158/0008-5472.CAN-11-3919]
  - 23 **Qiu J**, Cosmopoulos K, Pegtel M, Hopmans E, Murray P, Middeldorp J, Shapiro M, Thorley-Lawson DA. A novel persistence associated EBV miRNA expression profile is disrupted in neoplasia. *PLoS Pathog* 2011; **7**: e1002193 [PMID: 21901094 DOI: 10.1371/journal.ppat.1002193]
  - 24 **Lo AK**, To KF, Lo KW, Lung RW, Hui JW, Liao G, Hayward SD. Modulation of LMP1 protein expression by EBV-encoded microRNAs. *Proc Natl Acad Sci USA* 2007; **104**: 16164-16169 [PMID: 17911266 DOI: 10.1073/pnas.0702896104]
  - 25 **Lung RW**, Tong JH, Sung YM, Leung PS, Ng DC, Chau SL, Chan AW, Ng EK, Lo KW, To KF. Modulation of LMP2A expression by a newly identified Epstein-Barr virus-encoded microRNA miR-BART22. *Neoplasia* 2009; **11**: 1174-1184 [PMID: 19881953]
  - 26 **Shinozaki A**, Sakatani T, Ushiku T, Hino R, Isogai M, Ishikawa S, Uozaki H, Takada K, Fukayama M. Downregulation of microRNA-200 in EBV-associated gastric carcinoma. *Cancer Res* 2010; **70**: 4719-4727 [PMID: 20484038 DOI: 10.1158/0008-5472.CAN-09-4620]
  - 27 **Marquitz AR**, Mathur A, Chugh PE, Dittmer DP, Raab-Traub N. Expression profile of microRNAs in Epstein-Barr virus-infected AGS gastric carcinoma cells. *J Virol* 2014; **88**: 1389-1393 [PMID: 24227849 DOI: 10.1128/JVI.02662-13]
  - 28 **Kaneda A**, Kaminishi M, Yanagihara K, Sugimura T, Ushijima T. Identification of silencing of nine genes in human gastric cancers. *Cancer Res* 2002; **62**: 6645-6650 [PMID: 12438262]
  - 29 **Kusano M**, Toyota M, Suzuki H, Akino K, Aoki F, Fujita M, Hosokawa M, Shinomura Y, Imai K, Tokino T. Genetic, epigenetic, and clinicopathologic features of gastric carcinomas with the CpG island methylator phenotype and an association with Epstein-Barr virus. *Cancer* 2006; **106**: 1467-1479 [PMID: 16518809 DOI: 10.1002/cncr.21789]
  - 30 **Chang MS**, Uozaki H, Chong JM, Ushiku T, Sakuma K, Ishikawa S, Hino R, Barua RR, Iwasaki Y, Arai K, Fujii H, Nagai H, Fukayama M. CpG island methylation status in gastric carcinoma with and without infection of Epstein-Barr virus. *Clin Cancer Res* 2006; **12**: 2995-3002 [PMID: 16707594 DOI: 10.1158/1078-0432.CCR-05-1601]
  - 31 **Kang GH**, Lee S, Kim WH, Lee HW, Kim JC, Rhyu MG, Ro JY. Epstein-barr virus-positive gastric carcinoma demonstrates frequent aberrant methylation of multiple genes and constitutes CpG island methylator phenotype-positive gastric carcinoma. *Am J Pathol* 2002; **160**: 787-794 [PMID: 11891177 DOI: 10.1016/S0002-9440(10)64901-2]
  - 32 **Saito M**, Nishikawa J, Okada T, Morishige A, Sakai K, Nakamura M, Kiyotoki S, Hamabe K, Okamoto T, Oga A, Sasaki K, Suehiro Y, Hinoda Y, Sakaida I. Role of DNA methylation in the development of Epstein-Barr virus-associated gastric carcinoma. *J Med Virol* 2013; **85**: 121-127 [PMID: 23073987 DOI: 10.1099/0950-2688-85-2-121]

- 10.1002/jmv.23405]
- 33 **Ushiku T**, Chong JM, Uozaki H, Hino R, Chang MS, Sudo M, Rani BR, Sakuma K, Nagai H, Fukayama M. p73 gene promoter methylation in Epstein-Barr virus-associated gastric carcinoma. *Int J Cancer* 2007; **120**: 60-66 [PMID: 17058198 DOI: 10.1002/ijc.22275]
  - 34 **Okada T**, Nakamura M, Nishikawa J, Sakai K, Zhang Y, Saito M, Morishige A, Oga A, Sasaki K, Suehiro Y, Hinoda Y, Sakaida I. Identification of genes specifically methylated in Epstein-Barr virus-associated gastric carcinomas. *Cancer Sci* 2013; **104**: 1309-1314 [PMID: 23829175 DOI: 10.1111/cas.12228]
  - 35 **Sakuma K**, Chong JM, Sudo M, Ushiku T, Inoue Y, Shibahara J, Uozaki H, Nagai H, Fukayama M. High-density methylation of p14ARF and p16INK4A in Epstein-Barr virus-associated gastric carcinoma. *Int J Cancer* 2004; **112**: 273-278 [PMID: 15352040 DOI: 10.1002/ijc.20420]
  - 36 **Sherr CJ**. Cancer cell cycles. *Science* 1996; **274**: 1672-1677 [PMID: 8939849]
  - 37 **Jang TJ**, Kim DI, Shin YM, Chang HK, Yang CH. p16(INK4a) Promoter hypermethylation of non-tumorous tissue adjacent to gastric cancer is correlated with glandular atrophy and chronic inflammation. *Int J Cancer* 2001; **93**: 629-634 [PMID: 11477571]
  - 38 **Ohfuji S**, Osaki M, Tsujitani S, Ikeguchi M, Sairenji T, Ito H. Low frequency of apoptosis in Epstein-Barr virus-associated gastric carcinoma with lymphoid stroma. *Int J Cancer* 1996; **68**: 710-715 [PMID: 8980171]
  - 39 **Zhao J**, Liang Q, Cheung KF, Kang W, Dong Y, Lung RW, Tong JH, To KF, Sung JJ, Yu J. Somatostatin receptor 1, a novel EBV-associated CpG hypermethylated gene, contributes to the pathogenesis of EBV-associated gastric cancer. *Br J Cancer* 2013; **108**: 2557-2564 [PMID: 23722468 DOI: 10.1038/bjc.2013.263]
  - 40 **Kim J**, Lee HS, Bae SI, Lee YM, Kim WH. Silencing and CpG island methylation of GSTP1 is rare in ordinary gastric carcinomas but common in Epstein-Barr virus-associated gastric carcinomas. *Anticancer Res* 2005; **25**: 4013-4019 [PMID: 16309193]
  - 41 **Challouf S**, Ziadi S, Zaghdoudi R, Ksaa F, Ben Gacem R, Trimeche M. Patterns of aberrant DNA hypermethylation in nasopharyngeal carcinoma in Tunisian patients. *Clin Chim Acta* 2012; **413**: 795-802 [PMID: 22296674 DOI: 10.1016/j.cca.2012.01.018]
  - 42 **Guo X**, Zeng Y, Deng H, Liao J, Zheng Y, Li J, Kessing B, O'Brien SJ. Genetic Polymorphisms of CYP2E1, GSTP1, NQO1 and MPO and the Risk of Nasopharyngeal Carcinoma in a Han Chinese Population of Southern China. *BMC Res Notes* 2010; **3**: 212 [PMID: 20663217 DOI: 10.1186/1756-0500-3-212]
  - 43 **Li Z**, Mou H, Wang T, Xue J, Deng B, Qian L, Zhou Y, Gong W, Wang JM, Wu G, Zhou CF, Fang J, Le Y. A non-secretory form of FAM3B promotes invasion and metastasis of human colon cancer cells by upregulating Slug expression. *Cancer Lett* 2013; **328**: 278-284 [PMID: 23059759 DOI: 10.1016/j.canlet.2012.09.026]
  - 44 **Evangelista M**, Tian H, de Sauvage FJ. The hedgehog signaling pathway in cancer. *Clin Cancer Res* 2006; **12**: 5924-5928 [PMID: 17062662 DOI: 10.1158/1078-0432.CCR-06-1736]
  - 45 **Fan H**, Oro AE, Scott MP, Khavari PA. Induction of basal cell carcinoma features in transgenic human skin expressing Sonic Hedgehog. *Nat Med* 1997; **3**: 788-792 [PMID: 9212109 DOI: 10.1038/nm0797-788]
  - 46 **Blenk S**, Engelmann J, Weniger M, Schultz J, Dittrich M, Rosenwald A, Müller-Hermelink HK, Müller T, Dandekar T. Germinal center B cell-like (GCB) and activated B cell-like (ABC) type of diffuse large B cell lymphoma (DLBCL): analysis of molecular predictors, signatures, cell cycle state and patient survival. *Cancer Inform* 2007; **3**: 399-420 [PMID: 19455257]
  - 47 **Park SY**, Kook MC, Kim YW, Cho NY, Jung N, Kwon HJ, Kim TY, Kang GH. CpG island hypermethylator phenotype in gastric carcinoma and its clinicopathological features. *Virchows Arch* 2010; **457**: 415-422 [PMID: 20737169 DOI: 10.1007/s00428-010-0962-0]
  - 48 **Shin HJ**, Kim do N, Lee SK. Association between Epstein-Barr virus infection and chemoresistance to docetaxel in gastric carcinoma. *Mol Cells* 2011; **32**: 173-179 [PMID: 21626300 DOI: 10.1007/s10059-011-0066-y]
  - 49 **Mei YP**, Zhou JM, Wang Y, Huang H, Deng R, Feng GK, Zeng YX, Zhu XF. Silencing of LMP1 induces cell cycle arrest and enhances chemosensitivity through inhibition of AKT signaling pathway in EBV-positive nasopharyngeal carcinoma cells. *Cell Cycle* 2007; **6**: 1379-1385 [PMID: 17507800]
  - 50 **Qiu H**, Yashiro M, Shinto O, Matsuzaki T, Hirakawa K. DNA methyltransferase inhibitor 5-aza-CdR enhances the radiosensitivity of gastric cancer cells. *Cancer Sci* 2009; **100**: 181-188 [PMID: 19037991 DOI: 10.1111/j.1349-7006.2008.01004.x]
  - 51 **Bolden JE**, Peart MJ, Johnstone RW. Anticancer activities of histone deacetylase inhibitors. *Nat Rev Drug Discov* 2006; **5**: 769-784 [PMID: 16955068 DOI: 10.1038/nrd2133]
  - 52 **Fiskus W**, Pranpat M, Bali P, Balasis M, Kumaraswamy S, Boyapalle S, Rocha K, Wu J, Giles F, Manley PW, Atadja P, Bhalla K. Combined effects of novel tyrosine kinase inhibitor AMN107 and histone deacetylase inhibitor LBH589 against Bcr-Abl-expressing human leukemia cells. *Blood* 2006; **108**: 645-652 [PMID: 16537804 DOI: 10.1182/blood-2005-11-4639]
  - 53 **Jung EJ**, Lee YM, Lee BL, Chang MS, Kim WH. Lytic induction and apoptosis of Epstein-Barr virus-associated gastric cancer cell line with epigenetic modifiers and ganciclovir. *Cancer Lett* 2007; **247**: 77-83 [PMID: 16647201 DOI: 10.1016/j.canlet.2006.03.022]
  - 54 **Zhao J**, Jin H, Cheung KF, Tong JH, Zhang S, Go MY, Tian L, Kang W, Leung PP, Zeng Z, Li X, To KF, Sung JJ, Yu J. Zinc finger E-box binding factor 1 plays a central role in regulating Epstein-Barr virus (EBV) latent-lytic switch and acts as a therapeutic target in EBV-associated gastric cancer. *Cancer* 2012; **118**: 924-936 [PMID: 21717425 DOI: 10.1002/cncr.26184]
  - 55 **Yoo CB**, Cheng JC, Jones PA. Zebularine: a new drug for epigenetic therapy. *Biochem Soc Trans* 2004; **32**: 910-912 [PMID: 15506921 DOI: 10.1042/BST0320910]
  - 56 **Lübbert M**, Wijermans P, Kunzmann R, Verhoef G, Bosly A, Ravoet C, Andre M, Ferrant A. Cytogenetic responses in high-risk myelodysplastic syndrome following low-dose treatment with the DNA methylation inhibitor 5-aza-2'-deoxycytidine. *Br J Haematol* 2001; **114**: 349-357 [PMID: 11529854]
  - 57 **Yang AS**, Estecio MR, Garcia-Manero G, Kantarjian HM, Issa JP. Comment on "Chromosomal instability and tumors promoted by DNA hypomethylation" and "Induction of tumors in mice by genomic hypomethylation". *Science* 2003; **302**: 1153; author reply 1153 [PMID: 14615517 DOI: 10.1126/science.1089523]
  - 58 **Skalsky RL**, Corcoran DL, Gottwein E, Frank CL, Kang D, Hafner M, Nusbaum JD, Feederle R, Delecluse HJ, Luftig MA, Tuschl T, Ohler U, Cullen BR. The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS Pathog* 2012; **8**: e1002484 [PMID: 22291592 DOI: 10.1371/journal.ppat.1002484]
  - 59 **Riley KJ**, Rabinowitz GS, Yario TA, Luna JM, Darnell RB, Steitz JA. EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. *EMBO J* 2012; **31**: 2207-2221 [PMID: 22473208 DOI: 10.1038/emboj.2012.63]
  - 60 **Xia T**, O'Hara A, Araujo I, Barreto J, Carvalho E, Sapucaia JB, Ramos JC, Luz E, Pedroso C, Manrique M, Toomey NL, Brites C, Dittmer DP, Harrington WJ. EBV microRNAs in primary lymphomas and targeting of CXCL-11 by ebv-mir-BHRF1-3. *Cancer Res* 2008; **68**: 1436-1442 [PMID: 18316607 DOI: 10.1158/0008-5472.CAN-07-5126]
  - 61 **Barth S**, Pfuhl T, Mamiani A, Ehse C, Roemer K, Kremmer E, Jäker C, Höck J, Meister G, Grässer FA. Epstein-Barr virus-encoded microRNA miR-BART2 down-regulates the viral DNA polymerase BALF5. *Nucleic Acids Res* 2008; **36**: 666-675



- [PMID: 18073197 DOI: 10.1093/nar/gkm1080]
- 62 **Nachmani D**, Stern-Ginossar N, Sarid R, Mandelboim O. Diverse herpesvirus microRNAs target the stress-induced immune ligand MICB to escape recognition by natural killer cells. *Cell Host Microbe* 2009; **5**: 376-385 [PMID: 19380116 DOI: 10.1016/j.chom.2009.03.003]
- 63 **Dölken L**, Malterer G, Erhard F, Kothe S, Friedel CC, Suffert G, Marciniowski L, Motsch N, Barth S, Beitzinger M, Lieber D, Bailer SM, Hoffmann R, Ruzsics Z, Kremmer E, Pfeffer S, Zimmer R, Koszinowski UH, Grässer F, Meister G, Haas J. Systematic analysis of viral and cellular microRNA targets in cells latently infected with human gamma-herpesviruses by RISC immunoprecipitation assay. *Cell Host Microbe* 2010; **7**: 324-334 [PMID: 20413099 DOI: 10.1016/j.chom.2010.03.008]
- 64 **Choy EY**, Siu KL, Kok KH, Lung RW, Tsang CM, To KF, Kwong DL, Tsao SW, Jin DY. An Epstein-Barr virus-encoded microRNA targets PUMA to promote host cell survival. *J Exp Med* 2008; **205**: 2551-2560 [PMID: 18838543 DOI: 10.1084/jem.20072581]
- 65 **Iizasa H**, Wulff BE, Alla NR, Maragkakis M, Megraw M, Hatzigeorgiou A, Iwakiri D, Takada K, Wiedmer A, Showe L, Lieberman P, Nishikura K. Editing of Epstein-Barr virus-encoded BART6 microRNAs controls their dicer targeting and consequently affects viral latency. *J Biol Chem* 2010; **285**: 33358-33370 [PMID: 20716523 DOI: 10.1074/jbc.M110.138362]
- 66 **Marquitz AR**, Mathur A, Nam CS, Raab-Traub N. The Epstein-Barr Virus BART microRNAs target the pro-apoptotic protein Bim. *Virology* 2011; **412**: 392-400 [PMID: 21333317 DOI: 10.1016/j.virol.2011.01.028]
- 67 **Hino R**, Uozaki H, Murakami N, Ushiku T, Shinozaki A, Ishikawa S, Morikawa T, Nakaya T, Sakatani T, Takada K, Fukayama M. Activation of DNA methyltransferase 1 by EBV latent membrane protein 2A leads to promoter hypermethylation of PTEN gene in gastric carcinoma. *Cancer Res* 2009; **69**: 2766-2774 [PMID: 19339266 DOI: 10.1158/0008-5472.CAN-08-3070]
- 68 **Sudo M**, Chong JM, Sakuma K, Ushiku T, Uozaki H, Nagai H, Funata N, Matsumoto Y, Fukayama M. Promoter hypermethylation of E-cadherin and its abnormal expression in Epstein-Barr virus-associated gastric carcinoma. *Int J Cancer* 2004; **109**: 194-199 [PMID: 14750169 DOI: 10.1002/ijc.11701]
- 69 **Zazula M**, Ferreira AM, Czopek JP, Kolodziejczyk P, Sinczak-Kuta A, Klimkowska A, Wojcik P, Okon K, Bialas M, Kulig J, Stachura J. CDH1 gene promoter hypermethylation in gastric cancer: relationship to Goseki grading, microsatellite instability status, and EBV invasion. *Diagn Mol Pathol* 2006; **15**: 24-29 [PMID: 16531765]
- 70 **Kang GH**, Lee S, Cho NY, Gandamihardja T, Long TI, Weisenberger DJ, Campan M, Laird PW. DNA methylation profiles of gastric carcinoma characterized by quantitative DNA methylation analysis. *Lab Invest* 2008; **88**: 161-170 [PMID: 18158559 DOI: 10.1038/labinvest.3700707]
- 71 **Liu X**, Wang Y, Wang X, Sun Z, Li L, Tao Q, Luo B. Epigenetic silencing of WNT5A in Epstein-Barr virus-associated gastric carcinoma. *Arch Virol* 2013; **158**: 123-132 [PMID: 23001722 DOI: 10.1007/s00705-012-1481-x]
- 72 **Kuo HK**, Griffith JD, Kreuzer KN. 5-Azacytidine induced methyltransferase-DNA adducts block DNA replication in vivo. *Cancer Res* 2007; **67**: 8248-8254 [PMID: 17804739 DOI: 10.1158/0008-5472.CAN-07-1038]
- 73 **Singh N**, Dueñas-González A, Lyko F, Medina-Franco JL. Molecular modeling and molecular dynamics studies of hydralazine with human DNA methyltransferase 1. *ChemMedChem* 2009; **4**: 792-799 [PMID: 19322801 DOI: 10.1002/cmdc.200900017]
- 74 **Shanafelt TD**, Call TG, Zent CS, Leis JF, LaPlant B, Bowen DA, Roos M, Laumann K, Ghosh AK, Lesnick C, Lee MJ, Yang CS, Jelinek DF, Erlichman C, Kay NE. Phase 2 trial of daily, oral Polyphenon E in patients with asymptomatic, Rai stage 0 to II chronic lymphocytic leukemia. *Cancer* 2013; **119**: 363-370 [PMID: 22760587 DOI: 10.1002/cncr.27719]
- 75 **McLarty J**, Bigelow RL, Smith M, Elmajian D, Ankem M, Cardelli JA. Tea polyphenols decrease serum levels of prostate-specific antigen, hepatocyte growth factor, and vascular endothelial growth factor in prostate cancer patients and inhibit production of hepatocyte growth factor and vascular endothelial growth factor in vitro. *Cancer Prev Res (Phila)* 2009; **2**: 673-682 [PMID: 19542190 DOI: 10.1158/1940-6207.CAPR-08-0167]
- 76 **Wang JS**, Luo H, Wang P, Tang L, Yu J, Huang T, Cox S, Gao W. Validation of green tea polyphenol biomarkers in a phase II human intervention trial. *Food Chem Toxicol* 2008; **46**: 232-240 [PMID: 17888558 DOI: 10.1016/j.fct.2007.08.007]
- 77 **Beumer JH**, Parise RA, Newman EM, Doroshow JH, Synold TW, Lenz HJ, Egorin MJ. Concentrations of the DNA methyltransferase inhibitor 5-fluoro-2'-deoxycytidine (FdCyd) and its cytotoxic metabolites in plasma of patients treated with FdCyd and tetrahydrouridine (THU). *Cancer Chemother Pharmacol* 2008; **62**: 363-368 [PMID: 17899082 DOI: 10.1007/s00280-007-0603-8]
- 78 **Zhao Q**, Fan J, Hong W, Li L, Wu M. Inhibition of cancer cell proliferation by 5-fluoro-2'-deoxycytidine, a DNA methylation inhibitor, through activation of DNA damage response pathway. *Springerplus* 2012; **1**: 65 [PMID: 23397046 DOI: 10.1186/2193-1801-1-65]
- 79 **Zhou L**, Cheng X, Connolly BA, Dickman MJ, Hurd PJ, Hornby DP. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J Mol Biol* 2002; **321**: 591-599 [PMID: 12206775]
- 80 **Villar-Garea A**, Fraga MF, Espada J, Esteller M. Procaine is a DNA-demethylating agent with growth-inhibitory effects in human cancer cells. *Cancer Res* 2003; **63**: 4984-4989 [PMID: 12941824]
- 81 **Huisen TJ**, Kamble-Shripat T. Delayed toxicity of two chitinolytic enzyme inhibitors (psammaplin A and pentoxifylline) against eastern subterranean termites (Isoptera: Rhinotermitidae). *J Econ Entomol* 2013; **106**: 1788-1793 [PMID: 24020294]
- 82 **Baud MG**, Leiser T, Petrucci V, Gunaratnam M, Neidle S, Meyer-Almes FJ, Fuchter MJ. Thioester derivatives of the natural product psammaplin A as potent histone deacetylase inhibitors. *Beilstein J Org Chem* 2013; **9**: 81-88 [PMID: 23400330 DOI: 10.3762/bjoc.9.11]
- 83 **Brueckner B**, Garcia Boy R, Siedlecki P, Musch T, Kliem HC, Zielenkiewicz P, Suhai S, Wiessler M, Lyko F. Epigenetic reactivation of tumor suppressor genes by a novel small-molecule inhibitor of human DNA methyltransferases. *Cancer Res* 2005; **65**: 6305-6311 [PMID: 16024632 DOI: 10.1158/0008-5472.CAN-04-2957]
- 84 **Suzuki T**, Tanaka R, Hamada S, Nakagawa H, Miyata N. Design, synthesis, inhibitory activity, and binding mode study of novel DNA methyltransferase 1 inhibitors. *Bioorg Med Chem Lett* 2010; **20**: 1124-1127 [PMID: 20056538 DOI: 10.1016/j.bmcl.2009.12.016]
- 85 **Savickiene J**, Treigyte G, Jazdauskaite A, Borutinskaite VV, Navakauskiene R. DNA methyltransferase inhibitor RG108 and histone deacetylase inhibitors cooperate to enhance NB4 cell differentiation and E-cadherin re-expression by chromatin remodelling. *Cell Biol Int* 2012; **36**: 1067-1078 [PMID: 22845560 DOI: 10.1042/CBI20110649]
- 86 **Savickiene J**, Treigyte G, Borutinskaite VV, Navakauskiene R. Antileukemic activity of combined epigenetic agents, DNMT inhibitors zebularine and RG108 with HDAC inhibitors, against promyelocytic leukemia HL-60 cells. *Cell Mol Biol Lett* 2012; **17**: 501-525 [PMID: 22820861 DOI: 10.2478/s11658-012-0024-5]
- 87 **Cheng JC**, Matsen CB, Gonzales FA, Ye W, Greer S, Marquez VE, Jones PA, Selker EU. Inhibition of DNA methylation and reactivation of silenced genes by zebularine. *J Natl Cancer Inst* 2003; **95**: 399-409 [PMID: 12618505 DOI: 10.1093/

- jnci/95.5.399]
- 88 **Billam M**, Sobolewski MD, Davidson NE. Effects of a novel DNA methyltransferase inhibitor zebularine on human breast cancer cells. *Breast Cancer Res Treat* 2010; **120**: 581-592 [PMID: 19459041 DOI: 10.1007/s10549-009-0420-3]
- 89 **Flotho C**, Claus R, Batz C, Schneider M, Sandrock I, Ihde S, Plass C, Niemeyer CM, Lübbert M. The DNA methyltransferase inhibitors azacitidine, decitabine and zebularine exert differential effects on cancer gene expression in acute myeloid leukemia cells. *Leukemia* 2009; **23**: 1019-1028 [PMID: 19194470 DOI: 10.1038/leu.2008.397]
- 90 **Fatkins DG**, Zheng W. Substituting N(epsilon)-thioacetyl-lysine for N(epsilon)-acetyl-lysine in peptide substrates as a general approach to inhibiting human NAD(+)-dependent protein deacetylases. *Int J Mol Sci* 2008; **9**: 1-11 [PMID: 19325715 DOI: 10.3390/ijms9010001]
- 91 **Gui CY**, Ngo L, Xu WS, Richon VM, Marks PA. Histone deacetylase (HDAC) inhibitor activation of p21WAF1 involves changes in promoter-associated proteins, including HDAC1. *Proc Natl Acad Sci USA* 2004; **101**: 1241-1246 [PMID: 14734806 DOI: 10.1073/pnas.0307708100]
- 92 **Solomon JM**, Pasupuleti R, Xu L, McDonagh T, Curtis R, DiStefano PS, Huber LJ. Inhibition of SIRT1 catalytic activity increases p53 acetylation but does not alter cell survival following DNA damage. *Mol Cell Biol* 2006; **26**: 28-38 [PMID: 16354677 DOI: 10.1128/MCB.26.1.28-38.2006]
- 93 **Garcia-Manero G**, Kantarjian HM, Sanchez-Gonzalez B, Yang H, Rosner G, Verstovsek S, Rytting M, Wierda WG, Ravandi F, Koller C, Xiao L, Faderl S, Estrov Z, Cortes J, O'Brien S, Estey E, Bueso-Ramos C, Fiorentino J, Jabbour E, Issa JP. Phase 1/2 study of the combination of 5-aza-2'-deoxycytidine with valproic acid in patients with leukemia. *Blood* 2006; **108**: 3271-3279 [PMID: 16882711 DOI: 10.1182/blood-2006-03-009142]
- 94 **Nemunaitis JJ**, Orr D, Eager R, Cunningham CC, Williams A, Mennel R, Grove W, Olson S. Phase I study of oral CI-994 in combination with gemcitabine in treatment of patients with advanced cancer. *Cancer J* 2003; **9**: 58-66 [PMID: 12602769]
- 95 **Undevia SD**, Kindler HL, Janisch L, Olson SC, Schilsky RL, Vogelzang NJ, Kimmel KA, Macek TA, Ratain MJ. A phase I study of the oral combination of CI-994, a putative histone deacetylase inhibitor, and capecitabine. *Ann Oncol* 2004; **15**: 1705-1711 [PMID: 15520075 DOI: 10.1093/annonc/mdh438]
- 96 **Medda F**, Russell RJ, Higgins M, McCarthy AR, Campbell J, Slawin AM, Lane DP, Lain S, Westwood NJ. Novel cambinol analogs as sirtuin inhibitors: synthesis, biological evaluation, and rationalization of activity. *J Med Chem* 2009; **52**: 2673-2682 [PMID: 19419202 DOI: 10.1021/jm8014298]

**P- Reviewers:** Engin AB, Jung YD **S- Editor:** Zhai HH  
**L- Editor:** A **E- Editor:** Wang CH



# Hyperactive neutrophil chemotaxis contributes to anti-tumor necrosis factor- $\alpha$ treatment resistance in inflammatory bowel disease

**Tung On Yau**, Jayakumar Vadakekolathu, Gemma Ann Foulds, Guodong Du, Benjamin Dickins, Christos Polytarchou, Sergio Rutella

*Journal of Gastroenterology and Hepatology*. **2022**;37(3): 531–541.








Journal URL: <https://onlinelibrary.wiley.com/doi/10.1111/jgh.15764>

DOI: [10.1111/jgh.15764](https://doi.org/10.1111/jgh.15764)

PMID: [34931384](https://pubmed.ncbi.nlm.nih.gov/34931384/)

ORIGINAL ARTICLE - GASTROENTEROLOGY (EXPERIMENTAL)

# Hyperactive neutrophil chemotaxis contributes to anti-tumor necrosis factor- $\alpha$ treatment resistance in inflammatory bowel disease

Tung On Yau,<sup>\*,†</sup>  Jayakumar Vadakekolathu,<sup>\*,†</sup>  Gemma Ann Foulds,<sup>\*,†</sup>  Guodong Du,<sup>†</sup>   
Benjamin Dickins,<sup>\*,†</sup>  Christos Polytaichou,<sup>\*,†</sup>  and Sergio Rutella<sup>\*,†</sup> 

\*John van Geest Cancer Research Centre, School of Science and Technology, <sup>†</sup>Centre for Health, Ageing and Understanding Disease, Nottingham Trent University, Clifton Campus, Nottingham, United Kingdom; and <sup>‡</sup>Department of Artificial Intelligence, Xiamen University, Xiamen, China

## Key words

Anti-TNF- $\alpha$ , Chemotaxis, Diagnostic biomarkers, IL13RA2, Inflammatory bowel diseases.

Accepted for publication 7 December 2021.

## Correspondence

Professor Sergio Rutella and Tung On Yau, John van Geest Cancer Research Centre, School of Science and Technology, Nottingham Trent University, Clifton Campus, Nottingham NG11 8NS, United Kingdom.  
Email: sergio.rutella@ntu.ac.uk; payton.yau@ntu.ac.uk

**Declaration of conflict of interest:** The authors have no conflicts of interest to declare.

**Author contribution:** Study conception and design: Tung On Yau. Development of methodology: Tung On Yau. Data acquisition and analysis: Tung On Yau and Guodong Du. Data interpretation: Tung On Yau, Jayakumar Vadakekolathu, and Sergio Rutella. Writing—original draft preparation: Tung On Yau. Writing—review and editing: Jayakumar Vadakekolathu, Gemma Ann Foulds, and Benjamin Dickins. Administrative, technical, or material support: Benjamin Dickins, Christos Polytaichou, and Sergio Rutella. Study supervision: Christos Polytaichou and Sergio Rutella.

**Ethical approval:** No ethical clearance is required because this study retrieved and synthesized the data from already published studies, and researchers of each of the original studies obtained approval from their local ethics committee.

**Financial support:** This research received no specific grant from any funding bodies. The article publication charges were supported by Wiley's Jisc agreement.

[Correction added on 11 February 2022, after first online publication: The copyright line was changed.]

## Abstract

**Background and Aim:** Anti-tumor necrosis factor- $\alpha$  (anti-TNF- $\alpha$ ) agents have been used for inflammatory bowel disease; however, it has up to 30% nonresponse rate. Identifying molecular pathways and finding reliable diagnostic biomarkers for patient response to anti-TNF- $\alpha$  treatment are needed.

**Methods:** Publicly available transcriptomic data from inflammatory bowel disease patients receiving anti-TNF- $\alpha$  therapy were systemically collected and integrated. *In silico* flow cytometry approaches and Metascape were applied to evaluate immune cell populations and to perform gene enrichment analysis, respectively. Genes identified within enrichment pathways validated in neutrophils were tracked in an anti-TNF- $\alpha$ -treated animal model (with lipopolysaccharide-induced inflammation). The receiver operating characteristic curve was applied to all genes to identify the best prediction biomarkers.

**Results:** A total of 449 samples were retrieved from control, baseline, and after primary anti-TNF- $\alpha$  therapy or placebo. No statistically significant differences were observed between anti-TNF- $\alpha$  treatment responders and nonresponders at baseline in immune microenvironment scores. Neutrophil, endothelial cell, and B-cell populations were higher in baseline nonresponders, and chemotaxis pathways may contribute to the treatment resistance. Genes related to chemotaxis pathways were significantly upregulated in lipopolysaccharide-induced neutrophils, but no statistically significant changes were observed in neutrophils treated with anti-TNF- $\alpha$ . Interleukin 13 receptor subunit alpha 2 (*IL13RA2*) is the best predictor (receiver operating characteristic curve: 80.7%, 95% confidence interval: 73.8–87.5%), with a sensitivity of 68.13% and specificity of 84.93%, and significantly higher in nonresponders compared with responders ( $P < 0.0001$ ).

**Conclusions:** Hyperactive neutrophil chemotaxis influences responses to anti-TNF- $\alpha$  treatment, and *IL13RA2* is a potential biomarker to predict anti-TNF- $\alpha$  treatment response.

## Introduction

Inflammatory bowel disease (IBD) is an idiopathic and relapsing–remitting chronic inflammatory disorder characterized by a susceptible genetic background, causing immunological dysfunction and intestinal microbiome dysbiosis.<sup>1</sup> The long-standing mucosa inflammation destructs tight junctions, induces intestinal barrier injury and permeability, and increases the incidence of colonic neoplasia.<sup>1</sup> It is estimated that the prevalence of IBD exceeds 0.3% in North America, Oceania, and many countries in Europe.<sup>2</sup> With the incidence rising in the newly industrialized countries, including Brazil and Taiwan,<sup>3</sup> thus, IBD places a large burden on public health services and healthcare economies.

Tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) is a pleiotropic cytokine that participates in several pathological processes in IBD and is recognized as a pro-inflammatory cytokine. A soluble, biologically active homotrimer TNF- $\alpha$  originally from the monomeric TNF- $\alpha$  claved by TNF- $\alpha$ -converting enzyme (TACE) via proteolysis.<sup>4</sup> TNF- $\alpha$  activity is mediated through binding to the TNF receptors I and II (TNFRI and TNFRII).<sup>5</sup> This binding activates immune cells response and pro-inflammatory cytokine and chemokine productions, such as IL-1, IL-6, IL-8, and RANTES. It also increases the expression of adhesion molecules, production of matrix metalloproteinase, and induction of apoptosis.<sup>6</sup> The use of anti-TNF- $\alpha$  compounds such as full monoclonal IgG1 antibodies (infliximab and adalimumab), pegylated anti-TNF- $\alpha$  F[ab']<sub>2</sub> fragment (certolizumab), and IgG1 $\kappa$  monoclonal antibody derived from immunizing genetically engineered mice with human TNF- $\alpha$  (golimumab) has been approved for IBD patients,<sup>7</sup> including Crohn's disease (CD) and ulcerative colitis (UC) and IBD unclassified (IBD-U).<sup>8</sup>

Although CD and UC are the distinct subtypes of IBD, these diseases present a certain level of similarities, including symptoms, pathological features, immune response, risk factors, and the biological pathways producing TNF- $\alpha$ .<sup>9</sup> In addition, studies found that up to 3% of CD patients will be reclassified as UC and vice versa after their primary diagnosis, 5–15% of IBD patients classified as IBD-U, and a small portion of UC patients are later changed to CD or IBD-U.<sup>10</sup> More importantly, up to 30% of patients do not respond to anti-TNF- $\alpha$  blockers,<sup>1,11</sup> and the use of vedolizumab (anti-IL-12/23) and ustekinumab (anti-integrin) may be efficacious in many patients that failed anti-TNF- $\alpha$  therapy.<sup>12</sup> Thus, there is a clear need to identify potential anti-TNF- $\alpha$  treatment pathways in overall IBD patients with a view to better targeting anti-TNF- $\alpha$  treatment to more responsive cohorts and to minimize the adverse anti-TNF- $\alpha$  treatment effects.

## Methods

**Search strategy and data collection and integration.** A searching strategy for publicly available datasets related to IBD patients who received anti-TNF- $\alpha$  therapy was designed for the NCBI Gene Expression Omnibus database dated December 31, 2020, using the keywords, “TNF,” “Tumor Necrosis Factor,” “anti-TNF,” “anti-Tumor Necrosis Factor,” “Infliximab,” “Adalimumab,” “Golimumab,” “inflammatory bowel disease,” “IBD,” “ulcerative colitis,” “UC,” “Crohn Disease,” and “CD.” The included datasets have to meet the following inclusion criteria: (i) colonic sample from IBD patients, (ii) transcriptomic

data, (iii) raw data are available, (iv) anti-TNF- $\alpha$  treatment response status, (v) publicly accessible, and (vi) each of the original studies obtained approval from their local ethics committee and had written informed patient consent. Sample exclusion criteria were as follows: (i) subjects receive therapy other than anti-TNF- $\alpha$ , (ii) overlapped subjects, (iii) colonic samples other than large intestine, (iv) transcriptomic data other than Affymetrix, and (v) the posttreatment time point being over 3 months.

The eligible raw microarray datasets were collected and subjected to background correction, normalization, and summarization using the Robust Multichip Average (RMA) algorithm using Affy package version 1.66.0 individually.<sup>13</sup> Mean value of multiple probe sets representing the same gene was calculated. Next, the ComBat function from sva package version 3.36.0 was implemented on the datasets to eliminate the study-specific batch effects.<sup>14,15</sup>

### Composition of immune cells and immune-related score evaluation.

The evaluation of immune microenvironment scores and immune–stroma cell population are calculated using xCELL<sup>16</sup> and ESTIMATE (Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data)<sup>17</sup> algorithms. The immune cell types were evaluated from gene expression profile using five different algorithms, including CIBERSORT,<sup>18</sup> EPIC,<sup>19</sup> MCP-counter,<sup>20</sup> xCELL, and Deconvolution-To-Estimate-Immune-Cells (DTEIC).<sup>21</sup> Each of the algorithms was developed using their in-house or publicly immune cells expression data and different statistical learning approaches. For instance, DTEIC utilized  $\epsilon$ -support vector regression and CIBERSORT applied linear support vector regression<sup>18,21</sup>; MCP-counter is a single-sample scoring system, while xCELL requires heterogeneous dataset<sup>16,20</sup>; ESTIMATE utilizes single-sample Gene Set Enrichment Analysis (ssGSEA) to rank samples on the expression of two different 141 gene sets, and xCELL is based on the sets of cells values calculated from its algorithm.<sup>16,17</sup>

Value for each of the immune cell type estimated from EPIC version 1.1, MCP-counter version 1.2.0, and ESTIMATE version 1.0.13 were performed in R programming version 4.0.0, DTEIC was operated in Python 3.7 and CIBERSORT and xCELL were calculated by using their corresponding online tools. The source code can be found on the corresponding authors' GitHub page.

**Functional enrichment analysis.** Identification of differentially expressed genes (DEGs) between responders and nonresponders was calculated using limma package version 3.22.3,<sup>22</sup> and the threshold for the DEGs has a Benjamini–Hochberg adjusted  $P$ -value  $< 0.05$  with absolute  $\log_2$  fold change  $\geq 0.75$ . EnhancedVolcano package version 1.6.0 was applied for volcano plot.<sup>23</sup> Heatmap was generated by using pheatmap package version 1.0.12.<sup>24</sup> All the packages are applied within the R programming environment. The differentially overexpressed genes were utilized for the pathway enrichment analysis using Metascape (<http://metascape.org>),<sup>25</sup> a gene enrichment tool for understanding from previously pre-defined gene sets in different enriched biological themes, including GO terms, Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, BioCarta, and MSigDB. For each gene inputted into the server, the enrichment score was

calculated and clustered to match biological signaling pathways. Visualization of the selected pathways utilized Cytoscape version 3.8.0.

**Lipopolysaccharide-induced inflammation in neutrophils.** To further confirm the outcomes from the functional enrichment analysis, experimental neutrophil data from *Macaca mulatta* were applied. Briefly, neutrophils were collected from the target site (at approximately 130 days of gestation). Inflammation was subsequently induced at this site via lipopolysaccharide (LPS) treatment. Subjects were either treated or not treated with adalimumab at 3 and 1 h before LPS, with samples taken at 16 h after LPS.<sup>26</sup> The original study obtained approval from their local ethics committees. The pre-processed raw RNA-sequencing count data were retrieved from GSE145918. Then, the normalized log<sub>2</sub> + 1 values were calculated using per million reads mapped (CPM) from the count matrix using edgeR version 3.32.0 under R programming environment.

**Statistics.** Statistical analysis was performed using R version 4.0.0. The pROC package version 1.16.2 in R programming environment was applied to conduct receiver operating characteristic curve analysis to evaluate diagnostic accuracy. The statistical significance was evaluated using a Mann–Whitney *U*-test, and Benjamini and Hochberg adjustment was applied for the IBD treatment data. Differences were considered statistically significant at a *P*-value of < 0.05, and < 0.05, < 0.01, < 0.001, and < 0.0001 are indicated with one, two, three, and four asterisks, respectively.

## Results

### Characteristics of studies included in the analysis.

After the keyword searching, removal of ineligible and overlapped datasets from the total of 182 records, 5 transcriptomic data, including GSE16879: from the University Hospital of Gasthuisberg, Belgium, with ClinicalTrials.gov number NCT00639821<sup>27</sup>; GSE23597: the multicenter, randomized, double-blind, placebo-controlled ACT-1 study between March

2002 and March 2005 with ClinicalTrials.gov number NCT00036439<sup>28</sup>; GSE52746: the colonic samples collected between November 2010 and November 2013 from the Department of Gastroenterology, Hospital Clinic of Barcelona, Spain<sup>29</sup>; GSE73661, UC samples collected from two phase III clinical trials of Vedolizumab (VDZ)—GEMINI I and GEMINI LTS at Leuven University Hospitals, Belgium (the dataset included patients received anti-TNF- $\alpha$  blockers)<sup>30</sup>; and GSE92415: the PURSUIT golimumab study conducted in multicenters, with ClinicalTrials.gov number NCT01988961.<sup>31</sup> The five eligible microarray datasets were normalized and combined, and batch effects were corrected (Fig. S1). Eventually, a total of 449 samples, with 17 771 common gene symbols, were included in this study (Table 1).

### Immune microenvironment cell population is significantly higher in nonresponders.

The immune microenvironment scores from both ESTIMATE and xCELL identified that the baseline anti-TNF- $\alpha$  treatment nonresponders are significantly higher than the responders (ESTIMATE: *P* < 0.0001; xCELL: *P* = 0.0003) (Fig. 1a,b). The TNF- $\alpha$  treatment responders showed a significant drop after their treatments (ESTIMATE: *P* < 0.0001; xCELL: *P* = 0.0004) while no significant changes in the nonresponders (ESTIMATE: *P* = 0.0650; xCELL: *P* = 0.11) (Fig. 1a,b). The immune and stroma scores (the two calculation factors for the immune microenvironment) are also significantly higher in baseline anti-TNF- $\alpha$  treatment nonresponders compared with the responders (Fig. S2A–D).

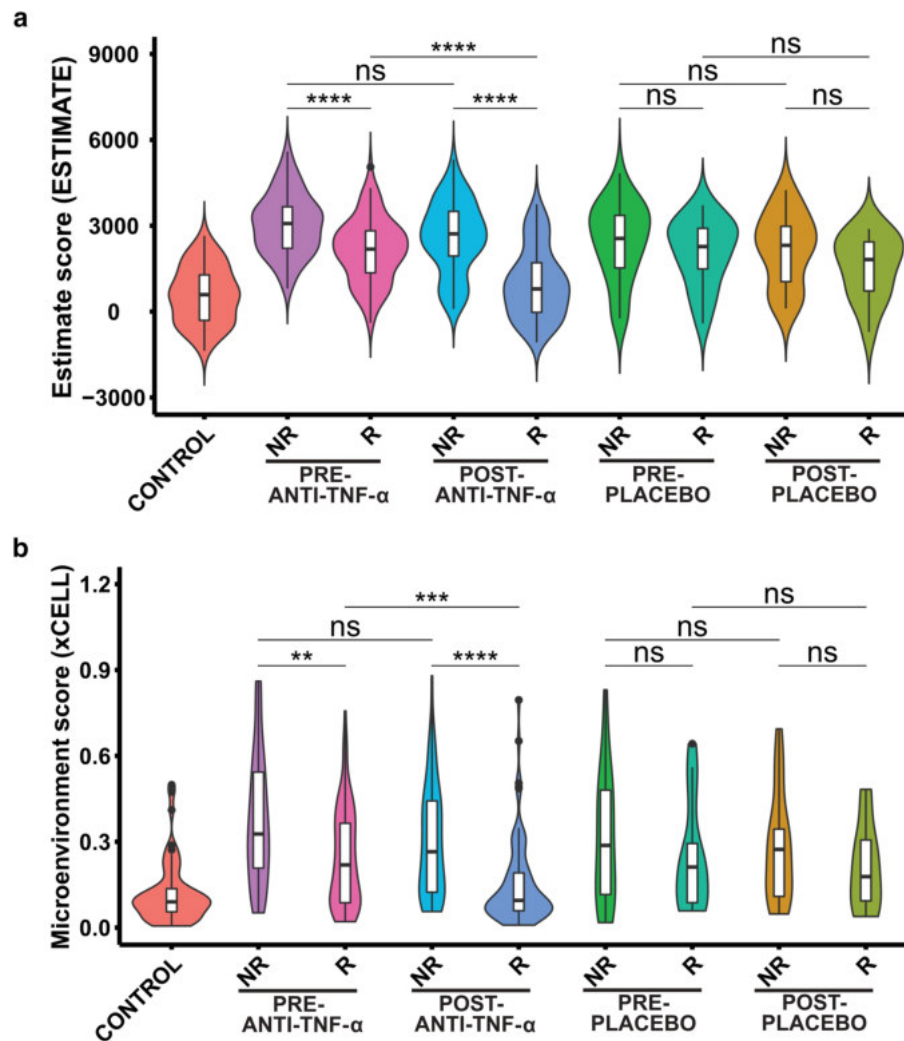
### Neutrophils, endothelial cells, and B cells are significantly higher in nonresponders.

To further our understanding of the immune cell-type composition between the treatment responders and nonresponders, five different *in silico* flow cytometry approaches, including CIBERSORT, xCELL, EPIC, MCP-counter, and DTEIC, were applied (Data S1). Across the algorithms, neutrophils (MCP-counter: *P* < 0.0001; xCELL: *P* = 0.0084; and CIBERSORT: *P* = 0.0021) (Fig. 2a–c), endothelial cells (MCP-counter: *P* = 0.0009; xCELL: *P* = 0.0183; and EPIC: *P* = 0.0337) (Fig. 2d–f), and B cells/B lineage

**Table 1** Summary of the included transcriptomic studies from large intestinal tissues in IBD patients

GSE no. (ref)	Affymetrix platform	Anti-TNF- $\alpha$ drug	Study location	IBD	Pretreatment				Posttreatment				Control	Time point (weeks)
					R		NR		R		NR			
					R	NR	Placebo	Placebo	R	NR	Placebo	Placebo		
16879 <sup>32</sup>	Human Genome U133 Plus 2.0	Infliximab	Belgium	UC	8	16	—	—	8	16	—	—	6	4–6
				CD	12	7	—	—	11	7	—	—		
23597 <sup>33</sup>	Human Genome U133 Plus 2.0	Infliximab	USA	UC	25	7	5	8	20	7	3	6	—	8
52746 <sup>34</sup>	Human Genome U133 Plus 2.0	Infliximab/adalimumab	Spain	CD	6	1	—	—	7	5	—	—	17	12
73661 <sup>35</sup>	Human Gene 1.0 ST	Infliximab	Belgium	UC	8	15	—	—	8	15	—	—	12	6/12
92415 <sup>36</sup>	HT HG-U133+ PM	Golimumab	USA	UC	32	27	11	17	29	21	10	15	21	8
Summary				UC	73	65	16	25	65	59	13	21		
Overall = 449				CD	18	8	—	—	18	12	—	—		
				Total	91	73	16	25	83	71	13	21	56	

CD, Crohn's disease; GSE, Gene Set Enrichment; IBD, inflammatory bowel disease; NR, nonresponder; R, responder; TNF- $\alpha$ , tumor necrosis factor- $\alpha$ ; UC, ulcerative colitis.

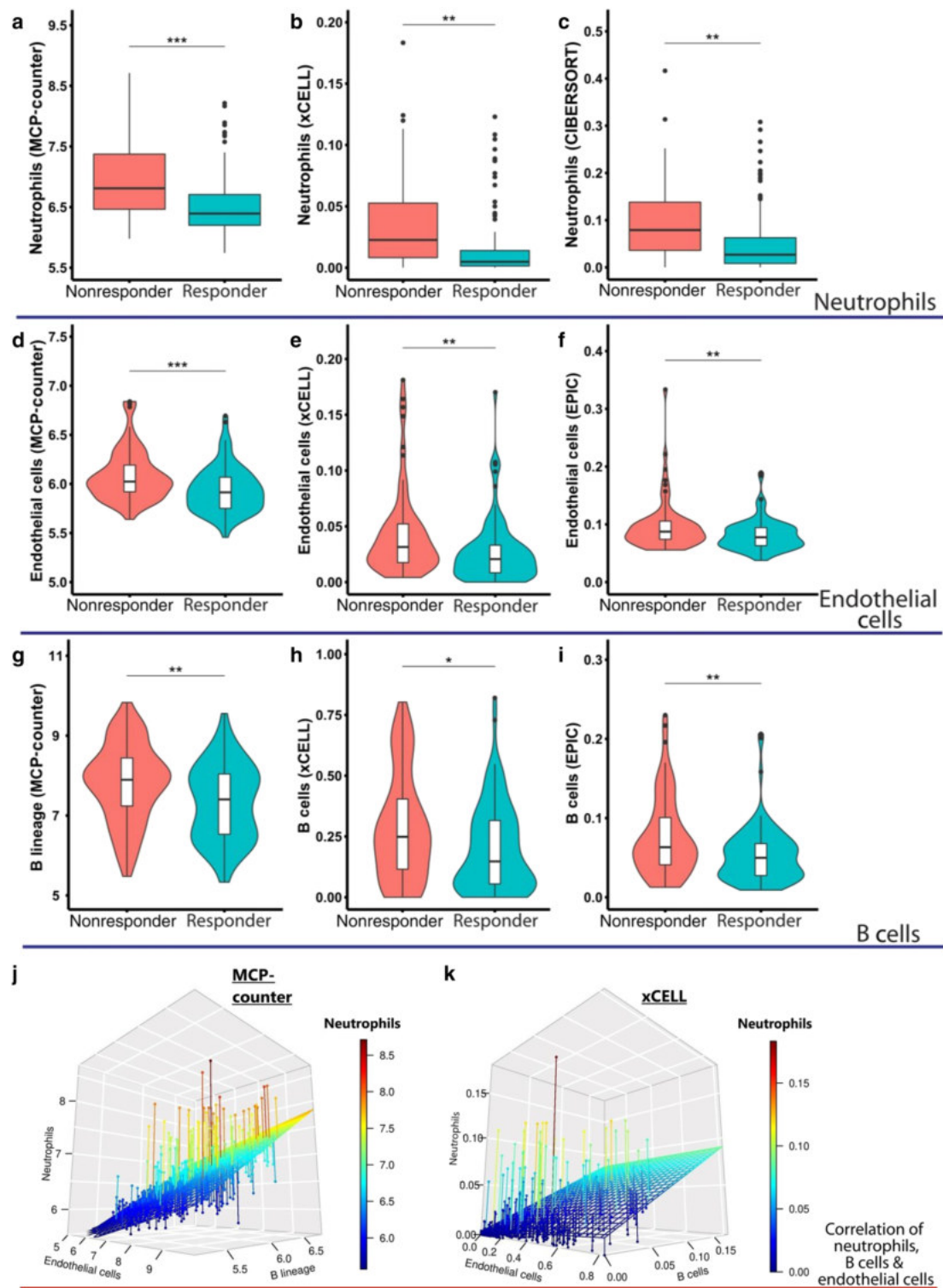


**Figure 1** Microenvironment scores are significantly higher on baseline nonresponders compared with the responders. Immune microenvironment scores evaluated via (a) ESTIMATE and (b) xCELL algorithms. NR, nonresponder; R, responder. The y-axes are the relative immune microenvironment scores from the corresponding algorithms. *P*-value determined by Mann–Whitney *U*-test with Benjamini and Hochberg adjustment. Asterisks denote statistically significant differences (\*\*\**P* < 0.001 and \*\*\*\**P* < 0.0001).

(MCP-counter: *P* = 0.0042; xCELL: *P* = 0.0251; and EPIC: *P* = 0.0042) (Fig. 2g–i) are significantly higher in baseline treatment nonresponders compared with the responders. The three-dimensional plots illustrated that neutrophils, endothelial cells, and B cells from both MCP-counter and xCELL have positive Pearson correlations with each other using all the eligible data (Figs 2j,k and S3A,B).

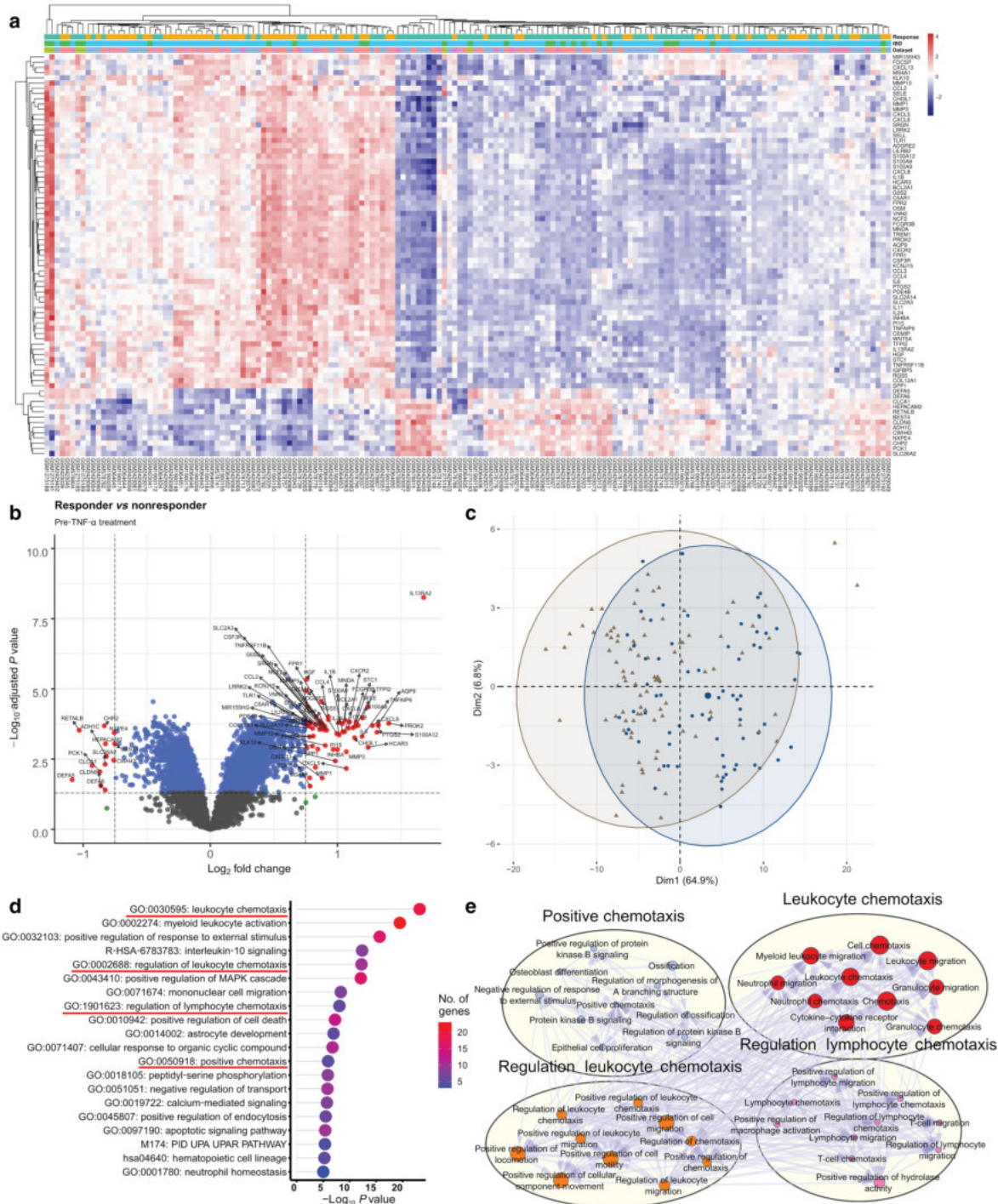
**Hyperactive chemotaxis contributes to anti-tumor necrosis factor- $\alpha$  treatment resistance in inflammatory bowel disease.** The pretreatment anti-TNF- $\alpha$  subjects (responder: *n* = 91 and nonresponder: *n* = 73) were utilized for DEGs analysis and identified a total of 77 DEGs (up-regulated genes = 64 and downregulated genes = 13) (Fig. 3a,b and Data S2). Principal component analysis does not have a

clear separation between responder and nonresponder subjects in the DEGs (Fig. 3c). The differently upregulated genes compose of several gene families, including cytokines (*CCL2*, *CCL3*, *CCL4*, *CXCL13*, *CXCL5*, *CXCL6*, and *CXCL8*), chemokines (*IL1B*, *IL6*, *IL11*, and *IL24*), S100 protein family (*S100A8*, *S100A9*, and *S100A12*), selectin (*SELE* and *SELL*), matrix metalloproteinases (*MMP1*, *MMP3*, and *MMP10*), and formyl peptide receptors (*FPR1* and *FPR2*). Metascape pathway enrichment analysis on the 64 highly expressed genes revealed that GO terms with chemotaxis are commonly found from the outcomes and may have a critical role affecting anti-TNF- $\alpha$  treatment (GO:0030595: leukocyte chemotaxis, GO:0002688: regulation of leukocyte chemotaxis, GO:1901623: regulation of lymphocyte chemotaxis, and GO:0050918: positive chemotaxis) (Fig. 3d,e and Data S3).



**Figure 2** Neutrophils, endothelial cells, and B cells are significantly higher on the baseline anti-TNF- $\alpha$  treatment nonresponders compared with responders. Immune cell population evaluated in five *in silico* flow cytometry, and (a–c) neutrophils, (d–f) endothelial cells, and (g–i) B cells can be recognized in three out of five algorithms. B-cell, endothelial cell, and neutrophil populations are higher on baseline anti-TNF- $\alpha$  nonresponders compared with responders. The three-dimensional plots illustrated that (j, k) neutrophil, endothelial cell, and B-cell populations from MCP-counter and xCELL algorithms have positive correlations with each other. The y-axes are the relative immune cell population abundance from the corresponding algorithms. P-value determined by Mann–Whitney U-test. Asterisks denote statistically significant differences (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , and \*\*\*\* $P < 0.0001$ ).





**Figure 3** Hyperactive chemotaxis may be involved in anti-TNF- $\alpha$  treatment resistance in inflammatory bowel disease. To identify the molecular mechanisms between anti-TNF- $\alpha$  blocker responders ( $n = 91$ ) and nonresponders ( $n = 73$ ), global gene expression analysis was applied from five combined and normalized microarray datasets. The differentially expressed genes were identified and presented using (a) heatmap (the relative expression values were z-score transformed) (response:  $\blacksquare$ , nonresponder;  $\blacksquare$ , responder; IBD:  $\blacksquare$ , CD;  $\blacksquare$ , UC; dataset:  $\blacksquare$ , GSE16879;  $\blacksquare$ , GSE23957;  $\blacksquare$ , GSE52746;  $\blacksquare$ , GSE73661;  $\blacksquare$ , GSE92415), (b) volcano plot, and (c) principal component analysis from a total of 64 upregulated and 13 downregulated genes based on the adjusted  $P$ -value  $< 0.05$  with absolute  $\log_2$  fold change  $\geq 0.75$ . TNF- $\alpha$  treatment:  $\blacksquare$ , nonresponder;  $\blacktriangle$ , responder. (d) The significantly upregulated genes were utilized for Metascape pathway enrichment analysis from the previously pre-defined gene set. Enriched terms related to the chemotaxis-related pathways are underlined. The y-axis represents the top 20 gene set category, the x-axis represents  $-\log_{10}$   $P$ -value, and the color intensity of the bar represents the number of genes identified in each hallmark category. (e) The four subsets of enriched terms under the chemotaxis-related pathways were selected and visualized using Cytoscape.

**Tumor necrosis factor- $\alpha$  blocker does not reduce chemotaxis in lipopolysaccharide-induced inflammation in neutrophils.** To demonstrate our finding in chemotaxis, we used the RNA-sequencing data from an animal study.<sup>26</sup> Briefly, neutrophils isolated from choriodecidua cells with LPS-induced inflammation were treated with or without adalimumab.<sup>26</sup> The list of genes from the four chemotaxis enrichment terms was matched with the neutrophil data to obtain the mean expression values of each sample (GO:0030595: leukocyte chemotaxis [31 out of 44 genes], GO:0002688: regulation of leukocyte chemotaxis [18 out of 22 genes], GO:1901623: regulation of lymphocyte chemotaxis [13 out of 16 genes], and GO:0050918: positive chemotaxis [18 out of 21 genes]) (Data S4). The data process workflow is in Figure S4. All the enrichment terms related to chemotaxis were significantly higher in the LPS-exposed neutrophils; three out of the four enrichment terms do not have significant reduction in the anti-TNF- $\alpha$ -treated group (Fig. 4).

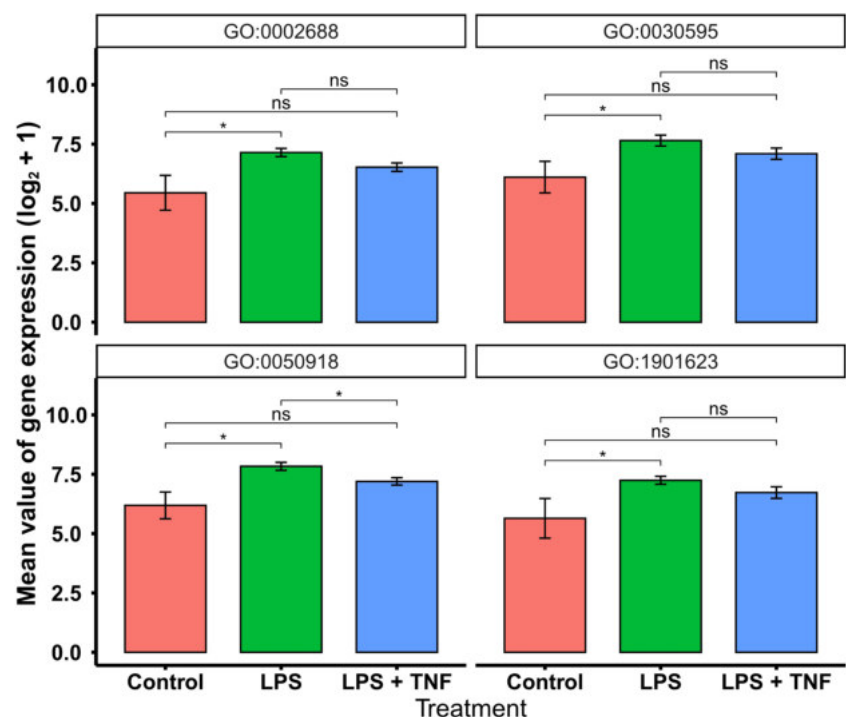
**Interleukin 13 receptor subunit alpha 2 is a diagnostic biomarker to predict tumor necrosis factor- $\alpha$  treatment response.** In order to find the best potential biomarker to predict anti-TNF- $\alpha$  respond IBD patients, receiver operating characteristic curve analysis was applied to all the genes using a for-loop with pROC package under the R programming environment. Among them, *IL13RA2* has the area under the curve of 80.7% (95% confidence interval: 73.8–87.5%) with the best sensitivity of 68.13% and specificity of 84.93% (Table 2, Fig. 5b, and Data S5). *IL13RA2* was stand-alone from the volcano plot ( $\log_2FC$ :1.678, adjusted  $P < 0.0001$ ) (Fig. 2b), and the expression of *IL13RA2* is significantly higher in pretreatment nonresponders compared with the pretreatment responders

( $P < 0.0001$ ). The responders showed a bigger drop after the treatment compared with the nonresponders (responder:  $P < 0.0001$ ; nonresponder:  $P = 0.0037$ ), and the responders restored the expression level to normal control after the treatment (mean  $\pm$  standard deviation: control:  $4.721 \pm 1.039$ ; posttreatment, responder:  $4.868 \pm 1.364$ ;  $P = 0.4969$ ) (Fig. 5a).

## Discussion

Treatment resistance of anti-TNF- $\alpha$  is a critical issue in IBD patients. By integrating the existing raw data and increasing the statistical power, we revealed that immune microenvironment scores are higher in treatment resistance patients on baseline level (Fig. 1a,b), indicating a higher inflammatory burden in anti-TNF- $\alpha$  treatment nonresponders. Further in-depth analysis uncovered that neutrophils, endothelial cells, and B cells contribute to the change in inflammatory burden (Fig. 3). Next, a total of 64 up-regulated genes were identified (Fig. 4a,b), and neutrophil chemotaxis (four out of the top 12 enrichment terms) may contribute to anti-TNF- $\alpha$  treatment resistance in IBD patients (Fig. 4d,e). Utilizing an animal study model, mean expression level (across samples) of genes matching the four chemotaxis GO terms is upregulated in LPS-induced neutrophils but no statistical changes in the three out of the four enrichment terms in the adalimumab-treated group (Fig. 4).

In a typical inflammatory response, immune cells such as macrophages, dendritic cells, natural killer cells, and T lymphocytes release TNF- $\alpha$  pro-inflammatory cytokines, leading to the activation of endothelial cells and neutrophils.<sup>32</sup> The activation of endothelial cells in colonic mucosa enhances vascular permeability and induces the recruitment of immune cells, leading to the activation of chemotaxis. The activation of neutrophils follows the tethering, rolling, crawling, and transmigration process from the blood vessel



**Figure 4** Anti-tumor necrosis factor- $\alpha$  (anti-TNF- $\alpha$ ) blocker does not reduce chemotaxis in lipopolysaccharide (LPS)-induced inflammation in neutrophils. Mean expression level (across samples) of genes matching indicated GO term.  $P$ -value determined by Mann-Whitney  $U$ -test. Asterisks denote statistically significant differences ( $^*P < 0.05$ ).

**Table 2** The top 15 genes to predict TNF- $\alpha$  treatment response

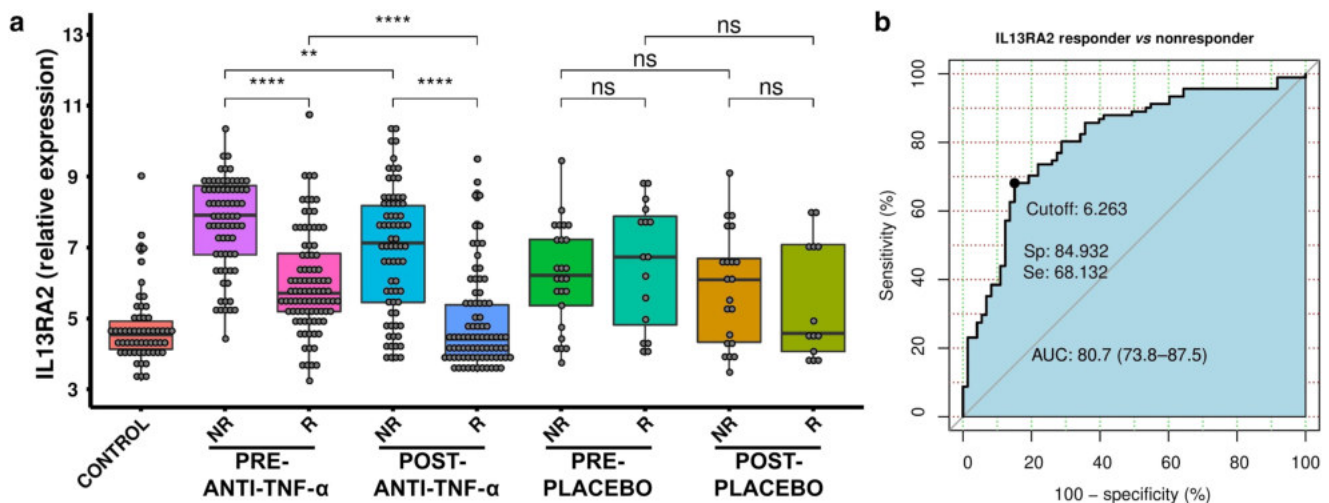
Rank	Gene name	AUC (95% CI)	Threshold	Best specificity	Best sensitivity	LR+	LR-
1	<i>IL13RA2</i>	0.807 (0.738–0.875)	6.263	0.849	0.681	4.510	0.222
2	<i>ADGRE2</i>	0.786 (0.716–0.857)	6.079	0.795	0.703	3.429	0.292
3	<i>ADGRL2</i>	0.771 (0.698–0.843)	6.013	0.753	0.736	2.980	0.336
4	<i>HGF</i>	0.768 (0.695–0.841)	5.598	0.753	0.725	2.935	0.341
5	<i>TLR1</i>	0.764 (0.69–0.838)	5.95	0.836	0.637	3.884	0.257
6	<i>NCF2</i>	0.757 (0.683–0.831)	7.696	0.671	0.769	2.337	0.428
7	<i>RGS5</i>	0.757 (0.682–0.832)	7.105	0.767	0.681	2.923	0.342
8	<i>FPR1</i>	0.756 (0.682–0.83)	6.947	0.836	0.626	3.817	0.262
9	<i>TMTC1</i>	0.754 (0.68–0.828)	6.085	0.849	0.593	3.927	0.255
10	<i>CCL4</i>	0.754 (0.679–0.829)	7.955	0.781	0.659	2.304	0.434
11	<i>PDE4B</i>	0.754 (0.679–0.829)	7.681	0.671	0.758	3.009	0.332
12	<i>IGFBP5</i>	0.754 (0.678–0.829)	7.185	0.712	0.725	2.517	0.397
13	<i>SRGN</i>	0.753 (0.678–0.828)	9.849	0.808	0.615	3.203	0.312
14	<i>PAPPA</i>	0.751 (0.675–0.827)	5.946	0.603	0.835	2.103	0.475
15	<i>TMEM71</i>	0.749 (0.674–0.824)	5.672	0.849	0.615	4.073	0.246

AUC, area under the curve; CI, confidence interval; LR+, positive likelihood ratio; LR-, negative likelihood ratio; TNF- $\alpha$ , tumor necrosis factor- $\alpha$ .

into the inflamed colonic tissues.<sup>33</sup> When neutrophils engulf invasive gut microbiome, they release granule proteins and chromatin to form neutrophil extracellular traps and secrete antimicrobial peptides to mediate extracellular killing of microbial pathogens.<sup>34</sup> However, hyperactive neutrophils trigger an unrestrained activity of the positive feedback amplification loops, leading to endothelial cells and the surrounding tissues damage, inducing resolution delay (IL-6, TNF- $\alpha$ , and IFN- $\gamma$ ) and chemokines (IL-8, CCL3, and CCL4), which further the recruitment of neutrophils, monocytes, and macrophages to the inflamed sites.<sup>35</sup> The use of anti-TNF- $\alpha$  blockers significantly suppresses the infiltration of neutrophil and B-cell population in the inflamed mucosa, and suppresses pro-inflammatory mediators, such as calprotectin (S100A8/A9), IL-8, IL-6, and TNF- $\alpha$  production,<sup>36,37</sup> and matched with our

finding only in responders (Fig. 2a–c,g–i). The unwanted immunogenicity, however, has a high level of B cells due to the presence of antidrug antibodies.<sup>38</sup> The presence of antidrug antibodies neutralizes, interferes, and/or alters the binding efficacy, as well as pharmacodynamic/pharmacokinetic properties of anti-TNF- $\alpha$  monotonical antibodies.<sup>39</sup>

Several S100 calcium-binding protein family genes are highly expressed and previously studied in anti-TNF- $\alpha$  treatment (Fig. 3a,b). Calprotectin is a calcium-binding protein from the S100A8 and S100A9 monomers, representing up to 40% of neutrophil cytosolic proteins and constantly released from the inflamed region(s).<sup>40</sup> *S100A12*, also known as calgranulin C, is released from neutrophils<sup>40</sup> and participates in pro-inflammatory process via the activation of the NF- $\kappa$ B.<sup>41</sup> A small-scale study



**Figure 5** Interleukin 13 receptor subunit alpha 2 (IL13RA2) can be a diagnostic biomarker to predict TNF- $\alpha$  treatment response. (a) The expression of IL13RA2 is significantly higher in the pre-TNF- $\alpha$  nonresponders compared with the pre-TNF- $\alpha$  responders. (b) The expression of IL13RA2 has an area under the curve (AUC) of 80.7% (73.8–87.5%) with a sensitivity (Se) of 68.13% and specificity (Sp) of 84.93%. NR, nonresponder; R, responder. Asterisks denote statistically significant differences (\*\* $P < 0.01$ , \*\*\* $P < 0.001$ , and \*\*\*\* $P < 0.0001$ ). Statistical significance was determined by two-tailed Student's *t*-test.

reported that the fecal calprotectin test (commonly used to distinguish between irritable bowel syndrome and IBD) and *S100A12* may predict relapse after 1 year of infliximab treatment,<sup>42</sup> while another fecal calprotectin study did not find the difference.<sup>43</sup>

*IL13RA2* is stand-alone in the volcano plot with the highest fold change and the lowest *P*-value (Fig. 3b and Data S4) and has the best area under the curve (80.7%, 95% confidence interval: 73.8–87.5%) outcome (Table 2, Fig. 5b, and Data S5). Early studies uncovered that *IL13RA2* is active in mucosal biopsies on the UC or CD anti-TNF- $\alpha$  treatment nonresponders compared with the responders.<sup>44,45</sup> A small-scale study demonstrated that soluble IL13RA2 protein cannot be detected in serum, and tissue expression of IL13RA2 could predict anti-TNF- $\alpha$  treatment in CD patients.<sup>46</sup> IL13RA2 is a decoy receptor enable to bind IL-13 cytokine, diminishes its JAK1/STAT6-mediated effector functions, and activates activator protein 1 (AP-1) to induce the secretion of TGF- $\beta$ .<sup>47,48</sup> The IL-13 pathway is also dependent on the production of TNF- $\alpha$ . Several IL-13 targeting drugs have been tested to inhibit hyperactive immune response on Th2-driven inflammatory diseases.<sup>48</sup> However, insufficient protection was demonstrated by the phase IIa anrukinzumab (an IL-13 monoclonal antibody) clinical trial on UC patients.<sup>49</sup> Thus, blocking the IL-13 pathway via IL13RA2 could be a new approach for treating IBD patients. IL13RA2 knockout mice in DSS-induced acute colitis model showed a better recovery rate compared with the wild-type mice and negatively regulate epithelial/mucosal healing.<sup>50</sup> By neutralizing IL13RA2 in DSS-induced IBD murine model using a monoclonal antibody, it presented a speedy recovery compared with the control group.<sup>47</sup>

The study here identified many strengths but should be considered in the context of shortcomings. Firstly, we only focused on data from large intestine and eliminated ileum data from the Arijis *et al.* study due to the low number of ileum samples that can be integrated,<sup>27</sup> and also reduce the gene expression variation between two different organ sites for downstream analysis. Secondly, our comparison does not include studies from vedolizumab and ustekinumab as it has limited datasets available online. Thirdly, the anti-TNF- $\alpha$  response criteria and the determination time points are slightly different between studies. As we can only rely on the information provided by the authors, and thus, our study has to accept the potential bias. Fourthly, anti-TNF- $\alpha$  is broadly used in colitis-based diseases with a high percentage of treatment failure, and the diagnosis criteria of UC/CD/IBD-U on each of the included studies may be slightly different with a certain percentage of misclassification.<sup>10</sup> Therefore, our priority is to find the common patterns to minimize the treatment resistance rate in this study. Last but not least, the animal study in neutrophils is not from the colonic tissue sites, and some of the chemotactic factor markers such as IL-8/CXCL8 and CSF3 have a significant reduction after the adalimumab treatment.<sup>26</sup> We believed that some single markers may not represent a whole picture of chemotaxis. Thus, in the early future, IBD subtype analysis and more in-depth study in the relation to hyperactive chemotaxis are needed.

In conclusion, nonresponders presented higher populations of neutrophils, endothelial cells, and B cells compared to responders at baseline level. *IL13RA2* is a potential biomarker to predict anti-TNF- $\alpha$  treatment response.

## Availability of data and materials

All the transcriptomics data were retrieved from publicly available datasets from the NCBI Gene Expression Omnibus (GEO) database. All the *in silico* cell sorting algorithms used in this study were based on the default settings recommended by the corresponding authors, either from web-based or retrieved from the corresponding authors' GitHub page for academic research purposes.

## References

- Bocchetti M, Ferraro MG, Ricciardiello F *et al.* The role of microRNAs in development of colitis-associated colorectal cancer. *Int. J. Mol. Sci.* 2021; **22**(8). Available from: <http://www.ncbi.nlm.nih.gov/pubmed/33921348>
- Alatab S, Sepanlou SG, Ikuta K *et al.* The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol. Hepatol.* 2020; **5**(1): 17–30. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2468125319303334>
- Ng SC, Shi HY, Hamidi N *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 2017; **390**(10114): 2769–78. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0140673617324480>
- Kalliolias GD, Ivashkiv LB. TNF biology, pathogenic mechanisms and emerging therapeutic strategies. *Nat. Rev. Rheumatol.* 2016; **12**(1): 49–62. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26656660>
- Gottlieb AB. Tumor necrosis factor blockade: mechanism of action. *J. Investig. Dermatol. Symp. Proc.* 2007; **12**(1): 1–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17502861>
- Gerriets V, Bansal P, Goyal A, Khaddour K. Tumor necrosis factor (TNF) inhibitors *StatPearls* 2020. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29494032>
- Berns M, Hommes DW. Anti-TNF- $\alpha$  therapies for the treatment of Crohn's disease: the past, present and future. *Expert Opin. Investig. Drugs* 2016; **25**(2): 129–43. Available from: <http://www.tandfonline.com/doi/full/10.1517/13543784.2016.1126247>
- Ashton JJ, Mossotto E, Ennis S, Beattie RM. Personalising medicine in inflammatory bowel disease—current and future perspectives. *Transl. Pediatr.* 2019; **8**(1): 56–69. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30881899>
- Lee SH, eun Kwon J Cho M-L. Immunological pathogenesis of inflammatory bowel disease. *Intest. Res.* 2018; **16**(1): 26. Available from: <http://irjournal.org/journal/view.php?doi=10.5217/ir.2018.16.1.26>
- Lamb CA, Kennedy NA, Raine T *et al.* British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut* 2019; **68**(Suppl 3): s1–06. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31562236>
- Ben-Horin S, Kopylov U, Chowers Y. Optimizing anti-TNF treatments in inflammatory bowel disease. *Autoimmun. Rev.* 2014; **13**(1): 24–30. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S156899721300102X>
- Singh S, Fumery M, Sandborn WJ, Murad MH. Systematic review with network meta-analysis: first- and second-line pharmacotherapy for moderate-severe ulcerative colitis. *Aliment. Pharmacol. Ther.* 2018; **47**(2): 162–75. Available from: <https://doi.org/10.1111/apt.14422>
- Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of *Affymetrix GeneChip* data at the probe level. *Bioinformatics* 2004;

- 20(3): 307–15. Available from: <https://doi.org/10.1093/bioinformatics/btg405>
- 14 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; **8**(1): 118–27. Available from: <https://academic.oup.com/biostatistics/article/8/1/118/252073>
  - 15 Leek JT, Johnson WE, Parker HS *et al.* sva: surrogate variable analysis. R package version 3.36.0. 2020. Available from: <https://bioconductor.org/packages/release/bioc/html/sva.html>
  - 16 Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017; **18**: 220. Available from: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1349-1>
  - 17 Yoshihara K, Shahmoradgoli M, Martínez E *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 2013; **4**(1): 2612. Available from: <http://www.nature.com/articles/ncomms3612>
  - 18 Newman AM, Liu CL, Green MR *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 2015; **12**(5): 453–7. Available from: <http://www.nature.com/articles/nmeth.3337>
  - 19 Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife* 2017; **6**. Available from: <https://elifesciences.org/articles/26476>
  - 20 Becht E, Giraldo NA, Lacroix L *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016; **17**(1): 218. Available from: <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1070-5>
  - 21 Chiu Y-J, Hsieh Y-H, Huang Y-H. Improved cell composition deconvolution method of bulk gene expression profiles to quantify subsets of immune cells. *BMC Med. Genomics* 2019; **12**(S8): 169. Available from: <https://bmcmmedgenomics.biomedcentral.com/articles/10.1186/s12920-019-0613-5>
  - 22 Ritchie ME, Phipson B, Wu D *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015; **43**(7): e47. Available from: <http://academic.oup.com/nar/article/43/7/e47/2414268/limma-powers-differential-expression-analyses-for>
  - 23 Blighe K, Rana S, Lewis M. Publication-ready volcano plots with enhanced colouring and labeling 2020. Available from: <https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html>
  - 24 Kolde R. pheatmap: pretty heatmaps 2019. Available from: <https://cran.r-project.org/web/packages/pheatmap/index.html>
  - 25 Zhou Y, Zhou B, Pache L *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 2019; **10**(1): 1523. Available from: <http://www.nature.com/articles/s41467-019-09234-6>
  - 26 Presicce P, Cappelletti M, Sentharamaiah P *et al.* TNF-signaling modulates neutrophil-mediated immunity at the fetomaternal interface during LPS-induced intrauterine inflammation. *Front. Immunol.* 2020; **11**. Available from: <https://www.frontiersin.org/article/10.3389/fimmu.2020.00558/full>
  - 27 Arijis I, De Hertogh G, Lemaire K *et al.* Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment. *PLoS ONE* 2009; **4**(11): e7984. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19956723>
  - 28 Toedter G, Li K, Marano C *et al.* Gene expression profiling and response signatures associated with differential responses to infliximab treatment in ulcerative colitis. *Am. J. Gastroenterol.* 2011; **106**(7): 1272–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21448149>
  - 29 Leal RF, Planell N, Kajekar R *et al.* Identification of inflammatory mediators in patients with Crohn's disease unresponsive to anti-TNF $\alpha$  therapy. *Gut* 2015; **64**(2): 233–42. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24700437>
  - 30 Arijis I, De Hertogh G, Lemmens B *et al.* Effect of vedolizumab (anti- $\alpha$ 4 $\beta$ 7-integrin) therapy on histological healing and mucosal gene expression in patients with UC. *Gut* 2018; **67**(1): 43–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27802155>
  - 31 Telesco SE, Brodmerkel C, Zhang H *et al.* Gene expression signature for prediction of golimumab response in a phase 2a open-label trial of patients with ulcerative colitis. *Gastroenterology* 2018; **155**(4): 1008–11.e8. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S001650851834719X>
  - 32 Balamayooran G, Batra S, Fessler MB, Happel KI, Jeyaseelan S. Mechanisms of neutrophil accumulation in the lungs against bacteria. *Am. J. Respir. Cell Mol. Biol.* 2010; **43**(1): 5–16. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19738160>
  - 33 Wéra O, Lancellotti P, Oury C. The dual role of neutrophils in inflammatory bowel diseases. *J. Clin. Med.* 2016; **5**(12): 118. Available from: <http://www.mdpi.com/2077-0383/5/12/118>
  - 34 Schmidt EP, Lee WL, Zemans RL, Yamashita C, Downey GP. On, around, and through: neutrophil-endothelial interactions in innate immunity. *Physiology* 2011; **26**(5): 334–47. Available from: <https://doi.org/10.1152/physiol.00011.2011>
  - 35 Mortaz E, Alipoor SD, Adcock IM, Mumby S, Koenderman L. Update on neutrophil function in severe inflammation. *Front. Immunol.* 2018; **9**. Available from: <https://www.frontiersin.org/article/10.3389/fimmu.2018.02171/full>
  - 36 Zhang C, Shu W, Zhou G *et al.* Anti-TNF- $\alpha$  therapy suppresses proinflammatory activities of mucosal neutrophils in inflammatory bowel disease. *Mediators Inflamm.* 2018; **2018**: 1–12. Available from: <https://www.hindawi.com/journals/mi/2018/3021863/>
  - 37 Timmermans WMC, van Laar JAM, van der Houwen TB *et al.* B-cell dysregulation in Crohn's disease is partially restored with infliximab therapy. Richard Y, editor. *PLoS ONE* 2016; **11**(7): e0160103. Available from: <https://doi.org/10.1371/journal.pone.0160103>
  - 38 Vaisman-Mentesh A, Rosenstein S, Yavzori M *et al.* Molecular landscape of anti-drug antibodies reveals the mechanism of the immune response following treatment with TNF $\alpha$  antagonists. *Front. Immunol.* 2019; **10**: 2921. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31921180>
  - 39 De Groot AS, Scott DW. Immunogenicity of protein therapeutics. *Trends Immunol.* 2007; **28**(11): 482–90. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17964218>
  - 40 Tardif MR, Chapeton-Montes JA, Posvandzic A, Pagé N, Gilbert C, Tessier PA. Secretion of S100A8, S100A9, and S100A12 by neutrophils involves reactive oxygen species and potassium efflux. *J. Immunol. Res.* 2015; **2015**: 1–16. Available from: <http://www.hindawi.com/journals/jir/2015/296149/>
  - 41 van de Logt F, Day AS. S100A12: a noninvasive marker of inflammation in inflammatory bowel disease. *J. Dig. Dis.* 2013; **14**(2): 62–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23146044>
  - 42 Boschetti G, Garnerio P, Moussata D *et al.* Accuracies of serum and fecal S100 proteins (calprotectin and calgranulin C) to predict the response to TNF antagonists in patients with Crohn's disease. *Inflamm. Bowel Dis.* 2015; **21**(2): 331–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25625487>
  - 43 Laharie D, Mesli S, El Hajbi F *et al.* Prediction of Crohn's disease relapse with faecal calprotectin in infliximab responders: a prospective study. *Aliment. Pharmacol. Ther.* 2011; **34**(4): 462–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21671970>
  - 44 Arijis I, Li K, Toedter G *et al.* Mucosal gene signatures to predict response to infliximab in patients with ulcerative colitis. *Gut* 2009; **58**(12): 1612–9. Available from: <https://doi.org/10.1136/gut.2009.178665>

- 45 Arijis I, Quintens R, Van Lommel L *et al.* Predictive value of epithelial gene expression profiles for response to infliximab in Crohn's disease<sup>‡</sup>. *Inflamm. Bowel Dis.* 2010; **16**(12): 2090–8. Available from: <https://academic.oup.com/ibdjournal/article/16/12/2090-2098/4628294>
- 46 Verstockt B, Verstockt S, Creyngs B *et al.* Mucosal IL13RA2 expression predicts nonresponse to anti-TNF therapy in Crohn's disease. *Aliment. Pharmacol. Ther.* 2019; **49**(5): 572–81. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30663072>
- 47 Karnele EP, Pasricha TS, Ramalingam TR *et al.* Anti-IL-13R $\alpha$ 2 therapy promotes recovery in a murine model of inflammatory bowel disease. *Mucosal Immunol.* 2019; **12**(5): 1174–86. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31308480>
- 48 Hoving JC. Targeting IL-13 as a host-directed therapy against ulcerative colitis. *Front. Cell. Infect. Microbiol.* 2018; **8**: 395. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30460209>
- 49 Reinisch W, Panés J, Khurana S *et al.* Anrukizumab, an anti-interleukin 13 monoclonal antibody, in active UC: efficacy and safety from a phase IIa randomised multicentre study. *Gut* 2015; **64**(6): 894–900. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25567115>
- 50 Verstockt B, Perrier C, De Hertogh G *et al.* Effects of epithelial IL-13R $\alpha$ 2 expression in inflammatory bowel disease. *Front. Immunol.* 2018; **9**: 2983. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30619339>

# Identification of microRNA-135b in Stool as a Potential Noninvasive Biomarker for Colorectal Cancer and Adenoma

Chung Wah Wu, Siew Chien Ng, Yujuan Dong, Linwei Tian, Simon Siu Man Ng, Wing Wa Leung, Wai Tak Law, **Tung On Yau**, Francis Ka Leung Chan, Joseph Jao Yiu Sung, Jun Yu.

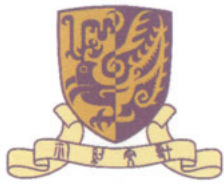
*Clinical Cancer Research*. 2014;20(11):2994–3002.

Journal URL:

[clincancerres.aacrjournals.org/content/20/11/2994](http://clincancerres.aacrjournals.org/content/20/11/2994)

DOI: [10.1158/1078-0432.CCR-13-1750](https://doi.org/10.1158/1078-0432.CCR-13-1750)

PMID: [24691020](https://pubmed.ncbi.nlm.nih.gov/24691020/)



3<sup>rd</sup> November, 2015

To Whom It May Concern,

**Statement of Joint Authorship**

**Title of publication:** Identification of microRNA-135b in stool as a potential noninvasive biomarker for colorectal cancer and adenoma

**Journal:** Clinical Cancer Research

**Publication date:** 1<sup>st</sup> June 2014

**Issue:** Volume 20, Issue 11, Pages 2994 – 3002

**PubMed ID:** 24691020

**Authors:** Chung Wah Wu, Siew Chien Ng, Yujuan Dong, Linwei Tian, Simon Siu Man Ng, Wing Wa Leung, Wai Tak Law, Tung On Yau, Francis Ka Leung Chan, Joseph Jao Yiu Sung, Jun Yu

I hereby confirm that Mr. Tung On YAU is a co-author in the above publication. He was a significant contributor to this project, assisting with experimental design, and data collection, as well as carrying out experiments. This included the acquisition of stool samples, collection and organisation of patient information, stool imRNA extraction and purification, as well as performing the TaqMan Probe based quantitative Real-Time PCR experiments. He also assisted with the subsequent analysis of results, and writing up of the paper.

Yours sincerely,

Jun Yu, MD, PhD

Professor

Institute of Digestive Disease

Prince of Wales Hospital

The Chinese University of Hong Kong, Hong Kong

Tel: 852-3763 6099; Fax: 852-2144 5330

Email: junyu@cuhk.edu.hk



## Identification of microRNA-135b in Stool as a Potential Noninvasive Biomarker for Colorectal Cancer and Adenoma

Chung Wah Wu<sup>1,4</sup>, Siew Chien Ng<sup>1</sup>, Yujuan Dong<sup>1,2</sup>, Linwei Tian<sup>3</sup>, Simon Siu Man Ng<sup>2</sup>, Wing Wa Leung<sup>2</sup>, Wai Tak Law<sup>1</sup>, Tung On Yau<sup>1,4</sup>, Francis Ka Leung Chan<sup>1</sup>, Joseph Jao Yiu Sung<sup>1</sup>, and Jun Yu<sup>1,4</sup>

### Abstract

**Purpose:** Detecting microRNA (miRNA) in stool is a novel approach for colorectal cancer (CRC) screening. This study aimed to identify stool-based miRNA as noninvasive biomarkers for detection of CRC and adenoma.

**Experimental Design:** A miRNA expression array covering 667 human miRNAs was performed on five pairs of CRC and two pairs of advanced adenoma tissues. The most upregulated miRNAs were validated in 40 pairs of CRC tissues, 16 pairs of advanced adenoma tissues, and 424 stool samples, including 104 CRCs, 169 adenomas, 42 inflammatory bowel diseases (IBD), and 109 healthy controls. miRNA levels were followed-up after removal of lesions.

**Results:** In an array analysis, miR-31 and miR-135b were the most upregulated miRNAs in CRC and advanced adenoma as compared with their adjacent normal tissues (>13-fold increase). In stool samples, level of miR-135b was significantly higher in subjects with CRC ( $P < 0.0001$ ) or adenomas ( $P < 0.0001$ ), but not in patients with IBD compared with controls. miR-135b showed a significant increasing trend across the adenoma to cancer sequence ( $P < 0.0001$ ). Levels of miR-31 were not significantly different among groups. The sensitivity of stool miR-135b was 78% for CRC, 73% for advanced adenoma, and 65% for any adenoma, respectively, with a specificity of 68%. No significant difference in the miR-135b level was found between proximal and distal colorectal lesions. Stool miR-135b dropped significantly upon removal of CRC or advanced adenoma ( $P < 0.0001$ ).

**Conclusion:** Stool-based miR-135b can be used as a noninvasive biomarker for the detection of CRC and advanced adenoma. *Clin Cancer Res*; 20(11); 2994–3002. ©2014 AACR.

### Introduction

Colorectal cancer (CRC) is the third most common cancer worldwide (1). Although the incidence of CRC is declining in the West, it remains the second most common overall cause of cancer death. Both the incidence and death rates from CRC are increasing rapidly in Asian countries (2). CRC screening allows the detection and removal of early-

stage lesions, and has been demonstrated to reduce both CRC morbidity and mortality (3, 4). However, mass screening efforts have been hindered by variable public acceptance and the limitations of existing tools. Colonoscopy, although considered a gold-standard test for CRC screening, is associated with high cost and relatively low patient acceptance rate. The fecal occult blood test (FOBT) is the most widely adopted screening method, but is compromised by a low sensitivity and specificity and poor patient adherence. The newer fecal immunochemical tests (FIT) have demonstrated a higher sensitivity for CRC, but sensitivity remains low for premalignant precursor lesions (5). CT colonography is less invasive and accurate for the detection of CRC and advanced adenomas but is associated with high cost (6–9). Therefore, there is a need to develop better screening tools to avoid nontherapeutic colonoscopies for adenomas or colonoscopies to confirm CRC.

Testing for molecular aberrations in the stool has emerged as a promising noninvasive approach for CRC screening. Among stool-based molecular tests, DNA testing is the most established test (10–12). A DNA panel which combined four methylation markers, seven reference mutations,  $\beta$ -actin, and a hemoglobin assay, achieved a sensitivity

**Authors' Affiliations:** <sup>1</sup>Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences; <sup>2</sup>Department of Surgery; <sup>3</sup>Stanley Ho Center for Emerging Infectious Disease, the Chinese University of Hong Kong, Hong Kong; and <sup>4</sup>Gastrointestinal Cancer Biology and Therapeutics Laboratory, CUHK-Shenzhen Research Institute, Shenzhen, China

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

C.W. Wu and S.C. Ng contributed equally to this article.

**Corresponding Author:** Jun Yu, Institute of Digestive Disease and Department of Medicine and Therapeutics, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China. Phone: 852-3763-6099; Fax: 852-2144-5330; E-mail: junyu@cuhk.edu.hk

doi: 10.1158/1078-0432.CCR-13-1750

©2014 American Association for Cancer Research.

### Translational Relevance

Colorectal cancer (CRC) screening allows the detection and removal of early-stage lesions, and has been demonstrated to reduce both CRC morbidity and mortality. Compared with DNA, microRNA (miRNA) represents an emerging class of biomolecule being utilized as stool-based marker for CRC screening. In the current study, we demonstrated for the first time, in a large cohort of patients with CRC or advanced adenoma that miRNA (miR-135b), elevated in tumor tissue and stool samples, can be used as stool-based biomarker for CRC as well as adenomas. We revealed that the detection of miR-135b in stool has a sensitivity of 78% for CRC and 73% for advanced adenoma, thus stool-based miR-135b can be used as a potential noninvasive biomarker for the diagnosis of CRC and adenoma.

of 85% for CRC and 54% for adenoma  $\geq 1$  cm. Each gene typically yielded an area under the curve (AUC) value ranging from 0.61 to 0.75 toward CRC.

microRNAs (miRNA) are short noncoding RNAs that regulate gene translation (13). Most tumor types, including CRC, are found to have altered miRNA expression profiles (14–17). As tumor cells shed from CRC tumor surface into the lumen, aberrantly expressed miRNA levels can be detected in the stool. We have previously demonstrated that stool miRNAs (miR-21 and miR-92a) are stable and reproducible in a small number of the samples using an optimized RNA extraction and quantitative reverse transcription PCR (qRT-PCR) protocol (18). This provides a rationale of utilizing stool miRNA for the detection of CRC and adenomas. So far, four other studies have reported the use of stool-based miRNA as a screening tool for CRC (19–22). Most of these studies involved small cohorts and were limited to the investigation of CRC.

In this study, we aimed to identify miRNA markers with higher sensitivity, particularly for the diagnosis of precancerous colorectal lesions in additional consecutive stool samples particularly from subjects with advanced adenomas. We investigated the miRNA expression profile in CRCs and in advanced adenomas to identify the most upregulated miRNAs. Candidate miRNA markers were validated in 40 paired primary CRC tumors and 16 paired advanced adenoma tissues, and then in a large cohort of 424 stool samples of 104 CRC, 169 adenomas, 42 inflammatory bowel disease (IBD) as disease control, and 109 healthy subjects as normal control. We identified and characterized that stool-based miR-135b accurately detects CRC and adenoma in this large case-control study.

## Materials and Methods

### Subjects and stool sample collection

Stool samples were collected from 424 subjects, including 104 patients with CRC (mean age,  $66.8 \pm 11.9$  years), 59 patients with advanced adenomas ( $62.1 \pm 9.5$  years), 110

subjects with adenomas of less than 1 cm in size ( $58.9 \pm 6.9$  years), 42 subjects with IBD ( $48.2 \pm 11.6$  years), and 109 individuals who had a normal colonoscopy ( $60.4 \pm 7.0$  years; Table 1). We have included patients with IBD as a control group to demonstrate that the markers are specific to CRC and precancerous adenoma but not to common inflammatory intestinal diseases. The mean age of the CRC group was significantly older than the control group ( $P < 0.0001$ ), whereas subjects with IBD were significantly younger than the control group ( $P < 0.0001$ ). There were more males in the CRC than control group (58% vs. 46%;  $P < 0.0001$ ; Table 1). Exclusion criteria included subjects with a family history of familial adenomatous polyposis or hereditary nonpolyposis CRC, previous colonic surgery, or adjuvant therapy for CRC before surgery. Patients who were passing liquid stool were also excluded (Supplementary Fig. S1). All patients were recruited from Prince of Wales Hospital and Alice Ho Miu Ling Nethersole Hospital (Hong Kong, China). All participants had signed informed consent for obtaining stool or tissue samples. The study protocol was approved by the Institutional Review Board of the Chinese University of Hong Kong and the Hong Kong Hospital Authority (Hong Kong, China).

All cancer stool samples were collected 7 days after colonoscopy. All normal and adenoma stool samples were collected before colonoscopy. Fresh human stool samples were collected from patients using a 30-mL universal sample container (height, 93 mm; cap diameter, 30.1 mm) with spoon cap. The container was aseptically manufactured under clean room condition to exclude microbiological contamination. All samples were stored at  $4^{\circ}\text{C}$  immediately and transferred to  $-80^{\circ}\text{C}$  within 24 hours. Stools were collected before bowel purgation and colonoscopy or 1 week after colonoscopy (but before resection of CRC or removal of advanced adenoma).

To investigate the changes in stool miRNA levels after removal of CRC or advanced adenomas, repeat stool collection was performed at least one month after surgical removal of CRC or at least 7 days after removal of the advanced adenoma. Advanced adenomas were defined as adenomas  $\geq 1$  cm in diameter, adenomas with villous or tubulovillous features, or high-grade dysplasia. Proximal lesions included lesions at or proximal to the splenic flexure, and distal lesions were lesions distal to the splenic flexure.

### Tissue collection

Forty pairs of CRC tissues and 16 pairs of advanced adenoma tissues were collected (Supplementary Table S1). CRC and advanced adenoma tissues and their respective adjacent normal tissues (at least 4 cm apart from the lesion) were biopsied during the initial colonoscopy or during surgical resection. "Paired lesions" refers to a lesion paired with adjacent normal mucosal sample. Tissue samples were snap frozen upon collection and stored in  $-80^{\circ}\text{C}$  freezer.

### miRNA extraction in tissue and stool samples

Frozen tissue of 10 to 20  $\mu\text{g}$  was added into 0.5 mL TRIzol reagent (Invitrogen) in a 1.5-mL tube. The tissue was

**Table 1.** Clinicopathological characteristics of subjects

Category	Normal	Adenoma < 1 cm	Advanced adenoma <sup>a</sup>	CRC	IBD
No. of cases	109	110	59	104	42
Age at enrollment, y					
Mean ± SD	60.4 ± 7.0	58.9 ± 6.9	62.1 ± 9.5	66.8 ± 11.9	48.2 ± 11.6
Gender, number (%)					
Female	59 (54.1)	51 (46.4)	29 (49.2)	44 (42.3)	16 (38.1)
Location <sup>b</sup> , number (%)					
Proximal	—	32 (29.1)	26 (44.1)	29 (27.9)	—
Distal	—	65 (59.1)	31 (52.5)	75 (72.1)	—
Both	—	13 (11.8)	2 (3.4)	0 (0.0)	—
Tumor histology, number (%)					
Adenocarcinoma	—	—	—	97 (93.3)	—
Mucinous adenocarcinoma	—	—	—	6 (5.8)	—
Signet ring cell and mucinous	—	—	—	1 (0.9)	—
TNM stage <sup>c</sup> , number (%)					
I and II	—	—	—	24 (23.1)	—
III and IV	—	—	—	76 (73.1)	—

<sup>a</sup>Advanced adenoma is defined as adenoma 1 cm or greater in diameter, adenoma with more than 25% villous feature, or adenoma with high-grade dysplasia.

<sup>b</sup>Proximal lesions include tumors at or proximal to the splenic flexure, and distal lesions are those distal to the splenic flexure.

<sup>c</sup>TNM stage data of four patients were unavailable.

homogenized by RNase-free pestles. Chloroform of 200  $\mu$ L was added to the 1.5 mL tube.

Fresh human stool sample (20–30 g) was collected with a 50-mL specimen cup and stored in  $-80^{\circ}\text{C}$ . Four categories of stool consistency were defined: "firm" (the stool has clear-cut edges, maintains its own shape during handling but deforms with pressure), "soft" (the stool has a uniform consistency but few or less apparent natural edges, it maintains its own shape but deforms with minimal handling), "loose" (the stool has a semisolid consistency and can take over the shape of the container), "watery" (no solid pieces, completely liquid). Only "firm," "soft," and "loose" stool samples were analyzed. More than 90% stool samples in each group belong to soft or loose. Stool sample of 200 to 300 mg (wet weight) was added to 1 mL TRIzol LS reagent in a 2-mL tube (Invitrogen), and homogenized mechanically by RNase-free pestles (USA Scientific) to deform it completely. Chloroform of 300  $\mu$ L was added to the 2 mL tube.

Total RNA was extracted from the TRIzol-chloroform mixture using the miRNeasy Mini Kit (Qiagen). Total RNA was eluted in 50  $\mu$ L nuclease-free water. RNA concentration was measured by Nanodrop 2000 (Thermo Fisher Scientific; the range of the initial RNA concentration RNA concentration was 300 to 500 ng/ $\mu$ L).

#### Reverse transcription and miRNA microarray in CRC and advanced adenoma tissues

RT for miRNA microarray was carried out using Megaplex Primer pools, Human Pools A and B v2.1 Kit (Applied Biosystems). Briefly, 2 ng total RNA was used in one RT

reaction with a total volume of 3  $\mu$ L. The RT product was diluted 4-fold by adding 9  $\mu$ L nuclease-free water (18).

Initial miRNA profiling was performed on 14 tissue samples from five pairs of CRC and two pairs of advanced adenoma. The characteristics of these patients are shown in Supplementary Table S2. Quantitation of 667 miRNAs in each of these samples was carried out using TaqMan Human MicroRNA Array Set version 2.0 (Applied Biosystems). In the array, miRNAs were normalized to mammalian U6 small RNA expression as based on the manufacturer's guide (Applied Biosystems). qRT-PCR was performed using Applied Biosystems 7900HT Real-Time PCR System. Results were analyzed by the SDS RQ Manager 1.2 software (Applied Biosystems; Supplementary Table S3).

#### miRNA quantitation by qRT-PCR

qRT-PCR of individual miRNA was performed using TaqMan miRNA Reverse Transcription Kit (Applied Biosystems) and TaqMan Human MiRNA Assay (Assay ID: RNU6B, 001093; miR-31, 002279; miR-135b, 002261). The quantitation of miRNA was based on standard curve plotted by known amount of synthetic miRNA, and normalized to per nanogram (ng) of input RNA (18). Assays were performed in a blinded fashion. On the basis of standard curves plotted from known amount of synthetic miR-135b, a technical detection limit of three copies of miR-135b would approximately give a  $C_t$  of 42. Therefore, we assigned all  $C_t$  larger than 42 as "0." For samples with no amplification of miR-135b at all, as long as that sample could be amplified with at least another miRNA tested (such as miR-18a, miR-20a, and miR-221), the sample was regarded to

have legitimate quality for qRT-PCR. Therefore, instead of excluding unamplifiable samples, we assigned the sample as "0" in the analysis of miR-135b.

### Sample size and statistics

Given the exploratory aspect of this initial miRNA profiling and limitations of resources, we did not use a formal statistical test to choose a sample size, and we analyzed five pairs of CRC samples and two pairs of advanced adenoma sample for initial miRNA profiling. In our validation study, we analyzed 40 pairs of CRCs and 16 pairs of advanced adenoma tissue samples. Differences in miRNA expression between paired lesion tissues and adjacent normal tissues were evaluated by the Wilcoxon matched pair test. Differences in stool miRNA levels between groups were analyzed by the Mann-Whitney *U* test. Receiver operating characteristics (ROC) curves were generated on the basis of the comparison with the control group. Differences in miRNA levels before and after removal of the CRC or advanced adenoma were determined by the Wilcoxon matched pair test. Significance in trend was tested by the Jonckheere-Terpstra test. Two cutoff values were selected using ROC curves for reference, based on a high sensitivity or a high specificity.  $P < 0.05$  was taken as statistically significant. The

Jonckheere-Terpstra test was done by SPSS 13.0 (SPSS Inc.). All other statistical tests were performed using GraphPad Prism 5.0 (GraphPad Software Inc.).

### Results

#### miRNA profiling in CRC and advanced adenoma tissues

miRNA expression profiles were performed in five pairs of CRC samples and two pairs of advanced adenoma samples. The miRNA expression levels were compared between the tumor and their adjacent nontumorous tissues in each case (Table 2). Amongst 667 miRNAs detected in each sample, miR-31 and miR-135b were identified to be the most upregulated miRNAs in both CRC (miR-31, 42.28-fold increase; miR-135b, 13.00-fold increase) and advanced adenomas (miR-31, 106.36-fold; miR-135b, 13.32-fold; Table 2). We have therefore focused on these two miRNA in subsequent experiments.

#### Validation of miRNA candidates in CRC and advanced adenoma tissues

In a validation set of 40 pairs of CRCs and 16 pairs of advanced adenoma tissue samples, miR-135b was demonstrated to be 555.4-fold higher in CRCs ( $P < 0.0001$ ) and

**Table 2.** Top 10 most upregulated miRNAs identified by profiling of 667 miRNAs in five patients with CRC and two patients with advanced adenoma

Rank	miRNA	Accession number	Chromosomal location	Average fold increase <sup>a</sup>
miRNAs with highest expression in colorectal tumor				
1	miR-31	MI0000089	9p21.3	42.28
2	miR-135b	MI0000810	1q32.1	13.00
3	miR-224	MI0000301	Xq28	5.48
4	miR-409-3p	MIMAT0001639	14q32.31	4.67
5	miR-18a	MI0000072	13q31.3	4.46
6	miR-452	MI0001733	Xq28	4.16
7	miR-221	MI0000298	Xp11.3	3.95
8	miR-21	MI0000077	17q23.1	3.39
9	miR-223	MI0000300	Xq12	3.18
10	miR-20a	MI0000076	13q31.3	3.02
miRNAs with highest expression in advanced adenoma				
1	miR-31	MI0000089	9p21.3	106.36
2	miR-135b	MI0000810	1q32.1	13.32
3	miR-20a-3p	MIMAT0004493	13q31.3	4.30
4	miR-182	MI0000272	7q32.2	3.99
5	miR-649	MI0003664	22q11.21	3.93
6	miR-26a-1-3p	MIMAT0004499	3p22.2	3.33
7	miR-625	MI0003639	14q23.3	3.13
8	miR-18a	MI0000072	13q31.3	2.78
9	miR-20a	MI0000076	13q31.3	2.75
10	miR-552	MI0003557	1p34.3	2.17

<sup>a</sup>Average fold change in colorectal tumor was expressed as ratio of the miRNA expression in tumor to adjacent nontumorous tissue. Average fold change in advanced adenoma was expressed as ratio of miRNA expression in advanced adenoma to adjacent nonadenomatous tissue.

**Table 3.** Expression of miR-135b and miR-31 in colorectal tumor or advanced adenoma

miRNA	Percentage of samples with elevated expression in tumor or advanced adenoma	Average fold increase	<i>P</i> <sup>a</sup>
CRC ( <i>n</i> = 40)			
miR-135b	92.5% (37/40)	555.4	<0.0001
miR-31	87.5% (35/40)	105.1	<0.0001
Advanced adenoma ( <i>n</i> = 16)			
miR-135b	93.8% (15/16)	33.1	0.0003
miR-31	87.5% (14/16)	86.1	0.0003

<sup>a</sup>*P* values were estimated by the Wilcoxon matched pair test.

33.1-fold higher in advanced adenoma tissues (*P* = 0.0003), whereas miR-31 was 105.1-fold higher in CRCs (*P* < 0.0001) and 86.1-fold higher in advanced adenoma (*P* = 0.0003) compared with their corresponding normal tissues (Table 3).

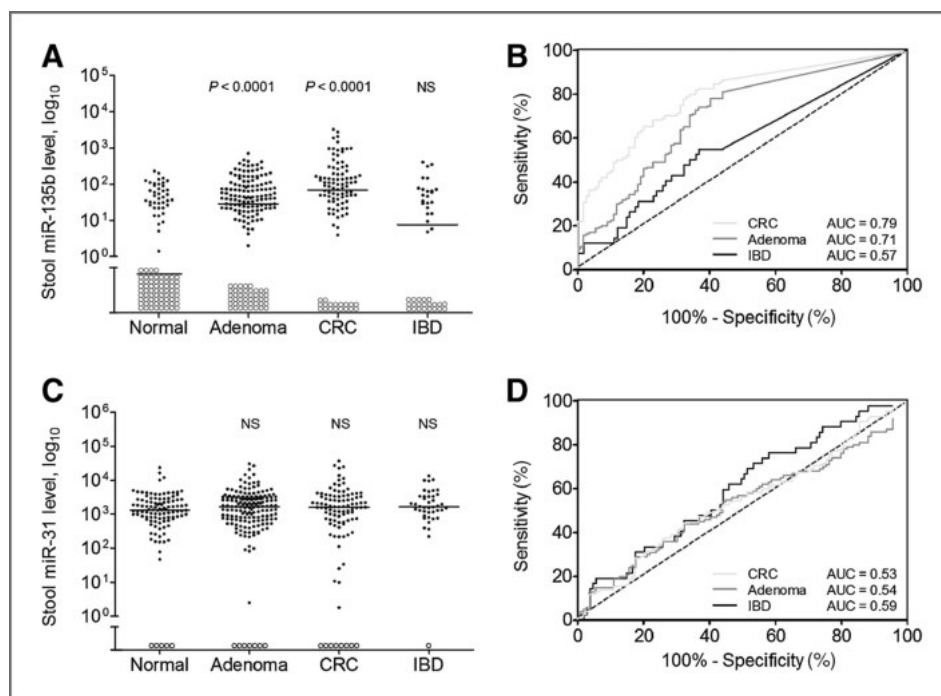
#### miR-135b is a potential noninvasive stool marker for CRC and adenoma

We then examined the levels of miR-135b and miR-31 in 424 stool samples, which included 104 CRC, 169 adenomas, 42 IBD, and 109 controls. The stool-based miR-135b level measured in a number of copies/ng stool extracted RNA was significantly higher in subjects with CRC [median,

67.9, interquartile range (IQR), 16.1–182.7; *P* < 0.0001] and adenomas (median, 28.4; IQR, 0.2–79.7; *P* < 0.0001) compared with controls (median, 0; IQR, 0–30.8; Fig. 1A). In contrast, there was no significant difference in the level of stool miR-135b in subjects with IBD (median, 7.57; IQR, 0–60) compared with controls. The AUC values for miR-135b were 0.79 and 0.71 for the detection of CRC and adenomas, respectively (Fig. 1B). As shown in Table 4, two cutoff values demonstrated the performance of this marker: a cutoff of 14 copies/ng of stool RNA provided the maximum sum of sensitivity and specificity; miR-135b has a sensitivity of 78% for CRC, 73% for advanced adenoma, 61% for adenoma < 1 cm in diameter, 65% for any adenoma, and a specificity of 68%. A cutoff of 38 copies/ng of stool RNA reflects its performance at a relatively high specificity (80%) level for reference, and the sensitivity was 44%, 46%, and 64% for adenoma < 1 cm, advanced adenoma, and CRC, respectively (Table 4). However, there were no significant differences in the levels of stool miR-31 among CRC (median, 1,583; IQR, 574.5–3,364), adenomas (median, 1,647; IQR, 661.9–3,148), IBD (median, 1,642; IQR, 1,066–3,345), and controls (median, 1,293; IQR, 721–2,612; Fig. 1C and 1D).

#### Stool miR-135b level is significantly reduced after removal of neoplasm

We repeated miRNA measurement in a subgroup of 28 patients (8 CRC, 20 advanced adenomas) after the removal of CRC or advanced adenomas. Upon removal of the primary lesions, levels of stool miR-135b dropped significantly compared with their initial levels (*P* < 0.0001; Fig. 2A). These findings suggested that the initial high levels of stool miR-135b were derived from the primary neoplasms.



**Figure 1.** Levels of miRNA markers in stool samples. A, miR-135b and (B) its ROC curve, (C) miR-31 and (D) its ROC curve. Subjects were categorized into four groups: individuals of normal colonoscopy results (*n* = 109), adenoma (*n* = 169), CRC (*n* = 104), and IBD (*n* = 42), respectively. The miRNA levels were expressed in the number of copies per nanogram of extracted RNA. Open circles represent samples with undetectable miRNA level. The lines denote the medians. *P* denotes significance measured by the Mann-Whitney test. NS denotes no statistical significance. ROC curves were plotted to discriminate CRC, adenoma, or IBD patients from individuals with normal colonoscopy findings.

**Table 4.** Stool-based miR-135b sensitivities and specificities

Cutoff value <sup>a</sup>	14 (high sensitivity)	38 (high specificity)
Sensitivities, % (95% CI)		
Adenoma	65 (57–72)	45 (37–53)
Adenoma <1 cm	61 (51–70)	44 (34–53)
Advanced adenoma	73 (60–84)	46 (33–60)
Size		
1–1.9 cm	76 (55–91)	40 (21–61)
≥2 cm	83 (52–98)	75 (43–95)
Location		
Proximal	65 (44–83)	38 (20–59)
Distal	81 (63–93)	52 (33–70)
CRC	78 (69–85)	64 (54–74)
Stage		
TNM I, II	67 (45–84)	50 (29–71)
TNM III, IV	80 (70–89)	67 (55–77)
Location		
Proximal	79 (60–92)	59 (39–76)
Distal	77 (66–86)	67 (55–77)
Specificity, % (95% CI)	68 (58–77)	80 (71–87)

<sup>a</sup>Cutoff value was expressed as copies of the miRNA per nanogram of extract stool RNA. Two cutoff values were selected using ROC curves for reference.

### Stool miR-135 level increases across the adenoma to cancer sequence

As shown in Fig. 2B, expression of stool miR-135 showed a significantly increasing trend across the disease transition from adenoma with diameter < 1 cm ( $n = 110$ ), advanced adenoma ( $n = 59$ ), tumor-node-metastasis (TNM) stage I and II cancer ( $n = 24$ ) to TNM stage III

and IV cancer ( $n = 76$ ) sequence ( $P < 0.0001$ ). In keeping with this, an increasing sensitivity for stool-based miR-135 detection was demonstrated with 61% for adenomas with diameter < 1 cm, 73% for advanced adenoma, 67% for TNM stage I and II cancer, and 80% for TNM stage III and IV cancer, and a specificity of 58.4%. On the basis of Table 3, the increase fold change of miR-135b in CRC tissue samples is approximately 17-fold higher than in advanced adenoma tissue samples.

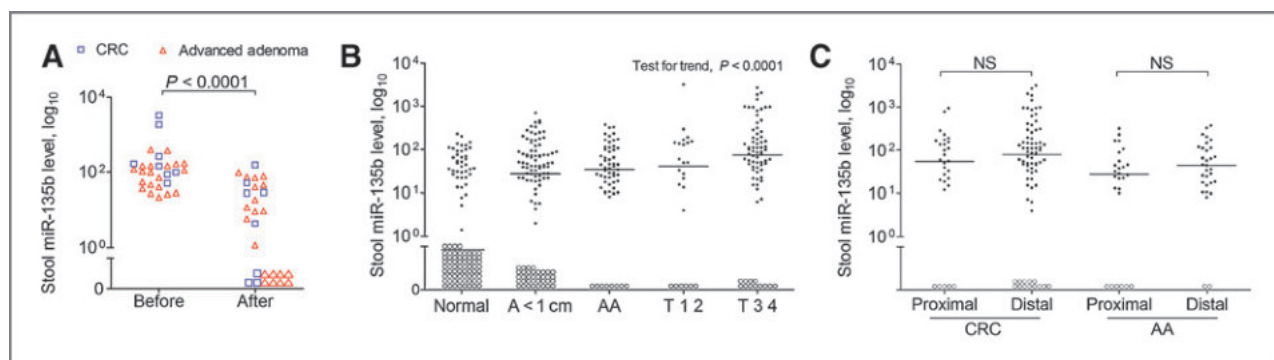
### Stool miR-135 level is not associated with the location of CRC or advanced adenoma

We compared the level of stool miR-135 between distal and proximal lesions. No significant difference was found in miR-135b levels for CRC or advanced adenomas located in proximal or distal colon (Fig. 2C), although at a specific cutoff value (14 copies/ng RNA), miR-135b showed better sensitivity in detecting distal advanced adenoma (81%) compared with proximal ones (65%), with specificity of 68% for both distal and proximal advanced adenoma (Table 4).

### Discussion

As CRC demonstrates high homogeneity in miRNA alteration, stool miRNA could represent a useful noninvasive tool for screening CRC and its precancerous lesions (18, 22, 23). For maximum benefit, a CRC screening tool should be effective in detecting CRC in its early stage and premalignant precursors throughout the entire colorectum. In this study, we performed systematical examinations with the aim to identify the best miRNA biomarker for the screening of CRC and advanced adenomas.

We first identified that miR-135b and miR-31 were the two most upregulated miRNAs both in CRCs and adenomas by miRNA expression arrays. In subsequent validation,



**Figure 2.** Associations between stool miR-135b and clinical features. A, changes of the stool miR-135b level after removal of tumors (squares) or advanced adenomas (triangles). Samples aligned on the x-axis represent those with undetectable miRNA level.  $P$  values indicate significant difference determined by the Wilcoxon matched pairs test. The stool miR-135b level in subjects categorized by lesion stage (B) and lesion location (C). On the basis of lesion stage, subjects were categorized into five groups: individuals of normal colonoscopy (normal,  $n = 109$ ), patients with adenoma of diameter smaller than 1 cm (A < 1 cm,  $n = 110$ ), patients with advanced adenoma (AA,  $n = 59$ ), patients of tumor-node-metastasis (TNM) stages I and II (T 1 2,  $n = 24$ ), and patients of TNM stages III and IV (T 3 4,  $n = 76$ ), respectively.  $P$  value for test for trend denotes significance measured by the Jonckheere–Terpstra nonparametric test for trend. On the basis of lesion location, CRC, and advanced adenoma (AA) patients were categorized into proximal CRC ( $n = 29$ ) and distal CRC ( $n = 75$ ), proximal AA ( $n = 26$ ) and distal AA ( $n = 31$ ). NS denotes no statistical significance. miRNA levels were expressed in number of copies per nanogram of extracted RNA. Open circles represent samples with undetectable miRNA level. The lines denote the medians.

miR-135b and miR-31 were also found to be significantly upregulated in CRC and advanced adenoma as compared with their adjacent normal tissues (Table 3). This is consistent with previous studies showing that miR-135b (24, 25) and miR-31 (14, 25–28) are upregulated miRNAs in CRC. The levels of miR-135b and miR-31 were therefore tested in a large cohort of stool samples. Significantly higher level of miR-135b, but not miR-31, was detected in the stool samples of CRC ( $P < 0.0001$ ) and advanced adenoma ( $P < 0.0001$ ) compared with healthy controls (Fig. 1A). In addition, the miR-135b level correlated positively with stages of lesions, with more advanced lesions having the highest miRNA level. A significant increasing trend across the histologic sequence was observed from small adenoma, advanced adenomas, TNM stage I and II, to TNM stage III and IV ( $P < 0.0001$ ; Fig. 2B). Areas under the ROC curve were 0.79 for the detection CRC and 0.71 for adenoma (Fig. 1B). At a cutoff of 14 copies/ng RNA, miR-135b had a sensitivity of 73% and 78% for the detection of advanced adenomas and CRC, respectively, with a specificity of 68%. We have included patients with IBD as a control group to demonstrate that the markers are specific to CRC and precancerous adenoma but not to common inflammatory intestinal diseases. The significant drop in the stool miR-135b level upon removal of advanced adenoma and CRC, and the relatively low level of miR-135b detected in stool of patients with IBD indicated that the upregulation of miR-135b is a specific biomarker for CRC and its precancerous lesion. In this study, we did not use RNU6B as internal control for stool miRNA detection. The choice and rationale of normalizing stool-based miRNA have been discussed in our previous study (18), which is mainly because RNU6B could only be detected in 83.3% (25/30) of stool samples, whereas the candidate miRNAs could be detected in 100% samples (18). In addition, under equal amount of input RNA and detection threshold, RNU6B was detected with a much lower abundance compared with candidate miRNAs (18). Thus, RNU6B may not be an ideal internal control for stool-based miRNA. In this regard, absolute quantitation with standard curve calibration was adopted for miRNA quantitation in the current study. Other upregulated miRNA candidates, including miR-18a, miR-20a, and miR-221 from, our miRNA array assay were also upregulated in CRC tissues compared with adjacent normal controls. As these miRNA candidates were not able to discriminate patients with adenoma from healthy individuals in stool samples (Supplementary Table S4), and the sensitivities and specificities of these miRNAs were lower compared with miR-135b for the stool detection, we did not include these miRNAs in this study. miR-135b was demonstrated to be the most discriminating marker for the detection of both advanced adenoma and CRC.

miR-135b originates from 1q32.1, a region of frequent copy number gain in CRC tumorigenesis. miR-135b targets the 3'untranslated region of the adenomatous polyposis coli (*APC*) gene, a well-known tumor suppressor, and

suppresses its expression (24). *APC* downregulation activates the Wnt signaling pathway, an important oncogenic pathway in regulating cell proliferation and apoptosis in CRC. Upregulation of miR-135b is consistent with its role in regulating *APC* gene, whose loss-of-function is well established to be an early event of the adenoma-cancer sequence in colorectum.

Results from this study have several clinical implications. First, effective detection of proximal colon neoplasms is an important criterion for an effective screening tool (29). Current screening modalities seem less sensitive for proximal than distal colonic neoplasms, and interval CRC is more likely to be found in the proximal colon (11, 30). In this study, the miR-135b level has comparable efficacy for the detection of both proximal and distal CRC and advanced adenoma. Second, miR-135b represents a stool-based test that is noninvasive, avoids the unpleasant bowel preparation, and allows the possibility of off-site sample collection. Therefore, miR-135b may act as a potential noninvasive diagnostic biomarker for CRC and its precancerous lesion (advanced adenoma), but more studies in different populations to validate miR-135b as a biomarker for CRC screening are required before it can be applied clinically. Ultimately, test adoption will depend on test performance in the screening setting, availability, affordability, and user appeal in a large population-based program. For instance, screening using fecal DNA markers demonstrated a higher sensitivity than the FOBT, although the two tests yielded similar specificities (10). There has been no direct comparison of stool DNA versus RNA tests in the screening setting. The evidence for the effectiveness of gFOBTs in reducing CRC incidence and mortality is strong, and is based on randomized controlled trials in large number of average risk populations. FIT is a more sensitive test for screen-relevant neoplasia than FOBT; FITs processing and interpretation are automated and objective. Compared with gFOBT and FIT, miRNA lacks a large-scale validation. Only if this validation is conducted, the possible advantage of miRNA in detecting more proximal neoplasms than fecal blood tests will be known.

Stool miR-31 levels did not differ significantly among controls, adenoma, and CRC (Fig. 1C and 1D), and it did not correlate with the stool miR-135b level. We excluded the possibility that the aberrant upregulation of miR-31 was restricted to the submucosal regions given that the upregulation was also present in CRC epithelial cells (14). The reason why stool miR-31 is not discriminating remains to be elucidated; however, this demonstrates the practical fact that abnormalities at the tissue level do not necessarily translate into a clinically useful marker in the stool. As a noninvasive marker, empirical testing in large number of stool samples is required to establish its usefulness.

This study has several limitations. Most patients were recruited from only two centers and some patients were symptomatic, results might not be representative of the screening setting, although we did not detect significant difference in miRNA results when patients were divided into

asymptomatic screening or symptomatic surgical cohorts. More late-stage CRC patients were identified in the study ( $n = 24$  for TNM I and II vs.  $n = 76$  for TNM III and IV) and the stool samples from CRC used in the study might be biased to late stage. Prompted by these findings, larger scale validation across multiple centers and different populations will be conducted. In addition, there were more male subjects in the CRC group and they were significantly older than the control group. Nonetheless, when we took into account age and gender in the data analysis for the stool miR-135b level, our results remained valid and were independent of age and gender. We have not assessed the effect of time point of sampling (before vs. after colonoscopy) on miRNA results. In all patients with advanced adenomas, stool was collected one week before colonoscopy, whereas in subjects with CRC, stool samples were collected one week after colonoscopy but before surgery. Finally, there are no head to head comparisons of miRNA and FOBTs.

In conclusion, this study demonstrates that stool-based miR-135b seems to be a potential noninvasive biomarker for the detection of CRC and advanced adenoma. Further validation in multiple cohorts of patients and comparison with the established tests are needed before it can be used as a biomarker in routine clinical practice.

## References

- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;61:69–90.
- Sung JJ, Lau JY, Young GP, Sano Y, Chiu HM, Byeon JS, et al. Asia Pacific consensus recommendations for colorectal cancer screening. *Gut* 2008;57:1166–76.
- Winawer SJ, Zauber AG, O'Brien MJ, Ho MN, Gottlieb L, Sternberg SS, et al. Randomized comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps. The National Polyp Study Workgroup. *N Engl J Med* 1993;328:901–6.
- Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegoijen M, Hankey BF, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med* 2012;366:687–96.
- Morikawa T, Kato J, Yamaji Y, Wada R, Mitsushima T, Shiratori Y, et al. A comparison of the immunochemical fecal occult blood test and total colonoscopy in the asymptomatic population. *Gastroenterology* 2005;129:422–8.
- Pickhardt PJ, Choi JR, Hwang I, Butler JA, Puckett ML, Hildebrandt HA, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med* 2003;349:191–2200.
- Kim DH, Pickhardt PJ, Taylor AJ, Leung WK, Winter TC, Hinshaw JL, et al. CT colonography versus colonoscopy for the detection of advanced neoplasia. *N Engl J Med* 2007;357:1403–12.
- Johnson CD, Chen MH, Toledano AY, Heiken JP, Dachman A, Kuo MD, et al. Accuracy of CT colonography for detection of large adenomas and cancers. *N Engl J Med* 2008;359:1207–17.
- Pickhardt PJ, Hassan C, Halligan S, Marmo R. Colorectal cancer: CT colonography and colonoscopy for detection—systematic review and meta-analysis. *Radiology* 2011;259:393–405.
- Imperiale TF, Ransohoff DF, Itzkowitz SH, Turnbull BA, Ross ME. Colorectal Cancer Study Group. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med* 2004;351:2704–14.
- Ahlquist DA, Sargent DJ, Loprinzi CL, Levin TR, Rex DK, Ahnen DJ, et al. Stool DNA and occult blood testing for screen detection of colorectal neoplasia. *Ann Intern Med* 2008;149:441–50.
- Ahlquist DA, Zou H, Domanico M, Mahoney DW, Yab TC, Taylor WR, et al. Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. *Gastroenterology* 2012;142:248–56.
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116:281–97.
- Bandres E, Cubedo E, Agirre X, Malumbres R, Zárate R, Ramirez N, et al. Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. *Mol Cancer* 2006;5:29.
- Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA Jr, Sjoblom T, et al. The colorectal microRNAome. *Proc Natl Acad Sci U S A* 2006;103:3687–92.
- Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature* 2005;435:834–8.
- Michael MZ, O' Connor SM, van Holst Pellekaan NG, Young GP, James RJ. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol Cancer Res* 2003;1:882–91.
- Wu CW, Ng SS, Dong YJ, Ng SC, Leung WW, Lee CW, et al. Detection of miR-92a and miR-21 in stool samples as potential screening biomarkers for colorectal cancer and polyps. *Gut* 2012;61:739–45.
- Luo X, Burwinkel B, Tao S, Brenner H. MicroRNA signatures: novel biomarker for colorectal cancer? *Cancer Epidemiol Biomarkers Prev* 2011;20:1272–86.
- Ahmed FE, Jeffries CD, Vos PW, Flake G, Nuovo GJ, Sinar DR, et al. Diagnostic microRNA markers for screening sporadic human colon cancer and active ulcerative colitis in stool and tissue. *Cancer Genomics Proteomics* 2009;6:281–95.
- Koga Y, Yasunaga M, Takahashi A, Kuroda J, Moriya Y, Akasu T, et al. MicroRNA expression profiling of exfoliated colonocytes isolated from feces for colorectal cancer screening. *Cancer Prev Res* 2010;3:1435–42.
- Link A, Balaguer F, Shen Y, Nagasaka T, Lozano JJ, Boland CR, et al. Fecal MicroRNAs as novel biomarkers for colon cancer screening. *Cancer Epidemiol Biomarkers Prev* 2010;19:1766–74.
- Schetter AJ, Leung SY, Sohn JJ, Zanetti KA, Bowman ED, Yanaihara N, et al. MicroRNA expression profiles associated with prognosis and

## Disclosure of Potential Conflicts of Interest

F.K.L. Chan reports receiving speakers bureau honoraria from the speakers' bureaus of AstraZeneca, Eisai, and Pfizer. No potential conflicts of interest were disclosed by the other authors.

## Authors' Contributions

**Conception and design:** C.W. Wu, S.C. Ng, J.J.Y. Sung, J. Yu  
**Development of methodology:** C.W. Wu, W.T. Law, J. Yu  
**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** C.W. Wu, S.C. Ng, Y. Dong, S.S.M. Ng, W.-W. Leung, W.T. Law, T.O. Yau, F.K.L. Chan  
**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** C.W. Wu, L. Tian, W.T. Law, J. Yu  
**Writing, review, and/or revision of the manuscript:** C.W. Wu, S.C. Ng, L. Tian, S.S.M. Ng, J.J.Y. Sung, J. Yu  
**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** C.W. Wu, T.O. Yau, J. Yu  
**Study supervision:** C.W. Wu, F.K.L. Chan, J. Yu

## Grant Support

This work was supported by research grants from a Technology and Innovation Project Fund Shenzhen (JSGG20130412171021059), a National High-tech R&D Program of China (863 Program, 2012AA02A506), a National Basic Research Program of China (973 Program, 2013CB531401), and an IIT fund Hong Kong (ITS/276/11).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 26, 2013; revised March 5, 2014; accepted March 12, 2014; published OnlineFirst April 1, 2014.



- therapeutic outcome in colon adenocarcinoma. *JAMA* 2008;299:425–36.
24. Nagel R, le Sage C, Diosdado B, van der Waal M, Oude Vrielink JA, Bolijn A, et al. Regulation of the adenomatous polyposis coli gene by the miR-135 family in colorectal cancer. *Cancer Res* 2008;68:5795–802.
  25. Sarver AL, French AJ, Borralho PM, Thayanyithy V, Oberg AL, Silverstein KA, et al. Human colon cancer profiles show differential microRNA expression depending on mismatch repair status and are characteristic of undifferentiated proliferative states. *BMC Cancer* 2009;9:401.
  26. Wang CJ, Zhou ZG, Wang L, Yang L, Zhou B, Gu J, et al. Clinicopathological significance of microRNA-31, -143 and -145 expression in colorectal cancer. *Dis markers* 2009;26:27–34.
  27. Motoyama K, Inoue H, Takatsuno Y, Tanaka F, Mimori K, Uetake H, et al. Over- and under-expressed microRNAs in human colorectal cancer. *Int J Oncol* 2009;34:1069–75.
  28. Slaby O, Svoboda M, Fabian P, Smerdova T, Knoflickova D, Bednarikova M, et al. Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer. *Oncology* 2007;72:397–402.
  29. Baxter NN, Goldwasser MA, Paszat LF, Saskin R, Urbach DR, Rabeneck L. Association of colonoscopy and death from colorectal cancer. *Ann Intern Med* 2009;150:1–8.
  30. Brenner H, Hoffmeister M, Arndt V, Stegmaier C, Altenhofen L, Haug U. Protection from right- and left-sided colorectal neoplasms after colonoscopy: population-based study. *J Natl Cancer Inst* 2010;102:89–95.

# MicroRNA-221 and MicroRNA-18a Identification in Stool as Potential Biomarkers for the Non-Invasive Diagnosis of Colorectal Carcinoma

**Tung On Yau**, Chung Wah Wu, Yujuan Dong, Ceen-Ming Tang, Simon Siu Man Ng, Francis Ka Leung Chan, Joseph Jao Yiu Sung, Jun Yu

*British Journal of Cancer.* **2014**;111(9):1765–71.

Journal URL: [nature.com/articles/bjc2014484](http://nature.com/articles/bjc2014484)

DOI: [10.1038/bjc.2014.484](https://doi.org/10.1038/bjc.2014.484)

PMID: [25233396](https://pubmed.ncbi.nlm.nih.gov/25233396/)

PMCID: [PMC4453736](https://pubmed.ncbi.nlm.nih.gov/PMC4453736/)

Keywords: microRNA; stool biomarker; colorectal cancer; diagnosis

# microRNA-221 and microRNA-18a identification in stool as potential biomarkers for the non-invasive diagnosis of colorectal carcinoma

T O Yau<sup>1,2</sup>, C W Wu<sup>1,2</sup>, Y Dong<sup>1,3</sup>, C-M Tang<sup>1,4</sup>, S S M Ng<sup>3</sup>, F K L Chan<sup>1</sup>, J J Y Sung<sup>1</sup> and J Yu<sup>\*,1,2</sup>

<sup>1</sup>Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China; <sup>2</sup>Gastrointestinal Cancer Biology & Therapeutics Laboratory, Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China; <sup>3</sup>Department of Surgery, The Chinese University of Hong Kong, Hong Kong, China and <sup>4</sup>Department of Pharmacology, University of Oxford, Oxford OX1 3QT, UK

**Background:** The detection of microRNA (miRNA) dysregulation in stool is a novel approach for the diagnosis of colorectal carcinoma (CRC). The aim of this study is to investigate the use of miR-221 and miR-18a in stool samples as non-invasive biomarkers for CRC diagnosis.

**Methods:** A miRNA expression array containing 667 miRNAs was performed to identify miRNA dysregulation in CRC tissues. We focused on miR-221 and miR-18a, two significantly upregulated miRNAs which were subsequently verified in 40 pairs of CRC tissues and 595 stool samples (198 CRCs, 199 polyps and 198 normal controls).

**Results:** miR-221 and miR-18a were upregulated in the miRNA expression array. miR-221 and miR-18a levels were also significantly higher in 40 CRC tumours compared with their respective adjacent normal tissues. In stool samples, miR-221 and miR-18a showed a significant increasing trend from normal controls to late stages of CRC ( $P < 0.0001$ ). The levels of stool miR-221 and miR-18a were both significantly higher in subjects with stages I + II (miR-221:  $P < 0.0001$ , miR-18a:  $P < 0.0001$ ) and stages III + IV of CRC (miR-221:  $P = 0.0004$ , miR-18a:  $P < 0.0001$ ) compared with normal controls. The AUC of stool miR-221 and miR-18a were 0.73 and 0.67 for CRC patients as compared with normal controls, respectively. No significant differences in stool miR-221 and miR-18a levels were found between patients with proximal and distal CRCs. The use of antibiotics did not influence stool miRNA-221 and miRNA-18a levels.

**Conclusions:** Stool-based miR-221 can be used as a non-invasive biomarker for the detection of CRC.

Colorectal carcinoma (CRC) is the third most common cancer, and the fourth highest cause of cancer-related mortality worldwide, with an estimated incidence of one million cases annually (Jemal *et al*, 2011). More significantly, the latest report from the Hong Kong Cancer Registry in 2011 identified CRC as the leading cause of cancer-related morbidity in Hong Kong. The pathogenesis of

CRC follows a stepwise progression from benign adenomas to malignant adenocarcinomas and distant metastasis. Survival is thus inversely related to cancer stages, with up to 90% of deaths preventable if detected early. Colorectal carcinoma, however, is often asymptomatic in its early stages and remains undiagnosed until advanced stages, where prognosis becomes poor. Thus, there

\*Correspondence: Professor J Yu; Email: junyu@cuhk.edu.hk

Received 21 April 2014; revised 6 August 2014; accepted 8 August 2014; published online 18 September 2014

© 2014 Cancer Research UK. All rights reserved 0007–0920/14

Table 1. Clinicopathological characteristics of recruited subjects

Category	Normal	Adenoma	AA	CRC
No. of cases	198	151	48	198
Age at enrolment, years (mean $\pm$ s.d.)	58.65 $\pm$ 6.87	60.39 $\pm$ 5.65	58.73 $\pm$ 6.78	66.53 $\pm$ 11.05
<b>Gender, no. (%)</b>				
Male	84 (42%)	83 (55%)	31 (65%)	116 (59%)
Female	114 (58%)	68 (45%)	17 (35%)	82 (41%)
<b>Location<sup>a</sup>, no. (%)</b>				
Proximal			20 (41.7%)	50 (25.3%)
Distal (excluding rectum)			18 (37.5%)	82 (41.4%)
Rectum			8 (16.6%)	66 (33.3%)
Others <sup>b</sup>			2 (4.2%)	0
<b>CRC staging</b>				
I + II				106 (53.5%)
III + IV				88 (44.5%)
Unknown				4 (2.0%)
<b>Antibiotic intake, no. (%)</b>				
Yes				26 (13%)
No				172 (87%)

Abbreviations: AA = advanced adenoma; CRC = colorectal carcinoma.

<sup>a</sup>Colorectal neoplasms were classified by three locations: proximal colon (caecum, ascending, hepatic flexure, and transverse), distal colon (splenic flexure, descending, sigmoid, and recto-sigmoid junction) and rectum.

<sup>b</sup>The carcinomas or advanced adenomas were found in more than one location.

is a compelling need to identify molecular biomarkers for mass screening and early diagnosis of CRC (Wu *et al*, 2012, 2014).

The advent of microRNA (miRNA) detection techniques has created a new focus in cancer biomarker research. MicroRNAs belong to a class of highly conserved short single-stranded segments (18–25 nucleotides) of non-coding ribonucleic acid, which induce messenger RNA degradation and/or inhibit translation of target genes via binding to the 3'-untranslated regions to regulate gene expression. Due to the unique clinicopathological features of cancer cells, miRNA expression profiles also vary from normal cell types (Ahmed *et al*, 2009; Chang *et al*, 2011). The two miRNA biomarkers (miR-221 and miR-18a) identified in this study have a well-established link with colon tumorigenesis. miR-221 belongs to the miR-221/222 family located on Xp11.3. It is a known suppressor of the p27 protein—a key regulator of the cell cycle. In support of its role in tumorigenesis, downregulation of p27 by miR-221 has been shown to induce cell proliferation and reduce apoptosis in multiple cancers including prostate (Galardi *et al*, 2007), glioblastoma (Gillies and Lorimer, 2007), hepatic (Pineau *et al*, 2010) and gastric carcinoma (Chun-Zhi *et al*, 2010). The oncogenic KRAS also induces increased expression of miR-221 in CRC cell lines (Tsunoda *et al*, 2011). miR-18a belongs to the miR-17-92 cluster located in the 13q31.1 region, an area partly regulated by the oncogenic transcription factor c-Myc (O'Donnell *et al*, 2005). The oncogenic role of the miR-17-92 cluster has also been well-documented, with its overexpression linked to accelerated tumour growth, cell proliferation and progression from benign adenomas to CRC (Hayashita *et al*, 2005; He *et al*, 2005; Diosdado *et al*, 2009).

The aim of this study was to evaluate two upregulated stool miRNAs (miR-221 and miR-18a) as non-invasive CRC diagnostic biomarkers. These two candidate miRNA biomarkers were validated in 40 pairs of primary CRC tumours, and then in a large cohort of 595 stool samples containing 198 CRCs, 199

adenomas and 198 healthy subjects. Through this large case-controlled study, we identified and characterised stool-based miR-221 as a potential biological marker for the diagnosis of CRC.

## MATERIALS AND METHODS

**Patients and sample collection.** Forty pairs of CRC and adjacent normal tissue specimens (at least 4 cm away from the tumour margin) were biopsied during the initial colonoscopy or during the surgical resection. Tissue specimens were snap frozen immediately in a liquid nitrogen filled vacuum flask and stored at  $-80^{\circ}\text{C}$ .

Stool samples were collected from 595 subjects (198 CRCs, 199 adenomas and 198 healthy subjects; Table 1) using a 30 ml disposable stool sample container with screw cap. The containers were manufactured under aseptic conditions to eliminate any biological contamination. Stool samples from CRC patients were collected 7 days after colonoscopy, whereas stool samples from healthy controls and adenoma patients were collected before bowel purgation and colonoscopy. Following collection, all stool samples were immediately stored at  $4^{\circ}\text{C}$ , and transferred to a  $-80^{\circ}\text{C}$  freezer within 24 h.

Advanced adenoma was defined as an adenoma  $\geq 10$  mm in diameter with villous, tubulovillous features or high-grade dysplasia (Winawer *et al*, 1993; Zarchy and Ershoff, 1994). Colorectal neoplasms were classified by three locations as follows: the proximal colon (caecum, ascending, hepatic flexure and transverse), distal colon (splenic flexure, descending, sigmoid and recto-sigmoid junction) and rectum. Exclusion criteria included subjects with a family history of familial adenomatous polyposis or hereditary non-polyposis CRC, previous colonic surgery or adjuvant therapy for CRC before surgery. Patients who were

passing type 7 stool on the Bristol Stool Chart (Lewis and Heaton, 1997) were also excluded. All participants had signed informed consent for obtaining stool and tissue specimens, and were recruited from Alice Ho Miu Ling Nethersole Hospital, Tai Po, Hong Kong and Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong. The study protocol was approved by the institutional review board of the Hong Kong Hospital Authority and The Chinese University of Hong Kong.

**MicroRNA extraction in tissue and stool specimens.** Frozen colon tissue (10–20 µg) was added into 500 µl of Trizol reagent (Invitrogen, Carlsbad, CA, USA) in a 1.5 ml RNase-free Microfuge tube. The tissue was homogenised by RNase-free pestles and vortexed for 30 s to allow for complete homogenisation. Chloroform (100 µl) was subsequently added to the 1.5 ml tube. Fresh human stool samples (10–20 g) were collected using 30 ml specimen containers and stored at –80 °C. The stool sample (200–300 mg (wet weight)) was scooped from the container and added into 1 ml of Trizol LS reagent (Invitrogen) in a 2 ml RNase-free microcentrifuge tube (Invitrogen). The stool sample was subsequently deformed by a RNase-free pestle (USA Scientific, Woodland, CA, USA) and homogenised by a vortex mixer in the Trizol LS reagent. After completing the homogenisation, 200 µl of chloroform was added into the 2 ml tube. Total stool RNA, including miRNA, was extracted from the Trizol LS–chloroform mixture using the miRNeasy Mini Kit (Qiagen, Valencia, CA, USA) as per the protocol provided. Total RNA was eluted in 50 µl of nuclease free water. RNA concentration was measured by Nanodrop 2000 (Thermo Fisher Scientific, Wilmington, DE, USA). Each total RNA sample was normalised to 2 ng µl<sup>-1</sup> based on the Nanodrop 2000 reading.

**Reverse transcription (RT) for microRNA microarray.** Reverse transcription for the miRNA microarray was carried out using Megaplex Primer pools, Human Pools A and B v2.1 (Applied Biosystems, Foster City, CA, USA). The cDNA product was subsequently used to perform the miRNA array. In the array, the miRNA profile of 667 human miRNAs in the tumour tissues and their adjacent normal tissues of five patients was obtained using TaqMan Human MicroRNA Array Set v2.0 (Applied Biosystems). Briefly, 6 µl of Megaplex RT product was added to 444 µl nuclease-free water and 450 µl TaqMan Universal PCR Master Mix with no AmpErase UNG (Applied Biosystems). Reaction mix (100 µl) was loaded into each of the eight fill reservoirs of the array. The array was then centrifuged to distribute the cDNA samples to the reaction wells. Real-time quantitative PCR was performed using Applied Biosystems 7900HT Real-Time PCR System. PCR profile was as follows: 95 °C for 10 min, and then 50 cycles of 95 °C for 15 s and 60 °C for 1 min. Data collection was carried out at the 60 °C step. Results were analysed by the SDS RQ Manager 1.2 software (Applied Biosystems).

**MicroRNA quantitation by quantitative real-time PCR.** Quantitative real-time PCR of each miRNA was performed using the TaqMan miRNA reverse transcription Kit (Applied Biosystems) and TaqMan Human miRNA Assay (miR-18a: 002422 and miR-221: 002279). In brief, 0.3 µl TaqMan miRNA reverse transcription primer, 3 nM dNTP (with dTTP), 10 units reverse transcriptase, 0.6 units RNase inhibitor, 0.3 µl 10 × RT buffer, and 2 ng total RNA were used in one RT reaction with a total volume of 3 µl. The thermal cycling conditions were as follows: 16 °C for 30 min, 42 °C for 30 min, 85 °C for 5 min, and hold at 4 °C. The reverse transcription product was subsequently diluted four-fold by adding 9 µl of nuclease free water.

The PCR reaction mix contains 10 µl TaqMan Universal PCR Master Mix with no AmpErase UNG, 0.5 µl miRNA TaqMan primers, 4 µl diluted RT product and 5.5 µl nuclease free water. Real-time PCR was carried out using the 7500 real-time PCR

system (Applied Biosystems). The PCR profile was as follows: 95 °C for 10 min, followed by 50 cycles of 95 °C for 15 s and 60 °C for 1 min. Data collection was carried out at each 60 °C step. The quantitation of the two target miRNAs (miR-221 and miR-18a) were based on standard curves plotted by known input among all of the miRNAs, and normalised to per nanogram of the total input RNA. Based on standard curves plotted from known amounts of synthetic miR-221 and miR-18a individually, a technical detection limit of two copies for miR-221 would give a Ct value of 42, and a technical detection limit of five copies for miR-18a would give a Ct value of 47. Consequently, we assigned '0' to all Ct values > 42 for miR-221, and 47 for miR-18a. Samples with no amplification of miR-221 or miR-18a were included and also assigned a value of '0' in the analysis, provided the sample could be amplified by another miRNA such as miR-135b or miR-20a (Wu, *et al*, 2014). All assays were performed in a blinded manner.

**Statistics.** The difference between miRNA expression in paired CRC and adjacent normal tissue specimens was evaluated by the Wilcoxon matched-pairs test. Receiver operating characteristic (ROC) curves were generated based on the stool miRNA levels of CRC and control groups. Significance in trend was tested by the Jonckheere trend test. Differences in stool miRNA levels between groups were analysed by the unpaired Student's *t*-test. Two cutoff values were selected using ROC curves for reference, based on either a high sensitivity or a high specificity. Combination analysis was calculated by binary logistic regression. *P* < 0.05 was taken as statistically significant. The Jonckheere trend test and ROC analysis were done by SPSS 16.0 (SPSS Inc., Chicago, IL, USA). All other statistical tests were done by Graphpad Prism 5.01 (Graphpad Software Inc., San Diego, CA, USA).

## RESULTS

**miR-221 and miR-18a are significantly upregulated in primary CRC compared with their adjacent normal tissues.** In addition to miR-135b and miR-31 (Wu *et al*, 2014), miR-221 and miR-18a were found to be two of the most upregulated miRNAs in the miRNA expression array. Thus miR-221 and miR-18a were selected for validation in 40 paired tumour and adjacent normal tissues from CRC patients. We found that miR-221 (1.96-fold, *P* < 0.0001) and miR-18a (2.65-fold, *P* = 0.0003) expression were significantly upregulated in tumours compared with their respective adjacent normal tissues (Table 2).

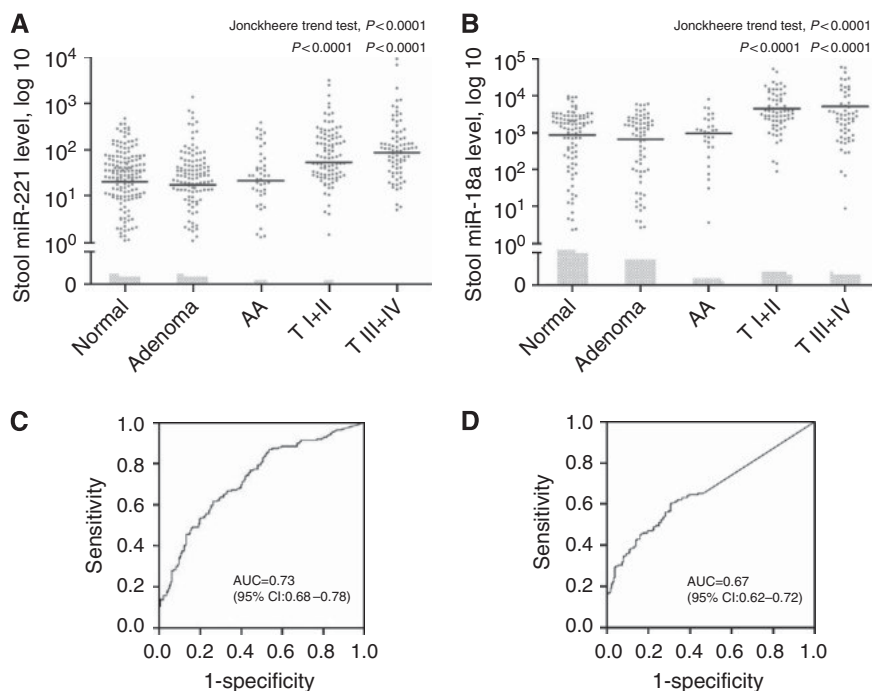
**Stool miR-221 and miR-18a are potential non-invasive biomarkers for CRC.** miR-221 and miR-18a were examined in stool samples from CRC and adenoma patients. Patients were stratified based on their histological stages, with individuals assigned to the groups normal colonoscopy (*n* = 198), adenoma (*n* = 151), advanced adenoma (*n* = 48), CRC stages I + II (*n* = 106) and CRC stages III + IV (*n* = 88) (Table 1). As shown in Figure 1, levels of stool miR-221 (Figure 1A) as well as miR-18a (Figure 1B)

Table 2. miRNAs are differentially expressed in colorectal carcinoma tissues compared with adjacent normal tissues

miRNA	Percentage of samples with elevated expression	Fold change (interquartile range)	P-value <sup>a</sup>
miR-221	77.5% (31/40)	1.96 (1.025–3.319)	<0.0001
miR-18a	77.5% (31/40)	2.65 (1.078–6.828)	0.0003

Abbreviation: miRNA, microRNA.

<sup>a</sup>*P*-values were analysed by Wilcoxon matched-pairs test.



**Figure 1.** Levels of candidate markers in stool (A) miR-221 and (B) miR-18a, and their respective receiver operating characteristic (ROC) curves (C) miR-221 and (D) miR-18a. Patients were categorised into four groups: individuals with a normal colonoscopy (normal) ( $n = 124$ ), adenoma ( $n = 72$ ), advanced adenoma (AN) ( $n = 48$ ), CRC stages I + II (T1, 2) ( $n = 24$ ) and CRC stages III + IV (T3, 4) ( $n = 76$ ). The miRNA levels were expressed as the number of copies per nanogram of extracted RNA. Open circles represent samples with an undetectable miRNA level. The lines denote the medians.  $P < 0.05$  denotes statistical significance. ROC curves were plotted to discriminate all CRC patients from individuals with normal colonoscopy findings.

displayed a significant trend of increase with lesion stages ( $P < 0.0001$ ). The two miRNAs were able to discriminate patients with CRC from healthy individuals. Stool miR-221 was significantly higher in CRC stages I + II (mean: 187.8, 95% CI: 105.6–270.0,  $P < 0.0001$ ) and in CRC stages III + IV (mean: 360.6, 95% CI: 100.6–620.6,  $P = 0.0004$ ) compared with healthy controls (mean: 47.0, 95% CI: 36.3–57.8) (Figure 1A). Stool miR-18a was also significantly increased in CRC stages I + II (mean: 4504, 95% CI: 2919–6088,  $P < 0.0001$ ) and CRC stages III + IV (mean: 5131, 95% CI: 2749–7513,  $P < 0.0001$ ) as compared with controls (mean: 872, 95% CI: 633–1111) (Figure 1B).

The area under the curve (AUC) values were 0.73 (95% CI: 0.68–0.78) and 0.67 (95% CI: 0.62–0.72) for miR-221 (Figure 1C) and miR-18a (Figure 1D) in the detection of CRC, respectively. The cutoff values for stool miR-221 and miR-18a were selected to maximise the sum of the sensitivity and specificity, and were 48 copies per ng and 446 copies per ng of extracted stool RNA, respectively. miR-221 has a sensitivity of 62% and specificity of 74%, whereas miR-18a has a sensitivity of 61% and specificity of 69%. The levels of miR-221 and miR-18a were correlated. Sixty-eight per cent of samples with upregulated miR-221 (cutoff: 48) also had upregulated miR-18a (cutoff: 446). Combining miR-221 and miR-18a produced an AUC of 0.75, with a sensitivity of 66% and specificity of 75% for CRC. Two relatively high specificity levels (80% and 90%) were chosen to reflect its performance for reference, with the cutoff values of 68 and 128 copies per ng for miR-221, demonstrating a sensitivity of 52% and 35%, respectively. The cutoff values of 1845 and 2886 copies per ng for miR-18a had a sensitivity of 46% and 35%, respectively (Table 3).

**Stool miR-221 and miR-18a are not associated with the location of CRC.** We evaluated the sensitivities of stool miR-221 and miR-18a based on the location of CRC. No significant differences were observed in sensitivities for these two miRNAs in detecting CRCs from the proximal colon, distal colon and the rectum. miR-221

showed a higher sensitivity in detecting rectal lesions than carcinomas situated in the proximal or distal colon, but not at a statistically significant level ( $P = 0.058$ ) (Figure 2).

**Stool miR-221 and miR-18a expressions are not associated with antibiotic intake.** We investigated the effects of antibiotic intake on stool miR-221 and miR-18a. Twenty-six CRC patients had taken antibiotics within 1 month of stool collection, whereas the remaining 162 CRC patients had not. There were no significant differences in stool miR-221 and miR-18a expression between the groups with or without antibiotic intake (Figure 3).

## DISCUSSION

Detecting aberrantly expressed miRNA in stool has emerged as a promising non-invasive approach to CRC screening (Ahmed *et al*, 2009; Koga *et al*, 2010; Link *et al*, 2010; Wu *et al*, 2012). Stool miRNAs demonstrate high stability and can be detected with high reproducibility by real-time PCR (Wu *et al*, 2012). Since CRC exhibits recognisable early stages and has high homogeneity in miRNA alterations (Luo *et al*, 2011), stool miRNA may be a useful non-invasive tool for CRC screening. We have previously investigated the expression profile of 667 miRNAs in a microarray, and reported miR-135b and miR-31 as potential biomarkers (Wu *et al*, 2014). In this study, miR-221 and miR-18a were verified in 40 paired tumour and adjacent normal tissues from CRC patients. Both miRNAs were confirmed to be more highly expressed in tumours than in the adjacent normal tissue (Table 2). We then investigated the expression of miR-221 and miR-18a in stool samples from 595 subjects, including 198 patients with CRC, 199 patients with polyps and 198 individuals with normal colonoscopy (Table 1). The two biomarkers were significantly upregulated in CRC stages I + II (miR-221:  $P < 0.0001$  and miR-18a:  $P < 0.0001$ ) and CRC stages III + IV (miR-221:  $P = 0.0004$  and miR-18a:

Table 3. The sensitivity and specificity of stool-based miR-18a and miR-221

Stool-based microRNA	microRNA-221			microRNA-18a		
Cutoff value <sup>a</sup>	48 (best)	68 (80% specificity)	128 (90% specificity)	446 (best)	1845 (80% specificity)	2886 (90% specificity)
<b>Sensitivities, % (95% CI)</b>						
CRC	62 (55–68)	52 (45–59)	35 (28–42)	61 (53–68)	46 (39–54)	35 (29–42)
<b>Stage</b>						
I + II	44 (33–53)	29 (20–40)	15 (8–26)	61 (54–71)	49 (39–59)	37 (28–27)
III + IV	63 (52–74)	54 (42–65)	28 (19–39)	58 (47–68)	40 (30–51)	31 (21–41)
<b>Location</b>						
Proximal	49 (35–63)	41 (28–58)	29 (18–44)	62 (47–75)	46 (32–61)	36 (23–51)
Distal	63 (52–74)	57 (46–69)	35 (25–47)	59 (47–69)	46 (35–58)	37 (26–48)
Rectum	70 (57–84)	55 (42–67)	38 (26–51)	62 (49–74)	48 (36–61)	33 (22–46)
<b>Antibiotic intake</b>						
Yes	65 (44–83)	58 (37–77)	33 (26–41)	62 (55–70)	48 (40–55)	35 (17–56)
No	61 (53–68)	51 (43–59)	46 (27–67)	50 (30–70)	38 (20–59)	35 (28–43)
Specificity, % (95% CI)	74 (67–80)	80 (74–86)	90 (85–94)	69 (62–75)	80 (74–86)	90 (85–94)

Abbreviations: CI = confidence interval; CRC = colorectal carcinoma; ROC = receiver operating characteristic.  
<sup>a</sup>Cutoff values were expressed as copies of two microRNAs per nanogram of extracted stool total RNA. The cutoff values were selected using ROC curves for reference.

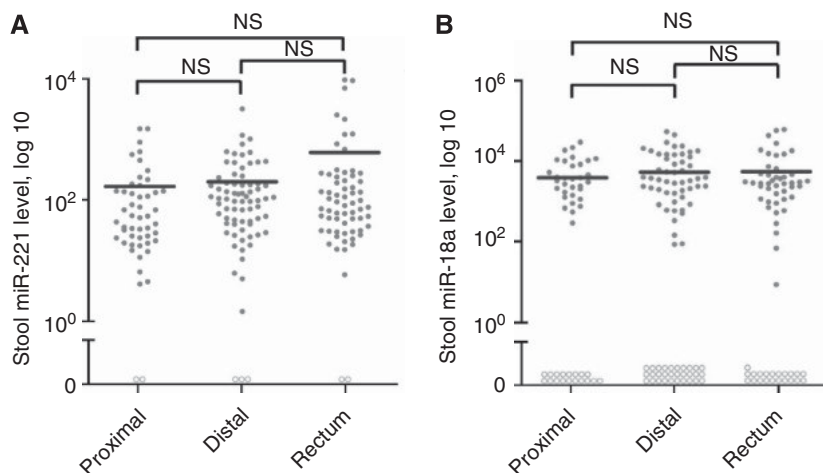


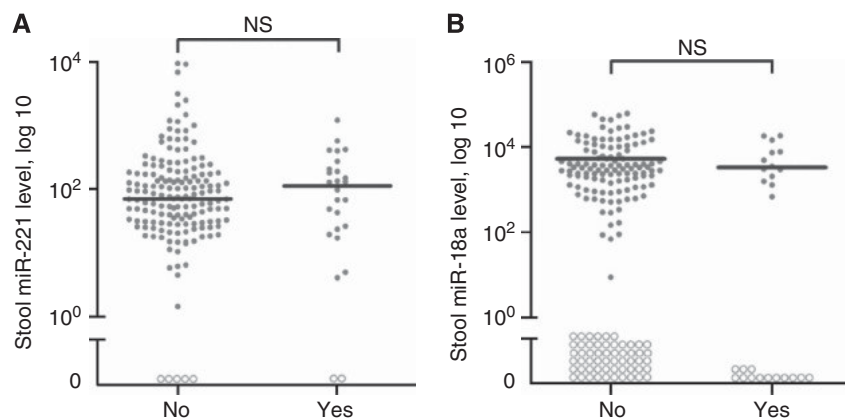
Figure 2. Based on lesion location, colorectal carcinoma (CRC) patients (A) miR-221 and (B) miR-18a were categorised into proximal (n = 50), distal (n = 82) and rectum (n = 66). Colorectal neoplasms were classified by three locations as follows: the proximal colon (caecum, ascending, hepatic flexure and transverse), distal colon (splenic flexure, descending and sigmoid and recto-sigmoid junction) and rectum. NS denotes no statistical significance. miRNA levels were expressed in number of copies per nanogram of extracted RNA. Open circles represent samples with an undetectable miRNA level. The lines denote the medians.

$P < 0.0001$ , Figure 1) compared with controls. There was no significant upregulation in adenoma or advanced adenoma for both miR-221 and miR-18a. This may be because miR-221 and miR-18a are not upregulated until more advanced stages of CRC. Collectively, these results suggest that the candidate miRNAs regulate key signalling pathways in colorectal tumorigenesis. Thus there is a strong rationale for using them as CRC biomarkers.

Our results revealed that stool miR-221 is a potential diagnostic biomarker. Plasma miR-221 was also previously investigated as a biomarker in CRC diagnosis. While it had a high sensitivity of 86%, it had a poor specificity of 41% and an AUC of only 0.61 (95% CI: 0.49–0.72) (Pu *et al*, 2010). By contrast, stool miR-221 had an AUC of 0.73, a sensitivity of 62% and a specificity of 74%. Hence the detection of miR-221 in stool is more specific to CRC than in plasma. This is because plasma miR-221 levels are

influenced by other factors such as digestive disease malignancies (Komatsu *et al*, 2011; Song *et al*, 2012; Cai *et al*, 2013; Kawaguchi *et al*, 2013), reproductive malignancies (Yaman Agaoglu *et al*, 2011; Hong *et al*, 2013) and haematological malignancies (Huang *et al*, 2012; Gimenes-Teixeira *et al*, 2013), as well as systemic diseases such as atherosclerosis, stroke (Tsai *et al*, 2013) and obesity (Ortega *et al*, 2013). In this study, levels of miR-221 have comparable efficacy for the detection of both proximal and distal CRC. Moreover, this stool-based detection method avoids the need for bowel preparation and invasive practice. Therefore, miR-221 is a potential non-invasive diagnostic biomarker for CRC.

Levels of stool miR-18a were also increased in CRC. However, the relatively low AUC value of 0.67 in detection of CRC suggested that miR-18a alone may not be a good biomarker for CRC. Instead, miR-18a may be used in combination with miR-221 to produce an



**Figure 3.** Evaluation of the effects of antibiotics on stool biomarkers **(A)** miR-221 and **(B)** miR-18a. Patients who took antibiotics within 30 days of specimen collection ( $n=26$ ) were compared with patients without any antibiotic intake ( $n=162$ ). miRNA levels were expressed in number of copies per nanogram of extracted RNA. Open circles represent samples with an undetectable miRNA level. The lines denote the medians.

AUC of 0.75 (95% CI: 0.70–0.80) (Supplementary Figure 1), with a sensitivity and specificity of 66% and 75% for CRC, respectively (Supplementary Table 1). This is better than either marker when considered alone. Therefore, combining miR-18a and/or miR-221 with other stool miRNA biomarkers, such as miR-135b (Wu *et al*, 2014), may also increase the sensitivity and specificity of CRC detection in screening programs (Supplementary Figure 1) (Supplementary Table 1).

Recent studies have also revealed that the gut microbiome influences miRNA levels (Yang *et al*, 2013) within host cells. Therefore, we investigated the effects of antibiotics on stool miRNA levels by comparing patients who took antibiotics within 30 days of the stool collection and those who had not. There were no significant differences in stool miR-221 and miR-18a expressions between the groups. To our knowledge, we are the first to report that antibiotic intake by patients does not affect or inhibit microRNA detection in stool (Figure 3). However, further studies are needed to determine the effects of antibiotics on the stool-based miRNAs reported by other groups. Nevertheless, this evidence is significant since antibiotic use is common. Thus restrictions to antibiotic use prior to testing are not required to optimise test performance.

Previously, stool miR-144\* was reported as a potential biomarker for CRC detection, with an AUC of 0.83, sensitivity of 74% and specificity of 85% (Kalimutho *et al*, 2011). The small sample size (35 CRCs and 40 healthy controls) on stool-based miR-144\* detection, however, undermines its use as a biomarker in CRC. More recently, stool miR-106a was identified as a molecular marker for identifying CRC patients from those with negative immunochemical faecal occult blood test results (Koga *et al*, 2013). However, their study was compromised by the use of miR-24 as an internal control. miR-24 belongs to the miR-17-92 cluster, which is known to be upregulated in CRC tissues. It is unclear why such abnormality at tissue level did not translate into a measurable difference in stool in their study, and this requires further investigation (Koga *et al*, 2013). From a more practical perspective, miR-221 also had greater amplification efficiency than miR-92a (Wu *et al*, 2012) and miR-135b (Wu *et al*, 2014), making it highly applicable to CRC screening. To our knowledge, this study tested the largest number of samples to date. Our findings add to an increasingly vast amount of knowledge about the use of stool miRNAs as tools in the non-invasive diagnosis of CRC.

Nevertheless, there are some limitations to this study. All patients were recruited from only two centres and some patients were symptomatic. While we did not detect any significant difference in stool miRNA expression when patients were divided into asymptomatic or symptomatic surgical cohorts, results may not accurately reflect the screening setting in the community.

The small sample sizes of advanced adenoma ( $n=48$ ) and antibiotic intake groups ( $n=26$ ) may have also caused bias in our study. Prompted by these findings, larger scale validation across multiple centres and different populations will be conducted.

In conclusion, our study demonstrated that stool-based miR-221 can be utilised as a potential biological marker. Its use in combination with other reported miRNA biomarkers can be an effective way of increasing the diagnostic accuracy of CRC screening.

#### ACKNOWLEDGEMENTS

We thank Mr Thomas Lam, Ms Joyce Choi and Ms Whitney Tang for their assistance in collecting clinical samples, and Dr JY Shen for his help in statistics. This project was supported by the Shenzhen Technology and Innovation Project Fund, Shenzhen (JSGG20130412171021059), the China 863 programme (2012AA02A506), the China 973 Programme (2013CB531401), the ITF fund Hong Kong (Project code: ITS/276/11), the Shenzhen Municipal Science and Technology R&D fund (JCYJ JCYJ20120619152326450) and the Shenzhen Virtual University Park Support Scheme to CUHK Shenzhen Research Institute.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### REFERENCES

- Ahmed FE, Jeffries CD, Vos PW, Flake G, Nuovo GJ, Sinar DR, Naziri W, Marquard SP (2009) Diagnostic microRNA markers for screening sporadic human colon cancer and active ulcerative colitis in stool and tissue. *Cancer Genomics Proteomics* 6: 281–295.
- Cai H, Yuan Y, Hao Y-F, Guo T-K, Wei X, Zhang Y-M (2013) Plasma microRNAs serve as novel potential biomarkers for early detection of gastric cancer. *Med Oncol* 30: 452.
- Chang KH, Miller N, Kheirleiseid EAH, Lemetre C, Ball GR, Smith MJ, Regan M, McAnena OJ, Kerin MJ (2011) MicroRNA signature analysis in colorectal cancer: identification of expression profiles in stage II tumors associated with aggressive disease. *Int J Colorectal Dis* 26: 1415–1422.
- Chun-Zhi Z, Lei H, An-Ling Z, Yan-Chao F, Xiao Y, Guang-Xiu W, Zhi-Fan J, Pei-Yu P, Qing-Yu Z, Chun-Sheng K (2010) MicroRNA-221 and microRNA-222 regulate gastric carcinoma cell proliferation and radioresistance by targeting PTEN. *BMC Cancer* 10: 367.



- Diosdado B, van de Wiel MA, Terhaar Sive Droste JS, Mongera S, Postma C, Meijerink WJH, Carvalho B, Meijer GA (2009) MiR-17-92 cluster is associated with 13q gain and c-myc expression during colorectal adenoma to adenocarcinoma progression. *Br J Cancer* **101**: 707–714.
- Galardi S, Mercatelli N, Giorda E, Massalini S, Frajese GV, Ciafrè SA, Farace MG (2007) miR-221 and miR-222 expression affects the proliferation potential of human prostate carcinoma cell lines by targeting p27Kip1. *J Biol Chem* **282**: 23716–23724.
- Gillies JK, Lorimer IAJ (2007) Regulation of p27 Kip1 by miRNA 221/222 in Glioblastoma. *Cell Cycle* **6**: 2005–2009.
- Gimenes-Teixeira HL, Lucena-Araujo AR, Dos Santos GA, Zanette DL, Scheucher PS, Oliveira LC, Dalmazzo LF, Silva-Júnior WA, Falcão RP, Rego EM (2013) Increased expression of miR-221 is associated with shorter overall survival in T-cell acute lymphoid leukemia. *Exp Hematol Oncol* **2**: 10.
- Hayashita Y, Osada H, Tatematsu Y, Yamada H, Yanagisawa K, Tomida S, Yatabe Y, Kawahara K, Sekido Y, Takahashi T (2005) A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res* **65**: 9628–9632.
- He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM (2005) A microRNA polycistron as a potential human oncogene. *Nature* **435**: 828–833.
- Hong F, Li Y, Xu Y, Zhu L (2013) Prognostic significance of serum microRNA-221 expression in human epithelial ovarian cancer. *J Int Med Res* **41**: 64–71.
- Huang J, Yu J, Li J, Liu Y, Zhong R (2012) Circulating microRNA expression is associated with genetic subtype and survival of multiple myeloma. *Med Oncol* **29**: 2402–2428.
- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. *CA Cancer J Clin* **61**: 69–90.
- Kalimutho M, Del Vecchio Blanco G, Di Cecilia S, Sileri P, Cretella M, Pallone F, Federici G, Bernardini S (2011) Differential expression of miR-144\* as a novel fecal-based diagnostic marker for colorectal cancer. *J Gastroenterol* **46**: 1391–1402.
- Kawaguchi T, Komatsu S, Ichikawa D, Morimura R, Tsujiura M, Konishi H, Takeshita H, Nagata H, Arita T, Hirajima S, Shiozaki A, Ikoma H, Okamoto K, Ochiai T, Taniguchi H, Otsuji E (2013) Clinical impact of circulating miR-221 in plasma of patients with pancreatic cancer. *Br J Cancer* **108**: 361–369.
- Koga Y, Yasunaga M, Takahashi A, Kuroda J, Moriya Y, Akasu T, Fujita S, Yamamoto S, Baba H, Matsumura Y (2010) MicroRNA expression profiling of exfoliated colonocytes isolated from feces for colorectal cancer screening. *Cancer Prev Res* **3**: 1435–1442.
- Koga Y, Yamazaki N, Yamamoto Y, Yamamoto S, Saito N, Kakugawa Y, Otake Y, Matsumoto M, Matsumura Y (2013) Fecal miR-106a is a useful marker for colorectal cancer patients with false-negative results in immunochemical fecal occult blood test. *Cancer Epidemiol Biomarkers Prev* **22**: 1844–1852.
- Komatsu S, Ichikawa D, Takeshita H, Tsujiura M, Morimura R, Nagata H, Kosuga T, Iitaka D, Konishi H, Shiozaki A, Fujiwara H, Okamoto K, Otsuji E (2011) Circulating microRNAs in plasma of patients with oesophageal squamous cell carcinoma. *Br J Cancer* **105**: 104–111.
- Lewis SJ, Heaton KW (1997) Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol* **32**: 920–924.
- Link A, Balaguer F, Shen Y, Nagasaka T, Lozano JJ, Boland CR, Goel A (2010) Fecal MicroRNAs as novel biomarkers for colon cancer screening. *Cancer Epidemiol Biomarkers Prev* **19**: 1766–1774.
- Luo X, Burwinkel B, Tao S, Brenner H (2011) MicroRNA signatures: novel biomarker for colorectal cancer? *Cancer Epidemiol Biomarkers Prev* **20**: 1272–1286.
- O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435**: 839–843.
- Ortega FJ, Mercader JM, Catalán V, Moreno-Navarrete JM, Pueyo N, Sabater M, Gómez-Ambrosi J, Anglada R, Fernández-Formoso JA, Ricart W, Frühbeck G, Fernández-Real JM (2013) Targeting the circulating microRNA signature of obesity. *Clin Chem* **59**: 781–792.
- Pineau P, Volinia S, McJunkin K, Marchio A, Battiston C, Terris B, Mazzaferro V, Lowe SW, Croce CM, Dejean A (2010) miR-221 overexpression contributes to liver tumorigenesis. *Proc Natl Acad Sci USA* **107**: 264–269.
- Pu X, Huang G, Guo H, Guo C, Li H, Ye S, Ling S, Jiang L, Tian Y, Lin T (2010) Circulating miR-221 directly amplified from plasma is a potential diagnostic and prognostic marker of colorectal cancer and is correlated with p53 expression. *J Gastroenterol Hepatol* **25**: 1674–1680.
- Song M, Pan K, Su H, Zhang L, Ma J, Li J, Yuasa Y, Kang D, Kim YS, You W (2012) Identification of serum microRNAs as novel non-invasive biomarkers for early detection of gastric cancer. *PLoS One* **7**: e33608.
- Tsai P-C, Liao Y-C, Wang Y-S, Lin H-F, Lin R-T, Juo S-HH (2013) Serum microRNA-21 and microRNA-221 as potential biomarkers for cerebrovascular disease. *J Vasc Res* **50**: 346–354.
- Tsunoda T, Takashima Y, Yoshida Y, Doi K, Tanaka Y, Fujimoto T, Machida T, Ota T, Koyanagi M, Kuroki M, Sasazuki T, Shirasawa S (2011) Oncogenic KRAS regulates miR-200c and miR-221/222 in a 3D-specific manner in colorectal cancer cells. *Anticancer Res* **31**: 2453–2459.
- Winawer SJ, Zauber AG, O'Brien MJ, Ho MN, Gottlieb L, Sternberg SS, Wayne JD, Bond J, Schapiro M, Stewart ET (1993) Randomized comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps. The National Polyp Study Workgroup. *N Engl J Med* **328**: 901–906.
- Wu CW, Ng SC, Dong Y, Tian L, Ng SS, Leung WW, Law WT, Yau TO, Chan FK, Sung JJ, Yu J (2014) Identification of microRNA-135b in stool as a potential non-invasive biomarker for colorectal cancer and adenoma. *Clin Cancer Res* **20**: 2994–3002.
- Wu CW, Ng SSM, Dong YJ, Ng SC, Leung WW, Lee CW, Wong YN, Chan FKL, Yu J, Sung JY (2012) Detection of miR-92a and miR-21 in stool samples as potential screening biomarkers for colorectal cancer and polyps. *Gut* **61**: 739–745.
- Yaman Agaoglu F, Kovancilar M, Dizdar Y, Darendeliler E, Holdenrieder S, Dalay N, Gezer U (2011) Investigation of miR-21, miR-141, and miR-221 in blood circulation of patients with prostate cancer. *Tumour Biol* **32**: 583–588.
- Yang T, Owen JL, Lightfoot YL, Kladd MP, Mohamadzadeh M (2013) Microbiota impact on the epigenetic regulation of colorectal cancer. *Trends Mol Med* **19**: 714–725.
- Zarchy TM, Ershoff D (1994) Do characteristics of adenomas on flexible sigmoidoscopy predict advanced lesions on baseline colonoscopy? *Gastroenterology* **106**: 1501–1504.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)

# MicroRNA-20a in Human Faeces as a Non-Invasive Biomarker for Colorectal Cancer

**Tung On Yau**, Chung Wah Wu, Ceen-Ming Tang, Yingxuan Chen, Jingyuan Fang, Yujuan Dong, Qiaoyi Liang, Simon Siu Man Ng, Francis Ka Leung Chan, Joseph Jao Yiu Sung, Jun Yu

*Oncotarget*. **2016**;7(2):1559–68.

Journal URL: [oncotarget.com/article/6403/](http://oncotarget.com/article/6403/)

DOI: [10.18632/oncotarget.6403](https://doi.org/10.18632/oncotarget.6403)

PMID: [26621842](https://pubmed.ncbi.nlm.nih.gov/26621842/)

PMCID: [PMC4811480](https://pubmed.ncbi.nlm.nih.gov/PMC4811480/)

## microRNA-20a in human faeces as a non-invasive biomarker for colorectal cancer

Tung On Yau<sup>1</sup>, Chung Wah Wu<sup>1</sup>, Ceen-Ming Tang<sup>1,2</sup>, Yingxuan Chen<sup>3</sup>, Jingyuan Fang<sup>3</sup>, Yujuan Dong<sup>1,4</sup>, Qiaoyi Liang<sup>1</sup>, Simon Siu Man Ng<sup>4</sup>, Francis Ka Leung Chan<sup>1</sup>, Joseph Jao Yiu Sung<sup>1</sup>, Jun Yu<sup>1</sup>

<sup>1</sup>Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, CUHK Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong

<sup>2</sup>Oxford University Clinical Academic Graduate School, John Radcliffe Hospital, Oxford, UK

<sup>3</sup>Renji Hospital, Shanghai Jiaotong University, Shanghai, China

<sup>4</sup>Department of Surgery, The Chinese University of Hong Kong, Hong Kong

**Correspondence to:** Jun Yu, e-mail: junyu@cuhk.edu.hk

**Keywords:** *microRNA, non-invasive, stool biomarker, colorectal cancer, diagnosis*

**Received:** July 25, 2015

**Accepted:** November 15, 2015

**Published:** November 26, 2015

### ABSTRACT

**Objective:** Detection of microRNA (miRNA) aberrations in human faeces is a new approach for colorectal cancer (CRC) screening. The aim of this study was to characterise miR-20a in faeces as a non-invasive biomarker for diagnosis of CRC.

**Design:** miR-20a was selected from an expression microarray containing 667 miRNAs. Further verification of miR-20a was performed in 40 pairs of primary CRC tissues, as well as 595 faecal samples (198 CRCs, 199 adenomas, and 198 healthy controls) using TaqMan probe based quantitative Real-Time PCR (qRT-PCR).

**Results:** miR-20a expression was significantly higher in the 40 CRC tumours compared to their respective adjacent normal tissues ( $P = 0.0065$ ). Levels of miR-20a were also significantly higher in faecal samples from CRC patients ( $P < 0.0001$ ). The area under receiver operating characteristic (AUROC) curve for miR-20a was 0.73, with a sensitivity of 55% and specificity of 82% for CRC patients compared with controls. No significant difference in the level of miR-20a was found between patients with proximal, distal, and rectal cancer. The use of antibiotics did not influence faecal miR-20a levels.

**Conclusions:** Faecal-based miR-20a can be utilised as a potential non-invasive biomarker for CRC screening.

### INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer worldwide, with incidence rates increasing by 6% over the past decade [1]. CRC typically develops from benign adenomas to malignant adenocarcinomas through a long and protracted stepwise process. Patient survival is inversely related to the cancer stage at diagnosis, with up to 90% of deaths preventable if diagnosed early [2]. However, colorectal cancer is frequently asymptomatic in its early stages. Hence, the development of non-invasive biomarkers for screening the populations at risk is urgently needed [3].

miRNAs belong to a class of highly conserved short single-stranded non-coding RNAs, which regulates

messenger RNA (mRNA) degradation, and inhibits translation of target genes via binding to the 3'-untranslated regions (3'UTR). Since miRNA expression profiles between normal and tumour cells, as well as between different subtypes of cancers vary due to their unique clinical histopathologic features, miRNAs are ideal cancer biomarkers [4]. miR-20a belongs to the miR-17/92 cluster located in the 13q31.1 region, and is up-regulated in numerous cancers, including anaplastic thyroid [5], ovarian [6], and prostate cancer [7, 8]. Notably, this area is partly regulated by the oncogenic transcription factor Myc [9] and TGF- $\beta$  [10]. Over-expression of the miR-17/92 cluster is thus associated with accelerated cell proliferation [11], tumour development [12], and transformation from benign adenomas to CRC [13].

Data from our miRNA microarray, which was previously reported [14], demonstrated that miR-20a was the one of most up-regulated miRNA in tumours compared to adjacent normal tissues. Thus, the purpose of this study was to evaluate the expression of miR-20a in faeces as a non-invasive CRC diagnostic biomarker. We began by using 40 paired clinical CRC tissues to validate miR-20a expression. Next, miR-20a expression was validated in faecal samples from a large cohort of 595 patients, including 198 with CRC, 199 with adenomas, and 198 healthy controls. Through this large case-controlled study, we identified and characterised faecal-based miR-20a as a potential non-invasive biomarker for CRC diagnosis.

## RESULTS

### miR-20a is significantly up-regulated in primary CRC compared to their adjacent normal tissues

Amongst the 667 miRNAs we screened using a microarray reported previously [14], miR-20a was the most up-regulated miRNA in tumour specimens compared to its adjacent normal. Thus, miR-20a was selected for further validation in 40 paired tumour and corresponding adjacent normal tissues from CRC patients. We found that miR-20a expression was significantly up-regulated (fold change: 2.063 (0.910–5.418),  $P = 0.0065$ ) in tumours compared to adjacent normal tissues (Table 1).

### Faecal-based miR-20a is a potential non-invasive marker for colorectal cancer

miR-20a was evaluated in three groups of participants, that is groups with normal colonoscopy ( $n = 198$ ), adenoma ( $n = 199$ ), and CRC ( $n = 198$ ) (Table 2). As shown in Figure 1A, miR-20a was able to discriminate between patients with CRC and healthy individuals. Statistically, faecal-based miR-20a levels were significantly higher in CRC (mean: 100,827 copies/ng, 95% confidence interval (CI): 114,870–86,783 copies/ng; median: 30,005 copies/ng;  $P < 0.0001$ ), but also significantly lower in adenoma (mean: 13,199 copies/ng, 95% CI: 15,033–11,365 copies/ng; median: 7,088 copies/ng;  $P = 0.0201$ ) compared to controls (mean: 18,051 copies/ng, 95% CI: 20,566–15,537 copies/ng; median: 10,776 copies/ng) (Figure 1A).

The AUROC values of faecal-based miR-20a were 0.73 (95% CI: 0.68–0.78) in CRC, and 0.41 (95% CI: 0.35–0.47) in adenoma (Figure 1B). The cut-off value of 27,493 copies/ng of extracted total faecal RNA for miR-20a was selected to maximise the sum of the sensitivity and specificity for CRC diagnosis (Table 3). miR-20a had a sensitivity of 55% and specificity of 82% for CRC detection. The second cut-off value of 43,312 copies/ng for miR-20a (Table 3) was chosen for its high specificity of 90% enabling assessment of its performance for reference.

Faecal-based miR-20a in combination with our previously reported faecal miRNA biomarkers miR-92a [15] or miR-135b [14] did not show a big improvement in sensitivity. When miR-20a is combined with miR-92a, the AUROC is 0.77 (95% CI: 0.72–0.82), with a sensitivity and specificity of 57% and 84% for CRC, respectively. If combined with miR-135b, it generates an AUROC of 0.79 (95% CI: 0.74–0.83), with a sensitivity and specificity of 79% and 65% for CRC, respectively (Supplementary Figure S1).

### Faecal-based miR-20a is not associated with the location of CRC

We evaluated the expression levels of faecal-based miR-20a in the context of tumour location in CRC patients. No significant differences were observed with regards to sensitivity for the detection of CRCs from the proximal colon, distal colon, and rectum (Figure 2).

### Faecal-based miR-20a expression is not associated with antibiotic intake

We investigated the effects of antibiotic intake on faecal miR-20a. Twenty-six CRC patients had taken antibiotics within one month of faecal collection, whereas the remaining 162 CRC patients had not. There were no significant differences in faecal-based miR-20a expression between the groups with or without antibiotic intake (Figure 3).

## DISCUSSION

CRC is associated with a highly recognisable, and homogenous pattern of miRNA alterations in human faeces [16]. miRNA in faeces is also stable in room temperature and in a 4°C refrigerator for up to 72 hours, with the results from faecal samples being highly repeatable [15, 17, 18]. Unlike the faecal occult blood test (FOBT), which is currently used for CRC screening, faecal-based miRNA tests do not require troublesome drug and dietary restrictions. Therefore, the uptake of faecal-based miRNA tests may be higher than that of the FOBT, which currently stands at 35% [19]. As a result, quantitation of miRNA biomarkers in human faeces by qRT-PCR is a promising non-invasive approach for screening CRC patients [14, 15, 20, 21]. We have previously investigated the expression profile of 667 mRNAs in a microarray, and reported miR-20a as a potential biomarker [14, 21].

Its potential as a biomarker is supported by various functional studies implicating miR-20a in tumourigenesis. miR-20a has been found to induce epithelial-mesenchymal transition (EMT) – a key step in cell migration and tumour metastasis-via down-regulation of E-cadherin, and up-regulation of matrix metalloproteinases [22, 23]. miR-20a has also been shown to diminish cellular response to the

**Table 1: miR-20a expression is elevated in colorectal carcinoma tissues compared with adjacent normal tissues**

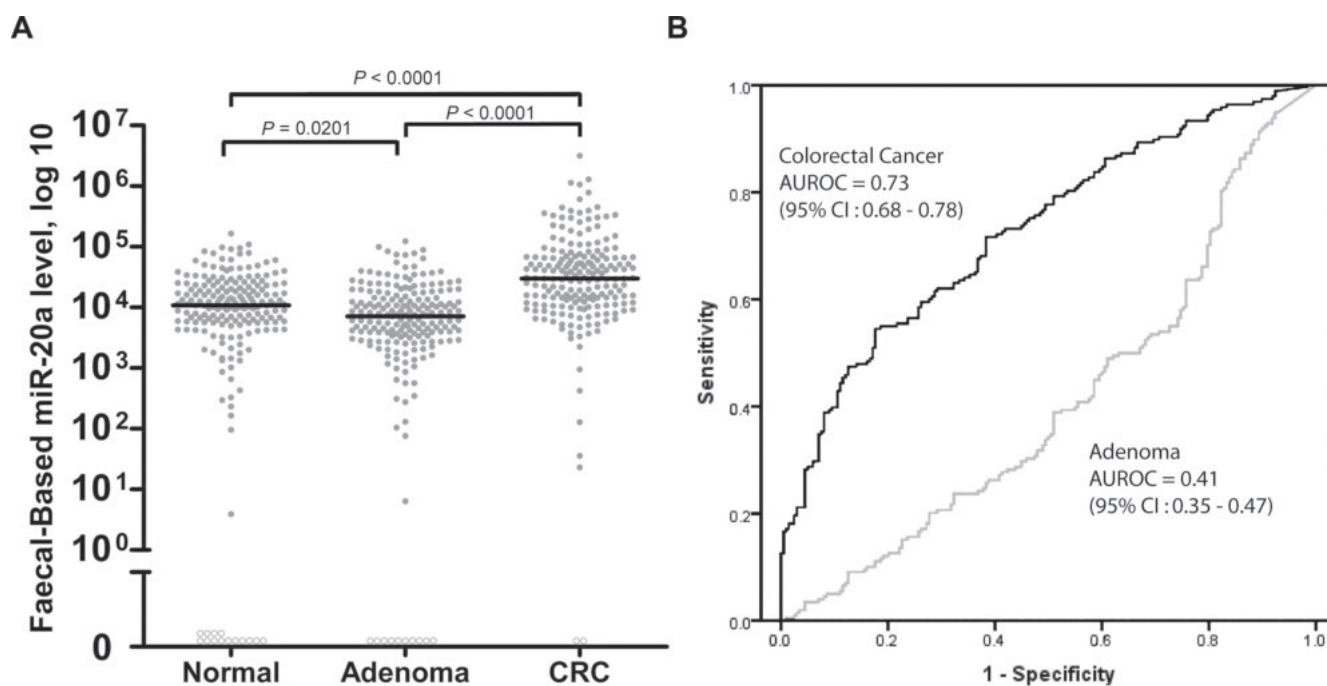
microRNA	Percentage of samples with elevated expression in tumours	Fold change (Interquartile range)	P value
miR-20a	70.0% (28/40)	2.063 (0.910–5.418)	0.0065

TGF- $\beta$  signalling pathway by preventing its delay of G1/S transition and promoting progression into the cell cycle [10, 22]. Mutational inactivation of the TGF- $\beta$  signalling pathway is critical in CRC progression, with restoration of the TGF- $\beta$  pathway in human CRC cells abrogating proliferation and tumorigenicity [24]. Collectively, these functional studies suggested a role for miR-20a in the pathogenesis of CRC, and supported the use of miR-20a as a non-invasive biomarker.

In this study, we began by verifying miR-20a expression levels in 40 paired tissues from CRC patients. miR-20a was confirmed to be more highly expressed in tumours than in their adjacent normal tissues (Table 2). Next, we quantitated miR-20a in human faecal samples from 595 subjects, including 198 patients with CRC, 199 patients with adenoma, and 198 individuals with a normal colonoscopy (Table 1). miR-20a was significantly increased in CRC patients ( $p < 0.0001$ , AUROC = 0.73) compared with the control group (Figure

1). No difference was found between different genders (Supplementary Figure S2A), and early stage (stages I + II) versus late stage (stages III + IV) CRC patients (Supplementary Figure S2B). Studies by other groups have also demonstrated that faecal miR-20a expression was significantly lower after curative CRC surgery, highlighting a potential role for miR-20a in surveillance of CRC recurrence [25]. Collectively, this data demonstrates the ability of miR-20a to differentiate patients with CRC from those without, supporting its use in CRC diagnostics.

Rather unexpectedly, miR-20a expression levels were lower in adenoma than in healthy controls ( $p = 0.0201$ , AUROC = 0.41) (Figure 1). A review of the literature revealed no published studies on faecal miR-20a expression in patients with colorectal adenomas. One study reported tissue miR-20a expression in colorectal adenomas, and found that expression was higher in paraffin-embedded colorectal adenoma samples ( $n = 7$ ) than healthy controls ( $n = 9$ ). The difference,

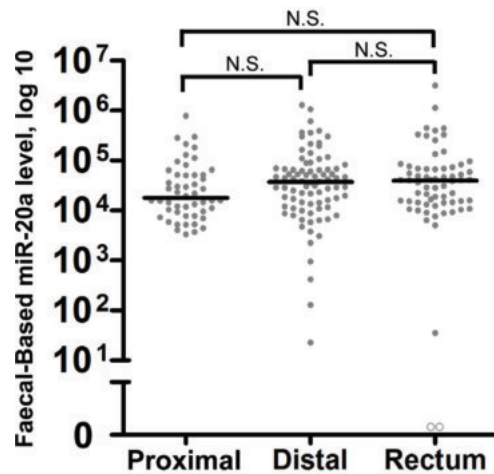


**Figure 1: Levels of (A) faecal-based miR-20a, and (B) the respective area under receiver operating characteristic (AUROC) curves for CRC and adenoma.** Patients were categorised into three subgroups: individuals with a normal colonoscopy (normal) ( $n = 198$ ), adenoma ( $n = 199$ ), and CRC ( $n = 198$ ). The miR-20a level was expressed as the number of copies per nanogram of extracted total RNA. Each open circle represents a sample with an undetectable miRNA level. The lines denote the median.  $P < 0.05$  denotes statistical significance. AUROC curves were plotted to discriminate all CRC and adenoma patients from individuals with normal colonoscopy findings.

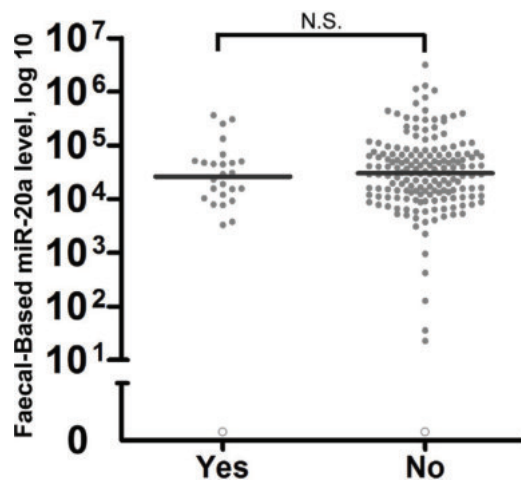
however, was not significant, and the small sample size made the findings unreliable [26]. We hypothesise that the lower expression levels are instead due to the influence of the gut microbiome on miRNA within host cells [27, 28]. This hypothesis is supported by recent studies which revealed the different, and unique microbiota profiles of healthy patients, patients with colorectal adenomas, and patients with CRC [29]. The dominant strains of bacteria in colorectal adenomas may degrade miR-20a in the bowel lumen, thus reducing miRNA expression in faecal samples. It is also known that over time, gut flora may alter gene expression in colonocytes [30, 31]. This may also result in lower expression levels of faecal miR-20a in colorectal adenoma patients. Further

research is needed to evaluate the relationship between the gut flora and expression of miR-20a in patients with colorectal adenomas.

We also investigated external factors which may affect the use of miR-20a as a faecal-based biomarker. We found that miR-20a levels have comparable efficacy for the detection of proximal colon, distal colon and rectal CRC. Whilst levels of faecal-based miR-20a were slightly lower in proximal CRC than distal and rectal CRC, this result was not statistically significant (Figure 2). Other research groups have demonstrated that antibiotics change the composition of intestinal microbiota, which may in turn alter miRNA expression in faeces [32]. Therefore, we also looked into the



**Figure 2: Tumour location does not significantly alter faecal miR-20a levels.** Colorectal neoplasms were classified by three locations as follows: the proximal colon (caecum, ascending, hepatic flexure and transverse) ( $n = 29$ ), distal colon (splenic flexure, descending and sigmoid and recto-sigmoid junction) ( $n = 75$ ), and rectum ( $n = 66$ ). The lines denote the median. N.S. denotes no statistical significance. miR-20a levels were expressed in number of copies per nanogram of extracted total RNA. Each open circle represents a sample with an undetectable miR-20a level.



**Figure 3: Evaluation of the effects of antibiotics on faecal-based biomarker miR-20a.** Patients who took antibiotics within 30 days of specimen collection ( $n = 26$ ) were compared with patients without any antibiotic intake ( $n = 162$ ). The lines denote the median. N.S. denotes no statistical significance. miR-20a levels were expressed in number of copies per nanogram of extracted total RNA. Each open circle represents a sample with an undetectable miR-20a level.

**Table 2: Pathological characteristics of recruited subjects**

Category	Healthy Controls	Adenoma	Colorectal Cancer
No. of Cases	198	199	198
Age at enrolment, Years (Mean ± SD)	58.65 ± 6.87	59.99 ± 5.97	66.53 ± 11.05
<b>Gender, Number (%)</b>			
Male	84 (42%)	114 (57%)	116 (59%)
Female	114 (58%)	85 (43%)	82 (41%)
<b>Location*, Number (%)</b>			
Proximal			50 (25.3%)
Distal			82 (41.4%)
Rectum			66 (33.3%)
<b>Cancer stage, Number (%)</b>			
I + II			106 (53.5%)
III + IV			88 (44.5%)
Unknown			4 (2.0%)
<b>Tumour histology, Number (%)</b>			
Adenocarcinoma			185 (93.4%)
Mucinous adenocarcinoma			11(5.6%)
Unknown			2 (1.0%)
<b>Differentiation, Number (%)</b>			
Poor			1 (0.5%)
Poor to Moderate			2 (1.0%)
Moderate			167 (84.3%)
Well to Moderate			3 (1.5%)
Well			3 (1.5%)
Unknown/No data			22 (11.2%)
<b>Antibiotic intake**, Number (%)</b>			
Yes			26 (13%)
No			172 (87%)

\*Colorectal neoplasms were classified by location into three groups: proximal colon (caecum, ascending, hepatic flexure and transverse), distal colon (splenic flexure, descending and sigmoid and recto-sigmoid junction) and rectum.

\*\*Antibiotic intake is defined as any antibiotic intake in the 30 days preceding faecal sample collection.

effects of antibiotics on faecal-based miR-20a levels by comparing patients who took antibiotics within 30 days of the faecal sample collection and those who had not. There were no significant differences in faecal-based miR-20a expressions between the groups (Figure 3). However, further studies are needed to determine the effects of antibiotics on the faecal-based miRNAs reported by other groups. Nevertheless, this evidence is significant since antibiotic use is common. Thus restrictions to antibiotic use prior to testing for miR-20a are not required to optimise test performance.

Thus far, our results suggest that faecal-based miR-20a is a potential non-invasive biomarker for CRC detection. Opponents, however, argue that faecal-based miRNA tests face similar challenges to the faecal occult blood test (FOBT) in terms of low patient acceptability. Other groups have thus investigated the use of circulating miR-20a in CRC diagnosis. The majority of studies found that circulating miR-20a was unable to differentiate CRC patients from healthy controls in a statistically significant manner [33, 34]. Only one study, in a cohort of 100 CRC and 79 cancer-free controls, reported a

**Table 3: The sensitivity and specificity of faecal-based miR-20a for colorectal cancer detection**

Category	Best	Reference
Specificity, % (95% CI)	82 (76–87)	90 (85–94)
Sensitivity, % (95% CI)	55 (47–62)	40 (33–47)
Cut-off value, copies/nanogram	27,493	43,312
Location*, Sensitivity % (95% CI)		
Proximal	42 (28–57)	30 (20–45)
Distal	60 (48–70)	45 (34–57)
Rectum	58 (45–70)	45 (33–58)
Antibiotic Intake**, Sensitivity % (95% CI)		
No	27 (10–40)	8 (3–14)
Yes	50 (30–70)	42 (23–63)

\*Colorectal neoplasms were classified by location into three groups: proximal colon (caecum, ascending, hepatic flexure and transverse), distal colon (splenic flexure, descending and sigmoid and recto-sigmoid junction) and rectum.

\*\*Antibiotic intake is defined as any antibiotic intake in the 30 days preceding faecal sample collection.

statistically significant difference ( $P = 0.038$ ). However, it had a low AUROC of 0.59, with a sensitivity of 46%, and specificity of 73% [35], making it an ineffective diagnostic tool. Moreover, the levels of circulating miR-20a may be influenced by other factors, including chronic diseases such as HCV-mediated liver disease [36], systemic lupus erythematosus [37] and chronic obstructive pulmonary disease (COPD) [38], as well as other malignancies [39–46]. We believe that the higher specificity of faecal miR-20a makes it a better choice for CRC diagnosis than circulating miRNAs.

Nevertheless, there were several shortcomings with our study. Several internal control genes such as 18S rRNA [47], endogenous control small RNAs (i.e. RNU19 [18] and U6 snRNA [48]), miR-16 [25], and miR-24 [49], were used in other faecal-based miRNA studies to determine the relative miRNA levels according to the  $2(-\Delta\Delta Ct)$  method. However, recent research has suggested that the use of internal controls for faecal-based miRNA detection may not be an ideal approach [15, 18]. This is firstly because 18S rRNA, RNU19, and U6 snRNA have longer sequences and degrade rapidly in faeces, thus potentially confounding results [15]. As the function of miR-16 and miR-24 itself are unknown [49], there may also be unintended repercussions to using it as an internal control. In our experiment, miRNA was quantified with a standard curve plotted by known amounts of synthetic miRNA and normalised to per nanogram of input RNA. Whilst this overcomes the faults of using internal control genes, this approach may also be problematic because the standard curve is only as good as the quantification method and does not eliminate the possibility of DNA contamination. Our laboratory is currently working on solutions to this

problem using multiplex PCR, as well as digital droplet PCR (ddPCR) to optimise performance, and to increase the sensitivity and specificity. Using multiplex PCR techniques, which facilitates detection of multiple targets in a single PCR reaction, our previously reported faecal miRNA biomarkers [14, 15, 21] can be combined with miR-20a in a panel to increase its overall sensitivity and specificity. Likewise, published studies suggest the use of ddPCR, which enables absolute quantification, would increase test performance by reducing the coefficient of variation by up to 86% compared to qRT-PCR [50]. Collectively, detection cost, time, and consumables would be minimised, whilst maximising test performance.

In summary, our study demonstrated that faecal-based miR-20a can be utilised as a potential non-invasive biological marker. Its use in combination with previously reported miRNA biomarkers can be an effective way of screening the population for CRC in a non-invasive manner.

## PATIENTS AND METHODS

### Tissue and faecal sample collection

Forty pairs of primary CRC and its adjacent normal tissues (at least 40 mm away from the tumour margin) were biopsied during the initial colonoscopy or the surgical resection. The specimens were snap frozen immediately in a liquid nitrogen filled vacuum flask, and transferred to a  $-80^{\circ}\text{C}$  freezer for storage.

Faecal samples were collected using a 30 mL disposable container with a screw cap from 595 subjects (198 CRCs, 199 adenomas, and 198 neoplasm-free controls) (Table 1). The containers were manufactured



under aseptic conditions to minimise the possibility of contamination. Faecal samples from CRC patients were collected 7 days after colonoscopy, whereas samples from adenoma and control groups were collected before bowel purgation and colonoscopy. All faecal samples were stored immediately at 4°C following collection, and transferred to a -80°C freezer for storage within 24 hours.

Colorectal neoplasms were categorised by three locations as follows: the proximal colon (caecum, ascending, hepatic flexure, and transverse), distal colon (splenic flexure, descending, sigmoid, and recto-sigmoid junction), and rectum. Exclusion criteria included: (i) patients who were passing type 7 stool on the Bristol stool chart [51], (ii) previous adjuvant therapy and/or colonic surgery for CRC as well as (iii) subjects with a family history of familial hereditary non-polyposis CRC and/or familial adenomatous polyposis. All participants had signed informed consent for obtaining tissue and/or faecal samples, and were recruited from The Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong and The Alice Ho Miu Ling Nethersole Hospital, Tai Po, Hong Kong. The institutional review board of the Hospital Authority of Hong Kong and the Chinese University of Hong Kong approved of the study protocol.

### MicroRNA extraction in tissue and faecal samples

Frozen colorectal tissue (10–20 µg) from biopsies were added into 500 µL of Trizol reagent (Invitrogen, Carlsbad, CA, USA) in a 1.5 mL RNase free microcentrifuge tube. The tissue was homogenised by RNase-free pestles and vortexed for 30 seconds to allow for complete homogenisation. 100 µL of chloroform was subsequently added to the 1.5 mL tube. Faeces (200–300 mg) were scooped from the container, and added into 1 mL of Trizol LS reagent (Invitrogen, Carlsbad, CA, USA) in a 2 mL RNase-free microcentrifuge tube (Invitrogen, Carlsbad, CA, USA). The faecal sample was subsequently deformed by a RNase-free pestle (USA Scientific, Woodland, CA, USA) and homogenised by a vortex mixer in the Trizol LS reagent. After completing the homogenisation, 200 µL of chloroform was added into the 2 mL microcentrifuge tube.

Total RNA, including miRNA from tissue and faeces, were extracted from the Trizol-chloroform and Trizol LS-chloroform mixture respectively using the miRNeasy Mini Kit (Qiagen, Valencia, CA, USA) as per the protocols provided. Total RNA was eluted in 50 µL of nuclease free water. Total RNA concentration was measured using the Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA). Each total RNA sample was normalised to 2 ng/µL based on the Nanodrop 2000 reading.

### MicroRNA quantitation by quantitative real-time PCR

Reverse transcription was performed using the TaqMan miRNA Reverse Transcription Kit (Thermo Fisher Scientific, Waltham, MA, USA). In brief, 2 ng total RNA, 0.3 µL TaqMan miRNA RT primer, 3 nM dNTP (with dTTP), 10 units reverse transcriptase, 0.6 units RNase inhibitor, and 0.3 µL 10X RT buffer were used in one RT reaction with a total volume of 3 µL. The thermal cycling conditions were as follows: 16°C for 30 minutes, 42°C for 30 minutes, 85°C for 5 minutes, and hold at 4°C. The RT product was subsequently diluted four-fold by adding 9 µL of nuclease free water.

qRT-PCR of miR-20a was carried out using the TaqMan has-miR-20a Assay (Assay ID: 000580; Mature sequence: UAAAGUGCUUAUAGUGCAGGUAG) (Thermo Fisher Scientific, Waltham, MA, USA), and the 7500 real-time PCR system (Thermo Fisher Scientific, Waltham, MA, USA). The PCR reaction mix contained 10 µL 2X TaqMan Universal PCR Master Mix with no AmpErase Uracil N-Glycosylase (UNG), 0.5 µL miRNA TaqMan primers, 4 µL diluted RT product, and 5.5 µL nuclease free water. The PCR profile was as follows: 95°C for 10 minutes, 50 cycles of 95°C for 15 seconds, and 60°C for 1 minute. Data collection was carried out at each 60°C step. The quantitation of miR-20a was based on a standard curve plotted by known input amongst all of the miRNAs, and normalised to per nanogram of the total input RNA. Based on standard curves plotted from known amounts of synthetic miR-20a, a technical detection limit of 6 copies for miR-20a would give an approximate Ct value of 48. Consequently, we assigned “0” to all Ct values larger than 48 for miR-20a. Samples with no amplification of miR-20a were also included and assigned a value of “0” in the analysis, provided the sample could be amplified by another miRNA such as miR-135b [14], miR-221, or miR-18a [21]. All assays were performed in a blinded fashion.

### Statistics

The difference between miRNA expression in paired CRC and adjacent normal tissue specimens was evaluated by the Wilcoxon matched-pairs test. AUROC curves were generated based on faecal miRNA levels in patients with CRC and adenoma compared to the control group. Differences in faecal miRNA levels between groups were analysed by the Mann Whitney *U* test. The best cut-off value, selected to maximise the sum of the sensitivity and specificity, and a cut-off with a high specificity of 90%, were selected using the AUROC curve for CRC. A *p* value < 0.05 was considered statistically significant. The AUROC analysis was done by SPSS 16.0 (SPSS Inc., Chicago, Illinois, USA). All other statistical tests

were performed using GraphPad Prism 5.01 (GraphPad Software Inc., San Diego, CA, USA).

## GRANT SUPPORTS

This project was supported by a National High-tech R&D Program China (863 Program, 2012AA02A506), the National Key Technology R&D Program (No. 2014BAI09B05), a Technology and Innovation Project Fund Shenzhen (JSGG20130412171021059), Shenzhen Municipal Science and Technology R & D fund (JCYJ20140414170821182) and Shenzhen Virtual University Park Support Scheme to CUHK Shenzhen Research Institute.

## CONFLICTS OF INTEREST

The authors declare no potential conflicts of interest.

## REFERENCES

1. Siegel R, Desantis C, Jemal A. Colorectal cancer statistics. *CA Cancer J Clin.* 2014; 64:104–117.
2. Smith RA, von Eschenbach AC, Wender R, Levin B, Byers T, Rothenberger D, Brooks D, Creasman W, Cohen C, Runowicz C, Saslow D, Cokkinides V, Eyre H, et al. American Cancer Society guidelines for the early detection of cancer: update of early detection guidelines for prostate, colorectal, and endometrial cancers. Also: update 2001—testing for early lung cancer detection. *CA Cancer J Clin.* 2001; 51:38–75.
3. Lieberman D. Progress and challenges in colorectal cancer screening and surveillance. *Gastroenterology.* 2010; 138:2115–2126.
4. Chang KH, Miller N, Kheirlehid EAH, Lemetre C, Ball GR, Smith MJ, Regan M, McAnena OJ, Kerin MJ. MicroRNA signature analysis in colorectal cancer: identification of expression profiles in stage II tumors associated with aggressive disease. *Int J Colorectal Dis.* 2011; 26:1415–1422.
5. Xiong Y, Zhang L, Kebebew E. MiR-20a Is Upregulated in Anaplastic Thyroid Cancer and Targets LIMK1. *PLoS One.* 2014; 9:e96103.
6. Fan X, Liu Y, Jiang J, Ma Z, Wu H, Liu T, Liu M, Li X, Tang H. miR-20a promotes proliferation and invasion by targeting APP in human ovarian cancer cells. *Acta Biochim Biophys Sin (Shanghai).* 2010; 42:318–324.
7. Qiang X-F, Zhang Z-W, Liu Q, Sun N, Pan L-L, Shen J, Li T, Yun C, Li H, Shi L-H. miR-20a Promotes Prostate Cancer Invasion and Migration Through Targeting ABL2. *J Cell Biochem.* 2014; 115:1269–1276.
8. Pesta M, Klecka J, Kulda V, Topolcan O, Hora M, Eret V, Ludvikova M, Babjuk M, Novak K, Stolz J, Holubec L. Importance of miR-20a expression in prostate cancer tissue. *Anticancer Res.* 2010; 30:3579–3583.
9. O'Donnell KA, Wentzel EA, Zeller KI, Dang C V, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature.* 2005; 435:839–843.
10. Sokolova V, Fiorino A, Zoni E, Crippa E, Reid JF, Gariboldi M, Pierotti MA. The Effects of miR-20a on p21: Two Mechanisms Blocking Growth Arrest in TGF- $\beta$ -Responsive Colon Carcinoma. *J Cell Physiol.* 2015; 230:3105–3114.
11. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM. A microRNA polycistron as a potential human oncogene. *Nature.* 2005; 435:828–833.
12. Hayashita Y, Osada H, Tatematsu Y, Yamada H, Yanagisawa K, Tomida S, Yatabe Y, Kawahara K, Sekido Y, Takahashi T. A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res.* 2005; 65:9628–9632.
13. Diosdado B, van de Wiel MA, Terhaar Sive Droste JS, Mongera S, Postma C, Meijerink WJHJ, Carvalho B, Meijer GA. MiR-17-92 cluster is associated with 13q gain and c-myc expression during colorectal adenoma to adenocarcinoma progression. *Br J Cancer.* 2009; 101:707–114.
14. Wu CW, Ng SC, Dong Y, Tian L, Ng SSM, Leung WW, Law WT, Yau TO, Chan FKL, Sung JY, Yu J. Identification of microRNA-135b in Stool as a Potential Noninvasive Biomarker for Colorectal Cancer and Adenoma. *Clin Cancer Res.* 2014; 20:2994–3002.
15. Wu CW, Ng SSM, Dong YJ, Ng SC, Leung WW, Lee CW, Wong YN, Chan FKL, Yu J, Sung JY. Detection of miR-92a and miR-21 in stool samples as potential screening biomarkers for colorectal cancer and polyps. *Gut.* 2012; 61:739–745.
16. Luo X, Burwinkel B, Tao S, Brenner H. MicroRNA signatures: novel biomarker for colorectal cancer? *Cancer Epidemiol Biomarkers Prev.* 2011; 20:1272–1286.
17. Yamazaki N, Koga Y, Yamamoto S, Kakugawa Y, Otake Y, Hayashi R, Saito N, Matsumura Y. Application of the fecal microRNA test to the residuum from the fecal occult blood test. *Jpn J Clin Oncol.* 2013; 43:726–733.
18. Kalimutho M, Vecchio Blanco G, Cecilia S, Sileri P, Cretella M, Pallone F, Federici G, Bernardini S. Differential expression of miR-144\* as a novel fecal-based diagnostic marker for colorectal cancer. *J. Gastroenterol.* 2011; 46: 1391–1402.
19. Tam TKW, Ng KK, Lau CM, Lai TC, Lai WY, Tsang LCY. Faecal occult blood screening: knowledge, attitudes, and practice in four Hong Kong primary care clinics. *Hong Kong Med J.* 2011; 17:35035–7.
20. Yau TO, Wu CW, Dong Y, Tang C-M, Ng SSM, Chan FKL, Sung JY, Yu J. microRNA-221 and microRNA-18a identification in stool as potential biomarkers for the

- non-invasive diagnosis of colorectal carcinoma. *Br J Cancer*. 2014; 111:1765–1771.
21. Ahmed FE, Jeffries CD, Vos PW, Flake G, Nuovo GJ, Sinar DR, Naziri W, Marcuard SP. Diagnostic microRNA markers for screening sporadic human colon cancer and active ulcerative colitis in stool and tissue. *Cancer Genomics Proteomics*. 2009; 6:281–295.
  22. Zhang GJ, Li Y, Zhou H, Xiao HX, Zhou T. miR-20a is an independent prognostic factor in colorectal cancer and is involved in cell metastasis. *Mol Med Rep*. 2014; 10:283–291.
  23. Xu T, Jing C, Shi Y, Miao R, Peng L, Kong S, Ma Y, Li L. microRNA-20a enhances the epithelial-to-mesenchymal transition of colorectal cancer cells by modulating matrix metalloproteinases. *Exp Ther Med*. 2015; 10:683–688.
  24. Wang J, Sun L, Myeroff L, Wang X, Gentry LE, Yang J, Liang J, Zborowska E, Markowitz S, Willson JK. Demonstration that mutation of the type II transforming growth factor beta receptor inactivates its tumor suppressor activity in replication error-positive colon carcinoma cells. *J Biol Chem*. 1995; 270:22044–22049.
  25. Rotelli MT, Di Lena M, Cavallini A, Lippolis C, Bonfrate L, Chetta N, Portincasa P, Altomare DF. Fecal microRNA profile in patients with colorectal carcinoma before and after curative surgery. *Int J Colorectal Dis*. 2015; 30:891–898.
  26. Tsuchida A, Ohno S, Wu W, Borjigin N, Fujita K, Aoki T, Ueda S, Takanashi M, Kuroda M. miR-92 is a key oncogenic component of the miR-17-92 cluster in colon cancer. *Cancer Sci*. 2011; 102:2264–2271.
  27. Archambaud C, Sismeiro O, Toedling J, Soubigou G, Bécavin C, Lechat P, Lebreton A, Ciaudo C, Cossart P. The intestinal microbiota interferes with the microRNA response upon oral *Listeria* infection. *MBio*. 2013; 4:e00707–00713.
  28. Yang T, Owen JL, Lightfoot YL, Kladd MP, Mohamadzadeh M. Microbiota impact on the epigenetic regulation of colorectal cancer. *Trends Mol Med*. 2013; 19:714–725.
  29. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z, Su L, Li X, Li X, et al. Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat Commun*. 2015; 6:6528.
  30. Zhong C-Y, Sun W-W, Ma Y, Zhu H, Yang P, Wei H, Zeng B-H, Zhang Q, Liu Y, Li W-X, Chen Y, Yu L, Song Z-Y. Microbiota prevents cholesterol loss from the body by regulating host gene expression in mice. *Sci Rep*. 2015; 5:10512.
  31. Schaubeck M, Clavel T, Calasan J, Lagkouvardos I, Haange SB, Jehmlich N, Basic M, Dupont A, Hornef M, Bergen MV, Bleich A, Haller D. Dysbiotic gut microbiota causes transmissible Crohn’s disease-like ileitis independent of failure in antimicrobial defence. *Gut*. 2015.
  32. Panda S, El Khader I, Casellas F, López Vivancos J, García Cors M, Santiago A, Cuenca S, Guarner F, Manichanh C. Short-term effect of antibiotics on human gut microbiota. *PLoS One*. 2014; 9:e95476.
  33. Brunet Vega A, Pericay C, Moya I, Ferrer A, Dotor E, Pisa A, Casalots À, Serra-Aracil X, Oliva J-C, Ruiz A, Saigi E. microRNA expression profile in stage III colorectal cancer: circulating miR-18a and miR-29a as promising biomarkers. *Oncol Rep*. 2013; 30:320–326.
  34. Yamada A, Horimatsu T, Okugawa Y, Nishida N, Honjo H, Ida H, Kou T, Kusaka T, Sasaki Y, Makato Y, Higurashi T, Yukawa N, Amanuma Y, et al. Serum miR-21, miR-29a and miR-125b are promising biomarkers for the early detection of colorectal neoplasia. *Clin Cancer Res*. 2015; 21:4234–4242.
  35. Chen W-Y, Zhao X-J, Yu Z-F, Hu F-L, Liu Y-P, Cui B-B, Dong X-S, Zhao Y-S. The potential of plasma miRNAs for diagnosis and risk estimation of colorectal cancer. *Int J Clin Exp Pathol*. 2015; 8:7092–7101.
  36. Shrivastava S, Petrone J, Steele R, Lauer GM, Di Bisceglie AM, Ray RB. Up-regulation of circulating miR-20a is correlated with hepatitis C virus-mediated liver disease progression. *Hepatology*. 2013; 58:863–871.
  37. Carlsen AL, Schetter AJ, Nielsen CT, Lood C, Knudsen S, Voss A, Harris CC, Hellmark T, Segelmark M, Jacobsen S, Bengtsson AA, Heegaard NHH. Circulating microRNA expression profiles associated with systemic lupus erythematosus. *Arthritis Rheum*. 2013; 65: 1324–1334.
  38. Akbas F, Coskunpinar E, Aynaci E, Oltulu YM, Yildiz P. Analysis of serum micro-RNAs as potential biomarker in chronic obstructive pulmonary disease. *Exp Lung Res*. 2012; 38:286–294.
  39. Moussay E, Wang K, Cho J-H, van Moer K, Pierson S, Paggetti J, Nazarov P V, Palissot V, Hood LE, Berchem G, Galas DJ. MicroRNA as biomarkers and regulators in B-cell chronic lymphocytic leukemia. *Proc Natl Acad Sci USA*. 2011; 108:6573–6578.
  40. Huang J, Yu J, Li J, Liu Y, Zhong R. Circulating microRNA expression is associated with genetic subtype and survival of multiple myeloma. *Med Oncol*. 2012; 29:2402–2408.
  41. Achberger S, Aldrich W, Tubbs R, Crabb JW, Singh AD, Triozzi PL. Circulating immune cell and microRNA in patients with uveal melanoma developing metastatic disease. *Mol Immunol*. 2014; 58:182–186.
  42. Wang M, Gu H, Wang S, Qian H, Zhu W, Zhang L, Zhao C, Tao Y, Xu W. Circulating miR-17-5p and miR-20a: molecular markers for gastric cancer. *Mol Med Rep*. 2012; 5:1514–1520.
  43. Zeng X, Xiang J, Wu M, Xiong W, Tang H, Deng M, Li X, Liao Q, Su B, Luo Z, Zhou Y, Zhou M, Zeng Z, et al. Circulating miR-17, miR-20a, miR-29c, and miR-223 combined as non-invasive biomarkers in nasopharyngeal carcinoma. *PLoS One*. 2012; 7:e46367.
  44. Shen J, Hrubby GW, McKiernan JM, Gurvich I, Lipsky MJ, Benson MC, Santella RM. Dysregulation of circulating microRNAs and prediction of aggressive prostate cancer. *Prostate*. 2012; 72:1469–1477.
  45. Lin H-M, Castillo L, Mahon KL, Chiam K, Lee BY, Nguyen Q, Boyer MJ, Stockler MR, Pavlakakis N, Marx G,

- Mallesara G, Gurney H, Clark SJ, et al. Circulating microRNAs are associated with docetaxel chemotherapy outcome in castration-resistant prostate cancer. *Br J Cancer*. 2014; 110:2462–2471.
46. Zhao S, Yao D, Chen J, Ding N. Circulating miRNA-20a and miRNA-203 for screening lymph node metastasis in early stage cervical cancer. *Genet Test Mol Biomarkers*. 2013; 17:631–636.
47. Ahmed FE, Ahmed NC, Vos PW, Bonnerup C, Atkins JN, Casey M, Nuovo GJ, Naziri W, Wiley JE, Mota H, Allison RR. Diagnostic microRNA markers to screen for sporadic human colon cancer in stool: I. Proof of principle. *Cancer Genomics Proteomics*. 2013; 10:93–113.
48. Koga Y, Yasunaga M, Takahashi A, Kuroda J, Moriya Y, Akasu T, Fujita S, Yamamoto S, Baba H, Matsumura Y. MicroRNA expression profiling of exfoliated colonocytes isolated from feces for colorectal cancer screening. *Cancer Prev Res (Phila)*. 2010; 3:1435–1442.
49. Koga Y, Yamazaki N, Yamamoto Y, Yamamoto S, Saito N, Kakugawa Y, Otake Y, Matsumoto M, Matsumura Y. Fecal miR-106a is a useful marker for colorectal cancer patients with false-negative results in immunochemical fecal occult blood test. *Cancer Epidemiol Biomarkers Prev*. 2013; 22:1844–1852.
50. Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, Vessella RL, Tewari M. Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat Methods*. 2013; 10:1003–1005.
51. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol*. 1997; 32:920–924.

# Faecal MicroRNAs as a Non-Invasive Tool in the Diagnosis of Colonic Adenomas and Colorectal Cancer: A Meta-Analysis

**Tung On Yau**, Ceen-Ming Tang, Elinor K. Harriss, Benjamin Dickins, Christos Polytarchou

*Scientific Reports.* **2019**;9(1):9491.

Journal URL: [nature.com/articles/s41598-019-45570-9](https://www.nature.com/articles/s41598-019-45570-9)

DOI: [10.1038/s41598-019-45570-9](https://doi.org/10.1038/s41598-019-45570-9)

PMID: [31263200](https://pubmed.ncbi.nlm.nih.gov/31263200/)

PMCID: [PMC6603164](https://pubmed.ncbi.nlm.nih.gov/pmc/articles/PMC6603164/)

# SCIENTIFIC REPORTS

OPEN

## Faecal microRNAs as a non-invasive tool in the diagnosis of colonic adenomas and colorectal cancer: A meta-analysis

Tung On Yau<sup>1</sup>, Ceen-Ming Tang<sup>2</sup>, Elinor K. Harriss<sup>3</sup>, Benjamin Dickins<sup>1</sup> & Christos Polytarchou<sup>1</sup>

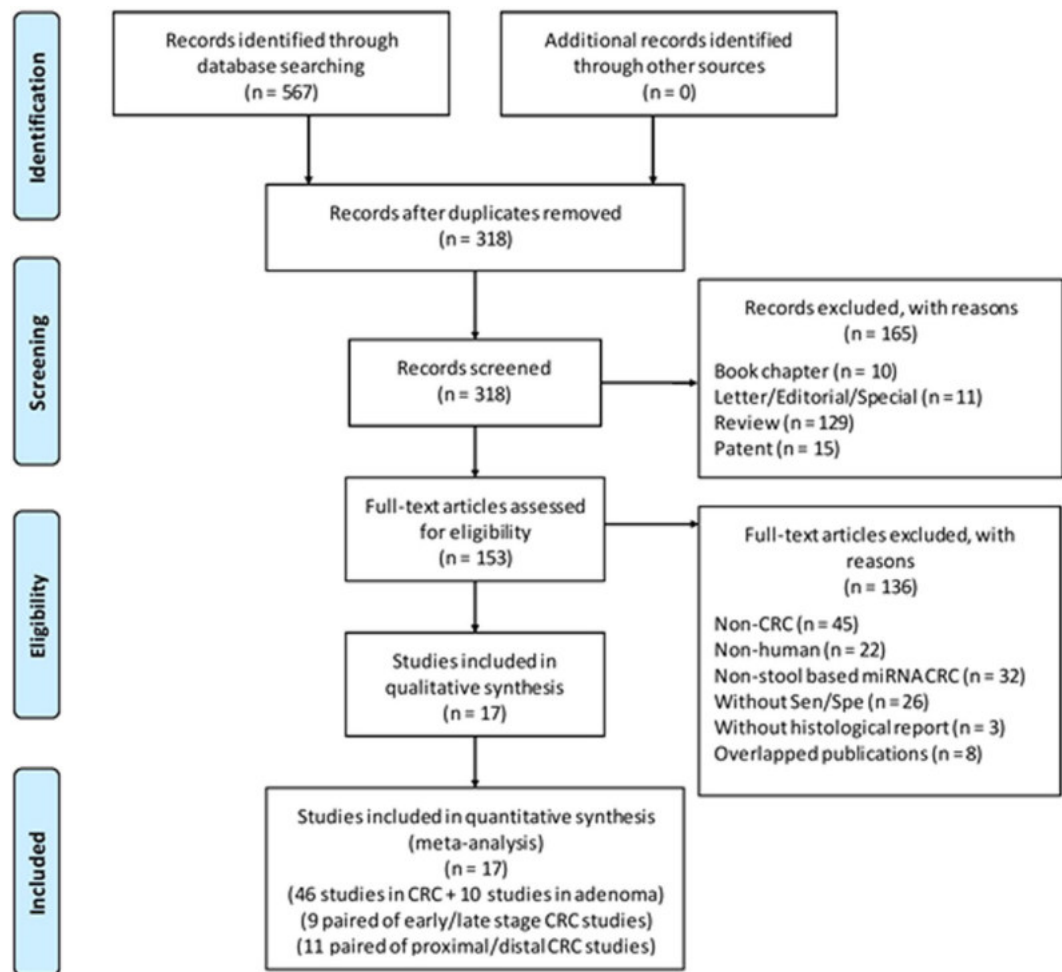
MicroRNAs (miRNAs) are proposed as potential biomarkers for the diagnosis of numerous diseases. Here, we performed a meta-analysis to evaluate the utility of faecal miRNAs as a non-invasive tool in colorectal cancer (CRC) screening. A systematic literature search, according to predetermined criteria, in five databases identified 17 research articles including 6475, 783 and 5569 faecal-based miRNA tests in CRC, adenoma patients and healthy individuals, respectively. Sensitivity, specificity, positive/negative likelihood and diagnostic odds ratios, area under curve (AUC), summary receiver operator characteristic (sROC) curves, association of individual or combinations of miRNAs to cancer stage and location, subgroup, meta-regression and Deeks' funnel plot asymmetry analyses were employed. Pooled miRNAs for CRC had an AUC of 0.811, with a sensitivity of 58.8% (95% confidence interval [CI]: 51.7–65.5%) and specificity of 84.8% (95% CI: 81.1–87.8%), whilst for colonic adenoma, it was 0.747, 57.3% (95% CI: 40.8–72.3%) and 76.1% (95% CI: 66.1–89.4%), respectively. The most reliable individual miRNA was miR-21, with an AUC of 0.843, sensitivity of 59.3% (95% CI: 26.3–85.6%) and specificity of 85.6% (95% CI: 72.2–93.2%). Paired stage analysis showed a better diagnostic accuracy in late stage CRC and sensitivity higher in distal than proximal CRC. In conclusion, faecal miR-21, miR-92a and their combination are promising non-invasive biomarkers for faecal-based CRC screening.

Colorectal cancer (CRC) is the second leading cancer-related cause of death in the United Kingdom (UK) and accounts for over 500,000 deaths annually worldwide<sup>1</sup>. The pathogenesis of CRC follows a protracted stepwise progression from benign colonic adenomas to malignant adenocarcinomas and distant metastasis. Patient survival inversely correlates to cancer stage during diagnosis, with up to 90% of deaths avertable if detected early<sup>2</sup>. However, CRC is often asymptomatic in its early stage and arises sporadically within the population, posing a challenge to the application of effective and timely treatments<sup>3</sup>. The mass screening of asymptomatic individuals for CRC utilising a non-invasive method is thus a high public health priority.

Under the National Health Service (NHS) Bowel Cancer Screening Program in the UK, currently, the faecal immunochemical test (FIT) is offered every two years to all asymptomatic men and women aged 60 to 74<sup>4</sup>. The FIT, which examines faecal samples for hidden blood, is appealing because the costs are low, the test is widely available, and does not pose an immediate risk to the screened population<sup>5</sup>. Although the recent changes from Faecal Occult Blood Test (FOBT) to FIT has improved the screening power by specific targeting to human haemoglobin, the effectiveness of FIT is still restricted by its relatively low sensitivity, with about half of all malignant large bowel tumours and most polyps undetected. This is due to the intermittent nature of bleeding<sup>6</sup> as well as degradation of haemoglobin in faeces<sup>7</sup>. Consequently, one in four CRC cases is only diagnosed at a late stage on emergency admission, resulting in poor prognosis<sup>8</sup>. Therefore, a more sensitive faecal-based non-invasive test is urgently needed.

miRNAs are a class of conserved endogenous, short non-coding RNAs with length of 18–24 nucleotides. miRNAs regulate gene expression through post-transcriptional processing by binding primarily to the 3'-untranslated

<sup>1</sup>Department of Biosciences, John van Geest Cancer Research Centre, School of Science and Technology, Nottingham Trent University, Nottingham, UK. <sup>2</sup>Oxford University Clinical Academic Graduate School, John Radcliffe Hospital, Oxford, UK. <sup>3</sup>Bodleian Health Care Libraries, University of Oxford, Oxford, UK. Correspondence and requests for materials should be addressed to C.P. (email: [christos.polytarchou@ntu.ac.uk](mailto:christos.polytarchou@ntu.ac.uk))



**Figure 1.** Flowchart diagram of study selection based on the inclusion and exclusion criteria.

region (3'UTR) of target mRNAs, resulting in mRNA degradation and/or translational repression<sup>9</sup>. Specific miRNAs (oncomiRs) through targeting tumour-suppressor genes have been found to be upregulated, while others targeting oncogenes are downregulated, in cancer. These alterations, through the regulation of intracellular signalling networks, induce cell proliferation, confer resistance to apoptosis and chemotherapy, and promote metastasis<sup>10</sup>. The expression of several miRNAs differs significantly between normal colonic tissues and CRC, and as colonocytes consistently exfoliate and shed into the lumen of the gastrointestinal (GI) tract, these changes in miRNA levels are represented in faecal specimens<sup>11–27</sup>. More recently, it was demonstrated that miRNAs are highly stable and detectable within samples throughout a 72 hour incubation period due to protection from ribonuclease degradation by exosomes<sup>28,29</sup>. Given that faeces contain genomic DNA and RNA derived from gut microbes, and miRNAs derived from blood cells released by tumours, the detection and utility of miRNAs for diagnostic purposes has been controversial. Therefore, this meta-analysis aims to assess the value of miRNAs as faecal-based biomarkers for CRC and colonic adenoma screening.

## Results

**Characteristics of selected studies.** The initial literature search from five different databases yielded a total of 567 articles, of which 249 were excluded as duplicated records. Next, 165 articles were deemed irrelevant and excluded based on the title and abstract. The full-text of the remaining publications were screened, resulting in the inclusion of 17 publications (16 in English, 1 in Chinese) (Fig. 1). These publications contained 46 studies on CRC and 10 studies on adenomas, corresponding to 6475, 783 and 5569 faecal-based miRNA tests in CRC patients, adenoma and healthy controls, respectively (Tables 1 and 2). The clinical data and collection procedures are summarised in Suppl. Table 1A,B, and methods of miRNA extraction and quantification in Suppl. Table 2A,B.

**Risk of bias.** All included articles were evaluated for the risk of bias using the QUADAS-2 tool (Fig. 2)<sup>30</sup>. The major risk of bias in this study was in the index test, where 10 out of 17 publications had a high risk of bias due to the unclear or lack of statement regarding interpretation of index test results without knowledge of the results of the reference. Additionally, 14 out of 17 studies had an unclear risk of bias in the “Patient Selection” domain. This is due to a lack of detail on whether a consecutive or random sample of patients were enrolled. There was a

Study ID	First author (Year) [Reference no.]	Origin of population	Sample size		miRNA profile	qPCR quantitation method	Proximal/Distal	Early/Late stage
			CRC	Control				
1	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-17-92 cluster*	Relative	Y	Y
2	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-17	Relative	—	—
3	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-18a	Relative	—	—
4	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-19a	Relative	—	—
5	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-19b	Relative	—	—
6	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-20a	Relative	—	—
7	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-92a	Relative	—	—
8	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-21	Relative	Y	Y
9	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-135a, miR-135b	Relative	Y	Y
10	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-135a	Relative	—	—
11	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-135b	Relative	—	—
12	Koga Y (2010) <sup>11</sup>	Japan	119	197	miR-17-92 cluster*, miR-21, miR-135a/b	Relative	Y	Y
13	Kalimutho M (2011) <sup>12</sup>	Italy	40	35	miR-144-5p	Relative	—	—
14	Wu CW (2012) <sup>20</sup>	Hong Kong	101	88	miR-21	Absolute	Y	Y
15	Wu CW (2012) <sup>20</sup>	Hong Kong	101	88	miR-92a	Absolute	Y	Y
16	Wu CW (2012) <sup>20</sup>	Hong Kong	101	88	miR-21, miR-92a	Absolute	Y	Y
17	Kuriyama S (2012) <sup>21</sup>	Japan	126	138	miR-106a	Relative	—	—
18	Kuriyama S (2012) <sup>21</sup>	Japan	126	138	miR-21, miR-92a, miR-106a	Relative	—	—
19	Kanaoka S (2013) <sup>22</sup>	Japan	126	138	miR-21	Relative	—	—
20	Kanaoka S (2013) <sup>22</sup>	Japan	126	138	miR-92a	Relative	—	—
21	Koga Y (2013) <sup>23</sup>	Japan	107	117	miR-106a	Relative	—	—
22	Zhao HJ (2014) <sup>24</sup>	China	20	28	miR-194	Relative	—	—
23	Wu CW (2014) <sup>25</sup>	Hong Kong	109	104	miR-135b	Absolute	—	—
24	Yau TO (2014) <sup>26</sup>	Hong Kong	198	198	miR-221	Absolute	—	—
25	Yau TO (2014) <sup>26</sup>	Hong Kong	198	198	miR-18a	Absolute	—	—
26	Yau TO (2014) <sup>26</sup>	Hong Kong	198	198	miR-221, miR-18a	Absolute	—	—
27	Yau TO (2014) <sup>26</sup>	Hong Kong	198	198	miR-221, miR-135b	Absolute	—	—
28	Yau TO (2014) <sup>26</sup>	Hong Kong	198	198	miR-18a, miR-135b	Absolute	—	—
29	Yau TO (2014) <sup>26</sup>	Hong Kong	198	198	miR-221, miR-18a, miR-135b	Absolute	—	—
30	Phua LC (2014) <sup>27</sup>	Singapore	28	17	miR-223	Relative	—	—
31	Phua LC (2014) <sup>27</sup>	Singapore	28	17	miR-451	Relative	—	—
32	Chang PY (2016) <sup>13</sup>	Taiwan	309	138	miR-223, miR-92a, miR-16, miR-106b	Relative	—	—
33	Chang PY (2016) <sup>13</sup>	Taiwan	309	138	miR-223, miR-92a	Relative	—	—
34	Yau TO (2016) <sup>14</sup>	Hong Kong	198	198	miR-20a	Absolute	—	—
35	Yau TO (2016) <sup>14</sup>	Hong Kong	198	198	miR-20a, miR-92a	Absolute	—	—
36	Yau TO (2016) <sup>14</sup>	Hong Kong	198	198	miR-20a, miR-135b	Absolute	—	—
37	Zhu Y (2016) <sup>15</sup>	China	51	80	miR-29a	Relative	—	—
38	Zhu Y (2016) <sup>15</sup>	China	51	80	miR-223	Relative	—	—
39	Zhu Y (2016) <sup>15</sup>	China	51	80	miR-224	Relative	—	—
40	Liu H (2016) <sup>16</sup>	China	150	98	miR-21	Relative	—	—
41	Liu H (2016) <sup>16</sup>	China	150	98	miR-146a	Relative	—	—
42	Liu H (2016) <sup>16</sup>	China	150	98	miR-21, miR-146a	Relative	—	—
43	Xue Y (2016) <sup>17</sup>	China	50	50	miR-141	Relative	—	—
44	Xue Y (2016) <sup>17</sup>	China	50	50	miR-92a	Relative	—	—
45	Bastaminejad S (2017) <sup>18</sup>	Iran	40	40	miR-21	Relative	—	—
46	Wu CW (2017) <sup>19</sup>	USA	29	115	miR-451a, miR-144-5p	Relative	Y	Y

**Table 1.** All publications on faecal-based microRNAs for the detection of colorectal cancer. \*The miR-17-92 cluster includes miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1, and miR-92a.

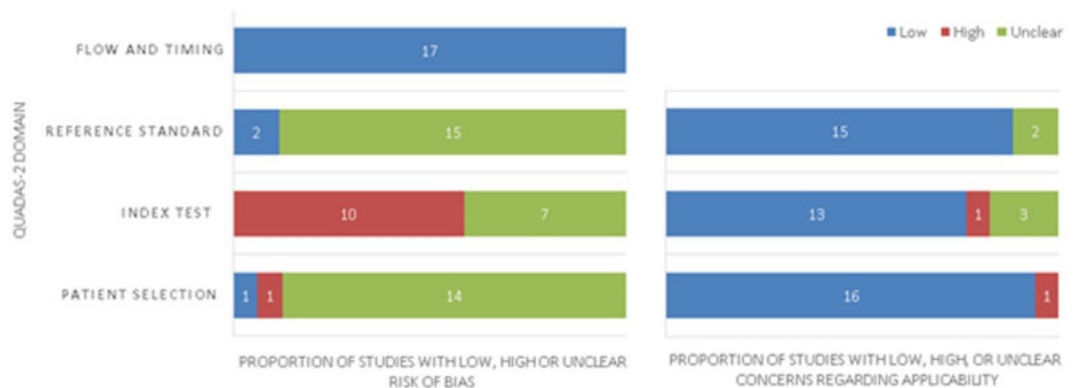
low risk of bias in the reference standard, since all studies were histopathologically confirmed prior to the index test using either TNM or Dukes' staging (Suppl. Table 1A,B). Concern about applicability in all domains was low.

**Pooled diagnostic accuracy of faecal-based microRNA for colorectal cancer and adenoma.** The detection accuracy of faecal-based miRNAs for CRC (Table 3) as well as colonic adenoma (Table 4) were pooled and analysed using the bivariate random effects model to evaluate the overall diagnostic measurements (Fig. 3). The DOR and log DOR were 8.32 (95% CI: 6.71–10.32) and 2.12 (95% CI: 1.90–2.33) for CRC and 5.31 (95% CI: 3.55–7.94) and 1.67 (95% CI: 1.27–2.07) for adenoma, respectively. The AUC value was 0.811 for the pooled



Study ID	First author (Year) [Reference no.]	Origin of population	Sample size		miRNA profile	qPCR quantitation method
			Adenoma	Control		
I	Wu CW (2012) <sup>20</sup>	Hong Kong	57	88	miR-21	Absolute
II	Wu CW (2012) <sup>20</sup>	Hong Kong	57	88	miR-92a	Absolute
III	Wu CW (2012) <sup>20</sup>	Hong Kong	57	88	miR-21, miR-92a	Absolute
IV	Kanaoka S (2013) <sup>22</sup>	Japan	26	138	miR-21	Relative
V	Kanaoka S (2013) <sup>22</sup>	Japan	26	138	miR-92a	Relative
VI	Wu CW (2014) <sup>25</sup>	Hong Kong	169	104	miR-135b	Absolute
VII	Liu H (2016) <sup>16</sup>	China	120	98	miR-21	Relative
VIII	Liu H (2016) <sup>16</sup>	China	120	98	miR-146a	Relative
IX	Liu H (2016) <sup>16</sup>	China	120	98	miR-21, miR-146a	Relative
X	Wu CW (2017) <sup>19</sup>	USA	31	115	miR-451a, miR-144-5p	Relative

**Table 2.** All publications on faecal-based microRNAs for the detection of colonic adenomas.



**Figure 2.** Quality assessment of included studies utilising the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) version 2. Summary of risk of bias and applicability concerns for faecal-based miRNAs in the detection of colorectal cancer.

CRC, and 0.747 for the pooled adenoma, respectively. The pooled studies of miRNAs in identification of CRC had a sensitivity of 58.8% (95% CI: 51.7–65.5%) and specificity of 84.8% (95% CI: 81.1–87.8%), whilst the pooled studies of miRNAs for identification of adenoma had a sensitivity of 57.3% (95% CI: 40.8–72.3%) and specificity of 76.1% (95% CI: 66.1–89.4%).

**Relative quantitation has a higher detection accuracy in colorectal cancer screening.** To investigate the potential of faecal-based miRNA in the non-invasive diagnosis of CRC, studies were subgrouped in different aspects to compare their detection accuracy (Table 3). For the individual/combination miRNA analysis, both presented a similar power of diagnostic accuracy. The individual miRNA analysis panel had a sensitivity of 53.5% (95% CI: 43.8–62.9%), specificity of 86.4% (95% CI: 81.8–89.9%), DOR of 7.71 (5.65–10.52), and log DOR of 2.04 (95% CI: 1.73–2.35). The combination of miRNAs had a sensitivity of 68.8% (95% CI: 63.0–74.0%), specificity of 81.6% (95% CI: 75.0–86.8%), DOR of 9.73 (95% CI: 7.51–12.60), and log DOR of 2.28 (95% CI: 2.02–2.53). The AUC value was 0.808 for the individual miRNAs, and 0.801 for the combination of miRNAs, respectively. The meta-regression analysis showed a significant effect on pooled sensitivity (Z-value: 2.310,  $P = 0.021$ ) but not in specificity (Z-value: 1.28,  $P = 0.199$ ) (Table 4). Comparing studies with large ( $n > 100$ ) versus small size ( $n \leq 100$ ), there was a sensitivity of 53.4% (95% CI: 44.4–62.3%) versus 70.6% (95% CI: 64.2–76.3%), specificity of 86.3% (95% CI: 82.3–89.6%) versus 80.8% (95% CI: 72.3–87.1%), DOR of 7.83 (95% CI: 6.11–10.04) versus 10.16 (95% CI: 6.45–16.00) and AUC of 0.811 versus 0.801 respectively. The meta-regression analysis indicated that the sample size did not significantly affect the pooled specificity (Z-value: 1.601,  $P = 0.109$ ), however it did affect the pooled sensitivity (z-value: 2.458,  $P = 0.014$ ).

The quantitation methods of absolute versus relative qPCR for faecal-based miRNAs were compared. The pooled relative quantitation qPCR method exhibited a better diagnostic accuracy in CRC (Table 3), with a sensitivity of 55.5% (95% CI: 45.9–64.7%), specificity of 88.8% (95% CI: 85.2–91.6%), DOR of 10.738 (95% CI: 7.718–14.940), log DOR of 2.37 (95% CI: 2.04–2.70) and AUC of 0.846. By contrast, the pooled absolute quantitation qPCR method exhibited sensitivity, specificity, DOR, log DOR and AUC of 67.3% (95% CI: 62.3–71.9%), 69.2% (95% CI: 69.4–77.0%), 5.706 (95% CI: 4.937–6.594), 1.74 (95% CI: 1.06–1.89) and 0.763, respectively. The meta-regression analysis revealed that the qPCR relative quantitation method in CRC affected only specificity (Z-value =  $-4.317$ ,  $P < 0.001$ ) when compared to the absolute quantification approach. The pooled relative quantitation qPCR approach exhibited a DOR of 8.32 (95% CI: 5.55–12.48) and log DOR of 2.12 (95% CI: 1.71–2.52).

	No. of Studies	AUC	Partial AUC	log DOR (95% CI)	DOR (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	+LR (95% CI)	-LR (95% CI)	Meta-Regression	
										Sensitivity Z-value, P	Specificity Z-value, P
Pooled miRNAs for CRC	46	0.811	0.624	2.12 (1.90–2.33)	8.32 (6.71–10.32)	58.8% (51.7–65.5%)	84.8% (81.1–87.8%)	3.34 (2.93–3.81)	0.47 (0.42–0.53)	—	—
<b>Sample size</b>											
Small (Case n < 100)	14	0.801	0.681	2.32 (1.86–2.77)	10.16 (6.45–16.00)	70.6% (64.2–76.3%)	80.8% (72.3–87.1%)	3.38 (2.52–4.52)	0.38 (0.32–0.46)	2.458, P = 0.014	1.601, P = 0.109
Large (Case n > 100)	32	0.811	0.561	2.06 (1.81–2.31)	7.83 (6.11–10.04)	53.4% (44.4–62.3%)	86.3% (82.3–89.6%)	3.36 (2.90–3.90)	0.51 (0.45–0.57)		
<b>Pooled individual / Combination miRNAs</b>											
Individual miRNA	31	0.808	0.593	2.04 (1.73–2.35)	7.71 (5.65–10.52)	53.5% (43.8–62.9%)	86.4% (81.8–89.9%)	3.32 (2.76–4.00)	0.54 (0.48–0.60)	2.310, P = 0.021	1.284, P = 0.199
Combination miRNAs	15	0.801	0.674	2.28 (2.02–2.53)	9.73 (7.51–12.60)	68.8% (63.0–74.0%)	81.6% (75.0–86.8%)	3.47 (2.87–4.20)	0.39 (0.34–0.44)		
<b>Quantitation method</b>											
Absolute	13	0.763	0.685	1.74 (1.06–1.89)	5.706 (4.937–6.594)	67.3% (62.3–71.9%)	69.2% (69.4–77.0%)	2.49 (2.26–2.74)	0.46 (0.41–0.51)	-1.632, P = 0.103	-4.317, P < 0.001
Relative	33	0.846	0.662	2.37 (2.04–2.70)	10.738 (7.718–14.940)	55.5% (45.9–64.7%)	88.8% (85.2–91.6%)	4.27 (3.48–5.23)	0.49 (0.43–0.55)		

**Table 3.** Subgroup analysis for pooled microRNAs in the identification of CRC. AUC, Area under curve; DOR, Diagnostic odds ratio; +LR, Positive likelihood ratio; -LR = Negative likelihood ratio; Z-value, regression coefficient.  $P < 0.05$  was considered statistically significant.

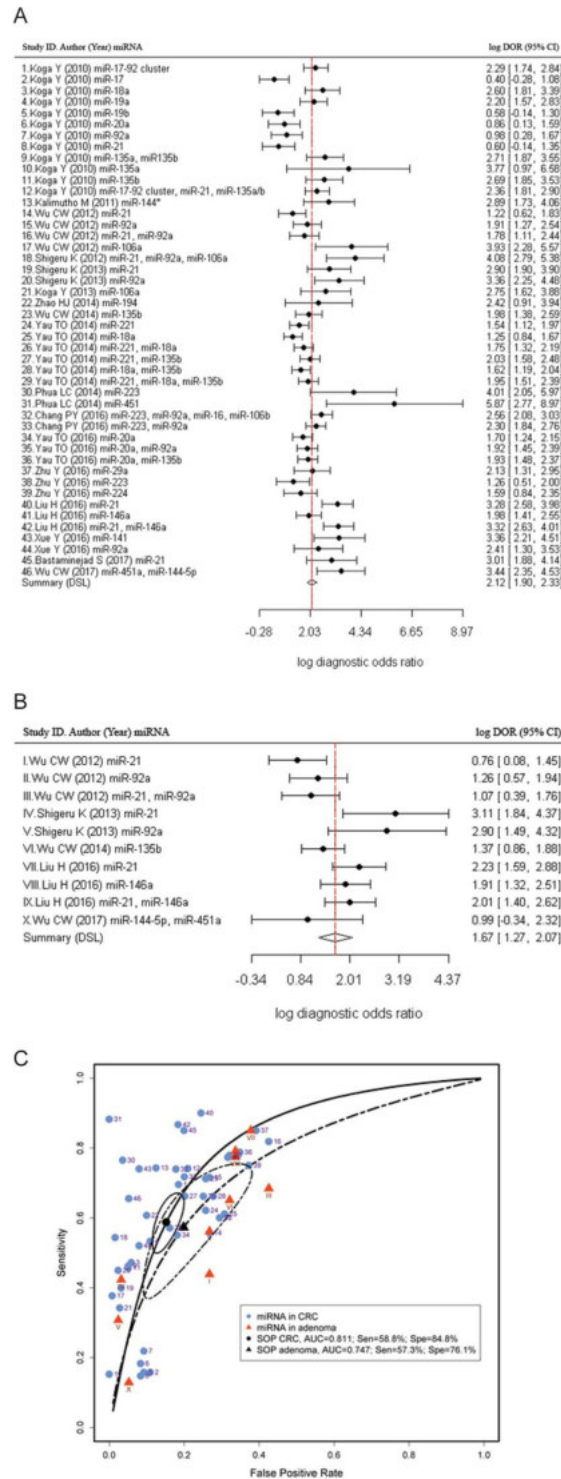
	No. of Studies	AUC	Partial AUC	log DOR (95% CI)	DOR (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	+LR (95% CI)	-LR (95% CI)	Meta-Regression	
										Sensitivity Z-value, P	Specificity Z-value, P
Pooled miRNAs for Adenoma	10	0.747	0.560	1.67 (1.27–2.07)	5.31 (3.55–7.94)	57.3% (40.8–72.3%)	76.1% (66.1–89.4%)	2.28 (1.84–2.84)	0.53 (0.41–0.68)	—	—
<b>Sample size</b>											
Small (Case n < 100)	4	0.647	0.490	1.53 (0.86–2.20)	6.36 (4.35–9.31)	42.0% (27.5–58.1%)	87.7% (70.0–95.6%)	2.23 (1.93–2.57)	0.35 (0.25–0.49)	-1.309, P < 0.001	-2.011, P = 0.044
Large (Case n > 100)	6	0.701	0.779	1.85 (1.47–2.23)	4.63 (2.37–9.04)	77.0% (67.7–84.2%)	65.7% (70.5–85.9%)	2.87 (1.71–4.83)	0.71 (0.59–0.85)		
<b>Pooled individual/Combination miRNAs</b>											
Individual miRNA	7	0.749	0.566	1.78 (1.26–2.30)	5.93 (3.52–10.00)	58.7% (41.6–73.9%)	81.6% (63.6–91.8%)	2.50 (1.86–3.36)	0.52 (0.40–0.67)	-0.185, P = 0.853	-0.194, P = 0.846
Combination miRNAs	3	0.729	0.520	1.45 (0.73–1.17)	4.26 (2.07–8.76)	52.6% (13.4–88.8%)	78.1% (42.2–94.5%)	1.97 (1.45–2.67)	0.55 (0.28–1.09)		
<b>Quantitation method</b>											
Absolute	4	0.687	0.626	1.16 (0.84–1.47)	3.18 (2.33–4.35)	59.4% (48.0–69.9%)	68.2% (36.7–52.0%)	1.83 (1.54–2.16)	0.61 (0.50–0.74)	-3.356, P = 0.001	-1.859, P = 0.063
Relative	6	0.802	0.622	2.12 (1.71–2.52)	8.32 (5.55–12.48)	56.9% (31.4–79.2%)	86.7% (67.6–95.3%)	3.01 (2.09–4.34)	0.48 (0.31–0.73)		

**Table 4.** Subgroup analysis for pooled microRNAs in the identification of adenomas. AUC, Area under curve; DOR, Diagnostic odds ratio; +LR, Positive likelihood ratio; -LR, Negative likelihood ratio; Z-value, regression coefficient.  $P < 0.05$  was considered as statistically significant.

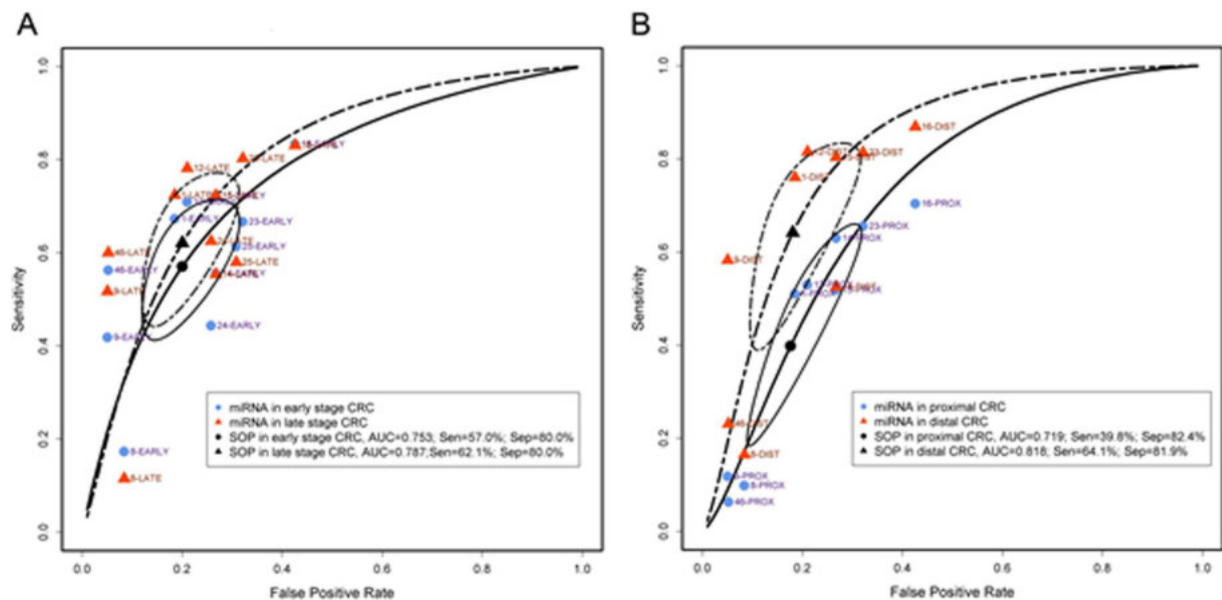
with a specificity of 86.7% (95% CI: 67.6–95.3%) and sensitivity of 56.9% (95% CI: 31.4–79.2%), compared with the absolute quantification approach (Table 3 and Suppl. Fig. 1).

Subgroup meta-regression analysis in colonic adenoma was also performed, looking at differences in sample size, pooled individual/combination miRNAs and quantitation method (Table 4). The meta-regression indicated that a small sample size significantly affected both the specificity (Z-value: -2.011,  $P = 0.044$ ) and sensitivity (Z-value: -1.309,  $P < 0.001$ ). With respect to the pooled individual/combination miRNAs, meta-regression analysis did not show significant effects in both the pooled sensitivity (Z-value: -10.85,  $P = 0.853$ ) and specificity (Z-value: -0.194,  $P = 0.846$ ). For the qPCR relative quantitation method, a significant effect was observed in the sensitivity (Z-value = -3.356,  $P < 0.001$ ) of pooled miRNAs.

**Differences in detecting CRC depending on tumour stage and location.** Meta-analysis on early versus late stage CRC as well as proximal versus distal CRC were performed to further evaluate the diagnostic ability of miRNAs. Pooled faecal miRNAs had a sensitivity of 57.0% (95% CI: 44.4–68.8%), specificity of 80.0% (95% CI: 71.1–86.7%), DOR of 5.58 (95% CI: 3.62–8.62) and log DOR of 1.72 (95% CI: 1.29–2.15) with respect to the diagnosis of early stage CRC, whilst in late stage CRC pooled miRNAs had a sensitivity of 62.1% (95% CI:



**Figure 3.** Diagnostic accuracy of pooled microRNAs in identification of colorectal cancer and colonic adenoma. Log Diagnostic Odds ratios in (A) CRC was 2.12 (95% CI: 1.90–2.33) and (B) adenoma was 1.67 (95% CI: 1.27–2.07). (C) Summary receiver operating characteristic curves (SROC) for pooled miRNAs in CRC and colonic adenoma. The pooled miRNAs for CRC (n = 46) had a sensitivity of 58.8% (95% CI: 51.7–65.5%), specificity of 84.8% (95% CI: 81.1–87.8%) and AUC of 0.811. The pooled miRNAs for adenoma (n = 10) had a sensitivity of 57.3% (95% CI: 40.8–72.3%), specificity of 76.1% (95% CI: 66.1–89.4%) and AUC of 0.747. The number next to the dot/triangle corresponds to the study ID in Table 1 (Blue dots: CRC) or Table 2 (Red triangles: colonic adenoma). The circular regions (95% confidence contour) contain likely combinations of the mean value of sensitivity and specificity. Sen, sensitivity; Spe, specificity; SOP, summary operating point.



**Figure 4.** Diagnostic accuracy in early stage versus late stage and proximal versus distal colorectal cancer. **(A)** Summary receiver operating characteristic curves (SROC) for early ( $n = 11$ ) and late ( $n = 11$ ) stage CRC. **(B)** SROC for proximal ( $n = 9$ ) and distal ( $n = 9$ ) CRC. The number next to the blue dot/red triangle corresponds to the study ID in Table 1. The circular regions (95% confidence contour) contain likely combinations of the mean value of sensitivity and specificity. Sen = sensitivity; Spe = specificity SOP, summary operating point. #Early stage CRC includes TMN stages 0 + I + II or Dukes' stage A + B; late stage CRC includes CRC stages III + IV or Dukes' stage C + D. Proximal CRC is defined as from cecum to transverse colon, and distal CRC is defined as from the splenic flexure to the rectum.

47.8–74.6%), specificity of 80.0% (95% CI: 71.1–86.7%), DOR of 6.70 (95% CI: 4.34–10.36) and log DOR of 1.90 (95% CI: 1.47–2.34) (Fig. 4A,B and Table 5). In proximal CRC, the pooled sensitivity, specificity, DOR, log DOR and AUC were 39.8% (95% CI: 21.8–61.0%), 82.4% (95% CI: 71.5–89.7%), 3.44 (95% CI: 2.53–4.66), 1.23 (95% CI: 0.93–1.54) and 0.719, respectively. For distal CRC, the pooled sensitivity, specificity, DOR, log DOR and AUC were 64.1% (95% CI: 43.9–80.3%), 81.9% (95% CI: 71.5–89.1%), 8.51 (95% CI: 4.97–14.57), 2.14 (95% CI: 1.60–2.68) and 0.818, respectively (Table 6 and Suppl. Fig. 2).

**The detection accuracy of individual microRNAs.** Each individual miRNA reported by more than one research group was pooled for an accuracy estimation (Table 6). miR-21 was reported by five CRC and three colonic adenoma studies<sup>11,16,18,20,22</sup>. Pooled miR-21 in CRC had an AUC of 0.843, DOR of 9.28 (95% CI: 2.97–28.97) and log DOR of 2.23 (95% CI: 1.09–3.37) whilst pooled miR-21 in adenoma had an AUC of 0.771, DOR of 7.10 (96% CI: 1.99–25.34) and log DOR of 1.96 (96% CI: 1.96–3.23) (Fig. 5A and Table 6). The miR-21-related combination pool for CRC detection had an AUC of 0.843, with a DOR of 16.73 (95% CI: 7.00–39.94) and log DOR of 2.82 (95% CI: 1.95–3.69) from four different CRC studies. miR-92a was reported in four CRC and two adenoma studies<sup>11,20–22</sup>. The AUC, DOR and log DOR were 0.794, 8.57 (95% CI: 3.30–22.27) and 2.15 (95% CI: 1.19–3.10) for pooled miR-92a alone in the diagnosis of CRC, and 0.635, 0.467, 7.08 (95% CI: 1.43–34.97) and 1.96 (95% CI: 0.36–3.55) for pooled miR-92a alone in the diagnosis of colonic adenoma, respectively (Fig. 5B and Table 6).

The miR-92a-related combination pool for CRC was reported in five CRC studies, with AUC 0.791, DOR 10.47 (95% CI: 6.46–16.98) and log DOR 2.35 (95% CI: 1.87–2.83). The pooled miR-21 plus miR-92a combination exhibited an AUC of 0.837, with a specificity of 87.8% (95% CI: 79.5–93.0%), sensitivity of 53.7% (95% CI: 33.4–74.8%) and DOR of 2.19 (95% CI: 1.48–2.91). miR-20a, miR-106a, miR-135b and miR-223 were reported in two different articles, with an AUC of 0.797, 0.416, 0.798 and 0.777, DOR of 3.87 (95% CI: 1.75–8.55), 29.33 (95% CI: 7.74–111.11), 10.18 (95% CI: 4.91–21.09) and 14.69 (95% CI: 0.66–328.44), respectively (Table 6 and Suppl. Fig. 2).

**Publication bias evaluation.** The Deeks' funnel plot asymmetry test was utilised to evaluate the potential publication bias from each included faecal-based miRNA study. The slope coefficient was associated with  $P = 0.03$  in the pooled miRNAs in CRC, and  $P = 0.61$  in pooled miRNAs in colonic adenoma studies, indicating that significant asymmetry was found in the CRC dataset (Fig. 6A) but not in the colonic adenoma dataset (Fig. 6B). The combination of CRC and colonic adenoma resulted in a  $P = 0.11$  (Suppl. Fig. 3).

## Discussion

To evaluate the diagnostic value of faecal-based miRNAs, data from 17 eligible publications, including 46 studies on miRNAs in CRC and 10 studies on colonic adenoma, corresponding to 6475, 783 and 5569 faecal-based miRNA tests in CRC patients, adenoma and healthy control volunteers, respectively, were subjected to meta-analysis. Our study reveals that pooled faecal-based miRNAs have a relatively high detection accuracy for

miRNA	No. of Studies	AUC	Partial AUC	log DOR (95% CI)	DOR (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	+LR (95% CI)	−LR (95% CI)
<b>Staging# (Based on the data from the included studies)</b>									
Early	11	0.753	0.582	1.72 (1.29–2.15)	5.58 (3.62–8.62)	57.0% (44.4–68.8%)	80.0% (71.1–86.7%)	2.68 (2.10–3.43)	0.54 (0.43–0.68)
Late	11	0.787	0.627	1.90 (1.47–2.34)	6.70 (4.34–10.36)	62.1% (47.8–74.6%)	80.0% (71.1–86.7%)	2.82 (2.22–3.59)	0.45 (0.33–0.61)
<b>Tumour location (Based on the data from the included studies)</b>									
Proximal	9	0.719	0.488	1.23 (0.93–1.54)	3.44 (2.53–4.66)	39.8% (21.8–61.0%)	82.4% (71.5–89.7%)	2.08 (1.76–2.44)	0.75 (0.64–0.88)
Distal	9	0.818	0.675	2.14 (1.60–2.68)	8.51 (4.97–14.57)	64.1% (43.9–80.3%)	81.9% (71.5–89.1%)	3.04 (2.31–4.01)	0.41 (0.28–0.61)

**Table 5.** Subgroup analysis for pooled microRNAs in association with CRC staging and tumour location. DOR, Diagnostic odds ratio; +LR, Positive likelihood ratio; −LR, Negative likelihood ratio. #Early stage CRC includes TNM stages 0 + I + II or Dukes' stage A + B; Late stage CRC includes CRC stages III + IV or Dukes' stage C + D.

CRC. However, the lack of consensus regarding the optimal quantitation method, data normalisation, and selection of control subjects, may present obstacles to clinical application.

qPCR-based studies were the subject of our meta-analysis analysis. This approach for miRNA level quantification is advantageous compared to others in that it is fast and easily adoptable in a clinical setting. However, it comes with limitations that relate to the method for miRNA isolation and the selection of the appropriate reference/normalisation control. Although reference quantitation method demonstrated a better diagnostic accuracy compared to absolute quantitation, it is important to acknowledge that a variety of internal controls were used as references, including RUN6B(U6)<sup>11,15–18,24</sup>, miR-24<sup>23</sup>, miR-200b-3p<sup>19</sup>, miR-378<sup>12</sup>, miR-1202<sup>27</sup> and miR-4257<sup>27</sup>. Increasing evidence suggests that RUN6B may not be a suitable endogenous control for miRNAs<sup>31,32</sup> due to its rapid degradation in faeces<sup>20</sup>. miRNAs used as internal controls also have functions in the host cell, and their deregulation could interfere with the detection accuracy. For example, miR-24, a proposed tumour suppressor miRNA in CRC, controls cellular proliferation independently of p53 by targeting the 3'UTR of dihydrofolate reductase (DHFR) mRNA<sup>33,34</sup>. Deregulation of plasma miR-378 was also found in CRC patients<sup>35</sup>. miR-4257 has been reported to be down-regulated in bladder cancer cell lines and up-regulated in the plasma of patients with recurrence of non-small cell lung cancer (NSCLC)<sup>36,37</sup>. Peripheral levels of miR-1202 predicts and mediates the response to anti-depressants, specifically regulating the expression of metabotropic glutamate receptor-4 (GRM4) with levels correlating to changes in brain activity<sup>38–40</sup>. miR-1202 is deregulated in different types of cancers, such as breast cancer<sup>41</sup>, gastric cancer<sup>42</sup> and clear cell papillary renal cell carcinoma<sup>43</sup>. Absolute quantitation was employed in several studies for faecal-based miRNA screening<sup>13,14,20,26</sup> (Table 1), however, this necessitates a standard curve which depends on the quantification detection method and does not eliminate potential contamination by gut bacteria DNA/RNA<sup>14</sup>.

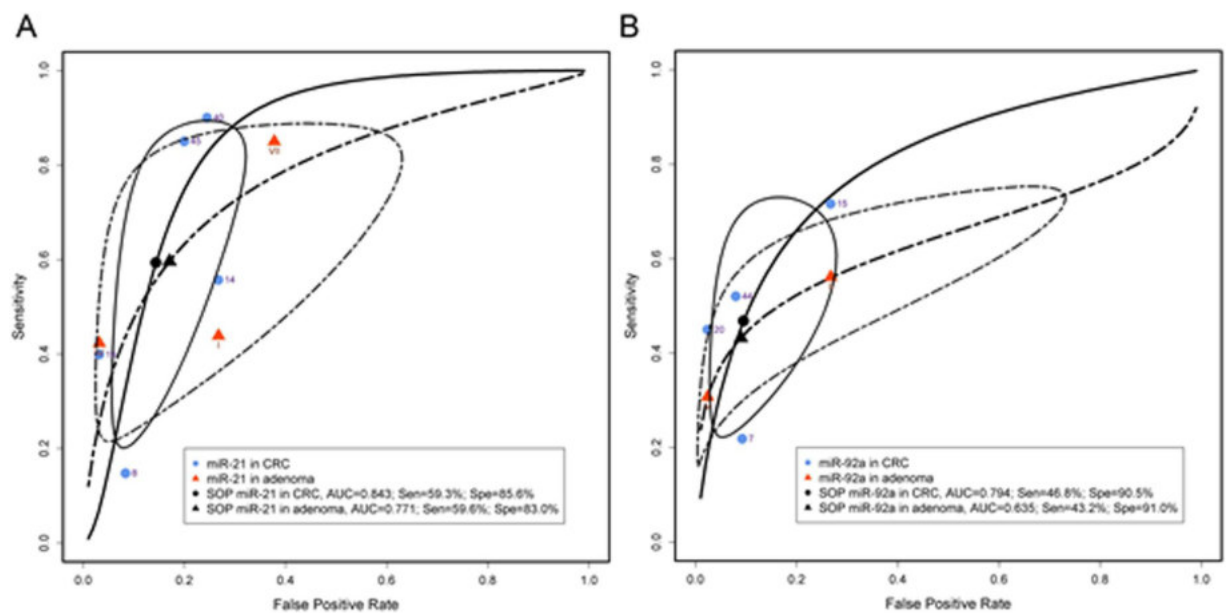
The combination of FIT and stool-based miRNA markers may increase detection accuracy to overcome this problem. A previous study indicated that the combination of miR-21 and miR-92a with FIT had a specificity of 96.8% and sensitivity of 78.4% while FIT alone only had a specificity of 98.4% and sensitivity of 66.7%<sup>22</sup>. A parameter that should be considered is the presence/absence of occult blood in samples as miRNAs expressed in blood cells may interfere with the assay, altering the levels of specific miRNAs. In an effort to assess the potential contribution of blood in faecal miRNA levels we have retrieved a list of circulating miRNAs<sup>44</sup>. Comparison showed that 8 miRNAs are detected in both blood and faecal specimens (Suppl. Fig. 4). This finding does not imply that blood cells are responsible for the alterations in the levels of these miRNAs, as their origin may as well be the tumour. Optimally, a controlled study including comparisons between samples positive and negative for FOBT/FIT could address the relative contribution. A more inclusive approach employing miRNA analyses and comparisons between matched blood, faecal specimens and tumours or colonic tissues would be most informative about the source of changes in miRNA levels. Furthermore, other colonic pathologies like inflammatory bowel diseases are characterised by deregulation of miRNAs detectable in tissues and serum<sup>45–48</sup> and the presence of occult blood in faeces. A comprehensive analysis would include samples from different pathologies of the colon, assess and identify disease-specific miRNA signatures and their diagnostic/prognostic properties.

A clear conclusion on which quantitation method is more suitable cannot be drawn with the currently available data. The use of multiple internal controls or the geometric mean, using a multiplex screening method, such as microarrays or next generation sequencing, would provide the optimal means of normalisation. Alternatively, the NanoString nCounter technology enables profiling of around 800 molecular targets in one single reaction by utilising molecular “barcodes”. This approach normalises the data by using multiple targets, and more importantly quantifies multiple miRNAs which can be used simultaneously as biomarkers to improve detection accuracy. In addition, this platform overcomes the need for data processing and bioinformatic analysis expertise, as in the case of microarrays or high-throughput sequencing, thus may be easily utilised in a clinical setting<sup>46</sup>.

To evaluate the potential detection efficiency for each individual miRNA, individual miRNAs reported in more than one study were grouped to evaluate its detection accuracy. In this meta-analysis, miR-21 and miR-92a were the most commonly reported faecal-based miRNAs (Table 6). Numerous studies have characterised the functional roles of these two miRNAs in CRC pathogenesis and aggressiveness. Up-regulation of miR-21 and miR-92a promotes CRC cell migration, invasion and proliferation<sup>11,16,18,20,22</sup>, and inhibition of apoptosis<sup>49–51</sup>. Several significant targets of miR-21 are associated with CRC malignancy, such as phosphatase and tensin homolog (PTEN)<sup>49,52</sup>,

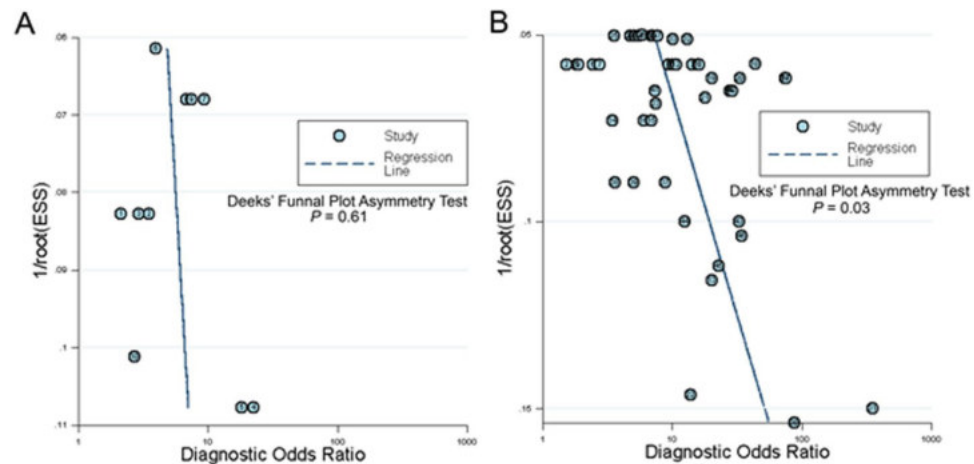
miRNA	Diagnosis	No. of Studies	AUC	Partial AUC	log DOR (95% CI)	DOR (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	+LR (95% CI)	-LR (95% CI)
miR-21	Adenoma	3	0.771	0.598	1.96 (1.96–3.23)	7.10 (1.99–25.34)	59.6% (27.7–85.0%)	83.0% (47.2–96.4%)	2.97 (1.45–6.07)	0.49 (0.27–0.90)
	CRC	5	0.843	0.549	2.23 (1.09–3.37)	9.28 (2.97–28.97)	59.3% (26.3–85.6%)	85.6% (72.2–93.2%)	3.38 (2.07–5.53)	0.43 (0.28–0.68)
miR-21-related combination	CRC	4	0.843	0.764	2.82 (1.95–3.69)	16.73 (7.00–39.94)	75.7% (60.3–86.5%)	85.0% (55.1–96.3%)	4.31 (2.22–8.39)	0.31 (0.20–0.46)
miR-92a	Adenoma	2	0.635	0.467	1.96 (0.36–3.55)	7.08 (1.43–34.97)	43.2% (20.1–69.8%)	91.0% (41.6–99.3%)	4.70 (0.80–27.60)	0.66 (0.54–0.81)
	CRC	4	0.794	0.537	2.15 (1.19–3.10)	8.57 (3.30–22.27)	46.8% (26.3–68.4%)	90.5% (77.1–96.4%)	4.53 (2.17–9.43)	0.57 (0.41–0.81)
miR-92a-related combination	CRC	5	0.791	0.685	2.35 (1.87–2.83)	10.47 (6.46–16.98)	68.2% (56.8–77.7%)	83.7% (65.9–93.2%)	3.71 (2.44–5.64)	0.40 (0.33–0.49)
miR-21 + miR-92a	CRC	9	0.837	0.548	2.19 (1.48–2.91)	8.97 (4.39–18.29)	53.7% (33.4–74.8%)	87.8% (79.5–93.0%)	3.68 (2.54–5.33)	0.51 (0.40–0.65)
miR-20a	CRC	2	0.797	0.367	1.35 (0.56–2.15)	3.87 (1.75–8.55)	34.4% (9.0–73.6%)	87.5% (73.6–94.6%)	2.84 (2.13–3.80)	0.70 (0.44–1.13)
miR-106a	CRC	2	0.416	0.356	3.38 (2.05–4.71)	29.33 (7.74–111.11)	36.1% (30.4–42.2%)	98.0% (94.7–99.2%)	18.85 (5.44–65.35)	0.65 (0.59–0.72)
miR-135b	CRC	2	0.798	0.656	2.32 (1.59–3.05)	10.18 (4.91–21.09)	63.1% (29.4–87.5%)	86.2% (41.8–98.2%)	4.44 (1.23–16.09)	0.44 (0.26–0.77)
miR-223	CRC	2	0.777	0.684	2.69 (-0.42–5.79)	14.69 (0.66–328.44)	67.9% (47.9–83.0%)	87.4% (39.3–98.7%)	5.43 (0.56–52.59)	0.41 (0.18–0.92)

**Table 6.** Diagnostic accuracy of individual microRNAs and microRNA combinations. DOR, Diagnostic odds ratio; +LR, Positive likelihood ratio; -LR, Negative likelihood ratio.



**Figure 5.** Diagnostic accuracy in pooled miR-21 and miR-92a. (A) SROC for pooled miR-21 in the detection of CRC (n = 5) and colonic adenoma (n = 3). (B) SROC for pooled miR-92a in the detection of CRC (n = 4) and colonic adenoma (n = 2). The number next to the dot/triangle corresponding to the study ID in Table 1 (Blue dots: CRC) or Table 2 (Red triangles: colonic adenoma). Sen, sensitivity; Spe, specificity; SOP, summary operating point. The circular regions (95% confidence contour) contain likely combinations of the mean value of sensitivity and specificity.

programmed cell death protein 4 (PDCD4)<sup>53,54</sup>, and *ras* homolog gene family member B (RhoB)<sup>55</sup>. Among these, PTEN was reported frequently silenced in CRC by miR-21, resulting in PI3K/AKT pathway activation and induction of tumour formation<sup>49,52</sup>. Recently, a long non-coding RNA (LINC00312) suppressed in CRC was shown to regulate miR-21 levels through its function as a miRNA sponge, thereby regulating PTEN expression<sup>56</sup>. miR-92a has been shown to disrupt the expression of several tumour suppressors such as PTEN<sup>57,58</sup>, Dickkopf WNT Signalling Pathway Inhibitor 3 (DKK3)<sup>57</sup>, Kruppel-like factor 4 (KLF4)<sup>59</sup> and mothers against decapentaplegic homolog 7 (SMAD7)<sup>60</sup>. Hence, miR-92a activates the PI3K/AKT, WNT/ $\beta$ -catenin and BMP/Smad pathways and enhances



**Figure 6.** Deeks' funnel plot asymmetry test for the assessment of potential bias in microRNA assays. (A) Pooled miRNAs for CRC and, (B) Pooled miRNAs for colonic adenoma.

tumorigenesis. Subject to this analysis five studies reported the use of miR-21 in the identification of CRC, and three studies reported its use in identification of adenomas<sup>11,16,18,20,22</sup>. Four studies reported the utility of miR-92a in the identification of CRC, and two studies in identification of adenomas<sup>11,17,20,22</sup>. miR-21 had a better detection accuracy range compared with miR-92a, with a DOR of 9.28 (95% CI: 2.97–28.97) and summary AUC of 0.843. Panels including a combination of either miR-21 or miR-92a, as well as panels including both miR-21 and miR-92a demonstrated a small improvement in detection (Fig. 5 and Table 6). However, due to the small number of published studies, with each having wide confidence intervals, a direct comparison between two faecal-based miRNAs may not be accurate. Additional data are needed to limit potential errors.

The FOBT or FIT, have limited sensitivity for detecting proximal compared with distal CRC<sup>61,62</sup>. This is due to the degradation of haemoglobin. Hence, tumour location analysis for faecal-based miRNA detection was also considered and reported by several studies – with none of them reporting a statistical difference. In this study, the results between pooled miRNAs for proximal and distal CRC reveal differences associated with tumour location, with an AUC of 0.719 versus 0.818, and DOR of 3.44 (95% CI: 2.53–4.66) versus 8.51 (95% CI: 4.97–14.57) (Fig. 4B and Table 5).

Our study is characterised by many strengths but should be interpreted in the context of specific shortcomings. Firstly, subgroup analysis suggested that the combination of faecal miRNAs exhibited a good accuracy for CRC and colonic adenoma patients screening (Tables 3, 4 and Fig. 3). However, certain combinations of miRNAs may not significantly improve the detection accuracy. For example, the panel containing miR-223, miR-92a, miR-16 and miR-106b had a sensitivity of 73.9%, specificity of 82.2% and AUC of 0.84<sup>13</sup>, whereas the combination of miR-18a and miR-135b only had a sensitivity of 66%, specificity of 72% and AUC of 0.75<sup>26</sup>. Therefore, an optimal miRNA combination panel should be prioritised. Secondly, the majority of studies were performed in East Asia (Hong Kong, Taiwan, China, Japan and Singapore) (Table 1) with only one study in the USA, Europe and the Middle East, making it unclear whether the ethnic background of participants has an influence on the expression of miRNAs in CRC. Thirdly, due to the high cost of colonoscopy, the majority of test subjects were recruited from the corresponding clinics. This may result in a degree of bias, since the subjects are not representative of the general population. Last but not least, the publication bias analysis revealed that pooled miRNAs in CRC have a significant asymmetry ( $P = 0.03$ ). This may be due to file-drawer effects, bias from the studies with small same sizes, lack of clarity in reporting the results for some publications, or the level of detail provided being lower than the one required for our analysis. Consequently, some studies were excluded, resulting in a possible bias in our meta-analysis (Fig. 6A).

In conclusion, faecal-based miRNAs show a relatively high accuracy for the non-invasive detection of colonic adenomas and CRC in the studied population. The use of a panel of miRNAs as biomarkers may result in a higher CRC detection rate, while the combination of miRNA biomarkers with FOBT or FIT may increase the detection accuracy. Large, ideally multi-centre, double-blinded randomised controlled trials are needed to establish the value of miRNAs as biomarkers in CRC screening within the general population.

## Methods

**Overview.** The study protocol followed the Cochrane Handbook for Diagnostic Test Accuracy Review<sup>63</sup> and the Preferred Reporting Items in Systematic Reviews and the Meta-Analysis statement (PRISMA)<sup>64</sup>. Investigators of each of the original studies obtained approval from their local ethics committee and had written, informed patient consent.

**Literature search strategy.** The search strategy was designed to identify any studies describing the diagnostic value of faecal-based miRNA for CRC and colonic adenoma patients. After an initial search for articles in PubMed, assessments of key terms within the title and abstract were conducted. A full systematic search using the established key terms was adopted for the following databases: PubMed, Ovid Embase, The Cochrane Library, Scopus and Web of Science. The search terms used were “miRNA OR microRNA OR miR” AND “colorectal

cancer OR colorectal tumor OR colorectal adenocarcinoma OR colorectal carcinoma OR colorectal neoplasm OR colon cancer OR colonic adenoma OR colonic adenocarcinoma OR stomach cancer OR rectal cancer OR CRC” AND “stool OR feces”. The search formulas are available as supplementary data (Supplementary data 1). Manual searching of related citations and reference lists was undertaken. Book chapters, letters to editors, commentaries, editorials, patents, and non-peer reviewed articles were excluded. Two investigators independently screened the search results, initially through articles’ title and abstract. The filtered candidate articles were then scrutinised independently through full-text reading. Discrepancies were resolved through discussion between the two investigators.

**Study selection criteria.** All research articles in any language published up to November 17, 2017 were eligible for inclusion. An electronic data extraction form was developed, and pre-tested, with data extracted by two researchers. Eligible studies included in this meta-analysis adhered to the following criteria: (1) studies evaluated the diagnostic value of miRNAs for detecting human CRC or colonic adenomas; (2) all CRC patients involved in the study had been confirmed by histology; (3) studies contained data on miRNAs’ sensitivity, specificity, and sample size to enable reconstruction of the diagnostic  $2 \times 2$  contingency table. Exclusion criteria were set as follows: (1) duplicated studies, the later ones were excluded; (2) publications that were unrelated to the diagnostic value of miRNAs for CRC; (3) incomplete data reporting. The detection accuracy of miRNAs between proximal (from cecum to transverse colon) and distal (from splenic flexure to rectum) CRC, as well as between early (CRC stages 0 + I + II or Dukes’ stages A + B) and late (CRC stages III + IV or Dukes’ stages C + D) stage CRC were evaluated separately if investigators reported the location and stage. Each individual miRNA and miRNA combinations were grouped together if found in more than two studies.

**Risk of bias.** The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) was utilised to assess the quality of included publications, evaluating four key domains (“Patient Selection”, “Index Test”, “Reference Standard”, and “Flow and Timing”) in two categories (risk of bias and applicability of diagnostic accuracy studies)<sup>30</sup>. Each category in all publications was judged as low, high or unclear based on the assessment criteria provided. Assessment of each included study was performed by two investigators, with disagreements resolved by consensus after discussion.

**Data synthesis.** A meta-analysis of diagnostic test accuracy was conducted on faecal-based non-invasive miRNA tests through a bivariate random effects modelling approach. The bivariate model accounts for the correlation between the studies’ sensitivity and specificity in two different levels. The first level represents a variability between sensitivity and specificity within one study; the second level represents the heterogeneity in diagnostic performance of the index test across the testing studies. Random effects meta-analysis methods were applied in our study as heterogeneity is presumed to exist.

Statistical analyses in this study were performed using the statistical package *mada* version 0.4.8 in R (version 3.4.3) to implement the bivariate normal approach of Reitsma *et al.*<sup>65</sup>. Sensitivity, specificity, diagnostic odds ratio (DOR), log DOR, positive likelihood ratio (+LR) and negative likelihood ratio (–LR) were calculated along with their 95% confidence interval (95% CI) based on the random effects model (DerSimonian and Laird method) with continuity correction<sup>66,67</sup>. Summary receiver operating characteristic (sROC) curves, area under the curve (AUC) and partial AUC were also utilised to examine the pooled faecal-based miRNAs in CRC, adenoma and the subgroups. Potential sources of heterogeneity were investigated using subgroup and bivariate meta-regression (restricted maximum likelihood (REML) estimators) analysis. The Deeks’ funnel plot asymmetry test was examined using the *midas* package in Stata (version 12).

**Interpretation of diagnostic test accuracy statistics.** The AUC was interpreted in four-grades: >0.97, excellent; 0.93–0.96, very good; 0.75–0.92, good; < 0.75, not accurate<sup>68</sup>. The values of –LR and +LR were also divided into four categories. The –LR values < 0.1, 0.1–0.2, 0.2–0.5 and > 0.5 were identified as large, moderate, small and not meaningful decreases in probability, respectively<sup>69</sup>. The +LR values > 10, 5–10, 2–5 and < 2 were classified as large, moderate, small and not meaningful increases in probability, respectively<sup>69</sup>.  $P < 0.05$  was considered statistically significant.

## References

1. Torre, L. A. *et al.* Global cancer statistics, 2012. *CA. Cancer J. Clin.* **65**, 87–108 (2015).
2. Smith, R. A. *et al.* American Cancer Society guidelines for the early detection of cancer: update of early detection guidelines for prostate, colorectal, and endometrial cancers. Also: update 2001–testing for early lung cancer detection. *CA. Cancer J. Clin.* **51**, 38–75 (2001).
3. Yau, T. O. Precision treatment in colorectal cancer: Now and the future. *JGH Open*, <https://doi.org/10.1002/jgh3.12153> (2019).
4. Halloran, S. P. Bowel cancer screening. *Surgery (Oxford)* **27**(9), 397–400 (2009).
5. Logan, R. F. A. *et al.* Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut* **61**, 1439–1446 (2012).
6. Simon, J. B. Should all people over the age of 50 have regular fecal occult-blood tests? Postpone population screening until problems are solved. *N. Engl. J. Med.* **338**, 1151–2; discussion 1154–5 (1998).
7. Doubeni, C. A. *et al.* Fecal Immunochemical Test (FIT) for Colon Cancer Screening: Variable Performance with Ambient Temperature. *J. Am. Board Fam. Med.* **29**, 672–681 (2016).
8. Mayor, S. One in four cases of bowel cancer in England are diagnosed only after emergency admission. *BMJ* **345**, e7117 (2012).
9. Liz, J. & Esteller, M. lncRNAs and microRNAs with a role in cancer development. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859**, 169–176 (2016).
10. Esquela-Kerscher, A. & Slack, F. J. Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer* **6**, 259–69 (2006).
11. Koga, Y. *et al.* MicroRNA expression profiling of exfoliated colonocytes isolated from feces for colorectal cancer screening. *Cancer Prev. Res. (Phila.)* **3**, 1435–42 (2010).



12. Kalimutho, M. *et al.* Differential expression of miR-144\* as a novel fecal-based diagnostic marker for colorectal cancer. *Journal of Gastroenterology* **46**, 1391–1402 (2011).
13. Chang, P.-Y. *et al.* MicroRNA-223 and microRNA-92a in stool and plasma samples act as complementary biomarkers to increase colorectal cancer detection. *Oncotarget* **7**, 10663–75 (2016).
14. Yau, T. O. *et al.* microRNA-20a in human faeces as a non-invasive biomarker for colorectal cancer. *Oncotarget* **7**, 1559–68 (2016).
15. Zhu, Y. *et al.* Fecal miR-29a and miR-224 as the noninvasive biomarkers for colorectal cancer. *Cancer Biomark.* **16**, 259–264 (2015).
16. Liu, H. *et al.* MicroRNA-21 and microRNA-146a identification in stool and its clinical significance in colorectal neoplasms. *Int. J. Clin. Exp. Med.* **9**, 16441–16449 (2016).
17. XUE, Y. *et al.* [Values of fecal microRNA-141, -17-3p and -92a-3p in the diagnosis and prognostic evaluation of colorectal cancer]. *Tumor* **36**, 901–907 (2016).
18. Bastaminejad, S. *et al.* Investigation of MicroRNA-21 Expression Levels in Serum and Stool as a Potential Non-Invasive Biomarker for Diagnosis of Colorectal Cancer. *Iran. Biomed. J.* **21**, 106–113 (2017).
19. Wu, C. W. *et al.* Novel Approach to Fecal Occult Blood Testing by Assay of Erythrocyte-Specific microRNA Markers. *Dig. Dis. Sci.* **62**, 1985–1994 (2017).
20. Wu, C. W. *et al.* Detection of miR-92a and miR-21 in stool samples as potential screening biomarkers for colorectal cancer and polyps. *Gut* **61**, 739–45 (2012).
21. Kuriyama, S. *et al.* Fecal MicroRNA Assays as a Marker for Colorectal Cancer Screening. *Gastroenterology* **142**, S-770 (2012).
22. Kanaoka, S. *et al.* Potential Usefulness of Fecal Immunochemical Test Plus Fecal MicroRNA Assay As a Marker for Colorectal Cancer Screening. *Gastroenterology* **144**, S-599-S-600 (2013).
23. Koga, Y. *et al.* Fecal miR-106a is a useful marker for colorectal cancer patients with false-negative results in immunochemical fecal occult blood test. *Cancer Epidemiol. Biomarkers Prev.* **22**, 1844–1852 (2013).
24. Zhao, H. J. *et al.* MiR-194 deregulation contributes to colorectal carcinogenesis via targeting AKT2 pathway. *Theranostics* **4**, 1193–1208 (2014).
25. Wu, C. W. *et al.* Identification of microRNA-135b in stool as a potential noninvasive biomarker for colorectal cancer and adenoma. *Clin. Cancer Res.* **20**, 2994–3002 (2014).
26. Yau, T. O. *et al.* microRNA-221 and microRNA-18a identification in stool as potential biomarkers for the non-invasive diagnosis of colorectal carcinoma. *Br. J. Cancer* **111**, 1765–1771 (2014).
27. Phua, L. C. *et al.* Global fecal microRNA profiling in the identification of biomarkers for colorectal cancer screening among Asians. *Oncol. Rep.* **32**, 97–104 (2014).
28. Hunter, M. P. *et al.* Detection of microRNA expression in human peripheral blood microvesicles. *PLoS One* **3**, e3694 (2008).
29. Mitchell, P. S. *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. USA* **105**, 10513–8 (2008).
30. Whiting, P. F. *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* **155**, 529–36 (2011).
31. Benz, F. *et al.* U6 is unsuitable for normalization of serum miRNA levels in patients with sepsis or liver fibrosis. *Exp. Mol. Med.* **45**, e42 (2013).
32. Xiang, M. *et al.* U6 is not a suitable endogenous control for the quantification of circulating microRNAs. *Biochem. Biophys. Res. Commun.* **454**, 210–214 (2014).
33. Gao, Y. *et al.* Down-regulation of miR-24-3p in colorectal cancer is associated with malignant behavior. *Med. Oncol.* **32**, 362 (2015).
34. Mishra, P. J. *et al.* MiR-24 tumor suppressor activity is regulated independent of p53 and through a target site polymorphism. *PLoS One* **4**, e8445 (2009).
35. Zanutto, S. *et al.* Circulating miR-378 in plasma: a reliable, haemolysis-independent biomarker for colorectal cancer. *Br. J. Cancer* **110**, 1001–1007 (2014).
36. Dejima, H., Iinuma, H., Kanaoka, R., Matsutani, N. & Kawamura, M. Exosomal microRNA in plasma as a non-invasive biomarker for the recurrence of non-small cell lung cancer. *Oncol. Lett.* **13**, 1256–1263 (2017).
37. Xu, X., Wang, X., Fu, B., Meng, L. & Lang, B. Differentially expressed genes and microRNAs in bladder carcinoma cell line 5637 and T24 detected by RNA sequencing. *Int. J. Clin. Exp. Pathol.* **8**, 12678–87 (2015).
38. Lopez, J. P. *et al.* Co-Variation of Peripheral Levels of miR-1202 and Brain Activity and Connectivity During Antidepressant Treatment. *Neuropsychopharmacology* **42**, 2043–2051 (2017).
39. Lopez, J. P. *et al.* miR-1202 is a primate-specific and brain-enriched microRNA involved in major depression and antidepressant treatment. *Nat. Med.* **20**, 764–8 (2014).
40. Fiori, L. M. *et al.* Investigation of miR-1202, miR-135a, and miR-16 in Major Depressive Disorder and Antidepressant Response. *Int. J. Neuropsychopharmacol.* <https://doi.org/10.1093/ijnp/pyx034> (2017).
41. Hamam, R. *et al.* microRNA expression profiling on individual breast cancer patients identifies novel panel of circulating microRNA for early detection. *Sci. Rep.* **6**, 25997 (2016).
42. Tokuhisa, M. *et al.* Exosomal miRNAs from Peritoneum Lavage Fluid as Potential Prognostic Biomarkers of Peritoneal Metastasis in Gastric Cancer. *PLoS One* **10**, e0130472 (2015).
43. Munari, E. *et al.* Clear cell papillary renal cell carcinoma: micro-RNA expression profiling and comparison with clear cell renal cell carcinoma and papillary renal cell carcinoma. *Hum. Pathol.* **45**, 1130–8 (2014).
44. Toiyama, Y., Okugawa, Y., Fleshman, J., Richard Boland, C. & Goel, A. MicroRNAs as potential liquid biopsy biomarkers in colorectal cancer: A systematic review. *Biochim. Biophys. Acta. Rev. cancer* **1870**, 274–282 (2018).
45. Oikonomopoulos, A., Polytarchou, C., Joshi, S., Hommes, D. W. & Iliopoulos, D. Identification of Circulating MicroRNA Signatures in Crohn's Disease Using the Nanostring nCounter Technology. *Inflamm. Bowel Dis.* **22**, 2063–9 (2016).
46. Polytarchou, C. *et al.* Assessment of Circulating MicroRNAs for the Diagnosis and Disease Activity Evaluation in Patients with Ulcerative Colitis by Using the Nanostring Technology. *Inflamm. Bowel Dis.* **21**, 2533–9 (2015).
47. Schönauen, K. *et al.* Circulating and Fecal microRNAs as Biomarkers for Inflammatory Bowel Diseases. *Inflamm. Bowel Dis.* **24**, 1547–1557 (2018).
48. Polytarchou, C. *et al.* MicroRNA214 Is Associated With Progression of Ulcerative Colitis, and Inhibition Reduces Development of Colitis and Colitis-Associated Cancer in Mice. *Gastroenterology* **149**, 981–92.e11 (2015).
49. Wu, Y. *et al.* MicroRNA-21 (Mir-21) Promotes Cell Growth and Invasion by Repressing Tumor Suppressor PTEN in Colorectal Cancer. *Cell. Physiol. Biochem.* **43**, 945–958 (2017).
50. Feng, Y.-H. *et al.* MicroRNA-21-mediated regulation of Sprouty2 protein expression enhances the cytotoxic effect of 5-fluorouracil and metformin in colon cancer cells. *Int. J. Mol. Med.* **29**, 920–6 (2012).
51. Ahmadi, S., Sharifi, M. & Salehi, R. Locked nucleic acid inhibits miR-92a-3p in human colorectal cancer, induces apoptosis and inhibits cell proliferation. *Cancer Gene Ther.* **23**, 199–205 (2016).
52. Xiong, Y. *et al.* Correlation of over-expressions of miR-21 and Notch-1 in human colorectal cancer with clinical stages. *Life Sci.* **106**, 19–24 (2014).
53. Peacock, O. *et al.* Inflammation and MiR-21 pathways functionally interact to downregulate PDCD4 in colorectal cancer. *PLoS One* **9**, e110267 (2014).
54. Saxena, A., Shoeb, M., Ramana, K. V. & Srivastava, S. K. Aldose reductase inhibition suppresses colon cancer cell viability by modulating microRNA-21 mediated programmed cell death 4 (PDCD4) expression. *Eur. J. Cancer* **49**, 3311–9 (2013).

55. Liu, M. *et al.* miR-21 targets the tumor suppressor RhoB and regulates proliferation, invasion and apoptosis in colorectal cancer cells. *FEBS Lett.* **585**, 2998–3005 (2011).
56. Li, G., Wang, C., Wang, Y., Xu, B. & Zhang, W. LINC00312 represses proliferation and metastasis of colorectal cancer cells by regulation of miR-21. *J. Cell. Mol. Med.* **22**, 5565–5572 (2018).
57. Zhang, G.-J. *et al.* MiR-92a promotes stem cell-like properties by activating Wnt/ $\beta$ -catenin signaling in colorectal cancer. *Oncotarget* **8**, 101760–101770 (2017).
58. Ke, T.-W., Wei, P.-L., Yeh, K.-T., Chen, W. T.-L. & Cheng, Y.-W. MiR-92a Promotes Cell Metastasis of Colorectal Cancer Through PTEN-Mediated PI3K/AKT Pathway. *Ann. Surg. Oncol.* **22**, 2649–55 (2015).
59. Lv, H. *et al.* MicroRNA-92a Promotes Colorectal Cancer Cell Growth and Migration by Inhibiting KLF4. *Oncol. Res.* **23**, 283–290 (2016).
60. Chen, E. *et al.* MiR-92a promotes tumorigenesis of colorectal cancer, a transcriptomic and functional based study. *Biomed. Pharmacother.* **106**, 1370–1377 (2018).
61. Strul, H. Fecal occult blood test for colorectal cancer screening. *Ann. Oncol.* **13**, 51–56 (2002).
62. Hirai, H. W. *et al.* Systematic review with meta-analysis: faecal occult blood tests show lower colorectal cancer detection rates in the proximal colon in colonoscopy-verified diagnostic studies. *Aliment. Pharmacol. Ther.* **43**, 755–764 (2016).
63. Leeflang, M. M. G., Deeks, J. J., Takwoingi, Y. & Macaskill, P. Cochrane diagnostic test accuracy reviews. *Syst. Rev.* **2**, 82 (2013).
64. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **6**, e1000097 (2009).
65. Reitsma, J. B. *et al.* Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J. Clin. Epidemiol.* **58**, 982–990 (2005).
66. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–88 (1986).
67. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials revisited. *Contemp. Clin. Trials* **45**, 139–145 (2015).
68. Jones, C. M. & Athanasiou, T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. *Ann. Thorac. Surg.* **79**, 16–20 (2005).
69. Grimes, D. A. & Schulz, K. F. Refining clinical diagnosis with likelihood ratios. *Lancet* **365**, 1500–1505 (2005).

## Acknowledgements

This study was partly supported by a grant from the Litwin Initiative at the Crohn's and Colitis Foundation and NTU QR funds.

## Author Contributions

T.O.Y. and C.M.T. contributed to the design of the study; E.K.H. designed and performed the literature search. T.O.Y. performed the data analysis and wrote the manuscript; B.D. reviewed the data analysis, C.M.T., E.K.H., B.D. and C.P. wrote and reviewed the manuscript. All authors have read and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-45570-9>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

# Metagenomic Analysis of Faecal Microbiome as a Tool towards Targeted Non-Invasive Biomarkers for Colorectal Cancer

Jun Yu\*, Qiang Feng\*, Sunny Hei Wong\*, Dongya Zhang\*, Qiao yi Liang\*, Youwen Qin, Longqing Tang, Hui Zhao, Jan Stenvang, Yanli Li, Xiaokai Wang, Xiaoqiang Xu, Ning Chen, William Ka Kei Wu, Jumana Al-Aama, Hans Jørgen Nielsen, Pia Kiilerich, Benjamin Anderschou Holbech Jensen, **Tung On Yau**, Zhou Lan, Huijue Jia, Junhua Li, Liang Xiao, Thomas Yuen Tung Lam, Siew Chien Ng, Alfred Sze-Lok Cheng, Vincent Wai-Sun Wong, Francis Ka Leung Chan, Xun Xu, Huanming Yang, Lise Madsen, Christian Datz, Herbert Tilg, Jian Wang, Nils Brünner, Karsten Kristiansen, Manimozhiyan Arumugam<sup>#</sup>, Joseph Jao-Yiu Sung<sup>#</sup>, Jun Wang<sup>#</sup>

*Gut.* **2017**;66(1):70–78.

Journal URL: [gut.bmj.com/content/66/1/70](http://gut.bmj.com/content/66/1/70)

DOI: [10.1136/gutjnl-2015-309800](https://doi.org/10.1136/gutjnl-2015-309800)

PMID: [26408641](https://pubmed.ncbi.nlm.nih.gov/26408641/)



3<sup>rd</sup> November, 2015

To Whom It May Concern,

**Statement of Joint Authorship**

**Title of publication:** Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer

**Journal:** Gut

**Publication date:** 25<sup>th</sup> September 2015 (Electronic publication ahead of print)

**PubMed ID:** 26408641

**Authors:** Jun Yu\*, Qiang Feng\*, Sunny Hei Wong\*, Dongya Zhang\*, Qiaoyi Liang\*, Youwen Qin, Longqing Tang, Hui Zhao, Jan Stenvang, Yanli Li, Xiaokai Wang, Xiaoqiang Xu, Ning Chen, William Ka Kei Wu, Jumana Al-Aama, Hans Jørgen Nielsen, Pia Kiilerich, Benjamin Anderschou Holbech Jensen, Tung On Yau, Zhou Lan, Huijue Jia, Junhua Li, Liang Xiao, Thomas Yuen Tung Lam, Siew Chien Ng, Alfred Sze-Lok Cheng, Vincent Wai-Sun Wong, Francis Ka Leung Chan, Xun Xu, Huanming Yang, Lise Madsen, Christian Datz, Herbert Tilg, Jian Wang, Nils Brüner, Karsten Kristiansen, Manimozhiyan Arumugam, Joseph Jao-Yiu Sung, Jun Wang (\* co-first authors)

I hereby confirm that Mr. Tung On YAU is a co-author in the above publication, He was a significant contributor to this project. His work included liaising with nurses and clinicians for the acquisition, processing, and storage of faecal samples, as well as the associated patient information. He then performed experimental validation of sequencing results using quantitative Real-Time PCR. He also performed clinical data analyses.

Yours sincerely,

Qiaoyi Liang, PhD

Research Assistant Professor

Institute of Digestive Disease

Prince of Wales Hospital

The Chinese University of Hong Kong, Hong Kong

Tel: 852-3763 6103; Fax: 852-2144 5330

Email: jessieQY@cuhk.edu.hk

Jun Yu, MD, PhD

Professor

Institute of Digestive Disease

Prince of Wales Hospital

The Chinese University of Hong Kong, Hong Kong

Tel: 852-3763 6099; Fax: 852-2144 5330

Email: junyu@cuhk.edu.hk

# Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer

Jun Yu,<sup>1†</sup> Qiang Feng,<sup>2,3†</sup> Sunny Hei Wong,<sup>1†</sup> Dongya Zhang,<sup>2†</sup> Qiao yi Liang,<sup>1†</sup> Youwen Qin,<sup>2</sup> Longqing Tang,<sup>2</sup> Hui Zhao,<sup>2</sup> Jan Stenvang,<sup>4</sup> Yanli Li,<sup>2</sup> Xiaokai Wang,<sup>2</sup> Xiaoqiang Xu,<sup>2</sup> Ning Chen,<sup>2</sup> William Ka Kei Wu,<sup>1</sup> Jumana Al-Aama,<sup>2,5</sup> Hans Jørgen Nielsen,<sup>6</sup> Pia Kiilerich,<sup>3</sup> Benjamin Anderschou Holbech Jensen,<sup>3</sup> Tung On Yau,<sup>1</sup> Zhou Lan,<sup>2</sup> Huijue Jia,<sup>2</sup> Junhua Li,<sup>2</sup> Liang Xiao,<sup>2</sup> Thomas Yuen Tung Lam,<sup>1</sup> Siew Chien Ng,<sup>1</sup> Alfred Sze-Lok Cheng,<sup>1</sup> Vincent Wai-Sun Wong,<sup>1</sup> Francis Ka Leung Chan,<sup>1</sup> Xun Xu,<sup>2</sup> Huanming Yang,<sup>2</sup> Lise Madsen,<sup>2,3,7</sup> Christian Datz,<sup>8</sup> Herbert Tilg,<sup>9</sup> Jian Wang,<sup>2</sup> Nils Brünner,<sup>2,4</sup> Karsten Kristiansen,<sup>2,3</sup> Manimozhayan Arumugam,<sup>2,10</sup> Joseph Jao-Yiu Sung,<sup>1</sup> Jun Wang<sup>2,3,5,11</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2015-309800>).

For numbered affiliations see end of article.

## Correspondence to

Professor Jun Wang, Beijing Genomics Institute at Shenzhen, Shenzhen 518000, China; wangj@genomics.org.cn and Professor Joseph JY Sung, The Chinese University of Hong Kong, Hong Kong; jjysung@cuhk.edu.hk and Dr Manimozhayan Arumugam, University of Copenhagen, 2200 Copenhagen, Denmark; arumugam@sund.ku.dk

JuY, QF, SHW, DZ, QL contributed equally.

Received 21 April 2015  
Revised 26 August 2015  
Accepted 1 September 2015  
Published Online First  
25 September 2015



CrossMark

**To cite:** Yu J, Feng Q, Wong SH, et al. *Gut* 2017;**66**:70–78.

## ABSTRACT

**Objective** To evaluate the potential for diagnosing colorectal cancer (CRC) from faecal metagenomes.

**Design** We performed metagenome-wide association studies on faecal samples from 74 patients with CRC and 54 controls from China, and validated the results in 16 patients and 24 controls from Denmark. We further validated the biomarkers in two published cohorts from France and Austria. Finally, we employed targeted quantitative PCR (qPCR) assays to evaluate diagnostic potential of selected biomarkers in an independent Chinese cohort of 47 patients and 109 controls.

**Results** Besides confirming known associations of *Fusobacterium nucleatum* and *Peptostreptococcus stomatis* with CRC, we found significant associations with several species, including *Parvimonas micra* and *Solobacterium moorei*. We identified 20 microbial gene markers that differentiated CRC and control microbiomes, and validated 4 markers in the Danish cohort. In the French and Austrian cohorts, these four genes distinguished CRC metagenomes from controls with areas under the receiver-operating curve (AUC) of 0.72 and 0.77, respectively. qPCR measurements of two of these genes accurately classified patients with CRC in the independent Chinese cohort with AUC=0.84 and OR of 23. These genes were enriched in early-stage (I–II) patient microbiomes, highlighting the potential for using faecal metagenomic biomarkers for early diagnosis of CRC.

**Conclusions** We present the first metagenomic profiling study of CRC faecal microbiomes to discover and validate microbial biomarkers in ethnically different cohorts, and to independently validate selected biomarkers using an affordable clinically relevant technology. Our study thus takes a step further towards affordable non-invasive early diagnostic biomarkers for CRC from faecal samples.

## INTRODUCTION

Colorectal cancer (CRC), the third most common cancer in the world affecting >1.36 million people every year,<sup>1</sup> arises due to complex interactions

## Significance of this study

### What is already known on this subject?

- Changes in the gut microbial composition are associated with colorectal cancer (CRC), but causality is yet to be established.
- *Fusobacterium nucleatum* potentiates intestinal tumorigenesis through recruitment of infiltrating immune cells and via activation of  $\beta$ -catenin signalling.
- Faecal microbiota holds promise for early non-invasive diagnosis of CRC.
- However, a simple and affordable targeted approach to diagnosing CRC from faecal samples is still lacking.

### What are the new findings?

- Discovery of significant enrichment of novel species, including *Parvimonas micra* and *Solobacterium moorei*, and a strong co-occurrence network between them in the faecal microbiomes of patients with CRC.
- Identification of 20 gene markers that significantly differentiate CRC-associated and control microbiomes in a Chinese cohort, and trans-continental validation of four of them in a Danish cohort.
- Further validation of the four gene markers in published cohorts from the French and Austrian cohorts with areas under the receiver-operating curve (AUC) of 0.72 and 0.77.
- Quantitative PCR abundance of two gene markers (butyryl-CoA dehydrogenase from *F. nucleatum*, and RNA polymerase subunit  $\beta$ , *rpoB*, from *P. micra*) clearly separates CRC microbiomes from controls in an independent Chinese cohort consisting of 47 cases and 109 healthy controls, with AUC=0.84 and odds ratio of 23.

## Significance of this study

**How might it impact on clinical practice in the foreseeable future?**

- ▶ The four microbial gene markers shared between the Chinese, Danish, Austrian and French cohorts suggest that even though different populations may have different gut microbial community structures, signatures of CRC-associated microbial dysbiosis could have universal features.
- ▶ Our study takes a step further towards affordable early diagnosis of CRC by targeted analysis of metagenomic biomarkers in faecal samples.

between genetic, lifestyle and environmental factors. Despite massive efforts in whole-genome sequencing and genome-wide association studies, genetic factors only explain a small proportion of disease variance<sup>2</sup>—heritability may account for up to 35% all CRCs,<sup>3</sup> but only about 5% of cancers occur in the setting of a known genetic predisposition syndrome.<sup>4</sup> These findings support lifestyle and environment as additional major disease determinants.

Emerging evidence indicates that microbial dysbiosis in the human gut may be an important environmental factor in CRC. Early evidence for gut microbial contribution to CRC pathogenesis came from *Apc<sup>min/+</sup>* mice, a genetic mouse model of CRC, where mice housed in germ-free conditions showed a reduction of tumour formation in the intestine compared with mice housed in specific pathogen-free conditions.<sup>5</sup> Further studies have suggested that several bacteria, including *Bacteroides fragilis* and a strain of *Escherichia coli*, may promote colorectal carcinogenesis.<sup>6–11</sup> In humans, bacterial culture-based studies have reported associations between CRC and clinical infections by specific bacteria such as *Streptococcus bovis*<sup>12</sup> and *Clostridium septicum*.<sup>13</sup> Additionally, culture-free 16S ribosomal RNA sequencing studies have associated faecal microbial composition with CRC.<sup>14–16</sup> Independent studies have identified *Fusobacterium nucleatum* to be more abundant in human CRC tissues,<sup>17, 18</sup> and follow-up studies showed that *F. nucleatum* potentiates intestinal tumorigenesis through recruitment of infiltrating immune cells<sup>19</sup> and by modulating  $\beta$ -catenin signalling.<sup>20</sup> Two recent studies investigated gut microbial dysbiosis in patients with CRC<sup>21, 22</sup> and reported diagnostic potential using metagenomic sequencing. These promising results are still far from directly translating to diagnostic tests for CRC, as a simple and affordable targeted approach to diagnosing CRC from faecal samples is still lacking.

Here we present the first study that (i) uses deep metagenomic profiling of CRC faecal microbiomes to discover and validate microbial gene biomarkers in ethnically different cohorts, and (ii) independently validates them using an affordable technology that can translate to clinical practice.

**MATERIALS AND METHODS****Sample collection and DNA preparation**

Cohorts C1 and C2 were from Hong Kong, China. C1 (see online supplementary table S1) comprised 128 individuals: 74 patients with CRC (15 stage I, 21 stage II, 34 stage III and 4 stage IV; median age 67 years; 26 were females) and 54 controls (median age 62 years; 21 were females). C2 (see online

supplementary table S16) comprised 156 individuals: 47 patients with CRC (4 stage I, 24 stage II, 15 stage III and 4 stage IV; median age 69 years; 22 were females) and 109 controls (median age 58 years; 69 were females). Cohort D from Copenhagen, Denmark (see online supplementary table S18), comprised 40 individuals: patients with CRC (n=16; 1 stage I, 9 stage II, 5 stage III and 1 stage IV; median age 67.5 years; 6 were females) and controls (n=24; median age 65.5 years; 17 were females). Cancer staging in all three cohorts was performed using the tumour, node, metastasis staging system<sup>23</sup> maintained by the American Joint Committee on Cancer and the International Union for Cancer Control. Stool samples were collected by individuals at home, followed by immediate freezing at  $-20^{\circ}\text{C}$ . DNA from Chinese samples was extracted using Qiagen QIAamp DNA Stool Mini Kit (Qiagen) according to manufacturer's instructions. DNA from Danish samples was extracted using previously published method.<sup>24</sup> For comprehensive description of sample collection and DNA extraction as well as ethical committee approval numbers, see online supplementary methods.

**Metagenomic sequencing and annotation**

Metagenomic sequencing using Illumina HiSeq 2000 platform, generating gene profiles using gene catalogue, constructing metagenomic linkage groups (MLGs), generating Kyoto Encyclopedia of Genes and Genomes (KEGG) ortholog, module and pathway profiles, were all done using previously published methods.<sup>25</sup> Species-level molecular operational taxonomic units (mOTUs) were obtained using mOTU profiling software.<sup>26</sup> Reads were mapped to the Integrated Microbial Genome (IMG) reference database<sup>27</sup> (v400) to generate IMG species and IMG genus profiles. Genes of MLGs were mapped to the IMG database, and MLGs were annotated to an IMG genome when >50% of genes were mapped. MLG species were constructed by grouping MLGs using this annotation. For comprehensive description of these procedures, see online supplementary methods.

**Data analysis**

Permutational multivariate analysis of variance (PERMANOVA) was used to assess effects of different phenotypes on gene profiles. Enrichments of genes, KEGG features, mOTUs, IMG species and MLG species were calculated using Wilcoxon rank-sum tests. When appropriate, we adjusted for confounding effects of sample collection before/after colonoscopy: Wilcoxon rank-sum tests were performed using 'colonoscopy before/after sampling' as a stratifying factor using COIN package in R, and ORs were estimated using Mantel–Haenszel test after stratifying by 'colonoscopy before/after sampling'. We controlled for multiple testing with Benjamini–Hochberg false discovery rate (FDR). Minimum-redundancy maximum-relevancy (mRMR) feature selection method<sup>28</sup> was used to select optimal gene markers, which were then used in constructing a CRC index. Co-occurrence networks were constructed using Spearman's correlation coefficient (>0.5 or <−0.5) and visualised in Cytoscape V3.0.2. Metagenomic sequences from French (F) and Austrian (A) cohorts were downloaded from NCBI Short Read Archive using study identifiers ERP005534 and ERP008729, respectively. For comprehensive description of biodiversity analysis, rarefaction analysis, identification of CRC-associated genes/species, estimation of FDR, mRMR feature selection framework, definition and validation of CRC index, and receiver operator characteristic (ROC) analysis, see online supplementary methods.

### Validation of gene markers by qPCR

Abundances of selected gene markers were estimated in stool samples using TaqMan probe-based quantitative PCR (qPCR). Primer and probe sequences were designed manually and then tested using Primer Express V3.0 (Applied Biosystems, Foster City, California, USA) for determination of Tm, guanine-cytosine (GC) content and possible secondary structures. Each probe carried a 5' reporter dye 6-carboxy fluorescein or 4,7,2'-trichloro-7'-phenyl-6-carboxyfluorescein and a 3' quencher dye 6-carboxytetramethylrhodamine. Primers and hydrolysis probes were synthesised by Invitrogen (Carlsbad, California, USA). Nucleotide sequences of primers and probes are listed in online supplementary table S27. qPCR was performed on an ABI7500 Real-Time PCR System using TaqMan Universal PCR Master Mix reagent (Applied Biosystems). Universal 16S rDNA was used as internal control and abundance of gene markers were expressed as relative levels to 16S rDNA.

## RESULTS

### Dysbiosis in CRC gut microbiome

We recruited 128 individuals (74 patients with CRC and 54 control subjects) from China (cohort C1; see online supplementary table S1), performed metagenomic sequencing on their stool samples and generated 751 million metagenomic reads (5.86 million reads per individual on average; see online supplementary table S2) using Illumina HiSeq 2000 platform. Among the recorded metabolic parameters, elevated fasting blood glucose and reduced high-density lipoproteins showed significant associations with CRC status (Wilcoxon rank-sum test,  $q=0.0014$  for both) agreeing with previous findings reporting them as risk factors.<sup>29,30</sup> We also observed that a significantly higher number of CRC patient samples were collected after colonoscopy than before (Fisher's exact test,  $q=0.0165$ ; see online supplementary table S1). We adjusted for this as a confounding factor in subsequent analyses when appropriate (see section 'Materials and methods'). Rarefaction analysis using a previously published gut microbial gene catalogue consisting of 4 267 985 genes<sup>25</sup> showed a curve reaching plateau, suggesting that this catalogue covers most prevalent microbial genes present in cohort C1 (see online supplementary figure S1A). Therefore, we based subsequent analyses on mapping the metagenomic reads to this catalogue. CRC patient microbiomes exhibited reduced gene richness (see online supplementary figure 1A, B; Wilcoxon rank-sum test,  $p<0.01$ ) and gene alpha diversity (Wilcoxon rank-sum tests on Shannon and Simpson indices:  $p=0.075$  and  $0.028$ , respectively; see online supplementary figure S1C,D and table S3). However, these differences exhibited  $p>0.5$  after correcting for colonoscopy.

To ensure robust comparison of gene content among 128 metagenomes from cohort C1, we created a set of 2 110 489 genes that were present in at least 6 subjects and generated 128 gene abundance profiles using these 2.1 million genes. When we performed multivariate analysis using PERMANOVA on 17 different covariates, only CRC status and CRC stage were significantly associated with these gene profiles ( $q<0.06$ , all other factors:  $q>0.27$ ; see online supplementary table S4). Thus, the data suggest an altered gene composition in CRC patient microbiomes that cannot be explained by other recorded factors. When we performed a principal component analysis (PCA) based on gene profiles, the first and fifth principal components, which explained 6.6% and 3.2% of total variance, respectively, were associated with CRC status (Wilcoxon rank-sum test, PC1:  $p=0.029$ ; PC5:  $p=1\times 10^{-6}$ ; see online supplementary figure S2

and table S5). Together, these results suggest a state of dysbiosis of the gut microbiome in patients with CRC.

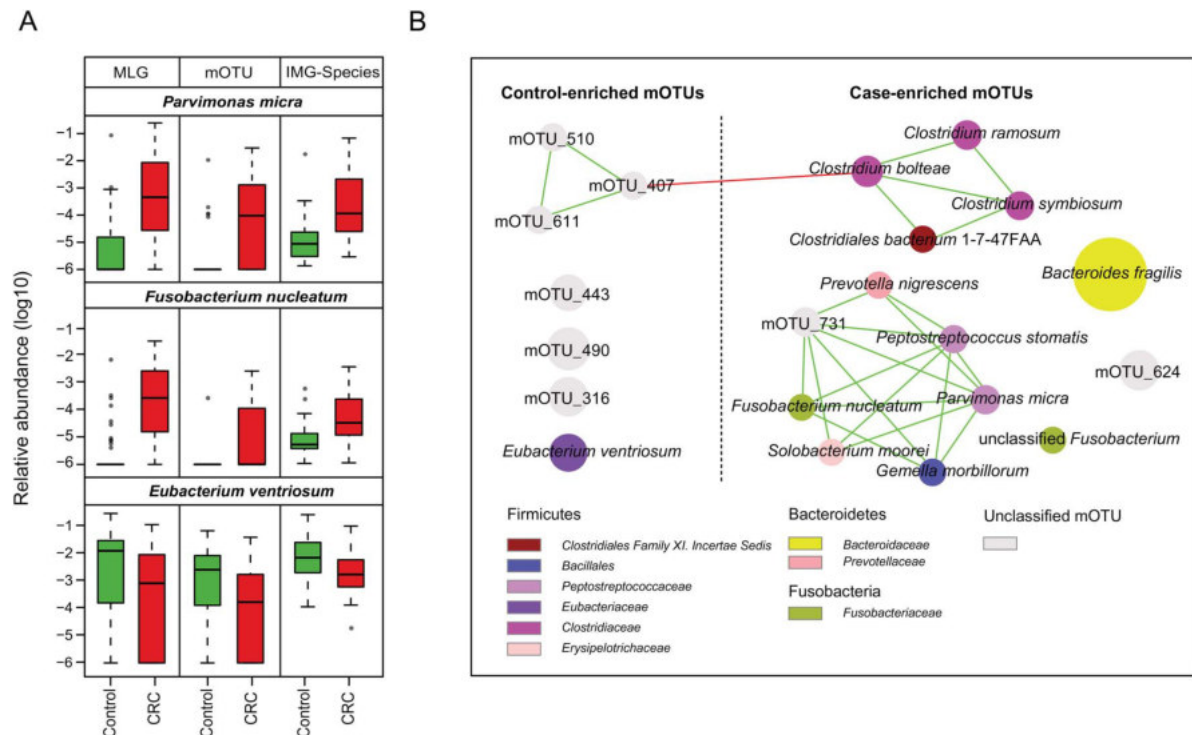
### Gut microbial genes associated with CRC

We performed a metagenome-wide association study (MGWAS) to identify genes contributing to the altered gene composition in CRC. From 2.1 million genes, we identified 140 455 genes that were associated with disease status (Wilcoxon rank-sum test  $p<0.01$  and FDR 11.03%; see online supplementary figure S3). Interestingly, CRC-enriched genes occurred less frequently and at lower abundance compared with control-enriched genes (see online supplementary figure S4), suggesting that microbial dysbiosis associated with CRC may not involve dominant species. Such patterns of frequency and occurrence have been observed in two earlier metagenomic case-control studies on type 2 diabetes<sup>25</sup> in Chinese individuals and CRC in Austrian individuals,<sup>31</sup> suggesting that this may be a common trend in disease-associated gut microbial dysbiosis.

We annotated the 140 455 genes using KEGG<sup>32</sup> functional database (V.59) to investigate whether certain microbial functions were associated with CRC. None of the KEGG pathways passed our stringent criteria (Wilcoxon rank-sum test,  $q<0.05$ ; see online supplementary table S6), suggesting that bacterial metabolic pathways present in KEGG database may not be involved in CRC pathogenesis. However, two KEGG modules were enriched in CRC microbiomes: leucine degradation ( $q=0.0148$ ) and guanine nucleotide biosynthesis ( $q=0.0241$ ; see online supplementary table S6). Leucine stimulates both protein synthesis and degradation,<sup>33,34</sup> suggesting possible links between leucine metabolism and cancer. At the gene level, several KEGG orthologous groups showed significant associations with disease status (Wilcoxon rank-sum test,  $q<0.05$ ; see online supplementary table S7).

### Taxonomic alterations in CRC gut microbiomes

We examined taxonomic differences between CRC-associated and control microbiomes to identify microbial taxa contributing to the dysbiosis. For this, we used species profiles derived from three different methods—IMG species, species-level mOTUs and MLG species (see section 'Materials and methods')—as supporting evidence from multiple methods would strengthen an association. Our analysis identified 28 IMG species, 21 mOTUs and 85 MLG species that were significantly associated with CRC status after adjusting for colonoscopy as a confounding factor (Wilcoxon rank-sum test,  $q<0.05$ ; see online supplementary table S8). *Eubacterium ventriosum* was consistently enriched in control microbiomes across all three methods (IMG:  $q=0.002$ ; mOTU:  $q=0.0049$ ; MLG:  $q=3.33\times 10^{-4}$ ). On the other hand, *Parvimonas micra* ( $q<7.73\times 10^{-6}$ ), *Solobacterium moorei* ( $q<0.011$ ) and *F. nucleatum* ( $q<0.00279$ ) were consistently enriched in CRC patient microbiomes across all three methods (figure 1A and online supplementary figure S5), while *Peptostreptococcus stomatis* ( $q<7.73\times 10^{-6}$ ) was enriched according to two methods. PERMANOVA analysis showed that only CRC status ( $p\leq 0.013$  from all three methods) and colonoscopy ( $p=0.079$  from two methods) explained the quantitative variation in the three CRC-enriched species. All other non-CRC-specific factors could not explain the variation with statistical significance ( $p>0.18$ ; see online supplementary table S9). *P. stomatis* has recently been shown to significantly associate with CRC,<sup>22</sup> and *S. moorei* has previously been associated with bacteraemia.<sup>35</sup> However, a highly significant enrichment of *P. micra*—an obligate anaerobic bacterium that can cause oral infections like *F. nucleatum*<sup>36</sup>—in CRC-associated microbiomes is a novel finding.

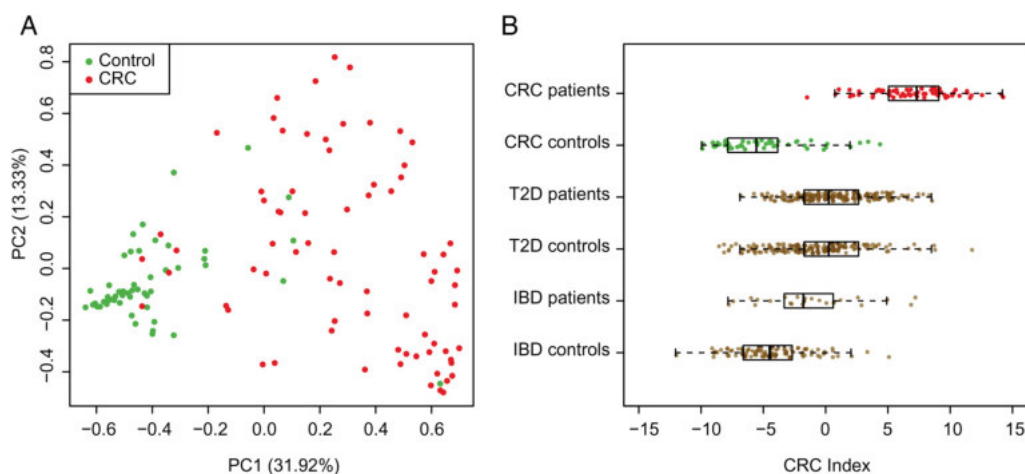


**Figure 1** Species involved in gut microbial dysbiosis associated with colorectal cancer (CRC). (A) Differential relative abundance of two CRC-enriched and one control-enriched microbial species consistently identified using three different methods: metagenomic linkage group (MLG), molecular operational taxonomic unit (mOTU) and Integrated Microbial Genome (IMG) database. (B) A co-occurrence network deduced from relative abundance of 21 mOTUs significantly associated with CRC. Species are rearranged in two sides based on their enrichment in CRC or control microbiomes. Spearman correlation coefficient values below  $-0.5$  (negative correlation) are indicated as red edges, and coefficient values above  $0.5$  (positive correlation) are indicated as green edges. Node size shows the average relative abundance for each species, and node colour shows their taxonomic annotation.

Species co-occurrence networks derived from pairwise correlations of species abundances showed a strong positive association between three oral pathogens: *P. micra*, *F. nucleatum* and *S. moorei* (figure 1B and online supplementary figure S6). Previous reports suggest that *P. micra* commonly occurs together with *F. nucleatum* in infected root canals, where they could account for up to 90% of the endodontic microbiome.<sup>36</sup> Given

this, our results could suggest cooperation between these two species in CRC-associated gut environment.

Although several bacterial genera corresponding to the CRC-associated species identified earlier (including *Parvimonas*, *Fusobacterium*, *Solobacterium* and *Peptostreptococcus*) showed significant associations with CRC status (see online supplementary table S10), we observed some exceptions as well. While we



**Figure 2** Discovering gut microbial gene markers associated with colorectal cancer (CRC). (A) Principal component analysis based on abundances of 20 gene markers separates CRC cases and control individuals in cohort C1. First and second principal components associate with CRC status (PC1 and PC2 explain 31.9% and 13.3% of variance, respectively). Compare this with online supplementary figure S2 based on 2.1 million genes, where no separation can be observed. (B) CRC index computed using a simple unweighed linear combination of log-abundance of 20 gene markers for patients with CRC (red) and control individuals (green) from this study, shown together with patients and control individuals (brown) from earlier studies on type 2 diabetes<sup>25</sup> and IBD.<sup>38</sup> CRC indices for CRC patient microbiomes are significantly different from the rest ( $p < 0.001$ ), suggesting that the 20 gene markers are CRC-specific. The box depicts the IQRs between the first and third quartiles, and the line inside denotes the median.



identified a significant over-representation of *B. fragilis* in patients with CRC (mOTU:  $q=0.0158$ ; MLG:  $q=3.02 \times 10^{-4}$ ; see online supplementary table S8), there was no association with *Bacteroides* genus. At the phylum level, only Fusobacteria and Basidiomycota were significantly enriched in CRC-associated microbiomes ( $q<0.0002$ ; see online supplementary table S11).

In order to evaluate the predictive power of these taxonomic associations, we used random forest ensemble learning method<sup>37</sup> to identify 17 IMG species, 7 species-level mOTUs and 27 MLG species that were highly predictive of CRC status (see online supplementary table S12), with predictive power of 0.86, 0.89 and 0.96 in ROC analysis, respectively (see online supplementary figure S7). *P. micra* was identified as a key species from all three methods, while *F. nucleatum*, *P. stomatis* and *S. moorei* were identified from two out of three methods, providing further statistical support for their association with CRC status.

### CRC biomarker discovery

We used the mRMR feature selection method<sup>28</sup> to identify potential CRC biomarkers from the 140 455 genes identified by MGWAS. First, to eliminate confounding effects of colonoscopy, we performed blocked independent Wilcoxon rank-sum tests on these genes with colonoscopy as a stratifying factor. This resulted in 102 514 genes at a significance level of  $p<0.01$  (FDR  $\leq 13\%$ ) and 24 960 genes at a significance level of  $p<0.001$  (FDR  $\leq 5.23\%$ ). Then, from the latter, we identified groups of genes that were highly correlated with each other (Kendall's  $\tau > 0.9$ ) and chose the longest gene in each group to generate a statistically non-redundant set of 11 128 significant genes. Finally, we used mRMR method and identified an optimal set of 20 genes that were strongly associated with CRC status (see online supplementary figure S8 and table S13). PCA using these 20 genes showed good separation of patients with CRC from controls (figure 2A). PERMANOVA analysis showed that only CRC status, stage and fasting blood glucose explained the variation in the 20 marker gene abundances with statistical significance ( $p \leq 0.01$ ; see online supplementary table S14). We computed a simple CRC index based on unweighted log relative abundance of these 20 markers, which clearly separated CRC patient microbiomes from control microbiomes, as well as from 490 faecal microbiomes from two previous studies on type 2 diabetes in Chinese individuals<sup>25</sup> and IBD in European individuals<sup>38</sup> (figure 2B; median CRC index for patients and controls in our study were 7.31 and  $-5.56$ , respectively; Wilcoxon rank-sum test,  $q < 6 \times 10^{-11}$  for all five comparisons; see online supplementary table S15).

### Evaluating CRC biomarkers using targeted qPCR

Translating our gene markers into diagnostic biomarkers would require reliable measurement by simple, affordable and targeted methods such as qPCR. To verify whether gene abundances measured by metagenomics sequencing and qPCR are comparable, we randomly selected two case-enriched and two control-enriched gene markers and measured their abundances by qPCR in a subset of cohort C1 (51 cases and 45 controls). Quantification by metagenomic sequencing and qPCR platforms showed strong correlations (Spearman  $r=0.81-0.95$ ; see online supplementary figure S9), suggesting that both measurements are reliable. Next, we measured the abundance of these four gene markers using qPCR in an independent Chinese cohort C2 (156 faecal samples; 47 cases and 109 controls; see online supplementary table S16). The two control-enriched genes did not

show significant associations in C2 ( $p>0.31$ ; see online supplementary table S17). On the other hand, CRC-enriched gene markers (m1704941, butyryl-CoA dehydrogenase from *F. nucleatum*; m482585, RNA-directed DNA polymerase from an unknown microbe) were also significantly enriched in CRC samples of C2 after adjusting for colonoscopy ( $p=0.0015$  and  $0.045$ , respectively, see online supplementary table S17). Among these, only the gene from *F. nucleatum* exhibited a significant OR after a Mantel-Haenszel test adjusted for colonoscopy (OR 18.5,  $p=0.0051$ ; see online supplementary table S17). CRC index based on abundances of the four genes only moderately classified CRC microbiomes from control microbiomes in C2 (areas under the receiver-operating curve (AUC)=0.73; see online supplementary figure S10), suggesting that choosing randomly from the list of 20 biomarkers was not an effective strategy. Nevertheless, the gene from *F. nucleatum* was present only in 4 out of 109 control microbiomes, suggesting a potential for developing specific diagnostic tests for CRC using faecal samples.

### Gene marker validation in independent metagenomic cohorts

To identify robust biomarkers that can have a more general applicability, we evaluated all 20 gene markers using faecal metagenomes from a cohort with different genetic background and lifestyle: 16 patients with CRC and 24 control individuals from Denmark (cohort D; see online supplementary table S18). When mapped to 4.3 million gut microbial genes, Danish metagenomes exhibited significantly higher gene richness and gene alpha diversity, both in cases (Wilcoxon rank-sum tests, gene count:  $p=1.94 \times 10^{-5}$ ; Shannon's index:  $p=5.85 \times 10^{-5}$ ) and controls (gene count:  $p=0.0017$ ; Shannon's index:  $p=9.34 \times 10^{-4}$ ; see online supplementary figure S11 and table S19), agreeing with a recent study and suggesting differences in gut microbial community structure between Chinese and Danish populations.<sup>39</sup> Among the 102 514 genes associated with CRC status in cohort C1, only 1498 genes could be validated in cohort D. However, CRC-enriched genes were shared significantly more between the two populations than control-enriched genes (1452 out of 35 735 CRC-enriched vs 46 out of 66 779 in control-enriched; two-tailed  $\chi^2$  test,  $\chi^2=2576.57$ ,  $p<0.0001$ ). Over half (53.6%) of the 1452 CRC-enriched genes were from just three species: *P. micra* (389 genes), *S. moorei* (204 genes) and *Clostridium symbiosum* (177 genes) (see online supplementary table S20). At the species level, *P. micra* was enriched in CRC microbiomes using all three methods, while *P. stomatis*, *Gemella morbillorum* and *S. moorei* were enriched according to two methods (Wilcoxon rank-sum test,  $q<0.05$ ; see online supplementary table S21). Notably, all species that were validated by at least one method were CRC-enriched. These results suggest that changes in colorectal environment during CRC development and progression may facilitate growth of similar species across the two populations, potentially leading to the reduced microbial diversity observed in patients with CRC (see online supplementary figure S1C), in line with earlier observations by others.<sup>40</sup> CRC index using 20 gene markers discovered in cohort C1 marginally differentiated Danish patient microbiomes from controls (Wilcoxon rank-sum test,  $p=0.029$ ) and exhibited moderate classification potential (area under ROC curve, AUC=0.71; see online supplementary figure S12). Only 4 out of 20 genes (two from *Peptostreptococcus anaerobius* and one each from *P. micra* and *F. nucleatum*) were associated with CRC status in cohort D (Wilcoxon rank-sum test,  $q \leq 0.05$ ; all CRC-enriched; see online

supplementary table S22). Among the factors we had recorded, only CRC status could explain the variation in these four genes (PERMANOVA  $p \leq 0.0001$ ; see online supplementary table S23).

For additional unbiased validation of the four gene markers, we used two recently published metagenomic datasets—an Austrian population (cohort A) consisting of 55 controls and 41 patients with CRC<sup>31</sup> and a French population (cohort F) consisting of 61 controls and 53 patients with CRC.<sup>22</sup> As our discovery cohort C1 only included carcinoma samples, we excluded all patients with adenoma and compared carcinoma patients with non-adenoma/non-carcinoma controls, contrary to the strategy used by the latter study<sup>22</sup> that included small adenomas in controls and excluded large adenomas. All four genes were significantly enriched in carcinoma faecal samples from both cohorts (Wilcoxon rank-sum test  $q < 0.0035$ ; see online supplementary table S24). CRC index using these four genes classified patients with CRC with AUC of 0.77 and 0.72 for cohorts A and F, respectively. When we checked association of all 20 markers, cohorts A and F each could validate an additional gene associated with CRC (see online supplementary table S25). Interestingly, one marker enriched in control samples in cohort C1 was enriched in CRC samples in cohort A.

### Accurate classification of CRC using qPCR

Two of the four cross-ethnically validated gene markers were transposases from *P. anaerobius*. The third gene (m1704941, butyryl-CoA dehydrogenase from *E. nucleatum*) was incidentally among the two genes successfully validated using qPCR in cohort C2. The fourth gene from *P. micra* was the highly conserved *rpoB* gene encoding RNA polymerase subunit  $\beta$ , often used as a phylogenetic marker.<sup>41</sup> We performed additional qPCR measurements of *rpoB* from *P. micra* in cohort C2, which showed a significant enrichment in CRC patient microbiomes (Wilcoxon rank-sum test adjusted for colonoscopy,  $p = 8.97 \times 10^{-8}$ ). Mantel–Haenszel OR adjusted for colonoscopy was 20.17 (95% CI 4.59 to 88.6,  $p = 3.36 \times 10^{-7}$ ). Combined qPCR measurements of the two genes clearly separated CRC from control samples in cohort C2 (Wilcoxon rank-sum test adjusted for colonoscopy,  $p = 1.384 \times 10^{-8}$ , figure 3A) and accurately classified CRC samples with an improved AUC of 0.84 (true-positive rate (TPR)=0.723; false-positive rate (FPR)=0.073; figure 3B). Accuracy was slightly better than that in a recent study (reporting AUC=0.836, TPR=0.58, FPR=0.08), even though they used a combination of abundances of 22 species using metagenomic sequencing.<sup>22</sup> Mantel–Haenszel OR, adjusted for colonoscopy, for detecting at least one of the two markers by qPCR in patients with CRC was 22.99 (95% CI 5.83 to 90.8,  $p = 5.79 \times 10^{-8}$ ). When stratifying cohort C2 into early-stage (stages I–II) and late-stage (stages III–IV) patients with cancer, classification potential and ORs were still significant (see online supplementary table S26). Abundance of these two genes was significantly higher compared with control samples starting from stage II of CRC (figure 3C, D), agreeing with our results from species abundances and providing proof-of-principle that faecal metagenomes may harbour non-invasive biomarkers for identification of early-stage CRC.

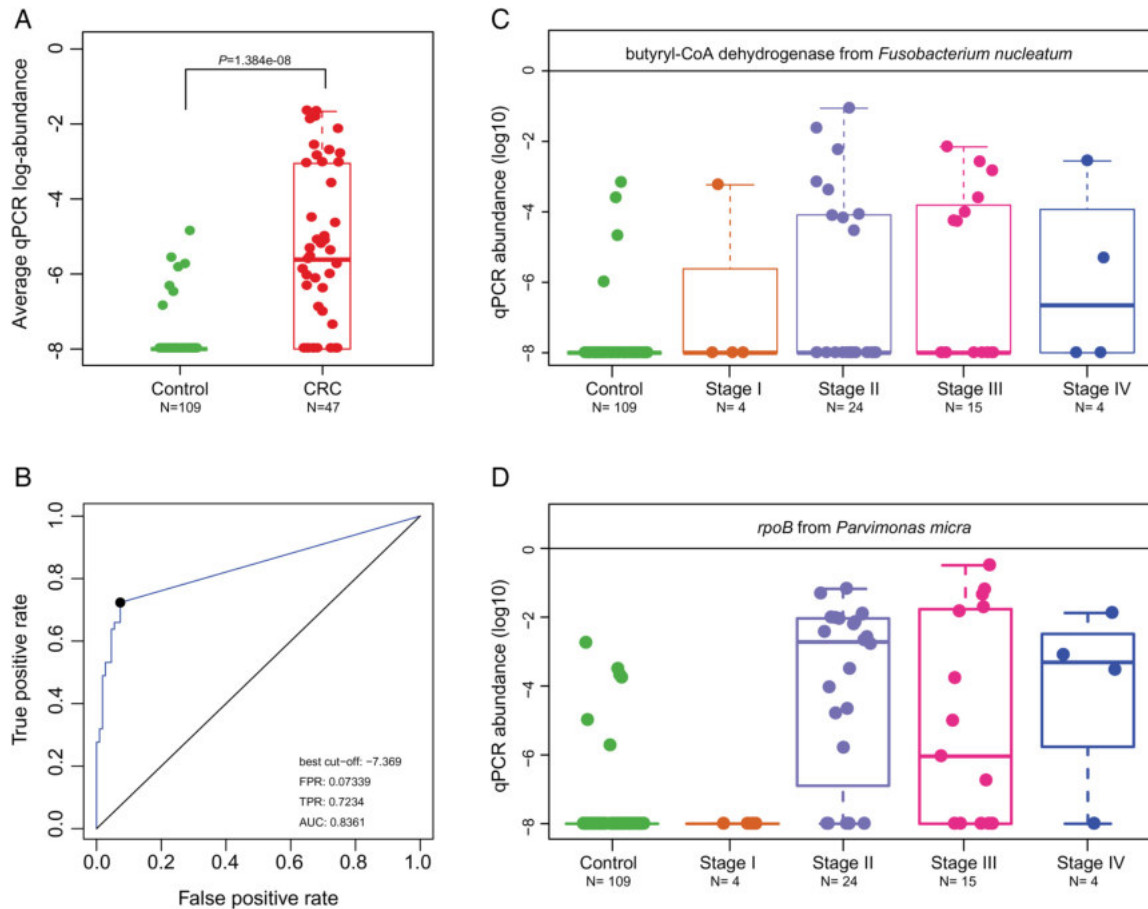
### DISCUSSION

We have reported the first successful cross-ethnic validation of metagenomic gene markers for CRC, notably including data from four countries. Two recent studies reported on potential CRC diagnosis using metagenomic sequencing of faecal microbiomes. The first study based on 16S ribosomal RNA gene used five operational taxonomic units to classify CRC from

healthy samples in a cohort from the USA.<sup>21</sup> As they did not perform any independent validation, we cannot compare our validation accuracy with theirs. The second study based on shotgun metagenomic sequencing used 21 species discovered in a French cohort to accurately classify patients with CRC in a German cohort.<sup>22</sup> Higher accuracy in their external validation (AUC=0.85 compared with our AUC of 0.77 and 0.72) could be because the validation cohort comes from the same ethnic group. Indeed, when two gene markers discovered in Chinese cohort C1 were validated in the independent Chinese cohort C2 using qPCR, we also achieved a high accuracy (AUC=0.84) even though we moved to a different platform. By doing so, we have also demonstrated, for the first time, the potential for CRC diagnosis through affordable targeted detection methods for microbial biomarkers in faecal samples. Significant improvement in the qPCR classification potential (from AUC=0.73 to AUC=0.84) by using a gene (*rpoB* gene from *P. micra*) validated in cohorts D, F and A reiterates the importance of validating newly discovered biomarkers in independent cohorts with different genetic and environmental background. Further work performing biomarker discovery in high-diversity cohorts or a meta-analysis of published cohorts could reveal whether it leads to increased predictive power. Combining metagenomic markers with the current clinical standard test (faecal occult blood test (FOBT)) has been shown to improve TPR from 49% to 72%.<sup>22</sup> The two markers reported here have reached a comparable TPR without using FOBT. It remains to be seen whether combining FOBT with these markers will further improve accuracy.

Gene markers shared between cohorts from China, Denmark, Austria and France suggest that even though different populations may have different microbial community structures, signatures of CRC-associated microbial dysbiosis could have universal features. Several important observations should be noted: (i) CRC-enriched gene markers had higher correlation between metagenomic and qPCR abundances ( $r=0.93$  and  $r=0.95$ ) compared with control-enriched genes ( $r=0.81$  and  $r=0.85$ ) in cohort C1; (ii) among four gene markers randomly tested using qPCR in cohort C2, only CRC-enriched genes were validated; (iii) all four gene markers validated in cohort D, all five markers validated in cohort A and four out of five markers validated in cohort F were CRC-enriched (see online supplementary table S25), even though there were 12 control-enriched markers compared with only 8 CRC-enriched markers; (iv) the only marker that switched enrichment during validation in different cohorts was control-enriched; (v) cohort D shared significantly more CRC-enriched genes than control-enriched genes with cohort C1; and (vi) all CRC-associated species from cohort C1 validated in cohort D were CRC-enriched. These features suggest that CRC-enriched biomarkers have a higher chance to be shared across populations and have better diagnostic potential than control-enriched biomarkers. One explanation could be that biomarkers for being healthy are harder to find than biomarkers for a specific disease, which goes against the Anna Karenina principle applied to gut microbiome that predicts higher number of disease-specific disturbed states than undisturbed states.<sup>42</sup> Although it is mandatory to have further validation for all biomarkers in larger cohorts across different populations, our results provide a proof of principle that development of an affordable diagnostic test using faecal microbial gene markers to identify patients with CRC may indeed be possible.

The finding that only two microbial metabolic modules associated with CRC status suggests that the role of microbial pathogens may be more important in disease development than that



**Figure 3** Validating robust gene markers associated with colorectal cancer (CRC). Quantitative PCR (qPCR) abundance of two gene markers (m1704941: butyryl-CoA dehydrogenase from *Fusobacterium nucleatum*, m1696299: RNA polymerase subunit  $\beta$ , *rpoB*, from *Parvimonas micra*) were measured in cohort C2 consisting of 47 cases and 109 healthy controls. Combined log-abundance of the two genes clearly separates CRC microbiomes from controls (A) and classifies CRC microbiomes with an area under the receiver operating characteristic curve of 0.84 (B). The two marker genes show relatively higher incidence and abundance in CRC stages II and III compared with control and stage I microbiomes (C and D). Abundances are plotted in log<sub>10</sub> scale, and zero abundance is plotted as  $-8$ . AUC, areas under the receiver-operating curve; FPR, false-positive rate; TPR, true-positive rate.

of functional abnormalities of the gut microbiome. Alternatively, expression levels of microbial genes may be more important than functional potential. Further research employing metatranscriptomic studies of microbial gene expression levels will clarify this.

The fact that only CRC-enriched genes and species could be validated across cohorts limits our conclusions on species depleted in CRC-associated microbiomes. We observed significant over-representations of several oral pathogens—*P. micra*, *P. stomatis*, *S. moorei* and *E. nucleatum* in the stool from patients with CRC, suggesting an oral–gut translocation route associated with CRC. Even though we cannot prove this route without further experiments, a recent study based on 300 healthy individuals reported that oral and gut microbiomes were predictive of each other, supporting this view.<sup>43</sup> While some of these species have been statistically associated with oral cancer in earlier studies,<sup>21 22 40</sup> only *E. nucleatum* has been shown to promote a proinflammatory environment leading to tumorigenesis.<sup>19</sup> Our study now introduces *P. micra* as a novel bacterial candidate involved in CRC-associated dysbiosis showing stronger associations with CRC across all five cohorts we investigated. Strong co-occurrence pattern between *P. micra* and the Gram-negative *E. nucleatum*,<sup>44</sup> and the former's ability to increase its capacity to induce inflammatory responses by binding to lipopolysaccharides from Gram-negative bacteria,<sup>45</sup>

could mean cooperation between the two, both in terms of colonisation strategies and in promoting a proinflammatory tumorigenic microenvironment. Enrichment of these species starts as early as in stage II of CRC, suggesting that they may play a role in the progression of CRC. Further work characterising *P. micra* could elucidate its role in CRC.

We have demonstrated consistent faecal microbial changes in CRC across four cohorts, identified novel bacterial candidates that may be involved in the development and progression of CRC, validated gene markers in three cohorts from three different countries and reported two bacterial genes that could serve as effective diagnostic biomarkers of CRC. Systematic investigation of key species and gene markers identified here might reveal further candidates. Additional work will be imperative (i) to benchmark these observations against currently used diagnostic approaches, (ii) to identify additional markers with improved predictive value and (iii) to eventually validate them in much larger cohorts. The ultimate goal would be to identify faecal metagenomic markers with strong predictive power to detect early stages of CRC, which would significantly reduce CRC-associated mortality.

#### Author affiliations

<sup>1</sup>Department of Medicine & Therapeutics, State Key Laboratory of Digestive Disease, Institute of Digestive Disease, LKS Institute of Health Sciences, CUHK Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong

<sup>2</sup>BGI-Shenzhen, Shenzhen, China

<sup>3</sup>Department of Biology, University of Copenhagen, Copenhagen, Denmark

<sup>4</sup>Department of Veterinary Disease Biology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

<sup>5</sup>Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia

<sup>6</sup>Department of Surgical Gastroenterology, Hvidovre Hospital, Hvidovre, Denmark

<sup>7</sup>National Institute of Nutrition and Seafood Research, Bergen, Norway

<sup>8</sup>Department of Internal Medicine, Hospital Oberndorf, Q3 Teaching Hospital of the Paracelsus Private University of Salzburg, Oberndorf, Austria

<sup>9</sup>First Department of Internal Medicine, Medical University Innsbruck, Innsbruck, Austria

<sup>10</sup>The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

<sup>11</sup>Macau University of Science and Technology, Macau, China

**Correction notice** This article has been corrected since it published Online First. The data sharing statement has been corrected.

**Contributors** JuY, QF, SHW, DZ and QL contributed equally. The project was designed by JuW, JoS, JuY, NB and MA. JuY, QF, JoS and JuW managed the project. JuY, TOY, JS, HJN, TYTL, SCN, QL, ASLC, VW-SW, WKKW and FKLC contributed to acquisition of clinical samples, patients' information and clinical data analyses. QL, SHW, JuY, ZL, PK and BAHJ performed DNA experiments. JuW, JuY, SHW, MA, KK, QF and DYZ designed the analysis. DYZ, MA, QF, YWQ, LQT, YLL, YL, NC, HJJ, JHL, LX and ZL performed the data analysis. DYZ, MA, QF, YWQ, LQT, YLL, YL, NC, HJJ, JHL, LX and ZL worked on metagenomic-wide association study. QL, JuY and T OY did the experimental validation. MA, JuY, SH W, QF, DY Z and LQT wrote the paper. JuW, KK, LM, JoS, JuY, NB, JiW, HMY, HJJ, JA-A and XX revised the paper.

**Funding** The project was supported by the National Basic Research Program of China (973 Program, 2011CB809203, 2013CB531401), SHHO foundation Hong Kong, theme-based Research Scheme of the Hong Kong Research Grants Council (T12-403-11), the introduction of innovative R&D team programme of Guangdong Province (no. 2009010016), the Danish Cancer Society (R72-A4659-13-S2) and the Shenzhen Municipal Government of China (CXB201108250098A).

**Competing interests** None declared.

**Patient consent** Obtained.

**Ethics approval** Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee (CUHK-NTEC CREC), Ethics Committee of the Capital Region of Denmark and Danish Data Protection Agency.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** Metagenomic sequence data set has been deposited in European Nucleotide Archive with accession number PRJEB10878.

## REFERENCES

- Ferlay J, Soerjomataram I, Ervik M, et al. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide*. IARC CancerBase. Lyon, France: International Agency for Research on Cancer, 2013.
- Galvan A, Ioannidis JP, Dragani TA. Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. *Trends Genet* 2010;26:132–41.
- Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78–85.
- Foulkes WD. Inherited susceptibility to common cancers. *N Engl J Med* 2008;359:2143–53.
- Dove WF, Clipson L, Gould KA, et al. Intestinal neoplasia in the ApcMin mouse: independence from the microbial and natural killer (beige locus) status. *Cancer Res* 1997;57:812–14.
- Arthur JC, Perez-Chanona E, Muhlbauer M, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 2012;338:120–3.
- Cuevas-Ramos G, Petit CR, Marq I, et al. *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc Natl Acad Sci USA* 2010;107:11537–42.
- Grivennikov SI, Wang K, Mucida D, et al. Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. *Nature* 2012;491:254–8.
- Toprak NU, Yagci A, Gulluoglu BM, et al. A possible role of *Bacteroides fragilis* enterotoxin in the aetiology of colorectal cancer. *Clin Microbiol Infect* 2006;12:782–6.
- Uronis JM, Muhlbauer M, Herfarth HH, et al. Modulation of the intestinal microbiota alters colitis-associated colorectal cancer susceptibility. *PLoS ONE* 2009;4:e6026.
- Wu S, Rhee KJ, Albesiano E, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat Med* 2009;15:1016–22.
- Boleij A, Schaeps RM, Tjalsma H. Association between *Streptococcus bovis* and colon cancer. *J Clin Microbiol* 2009;47:516.
- Seder CW, Kramer M, Long G, et al. Clostridium septicum aortitis: Report of two cases and review of the literature. *J Vasc Surg* 2009;49:1304–9.
- Scanlan PD, Shanahan F, Clune Y, et al. Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis. *Environ Microbiol* 2008;10:789–98.
- Sobhani I, Tap J, Roudot-Thoraval F, et al. Microbial dysbiosis in colorectal cancer (CRC) patients. *PLoS ONE* 2011;6:e16393.
- Chen W, Liu F, Ling Z, et al. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS ONE* 2012;7:e39743.
- Castellari M, Warren RL, Freeman JD, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res* 2012;22:299–306.
- Kostic AD, Gevers D, Pedamallu CS, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res* 2012;22:292–8.
- Kostic AD, Chun E, Robertson L, et al. Fusobacterium nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. *Cell Host Microbe* 2013;14:207–15.
- Rubinstein MR, Wang X, Liu W, et al. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. *Cell Host Microbe* 2013;14:195–206.
- Zackular JP, Rogers MA, Ruffin MT, et al. The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prev Res (Phila)* 2014;7:1112–21.
- Zeller G, Tap J, Voigt AY, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* 2014;10:766.
- Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* 2010;17:1471–4.
- Godon JJ, Zumstein E, Dabert P, et al. Molecular microbial diversity of an anaerobic digester as determined by small-subunit rDNA sequence analysis. *Appl Environ Microbiol* 1997;63:2802–13.
- Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 2012;490:55–60.
- Sunagawa S, Mende DR, Zeller G, et al. Metagenomic species profiling using universal phylogenetic marker genes. *Nat Methods* 2013;10:1196–9.
- Markowitz VM, Chen IM, Palaniappan K, et al. IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* 2012;40:D115–22.
- Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 2005;27:1226–38.
- Kabat GC, Kim MY, Strickler HD, et al. A longitudinal study of serum insulin and glucose levels in relation to colorectal cancer risk among postmenopausal women. *Br J Cancer* 2012;106:227–32.
- van Duynhoven FJ, Bueno-De-Mesquita HB, Calligaris M, et al. Blood lipid and lipoprotein concentrations and colorectal cancer risk in the European Prospective Investigation into Cancer and Nutrition. *Gut* 2011;60:1094–102.
- Feng Q, Liang S, Jia H, et al. Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun* 2015;6:6528.
- Kanehisa M, Goto S, Sato Y, et al. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 2012;40:D109–14.
- Baracos VE, Mackenzie ML. Investigations of branched-chain amino acids and their metabolites in animal models of cancer. *J Nutr* 2006;136:237S–42S.
- Gonçalves EM, Salomão EM, Gomes-Marcondes MCC. Leucine modulates the effect of Walker factor, a proteolysis-inducing factor-like protein from Walker tumours, on gene expression and cellular activity in C2C12 myotubes. *Cytokine* 2013;64:343–50.
- Pedersen RM, Holt HM, Justesen US. Solobacterium moorei bacteremia: identification, antimicrobial susceptibility, and clinical characteristics. *J Clin Microbiol* 2011;49:2766–8.
- Sundqvist G. Taxonomy, ecology, and pathogenicity of the root canal flora. *Oral Surg Oral Med Oral Pathol* 1994;78:522–30.
- Knights D, Costello EK, Knight R. Supervised classification of human microbiota. *FEMS Microbiol Rev* 2011;35:343–59.
- Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;464:59–65.
- Li J, Jia H, Cai X, et al. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 2014;32:834–41.
- Ahn J, Sinha R, Pei Z, et al. Human gut microbiome and risk for colorectal cancer. *J Natl Cancer Inst* 2013;105:1907–11.

- 41 Ciccarelli FD, Doerks T, von Mering C, *et al.* Toward automatic reconstruction of a highly resolved tree of life. *Science* 2006;311:1283–7.
- 42 Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS ONE* 2012;7:e30126.
- 43 Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. *Nature* 2014;509:357–60.
- 44 Kremer BH, van Steenberghe TJ. Peptostreptococcus micros coaggregates with Fusobacterium nucleatum and non-encapsulated Porphyromonas gingivalis. *FEMS Microbiol Lett* 2000;182:57–62.
- 45 Yoshioka M, Grenier D, Mayrand D. Binding of Actinobacillus actinomycetemcomitans lipopolysaccharides to Peptostreptococcus micros stimulates tumor necrosis factor alpha production by macrophage-like cells. *Oral Microbiol Immunol* 2005;20:118–21.

# A Novel Faecal *Lachnoclostridium* Marker for the Non-Invasive Diagnosis of Colorectal Adenoma and Cancer

Jessie Qiaoyi Liang, Tong Li, Geicho Nakatsu, Ying-Xuan Chen, **Tung On Yau**, Eagle Chu, Sunny Wong, Chun Ho Szeto, Siew C. Ng, Francis K. L. Chan, Jing-Yuan Fang, Joseph J. Y. Sung, Jun Yu.

*Gut.* **2020**;69(7):1248-1257.

Journal URL: [gut.bmj.com/content/69/7/1248](https://gut.bmj.com/content/69/7/1248)

DOI: [10.1136/gutjnl-2019-318532](https://doi.org/10.1136/gutjnl-2019-318532)

PMID: [31776231](https://pubmed.ncbi.nlm.nih.gov/31776231/)

PMCID: [PMC7306980](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC7306980/)



Statement of Joint Authorship

To Whom It May Concern:

I am writing to confirm that Tung On (Payton) Yau, a Research Fellow at Nottingham Trent University, collaborated with my team on colorectal cancer projects in 2017 and 2018. This included work on faecal-based miRNA and the gut microbiome. He was a significant contributor with responsibilities in protocol optimization, data management, and in carrying out experiments (i.e. DNA / miRNA extraction, cDNA conversion, real-time PCR).

From this work, he fulfils criteria for being a co-author in two submitted publications:

1. *A novel microRNA panel for non-invasive diagnosis and prognosis of colorectal cancer*

Conference abstract submitted to Digestive Disease Week 2019

Notification of outcome of abstract submission: February 12th, 2019

Anticipated publication of abstract: Gastroenterology (Impact Factor = 20.87), May 2019

2. *A Novel Fecal Bacterial Marker for the Non-invasive Diagnosis of Colorectal Adenoma*

*JQ Liang, T Li, Y Chen, **TO Yau**, ESH Chu, S Wong, CH Szeto, SC Ng, FKL Chan, JY Fang, JJY Sung, and J Yu*

Manuscript Number: CCR-18-3649

Journal: Clinical Cancer Research (Impact Factor = 10.2)

Date of Submission: November 2018

Current Status: Under review

I anticipate publication of both pieces of work within the next 6 to 12 months.

Yours Sincerely,

Dr Jessie Qiaoyi Liang

Research Assistant Professor

Institute of Digestive Disease and Department of Medicine and Therapeutics

Chinese University of Hong Kong



OPEN ACCESS

ORIGINAL RESEARCH

# A novel faecal *Lachnoclostridium* marker for the non-invasive diagnosis of colorectal adenoma and cancer

Jessie Qiaoyi Liang,<sup>1</sup> Tong Li,<sup>1</sup> Geicho Nakatsu,<sup>1</sup> Ying-Xuan Chen,<sup>2</sup> Tung On Yau,<sup>1</sup> Eagle Chu,<sup>1</sup> Sunny Wong ,<sup>1</sup> Chun Ho Szeto,<sup>1</sup> Siew C Ng ,<sup>1</sup> Francis K L Chan ,<sup>1</sup> Jing-Yuan Fang,<sup>2</sup> Joseph J Y Sung,<sup>1</sup> Jun Yu

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2019-318532>).

<sup>1</sup>Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, CUHK Shenzhen Research Institute, The Chinese University of Hong Kong, Shatin, Hong Kong

<sup>2</sup>Division of Gastroenterology, Shanghai Jiaotong University School of Medicine Renji Hospital, Shanghai Institute of Digestive Disease, Shanghai, China

## Correspondence to

Dr Jun Yu, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, Hong Kong; [junyu@cuhk.edu.hk](mailto:junyu@cuhk.edu.hk)  
Dr Jessie Qiaoyi Liang; [jessieq@cuhk.edu.hk](mailto:jessieq@cuhk.edu.hk)

Received 17 February 2019

Revised 11 November 2019

Accepted 14 November 2019

Published Online First

27 November 2019

## ABSTRACT

**Objective** There is a need for early detection of colorectal cancer (CRC) at precancerous-stage adenoma. Here, we identified novel faecal bacterial markers for diagnosing adenoma.

**Design** This study included 1012 subjects (274 CRC, 353 adenoma and 385 controls) from two independent Asian groups. Candidate markers were identified by metagenomics and validated by targeted quantitative PCR.

**Results** Metagenomic analysis identified 'm3' from a *Lachnoclostridium* sp., *Fusobacterium nucleatum* (*Fn*) and *Clostridium hathewayi* (*Ch*) to be significantly enriched in adenoma. Faecal m3 and *Fn* were significantly increased from normal to adenoma to CRC ( $p < 0.0001$ , linear trend by one-way ANOVA) in group I ( $n = 698$ ), which was further confirmed in group II ( $n = 313$ ;  $p < 0.0001$ ). Faecal m3 may perform better than *Fn* in distinguishing adenoma from controls (areas under the receiver operating characteristic curve (AUROCs)  $m3 = 0.675$  vs  $Fn = 0.620$ ,  $p = 0.09$ ), while *Fn* performed better in diagnosing CRC (AUROCs  $Fn = 0.862$  vs  $m3 = 0.741$ ,  $p < 0.0001$ ). At 78.5% specificity, m3 and *Fn* showed sensitivities of 48.3% and 33.8% for adenoma, and 62.1% and 77.8% for CRC, respectively. In a subgroup tested with faecal immunochemical test (FIT;  $n = 642$ ), m3 performed better than FIT in detecting adenoma (sensitivities for non-advanced and advanced adenomas of 44.2% and 50.8% by m3 (specificity = 79.6%) vs 0% and 16.1% by FIT (specificity = 98.5%)). Combining with FIT improved sensitivity of m3 for advanced adenoma to 56.8%. The combination of m3 with *Fn*, *Ch*, *Bacteroides clarus* and FIT performed best for diagnosing CRC (specificity = 81.2% and sensitivity = 93.8%).

**Conclusion** This study identifies a novel bacterial marker m3 for the non-invasive diagnosis of colorectal adenoma.

## INTRODUCTION

Colorectal cancer (CRC) is one of the most common malignancies worldwide.<sup>1</sup> A higher incidence of CRC has been observed in more developed regions than less developed regions, and an increased incidence of CRC is believed to have attributed to changes in diet.<sup>1,2</sup> Recent evidences have shown that an altered microbiome environment in the gut is associated with colorectal tumourigenesis. Abnormality in the composition of the gut microbiota has been implicated as a potentially important

## Significance of this study

### What is already known on this subject?

- Early detection of colonic adenomas and cancer can facilitate the successful treatment and significantly reduces the incidence of colorectal cancer (CRC).
- Molecular markers for adenoma, especially non-advanced adenoma, is limited.

### What are the new findings?

- A new gene marker from a *Lachnoclostridium* sp., labelled as m3, was identified to be enriched in faecal samples of patients with adenoma by metagenomic analysis.
- m3 showed the best performance in diagnosing adenoma in two independent Asian groups of 1012 subjects by quantitative PCR, which is superior to currently available stool-based tests.
- Combination of m3 with faecal immunochemical test (FIT) improved diagnostic sensitivity from 50.8% to 56.8% (specificity 79.6%) for advanced adenoma, while combination of m3 with other bacterial markers (*Fn*, *Ch*, *Bc*) and FIT showed good diagnostic performance for CRC (specificity = 81.2% and sensitivity = 93.8%).

### How might it impact on clinical practice in the foreseeable future?

- m3 is a novel stool-based non-invasive biomarker for patients with adenoma and CRC.

aetiological factor in the initiation and progression of CRC.<sup>3</sup> With the widespread application of metagenomic analyses in the investigation of intestinal microbiota, an increasing number of bacteria have been identified to be positively associated with CRC.<sup>4-7</sup> Recent basic research has established a critical function for the intestinal microbiota<sup>8</sup> and specific bacterial species, such as *Fusobacterium nucleatum* (*Fn*)<sup>9-11</sup> and *Peptostreptococcus anaerobius*,<sup>12</sup> in promoting colorectal tumourigenesis. Bacteria such as *Fn*,<sup>13</sup> *Clostridium symbiosum*<sup>14</sup> and species within the genera *Parvimonas*, *Porphyromonas* and *Parabacteroides*<sup>15</sup> have been shown to be potential markers for the diagnosis of patients with CRC. However, current knowledge on biomarkers for colorectal adenoma detection is limited.

Check for updates

© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Liang JQ, Li T, Nakatsu G, et al. *Gut* 2020;**69**:1248–1257.



Early detection of cancer can facilitate successful treatment. Endoscopic removal of colorectal adenomas, precursors of most CRCs, significantly reduces the risk of CRC. Early detection of adenomas is thus important for decreasing CRC morbidity and mortality. The most widely used non-invasive stool test is the faecal immunochemical test (FIT), which shows unsatisfying sensitivities for CRC (0.79 (95% CI 0.69 to 0.86); differed greatly among various studies) and is not sensitive for adenoma.<sup>16</sup> Sensitivity of FIT for advanced adenoma varied from 6% to 56%, with screening studies involving cohort sizes over 8000 all showing sensitivities of less than 28%.<sup>17</sup> Therefore, identification of molecular markers that improve the diagnostic sensitivity for adenoma is warranted.

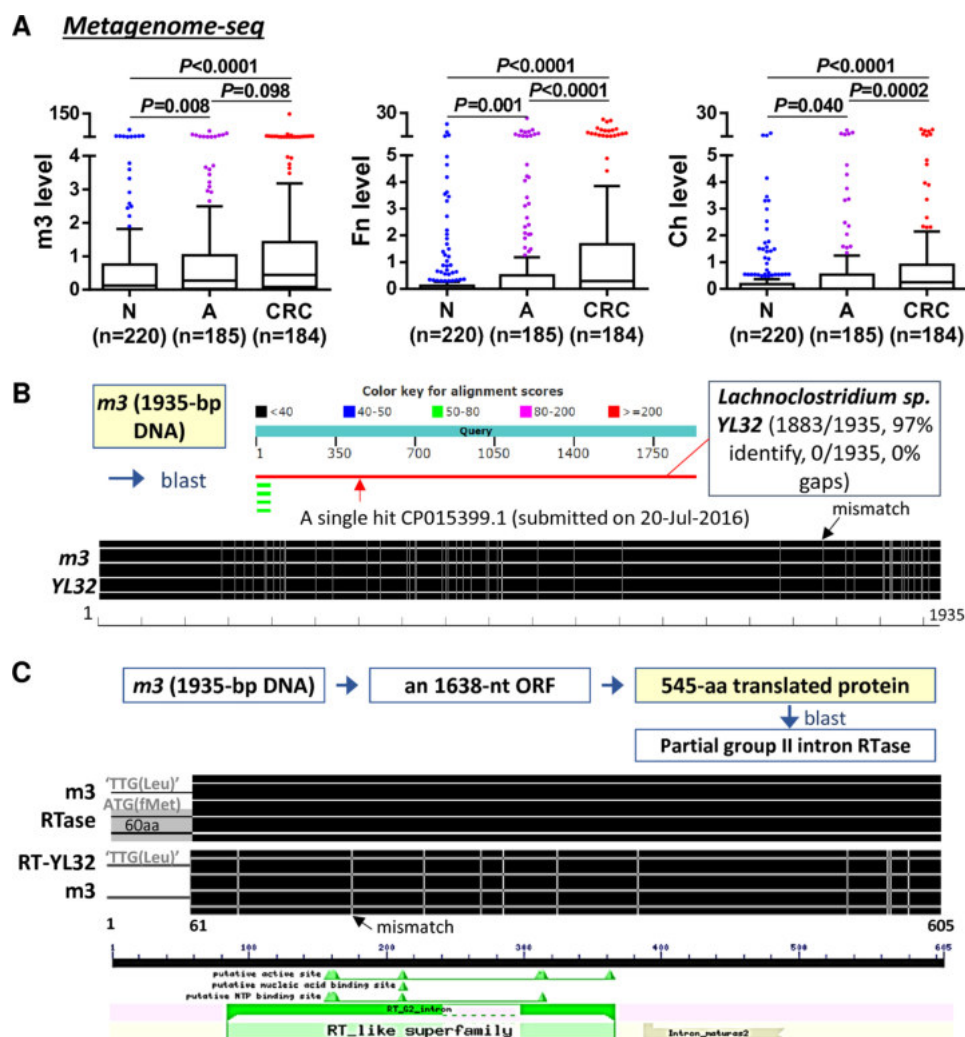
Using metagenomic analysis to compare the faecal microbiome of patients with CRC and healthy subjects, we identified 20 bacterial gene marker candidates that may serve as non-invasive biomarkers for CRC.<sup>4</sup> We further showed that stool-based bacteria could serve as non-invasive diagnostic biomarkers for CRC by targeted quantification using quantitative PCR (qPCR).<sup>13</sup> We showed that *Fn* was a good marker for CRC, and combination with three others (*Clostridium hathewayi* (*Ch*), undefined '*m7*' and *Bacteroides clarus* (*Bc*)) could further

improve the diagnostic performance of *Fn*. However, the diagnostic performance of these bacterial gene markers for adenoma was limited. In this study, we identified and evaluated the utility of a new *Lachnospirillum* gene marker (labelled as '*m3*') for the diagnosis of colorectal adenoma. The diagnostic performance of *m3*, comparing with and in combination with other bacterial gene markers and FIT, was tested in 1012 subjects from two independent groups.

## RESULTS

### Identification of '*m3*' from a *Lachnospirillum* species as a potential biomarker for colorectal neoplasm

To investigate whether our previously identified 20 bacterial gene markers for CRC<sup>4</sup> may also serve as biomarkers for adenoma, we analysed their abundances in our in-house metagenomics data from 589 Asian subjects (184 CRC, 185 adenoma and 220 control subjects) (online supplementary table S1). Among them, the marker labelled as '*m3*', which was not assignable to any known species at the time of the previous discovery study,<sup>4</sup> was found to be significantly enriched in patients with CRC and adenoma as compared with control subjects, as well



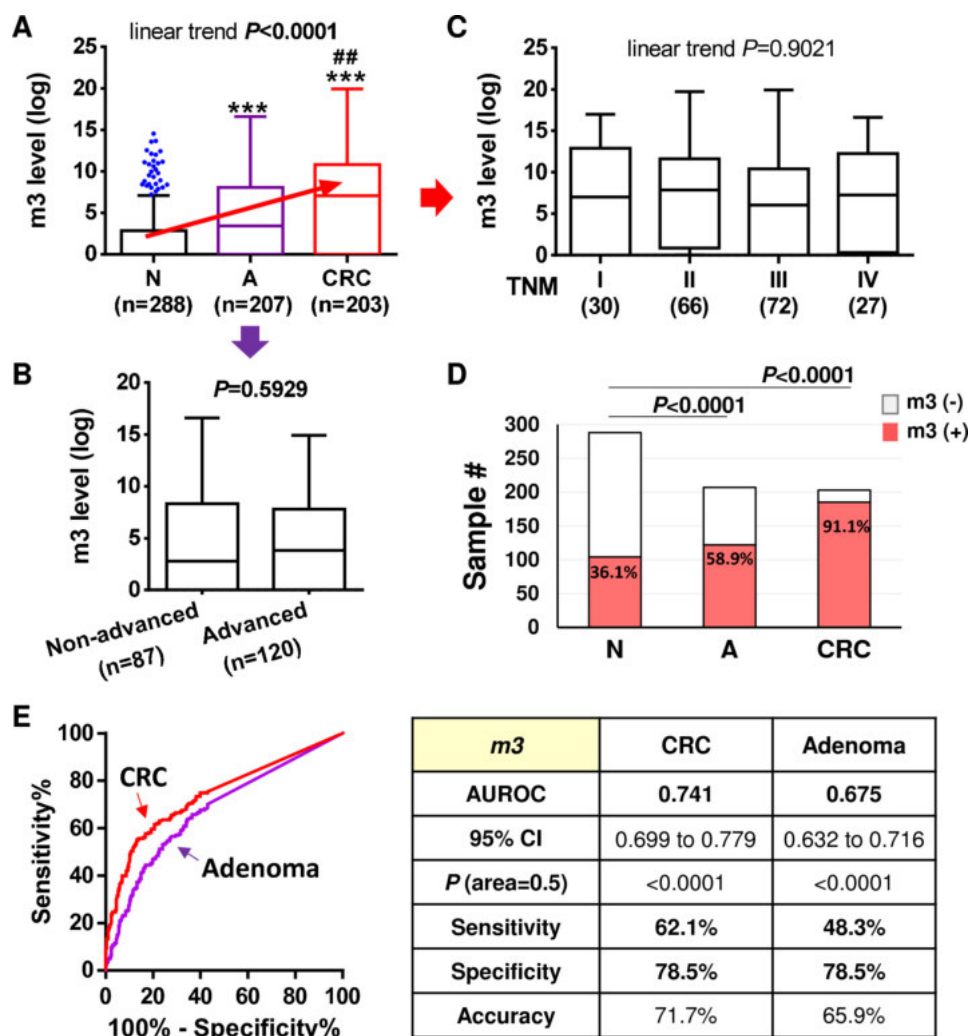
**Figure 1** Identification and characterisation of *m3*. (A) Metagenome sequencing identified *m3*, as well as *Fusobacterium nucleatum* (*Fn*) and *Clostridium hathewayi* (*Ch*), to be significantly increased in faecal samples of patients with adenoma. (B) DNA sequence of *m3* showed high similarity to *Lachnospirillum* sp. YL32. (C) *m3* encodes a putative reverse transcriptase (RTase) that maps to a group II intron RTase, lacking the first 60 amino acids but retaining the RTase conserved domain. A, adenoma; CRC, colorectal cancer; N, normal control.

as two previously verified CRC markers (*Fn* and *Ch*) (figure 1A; other gene markers in online supplementary figure 1). A blast search for the 1935 nt *m3* sequence in the non-redundant nucleotide collection of NCBI identified *Lachnospirillum* sp. YL32, a new species with genome sequence deposited in GenBank in July 2016 (accession no. CP015399). *m3* and *Lachnospirillum* sp. YL32 shared 97% (1883/1935) DNA sequence similarity (figure 1B). The *m3* DNA contains a 1638 nt open reading frame (nt 298–1935), encoding a putative 545 aa protein with 100% sequence similarity to a group II RTase (GenBank accession no. WP\_055650193). Although *m3* protein lacks the first 60 aa due to 'TTG' codon instead of 'ATG' at the corresponding translation start site, it retains the RTase conserved domain (figure 1C). The corresponding sequence in *Lachnospirillum* sp. YL32 genome also encodes a partial group II RTase, showing 98% (534/545) sequence similarity with *m3*-RTase and the group II RTase. We further analysed the abundance of *Lachnospirillum* sp. YL32 genome based on the Prokka-annotated protein coding gene sequences in our in-house metagenomics data. The result

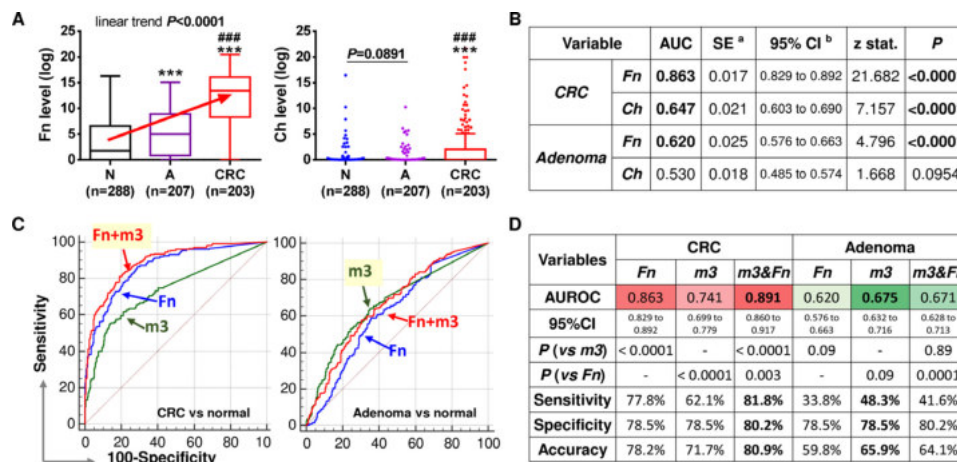
showed that *Lachnospirillum* sp. YL32 was significantly increased in adenoma but to a less extent in CRC as compared with control subjects (online supplementary figure 2A). Therefore, the candidate gene marker *m3* may belong to *Lachnospirillum* species close to *Lachnospirillum* sp. YL32.

#### Validation of *m3* as a novel faecal biomarker for colorectal adenoma by qPCR

We further quantitatively examined the abundance of *m3* in an enlarged group of stool samples from Hong Kong Chinese by using our previously established duplex-qPCR platform.<sup>13</sup> The results showed that faecal *m3* level was significantly higher in patients with adenoma (n=207) versus control subjects (n=288), and was significantly higher in patients with CRC (n=203) versus control subjects or patients with adenoma (all p<0.0001 by multiple comparison). There was a significant linear trend of *m3* increasing from control to adenoma to cancer (p<0.0001, one-way ANOVA) (figure 2A). Interestingly, *m3* level was similar



**Figure 2** Quantitative detection of faecal *m3* in the diagnosis of patients with colorectal cancer (CRC) and adenoma. (A) Relative abundance of *m3* in faecal samples differed significantly between healthy control subjects (N, n=288), patients with adenoma (A, n=207) and patients with CRC (n=203). \*\*\*p<0.0001 as compared with N; ##p<0.001 as compared with A. (B) No significant difference in faecal abundance of *m3* was observed between non-advanced and advanced adenomas. (C) No difference in faecal abundance of *m3* was observed among patients with CRC of different tumour-node-metastasis (TNM) stages. (D) Occurrence rates of *m3* was significantly higher in patients with adenoma compared with control subjects, and highest in patients with CRC. (E) Receiver operating characteristic (ROC) curves and diagnostic performance of *m3* in discriminating patients with CRC and adenoma from control subjects, respectively. AUROC, area under ROC.



**Figure 3** Comparison and combination of bacterial markers for non-invasive diagnosis of colorectal cancer (CRC) and adenoma. (A) Relative abundances of *Fusobacterium nucleatum* (*Fn*), *Clostridium hathewayi* (*Ch*) in faecal samples of control subjects, patients with adenoma and patients with CRC. N, normal control; A, adenoma; \*\*\* $p < 0.0001$  as compared with N; #### $p < 0.0001$  as compared with A. (B) ROC curve analyses showed *Fn* could discriminate adenoma and CRC from controls, while *Ch* could discriminate CRC but not adenoma from controls. (C) Comparison of ROC curves of *Fn*, *m3* and their combination. (D) Diagnostic performances of *Fn*, *m3* and their combination. *Fn* performed better than *m3* in diagnosing CRC, and *m3* was superior to *Fn* in diagnosing adenoma. Combination with *Fn* improved the diagnostic performance of *m3* for CRC but not for adenoma.

between non-advanced and advanced adenomas (figure 2B). Similarly, *m3* level showed no significant change across tumour-node-metastasis staging in patients with CRC (figure 2C). The occurrence rate of faecal *m3* was significantly higher in patients with adenoma as compared with control subjects, and highest in patients with CRC (both  $p < 0.0001$  vs control subjects; figure 2D). Receiver operating characteristic (ROC) curve analysis showed that *m3* could significantly discriminate CRC and adenoma patients from control subjects, with areas under ROC (AUROCs) of 0.741 and 0.675 for CRC and adenoma, respectively (both  $p < 0.0001$ ; figure 2E). At specificity of 78.5%, *m3* showed sensitivities of 62.1% for CRC and 48.3% for adenoma; the accuracies were 71.7% and 65.9% for distinguishing patients with CRC and adenoma from control subjects, respectively. These results demonstrated that *m3* may serve as a new stool-based biomarker to assist the non-invasive diagnosis of CRC and adenoma.

### *m3* performs better than other bacterial markers (*Fn* and *Ch*) in diagnosing colorectal adenoma

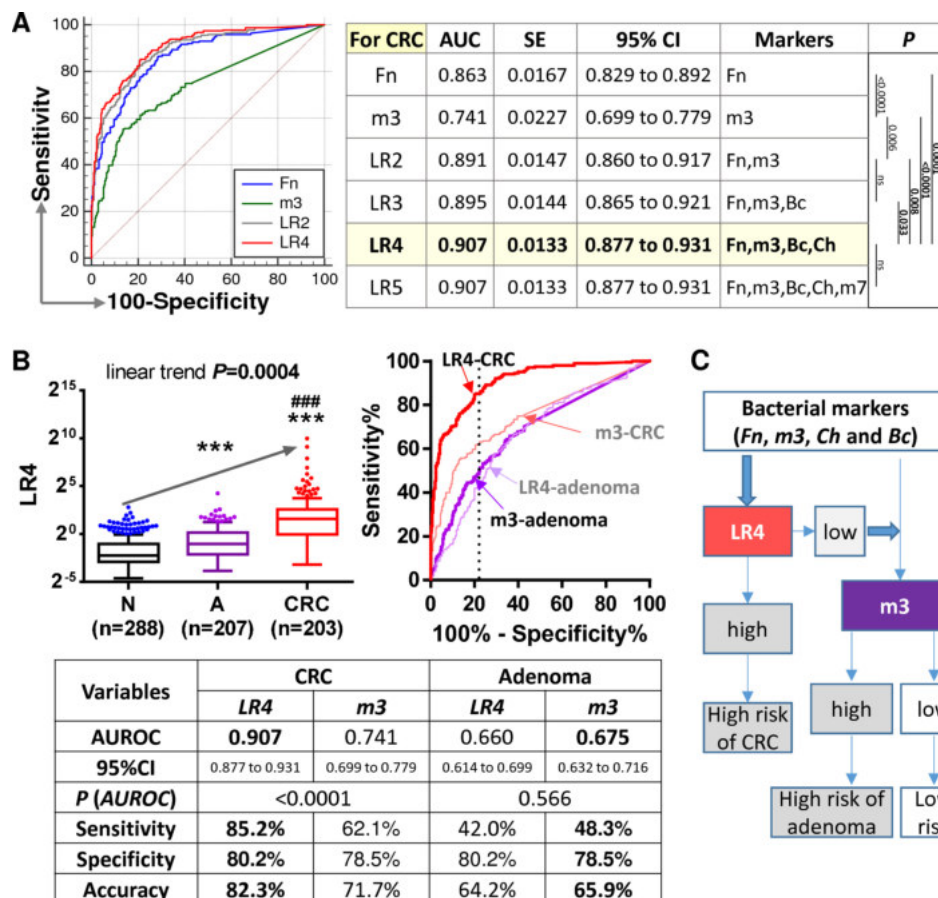
As *Fn* and *Ch* were also identified to be significantly increased in patients with adenoma by metagenome sequencing, we further examined the levels of *Fn* and *Ch* by qPCR and compared their diagnostic performances with *m3*. Results confirmed that,

similar to *m3*, the relative faecal abundance of *Fn* was significantly elevated in patients with adenoma compared with control subjects, and highest in patients with CRC, with a significant linear trend of increase during CRC development (all  $p < 0.0001$ ; figure 3A). However, faecal *Ch* was not significantly increased in patients with adenoma as compared with control subjects by qPCR ( $p > 0.05$ ). *Ch* was significantly enriched in patients with CRC compared with patients with adenoma and control subjects (both  $p < 0.0001$ , figure 3A). ROC curve analyses also showed that both *Fn* and *Ch* performed well in diagnosing CRC, but only *Fn* could significantly distinguish patients with adenoma from control subjects ( $p < 0.0001$ ; figure 3B). The abundances of both *Fn* and *m3* were not associated with gender, CRC staging, lesion location or body mass index. Multivariate analysis showed that *Fn* and *m3* were significantly associated with CRC and adenoma diagnosis, as well as *Fn* with age (table 1). Although *Fn* performed better than *m3* in CRC diagnosis (AUROC of  $Fn = 0.863$  vs  $m3 = 0.741$ ;  $p < 0.0001$ ), *m3* may work better than *Fn* for adenoma diagnosis as shown by comparison of ROC curves (AUROCs:  $m3 = 0.675$  vs  $Fn = 0.620$ ,  $p = 0.09$ ; figure 3C,D). As single diagnostic factors at specificity of 78.5%, *Fn* discriminated patients with CRC from control subjects with a sensitivity of 77.8% and accuracy of 78.2% (vs 62.1% and 71.7%, respectively, by *m3*). At specificity of 78.5%, *m3* discriminated patients

**Table 1** Correlations between faecal marker abundances and clinical characteristics

Markers	Univariate				Multivariate			
	<i>Fn</i>		<i>m3</i>		<i>Fn</i>		<i>m3</i>	
Variable	Coef	P value	Coef	P value	Coef	P value	Coef	P value
Age	0.217	<0.0001	0.111	<0.0001	0.064	0.007	0.025	0.269
Gender	-0.348	0.443	-0.143	0.706				
Diagnosis	2.768	<0.0001	1.535	<0.0001	2.597	<0.0001	1.472	<0.0001
Pre/post-colonoscopy	-0.181	0.729	-0.735	0.095				
BMI	0.052	0.761	-0.069	0.633				
CRC staging	0.173	0.689	-0.227	0.620				
Lesion location	-0.153	0.862	-0.201	0.830				

BMI, body mass index; CRC, colorectal cancer.



**Figure 4** Combination of four markers for colorectal cancer (CRC) and *m3* alone for adenoma. (A) Receiver operating characteristic (ROC) curve analysis of combination of the five bacterial markers of interest showed that combination of *Fn*, *m3*, *Ch* and *Bc* by a logistic regression (LR) model worked best for CRC diagnosis. Shown p values are by comparison ROC curves. (B) Level of the combination of *Fn*, *m3*, *Ch* and *Bc* (LR4) in faecal samples and comparison of its diagnostic performance with *m3*. N, normal control; A, adenoma; \*\*\* $p < 0.0001$  as compared with N; #### $p < 0.0001$  as compared with A. (C) Proposed strategy for the application of *Fn*, *m3*, *Ch* and *Bc* in the diagnosis of CRC and adenoma.

with adenoma from control subjects with a sensitivity of 48.3% and accuracy of 65.9% (vs 33.8% and 59.8%, respectively, by *Fn*) (figure 3C,D). Interestingly, combination of *Fn* and *m3* by a logistic regression model significantly improved their individual diagnostic performances for CRC (AUROC=0.891,  $p=0.003$  vs *Fn*) but not for adenoma (AUROC=0.671,  $p=0.89$  vs *m3*). These results demonstrate that *m3* is a potential good diagnostic biomarker especially for adenoma.

#### Combination with other bacterial markers (*Fn*, *Bc* and *Ch*) increases the diagnostic performance of *m3* for CRC, while *m3* alone works best for adenoma

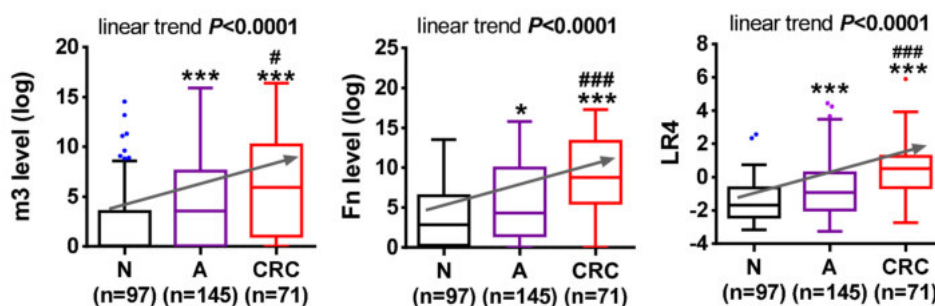
As we have previously reported bacterial markers *Fn*, *Ch*, *Bc* and *m7* for diagnosis of CRC, we further tested the performance of the combination of these markers with *m3* for CRC. The results showed that combination of *Fn*, *m3*, *Ch* and *Bc* by a logistic regression model performed best in diagnosing CRC, with an AUROC of 0.907 (all  $p < 0.05$  as compared with combinations of fewer markers by comparison of ROC curves; figure 4A). At specificity of 80.2%, combination of *Fn*, *m3*, *Ch* and *Bc* showed a sensitivity of 85.2% and accuracy of 82.3% for CRC. Although combination of *Fn*, *m3*, *Ch* and *Bc* showed significantly increased score in patients with adenoma as compared with control subjects, its diagnostic performance for adenoma was not better than *m3* (AUROC=0.660 vs *m3*=0.675,  $p=0.086$ ; figure 4B). Therefore, combination of *Fn*, *m3*, *Ch* and *Bc* may

serve as a novel tool for the non-invasive diagnosis of CRC, while *m3* alone may be applied for further detection of adenoma (figure 4C).

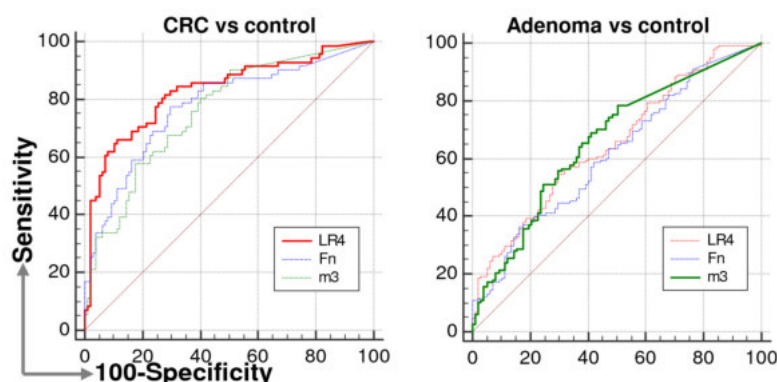
#### Verification of diagnostic performance of bacterial markers in a second independent group

We further tested the bacterial markers in a second independent group of 313 samples from Shanghai China (71 CRC, 145 adenoma and 97 controls). Results showed that both *Fn* and *m3* were significantly increased in patients with adenoma compared with control subjects, and further increased in patients with cancer, with a significant linear trend of increase from control to adenoma to cancer by multiple comparison ( $p < 0.0001$ ; figure 5A). Using the same logistic regression model established in the Hong Kong group, the combined score of the four markers (*Fn*, *m3*, *Ch* and *Bc*) also showed a significant increase during the normal-adenoma-CRC sequence ( $p < 0.0001$ ). *Fn* performed better than *m3* in discriminating CRC from controls although not significantly (AUROCs for CRC: *Fn*=0.776 vs *m3*=0.759), and the four-marker combination showed the best performance in diagnosing CRC (AUROC=0.830, sensitivity=77.5% and specificity=75.3%; figure 5B). *m3* performed better than *Fn* and the four-marker combination in distinguishing patients with adenoma from controls although not significantly (AUROCs: *m3*=0.662, *Fn*=0.616 and four-marker=0.652), and *m3* showed a sensitivity of 51.0% and specificity of 75.3% for

## A Group II - Shanghai



## B



Variables	CRC			Adenoma		
	<i>Fn</i>	<i>m3</i>	<i>LR4</i>	<i>Fn</i>	<i>m3</i>	<i>LR4</i>
<b>AUROC</b>	0.776	0.759	<b>0.830</b>	0.616	<b>0.662</b>	0.652
<b>95%CI</b>	0.705 to 0.837	0.687 to 0.822	0.765 to 0.884	0.552 to 0.678	0.599 to 0.722	0.589 to 0.712
<b>P (0.5)</b>	<0.0001	<0.0001	<0.0001	0.0014	<0.0001	<0.0001
<b>P (vs <i>m3</i>)</b>	0.7475	-	0.0470	0.3475	-	0.7821
<b>P (vs <i>Fn</i>)</b>	-	0.7475	0.0465	-	0.3475	0.1305
<b>Sensitivity</b>	59.2%	62.0%	<b>77.5%</b>	35.9%	<b>51.0%</b>	42.8%
<b>Specificity</b>	83.5%	75.3%	<b>75.3%</b>	83.5%	<b>75.3%</b>	75.3%
<b>Accuracy</b>	73.2%	69.6%	<b>76.2%</b>	55.0%	<b>60.7%</b>	55.8%

**Figure 5** Validation of bacterial markers in diagnosing colorectal cancer (CRC) and adenoma in a second independent group of faecal samples. (A) Relative faecal abundances of *Fn* and *m3* and level of the combination of *Fn*, *m3*, *Ch* and *Bc* (*LR4*) in patients with CRC and adenoma compared with control subjects of the second group. N, normal control; A, adenoma; \* $p < 0.05$  and \*\*\* $p < 0.0001$  as compared with N; # $p < 0.05$  and ### $p < 0.001$  as compared with A. (B) Comparison of ROC curves and diagnostic performances of *Fn*, *m3* and *LR4*.

adenoma (figure 5B). These results further confirm the diagnostic values of the four bacterial markers for CRC and *m3* for adenoma.

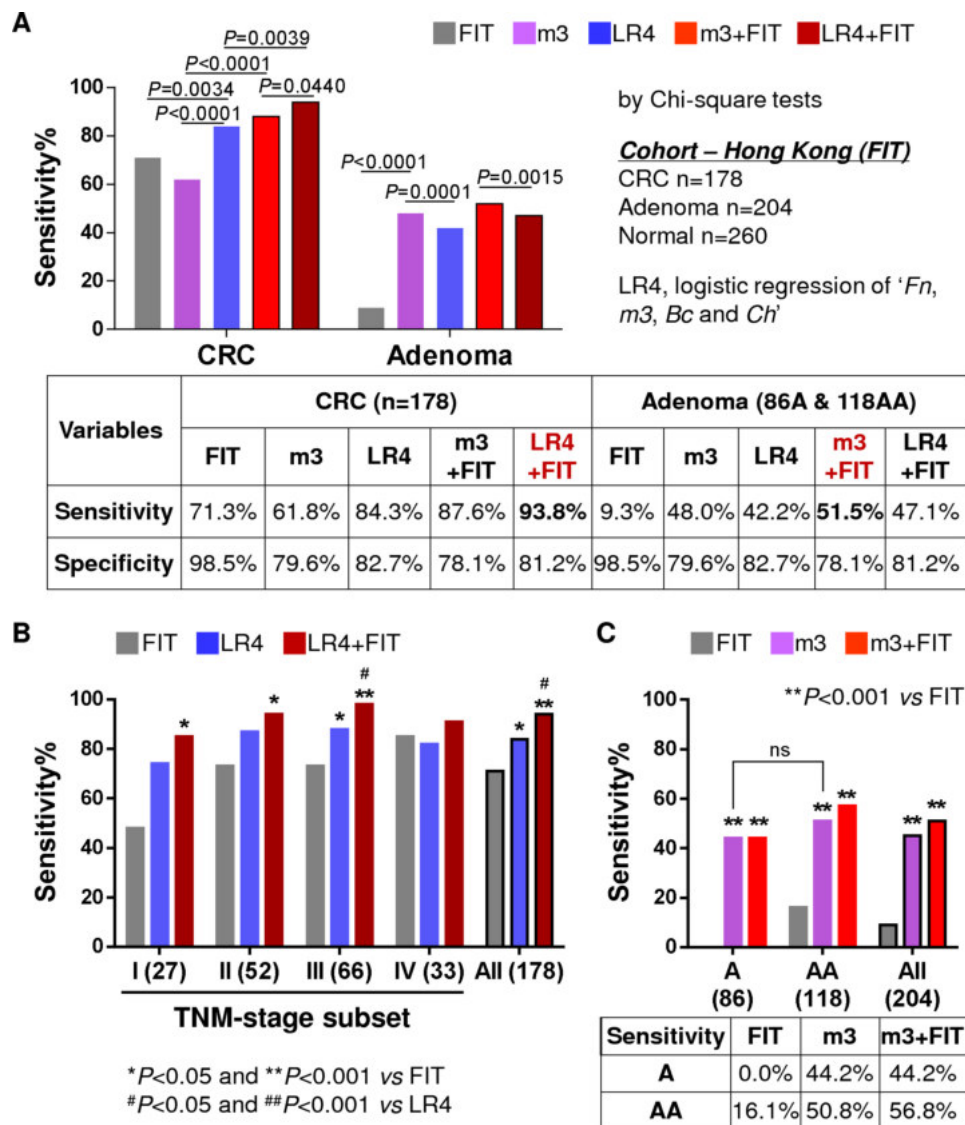
### Combination with FIT improves the diagnostic ability of bacterial markers for CRC

To compare the diagnostic performance of bacterial markers with the most widely used non-invasive stool test in CRC screening, FIT was performed in a subgroup of 642 samples from Hong Kong (178 CRC, 204 adenoma and 260 controls). The CRC detection rate by combination of *Fn*, *m3*, *Ch* and *Bc* (84.7%) was significantly higher than *m3* alone (61.8%) or FIT (71.3%) (both  $p < 0.01$ ). The combination of FIT with four-marker showed the best performance in detecting CRC, with a sensitivity of 93.8% and specificity of 81.2% (figure 6A). The four-marker showed higher sensitivities than FIT for stages I–III cancers but not late stage IV. Combination of the four-marker and FIT showed significantly increased sensitivities than FIT for stages I–III cancers, and also elevated the detection rate for stage

IV cancers (figure 6B). These results demonstrate that the bacterial marker panel is superior to FIT for detection of stages I–III CRC, and their combination further improves the non-invasive diagnosis of CRC.

### *m3* performs significantly better than FIT in diagnosing adenoma

In the subgroup of 204 adenoma (86 non-advanced and 118 advanced) cases with FIT results, *m3* alone (sensitivity=48.0%) showed a significantly higher detection rate than FIT (9.3%) and the four-marker (42.2%) (both  $p < 0.001$ ). Combination of *m3* with FIT showed the best diagnostic performance for adenoma, with a sensitivity of 51.5% and specificity of 78.1% (figure 6A). FIT failed to detect any non-advanced adenoma (0/86, 0%) and detected 16.1% (19/118) of the advanced adenoma. On the other hand, *m3* showed no difference in the detection between non-advanced and advanced adenomas (figures 2B and 6C), and the detection rates of *m3* were significantly higher for both non-advanced (44.2%) and advanced (50.8%) adenomas than FIT



**Figure 6** Comparison and combination of bacterial markers with faecal immunochemical test (FIT). (A) Comparison of sensitivity and specificity of FIT, *m3*, combination of four makers (*Fn*, *m3*, *Ch* and *Bc*; LR4) and combination of bacterial markers with FIT in a subgroup of Hong Kong samples. LR4 combined with FIT performed best for colorectal cancer (CRC) detection, while *m3* combined with FIT performed best for detecting adenoma. (B) Comparison of the sensitivities of FIT, LR4 and their combination in detecting CRC according to tumour-node-metastasis (TNM) stage subsets. (C) Comparison of the sensitivities of FIT, *m3* and their combination in detecting non-advanced and advanced adenomas. All comparison of sensitivities was conducted by  $\chi^2$  tests. A, non-advanced adenoma; AA, advanced adenoma.

(both  $p < 0.001$ ; figure 6C). Combination with FIT increased the sensitivity of *m3* for advanced adenoma to 56.8%. These results demonstrate that *m3* alone shows good performance for stool-based detection of adenoma.

## DISCUSSION

In this study, we screened a previously identified panel of CRC-associated gene markers in patients with CRC or adenoma compared with control subjects by metagenomics analysis. Focusing on the candidate gene markers that were significantly changed in patients with adenoma as compared with control subjects, we further validated their application values in non-invasive diagnosis of adenoma and CRC by qPCR. We demonstrated a faecal bacterial marker *m3* that is useful for adenoma detection and at the mean time devised a new panel of faecal bacterial markers (*m3+Fn+Ch+Bc*) to achieve improved

diagnostic capacity for CRC as compared with markers reported in our previous studies.<sup>4 13</sup>

*m3* is a well-assembled DNA sequence (1935 bp) from shotgun sequencing data. As *m3* DNA is long enough and distinct from DNA polymerase genes of other micro-organisms, with a single specific hit to *Lachnoclostridium sp. YL32* with high score (97% identify), it is reliable to conclude that the host bacterium of '*m3*' belong to the bacterial genus *Lachnoclostridium*. We further analysed the abundances of known *Lachnoclostridium* genomes in our in-house metagenomics data, with 160 strains from GenBank and 27 species from ChocoPhlAn pangenome database (online supplementary figure 2A,B). We found stepwise increase from control to adenoma to cancer in some species, such as *Clostridium (C.) aldenense*, *C. bolteae*, *C. citroniae* and *C. clostridioforme* (online supplementary figure 2C), demonstrating the potential of *Lachnoclostridium* species

in discriminating patients with colorectal neoplasm from control subjects. *Lachnospirillum* is a newly defined genus under the highly polyphyletic class *Clostridia*,<sup>18</sup> with an increasing number of new species identified from human gut microbiota in recent few years, such as *Lachnospirillum* (*L.*) *edouardi*,<sup>19</sup> *L. pacaense*<sup>20</sup> and *L. touaregense*.<sup>21</sup> The *Lachnospirillum* species carrying *m3* and its roles in colorectal tumourigenesis warrant further characterisation in future studies.

We have reported that combination of *Fn*, *Bc*, *Ch* and the undefined 'm7' showed good diagnostic performance for CRC.<sup>13</sup> Comparison of ROC curves showed that combination of 'Fn, Bc, Ch and m3' (AUROC=0.907 (0.877 to 0.931)) showed an increased AUROC than the combination of 'Fn, Bc, Ch and m7' (AUROC=0.892 (0.856 to 0.921)) in our Hong Kong group and in Shanghai group (AUROCs: 0.830 (0.765 to 0.884) with *m3* vs 0.795 (0.705 to 0.867) with *m7*). These results suggest the impact of *m3* on improving other bacterial markers for the non-invasive diagnosis of CRC.

*Fn* is prevalently detected in human CRC, with important roles in the initiation and progression of CRC. *Fn* level in colonic adenoma and adenocarcinoma tissues was found to be >10–100 times higher than normal colonic mucosa,<sup>10</sup> demonstrating *Fn* accumulation may occur at an early stage of colonic tumourigenesis. However, there is disagreement about the relationship between *Fn* and colorectal adenoma.<sup>22</sup> *Fn* was found to be enriched in cancerous versus matched normal tissues, but not significantly higher in adenoma versus normal tissues in a European cohort.<sup>23</sup> Similarly, faecal abundance of *Fn* was found to be strongly associated with CRC but not adenoma in a German cohort.<sup>24</sup> Although we observed a significant increase of faecal *Fn* in patients with adenoma compared with control subjects, the diagnostic value of *Fn* for adenoma is not as good as *m3*, and combination with *Fn* could not improve the diagnostic performance of *m3* for adenoma.

We have showed that the gram-negative bacterium *Bc*<sup>25</sup> was significantly decreased in patients with CRC as compared with healthy subjects and thus could help improve diagnostic specificity. The gram-positive bacterium *Ch*, which participates in glucose metabolism using carbohydrates as fermentable substrates to produce acetate, ethanol, carbon dioxide and hydrogen,<sup>26</sup> was significantly increased in patients with CRC compared with healthy subjects. However, faecal abundances of *Bc* and *Ch* showed no differences between patients with adenoma and control subjects.

On the other hand, *m3* is superior to other bacterial markers in discriminating patients with adenoma from control subjects according to our results from two independent Chinese groups, although its diagnostic capacity for CRC is not as good as *Fn*. FIT only detected 16.1% advanced adenoma and none of non-advanced adenoma. The multitarget stool DNA test approved by the US Food and Drug Administration, which combines mutant and methylated DNA markers and a FIT, shows sensitivities of 42.4% for advanced adenoma and 17.2% for non-advanced adenoma.<sup>27</sup> Although the sensitivity of *m3* (48.3%) is still low for adenoma, *m3* showed no significant difference in the detection between advanced and non-advanced adenomas, with sensitivities of 50.8% and 44.2%, respectively. Therefore, *m3* may outperform all other available stool-based tests in detecting non-advanced adenoma. Moreover, combination with FIT improved the detection rate of *m3* for advanced adenoma from 50.8% to 56.8%.

Some of the faecal samples were collected after colonoscopy, with 40.6%, 36.2% and 40.4% in control, adenoma and cancer groups, respectively ( $p=0.577$ ). However, these

post-colonoscopy samples were collected at least 1 month after colonoscopy when gut microbiome should have recovered to baseline.<sup>28</sup> Furthermore, we have adjusted for confounding effects of sample collection before/after colonoscopy during marker discovery.<sup>4</sup> There was no significant difference in *m3* level between pre-colonoscopy and post-colonoscopy adenoma samples by qPCR or metagenome sequencing. There were also no difference in *Fn*, *m3* or the four-marker combination between pre-colonoscopy and post-colonoscopy samples of the control, adenoma or CRC groups (online supplementary figure 3). Therefore, the markers involved in this study may not be affected by colonoscopic/bowel-prep status, given enough time for gut microbiome to recover after colonoscopy. Although age and gender differed significantly among the groups, inclusion of age and gender in the logistic regression model did not affect the ROC curves for CRC and adenoma significantly (online supplementary figure 4).

Recent studies of CRC have identified a large number of faecal microbial markers, and attempts to combine such markers from shotgun metagenomics data showed good diagnostic performance.<sup>6 29 30</sup> Our recent meta-analysis of multicohort metagenomics data, covering 526 samples from Chinese, Austrian, American, and German and French cohorts, identified seven CRC-enriched bacterial species showing an AUROC of 0.8 in discriminating patients with CRC from control subjects, which was increased to 0.88 when the clinical data were added.<sup>6</sup> Application of direct shotgun metagenomics to diagnosis is not cost-efficient due to cumbersome experimental procedure and heavy computing workload. Targeted detection of identified microbial marker candidates based on shotgun metagenomics for clinical application is a more promising strategy. Based on our bacterial gene markers identified by metagenomics investigation, quantification of four bacterial gene markers by qPCR shows an AUROC of 0.907 for CRC diagnosis in this study. However, as the true performance of the markers cannot be established from these case-control samples, future validation is required in large sample cohorts representative of the CRC screening populations. We have also reported for the first time that faecal CRC-enriched virome and mycobiome biomarkers distinguished CRC from controls with AUROCs of 0.802 and 0.93, respectively.<sup>29 30</sup> The application of these viral and fungal markers to non-invasive diagnosis of CRC by targeted quantification needs further exploration.

In conclusion, we identified a novel bacterial marker *m3*, from a *Lachnospirillum* species, for the non-invasive diagnosis of colorectal adenoma. *m3* is superior to other bacterial markers and currently available stool-based tests for adenoma detection.

## METHODS

### Metagenomic marker gene sequence analysis

Metagenomic sequencing data from 589 Hong Kong Chinese subjects (184 CRC, 185 adenoma and 220 control subjects) from our previous study were analysed,<sup>29</sup> which included the discovery cohort of 74 CRC and 54 controls for the identification of the 20 CRC-related markers.<sup>4</sup> Raw faecal shotgun metagenomic sequences were quality-trimmed and decontaminated as described previously.<sup>29</sup> Low complexity subsequences of bacterial genes were hard-masked with the DUST program and indexed using the Burrows-Wheeler Aligner (BWA; V.0.7.17) to create the gene database for short read alignment.<sup>31 32</sup> Post-quality control sequences in FASTQ format were mapped against the BWA database with maximal exact match (mem) algorithm and default parameters of penalty scoring. Histograms of aligned

sequence coverage were reported using the ‘genomecov’ module of BEDTools suite (V.2.27.0).<sup>33</sup> Mean sequence coverage table of metagenomic samples was constructed by computing summed products of coverage depth and base-pair fraction of marker gene length for positional features in input BAM files. Multiple group comparison of clinical phenotype was performed by pairwise Wilcoxon’s rank-sum tests, and *p* values were corrected by Benjamini-Hochberg step-up procedure. We then derived average weighted contribution (AWC) scores to estimate differential genomic enrichment and depletion of *Lachnospirillum* species using marker gene sequences originating from the ChocoPhlAn pangenome database (V.293)<sup>34</sup> as well as Prokka-annotated protein coding gene sequences representing 160 *Lachnospirillum* genomes at all assembly levels from the NCBI GenBank (release 234.0; accessed 16 Oct 2019).<sup>35</sup> The AWC of species *i* with gene set *j* to phenotype *k* was computed as follows:

$$AWC_{ijk} = \frac{\sum_{i \in k} NE_{ij} - ND_{ij}}{N_{ij}^2},$$

where  $NE_{ij}$  (or  $ND_{ij}$ ) is the total count of significant enrichment (or depletion) of a genomic sequence in gene set *j* for species *i* in a one-versus-all comparative statistical analysis of clinical phenotype *k* at 5% false discovery rate, respectively.  $N_{ij}$  denotes the number of gene sequences of species *i*.

### Human faecal sample collection

Faecal samples (n=1012) were collected from two independent groups of subjects, including group I—Hong Kong (698 subjects: 203 CRC, 207 adenoma and 288 normal controls) at the Prince of Wales Hospital, the Chinese University of Hong Kong between 2009 and 2014 and group II—Shanghai (313 subjects: 71 CRC, 145 adenoma and 97 normal controls) at Renji Hospital, Shanghai Jiaotong University between 2014 and 2018 (detailed clinical characteristics in online supplementary table S2). Subjects recruited for faecal sample collection included individuals presenting symptoms such as change of bowel habit, rectal bleeding, abdominal pain or anaemia, and asymptomatic individuals aged 50 or above undergoing screening colonoscopy as in our previous metagenomic study.<sup>4</sup> Samples were collected before or 1 month after colonoscopy, when gut microbiome should have recovered to baseline.<sup>28</sup> The exclusion criteria were (1) use of antibiotics within the past 3 months, (2) on a vegetarian diet, (3) had an invasive medical intervention within the past 3 months and (4) had a history of any cancer or inflammatory disease of the intestine. Subjects were asked to collect stool samples in standardised containers at home and store the samples in their home freezer at  $-20^{\circ}\text{C}$  immediately. Frozen samples were then delivered to the hospitals in insulating polystyrene foam containers and stored at  $-80^{\circ}\text{C}$  immediately until further analysis. Patients were diagnosed by colonoscopic examination and histopathological review of any biopsies taken.

### DNA extraction, design of primers and probes and qPCR

DNA extraction, design of primer and probe sequences and qPCR amplifications on an ABI QuantStudio sequence detection system were conducted as our previous description.<sup>13</sup> Primer and probe sequences specifically targeting *m3* are as following: forward 5'-AATGGGAATGGAGCGGATTC-3'; reverse 5'-CCTGCACCAGCTTATCGTCAA-3'; probe 5'-AAGCCTGCGGAACCACAGTTACCAGC-3'. Primer and probe sequences targeting other bacterial gene markers and 16s rDNA internal control are as in our previous study.<sup>13</sup> Each probe carried a 5' reporter dye FAM (6-carboxy fluorescein)

or VIC (4,7,2'-trichloro-7'-phenyl-6-carboxyfluorescein) and a 3' quencher dye TAMRA (6-carboxytetramethyl-rhodamine). Primers and hydrolysis probes were synthesised by Invitrogen (Carlsbad, CA). PCR amplification specificity was confirmed by direct Sanger sequencing of the PCR products or by sequencing randomly picked TA clones. Relative abundance of each marker was calculated by using delta Cq method as compared with internal control and shown as Log value of  $*10e6+1$ .

### Faecal immunochemical test

A subgroup of Hong Kong samples (n=642; 178 CRC, 118 advanced adenoma, 86 non-advanced adenoma and 260 control subjects) were examined by FIT using the automated quantitative OC-Sensor test (Eiken Chemical, Japan). The quantitative OC-Sensor test was performed as our previous description,<sup>36</sup> with a positive cut-off value equivalent to a concentration of 100 ng of haemoglobin per millilitre.

### Statistical analyses

Values were all expressed as mean  $\pm$  SD or median (IQR) as appropriate. The differences in bacterial abundances were determined by Mann-Whitney U test. One-way ANOVA multiple comparison with test for linear trend was used to evaluate the changes of marker levels during disease progression (from control to adenoma to cancer). Simple and multiple regression analyses were used to estimate the association between marker levels and factors of interest. Occurrence rates between different groups and sensitivities by different markers were analysed using the  $\chi^2$  test. Combination of multiple biomarkers was performed by applying logistic regression model to obtain values for estimating the incidence of CRC as compared with controls. The scores of the combination of four markers were calculated as follows:  $LR4 = \text{Power}(2, (\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4))$ , where  $\alpha$  represented the intercept,  $\beta$  represented the regression coefficients and  $X$  represented the levels of the corresponding markers. ROC curves were used to evaluate the diagnostic value of bacterial markers/models in distinguishing CRC/adenoma and controls. Pairwise comparison of ROC curves was performed using a non-parametric approach.<sup>37</sup> The best cut-off values were determined by ROC analyses that maximised the Youden index ( $J = \text{Sensitivity} + \text{Specificity} - 1$ ).<sup>38</sup> All tests were done by GraphPad Prism V.5.0 (GraphPad Software, San Diego, CA) or MedCalc Statistical Software V.18.5 (MedCalc Software bvba, Ostend, Belgium; <http://www.medcalc.org>; 2018). A *p* value  $< 0.05$  was taken as statistical significance.

**Contributors** Study conception and design: QL, JY. Development of methodology: QL. Acquisition of data: TL, Y-XC, TOY, GN, EC, SW, SCN, FKLC, Y-YF. Analysis and interpretation of data: QL. Writing, review and/or revision of the manuscript: QL, FKLC, JJYS, JY. Administrative, technical or material support: QL, JY. Study supervision: QL, JJYS, JY.

**Funding** This project was supported by HRMF research fellowship scheme (02160037), China MOST fund (2016YFC1303200), Science and Technology Program Grant Shenzhen (JCYJ20170413161534162), National Key R&D Program of China (2017YFE0190700, 2018YFC135000, 2018YFC1315004), National Natural Science Foundation of China (81773000) and Shenzhen Virtual University Park Support Scheme to CUHK Shenzhen Research Institute.

**Competing interests** None declared.

**Patient consent for publication** Obtained.

**Ethics approval** The study was approved by the Clinical Research Ethics Committee of the Chinese University of Hong Kong and the Ethics Committee of Renji Hospital, Shanghai Jiaotong University.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as online supplementary information.



**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Sunny Wong <http://orcid.org/0000-0002-3354-9310>

Siew C Ng <http://orcid.org/0000-0002-6850-4454>

Francis K L Chan <http://orcid.org/0000-0001-7388-2436>

Jun Yu <http://orcid.org/0000-0001-9239-2416>

#### REFERENCES

- Allemani C, Matsuda T, Di Carlo V, *et al*. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 2018;391:1023–75.
- The Lancet. GLOBOCAN 2018: counting the toll of cancer. *Lancet* 2018;392:985.
- Irrazábal T, Belcheva A, Girardin SE, *et al*. The multifaceted role of the intestinal microbiota in colon cancer. *Mol Cell* 2014;54:309–20.
- Yu J, Feng Q, Wong SH, *et al*. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 2017;66:70–8.
- Nakatsu G, Li X, Zhou H, *et al*. Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun* 2015;6:8727.
- Dai Z, Coker OO, Nakatsu G, *et al*. Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome* 2018;6.
- Tilg H, Adolph TE, Gerner RR, *et al*. The intestinal microbiota in colorectal cancer. *Cancer Cell* 2018;33:954–64.
- Wong SH, Zhao L, Zhang X, *et al*. Gavage of fecal samples from patients with colorectal cancer promotes intestinal carcinogenesis in germ-free and conventional mice. *Gastroenterology* 2017;153:1621–33.
- Kostic AD, Chun E, Robertson L, *et al*. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* 2013;14:207–15.
- Rubinstein MR, Wang X, Liu W, *et al*. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/ $\beta$ -catenin signaling via its FadA adhesin. *Cell Host Microbe* 2013;14:195–206.
- Yu T, Guo F, Yu Y, *et al*. *Fusobacterium nucleatum* promotes chemoresistance to colorectal cancer by modulating autophagy. *Cell* 2017;170:548–63. e16.
- Tsoi H, Chu ESH, Zhang X, *et al*. Peptostreptococcus anaerobius induces intracellular cholesterol biosynthesis in colon cells to induce proliferation and causes dysplasia in mice. *Gastroenterology* 2017;152:1419–33.
- Liang Q, Chiu J, Chen Y, *et al*. Fecal bacteria act as novel biomarkers for noninvasive diagnosis of colorectal cancer. *Clin Cancer Res* 2017;23:2061–70.
- Xie Y-H, Gao Q-Y, Cai G-X, *et al*. Fecal *Clostridium symbiosum* for noninvasive detection of early and advanced colorectal cancer: test and validation studies. *EBioMedicine* 2017;25:32–40.
- Shah MS, DeSantis TZ, Weinmaier T, *et al*. Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer. *Gut* 2018;67:882–91.
- Lee JK, Liles EG, Bent S, *et al*. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. *Ann Intern Med* 2014;160:171.
- Robertson DJ, Lee JK, Boland CR, *et al*. Recommendations on fecal immunochemical testing to screen for colorectal neoplasia: a consensus statement by the US Multi-Society Task force on colorectal cancer. *Gastroenterology* 2017;152:1217–37.
- Yutin N, Galperin MY. A genomic update on clostridial phylogeny: gram-negative spore formers and other misplaced clostridia. *Environ Microbiol* 2013;140:2631–41.
- Traore SI, Azhar EI, Yasir M, *et al*. Description of '*Blautia phocaeensis*' sp. nov. and '*Lachnoclostridium edouardi*' sp. nov., isolated from healthy fresh stools of Saudi Arabia Bedouins by culturomics. *New Microbes New Infect* 2017;19:129–31.
- Pham T-P-T, Cadoret F, Alou MT, *et al*. '*Urmitella timonensis*' gen. nov., sp. nov., '*Blautia marasmii*' sp. nov., '*Lachnoclostridium pacaense*' sp. nov., '*Bacillus marasmii*' sp. nov. and '*Anaerotruncus rubiinfantis*' sp. nov., isolated from stool samples of undernourished African children. *New Microbes New Infect* 2017;17:84–8.
- Tidjani Alou M, Khelaifia S, La Scola B, *et al*. "*Lachnoclostridium touaregense*," a new bacterial species isolated from the human gut microbiota. *New Microbes New Infect* 2016;14:51–2.
- Zhang S, Cai S, Ma Y. Association between *Fusobacterium nucleatum* and colorectal cancer: progress and future directions. *J Cancer* 2018;9:1652–9.
- Flanagan L, Schmid J, Ebert M, *et al*. *Fusobacterium nucleatum* associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. *Eur J Clin Microbiol Infect Dis* 2014;33:1381–90.
- Amitay EL, Werner S, Vital M, *et al*. *Fusobacterium* and colorectal cancer: causal factor or passenger? Results from a large colorectal cancer screening study. *Carcinogenesis* 2017;38:781–8.
- Watanabe Y, Nagai F, Morotomi M, *et al*. *Bacteroides clarus* sp. nov., *Bacteroides fluxus* sp. nov. and *Bacteroides oleiciplenus* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* 2010;60:1864–9.
- Steer T, Collins MD, Gibson GR, *et al*. *Clostridium hathewayi* sp. nov., from human faeces. *Syst Appl Microbiol* 2001;24:353–7.
- Imperiale TF, Ransohoff DF, Itzkowitz SH, *et al*. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med* 2014;370:1287–97.
- Jalanka J, Salonen A, Salojärvi J, *et al*. Effects of bowel cleansing on the intestinal microbiota. *Gut* 2015;64:1562–8.
- Nakatsu G, Zhou H, Wu WKK, *et al*. Alterations in enteric virome are associated with colorectal cancer and survival outcomes. *Gastroenterology* 2018;155:529–41.
- Coker OO, Nakatsu G, Dai RZ, *et al*. Enteric fungal microbiota dysbiosis and ecological alterations in colorectal cancer. *Gut* 2019;68:654–62.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- Morgulis A, Gertz EM, Schäffer AA, *et al*. A fast and symmetric dust implementation to mask low-complexity DNA sequences. *J Comput Biol* 2006;13:1028–40.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2.
- Franzosa EA, McIver LJ, Rahnavard G, *et al*. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* 2018;15:962–8.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–9.
- Wong SH, Kwong TNY, Chow T-C, *et al*. Quantitation of faecal *Fusobacterium* improves faecal immunochemical test in detecting advanced colorectal neoplasia. *Gut* 2017;66:1441–8.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.
- Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.

# Multimomics Profiling Reveals Signatures of Dysmetabolism in Urban Populations in Central India

Tanya M. Monaghan, Rima N. Biswas, Rupam R. Nashine, Samidha S. Joshi, Benjamin H. Mullish, Anna M. Seekatz, Jesus Miguens Blanco, Julie A. K. McDonald, Julian R. Marchesi, **Tung On Yau**, Niki Christodoulou, Maria Hatziapostolou, Maja Pucic-Bakovic, Frano Vuckovic, Filip Klicek, Gordan Lauc, Ning Xue, Tania Dottorini, Shrikant Ambalkar, Ashish Satav, Christos Polytarchou, Animesh Acharjee, Rajpal Singh Kashyap

*Microorganisms*. **2021**;9(7):1485.

Journal URL: [mdpi.com/2076-2607/9/7/1485](https://mdpi.com/2076-2607/9/7/1485)

DOI: [10.3390/microorganisms9071485](https://doi.org/10.3390/microorganisms9071485)

PMID: [34361920](https://pubmed.ncbi.nlm.nih.gov/34361920/)

PMCID: [PMC8307859](https://pubmed.ncbi.nlm.nih.gov/pmc/articles/PMC8307859/)

31<sup>st</sup> July, 2021

To Whom It May Concern,

**Statement of Joint Authorship**

**Title of publication:** Multiomics profiling reveals signatures of dysmetabolism in urban populations in Central India

**Journal:** Microorganisms

**Publication date:** July, 2021

**Issue:**

**PubMed ID:**

**Authors:** Tanya M Monaghan\*, Rima Biswas, Rupam Nashine, Samidha S Joshi, Benjamin H Mullish, Anna M Seekatz, Jesus Miguens Blanco, Julie AK McDonald, Julian Marchesi, **Tung On Yau**, Niki Christodoulou, Maria Hatziapostolou, Maja Pučić-Baković, Frano Vučković, Filip Klicek, Gordan Lauc, Ning Xue, Tania Dottorini, Shrikant Ambalkar, Ashish Satav, Christos Polytarchou\*, Animesh Acharjee\*, Rajpal Singh Kashyap\*

\* co-corresponding authors

I hereby confirm that Mr. Tung On YAU is a co-author in the above publication. He was a significant contributor to this project, carrying out experiments, data acquisition and collection. This included performing Bio-Plex Pro Human Inflammation Panel 1 assay and Bio-Plex Pro™ Human diabetes 10-plex immunoassay on sera samples using a Bio-Plex 200 System.

Yours sincerely,



Christos Polytarchou, PhD  
Associate Professor  
John van Geest Cancer Research Centre  
School of Science and Technology  
Nottingham Trent University  
Nottingham  
NG11 8NS  
Email: [christos.polytarchou@ntu.ac.uk](mailto:christos.polytarchou@ntu.ac.uk)



**University of  
Nottingham**  
UK | CHINA | MALAYSIA

Dr Tanya Monaghan  
Faculty of Medicine & Health Sciences  
Room W/E 1381 NDDC  
Queen's Medical Centre  
Nottingham  
NG7 2UH  
UK

[Tanya.Monaghan@nottingham.ac.uk](mailto:Tanya.Monaghan@nottingham.ac.uk)  
+44(0)115 9249924 x60589

April 12<sup>th</sup>, 2022

Re: Full Publications

To Whom It May Concern:

I am writing to confirm that Tung On (Payton) Yau, previously a Research Fellow in the laboratory of Dr Christos Polytarchou at Nottingham Trent University (currently serving as an Hourly-Paid Lecturer at Nottingham Trent University and a Post-doctoral Researcher at Scotland's Rural College), collaborated with my team on projects relating to microRNAs, microbiota, public health and *Clostridioides difficile* infection from the year 2017 to 2021. He was a significant contributor, with a role in literature search and acquisition, experimental design, clinical sample management, and performing experiments. These experiments include cell culture, RNA extraction from serum and peripheral blood mononuclear cells, high-throughput Nanostring miRNA analysis, Bio-Plex cytokine assays, and luciferase reporter assays for the identification of miRNA targets. He was subsequently involved in data analysis and visual representations of the data for publications.

He is a co-author of three published publications:

1. Fecal microbiota transplantation for recurrent *Clostridioides difficile* infection associates with functional alterations in circulating microRNAs. **Gastroenterology** **161** (1), 255-270
2. Multiomics profiling reveals signatures of dysmetabolism in urban populations in Central India. **Microorganisms** **9** (7), 1485
3. A multi-factorial observational study on sequential fecal microbiota transplant in patients with medically refractory *Clostridioides difficile* infection. **Cells** **10** (11), 3234

Yours Sincerely,



Article

# Multimomics Profiling Reveals Signatures of Dysmetabolism in Urban Populations in Central India

Tanya M. Monaghan <sup>1,2,\*</sup>, Rima N. Biswas <sup>3</sup>, Rupam R. Nashine <sup>3</sup>, Samidha S. Joshi <sup>3</sup>, Benjamin H. Mullish <sup>4</sup>, Anna M. Seekatz <sup>5</sup>, Jesus Miguens Blanco <sup>4</sup>, Julie A. K. McDonald <sup>4,6</sup>, Julian R. Marchesi <sup>4</sup>, Tung on Yau <sup>7</sup>, Niki Christodoulou <sup>7</sup>, Maria Hatzia Apostolou <sup>7</sup>, Maja Pucic-Bakovic <sup>8</sup>, Frano Vuckovic <sup>8</sup>, Filip Klicek <sup>8</sup>, Gordan Lauc <sup>8,9</sup>, Ning Xue <sup>10</sup>, Tania Dottorini <sup>10</sup>, Shrikant Ambalkar <sup>11</sup>, Ashish Satav <sup>12</sup>, Christos Polytarchou <sup>7,\*</sup>, Animesh Acharjee <sup>13,14,15,\*</sup> and Rajpal Singh Kashyap <sup>3,\*</sup>

- <sup>1</sup> NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham NG7 2UH, UK
- <sup>2</sup> Nottingham Digestive Diseases Centre, School of Medicine, University of Nottingham, Nottingham NG7 2UH, UK
- <sup>3</sup> Biochemistry Research Laboratory, Dr. G.M. Taori Central India Institute of Medical Sciences, Nagpur 440010, India; rimabiswas13@gmail.com (R.N.B.); rpmnashine@gmail.com (R.R.N.); sjoshi.res@gmail.com (S.S.J.)
- <sup>4</sup> Division of Digestive Diseases, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London SW7 2AZ, UK; b.mullish@imperial.ac.uk (B.H.M.); j.miguens-blanco18@imperial.ac.uk (J.M.B.); julie.mcdonald@imperial.ac.uk (J.A.K.M.); j.marchesi@imperial.ac.uk (J.R.M.)
- <sup>5</sup> Department of Biological Sciences, Clemson University, Clemson, SC 29631, USA; aseekatz@clemson.edu
- <sup>6</sup> MRC Centre for Molecular Bacteriology and Infection, Imperial College London, London SW7 2AZ, UK
- <sup>7</sup> Department of Biosciences, John van Geest Cancer Research Centre, Centre for Health Aging and Understanding Disease, School of Science and Technology, Nottingham Trent University, Nottingham NG7 2UH, UK; payton.yau@ntu.ac.uk (T.o.Y.); niki.christodoulou2019@my.ntu.ac.uk (N.C.); maria.hatzia Apostolou@ntu.ac.uk (M.H.)
- <sup>8</sup> Glycoscience Research Laboratory, Genos Ltd., Borongajska cesta 83H, 10000 Zagreb, Croatia; mpucicbakovic@genos.hr (M.P.-B.); fvuckovic@genos.hr (F.V.); fklicek@genos.hr (F.K.); glauc@genos.hr (G.L.)
- <sup>9</sup> Faculty of Pharmacy and Biochemistry, University of Zagreb, 10000 Zagreb, Croatia
- <sup>10</sup> School of Veterinary Medicine and Science, University of Nottingham, Nottingham NG7 2UH, UK; ning.xue@microlise.com (N.X.); tania.Dottorini@nottingham.ac.uk (T.D.)
- <sup>11</sup> Department of Microbiology and Infection, King’s Mill Hospital, Sherwood Forest Hospitals NHS Trust, Sutton in Ashfield NG17 4JL, UK; shrikant.ambalkar@nhs.net
- <sup>12</sup> Mahatma Gandhi Tribal Hospital, MAHAN Trust Melghat, Amravati 605006, India; drashish@mahantrust.org
- <sup>13</sup> Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, UK
- <sup>14</sup> Institute of Translational Medicine, University Hospitals Birmingham, Foundation Trust, Birmingham B15 2TT, UK
- <sup>15</sup> NIHR Surgical Reconstruction and Microbiology Research Centre, University Hospital Birmingham, Birmingham B15 2WB, UK
- \* Correspondence: tanya.monaghan@nottingham.ac.uk or tanyamonaghan@gmail.com (T.M.M.); Christos.polytarchou@ntu.ac.uk (C.P.); a.acharjee@bham.ac.uk (A.A.); rajpalsingh.kashyap@gmail.com (R.S.K.)



**Citation:** Monaghan, T.M.; Biswas, R.N.; Nashine, R.R.; Joshi, S.S.; Mullish, B.H.; Seekatz, A.M.; Blanco, J.M.; McDonald, J.A.K.; Marchesi, J.R.; Yau, T.o.; et al. Multimomics Profiling Reveals Signatures of Dysmetabolism in Urban Populations in Central India. *Microorganisms* **2021**, *9*, 1485. <https://doi.org/10.3390/microorganisms9071485>

Academic Editor: Fabio Pace

Received: 8 May 2021

Accepted: 7 July 2021

Published: 12 July 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Background: Non-communicable diseases (NCDs) have become a major cause of morbidity and mortality in India. Perturbation of host–microbiome interactions may be a key mechanism by which lifestyle-related risk factors such as tobacco use, alcohol consumption, and physical inactivity may influence metabolic health. There is an urgent need to identify relevant dysmetabolic traits for predicting risk of metabolic disorders, such as diabetes, among susceptible Asian Indians where NCDs are a growing epidemic. Methods: Here, we report the first in-depth phenotypic study in which we prospectively enrolled 218 adults from urban and rural areas of Central India and used multimomic profiling to identify relationships between microbial taxa and circulating biomarkers of cardiometabolic risk. Assays included fecal microbiota analysis by 16S ribosomal RNA gene amplicon sequencing, quantification of serum short chain fatty acids by gas chromatography-mass spectrometry, and multiplex assaying of serum diabetic proteins, cytokines, chemokines, and multi-isotype antibodies. Sera was also analysed for N-glycans and immunoglobulin G Fc N-glycopeptides. Results: Multiple hallmarks of dysmetabolism were identified in urbanites and young overweight

adults, the majority of whom did not have a known diagnosis of diabetes. Association analyses revealed several host–microbe and metabolic associations. Conclusions: Host–microbe and metabolic interactions are differentially shaped by body weight and geographic status in Central Indians. Further exploration of these links may help create a molecular-level map for estimating risk of developing metabolic disorders and designing early interventions.

**Keywords:** geography; host–microbe interactions; glycome; dysmetabolism; multiomics; diabetes mellitus

## 1. Introduction

Whilst communicable diseases caused by infectious microbes continue to exert a significant public health burden in India, existing evidence now indicates a marked shift to non-communicable diseases (NCDs) [1–4]. The consumption of Western-type energy-intensive, nutrient poor, high glycaemic index carbohydrate enriched diets, increasingly sedentary occupations, and low levels of recreational activity, particularly in urbanised populations, all lead to a higher body mass index (BMI), evoking a state of chronic metabolic inflammation, termed metainflammation [5,6]. Metainflammation contributes to the development of many NCDs, including diabetes, which has increased rapidly in India over the last quarter of a century, rising from 26 million prevalent cases in 1990 to 65 million in 2016 [7]. The 9th Edition of the IDF Diabetes Atlas in 2019 reported that India is currently home to 77 million diabetics and this number is projected to soar to 134 million cases in the next 25 years. Asian Indians have one of the highest rates of diabetes among major ethnic groups, and the progression from prediabetes to diabetes appears to occur faster in this population [8]. According to the National Urban Diabetes Survey, the estimated prevalence of prediabetes is 14 per cent in India [9]. A more recent study reported that 6 in 10 adults in large South Asian cities have either diabetes or prediabetes [10]. Concerningly, an Indian multistate study has reported that a high percentage of the diabetes cases in the Indian population remain undiagnosed, highlighting issues of poor awareness and detection of diabetes [11]. An important epidemiologic aim going forward will be to identify at-risk individuals, to facilitate an early therapeutic impact.

New multi-biomarker approaches which detect dysmetabolic traits are urgently being sought to predict risk of metabolic diseases such as diabetes and its complications [12]. In this regard, emerging evidence leads us to conclude that metabolic syndrome, which often accompanies obesity and hyperglycaemia, also leads to increased risk of enteric and systemic infections [13]. A recent study suggested that this increased risk may be due to hyperglycaemia, either genetically, chemically or diet-induced, rather than obesity itself, which provides the mechanistic basis for intestinal barrier dysfunction [14].

Alterations in the gut microbiome, metabonome, immune system and, more recently, the total serum and IgG *N*-glycome have been separately described in various dysmetabolic states, with a predominant focus on humans residing in developed nations [15–24]. However, it remains unclear how these molecular signatures interact, and whether such interactions can offer novel pathophysiological insight into the earliest stages of a dysregulated metabolism that is often associated with an elevated BMI and insulin resistance (IR) state.

To explore this gap in knowledge, we used a multiomics strategy to deeply phenotype rural and urban populations in Central India, unbiasedly sampled in terms of their metabolic state; this unbiased approach allows us to gauge a ‘real world’ cohort without systematically favouring certain populations (e.g., those with metabolic syndrome) over others. We report the first association study investigating the interplay between the circulating immune-metabolic proteome, metabonome, glycome and gut microbiome in previously poorly phenotyped Central Indians. Notably, we associate urban living with multiple hallmarks of metabolic dysregulation, a critical precursor to metabolic disease.

## 2. Materials and Methods

### 2.1. Participant Recruitment

In this observational cohort study, which was carried out during 2019, we prospectively recruited adult ( $\geq 18$  years of age) participants from both in- and outpatient settings in urban and rural settings of Central India. Health records were also reviewed for each participant where available. Basic demographic details including age, gender, geographic location, as well as information on hospitalisation exposure, antibiotic usage during and before (within 3 months) study recruitment, antacids usage, smoking status, co-morbidities, toilet access, use of hand soap, and presence of domestic animals were recorded for rural and urban participants. Body mass index (BMI) was also recorded for all participants. BMI ranges were pre-defined using WHO Asian BMI classifications: underweight ( $< 18.5$ ), normal ( $18.5\text{--}22.9$ ), overweight ( $23\text{--}24.9$ ), pre-obese ( $25\text{--}29.9$ ) and obese ( $\geq 30$ ) categories.

Site-specific project coordinators based at the Central Indian Institute of Medical Sciences (CIIMS), Nagpur, and MAHAN Trust, Melghat, supervised recruitment of urban and rural participants and sample acquisition across 25 urban and 35 rural sampling sites, respectively (Table S1). Nagpur is India's 13th largest city by population (2.5M) and is located at the exact centre of the Indian peninsula. Project fellows approached all consecutive in- and outpatients at CIIMS and also processed samples received from other participating hospitals or private clinical laboratories within a 20 km radius of Nagpur as well as within rural Melghat, Amravati district.

In MAHAN Trust in rural Melghat, which is located approximately 293 km from Nagpur, and is home to a community of 250,000 members of the Korku tribe, all participants were directly recruited by community village healthcare workers and counsellors trained by MAHAN Trust who liaised closely with project fellows from CIIMS. Stool samples were also collected from the rural extensions within a 50 km radius from the satellite centre at MAHAN Trust, Melghat. Here, patient recruitment and sample acquisition were facilitated by village healthcare workers and counsellors working in the subdivisional hospitals (SDH) and public health centres. The counsellors then contacted the research fellows at the rural satellite hospital in MAHAN Trust.

In contrast to the emerging metropolis of Nagpur, which was declared open defecation free in 2018, and is one of the cleanest and most liveable cities in India, rural agriculturalist communities within Melghat and its rural extension zones are of lower socioeconomic status, display high rates of illiteracy and malnutrition, and possess poor access to medical and educational facilities. Their small hut dwellings are typically composed of mud, grass and bamboo frames which lack an electricity or running water supply or proper sanitation systems. They live in close proximity to their animals (chickens, goats, pigs, cows, buffalo), often in the same one-room dwelling.

### 2.2. Inclusion and Exclusion Criteria

During participant selection, inclusion criteria were (i) adults aged 18 to 70 years of age who could provide written or thumb-print consent, (ii) HIV, hepatitis B or C negative, and (iii) not pregnant or breastfeeding. Participants who were immunosuppressed were not excluded. Immunosuppression was defined as those with cancer, receiving chemotherapy or on prednisolone ( $> 5$  mg/d), immunomodulators (azathioprine, methotrexate, calcineurin inhibitor) or biologics. We excluded subjects that were unable to provide a stool sample.

### 2.3. Ethics Statement

This study was approved by the Faculty of Medicine and Health Sciences Research Ethics Committee at the University of Nottingham (REC no. 199-1901) and the Ethical Committee of the Central India Institute of Medical Sciences, Nagpur. All subjects provided verbal and written (or thumbprint) consent.

#### 2.4. Sample Preparation

All clinical samples were anonymised and assigned a study code number linked to participant demographic details. Up to two faecal samples (3–5 g each) were collected in UV sterilised dry plastic containers at the time of recruitment from each participant and placed in a cool box. As per the standard operating procedures, all stool specimens were stored at 4 °C immediately after collection to avoid enzymatic degradation prior to genomic DNA extraction which was performed within 24 hours of sample collection. Whole blood samples were drawn from all participants into vacutainer tubes with EDTA as anticoagulant. These were centrifuged for 10 min at 2400× *g* within 30 min of being taken. Serum was then carefully aspirated at room temperature and aliquoted accordingly into single-use cryotubes to avoid repeated freeze–thaw cycles prior to sample storage at –20 °C.

#### 2.5. Gut Bacterial Community Profiling by 16S rRNA Gene Sequencing

Stool samples were randomised for processing and DNA was extracted from 1–1.5 g of faeces and homogenised in lysis buffer (Tris HCl, EDTA, NaCl and SDS) using phenol-chloroform method. Briefly, the content was centrifuged at 7000× *g* for 10 min. The supernatant was then transferred to a 1.5 mL tube containing a mixture of isopropanol and sodium acetate (5M) and incubated at –20 °C for 30 min. Following removal of the supernatant the pellet was dried for about an hour. The pellet was suspended in 1X Tris EDTA buffer (pH 8) and incubated at 65 °C for 15 min. An approximate equal volume (0.5–0.7 mL) of phenol: chloroform-isoamyl alcohol (24:1) was added, mixed thoroughly and centrifuged for 10 min at 12,000× *g*. The aqueous viscous supernatant was carefully transferred to a new 1.5 mL tube. An equal volume of chloroform-isoamyl alcohol (1:1) was added, followed by centrifugation for 10 min at 12,000× *g*. The supernatant was mixed with 0.6× volume of isopropanol to aid precipitation. The precipitated nucleic acids were washed with 75% ethanol, dried and re-suspended in 50µL of TE buffer.

Extracted DNA was quantified using a Qubit 2.0 Fluorometer (ThermoFischer Scientific, Hemel Hempstead, UK), and stored at –80 °C pending downstream assays. Gene-sequencing sample libraries for 16S rRNA were generated via Illumina’s 16S Metagenomic Sequencing Library Preparation Protocol, but with some modifications. Amplification was performed of the V1-V2 16S rRNA gene regions from the faecal DNA, using primers as previously described [25]. Products from the index PCR reactions were cleaned and normalised via the SequalPrep Normalization Plate Kit (Life Technologies, Carlsbad, CA, USA), and library quantification was performed using the NEBNext Library Quant Kit for Illumina (New England Biolabs, Ipswich, MA, USA). Sequencing data were obtained using paired-end 300 bp chemistry on an Illumina MiSeq (Illumina Inc, San Diego, CA, USA), with MiSeq Reagent Kit Volume 3 (Illumina Inc). Sequenced libraries included both negative controls (PCR grade water, Roche, Basel, Switzerland) and positive controls, with the latter using a mock community of 10 bacterial strains (LGC Group, Teddington, UK). Processing of sequencing data was performed via the DADA2 pipeline as previously described [26], using the SILVA bacterial database Volume 132 (<https://www.arb-silva.de/> (accessed on 15 February 2021)).

A combination of R packages was used to analyse and visualise microbiota relative abundance data. The inverse Simpson index, non-metric multidimensional scaling (NMDS) and Analysis of Similarities (ANOSIM) were implemented in the R package ‘vegan’ [27], using the Bray–Curtis distance metric based on normalized ASV counts. Partitioning Around Medoids (PAM) clustering [28] on the Jensen–Shannon divergence calculated from normalised ASV counts was used to identify two optimal community types, as defined by best-fit silhouette score (mean silhouette score = 0.47). Linear discriminant analysis Effect Size (LEfSe) [29] as implemented in mothur [30] was used to identify differentially abundant genera in urban vs. rural, or high ( $\geq 23$ ) vs. low/normal ( $< 23$ ) BMI score. Kruskal–Wallis and Pearson’s chi-squared tests were run in standard R.



## 2.6. Serum Short Chain Fatty Acid Identification and Quantification

This was performed using a targeted gas chromatography-mass spectrometry protocol, as previously described [31]. Sample analysis was performed on an Agilent 7890B GC system coupled to an Agilent 5977A mass selective detector (Agilent, Santa Clara, CA, USA). Patient samples were run alongside negative controls and quality control samples (pooled aliquots of all patient samples; one run after every ten patient samples) to ensure no source contamination and to assess for signal drift. Three injections were undertaken for each sample. Analysis of data was performed using MassHunter software (Agilent), with SCFA levels calculated via integration of spectra from patient samples and comparison with freshly prepared calibration curves using SCFA standards (Merck, Darmstadt, Germany).

## 2.7. Serum N-Glycome Profiling

### 2.7.1. Experimental Design

Participant serum samples and in-house serum standards were thawed, vortexed and centrifuged for 3 min at 12,100 g. Each sample (100 µL) was aliquoted to 2 mL 96-well collection plates (Waters, Milford, MA, USA) following a predetermined, established experimental design [32] which included blocking of all known sources of variation (age, sex, diarrheal/non-diarrheal and urban/rural status) and sample randomization between the plates to reduce experimental error. In-house serum standards were aliquoted in seven to eight replicates per plate, to evaluate experimental error and integrity of generated data. An aliquot (10 µL) of each sample was transferred to 1 mL 96-well collection plates (Waters, Milford, MA, USA) for N-glycome analysis and the rest was used for isolation of IgG followed by IgG Fc N-glycopeptide analysis.

### 2.7.2. Serum N-Glycome Analysis

Serum N-glycans were enzymatically released from proteins by PNGase F, fluorescently labelled with 2-aminobenzamide and cleaned-up from the excess of reagents by hydrophilic interaction liquid chromatography-solid phase extraction (HILIC-SPE), as previously described [33]. Fluorescently labelled and purified N-glycans were separated by HILIC on a Waters BEH Glycan chromatography column, 150 × 2.1 mm i.d., 1.7 µm BEH particles, installed on an Acquity ultra-performance liquid chromatography (UPLC) H-class system (Waters, Wilmslow, UK), consisting of a quaternary solvent manager, sample manager and a fluorescence detector set with excitation and emission wavelengths of 250 nm and 428 nm, respectively. Obtained chromatograms were separated into 39 peaks. The amount of N-glycans in each chromatographic peak was expressed as a percentage of total integrated area. From 39 directly measured glycan peaks we calculated 12 derived traits which average particular glycosylation traits such as galactosylation, sialylation and branching across different individual glycan structures and are, consequently, more closely related to individual enzymatic activities and underlying genetic polymorphisms. Derived traits used: the proportion of low branching (LB); defined as di-antennary complex type N-glycans with two N-acetylglucosamine residues attached to the core pentasaccharide, (Man3GlcNAc2) at both the α-3 and α-6 mannose sites and high branching (HB); tri- and tetra-antennary complex type N-glycans with three of four N-acetylglucosamine (GlcNAc) residues attached to the core pentasaccharide. The majority of antennas are further elongated by the addition of galactose, sialic acid and fucose. Additional modifications such as the addition of bisecting GlcNAc and/or a fucose residue on the core pentasaccharide are also possible. N-glycans, the proportion of a-, mono-, di-, tri- and tetra-galactosylated N-glycans (G0, G1, G2, G3 and G4, respectively), and a-, mono-, di-, tri- and tetra-sialylated N-glycans (S0, S1, S2, S3 and S4, respectively).

### 2.7.3. IgG Fc N-Glycopeptides Analysis

Sample preparation and analysis of IgG N-glycopeptides was done following a previously described protocol with minor changes [34]. Briefly, IgG was isolated from 90 µL of serum samples by affinity chromatography using CIM<sup>®</sup> 96-well Protein G monolithic

plate (BIA Separations, Ajdovščina, Slovenia). IgG *N*-glycopeptides were prepared by trypsin digestion of an aliquot of IgG isolates (25 µg on average per sample) followed by reverse-phase solid phase extraction (RP-SPE). Purified tryptic IgG *N*-glycopeptides were separated and measured on nanoAcquity chromatographic system (Waters, Wilmslow, UK) coupled to Compact Q-TOF mass spectrometer (Bruker, Bremen, Germany), equipped with Apollo II source and operated under HyStar software version 3.2. The first four isotopic peaks of doubly and triply charged signals, belonging to the same glycopeptide species, were summed together, resulting in 20 Fc *N*-glycopeptides per IgG subclass. Predominant allotype variant of IgG3 tryptic peptide carrying *N*-glycans in the Caucasian population has the same amino acid sequence as IgG2 [35]. Therefore, IgG glycopeptides were separated into three chromatographic peaks designated as IgG1, IgG2/3 and IgG4. Signals of interest were normalised to the total area of each IgG subclass.

### 2.8. Immune and Diabetic Protein Profiling of Sera

Patient sera were analysed for the quantifications of 37 key biomarkers of inflammation from the TNF superfamily proteins, IFN family proteins, Treg cytokines, and MMPs: APRIL/TNFSF13, BAFF/TNFSF13B, sCD30/TNFRSF8, sCD163, Chitinase-3-like 1, gp130/sIL-6Rβ, IFN-α2, IFN-β, IFN-γ, IL-2, sIL-6Rα, IL-8, IL-10, IL-11, IL-12 (p40), IL-12 (p70), IL-19, IL-20, IL-22, IL-26, IL-27 (p28), IL-28A/IFN-λ2, IL-29/IFN-λ1, IL-32, IL-34, IL-35, LIGHT/TNFSF14, MMP-1, MMP-2, MMP-3, Osteocalcin, Osteopontin, Pentraxin-3, sTNF-R1, sTNF-R2, TSLP, TWEAK/TNFSF12 using the Bio-Plex Pro Human Inflammation Panel 1 (171AL001M, Bio-Rad, Hercules, CA, USA); immunoglobulins IgG1, IgG2, IgG3, IgG4, IgA, IgM, using the Bio-Plex Pro™ Human Isotyping Panel (171A3100M, Bio-Rad); and C-peptide, ghrelin, GIP, GLP-1, glucagon, insulin, leptin, PAI-1 (total), resistin and visfatin, using the Bio-Plex 10 Pro™ Human diabetes 10-plex immunoassay (171A7001M, Bio-Rad), respectively. Samples were analysed in a Bio-Plex 200 System using the Bio-Plex manager software, according to manufacturer's instructions. The concentrations were calculated by standard curves developed in parallel and are expressed as pg/mL for the inflammatory biomarkers and diabetic proteins, and ng/mL for the immunoglobulins.

Glycated serum protein (GSP) levels (µmol/L), which provide a short to medium-term assessment of glycaemia and diabetes risk [36], were assayed in sera by enzymatic assay (Crystal Chem, Elk Grove Village, IL, USA).

### 2.9. Statistical Analysis

As per the manufacturer's guidelines, all sera were assayed in duplicate in immune (antibody and inflammation panels), diabetic protein, and GSP assays. Descriptive statistics including median and interquartile range (IQR) are presented for demographic variables. Student's *t*-tests were used to detect differences in the abundance of microbial and immunometabolic features across the groups assessed. The association between the metavariables and microbial taxa was assessed using Pearson's correlation analysis. Identification and selection of the candidate biomarkers associated with urban, rural and BMI status, together with the performance of markers, was investigated using the elastic net method (see below) [37]. All *p*-values were adjusted where necessary to control for the false discovery rate (FDR) according to the Benjamini–Hochberg method. All analyses were performed in the R statistical computing (R version 3.4.3) environment. Statistical significance was set at an alpha of 5% for a two-sided *p*-value for all analyses.

#### Elastic Net Machine Learning Method

We applied elastic net (EN) machine learning method [37] to help select important features which may discriminate between the urban and rural population, and BMI groups. Elastic net automatically selects the best features linked with the outcome or response variable from the dataset-based penalty applied, and hence provides a sparse solution [38–40]. Penalty parameters,  $\lambda$  (Range of  $\lambda$ : 0 to 1), are optimized using 10-fold cross validation. The stronger the penalty (close to 1), the smaller the number of variables selected, while if

the penalty is weaker (close to 0), a higher number of variables are selected. In other words, the penalty function  $\lambda$  controls the trade-off between likelihood and penalty, thereby influencing the variables to be selected. Elastic net employs a mixed version of penalty called L1 (Least Absolute Shrinkage and Selection Operator also called as LASSO penalty) and L2 penalty (Ridge penalty). The L1 penalty encourages the sparse representation, whereas L2 stabilises the solution. The process was repeated 100 times and the features were ranked according to their respective selection frequency associated with each run. We then selected the first quartile from the EN-selected features over 100 runs. These selected features were then further modeled by generating area under curve (AUC) curves. We performed stability analysis [39] (also called a permutation analysis) after randomizing the class label (for rural vs. urban populations). We compared a random AUC based on each iteration and averaged over 100 iterations with the true AUC (without changing the class label). From these calculations, we generated two AUC distributions and compared mean values of the distributions, and generated  $p$ -values accordingly. All analyses were performed in the R statistical computing (R version 3.4.3) environment [41] and MetaboAnalyst web tool [42].

### 3. Results

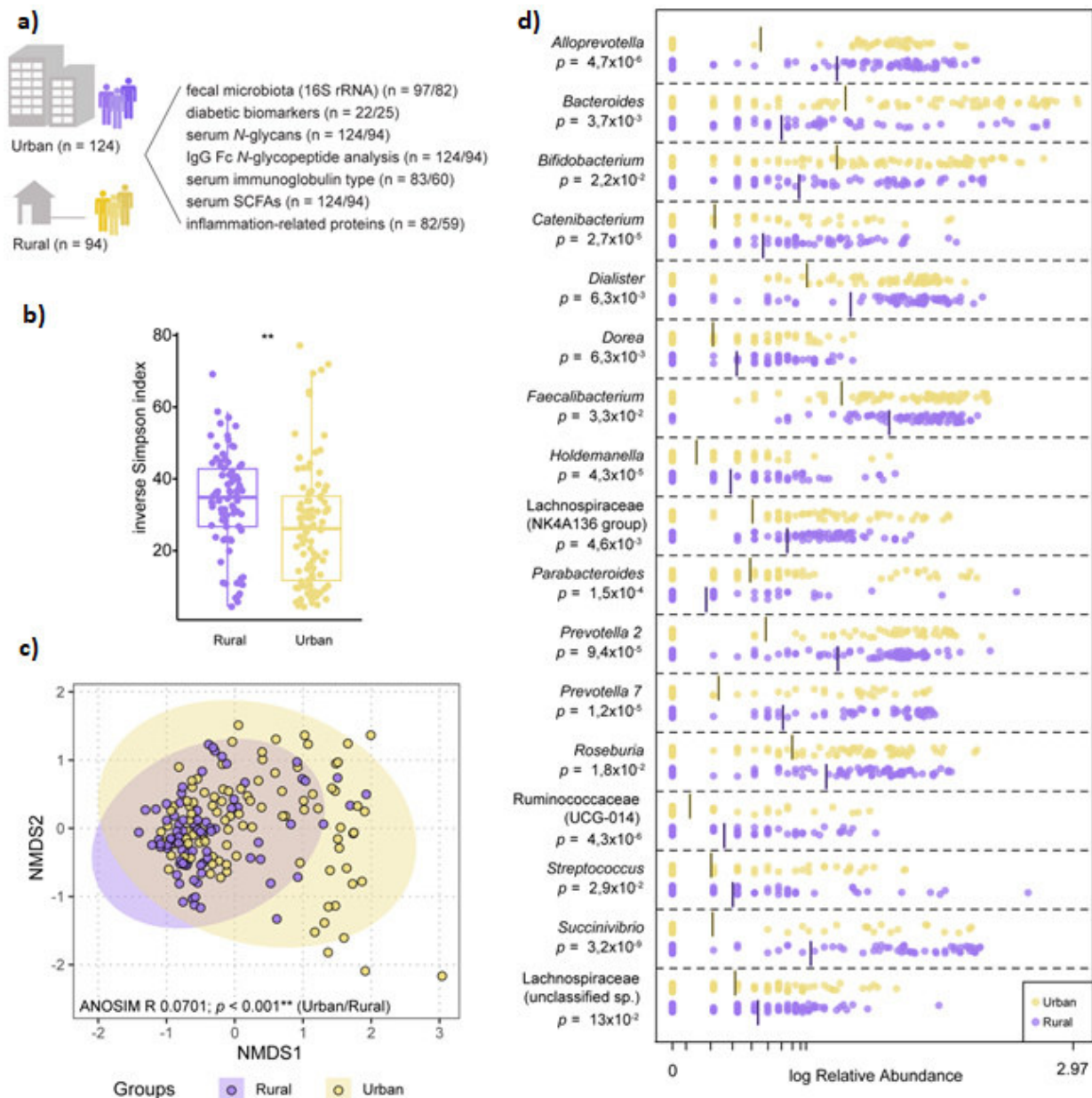
#### 3.1. Characteristics of the Study Participants

Clinical and demographic characteristics of the cohort are presented in Table 1 and geographic sites specified in Table S1. In total, 218 participants, of which 26.6% were inpatients, were enrolled into this prospective cohort study during 2019. A survey of the medical records of the urban cohort revealed that 10.5% of participants had diabetes mellitus at enrolment. No co-morbidity data were available for the rural cohort, highlighting a lack of diagnostic hospital facilities, a general reluctance to engage with Western medicine and a reliance on local faith healers and alternative medicines. In terms of cardiometabolic risk factors, 24.5% were active smokers (rural  $n = 23$  vs. urban  $n = 31$ ), and over half of the cohort were overweight (BMI  $\geq 23$ ) by Asian WHO standards. The urban Nagpurian cohort displayed significantly higher BMIs compared to their rural counterparts ( $p < 0.001$ ).

**Table 1.** Baseline characteristics of study population. Descriptive statistics presented as the number of samples ( $n$ ) and percentage (%) or median (interquartile range, IQR).

Characteristic	Rural, $n = 94$	Urban, $n = 124$	$p$ -Value
Age, yrs (median (IQR))	39 (27, 53)	38 (30, 49)	>0.9
Gender			0.3
Female	47 (50%)	52 (42%)	
Male	47 (50%)	72 (58%)	
BMI (median (IQR))	21.0 (19.2, 22.3)	25.0 (23.5, 26.0)	<0.001
BMI Class			<0.001
Underweight	10 (11%)	0 (0%)	
Normal	68 (72%)	20 (16%)	
Overweight	12 (13%)	38 (31%)	
Pre-Obese	2 (2.1%)	62 (50%)	
Obese	2 (2.1%)	4 (3.2%)	
Smoker	23 (24%)	31 (25%)	>0.9
Hospitalized	13 (14%)	45 (36%)	
Drugs			0.017
Antacid	24 (26%)	12 (9.7%)	
PPI	1 (1.1%)	1 (0.8%)	
Co-morbidities			<0.001
Diabetes mellitus	8 (8.5%)	15 (12%)	
Epilepsy	3 (3.2%)	12 (9.7%)	
High cholesterol	0 (0%)	1 (0.8%)	
Hypertension	0 (0%)	7 (5.6%)	
Hypothyroidism	0 (0%)	1 (0.8%)	
Seizure disorder	0 (0%)	1 (0.8%)	
Tuberculosis	0 (0%)	1 (0.8%)	
Toilet facilities	80 (85%)	124 (100%)	<0.001
Hand soap	80 (85%)	124 (100%)	<0.001
Domestic animals	42 (45%)	21 (17%)	<0.001
Water supply			<0.001
Borewell	0 (0%)	18 (15%)	
Corporation water connection	6 (6.4%)	101 (81%)	
Corporation water tank	78 (83%)	3 (2.4%)	
Well water	10 (11%)	2 (1.6%)	

Following quality control checks, we analysed 179 fecal samples for taxonomic composition by 16S rRNA gene amplicon sequencing. Sera were profiled for relative abundance of total serum *N*-glycans and IgG Fc *N*-glycopeptides ( $n = 218$ ), detection and quantification of short chain fatty acids ( $n = 218$ ), an inflammation panel of immune proteins ( $n = 141$ ), a multi-isotype antibody panel ( $n = 143$ ), glycosylated serum protein levels ( $n = 135$ ), and a diabetes panel ( $n = 47$ ); see Figure 1A for study schematic with urban/rural sampling numbers and Supplementary Table S2 for study metrics.



**Figure 1.** The microbiota is structurally distinct in participants from rural vs. urban areas. (a) Schematic of overall study design ( $n$  = number of urban/rural samples). (b) Diversity as determined by inverse Simpson index based on normalized ASV counts in participants from rural vs. urban areas (Kruskal–Wallis nonparametric test,  $p < 0.001$ ). (c) Non-metric multidimensional scaling (NMDS) visualization of Bray–Curtis distance (based on normalized ASV counts) of the microbiota in participants based on geography (rural vs. urban; purple vs. yellow). Analysis of similarities (ANOSIM) was conducted using Bray–Curtis distance, 9999 permutations. (d) Log-transformed relative abundance of significantly differential genera between participants from rural or urban areas, as determined by Linear discriminant analysis Effect Size (LEfSe).

### 3.2. Microbiota Composition Varies by Geographic-Specific Factors

Significant differences in microbiota diversity, structure, and composition were observed between urban and rural participants. Overall, microbiota diversity was increased in the rural population (Figure 1B), and ANOSIM on NMDS ordination indicated significant separation between the two groups (Figure 1C). LEfSe identified several overrepresented genera belonging to the Firmicutes phylum in the rural population, including significant differences in relative abundance of *Faecalibacterium*, *Roseburia*, unclassified *Lachnospiraceae* and *Ruminococcaceae* groups. Within Bacteroidetes, the rural microbiota was dominated by *Prevotella* and *Alloprevotella* genera, while *Bacteroides* and *Parabacteroides* were overrepresented in the urban microbiota (Figure 1D). Community type analysis using PAM clustering revealed two major clusters, with an overrepresentation of rural samples clustering within one cluster (69/82) compared to urban samples, which were more evenly distributed between both clusters (56 vs. 41 samples; Pearson's chi-squared test,  $p < 0.001$ ).

BMI (defined as 'low/normal' <18.5/18.5–22.9 vs. 'high' >23) was not a significant factor in differentiating microbiota composition or diversity; however, an unclassified *Ruminococcaceae* group (*Ruminococcaceae\_UCG-014*) was overrepresented in participants with a high BMI across all samples and within rural participants (online Supplementary Figure S1).

### 3.3. Dysmetabolic Hallmarks and Urban Living

Diabetic biomarker panel profiling revealed substantially higher levels of proteins linked to diabetes including C-peptide, insulin and leptin in the peripheral circulation of the sampled urban population (Table 2). Accompanying serum *N*-glycan profiles demonstrated glycan structural features of increased branching, galactosylation and sialylation in the urban cohort, in line with increasing plasma *N*-glycome complexity typically observed in individuals with increased risk of type 2 diabetes development [17]. Specifically, we found a statistically significant increase in levels of high-branching, tri- and tetragalactosylated glycans, tri- and tetrasialylated glycans and increase in levels of glycans with antennary fucosylation in inhabitants of Nagpur. For IgG Fc *N*-glycopeptide analysis, IgG1 glycopeptides with agalactosylated and monogalactosylated glycans were detected at a significantly higher relative abundance in the sera of tested urban inhabitants.

**Table 2.** Features which show significant differential responses between rural and urban cohorts are shown using two-tailed Student's *t*-test. An FDR corrected *p*-value is shown in the last column. Arrows (↑/↓) represent features that were increased/decreased in the corresponding population.

Feature	tstat	Rural	Urban	<i>p</i> -Value (FDR Corrected)
Serum Short-chain Fatty Acids				
Caproate	6.679	↑	↓	0.000000
Valerate	5.5217	↑	↓	0.000001
Acetate	3.1602	↑	↓	0.006598
Propionate	3.0367	↑	↓	0.007375
Serum Diabetic panel				
BMI	−3.9651	↓	↑	0.003120
C-peptide	−3.4949	↓	↑	0.006466
Insulin	−3.0994	↓	↑	0.013355
Leptin	−2.9744	↓	↑	0.014119
Serum IgG Fc <i>N</i> -Glycopeptides				
IgG1 H4N4F1: IgG1 glycopeptide with monogalactosylated glycan with core fucose	−3.6748	↓	↑	0.004191
IgG4 H5N4F1: IgG4 glycopeptide with digalactosylated glycan with core fucose	3.4585	↑	↓	0.004569
IgG1 H3N4F1: IgG1 glycopeptide with agalactosylated glycan with core fucose	−2.9742	↓	↑	0.014886
IgG4 H5N4F1S1: IgG4 glycopeptide with digalactosylated and monosialylated glycan with core fucose	2.889	↑	↓	0.014886
IgG1_H5N4F1S1: IgG1 glycopeptide with digalactosylated and monosialylated glycan with core fucose.	2.5309	↑	↓	0.033823

Table 2. Cont.

Feature	tstat	Rural	Urban	p-Value (FDR Corrected)
Serum Immunoglobulin isotype				
IgG1	−3.5703	↓	↑	0.003905
IgM	2.5608	↑	↓	0.045976
Inflammation-related Protein				
IFN- $\gamma$	3.077	↑	↓	0.051323
Osteocalcin	−3.063	↓	↑	0.051323
Serum N-Glycans				
S4: Tetrasialylated glycans	−5.2077	↓	↑	0.000004
G4: Tetragalactosylated glycans	−5.1823	↓	↑	0.000004
AF: Antennary fucosylation	−4.7813	↓	↑	0.000019
S1: Monosialylated glycans	3.9387	↑	↓	0.000413
HB: High branching glycans	−3.9283	↓	↑	0.000413
LB: Low branching glycans	3.8475	↑	↓	0.000470
S3: Trisialylated glycans	−3.25	↓	↑	0.003435
G2: Digalactosylated glycans	2.9324	↑	↓	0.008372
G3: Trigalactosylated glycans	−2.7838	↓	↑	0.011686
B: Bisection (Glycans with bisecting GlcNAc)	2.403	↑	↓	0.030770
HM: High mannose glycans	2.2316	↑	↓	0.043612

### 3.4. Rural Living Associates with Contrasting Serum Immunometabolic Features

Levels of a number of short chain fatty acids (including caproate, valerate, acetate and propionate) were significantly higher in the sera of rural inhabitants. Rural inhabitants showed a significantly higher relative abundance of low branching, monosialylated and digalactosylated serum glycans, as well as a higher abundance of bisected and high mannose serum glycans (Table 2). Analysis of IgG Fc N-glycopeptides revealed a higher relative abundance of IgG1 and IgG4 glycopeptides with digalactosylated and monosialylated glycans with core fucose in the circulation of rural subjects.

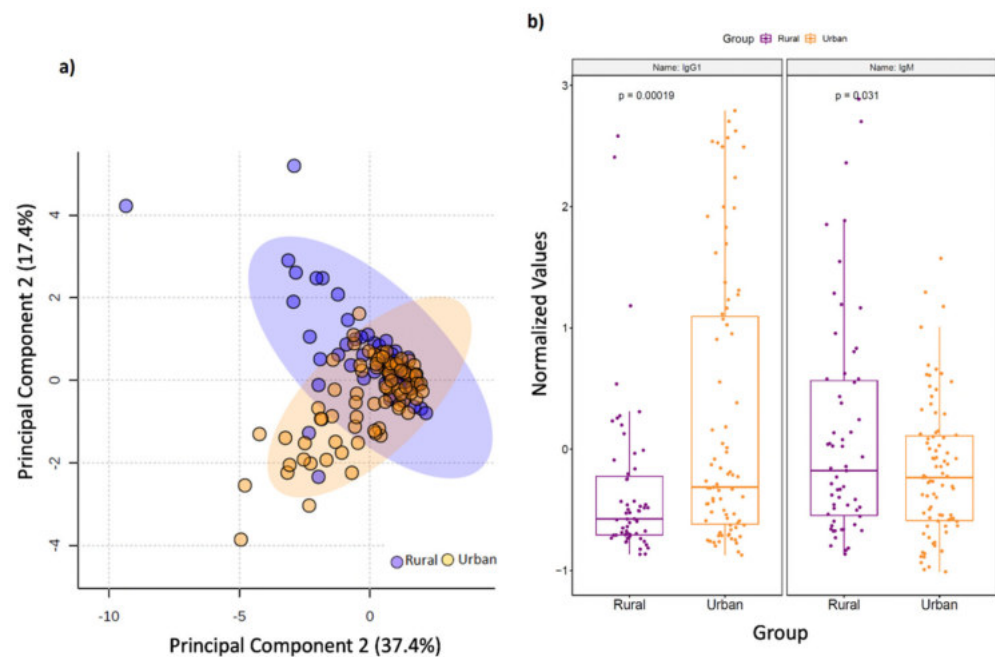
Geographic differences also extended to circulating immunoglobulin responses. Principal component analysis (PCA) (Figure 2A) demonstrated a clear separation of multi-isotype antibody responses between rural and urban cohorts. In particular, the rural cohort displayed significantly higher levels of circulating total IgM antibodies, whereas IgG1 antibodies were significantly higher in the urban cohort ( $p < 0.05$ ; Figure 2B). Correlation analyses also focussed on studying connections between immunoglobulin responses and SCFAs, the latter of which are known to fuel antibody responses. Here, we found that serum 2-hydroxybutyrate positively correlated with IgG4 levels in the rural cohort ( $p < 0.05$ ), and IgG4 strongly positively associated with *Porphyromonas*, *Campylobacter*, *Gemella*, *Streptococcus*, *Leptotrichia* and *Neisseria* ( $p = 0$ ).

### 3.5. Diabetic Protein-Microbe Interactions Vary by Geography

In the urban group (see online Supplementary Table S3), the strongest positive Pearson correlations were detected between visfatin and *Bacillales*, *Marinifilaceae*, *Staphylococcaeae*, *Odoribacter*, *Macelibacteroides*, *Staphylococcus*, *Hungatella*, *Ruminiclostridium\_6*, *Erysipelatoclostridium*, *Acidaminococcus*, and *Lactobacillaceae* ( $p < 0.0001$ ); followed by leptin with *Actinobacteria*, and *Bifidobacteriales* at class, family and genus. Positive correlations were observed for GLP-1 with *Escherichia/Shigella*, *Enterobacteriaceae*, *Proteobacteria* and *Gammaproteobacteria*. C-peptide and insulin also positively associated with *Gammaproteobacteria*, *Proteobacteria* and *Enterobacteriaceae*.

In contrast, in the rural group, (see online Supplementary Table S4), the strongest positive correlations were detected for GLP-1 with unclassified *Erysipelotrichaceae\_unclassified*, *Anaeroplasmatales* at class, family and genus, and *Erysipelotrichaceae\_UCG.004*, and for C-peptide, with *Paraprevotella*, *Flavonifractor*, *UBA1819* and *Erysipelatoclostridium* ( $p < 0.0001$ ).

Further significant diabetic protein–microbiota–immune correlations for the urban and rural groups are presented in online Supplementary Tables S3 and S4, respectively.



**Figure 2.** Serum immunoglobulin levels vary by geography. (a) Principal component analysis (PCA) score plot on the selected features demonstrates a clear separation in serum multi-isotype antibody responses in terms of geographic setting of sampled population. Dots represent patients and are coloured according to the subject cohort. Ellipse represents 95% confidence. Results are plotted according to the Principal component-1 (PC1) and Principal component-2 (PC2) scores, with the percent variation of the cohort explained by the respective x and y axes. (b) Box plots showing levels of serum IgM and IgG1 antibodies in rural and urban cohorts, respectively.

### 3.6. Differential Impact of Glycated Serum Protein Levels on Immunometabolic and Gut Bacterial Features

In subjects where glycated serum protein (GSP) concentrations were assessed ( $n = 135$ ), levels were significantly higher in overweight (BMI 23–24.9) and pre-obese (BM 25–29.9) test subjects compared to those with normal BMI (18.5–22.9), and in urban subjects compared to rural participants;  $p < 0.001$  (Table S5). Across the whole cohort, high GSP levels were associated with significantly lower circulating IgG2, IgM, caproate, and valerate levels, and lower relative abundance of *Roseburia* and *Dorea* ( $p < 0.05$ ). See Tables 3–5.

In urban BMI comparisons, circulating levels of 2-hydroxybutyrate, isobutyrate, propionate, and valerate were significantly higher in overweight subjects compared to those with normal BMI (see Supplementary Table S6;  $p < 0.05$ ). Similarly, levels of isobutyrate, propionate and valerate, alongside acetate were higher in pre-obese vs. normal BMI subjects (see online Supplementary Table S7). Contrastingly, in rural-BMI group comparisons, pre-obese subjects displayed significantly lower levels of 2-methylbutyrate, acetate, caproate, isobutyrate and isovalerate, but a higher relative abundance of *Collinsella*, *Prevotella\_9*, *Agathobacter*, *Roseburia*, *Faecalibacterium*, *Ruminococcaceae unclassified*, *Ruminococcaceae\_UCG.014*, *Catenibacterium*, *Megasphaera* and *Mitsuokella* compared to subjects with a normal BMI ( $p < 0.05$ ; see online Supplementary Table S8). In underweight rural subjects ( $n = 8$ ), there was a higher representation of *Collinsella*, *Roseburia* and Pentraxin 3 compared to the normal BMI group (see online Supplementary Table S9).

**Table 3.** Features which demonstrate differential responses between normal and low glycosylated serum protein (GSP) levels ( $\mu\text{mol/L}$ ). Low GSP = 0–199; Normal GSP = 200–285; MMP-2 = Matrix metalloproteinase-2; MMP-3 = Matrix metalloproteinase-3; sCD163 = Soluble CD163; sIL-6R $\alpha$  = Soluble interleukin 6 receptor alpha; IFN- $\alpha$ 2 = Interferon alpha-2; sCD30/TNFRSF8 = Tumour necrosis factor receptor superfamily member 8; Two-tailed Student's t- test. An FDR corrected *p*-value is shown in the last column. Arrows ( $\uparrow/\downarrow$ ) represent features that were increased/decreased in the corresponding population.

Feature	tstat	Normal GSP (n = 30)	Low GSP (n = 54)	<i>p</i> -Value (FDR Corrected)
MMP-2	−3.5975	$\uparrow$	$\downarrow$	0.000548
HM: High mannose glycans	2.8571	$\downarrow$	$\uparrow$	0.005416
MMP-3	2.8315	$\downarrow$	$\uparrow$	0.005827
sCD163	−2.7054	$\uparrow$	$\downarrow$	0.008297
sIL-6R $\alpha$	−2.6473	$\uparrow$	$\downarrow$	0.009727
IFN- $\alpha$ 2	−2.4229	$\uparrow$	$\downarrow$	0.017598
IgG4 H5N4F1S1: IgG4 glycopeptide with digalactosylated and monosialylated glycan with core fucose	2.3389	$\uparrow$	$\downarrow$	0.021773
<i>Cyanobacteria</i>	−2.2579	$\uparrow$	$\downarrow$	0.026608
<i>Melainabacteria</i>	−2.2579	$\uparrow$	$\downarrow$	0.026608
2-methylbutyrate	−2.196	$\uparrow$	$\downarrow$	0.030914
AF: Antennary Fucosylation	−2.1194	$\uparrow$	$\downarrow$	0.03708
<i>Gastranaerophilales_unclassified</i>	−2.0844	$\uparrow$	$\downarrow$	0.040231
<i>Gastranaerophilales</i>	−2.0666	$\uparrow$	$\downarrow$	0.041926
sCD30/TNFRSF8	−2.0552	$\uparrow$	$\downarrow$	0.043046

**Table 4.** Features which demonstrate differential responses between normal and high GSP levels. Normal GSP = 200–285; High GSP = 286–400; APRIL/TNFSF13 = A proliferation-inducing ligand/Tumor necrosis factor ligand superfamily member, 13; Two-tailed Student's t- test. Arrows ( $\uparrow/\downarrow$ ) represent features that were increased/decreased in the corresponding population. An FDR corrected *p*-value is shown in the last column.

Feature	tstat	Normal GSP (n = 30)	High GSP (n = 33)	<i>p</i> -Value (FDR Corrected)
IgG2	−2.7269	$\uparrow$	$\downarrow$	0.008335
Caproate	−2.6832	$\uparrow$	$\downarrow$	0.009373
Roseburia	−2.4077	$\uparrow$	$\downarrow$	0.019095
Valerate	−2.2378	$\uparrow$	$\downarrow$	0.028897
Dorea	−2.2193	$\uparrow$	$\downarrow$	0.030193
IgM	−2.1594	$\uparrow$	$\downarrow$	0.034761
APRIL/TNFSF13	2.141	$\downarrow$	$\uparrow$	0.036276

**Table 5.** Features which demonstrate differential responses between normal and very high GSP levels; Normal GSP = 200–285; Very high GSP = >400. Two-tailed student's t- test. Arrows ( $\uparrow/\downarrow$ ) represent features that were increased/decreased in the corresponding population. An FDR corrected *p*-value is shown in the last column.

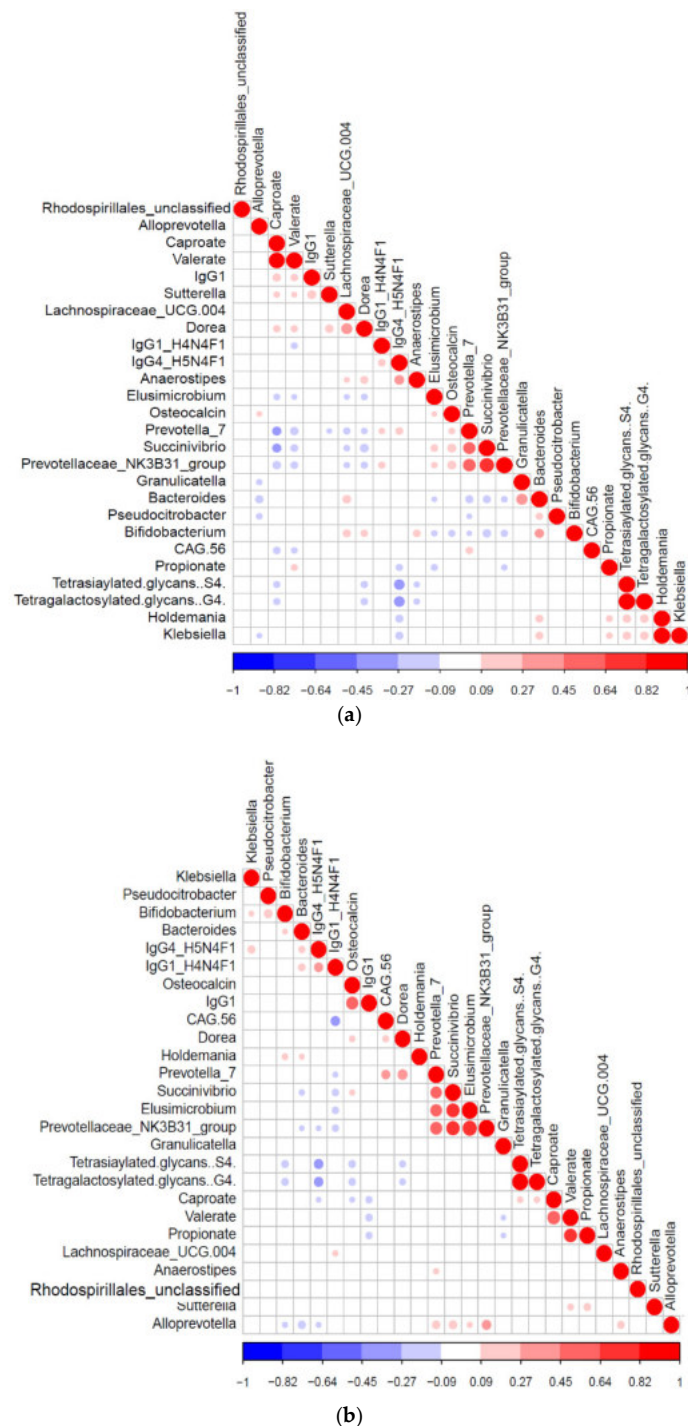
Feature	tstat	Normal GSP (n = 30)	Very High GSP (n = 18)	<i>p</i> -Value (FDR Corrected)
Caproate	2.4758	$\uparrow$	$\downarrow$	0.017035
Blautia	−2.0712	$\downarrow$	$\uparrow$	0.04398
Osteopontin	2.0162	$\uparrow$	$\downarrow$	0.049643

### 3.7. Multiomics Data Integration Identified Potential Biomarkers Distinguishing Urban vs. Rural Cohort

We constructed Pearson correlation-based heatmaps to reveal interactions between microbial taxa and immunometabolic features. These were filtered by geographic status (rural vs. urban, Figure 3). In the urban group (Figure 3B), positive associations (red circles) were seen for tetrasialylated and tetragalactosylated serum glycans with serum caproate.



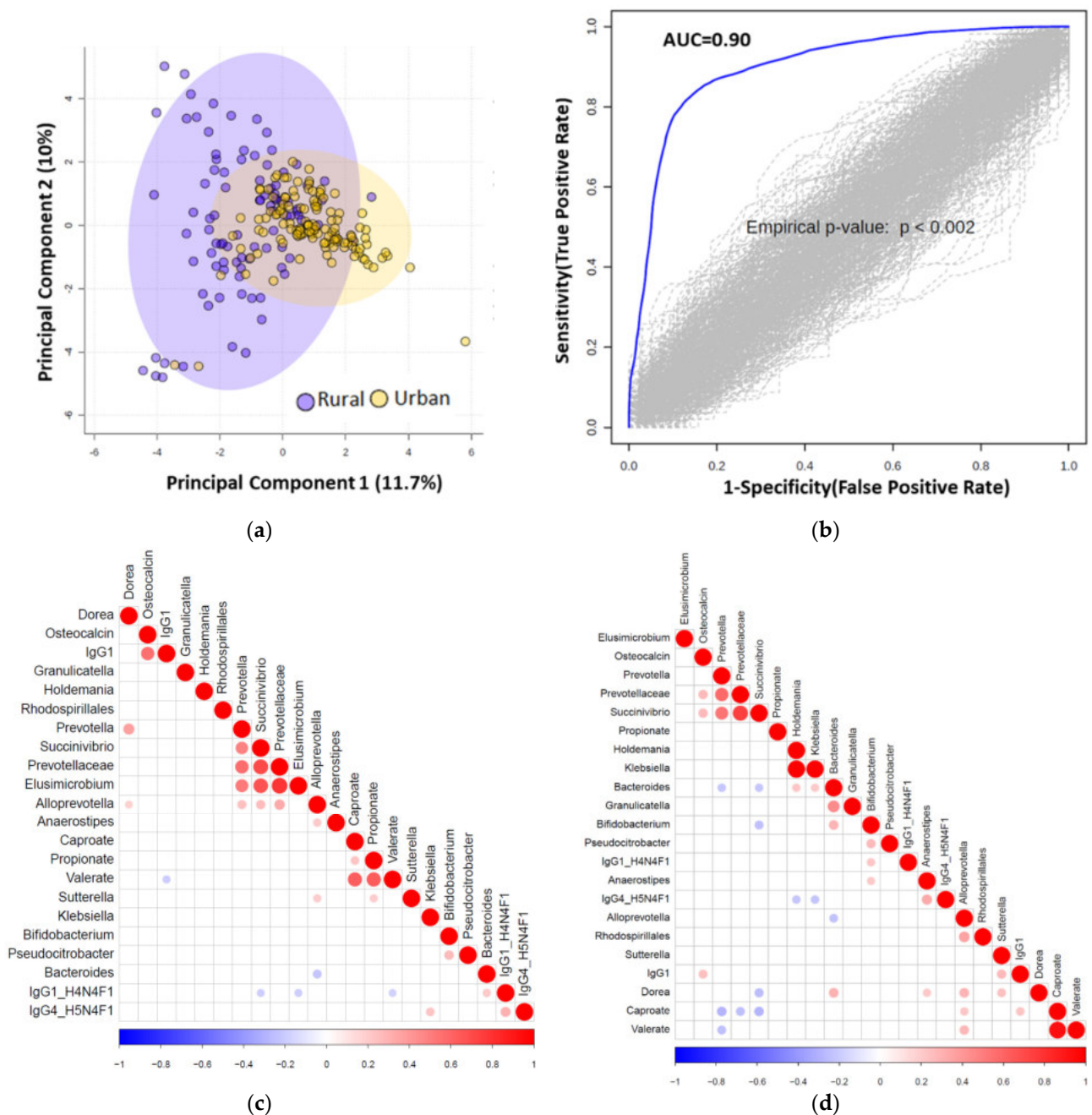
Negative correlations (blue circles) were seen between these same complex glycans and *Bifidobacterium*, *Dorea*, osteocalcin and IgG4 glycopeptides with digalactosylated glycan with core fucose. Notable clusters in the rural group (Figure 3A) included positive correlations with *Holdemania* and *Klebsiella* with propionate, tetrasialylated and tetragalactosylated serum glycans.



**Figure 3.** Significant Pearson correlation ( $p < 0.05$ ) of the selected features for the (a) rural ( $n = 94$ ) and (b) urban samples (124). Correlated variables are either highly positively correlated (in blue circles) or negatively correlated (red circles).

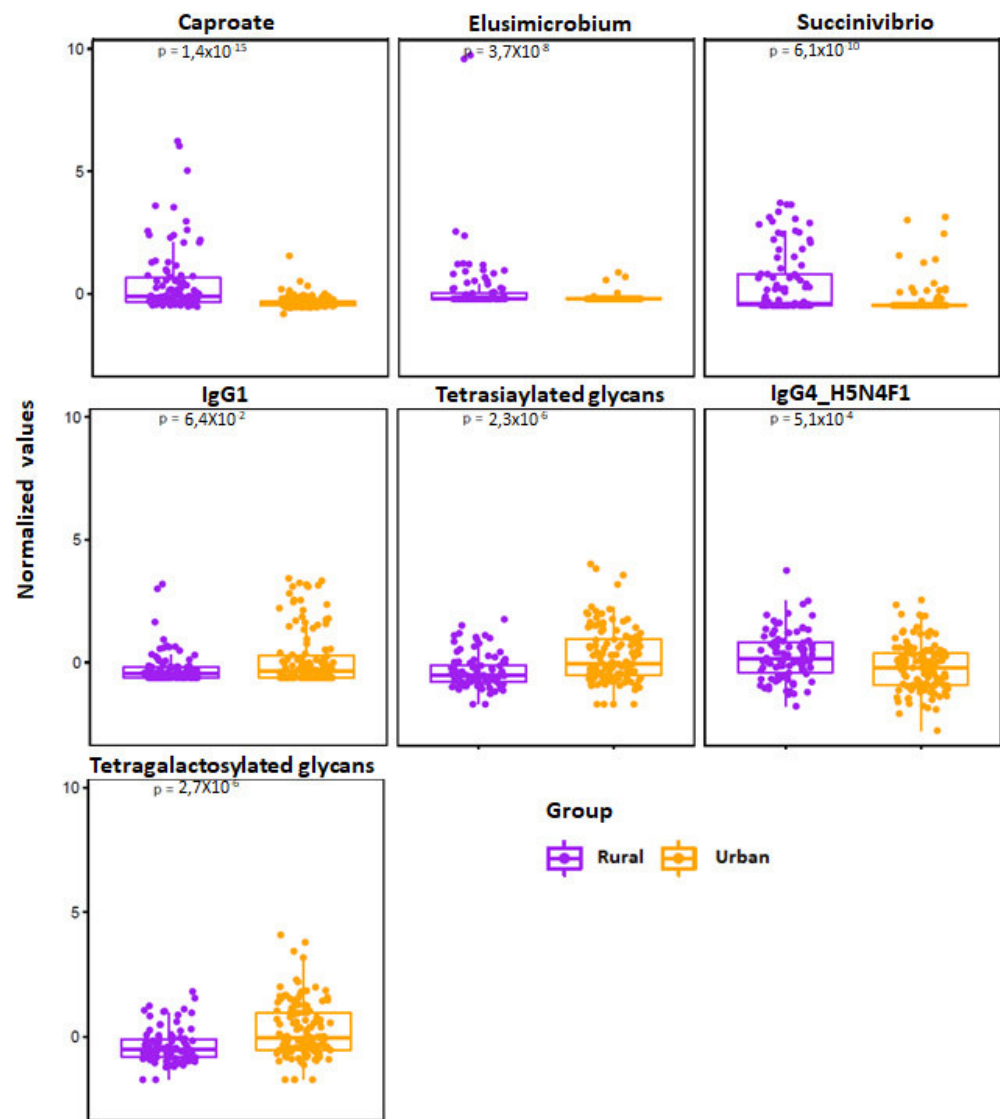
We next used elastic net (EN) algorithm to identify and select the most important features representing potential biomarker candidates distinguishing rural vs. urban (Figure 4), and

normal BMI vs. overweight or pre-obese groups (see online Supplementary Figures S1 and S2). We ran the model 100 times using different training sets and ranked the selected features based on the selection frequency and chose the first quartile features. We then compared area under the curve (AUC) value with the selected features and 1000 random permuted data sets.



**Figure 4.** (a) Principal component analysis (PCA) score plot performed on the selected omics features demonstrating clustering of the rural vs. urban cohorts. Dots represent patients and are coloured according to the subject cohort. Ellipse represents 95% confidence. Results are plotted according to the Principal component-1 (PC1) and Principal component-2 (PC2) scores, with the percent variation of the cohort explained by the respective x and y axes. (b) Permutation test to show the stability of the AUC value after randomizing the urban and rural samples 100 times. (c) Significant correlation ( $p < 0.05$ ) heatmap of the elastic net selected features is shown for urban samples. (d) Significant Pearson correlation ( $p < 0.05$ ) heatmap of the elastic net selected features is shown for rural ( $n = 94$ ) samples.

We identified multiple distinguishing features across the rural vs. urban groups and present their frequency in online Supplementary Table S10. We show significance levels and directionality of response for identified discriminatory features in Figure 5. Using those features, we found an AUC value of 0.90 between urban vs. rural population using logistic regression (Figure 4B).



**Figure 5.** Jitter plot of the normalized selected features from elastic net analysis are shown for the rural ( $n = 94$ ) vs. urban ( $n = 124$ ) cohorts. Wilcoxon Rank test was performed and results were obtained, and  $p$ -values are shown in the respective plots.

To integrate and explore associations between the selected features using the elastic net method, we generated heatmap plots (correlation method based) for urban and rural groups, as shown in Figures 4C and 4D, respectively. As a representative example, the most important rural classifiers in EN were serum caproate and relative abundance of two taxa: *Elusimicrobium* and *Succinivibrio*. Caproate positively correlated with *Alloprevotella* and serum IgG1 and *Succinivibrio* with osteocalcin, *Prevotella* and *Prevotellaceae*.

For the BMI EN group comparisons, features which could distinguish between overweight and pre-obese subjects, and those with normal BMI, are displayed in online Supplementary Tables S11 and S12. PCA analysis and AUC values are displayed in Supplementary Figures S2 and S3. In particular, *Dialister* was significantly underrepresented in the pre-obese group compared to the group with normal BMI;  $p < 0.036$ .

#### 4. Discussion

This study is the first integrative omics-based population study in which baseline gut bacterial, as well as systemic immunometabolic and glycomic traits, have been captured in geographically divergent populations in Central India. Strikingly, our findings identify a constellation of biomolecular traits that associate with metabolic dysregulation. These are principally seen in unselected urbanised populations without known diabetes. Although there is now substantial evidence connecting the gut microbiome to physiological parameters related to metabolic disorders such as diabetes [24], very few deep phenomic studies have been undertaken in Asian populations to obtain a better understanding of the biological processes associated with both healthy individuals and those at potential risk of developing diabetes. Identification of individuals at higher risk of developing diabetes is of great importance as early interventions may delay or even prevent overt diabetes. By unravelling and understanding the immunometabolic interplay between gut microbiome and the host, individualised therapeutic strategies including novel prebiotics, probiotics, synbiotics, and postbiotics could be explored to prevent or treat cardiometabolic disorders [24,43].

Several observational studies have indicated that obesity, estimated by BMI and an insulin resistance state, is a very important risk factor for T2DM [44–46]. Since variations in glucose metabolism are known to directly affect glycosylation, we studied serum N-glycan profiles and observed a more complex glyco phenotype, that has previously been reported to be associated with a higher risk of developing T2DM and poorer regulation of blood sugar levels [18], in the urban population group. These pathogenic complex glycans (tetragalactosylated and tetrasialylated glycans) positively correlated with serum caproate in the urban population, and *Holdemania* and *Klebsiella* in the rural population, suggesting that these serum metabolites and genera are potentially diabetogenic. Similarly, circulating IgG N-glycopeptide profiles revealed a higher relative abundance of pro-inflammatory IgG glycoforms (IgG1 glycopeptides with agalactosylated glycans) in urban participants, which is consistent with that seen in other inflammatory diseases [19]. It has been suggested that agalactosylated IgG species have an enhanced capacity to activate the complement system via the lectin pathway, thereby contributing to the development of inflammation as an underlying pathological mechanism of autoimmune diseases [19]. In contrast, C-peptide, GIP, insulin and leptin correlated negatively with the anti-inflammatory gut commensal *Faecalibacterium*, a beneficial microbe which produces SCFAs [47–49].

Analysis of the faecal taxonomic compositional profiles revealed a dominant prevalence of *Prevotella* and *Alloprevotella* genera in rural microbiota and overrepresentation of *Bacteroides* and *Parabacteroides* in the urban microbiota, a finding substantiated by our earlier microbiome observations in Central India [50].

We observed geographic-specific variation in immunoglobulin levels, which may be due to as yet unidentified genetic and environmental factors. We surmise that frequent exposure to a wide range of infectious agents, and other environmental stressors in the rural cohort, may have skewed the humoral response towards IgM to help protect the host from invading pathogens not previously encountered. However, there is mounting evidence that natural IgM antibodies also contribute to critical innate immune functions involved in the maintenance of tissue homeostasis, including augmenting the clearance of apoptotic cells and mediating specific anti-inflammatory signaling pathways [51,52]. Higher levels of IgG in the urban cohort could reflect (meta)inflammation-associated immunosenescence, in which there is a shift towards immunoglobulins being produced by naive B cells (IgD, IgM) to immunoglobulin produced by memory B cells (IgG, IgA) [51,52]. The presence of a higher burden of diabetic-related proteins and a complex glyco phenotype in the circulation of urban populations is consistent with MetS phenotype, which in younger adults may be a sign of premature ageing [53]. Thus, preventing and treating MetS and cardiovascular disease would be useful in promoting normal ageing. These findings are also in keeping with the prevalence of type 2 diabetes mellitus and metabolic syndrome which strongly associate with urban residency in Southern Asia and India, respectively [54,55]. We previously demonstrated

that FMT for successful recurrent *Clostridioides difficile* associates with a reduction in the complexity of serum N-glycosylation profiles, which is mainly driven through a significant reduction in the relative abundance of high branching, tetragalactosylated and trisialylated glycans [56]. We therefore infer that faecal microbiota-based interventions may be useful in helping to reverse a complex glyco-phenotype, which may lead to improved metabolic health.

We also detected geographic-specific variation in levels of serum SCFAs which was significantly higher in the rural cohort. This particular observation has been reported in other studies where, compared to industrial human microbiomes, non-industrial gut microbiomes show greater diversity of genes involved in complex carbohydrate metabolism, and demonstrate higher amounts of SCFAs in stool [57,58]. These trends have been linked to plant-based diets rich in fibers, infrequent consumption of highly processed foods, and low exposure to pharmaceutical drugs, such as antibiotics, in non-industrialised populations [59]. Of note, we observed a positive correlation between circulating levels of IgG1 and caproate and valerate in the rural cohort, whereas a negative correlation was seen for caproate and IgG1 for the urban cohort. The former observation is supported by evidence which shows that short SCFAs function as commensal-derived stimulators of host antibody responses, by accelerating cellular metabolism and regulating gene expression to promote B cell differentiation into antibody-producing cells [60]. It remains unclear why caproate was found to negatively correlate with IgG1 in the urban population.

In terms of study limitations, not all our omics data sets were complete, a limitation which arose due to small volume of blood samples (2 mL) permitted to be collected. We were unable to acquire fasting blood samples for GSP and diabetes panel measurements and did not assess fasting blood glucose or HbA1C levels, again due to considerable practical challenges imposed by working in under-resourced and remote areas of Central India. For this same reason, we could not assess dietary or genetic effects which are likely to be important drivers of metabolic health. Moreover, we are mindful that our study findings are largely associative and not causal and, thus, will require a follow-on validation cohort study to assess their translational potential. Future work should focus on designing larger longitudinal meta-omics studies to decipher host–microbe interactions in health and disease using multi-ethnic cohorts.

## 5. Conclusions

In conclusion, we present multi-level evidence which suggests that urban living, rather than an elevated BMI, drives dysmetabolic phenotypes in young urban and rural populations in Central India. These findings start to deconvolute the complex interaction between the environment, gut microbiota, immunometabolism and dysmetabolism in a non-Western population. Our observations may serve as a launchpad for novel approaches to prediction and intervention to minimize the risk of T2DM within these vulnerable populations.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/microorganisms9071485/s1>, Table S1: Sampling sites from towns and villages in rural Melghat, Amravati district, sub-divisional hospitals (SDH) and public health centres (PHC), and urban Nagpur district, Maharashtra state, India. Table S2: Study metrics (microbial and immunometabolic) Table S3: Pearson correlation coefficients correlating diabetic proteins with microbiota and molecular features in urban cohort. A corrected *p*-value was obtained from the correlation test. Table S4: Pearson correlation coefficients correlating diabetic proteins with microbiota and molecular features in rural cohort. A corrected *p*-value was obtained from the test. Table S5: Glycated serum protein levels (GSP;  $\mu\text{mol/L}$ ). GSP levels were categorized into low, normal, high or very high categories and assessed by BMI, gender and geography. Low GSP = 0–199; Normal GSP = 200–285; High GSP = 286–400; Very high GSP = >400. BMI ranges were pre-defined using WHO Asian BMI classifications: underweight <18.5, normal (18.5–22.9), overweight (23–24.9), pre-obese (25–29.9) and obese ( $\geq 30$ ) categories. Number of samples and percentage is represented. A chi-square test of independence was performed and a corrected *p*-value was obtained from the test. Table S6: Urban BMI group comparisons showing differential features in normal BMI vs. overweight groups. BMI ranges were pre-defined using WHO Asian BMI classifications: underweight <18.5, normal (18.5–22.9), overweight (23–24.9), pre-obese

(25–29.9) and obese ( $\geq 30$ ) categories. Median and interquartile range (IQR). Kruskal–Wallis test (for continuous data) or chi-square test of independence (for ordinal data) and corrected  $p$ -values shown. Table S7: Urban BMI group comparisons showing differential features in normal BMI vs. pre-obese groups. BMI ranges were pre-defined using WHO Asian BMI classifications: normal (18.5–22.9); pre-obese (25–29.9) categories. Median and interquartile range (IQR). Kruskal–Wallis test (for continuous data) or chi-square test of independence (for ordinal data) and corrected  $p$ -values shown. Table S8: Rural-BMI group comparisons showing differential features in normal BMI vs. pre-obese groups. BMI ranges were pre-defined using WHO Asian BMI classifications: normal (18.5–22.9); pre-obese (25–29.9) categories. Median and interquartile range (IQR). Kruskal–Wallis test (for continuous data) or chi-square test of independence (for ordinal data) and corrected  $p$ -values shown. Table S9: Rural-BMI group comparisons showing differential features in normal BMI vs. underweight groups. BMI ranges were pre-defined using WHO Asian BMI classifications: underweight  $< 18.5$ ; normal (18.5–22.9) categories. Median and interquartile range (IQR). Kruskal–Wallis test (for continuous data) or chi-square test of independence (for ordinal data) and corrected  $p$ -values shown. Table S10: Elastic net selected frequency of all the features for urban vs. rural group with rankings in decreasing order over 100 iterations. Table S11: Elastic net selected frequency of all the features for normal vs. overweight groups with rankings in decreasing order over 100 iterations. Table S12: Elastic net selected frequency of all the features for normal vs. pre-obese groups with rankings in decreasing order over 100 iterations. Figure S1: Log-transformed relative abundance of significantly differential genera between participants with high ( $\geq 23$ ) or low/normal ( $< 23$ ) BMI score in rural and/or urban groups, as determined by Linear discriminant analysis effect size (LEfSe). *Ruminococcaceae* (group UCG-014) was significantly different in comparisons using both all samples and within urban samples alone. Figure S2: A) Principal component analysis (PCA) score plot performed on the overweight vs. normal population demonstrating clustering of subjects within samples. BMI ranges were pre-defined using WHO Asian BMI classifications: underweight normal (18.5–22.9); overweight (23–24.9) categories. PCA analysis used five selected discriminatory features (IFN-gamma, C-peptide, *Lachnospira*, *Bifidobacterium*, *Lachnoclostridium*) between normal vs. overweight using elastic net analysis. Overweight samples (in green) are more dispersed compared to the normal samples (in red). B) Area under the curve (AUC) is shown using the logistic regression method and five discriminatory features from the elastic net method. A corresponding confidence interval is also calculated as a shaded area. Figure S3: A) Principal component analysis (PCA) score plot performed on the pre-obese vs. normal population demonstrating clustering of subjects within samples. BMI ranges were pre-defined using WHO Asian BMI classifications: normal (18.5–22.9); pre-obese (25–29.9) categories. PCA analysis used twelve selected discriminatory features (*Prevotellaceae\_NK3B31\_group*, *Dialister*, Glucagon, C-peptide, *Prevotella\_2*, Antennary fucosylation (AF), CAG-56, *Muribaculaceae\_unclassified*, *Haemophilus*, *Sutterella*, *Treponema\_2*, *Gastranaerophilales\_unclassified*) between Normal vs. Pre-Obese using elastic net analysis. Pre-Obese (in green) and normal BMI samples (in red) seem to be separating from each other. B) Area under the curve (AUC) is shown using the logistic regression method and twelve discriminatory features from the elastic net method. A corresponding confidence interval is also calculated as a shaded area.

**Author Contributions:** Conceptualization, T.M.M., C.P., A.A., R.S.K.; methodology, B.H.M., A.M.S., J.M.B., M.P.-B., C.P.; software, A.A., N.X., T.D.; validation, A.A., C.P.; formal analysis, B.H.M., J.M.B., A.M.S., F.V., N.X., T.D., A.A., C.P.; investigation, R.N.B., R.R.N., S.S.J., B.H.M., A.M.S., J.M.B., J.R.M., T.o.Y., N.C., M.H., F.K., J.A.K.M.; resources, T.M.M.; data curation, T.M.M., R.N.B.; writing—original draft preparation, T.M.M.; writing—review and editing, B.H.M., A.M.S., J.R.M., M.P.-B., G.L., S.A., A.S., C.P., A.A., R.S.K., J.A.K.M.; visualization, T.M.M., A.A., A.M.S., B.H.M., J.M.B.; supervision, T.M.M., R.S.K.; project administration, T.M.M., R.S.K.; funding acquisition, T.M.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by a University of Nottingham Anne McLaren Fellowship and Research Priority Area (RPA) grant to T.M.M. and supplemented by the National Institute for Health Research (NIHR) Nottingham Digestive Diseases Biomedical Research Centre based at Nottingham University Hospitals NHS Trust and University of Nottingham, Litwin Initiative at the Crohn’s and Colitis Foundation to C.P, NTU Quality Research (QR) funds to C.P. and M.H, and B.H.M. is the recipient of an NIHR Academic Clinical Lectureship (reference: CL-2019-21-002). A.A. was supported by National Institute for Health Research (NIHR) Surgical Reconstruction and Microbiology Research Centre (SRMRC), Birmingham, UK. The views expressed in this publication

are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, the Medical Research Council or the Department of Health, UK. The Division of Digestive Diseases at Imperial College London receives financial support from the NIHR Imperial Biomedical Research Centre. Metabonomics studies were performed at the MRC-NIHR National Phenome Centre at Imperial College London; they receive financial support from the MRC and NIHR (grant number: MC\_PC\_12025). Genos Ltd. receives financial support from European Structural and Investment Funds CEKOM grant (#KK.01.2.2.03.0006), IRI grant (#KK.01.2.1.01.0003), and Croatian National Centre of Research Excellence in Personalized Healthcare grant (#KK.01.1.1.01.0010). The funders were not involved in study design, writing the report or decision for publication.

**Institutional Review Board Statement:** Institutional ethics for this project was approved by Faculty of Medicine and Health Sciences Research Ethics Committee of the University of Nottingham on 18 January 2019. Ethics Reference number: 199-1901.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Sequencing data from this study (in fastq-format) are publicly available for download at the European Nucleotide Archive (ENA) database using study accession number PRJEB42528 (<http://www.ebi.ac.uk/ena/data/view/PRJEB42528> (accessed on 15 February 2021)).

**Acknowledgments:** We acknowledge Melanie Lingaya for her technical assistance in sample preparation. We would also like to thank the village health care workers of the rural sampling sites who facilitated patient consent and sample acquisition with the assistance of the project fellows for this study. Finally, we are grateful to our participants that made this research possible.

**Conflicts of Interest:** T.M. has received consultancy fees from Takeda. B.H.M. has received consultancy fees from Finch Therapeutics Group. J.R.M. has received consultancy fees from Cultech Ltd., and Enterobiotix Ltd. G.L. is founder and CEO of Genos, a private research organization that specializes in high-throughput glycomic analysis and has several patents in this field. M.P.-B., F.K. and F.V. are employees of Genos. The remaining authors declare no competing interests.

## References

1. Mohan, P.; Mohan, S.B.; Dutta, M. Communicable or non-communicable diseases? Building strong primary health systems to address double burden of disease in India. *J. Fam. Med. Prim. Care* **2019**, *8*, 326–329. [[CrossRef](#)]
2. Arokiasamy, P. India's escalating burden of non-communicable diseases. *Lancet Glob. Health* **2018**, *6*, e1262–e1263. [[CrossRef](#)]
3. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **2017**, *390*, 1211–1259. [[CrossRef](#)]
4. India State-level Disease Burden Initiative Collaborators. Nations within a nation: Variations in epidemiological transition across the states of India, 1990–2016 in the Global Burden of Disease Study. *Lancet* **2017**, *390*, 2427–2460.
5. Lumeng, C.N.; Saltiel, A.R. Inflammatory links between obesity and metabolic disease. *J. Clin. Investig.* **2011**, *121*, 2111–2117. [[CrossRef](#)] [[PubMed](#)]
6. Furman, D.; Campisi, J.; Verdin, E.; Carrera-Bastos, P.; Targ, S.; Franceschi, C.; Ferrucci, L.; Gilroy, D.W.; Fasano, A.; Miller, G.W.; et al. Chronic inflammation in the etiology of disease across the life span. *Nat. Med.* **2019**, *25*, 1822–1832. [[CrossRef](#)] [[PubMed](#)]
7. India State-Level Disease Burden Initiative Diabetes Collaborators. The increasing burden of diabetes and variations among the states of India: The Global Burden of Disease Study 1990–2016. *Lancet Glob. Health* **2018**, *6*, e1352–e1362. [[CrossRef](#)]
8. International Diabetes Federation. 9th Edition. International Diabetes Federation; 2019. IDF Diabetes Atlas. Available online: <https://www.diabetesatlas.org/en/> (accessed on 15 February 2021).
9. Ramachandran, A.; Snehalatha, C.; Kapur, A.; Vijay, V.; Mohan, V.; Das, A.K.; Rao, P.V.; Yajnik, C.S.; Prasanna Kuman, K.M.; Nair, J.D.; et al. High prevalence of diabetes and impaired glucose tolerance in India: National Urban Diabetes Survey. *Diabetologia* **2001**, *44*, 1094–1101. [[CrossRef](#)]
10. Deepa, M.; Grace, M.; Binukumar, B.; Pradeepa, R.; Roopa, S.; Khan, H.M.; Fatmi, Z.; Kadir, M.M.; Naeem, I.; Ajay, V.S.; et al. High burden of prediabetes in three large cities in South Asia: The Center for Cardio-metabolic Risk Reduction in South Asia (CARRS) Study. *Diabetes Res. Clin. Pract.* **2015**, *110*, 172–182. [[CrossRef](#)]
11. Anjana, R.M.; Deep, M.; Pradeepa, R. Prevalence of diabetes and prediabetes in 15 states of India: Results from the ICMR-INDIAB population-based cross-sectional study. *Lancet Diabetes Endocrinol.* **2017**, *5*, 585–596. [[CrossRef](#)]
12. Ahluwalia, T.S.; Kilpelainen, T.O.; Singh, S.; Rossing, P. Editorial: Novel Biomarkers for Type 2 Diabetes. *Front. Endocrinol.* **2019**, *10*, 649. [[CrossRef](#)]
13. Yao, M.J.; Li, J.Y.; Li, J.Z.; Wu, T.F.; Xu, J.-H.; Huang, C.Z.; Cheng, D.; Chen, Q.K.; Yu, T. Diabetes mellitus increases the risk of enteric infections: A meta-analysis. *Int. J. Clin. Exp. Med.* **2018**, *11*, 5457–5468.

14. Thaiss, C.A.; Levy, M.; Grosheva, I.; Zheng, D.; Soffer, E.; Blacher, E.; Braverman, S.; Tengeler, A.C.; Barak, O.; Elazer, M.; et al. Hyperglycemia drives intestinal barrier dysfunction and risk for enteric infection. *Science* **2018**, *359*, 1376–1383. [[CrossRef](#)] [[PubMed](#)]
15. Dotz, V.; Wuhrer, M. N-glycome signatures in human plasma: Associations with physiology and major diseases. *FEBS Lett.* **2019**, *593*, 2966–2976. [[CrossRef](#)]
16. Lauc, G.; Pezer, M.; Rudan, I.; Campbell, H. Mechanisms of disease: The human N-glycome. *Biochim. Biophys. Acta* **2016**, *1860*, 1574–1582. [[CrossRef](#)] [[PubMed](#)]
17. Keser, T.; Gornik, I.; Vuckovic, F.; Selek, N.; Pavic, T.; Lukic, E.; Gudelj, I.; Gasparovic, H.; Biocina, B.; Tilin, T.; et al. Increased plasma N-glycome complexity is associated with higher risk of type 2 diabetes. *Diabetologia* **2017**, *60*, 2352–2360. [[CrossRef](#)] [[PubMed](#)]
18. Rudman, N.; Gornik, O.; Lauc, G. Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes. *FEBS Lett.* **2019**, *593*, 1598–1615. [[CrossRef](#)]
19. Gudelj, I.; Lauc, G.; Pezer, M. Immunoglobulin G glycosylation in aging and diseases. *Cell Immunol.* **2018**, *333*, 65–79. [[CrossRef](#)] [[PubMed](#)]
20. Gurung, M.; Li, Z.; You, H.; Rodrigues, R.; Jump, D.B.; Morgun, A.; Shulzhenko, N. Role of gut microbiota in type 2 diabetes pathophysiology. *EBioMedicine* **2020**, *51*, 102590. [[CrossRef](#)] [[PubMed](#)]
21. Guasch-Ferre, M.; Hruby, A.; Toledo, E.; Clish, C.B.; Martinez-Gonzalez, M.A.; Salas-Salvado, J.; Hu, F.B. Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. *Diabetes Care* **2016**, *39*, 833–846. [[CrossRef](#)]
22. Arneth, M.; Arneth, R.; Shams, M. Metabolomics of Type 1 and Type 2 Diabetes. *Int. J. Mol. Sci.* **2019**, *20*, 2467. [[CrossRef](#)]
23. Zhou, W.; Sailani, M.R.; Contrepoi, K.; Zhou, Y.; Ahadi, S.; Leopold, S.R.; Zhang, M.J.; Rao, V.; Avina, M.; Mishra, T.; et al. Longitudinal multi-omics of host-microbe dynamics in prediabetes. *Nature* **2019**, *569*, 663–671. [[CrossRef](#)] [[PubMed](#)]
24. Vallianou, N.; Stratigou, T.; Christodoulatos, G.S.; Dalamaga, M. Understanding the Role of the Gut Microbiome and Microbial Metabolites in Obesity and Obesity-Associated Metabolic Disorders: Current Evidence and Perspectives. *Curr. Obes. Rep.* **2019**, *8*, 317–332. [[CrossRef](#)] [[PubMed](#)]
25. Mullish, B.J.; Pechlivanis, A.; Barker, G.F.; Thursz, M.R.; Marchesi, J.R.; McDonald, J.A.K. Functional Microbiomics: Evaluation of Gut Microbiota-Bile Acid Metabolism Interactions in Health and Disease. *Methods* **2018**, *149*, 49–58. [[CrossRef](#)]
26. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.; Holmes, S.P. DADA2: High resolution samples inference from Illumina amplicon data. *Nat. Methods* **2016**, *13*, 581–583. [[CrossRef](#)]
27. Oksanen, J.; Blanchett, F.G.; Kindt, R.; Legendre, P.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; Stevens, M.H.H.; Wagner, H. The vegan package. *Community Ecol. Package* **2014**. R Package Version 2.2-0. Available online: <http://CRAN.Rproject.org/package=vegan> (accessed on 15 February 2021).
28. Kaufman, L.; Rousseeuw, P.J. Partitioning around Medoids (Program PAM). In *Finding Groups in Data: An Introduction to Cluster Analysis*; Kaufman, L., Rousseeuw, P.J., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1990; pp. 68–125.
29. Segata, N.; Izard, J.; Waldron, L.; Gevers, D.; Miropolsky, L.; Garrett, W.S.; Huttenhower, C. Metagenomic biomarker discovery and explanation. *Genome Biol.* **2011**, *12*, R60. [[CrossRef](#)] [[PubMed](#)]
30. Schloss, P.D.; Westcott, S.L. Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis. *Appl. Environ. Microbiol.* **2011**, *77*, 3219–3226. [[CrossRef](#)]
31. Moreau, N.M.; Goupy, S.M.; Antignac, J.P.; Monteau, F.J.; Le Bizec, B.J.; Champ, M.M.; Martin, L.J.; Dumon, H.J. Simultaneous measurement of plasma concentrations and <sup>13</sup>C-enrichment of short-chain fatty acids, lactic acid and ketone bodies by gas chromatography coupled to mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2003**, *784*, 395–403. [[CrossRef](#)]
32. Ugrina, I.; Campbell, H.; Vučković, F. Laboratory Experimental Design for a Glycomic Study. In *High-Throughput Glycomics and Glycoproteomics*; Lauc, G., Wuhrer, M., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2017; Volume 1503, pp. 13–19.
33. Akmačić, I.T.; Ugrina, I.; Štambuk, J.; Gudelj, I.; Vučković, F.; Lauc, G.; Pučić-Baković, M. High-throughput glycomics: Optimization of sample preparation. *Biochemistry* **2015**, *80*, 934–942. [[PubMed](#)]
34. Simurina, M.; de Haan, N.; Vuckovic, F.; Kennedy, N.A.; Štambuk, J.; Falck, D.; Trbojević-Akmačić, I.; Clerc, F.; Razdorov, G.; Khon, A.; et al. Glycosylation of Immunoglobulin G Associates with Clinical Features of Inflammatory Bowel Diseases. *Gastroenterology* **2018**, *154*, 1320–1333.e.10. [[CrossRef](#)]
35. Balbin, M.; Grubb, A.; de Lange, G.G.; Grubb, R. DNA sequences specific for Caucasian G3m(b) and (g) allotypes: Allotyping at the genomic level. *Immunogenetics* **1994**, *39*, 187–193. [[CrossRef](#)] [[PubMed](#)]
36. Johnson, M.P.; Keyho, R.; Blackburn, N.B.; Laston, S.; Kumar, S.; Peralta, J.; Thapa, S.S.; Towne, B.; Subedi, J.; Blangero, J.; et al. Glycated Serum Protein Genetics and Pleiotropy with Cardiometabolic Risk Factors. *J. Diabetes Res.* **2019**, *2019*, 2310235. [[CrossRef](#)] [[PubMed](#)]
37. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc.* **2005**, *67*, 301–320. [[CrossRef](#)]
38. Bravo-Merodio, L.; Williams, J.A.; Gkoutos, G.C.; Acharjee, A. Omics biomarker identification pipeline for translational medicine. *J. Transl. Med.* **2019**, *17*, 155. [[CrossRef](#)]
39. Bravo-Merodio, L.; Acharjee, A.; Hazeldine, J.; Bentley, C.; Foster, M.; Gkoutos, G.V.; Lord, J.M. Machine Learning for the detection of early immunological markers as predictors of multi-organ dysfunction. *Sci. Data* **2019**, *6*, 328. [[CrossRef](#)]



40. Stapleton, C.J.; Acharjee, A.; Irvine, H.J.; Wolcott, Z.C.; Patel, A.B.; Kinberly, W.T. High-throughput metabolite profiling: Identification of plasma taurine as a potential biomarker of functional outcome after aneurysmal subarachnoid hemorrhage. *J. Neurosurg.* **2020**, *133*, 1635–1978. [[CrossRef](#)] [[PubMed](#)]
41. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna: R Foundation for Statistical Computing. 2013. Available online: <http://www.R-project.org/> (accessed on 15 February 2021).
42. Hong, J.; Wishart, D.S.; Xia, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr. Protoc. Bioinform.* **2019**, *68*, e86.
43. Barengolts, E. Gut microbiota, Prebiotics, Probiotics, and Synbiotics in Management of Obesity and Prediabetes: Review of Randomized controlled Trials. *Endocr. Pract.* **2016**, *22*, 1224–1234. [[CrossRef](#)]
44. De Groot, P.; Nikolic, T.; Pellegrini, S.; Sordi, V.; Imangaliyev, S.; Rampanelli, E.; Hanssen, N.; Attaye, I.; Bakker, G.; Duinkerken, G.; et al. Faecal microbiota transplantation halts progression of human new-onset type 1 diabetes in a randomised controlled trial. *Gut* **2020**, *70*, 92–105. [[CrossRef](#)]
45. Ganz, M.L.; Wintfeld, N.; Li, Q.; Ala, V.; Langer, J.; Hammer, M. The association of body mass index with the risk of type 2 diabetes: A case-control study nested in an electronic health records system in the United States. *Diabetol. Metab. Syndr.* **2014**, *6*, 50. [[CrossRef](#)]
46. Eckel, R.H.; Kahn, S.E.; Ferrannini, E.; Goldfine, A.B.; Nathan, D.M.; Schwartz, M.W.; Smith, R.J.; Smith, S.R. Obesity and Type 2 Diabetes: What Can be Unified and What Needs to be Individualized? *J Clin Endocrinol Metab.* **2011**, *96*, 1654–1663. [[CrossRef](#)] [[PubMed](#)]
47. Al-Goblan, A.S.; Al-Alfi, M.; Khan, M.Z. Mechanisms linking diabetes mellitus and obesity. *Diabetes Metab. Syndr. Obes.* **2014**, *7*, 587–591. [[CrossRef](#)] [[PubMed](#)]
48. Martin, R.; Miquel, S.; Benevides, L.; Bridonneau, C.; Robert, V.; Hudault, S.; Chain, F.; Berteau, O.; Azevedo, V.; Chatel, J.M.; et al. Functional Characterization of Novel *Faecalibacterium prausnitzii* Strains Isolated from Healthy Volunteers: A Step Forward in the Use of *F. prausnitzii* as a Next Generation Probiotic. *Front. Microbiol.* **2017**, *8*, 1226. [[CrossRef](#)] [[PubMed](#)]
49. Parada Venegas, D.; De la Funete, M.; Landskron, G.; Gonzalez, M.J.; Quera, R.; Dijkstra, G.; Harmsen, H.J.M.; Faber, K.N.; Hermoso, M.A. Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and its Relevance for Inflammatory Bowel Diseases. *Front. Immunol.* **2019**, *10*, 277. [[CrossRef](#)] [[PubMed](#)]
50. Monaghan, T.M.; Sloan, T.J.; Stockdale, S.R.; Blanchard, A.M.; Emes, R.D.; Wilcox, M.; Biswas, R.; Nashine, R.; Manke, S.; Gandhi, J.; et al. Metagenomics reveals impact of geography and acute diarrheal disease on the Central Indian human gut microbiome. *Gut Microbes* **2020**, *12*, 1752606. [[CrossRef](#)] [[PubMed](#)]
51. Gronwall, C.; Vas, J.; Silverman, G.J. Protective roles of natural IgM antibodies. *Front. Immunol.* **2012**, *3*, 66. [[CrossRef](#)]
52. Gronwall, C.; Silverman, G.J. Natural IgM: Beneficial autoantibodies for the control of inflammatory and autoimmune disease. *J. Clin. Immunol.* **2014**, *34*, S12–S21. [[CrossRef](#)]
53. Tachang, G.K. Metabolic Syndrome May be a Sign of Rapid Aging. *J. Diabetes Metab.* **2016**, *7*, 5. [[CrossRef](#)]
54. Cheema, A.; Adeloje, D.; Sidhu, S.; Sridhar, D.; Chan, K.Y. Urbanization and prevalence of type 2 diabetes in Southern Asia: A systematic analysis. *J. Glob. Health* **2014**, *4*, 010404. [[CrossRef](#)]
55. Krishnamoorthy, Y.; Rajaa, S.; Murali, S.; Rehman, T.; Sahoo, J.; Kar, S.S. Prevalence of metabolic syndrome among adult population in India: A systematic review and meta-analysis. *PLoS ONE* **2020**, *15*, e0240971. [[CrossRef](#)] [[PubMed](#)]
56. Monaghan, T.M.; Pucic-Bakovic, M.; Vuckovic, F.; Lee, C.; Kao, D. Human Glycome Project. Decreased Complexity of Serum N-glycan Structures Associates with successful Fecal Microbiota Transplantation for Recurrent *Clostridioides difficile* infection. *Gastroenterology* **2019**, *157*, 1676–1678. [[CrossRef](#)]
57. Obregon-Tito, A.J.; Tito, R.Y.; Metcalf, J.; Sankaranarayan, K.; Clement, J.C.; Ursell, L.K.; Xu, Z.Z.; Van Treuren, W.; Knight, R.; Gaffney, P.M.; et al. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat. Comm.* **2015**, *25*, 6505. [[CrossRef](#)] [[PubMed](#)]
58. Ou, J.; Carbonero, F.; Zoetendal, E.G.; Delany, J.P.; Wang, M.; Newton, K.; Gaskins, H.R.; O’Keefe, S.J. Diet, microbiota, and microbial metabolites on colon cancer in rural Africans and African Americans. *Am. J. Clin. Nutr.* **2013**, *98*, 111–120. [[CrossRef](#)] [[PubMed](#)]
59. Schnorr, S.L. The diverse microbiome of the hunter-gatherer. *Nature* **2015**, *518*, S14–S15. [[CrossRef](#)]
60. Kim, M.; Qie, Y.; Park, J.; Kim, C.H. Gut Microbial Metabolites Fuel Antibody Responses. *Cell Host Microbe* **2016**, *20*, 202–214. [[CrossRef](#)] [[PubMed](#)]

# **A multi-factorial observational study on sequential fecal microbiota transplant in patients with medically refractory *Clostridioides difficile* infection**

Tanya M Monaghan\*<sup>#</sup>, Niharika A Duggal\*, Elisa Rosati\*, Ruth Griffin, Jamie Hughes, Brandi Roach, David Y Yang, Christopher Wang, Karen Wong, Lynora Saxinger, Maja Pučić-Baković, Frano Vučković, Filip Klicek, Gordan Lauc, Paddy Tighe, Benjamin H Mullish, Jesus Miguens Blanco, Julie AK McDonald, Julian R Marchesi, Ning Xue, Tania Dottorini, Animesh Acharjee, Andre Franke, Yingrui Li, Gane Ka-Shu Wong, Christos Polytarchou, **Tung On Yau**, Niki Christodoulou, Maria Hatziapostolou, Minkun Wang, Lindsey A Russell<sup>#</sup>, Dina H Kao<sup>#</sup>

*Cells* **2021**, 10(11), 3234.

Journal URL: <https://www.mdpi.com/2073-4409/10/11/3234>

DOI: [10.3390/cells10113234](https://doi.org/10.3390/cells10113234)

PMID: [34831456](https://pubmed.ncbi.nlm.nih.gov/34831456/)

PMCID: [PMC8624539](https://pubmed.ncbi.nlm.nih.gov/pmc/articles/PMC8624539/)



**University of  
Nottingham**  
UK | CHINA | MALAYSIA

Dr Tanya Monaghan  
Faculty of Medicine & Health Sciences  
Room W/E 1381 NDDC  
Queen's Medical Centre  
Nottingham  
NG7 2UH  
UK

[Tanya.Monaghan@nottingham.ac.uk](mailto:Tanya.Monaghan@nottingham.ac.uk)  
+44(0)115 9249924 x60589

April 12<sup>th</sup>, 2022

Re: Full Publications

To Whom It May Concern:

I am writing to confirm that Tung On (Payton) Yau, previously a Research Fellow in the laboratory of Dr Christos Polytarchou at Nottingham Trent University (currently serving as an Hourly-Paid Lecturer at Nottingham Trent University and a Post-doctoral Researcher at Scotland's Rural College), collaborated with my team on projects relating to microRNAs, microbiota, public health and *Clostridioides difficile* infection from the year 2017 to 2021. He was a significant contributor, with a role in literature search and acquisition, experimental design, clinical sample management, and performing experiments. These experiments include cell culture, RNA extraction from serum and peripheral blood mononuclear cells, high-throughput Nanostring miRNA analysis, Bio-Plex cytokine assays, and luciferase reporter assays for the identification of miRNA targets. He was subsequently involved in data analysis and visual representations of the data for publications.

He is a co-author of three published publications:

1. Fecal microbiota transplantation for recurrent *Clostridioides difficile* infection associates with functional alterations in circulating microRNAs. **Gastroenterology** **161** (1), 255-270
2. Multiomics profiling reveals signatures of dysmetabolism in urban populations in Central India. **Microorganisms** **9** (7), 1485
3. A multi-factorial observational study on sequential fecal microbiota transplant in patients with medically refractory *Clostridioides difficile* infection. **Cells** **10** (11), 3234

Yours Sincerely,

Article

# A Multi-Factorial Observational Study on Sequential Fecal Microbiota Transplant in Patients with Medically Refractory *Clostridioides difficile* Infection

Tanya M. Monaghan <sup>1,2,\*</sup>, Niharika A. Duggal <sup>3,†</sup>, Elisa Rosati <sup>4,†</sup>, Ruth Griffin <sup>2,5</sup>, Jamie Hughes <sup>5</sup>, Brandi Roach <sup>6</sup>, David Y. Yang <sup>6</sup>, Christopher Wang <sup>6</sup>, Karen Wong <sup>6</sup>, Lynora Saxinger <sup>7</sup>, Maja Pučić-Baković <sup>8</sup>, Frano Vučković <sup>8</sup>, Filip Klicek <sup>8</sup>, Gordan Lauc <sup>8,9</sup>, Paddy Tighe <sup>10</sup>, Benjamin H. Mullish <sup>11</sup>, Jesus Miguens Blanco <sup>11</sup>, Julie A. K. McDonald <sup>11,12</sup>, Julian R. Marchesi <sup>11</sup>, Ning Xue <sup>13</sup>, Tania Dottorini <sup>13</sup>, Animesh Acharjee <sup>14</sup>, Andre Franke <sup>4</sup>, Yingrui Li <sup>15</sup>, Gane Ka-Shu Wong <sup>16</sup>, Christos Polytarchou <sup>17</sup>, Tung On Yau <sup>17</sup>, Niki Christodoulou <sup>17</sup>, Maria Hatziapostolou <sup>17</sup>, Minkun Wang <sup>15,18,\*</sup>, Lindsey A. Russell <sup>19,\*</sup> and Dina H. Kao <sup>6,\*</sup>



**Citation:** Monaghan, T.M.; Duggal, N.A.; Rosati, E.; Griffin, R.; Hughes, J.; Roach, B.; Yang, D.Y.; Wang, C.; Wong, K.; Saxinger, L.; et al. A Multi-Factorial Observational Study on Sequential Fecal Microbiota Transplant in Patients with Medically Refractory *Clostridioides difficile* Infection. *Cells* **2021**, *10*, 3234. <https://doi.org/10.3390/cells10113234>

Academic Editors: Walter Wahli and Pengfei Xu

Received: 16 October 2021  
Accepted: 17 November 2021  
Published: 19 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- 1 NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham NG7 2UH, UK
- 2 Nottingham Digestive Diseases Centre, School of Medicine, University of Nottingham, Nottingham NG7 2UH, UK; ruth.griffin1@nottingham.ac.uk
- 3 MRC-Arthritis Research UK Centre for Musculoskeletal Ageing Research, Institute of Inflammation and Ageing, University of Birmingham, Birmingham B15 2TT, UK; n.arora@bham.ac.uk
- 4 Institute of Clinical Molecular Biology, Universitätsklinikum Schleswig-Holstein, Christian-Albrecht University of Kiel, 24105 Kiel, Germany; e.rosati@ikmb.uni-kiel.de (E.R.); a.franke@mucosa.de (A.F.)
- 5 Synthetic Biology Research Centre, The University of Nottingham Biodiscovery Institute, University of Nottingham, Nottingham NG7 2RD, UK; svzjh2@exmail.nottingham.ac.uk
- 6 Division of Gastroenterology, Department of Medicine, University of Alberta; Edmonton, Alberta, AB T6G 2G3, Canada; brandi.roach@albertahealthservices.ca (B.R.); dyayang@ualberta.ca (D.Y.Y.); cwang1@ualberta.ca (C.W.); kwong3@ualberta.ca (K.W.)
- 7 Division of Infectious Diseases, Department of Medicine, University of Alberta; Edmonton, Alberta, AB T6G 2G3, Canada; saxinger@ualberta.ca
- 8 Glycoscience Research Laboratory, Genos Ltd., Borongajska cesta 83H, 10000 Zagreb, Croatia; mpucicbakovic@genos.hr (M.P.-B.); fvuckovic@genos.hr (F.V.); fklicek@genos.hr (F.K.); glauc@genos.hr (G.L.)
- 9 Faculty of Pharmacy and Biochemistry, University of Zagreb, 10000 Zagreb, Croatia
- 10 School of Life Sciences, University of Nottingham, Nottingham NG7 2RD, UK; mrzpj1@exmail.nottingham.ac.uk
- 11 Division of Digestive Diseases, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London SW7 2AZ, UK; b.mullish@imperial.ac.uk (B.H.M.); j.miguens-blanco18@imperial.ac.uk (J.M.B.); julie.mcdonald@imperial.ac.uk (J.A.K.M.); j.marchesi@imperial.ac.uk (J.R.M.)
- 12 MRC Centre for Molecular Bacteriology and Infection, Imperial College London, London SW7 2AZ, UK
- 13 School of Veterinary Medicine and Science, University of Nottingham, Nottingham NG7 2UH, UK; ning.xue@microlise.com (N.X.); svztd@exmail.nottingham.ac.uk (T.D.)
- 14 College of Medical and Dental Sciences, Institute of Cancer and Genomic Sciences, Centre for Computational Biology, University of Birmingham, Birmingham B15 2TT, UK; a.acharjee@bham.ac.uk
- 15 Shenzhen Digital Life Institute, Shenzhen 518016, China; liyr@icarbonx.com
- 16 Department of Biological Sciences, Department of Medicine, University of Alberta, Edmonton, AB T6G 2E1, Canada; gane@ualberta.ca
- 17 Department of Biosciences, John van Geest Cancer Research Centre, Centre for Health Aging and Understanding Disease, School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK; Christos.polytarchou@ntu.ac.uk (C.P.); payton.yau@ntu.ac.uk (T.O.Y.); niki.christodoulou2019@my.ntu.ac.uk (N.C.); maria.hatziapostolou@ntu.ac.uk (M.H.)
- 18 Innovation Lab, Innovent Biologics, Inc., Suzhou 215011, China
- 19 Division of Gastroenterology, Department of Medicine, McMaster University, Hamilton, ON L8N 3Z5, Canada
- \* Correspondence: tanya.monaghan@nottingham.ac.uk (T.M.M.); minkunw@gmail.com (M.W.); lindsey.russell@medportal.ca (L.A.R.); dkao@ualberta.ca (D.H.K.); Tel.: +115-8231090 (T.M.M.)
- † These authors contributed equally to this work.

**Abstract:** Fecal microbiota transplantation (FMT) is highly effective in recurrent *Clostridioides difficile* infection (CDI); increasing evidence supports FMT in severe or fulminant *Clostridioides difficile*

infection (SFCDI). However, the multifactorial mechanisms that underpin the efficacy of FMT are not fully understood. Systems biology approaches using high-throughput technologies may help with mechanistic dissection of host-microbial interactions. Here, we have undertaken a deep phenomics study on four adults receiving sequential FMT for SFCDI, in which we performed a longitudinal, integrative analysis of multiple host factors and intestinal microbiome changes. Stool samples were profiled for changes in gut microbiota and metabolites and blood samples for alterations in targeted epigenomic, metabonomic, glycomic, immune proteomic, immunophenotyping, immune functional assays, and T-cell receptor (TCR) repertoires, respectively. We characterised temporal trajectories in gut microbial and host immunometabolic data sets in three responders and one non-responder to sequential FMT. A total of 562 features were used for analysis, of which 78 features were identified, which differed between the responders and the non-responder. The observed dynamic phenotypic changes may potentially suggest immunosenescent signals in the non-responder and may help to underpin the mechanisms accompanying successful FMT, although our study is limited by a small sample size and significant heterogeneity in patient baseline characteristics. Our multi-omics integrative longitudinal analytical approach extends the knowledge regarding mechanisms of efficacy of FMT and highlights preliminary novel signatures, which should be validated in larger studies.

**Keywords:** fecal microbiota transplantation; *Clostridioides difficile*; immunosenescence; host-microbial interactions; systems biology

## 1. Introduction

*Clostridioides difficile* infection (CDI) is the most common cause of diarrhea acquired in acute healthcare settings. Hospital-acquired CDI increases healthcare costs 4 times over matched hospitalization, resulting in an added annual cost of \$1.1 billion in North America [1,2]. Most patients with CDI have mild to moderate disease and respond to either oral vancomycin or fidaxomicin; a subset of these patients may develop recurrent infections necessitating a different therapeutic approach [3]. Severe infection is defined by an elevated white blood cell count of over 15,000 cells/mL or serum creatinine level >1.5 mg/dL, while fulminant infection is characterized by hypotension or shock, ileus or toxic megacolon [3]. These patients usually require hospital admission for treatment with oral vancomycin and intravenous metronidazole or surgery if refractory to medical therapy. In CDI, perturbation of the gut microbiome has a clear causation in disease pathogenesis, triggered by antibiotic exposure [4,5]. Restitution of the gut microbiome with fecal microbiota transplant (FMT) is highly effective in the treatment of recurrent CDI. To prevent mild-moderate CDI recurrence, a single FMT is administered once vancomycin has been discontinued, and the success rate is in the range of 80–90% [6]. In patients with antibiotic-refractory severe or fulminant CDI (SFCDI) who are poor surgical candidates, treatment options are limited and sequential FMT by colonoscopy with concomitant vancomycin has been shown to be effective in several small case series and a single randomized trial [7–10]. However, despite its effectiveness, significant knowledge gaps remain in our understanding of how FMT exerts these beneficial effects, and what molecular features, particularly immunological, may predict treatment outcomes in this unique population of antibiotic refractory SFCDI [11].

Previous data generated by our laboratory and collaborative network has demonstrated that successful FMT for mild-moderate CDI is associated with significantly decreased stool levels of the primary bile acids chenodeoxycholic acid and cholic acid, and significantly increased levels of the secondary bile acids deoxycholic acid and lithocholic acid [12]. In addition to restoration of gut microbiota and bile acid profiles, these findings were accompanied by increased levels of circulating fibroblast growth factor (FGF)-19, consistent with the upregulation of the farnesoid X receptor (FXR)-fibroblast growth factor (FGF) pathway [12]. Interestingly, microbially mediated production of certain secondary bile acids, which predominate in post-FMT stool [13–15], has been demonstrated to promote the generation of peripheral regulatory T cells [16], linking these metabolites with

colonic immunity. Furthermore, successful FMT for recurrent CDI is also associated with restoration of short chain fatty acids (SCFA) and inhibition of *C. difficile* growth [17,18]. Similar to secondary bile acids, SCFA can regulate colonic regulatory T cell and modulate regulatory B cell immunosuppressive function in mice, which has been directly demonstrated to be a protective mechanism against colitis [19]. Our group has also detected alterations in the circulating host *N*-glycome with successful FMT for recurrent CDI [20].

Beyond gut microbiota-metabolite changes, there is compelling evidence that the immune response to *C. difficile* is a predominant factor determining clinical outcome. A picture is emerging of an exaggerated host inflammatory immune response particularly in the context of severe disease [21]. By contrast, natural antibody responses to *C. difficile* toxins, the major virulence factors associated with disease pathogenesis in shaping clinical outcomes, have been conflicting [5], suggesting that antibody titers may not be an accurate predictor of natural immunity to CDI and recurrence [22]. Nonetheless, the recent success of bezlotoxumab (a human monoclonal antibody that binds and neutralizes *C. difficile* toxin B) and the promise of ongoing vaccine trials suggest that antibody, B cell, and T cell responses to toxin B (TcdB), in particular, contribute to the protection against CDI [22]. Although most adaptive immunity CDI research has focused on components of humoral immunity, we and others have detected TcdA- and TcdB-specific memory B cell responses in CDI patients and a murine model of *C. difficile* disease recurrence [23–25].

Recent technological advancements in high throughput next-generation sequencing technologies and omics-based sciences have expanded our fundamental understanding of the complex biology and pathogenesis of CDI. The study of *Clostridioides difficile*, which represents a severe perturbation of the gut microbiota, and the impact of sequential FMT, provides a unique opportunity to directly interrogate microbiome-metabonome-immune interactions directly in humans. Here, we present the observations of a highly multidimensional analysis in a small case series including four adults with antibiotic refractory SFCDI who were treated with sequential FMT during which blood and stool samples were collected prospectively. We analyzed multiple host factors and intestinal microbiome changes longitudinally and related these biological metrics to clinical outcomes in this distinctive population.

## 2. Materials and Methods

### 2.1. Study Cohort, Treatment Regimen and Outcome Definitions

A total of 4 patients with SFCDI were included in this study, of which 3 patients (patients 1–3) were enrolled into an open-label trial of sequential FMT by enemas with fidaxomicin. Patient 4 was treated with sequential FMT via colonoscopy as part of usual clinical care with vancomycin and metronidazole prior to the open-label study.

Three adult patients with SFCDI unresponsive to vancomycin and metronidazole were eligible for this open label study (NCT03760484) at the University of Alberta. The study protocol was approved by the local research ethics board (Pro81229) and Health Canada (Control#220509). Key exclusion criteria were bowel perforation and planned colectomy. Patients were treated with 2 cycles of FMTs and fidaxomicin; each cycle consists of 3 consecutive days of FMT with fecal slurry (day 1 = 720 mL, day 2 = 260 mL and day 3 = 180 mL) delivered by enema, with concomitant oral fidaxomicin 200 mg twice daily up to 10 days. Worsening clinical symptoms and/or elevation in inflammatory markers, including leukocyte count and C-reactive protein, triggered the second treatment cycle. The protocol was intentionally flexible since it is not always predictable how quickly or slowly a patient with SFCDI may respond to proposed intervention. We wanted to suppress *C. difficile* burden as much as possible with fidaxomicin prior to initiating the next FMT cycle, but we also did not want to persist in a fixed fidaxomicin treatment duration in a treatment cycle if a patient was not clinically improving. With resolution of diarrhea and normalization of inflammatory markers, fidaxomicin was discontinued and a final enema of 180 mL was administered. Each patient was then assessed clinically at 1, 2, 4 and 8 weeks after final FMT. Blood and stool samples were collected at screening, at the end

of each treatment cycle, and 1 and 2–4 weeks post-final enema and stored at  $-80^{\circ}\text{C}$  until analyses. Treatment success (response) was defined as resolution of diarrhea without the need for anti-CDI therapy 2 weeks after final FMT. Treatment failure (non-response) was defined as recurrence of diarrhea requiring anti-CDI therapy within 2 weeks of final FMT.

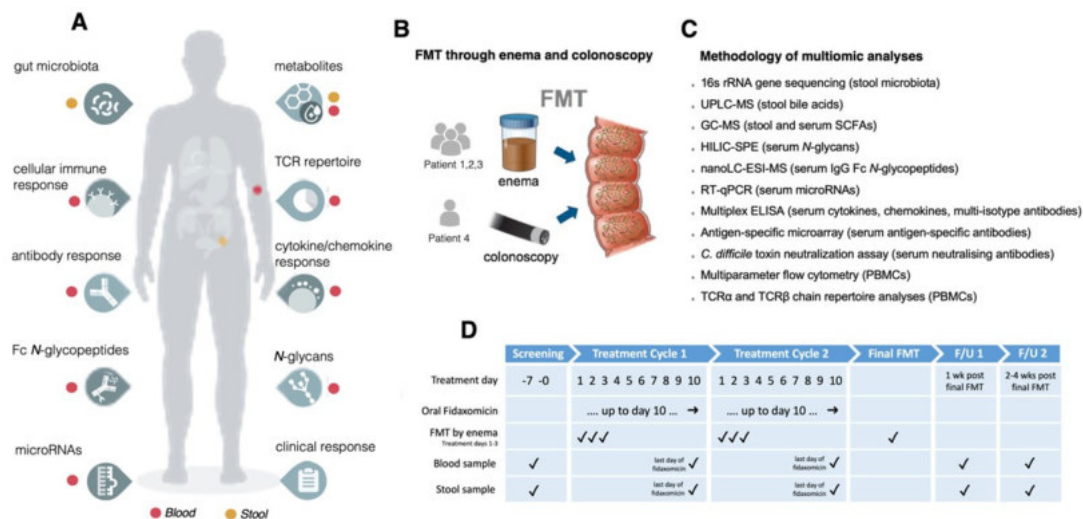
We also included a fourth patient with fulminant CDI who was treated with oral vancomycin and intravenous metronidazole and sequential FMT by colonoscopy every 5–7 days until resolution of pseudomembranous colitis [7]. Blood and stool samples were collected at similar intervals to the fidaxomicin patients.

## 2.2. FMT Preparation

A total of 3 universal stool donors registered with the Edmonton FMT program (2 females aged 56 and 37 years and 1 male aged 31 years) provided stool. The donor screening process complied with Health Canada regulations. Each donation of 100 g of stool was manufactured into 360 mL of fecal slurry as previously described [26].

## 2.3. Multiomics Studies

Whole blood was separated into peripheral blood mononuclear cells (PBMCs) and sera. Sera from all four patients was assayed for targeted epigenomic (microRNA; miR), metabonomic (short-chain fatty acids; SCFA), glycomic (total serum *N*-glycans, IgG Fc *N*-glycopeptides), and immune proteomic (cytokines, chemokines, total isotype, antigen-specific and neutralizing antibodies to *C. difficile* toxins) changes. Multi-parameter flow cytometry was used to profile PBMCs and the circulating PBMC TCR $\alpha$  and TCR $\beta$  repertoire in the three patients who were treated with the fidaxomicin protocol. Fecal samples were profiled for microbiota and metabolite (SCFA and bile acids; BAs) changes. Figure 1 describes the analytical pipeline.



**Figure 1.** Schematic of pipeline. (A) Multidimensional, longitudinal assays performed in patients receiving sequential fecal microbiota transplantation (FMT) either by enema or colonoscopy with severe or fulminant *Clostridioides difficile* infection; (B) Delivery route for each patient; (C) Methodologies utilized; (D) Treatment and sampling strategy with timelines.

## 2.4. 16S rRNA Gene Sequencing

DNA extraction methods are described in the Supplementary Methods. Gene-sequencing sample libraries for 16S rRNA gene were generated via Illumina's 16S Metagenomic Sequencing Library Preparation Protocol, but with modifications. Amplification was performed of the V1-V2 16S rRNA gene regions from the fecal DNA, using primers as previously described [27]. Products from the index PCR reactions were cleaned and normalized via the SequalPrep Normalization Plate Kit (ThermoFisherScientific, Waltham, MA, USA; library quantification was performed using the NEBNext Library Quant Kit for

Illumina (New England Biolabs, Ipswich, MA, USA). Sequencing data were obtained using paired-end 300 bp chemistry on an Illumina MiSeq (Illumina Inc, San Diego, CA, USA), with MiSeq Reagent Kit v.3 (Illumina). Processing of sequencing data was performed via the DADA2 pipeline as previously described [28], using the SILVA bacterial database v.132 (<https://www.arb-silva.de/>, accessed on 25 May 2021).

### 2.5. Metabolomic Analysis

Both ultra-performance liquid chromatography-mass spectrometry (UPLC-MS; for the profiling and analysis of fecal bile acids) and gas chromatography-mass spectrometry (GC-MS; for detection, identification and quantification of short chain fatty acids in feces and serum) were performed. Further details are within the Supplementary Methods.

### 2.6. Serum N-Glycome Analysis

Serum N-glycans were enzymatically released from proteins by PNGase F, fluorescently labelled with 2-aminobenzamide and cleaned up from the excess of reagents by hydrophilic interaction liquid chromatography solid phase extraction (HILIC-SPE), as previously described [29]. Fluorescently labelled and purified N-glycans were separated by HILIC on a Waters BEH Glycan chromatography column, 150 × 2.1 mm i.d., 1.7 µm BEH particles, installed on an Acquity ultra-performance liquid chromatography (UPLC) H-class system (Waters, Mississauga, ON, Canada) under the control of Empower 3 build 3471 software (Waters). Obtained chromatograms were separated into 39 chromatographic peaks, and the amount of N-glycans in each chromatographic peak was expressed as a percentage of total integrated area. From 39 directly measured glycan peaks, we calculated 12 derived traits which average particular glycosylation traits across different individual glycan structures and are, consequently, more closely related to individual enzymatic activities and underlying genetic polymorphisms. The derived traits used were the proportion of low branching (LB) and high branching (HB) N-glycans; the proportion of a-, mono-, di-, tri- and tetra-galactosylated N-glycans (G0, G1, G2, G3 and G4, respectively); and a-, mono-, di-, tri- and tetra-sialylated N-glycans (S0, S1, S2, S3 and S4, respectively).

### 2.7. IgG Fc N-Glycopeptides Analysis

Sample preparation and analysis of IgG N-glycopeptides was done following a previously described protocol with minor changes [30]. Briefly, IgG was isolated from 90 µL of serum samples by affinity chromatography using a CIM 96-well Protein G monolithic plate (BIA Separations, Ajdovscina, Slovenia) and vacuum manifold. IgG N-glycopeptides were prepared by trypsin digestion of 25 µg of IgG isolates and purified with reverse-phase solid phase extraction (RP-SPE) using Chromabond C18 beads suspension applied to the wells of an OF1100 96-well polypropylene filter plate (Orochem Technologies Inc., Naperville, IL, USA) and vacuum manifold. Purified tryptic IgG N-glycopeptides were separated and measured on a nanoAcquity chromatographic system (Waters) coupled to a Compact Q-TOF mass spectrometer (Bruker, Bremen, Germany) equipped with an Apollo II source and operated under HyStar software version 3.2. After calibration, the first four isotopic peaks of doubly and triply charged signals belonging to the same glycopeptide species were extracted and summed together, resulting in 20 Fc N-glycopeptides per IgG subclass. Signals of interest were normalised to the total area of each IgG subclass.

### 2.8. RT-qPCR for miRNAs

We chose to undertake targeted profiling of a 6-plex panel of miRNAs that we had previously determined to be most upregulated in the circulation of patients with recurrent *C. difficile* infection following successful FMT. These were comprised of Let-7b, miR-16, miR-22-3p, miR-23a-3p, miR-4454, and miR-451a [31]. miRNAs were isolated from 200 µL of serum samples using the miRNeasy Serum/Plasma Kit (217184, Qiagen, Hilden, Germany) upon addition of Serum/Plasma Spike-In Control (219610, Qiagen) according to the manufacturer's instructions. Eluted RNA from serum samples was further purified and



concentrated by using Amicon Ultra YM-3 columns (3 kDa MWCO; UFC5003, Millipore, Burlington, MA, USA). Reverse transcription was performed using a miRCURY LNA RT Kit (339340) and quantitative polymerase chain reaction (qPCR) using a miRCURY LNA SYBR Green PCR Kit (339346) and miRCURY LNA primer sets for hsa-let-7b-5p (YP00204750), hsa-miR-16-5p (YP00205702), hsa-miR-22-3p (YP00204606), hsa-miR-23a-3p (YP00204772), hsa-miR-4454 (YP02114119), hsa-miR-451a (YP02119305), cel-miR-39-3p (YP00203952), RNU1A1 (YP00203909), and 5S rRNA (YP00203906, Qiagen) on a CFX384 real-time PCR detection system (Bio-Rad, Hercules, California, CA, USA). The qPCR conditions applied were 95 °C for 10 min, and 40 cycles of 95 °C for 10 s and 60 °C for 1 min, followed by melting curve analysis. qPCR reactions were performed in quadruplicates, and miRNA levels were normalized against cel-miR-39-3p (spike-in), RNU1A1 and 5S rRNA. Normalized miRNA levels are expressed as 'Relative to 1st time point', in comparison with time point 0 (D0-Start) for each patient, or 'Relative to 001\_01-22', in comparison with the non-responder (patient 1) at D0-Start (set as 1).

### 2.9. Multiplex ELISA for Profiling Cytokine and Multi-Isotype Antibody Responses

Patient sera were analyzed for the quantifications of 37 key biomarkers of inflammation from the TNF superfamily proteins, IFN family proteins, Treg cytokines, MMPs, and immunoglobulins IgG1, IgG2, IgG3, IgG4, IgA, IgM using the Bio-Plex Pro Human Inflammation Panel 1 (171AL001M) and Bio-Plex Pro Human Isotyping Panel (171A3100M, Bio-Rad), respectively. Samples were analyzed in a Bio-Plex 200 System using the Bio-Plex manager software according to manufacturer's instructions. The concentrations were calculated by standard curves developed in parallel and are expressed as pg/mL for the inflammatory biomarkers and mg/mL for the immunoglobulins.

### 2.10. Antigen-Specific Microarray

We have previously established and validated a multiplex protein microarray system to measure antibodies to specific *C. difficile* antigens in human sera [32–34]. Briefly and unless otherwise stated, all target antigens and protein homogenates used in this procedure were diluted to 100 µg/mL in sterile print buffer (1× PBS containing 50 mM trehalose, 0.01% Tween 20) prior to application to the antigen microarrays. Corner marker solution was a 1:100 dilution of anti-mouse antibody-IR680 conjugate (Licor BioSciences, Lincoln, NE, USA) in the print buffer. Information on control and test antigens and full microarray experimental procedures is detailed in the Supplementary Methods.

### 2.11. Toxin Neutralization Assay

VERO cells were seeded at  $1 \times 10^4$  per well in a 96-well plate in 50 µL phenol red-free DMEM supplemented with 4.5 g/L D-glucose, 584 mg/L L-glutamine, 25 mM HEPES and 10% FBS and incubated for 18–20 h at 37 °C in 5% CO<sub>2</sub>. Sterile-filtered patient serum was serially diluted 2-fold (1:4 to 1:32) in serum-free, phenol red-free DMEM and mixed with either an equal volume of Toxin A at 200 ng/mL or Toxin B at 1 ng/mL for 1 h at 37 °C. The serum-toxin mixtures were added to VERO cells to give a total well volume of 100 µL and plates were incubated for 18 h at 37 °C. The final concentration of Toxin A and Toxin B was 50 ng/mL and 0.25 ng/mL, respectively. All combinations, including negative controls, were carried out in triplicate. Toxin neutralization was assessed by counting the number of rounded cells versus non-rounded cells in one randomly chosen field of view in each well. Data is displayed as the percentage of cells protected against toxicity.

### 2.12. Isolation and Freezing of Peripheral Blood Mononuclear Cells

Peripheral blood mononuclear cells (PBMCs) were isolated from peripheral blood by density centrifugation using Ficoll-Paque PLUS (GE Healthcare, Chicago, IL, USA). Isolated PBMCs were frozen by re-suspending cells in freezing medium consisting of 10% DMSO (Sigma Aldrich, St Louis, MO, USA) in heat-inactivated fetal calf serum (Biosera, Marikina, Philippines and stored at –80 °C.

### 2.13. Immunostaining via Flow Cytometry

Frozen PBMCs were thawed at 37 °C and washed in 10 mL of RPMI containing FCS (10%) (Sigma Aldrich). The pelleted cells were re-suspended in PBS ( $1 \times 10^6$  cells/mL), were stained with combinations of antibodies (Supplementary Table S1) for 30 min at 4 °C and followed by 2 washes with PBS. For intracellular transcription factor staining for regulatory and follicular helper T cell staining, cells were surface-stained (anti-human CD3, anti-human CD4) and fixed with Foxp3 Fix Perm solution (eBiosciences, San Diego, CA, USA) for 30 min. This was followed by washing the cells, permeabilization with diluted permeabilization buffer (eBiosciences) and staining with antibodies for anti-human foxp3 and anti-human bcl6 for 30 min at 4 °C followed by 2 washes with PBS.

Samples were acquired using a Cyan ADP flow cytometer (Dako, Glostrup, Denmark). Data analysis was done using Summit V 4.3 software. Spectral overlap when using more than one colour was corrected via compensation. Appropriate isotype controls were used for setting gates. The gating strategy used to identify the T cell subset has been shown in Figure S1; the gating strategy for B cell subset distribution has been published [35,36]. The detailed methods for stimulation of PBMCs to induce cytokine production by CD4 T cells and staining for toxin expressing immune cells can be found in the Supplementary Methods.

### 2.14. RNA Isolation, TCR Library Preparation and Sequencing

Total RNA was isolated from PBMCs using the RNeasy mini kit (Qiagen) following manufacturer instructions. Starting from up to 1500 ng of total RNA, molecular-barcoded TCR cDNA libraries were prepared as previously described [37], with minor modifications for both TCR $\alpha$  and TCR $\beta$  chains. See Supplementary Methods for a detailed protocol. Libraries were pooled using 5 ng per library and sequenced on an Illumina NovaSeq6000 SP 2  $\times$  150 bp flow cell. Custom sequencing primers were added to the Illumina primers.

### 2.15. TCR Data Analysis

PCR and sequencing error correction were performed through identification and selection of unique molecular identifiers using the software MiGEC [38], version 1.2.6. Filtered sequences were aligned on a TCR gene reference, clonotypes were identified and grouped, and CDR3 sequence was identified using the software MiXCR [39], version 2.1.1.

Further analysis was performed using R software and packages Vegan [40] for diversity analysis, Mfuzz [41] for temporal trajectory clustering, and ggplot2 for visualization. For temporal trajectory clustering, the most abundant 50 clonotypes of each patient and time point were selected. Clonotypes present in less than 3 time points were excluded from temporal clustering analysis. For clustering of TCRs together with all other experimental measures, K-means clustering was applied as described below.

### 2.16. Statistical Analysis

Time points for all patients were aligned (D0—Start, D1—Cycle 1, D2—Cycle 2, D3—1WeekPost, D4—2/4WeeksPost), and features observed in the one non-responder patient and at least one of the 3 FMT patient responders were then considered. Valid features were normalized, and responders vs. non-responder were compared using statistical *t*-tests between all time points, early time points (D0–D2), and post-FMT time points (D3–D4). Measurements across time points were combined and evaluated using *t*-tests and did not account for any potential correlation present within individuals. This combination of time points was necessitated due to the single individual in the non-responder group. Features were categorized into two groups: “divergent” features showed no significant difference between responders and non-responder at the early time points but demonstrated significant difference at the post-FMT time-points; “convergent” features showed significant difference at early time points and no difference at later time points. The measurements of the same feature in each group were averaged before clustering. To harmonize the heterogeneous profiles for trendline clustering, min-max scaling was performed to map each feature individually to the same range between zero and one. These change patterns were

first categorized by K-means clustering and further grouped into 4 major response models according to whether they increased, decreased, increased then recovered or decreased and then recovered following treatment. Pairwise correlations were performed between selected parameters using Spearman's rank correlation. The features were reordered for visualization using hierarchical clustering [42].  $p$ -values  $\leq 0.05$  were considered statistically significant.  $p$ -values were not adjusted for multiple testing due to the small sample size and exploratory nature of this work. A heatmap was generated using MetaboAnalyst (version 5.0, [www.metaboanalyst.ca](http://www.metaboanalyst.ca), accessed on 13 June 2021) to visualize the immune parameters in the responders vs. non-responder [43].

### 3. Results

#### 3.1. Clinical Outcomes

Four patients treated with sequential FMT and concurrent antibiotics for SFCDI were included. Baseline characteristics of included patients are described in Table 1. A total of 3 patients (patients 1–3) followed a similar treatment protocol with the use of fidaxomicin with FMT by enema, and the fourth patient had FMT by colonoscopy and used vancomycin and metronidazole (Table 1). Information pertaining to multidimensional longitudinal datasets, methodologies employed, and delivery routes for FMT, in addition to treatment cycles and sampling time points, are outlined in Figure 1.

**Table 1.** Participant characteristics and treatment outcomes.

Participant ID	Patient 1	Patient 2	Patient 3	Patient 4
Sex	Male	Female	Female	Male
Age	70	61	85	84
Comorbidities	Chronic pain, NASH cirrhosis (MELD score 9), bariatric Roux-en-Y surgery, chronic obstructive pulmonary disease, depression, atrial fibrillation, hypothyroidism	Congenital blindness in left eye, anxiety	Hypertension, moderate aortic stenosis, oosteoarthritis	Hypothyroidism, type 2 diabetes, hypertension, myocardial infarction, abdominal aortic aneurysm, benign prostatic hypertrophy, chronic kidney disease, prior laparotomy for diverticulosis and small bowel obstruction
Pertinent Medications	Hydromorphone, Flomax, Furosemide, Breo-Ellipta, Synthroid, Apixaban	Temazepam, Citalopram, Gabapentin	Rosuvastatin, Perindopril, Hydrochlorothiazide	Lipitor, Fenofibrate, Synthroid, Lopressor, Flomax
Treatment outcome	Failure	Success	Success	Success
Number of prior CDI	4	1	None	None
FMT prior to study enrolment	>5	None	None	None
CDI Severity	Fulminant	Fulminant	Severe	Fulminant
Anti-CDI Antibiotics during treatment cycles (Total Days)	Fidaxomicin (16 days)	Fidaxomicin (18 days)	Fidaxomicin (11 days)	Metronidazole IV and Vancomycin PO (25 days)
Number of treatment cycles *	2	2	2	5 FMTs by colonoscopy

\* Each treatment cycle consists of 3 consecutive daily FMTs by retention enema plus concurrent fidaxomicin for up to 10 days. Patient 4 received FMT by colonoscopy every 5–7 days with concurrent IV metronidazole and oral (PO) vancomycin, until resolution of pseudomembranous colitis. CDI: *C. difficile* infection; FMT: fecal microbiota transplant.

A total of 2 of the 3 fidaxomicin protocol patients achieved CDI resolution (responders) with 2 treatment cycles (patients 2 and 3). The failure case (non-responder; patient 1) had transient resolution of diarrhea after 2 treatment cycles, but diarrhea recurred within 2 weeks after the final FMT. Repeat *C. difficile* toxin testing was positive, requiring long-

term vancomycin suppression after treating the active CDI episode with vancomycin and metronidazole. Patient 4 achieved CDI resolution after 5 FMT cycles.

### 3.2. Extensive Multi-Analyte Changes Occur with Sequential FMT

681 features were examined (Table S1). After quality control and data completion filtering, 562 features were used for analysis. 78 of these were significantly different ( $p < 0.05$ ) between responders and non-responder at all time points (Table 2). The trendlines and flow cytometry plots for selected immune cell features are presented in Figure 2. Of particular interest are selected immune features which showed the highest positive fold changes for the FMT responders: % CD8 naïve T cells (Figure 2B), CD8 naïve:memory T cell ratio, % B cells (Figure 2C), % unswitched memory B cells (Figure 2D), % regulatory B cells (Figure 2E), in addition to other features, including microRNA-451a, microRNA-16 and toxin B IgG ( $p < 0.05$ ); see Table 2. Noteworthy features which showed higher fold changes for the non-responder included: *Phascolarctobacterium*, unclassified *Enterobacteriaceae*, *Pseudocitrobacter* and *Enterococcus*;  $p < 0.05$ .

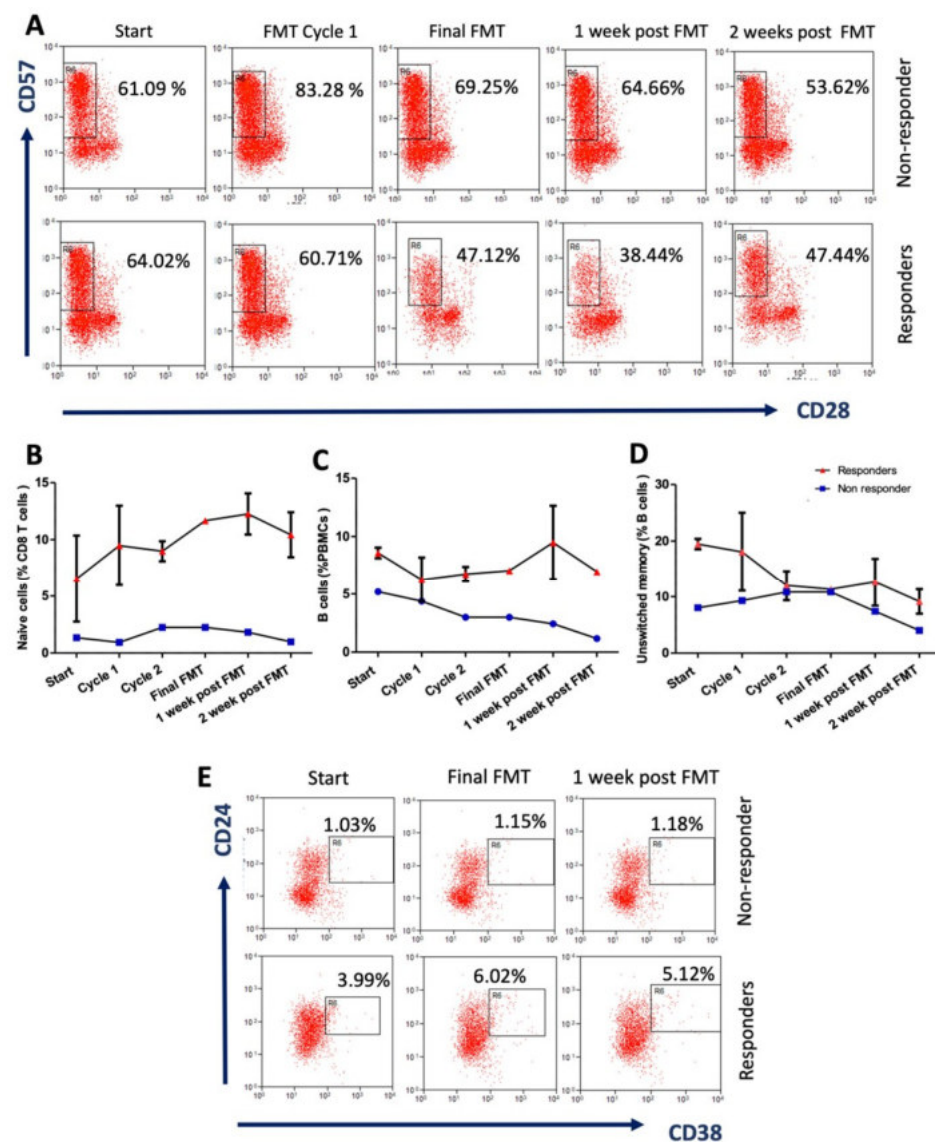
**Table 2.** Features demonstrating statistically significant threshold difference between responders and non-responder (average across all timepoints).

Feature	Category	p-Value	Fold Change (log2) (Succ/Fail)	Mean Value (Responders)	
Features with higher mean in FMT responders					
Naïve:memory CD8 T cell ratio	Flow cytometry	0.0007	2.8982	0.1109	0.0149
Naïve CD8 T cells (%)	Flow cytometry	0.0005	2.6972	9.4820	1.4620
miR-451a	Serum microRNA	0.0040	2.3704	2.2922	0.4433
Regulatory B cells; Bregs (%)	Flow cytometry	0.0008	2.1593	4.2080	0.9420
Toxin B IgG *	Antigen-specific antibody panel	0.0471	1.4260	8.7778	3.2667
Total B cell (%)	Flow cytometry	0.0007	1.3481	7.6880	3.0200
miR-16	Serum microRNA	0.0071	1.3251	1.5138	0.6042
CD4:CD8 T cell ratio	Flow cytometry	0.0002	1.2064	1.7049	0.7388
IgM	Isotype panel	0.0395	0.9896	1.1970	0.6028
EMRA CD4 T cells (%)	Flow cytometry	0.0390	0.9057	25.0740	13.3840
CD28 expression levels on CD4 T cells (MFI)	Flow cytometry	0.0022	0.8997	44.6770	23.9460
Unswitched memory B cells (%)	Flow cytometry	0.0106	0.8939	15.3640	8.2680
IL4 <sup>+</sup> ve stimulated CD8 T cells (%)	Flow cytometry	0.0323	0.8806	2.0290	1.1020
Glycodeoxycholic acid	Stool bile acids	0.0221	0.7647	6.4343	3.7870
A027 IgM ['A' = surface layer proteins (SLP) of ribotype 027]	Antigen-specific antibody panel	0.0240	0.6851	5.1609	3.2099
Stimulated CD4 T cells IL4 expression levels (MFI)	Flow cytometry	0.0277	0.6786	10.0230	6.2620
IL4 <sup>+</sup> ve stimulated CD4 T cells (%)	Flow cytometry	0.0451	0.6227	2.5160	1.6340
Total CD4 T cells (%)	Flow cytometry	0.0001	0.4822	55.8040	39.9500
CD28 expression levels on CD4 T cells (MFI)	Flow cytometry	0.0001	0.4684	28.6430	20.7020
Total memory B cells (%)	Flow cytometry	0.0241	0.4349	51.7350	38.2700
IgGII1H5N4F1S1: IgG2&3 glycopeptide with digalactosylated and monosialylated glycan with core fucose	IgG glycoprofiling	0.0289	0.3593	7.8840	6.1460
Stimulated CD8 T cells IL4 expression levels (MFI)	Flow cytometry	0.0030	0.3565	2.3670	1.8488
IgGII1H4N4F1: IgG2&3 glycopeptide with monogalactosylated glycan with core fucose	IgG glycoprofiling	0.0059	0.3545	14.2693	11.1605
CD28 expression levels on CD8 T cells (MFI)	Flow cytometry	0.0315	0.3306	23.2890	18.5200
IgGIV1H5N4F1: IgG4 glycopeptide with digalactosylated glycan with core fucose	IgG glycoprofiling	0.0471	0.2910	4.0918	3.3443
Monosialylated glycans	Serum glycan traits	0.0014	0.2556	16.6257	13.9260
NKG2D expression levels on CD4 T cells (MFI)	Flow cytometry	0.0003	0.2181	5.2907	4.5483
IgGI1H4N4F1: IgG1 glycopeptide with monogalactosylated glycan with core fucose	IgG glycoprofiling	0.0140	0.1917	20.8725	18.2761
Candida IgM	Antigen-specific antibody panel	0.0009	0.1841	7.2038	6.3409

Table 2. Cont.

Feature	Category	p-Value	Fold Change (log2) (Succ/Fail)	Mean Value (Responders)	
Senescent CD4 T cells NKG2D expression levels (MFI)	Flow cytometry	0.0011	0.1781	5.2329	4.6252
Digalactosylated glycans	Serum glycan traits	0.0042	0.1467	54.6643	49.3780
MMP-1: matrix metalloproteinase-1	Inflammation panel	0.0072	0.1443	7.3383	6.6399
Low-branching glycans	Serum glycan traits	0.0155	0.1114	72.3240	66.9480
	Features with higher mean in FMT non-responder				
<i>Acidaminococcaceae</i>	Family	0.0212	−3.6351	0.1625	2.0193
<i>Phascolarctobacterium</i>	Genus	0.0212	−3.6351	0.1625	2.0193
<i>Enterobacteriaceae_unclassified</i>	Genus	0.0013	−2.2113	1.0058	4.6577
<i>Pseudocitrobacter</i>	Genus	0.0080	−2.2079	0.7009	3.2383
<i>Enterococcaceae</i>	Family	0.0035	−1.8467	1.4509	5.2185
<i>Enterococcus</i>	Genus	0.0035	−1.8467	1.4509	5.2185
L001 IgA ('L' = lysates of ribotype 001)	Antigen-specific antibody panel	0.0000	−1.5667	24.6667	73.0667
3-alpha-hydroxy-7,12-diketocholanic acid	Stool bile acids	0.0189	−1.4326	2.0780	5.6094
CMV IgG	Antigen-specific antibody panel	0.0000	−1.0003	3350.2467	6702.0667
Toxin B IgA*	Antigen-specific antibody panel	0.0194	−0.9586	1.5437	3.0000
A001 IgA ['A' = surface layer proteins (SLP) of ribotype 001]	Antigen-specific antibody panel	0.0381	−0.8841	2.1246	3.9212
3 dehydrocholic acid	Stool bile acids	0.0082	−0.8481	4.0741	7.3340
Beta muricholic acid	Stool bile acids	0.0004	−0.8238	4.3596	7.7168
A027 IgG ['A' = surface layer proteins (SLP) of ribotype 027]	Antigen-specific antibody panel	0.0499	−0.7879	3.1054	5.3618
CD28 <sup>−ve</sup> T cells (%)	Flow cytometry	0.0000	−0.7757	35.0250	59.9640
sTNF-R1: soluble tumor necrosis factor receptor-1	Inflammation panel	0.0019	−0.7465	3079.3947	5166.5200
IL-26	Inflammation panel	0.0267	−0.7191	1044.0507	1718.7220
Integrin <sup>+</sup> ve dendritic cells (%)	Flow cytometry	0.0000	−0.7071	2.0863	3.4059
sTNF-R2	Inflammation panel	0.0150	−0.6979	1274.3983	2067.2920
Total CD8 T cells (%)	Flow cytometry	0.0000	−0.6956	33.6670	54.5240
12 dehydrocholic acid	Stool bile acids	0.0135	−0.6126	6.7393	10.3044
Chenodeoxycholic acid	Stool bile acids	0.0000	−0.5788	8.4395	12.6054
CD28 <sup>−ve</sup> CD57 <sup>+</sup> ve senescent CD8 T cells (%)	Flow cytometry	0.0008	−0.5387	46.6120	67.7120
Antennary fucosylation	Serum glycan traits	0.0017	−0.5289	9.8970	14.2800
CD28 <sup>−ve</sup> senescent CD8 T cells (%)	Flow cytometry	0.0002	−0.4370	61.7170	83.5540
CD28 <sup>−ve</sup> CD57 <sup>+</sup> ve senescent CD4 T cells (%)	Flow cytometry	0.0497	−0.4303	23.3810	31.5060
Tetragalactosylated glycans	Serum glycan traits	0.0017	−0.4232	4.9430	6.6280
Cholic acid-3-sulfate	Stool bile acids	0.0362	−0.4122	5.5642	7.4042
IgGIV1H3N5F1: IgG4 glycopeptide with bisected agalactosylated glycan with core fucose	IgG glycoprofiling	0.0161	−0.3958	8.3615	11.0012
Tetrasialylated glycans	Serum glycan traits	0.0074	−0.3942	4.2567	5.5940
Chenodeoxycholic acid-3-sulfate	Stool bile acids	0.0412	−0.3777	7.2848	9.4649
CD8 effector memory T cells (%)	Flow cytometry	0.0009	−0.3580	58.0380	74.3860
CD57 <sup>+</sup> ve senescent CD8 T cells (%)	Flow cytometry	0.0290	−0.3452	55.0520	69.9360
IgGIII1H3N4: IgG2&3 glycopeptide with agalactosylated glycan without core fucose	IgG glycoprofiling	0.0051	−0.3291	1.5296	1.9215
High-branching glycans	Serum glycan traits	0.0071	−0.2972	25.0977	30.8400
IgGIII1H4N4: IgG2&3 glycopeptide with monogalactosylated glycan without core fucose	IgG glycoprofiling	0.0276	−0.2911	3.3512	4.1005
Trisialylated glycans	Serum glycan traits	0.0207	−0.2873	16.2837	19.8720
Trigalactosylated glycans	Serum glycan traits	0.0327	−0.2646	20.1547	24.2120
Total T cells (%)	Flow cytometry	0.0077	−0.2526	57.9910	69.0860
IgGIII1H4N5S1: IgG2&3 glycopeptide with bisected monogalactosylated and monosialylated glycan without core fucose	IgG glycoprofiling	0.0458	−0.2491	1.8785	2.2325
MMP-2: matrix metalloproteinase-2	Inflammation panel	0.0000	−0.2025	9.3344	10.7411
Candida IgG	Antigen-specific antibody panel	0.0001	−0.1961	8.0645	9.2384
Total CD8 memory T cells (%)	Flow cytometry	0.0015	−0.1777	87.0120	98.4160
TWEAK/TNFSF12: TNF-like weak inducer of apoptosis/tumor necrosis factor superfamily	Inflammation panel	0.0033	−0.1684	5.4639	6.1403
Integrin expression levels on dendritic cells (MFI)	Flow cytometry	0.0235	−0.1316	3.7044	4.0582

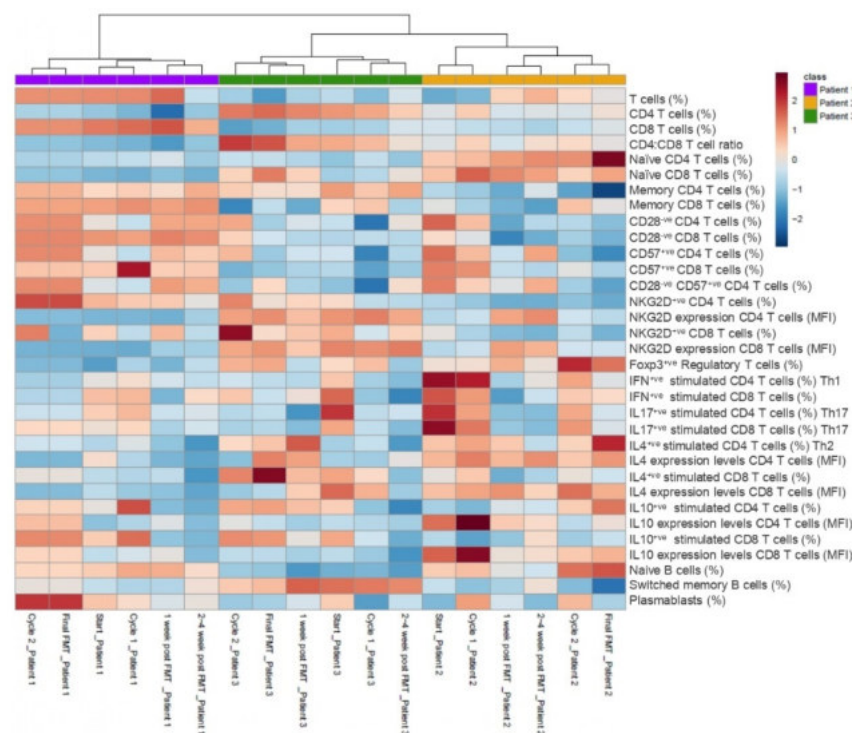
MFI—Median fluorescence intensity. \* Whole toxins A and B purified from toxinotype 0, strain VPI 10463, Public Health England.



**Figure 2.** Reversal of immunosenescence features in patients with severe or fulminant *Clostridioides difficile* infection post-sequential FMT. (A) Representative flow cytometry plots show the kinetics of peripheral CD28<sup>-ve</sup> CD57<sup>+</sup> senescent CD8 T cells in FMT responders ( $n = 2$ ) (mean  $\pm$  S.D data for patient's 2 and 3) and a non-responder patient ( $n = 1$ ) (patient 1). Percentage of peripheral (B) naive CD8 T cells; (C) B cells; and (D) Unswitched memory B cells in responders at the start of the trial, post-FMT cycle 1, post-FMT C\cycle 2, post-final FMT cycle, and 1 week and 2 weeks after FMT. (E) Representative flow cytometry plots show the kinetics of peripheral CD24<sup>hi</sup> CD38<sup>hi</sup> regulatory B cells in FMT responders and non-responder patient at the start of the trial, post final FMT cycle and 1 week after FMT.

Features demonstrating convergence and divergence are displayed in Tables S2 and S3. A total of 114 immune parameters were profiled by multi-color flow cytometry in patients 1–3 (Table S1). Further details of gating strategies for CD4 T cell subsets (Figure S1) and antibodies used are detailed below. Temporal trends in the normalized % frequency values of other selected immune cytometric parameters for responders (patient 2 and 3 combined) and non-responder (patient 1) are presented in Figure 3. Higher normalized % frequencies of CD4 T cells and naive CD4 T cells were observed in the FMT responders. In contrast, the non-responder's peripheral immune system was composed of higher frequencies of total T cells, CD8 T cells, memory CD4 and CD8 T cells, as well as senescent CD4 and CD8 T cell populations. Of interest, circulating senescent T cells (CD28<sup>-ve</sup>CD57<sup>+</sup>CD4 T cell and

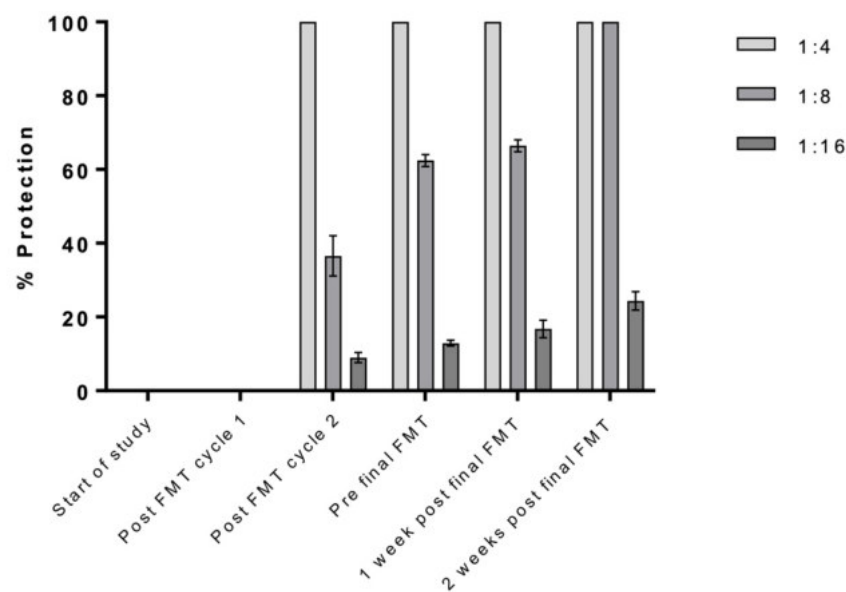
NKG2D<sup>+</sup> CD4 T cell) peaked at the end of cycle 2 and final FMT timepoints, particularly in the non-responder (Figure 3). Peripheral frequencies of anti-inflammatory regulatory T cells peaked at the final FMT time point in the responders (Figure 3). Functional immune assays revealed a higher proportion of anti-inflammatory cytokine IL4 production by CD4 T cells (Th2) and CD8 T cells post-stimulation in the FMT responders, particularly following FMT treatment (Figure 3). Furthermore, anti-inflammatory cytokine IL10 production by CD4 T cells peaked at the end of cycle 1 and was detectable in the circulation of the non-responder until the end of the final FMT, when the patient relapsed, and thereafter fell. In the responders, IL10-producing CD4 T cells peaked in frequency at the end of final FMT before being maintained at a lower but detectable frequency, at least until 2–4 weeks post FMT (Figure 3).



**Figure 3.** Heat map of normalized frequency values of selected immune subset parameters in the responders (patient 2 and 3 combined) and non-responder patient (patient 1) at different time points. Patients 1, 2 and 3 were clustered using hierarchical clustering (Euclidian distance based). High and low normalized frequency values are indicated in red and blue, respectively. Different immune subset percentages for FMT responders ( $n = 2$ ; patients 2 and 3)) and non-responder ( $n = 1$ ; patient 1) are indicated at the end of the column.

In terms of functional antibody responses, interestingly, only patient 2 of the 4 patients assessed displayed neutralizing anti-*C. difficile* toxin antibodies in their sera (Figure 4), and thus, readouts from this assay were not incorporated into integrative analyses.

Further analysis of toxin-expressing immune cells revealed that the FMT responders were characterized by a decline in peripheral frequency of Toxin A-expressing naïve CD4 T cells (Figure S2A) and Toxin A-expressing memory CD4 T cells (Figure S2B) at the end of cycle 1 and post-final FMT, respectively. Similarly, a decline in peripheral Toxin A- and B- expressing total B cells (Figure S2D,E) and memory B cells (Figure S2F,G) occurred post-second cycle of FMT; the lowest frequencies were observed post-final FMT cycle and thereafter increased 1 week post-FMT.

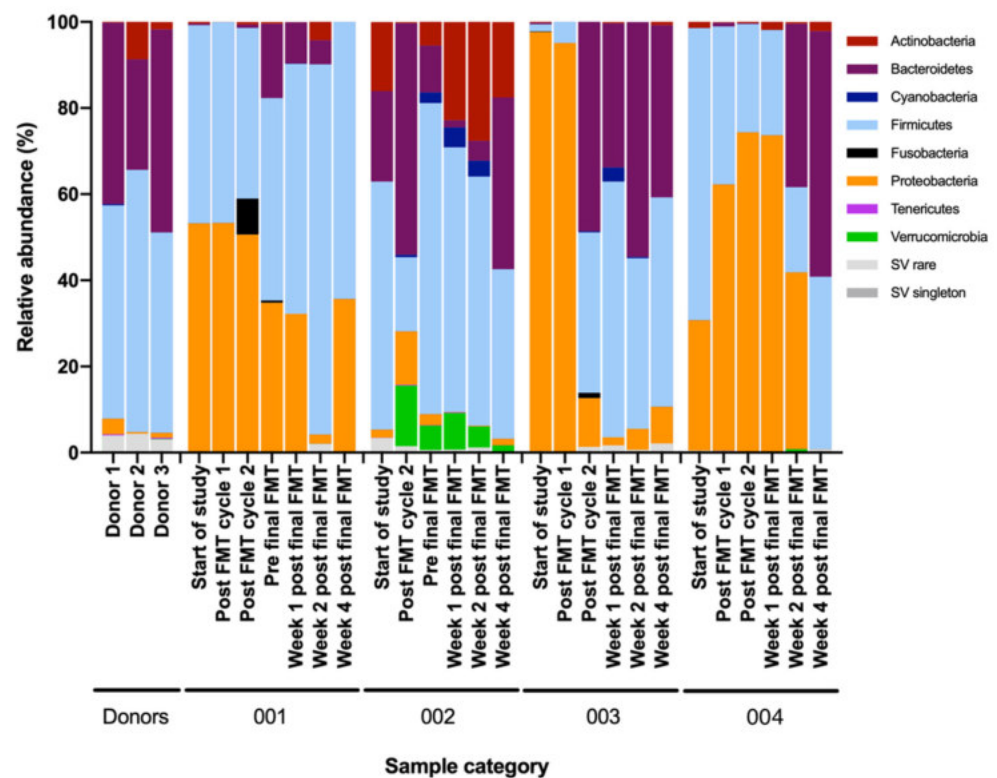


**Figure 4.** In vitro evaluation of antibody-mediated neutralization of Toxin B (patient 2, FMT responder). Sera were serially diluted and incubated with whole purified toxin B (toxintype 0, strain VPI 10463, ribotype 087) before addition to VERO cells. Cytotoxicity was assessed by counting the number of rounded and non-rounded healthy cells, expressed as percentage protection. For patient 2, detection of neutralization against Toxin B became apparent post-FMT cycle 2 with 100% protection from Vero cell rounding observed with the most concentrated serum tested (1:4 dilution) compared to 0% at the 2 earlier time points. The degree of protection increased from this point over the course of treatment, with protective efficacy clearly detected even in the lowest dilution tested, 1:16. For 1:8 diluted sera, the mean percentage of healthy, non-rounded, protected cells increased from 36.57% at post-FMT cycle 2 to 62.43% pre-final FMT to 66.42% 1 week post-final FMT to 100% 2 weeks post-final FMT. Unlike this patient, patients 1 and 3 displayed no neutralization against Toxin B, and none of the patients showed neutralization against Toxin A throughout the treatment. Controls for this assay showed 100% rounding for cells incubated with the appropriate toxin alone and 100% healthy non-rounding for cells incubated with the respective serum dilution alone. Sera from patients 1 and 3 showed no neutralization against Toxin B. No neutralization was shown against Toxin A for patients 1, 2 or 3. Data from triplicate values  $\pm$  SD.

Changes in stool microbiota at the class, order, family and genus levels for all patients are described in Figure S3A–D. At the phylum level, responders' stool microbiota shifted from Proteobacteria predominance at baseline to Bacteroidetes predominance 2–4 weeks after treatment. Conversely, the non-responder demonstrated only a modest decrease in Proteobacteria and no consistent enrichment of Bacteroidetes (Figure 5).

Stool donors' microbial composition was predominated by Firmicutes and Bacteroidetes. Severe or fulminant *Clostridioides* patients had higher relative abundance of Proteobacteria or Actinobacteria prior to fecal microbiota transplantation (FMT). In treatment responders (patient 2, 3 and 4), the relative abundance of Proteobacteria reduced, but this did not occur in the non-responder (patient 1) after sequential FMT.





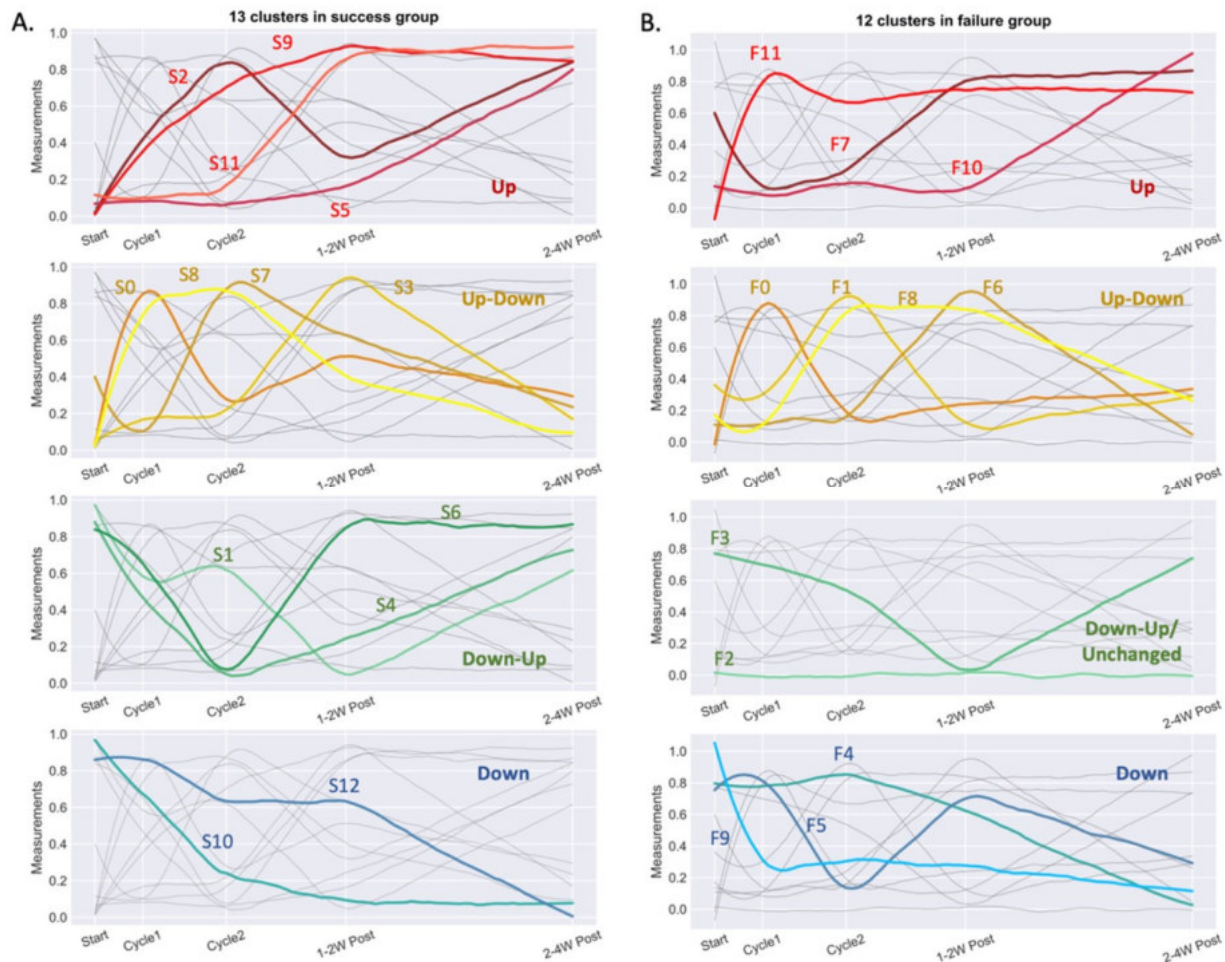
**Figure 5.** Fecal metataxonomic changes at the phylum level in relation to sequential FMT in severe or fulminant CDI patients. 16S rRNA gene sequencing of DNA extracted from stool samples, presented as relative abundance plots. Participant samples presented as: three stool donors; patient 1(001), earliest to latest timepoint; patient 2 (002), earliest to latest timepoint; patient 3 (003), earliest to latest timepoint; and the fourth patient, earliest to latest timepoint.

### 3.3. Multiomics Longitudinal Patterns Possibly Associated with FMT Response

Analyzed features from FMT responders were longitudinally clustered and classified into 13 temporal behavioral clusters, S0–S12. In total, 4 clusters (S2, S5, S9 and S11) included features which increased, while clusters S10 and S12 decreased over the course of the study intervention (Figure 6A and Table S4A). It was interesting to note that while some clusters were dominated by certain categories of features, others contained mixed feature groups with correlating behaviors. Among these, cluster S10 was particularly noteworthy. This cluster contained features which sharply decreased in abundance from the end of cycle 1, and comprised serum pro-inflammatory proteins; frequency of circulating CD8 memory T cells; frequencies of senescent CD57<sup>+</sup> T cells and CD28<sup>-</sup>CD57<sup>+</sup>CD8 T cells; IFN<sup>+</sup> and IL17<sup>+</sup> stimulated CD4 and CD8 T cells; IL-4 and IL-10 expression levels in CD8 and CD4 T cells, respectively; disialylated and trigalactosylated glycans; IgG Fc N-glycopeptides; fecal 2-hydroxybutyrate; tauro- and glycoconjugated fecal bile acids; as well as bacterial genera comprising *Cutibacterium*, *Collinsella*, *Barnesiella*, *Prevotella\_7*, S5-A14a, *Tyzzarella\_4*, *Fastidiosipila*, *Ruminococcus\_1*, *Phascolarctobacterium*, *Suterella*, *Citrobacter*, and unclassified *Mollicutes\_RF39*.

The multiomics profile of the non-responder was classified into 12 clusters, F0–F11 (Figure 6B and Table S4B). Of note, features within cluster F1 of the non-responder increased at the end of cycle 2, when the patient transiently improved clinically, before falling to low pre-intervention baseline levels after 1–2 weeks post-FMT when the patient became feverish and tested positive for fecal *C. difficile* toxin. This cluster was comprised of serum butyrate, isobutyrate and fecal 2-hydroxybutyrate; several fecal bile acids; peripheral memory T cells; plasmablasts; IL-4-producing CD8 T cells; and bacterial genera *Faecalibacterium*, *Fusobacterium* and *Klebsiella*. Other notable clusters (F7 and F10) comprised features which gradually increased after the end of FMT cycle 2. F10 contained serum acetate,

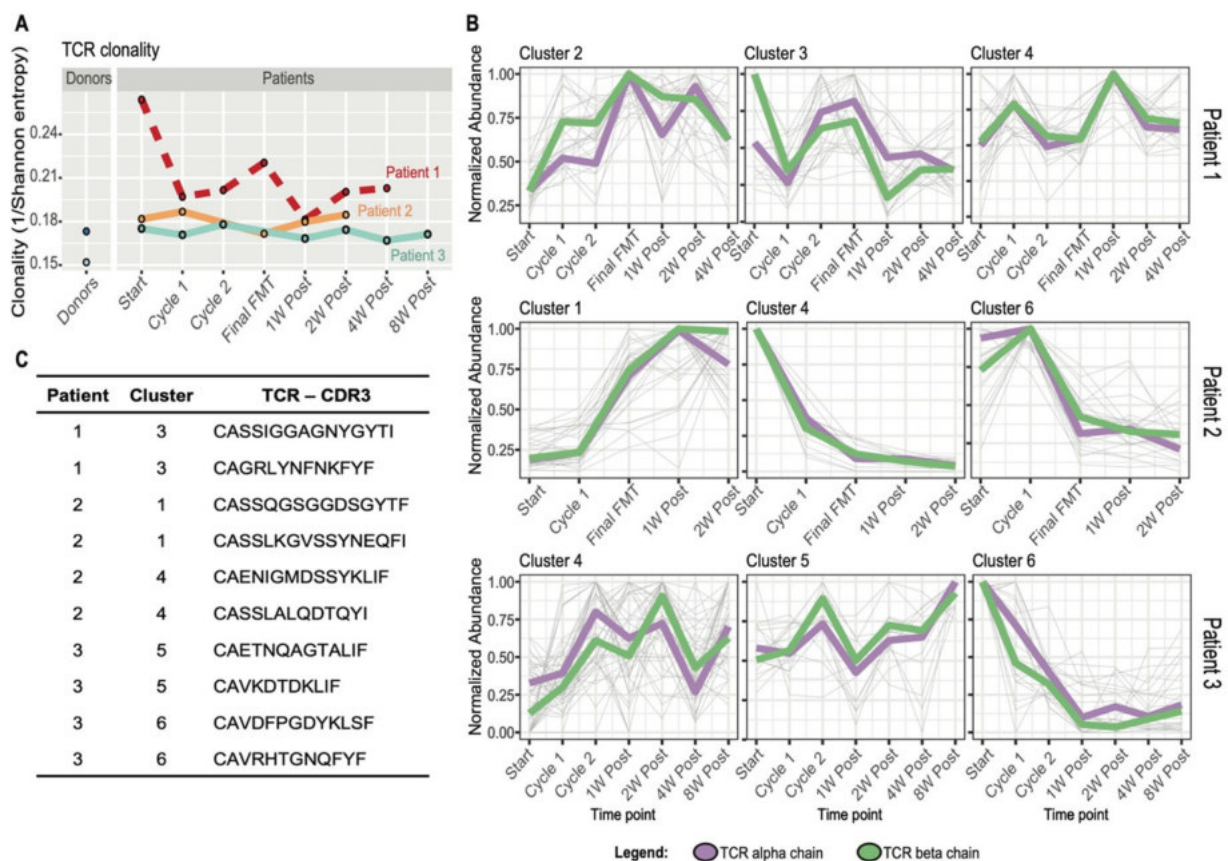
serum microRNA-4454, switched memory B cells, tauro- and glyco-conjugated fecal bile acids, *Enterococcus*, *Agathobacter*, *Tyzzrella*, *Dialister*, *Dorea*, *Collinsella*, and members of the *Ruminococcaceae* family.



**Figure 6.** K-means clustering of trendlines of 562 valid measurements. Results shown for FMT responders ( $n = 2$ ) (success group; **A**) and FMT non-responder ( $n = 1$ ) (failure case; **B**) with each cluster indexed. Clusters are regrouped into four categories as highlighted in each row of the subplots: increased after FMT (up, red), increased after FMT but recovered (up-down, yellow), decreased after FMT but recovered or unchanged (down-up or unchanged, green), and decreased after FMT (down, blue).

### 3.4. FMT Impact on T Cell Receptor Repertoire and Multiomics Integration

To investigate the T cell immune response more specifically, longitudinal TCR $\alpha$  and TCR $\beta$  repertoire analysis was performed in PBMCs in patients 1–3 (no PBMCs available for patient 4) and two FMT donors. Both FMT donors displayed lower clonality and higher diversity compared to all patients. Both responders (patients 2 and 3) exhibited stable clonality profiles over time. Interestingly, clonality was much higher at screening (pre-FMT) but drastically declined following the first treatment cycle when diarrhea resolved transiently in the only non-responder (patient 1) (Figure 7A). The most abundant TCRs for each patient were clustered based on their temporal behavior using Mfuzz R software. Clusters were identified which appeared to correlate with therapy response, based on the observation that they contained TCRs which sharply increased or decreased after FMT in patients 2 and 3. However, patient 1 (non-responder) did not show clear temporal trajectories in any cluster, but was rather characterized by an intermittent or fluctuating TCR temporal profile (Figure 7B,C and Figure S4A–C).



**Figure 7.** Longitudinal TCR repertoire analysis. (A) TCR repertoire clonality calculated as the inverse of the Shannon entropy on sampled peripheral blood mononuclear samples to 1000 TCR sequences. (B) Temporal clustering performed with Mfuzz R package for the 50 most abundant TCRs, alpha and beta, for each time point for patient 1 (FMT failure/non-responder), patient 2 (FMT success/responder) and patient 3 (FMT success/responder). The figure shows 3 exemplary clusters out of 6 for each patient. Thin grey lines in the background represent single clonotypes. The median value of the temporal trajectories of TCR alpha (violet) and beta (green) chains for each cluster is displayed. (C) Exemplary TCR CDR3 amino acid sequences of the most interesting temporal clusters containing TCRs increasing or decreasing in abundance during and after FMT for patients 2 and 3 and TCRs with high abundance at baseline and during disease recurrence for patient 1.

A dedicated integrative multiomics analysis was performed separately for each patient to include the TCR information, which is unique to each individual. As described in the multiomics paragraph, K-means clustering was performed, and TCRs showing correlative temporal behavior with other increasing-decreasing features were identified (Figure S5A–C and Table S5A–C).

### 3.5. Temporal Correlation among Features: A Closer Look at T Cell Immunosenescence Signatures, Gut Microbiome and Immunometabolic Features

As an exploratory analysis to identify potentially correlating features which may be associated with treatment outcome, Spearman’s rank correlation coefficients were assessed between selected features of interest for disease pathogenesis and progression and all omics measurements in patients 2 and 3 (FMT responders) (Table S6A) and patient 1 (non-responder) (Table S6B). Of particular interest was the correlation analysis of peripheral senescent T cells (Table S7). Observations in the non-responder included strong positive associations of peripheral senescent T cells and several host factors, including fecal butyrate; serum hydroxybutyrate; fecal urso-, iso- and hydoxycholeic acids; serum IgG Fc N-glycopeptides; and microbial taxa, particularly *Pseudomonas* at the genus, family and order level, *Coprococcus\_1* and *Ruminococcaceae\_014*, *Solibacterium*, and *Mollicutes*, which were also positively associated with titers of serum IgA anti-toxin B.

#### 4. Discussion

In this longitudinal multiomics study, we observed temporal changes in immune, metabolic and gut bacterial features in a small cohort of patients with antibiotic refractory SFCDI during sequential FMT. We identified 78 features which could help in differentiating responders from the non-responder. Our results hint that non-response may be associated with immunosenescent signals in the non-responder, including: a higher frequency of circulating senescent T cells characterized by loss of CD28 surface antigen and acquisition of CD57; lower B cell and regulatory B cell frequencies; and higher levels of MMP-2, TWEAK, IL-26, sTNF-R1, sTNF-R2, and effector memory CD8 T cells. Furthermore, a higher relative abundance of *Enterococcus*, unclassified *Enterobacteriaceae*, and a lower peripheral TCR repertoire diversity coincided with CDI recurrence, higher levels of fecal primary BAs, and agalactosylated serum IgG N-glycopeptides. In support of these immune aging-related observations, enhanced expression of CD57 on CD8<sup>+</sup>ve T cells has been linked to rejection in renal transplant recipients, highlighting a role of immunosenescence beyond CDI [44]. Aging is accompanied by the decline of CD28 expression on CD4 and CD8 T cells, loss of naïve cells, accumulation of memory T cells, and reduced diversity of the TCR repertoire [45]. The increased secretion of circulating pro-inflammatory molecules, such as MMP-2, sTNF-R1, and IL-26, the master regulator of inflammation in the non-responder, provides further evidence of the senescence-associated secretory phenotype (SASP) [46] and suggests immunosenescence may be a critical driver of treatment outcome in antibiotic refractory CDI patients receiving FMT. Indeed, IL-26 is known to induce the production of SASP proteins, including IL-1 $\beta$ , IL-6, and TNF $\alpha$ , by human monocytes and can trigger NK cell activation, inducing expression of IL-1 $\beta$ , TNF $\alpha$  and type 1 and 2 interferons [47].

The positive correlations observed separately between butyrate, *Pseudomonas* and T cell senescence in the non-responder is indirectly supported by recent studies which demonstrate the ability of sodium butyrate and *Pseudomonas aeruginosa* to induce cellular senescence in human glioblastoma cells and murine lung tissues [48,49], and supports the contention that immunosenescence in the FMT non-responder may be driven by persistent infections; the parallel and increased detection of anti-CMV IgG antibodies in the non-responder provides additional support for this hypothesis. Cytomegalovirus may contribute markedly to immune dysfunction with age [50] and CMV infection has emerged as a pathogenic accelerant of immune cell proliferation underlying immune senescence [51,52]. Human CMV can also cause poor outcomes in solid organ transplant recipients [53,54], and thus, anti-CMV antibody status may be useful in identifying the risk of immunosenescence and FMT failure. Our correlation analysis showing positive associations between senescent T cell frequencies and the *Rikenellaceae* family, which includes the *Alistipes* genus, are in keeping with previous human studies which report that both are over-represented in older adults [55].

Consistent with our findings, bacterially-derived primary bile acids (BAs), such as taurocholic acid, promote *C. difficile* spore germination, while secondary BA inhibit vegetative growth [56] and downregulate toxin activity [57]. Following successful FMT for recurrent mild CDI, we and others have demonstrated that primary BA diminish and are replaced by protective secondary BA [12,58,59]. Secondary BA deficiency in CDI and other forms of colitis “cross-talk” to the host via changes in activation of TGR5- and FXR-mediated signaling pathways [60]. SCFAs stably increase following FMT [17,18,59] and exert an epigenetic effect on the host through the inhibition of histone deacetylases (HDACs); the transcriptional changes arising through this inhibition result in a net anti-inflammatory effect [61]. Valerate is among the SCFAs that have been demonstrated to diminish colitis via HDAC inhibition [61]. Our results also align with our previous study, in which we reported a significantly higher relative abundance of monosialylated and digalactosylated serum N-glycans following successful FMT for rCDI [20], supporting a protective role for these serum N-glycans. Our observation of a higher relative abundance of circulating IgG4 glycopeptides with agalactosylated and bisected N-glycans with core fucose in non-response is in keeping with pathogenic pro-inflammatory and aging states [62].

In contrast, the increase in CD8 naïve T cells in responders supports the notion that these ‘foot soldiers’ of the immune system may be important players in CDI resolution, although the role of the memory T cell response in CDI remains incompletely understood. The detection of more abundant circulating B cells in responders is congruent with a previous report which showed that mucosal IgA-producing B cells are reduced in patients with rCDI [63], and leads us to suggest that FMT may restore B cell frequencies. Nevertheless, only one responder (patient 2) displayed detectable neutralizing antibodies, indicating that other factors, such as the indigenous gut microbiota, may facilitate *C. difficile* clearance, independent of adaptive immunity [64]. However, evidence suggests that, at least in the context of certain viral infections, in vitro toxin neutralization studies may not accurately predict in vivo protection [65,66]. Moreover, it is likely that study participants could have been infected with a *C. difficile* strain other than the historical VPI-10463 strain, from which toxins were derived for in vitro assays described herein. As such, their memory B cell-encoded antibodies may not have been able to neutralize the VPI-10463 toxins. Toxin concentration and the availability of neutralizing antibodies in the gut lumen as well as the rate of toxin clearance may influence disease progression, severity and treatment of CDI [67–69]. Interestingly, circulating toxin A- and B-expressing memory B cell frequencies fell at the end of final FMT for the responders before recovering to lower frequencies than those detected prior to FMT. These antigen-specific memory B cells may no longer be required in the context of presumed *C. difficile* clearance and may have transited into the lymphoid organs. Alternatively, these observations may reflect memory B cell apoptosis, which has been reported to contribute to impaired germinal center formation of memory B cells after allogeneic hematopoietic stem cell transplantation [70]. In contrast, peripheral toxin A- and B-expressing memory B cells seemed to peak in frequency two weeks post-FMT for the non-responder. It is possible that a lack of T cell help may have limited memory B cell expansion and differentiation into antibody-secreting plasma cells in this non-responder, or that their toxin A- and B-expressing memory B cells, which increased 2 weeks post-final FMT, only encoded low-affinity antibodies incapable of toxin neutralization. In support of this latter hypothesis, Shah and colleagues recently demonstrated that human *C. difficile* toxin-specific memory B cell repertoires encode poorly neutralizing antibodies, with limited class switching and IgM dominance [71]. These findings indicate that memory B cells may be an important factor in *C. difficile* disease recurrence.

FMT responders also exhibited higher fold changes in frequency of tolerogenic regulatory B cell populations and IL-4 producing CD4 T cells across all time points compared to the non-responder. These findings concur with evidence which demonstrates the protective capacity of immunosuppressive Bregs in the prevention of allograft rejection in renal, liver and lung transplantation recipients and in the development of graft-versus-host disease in stem cell recipients, and with increased frequencies of CD4<sup>+</sup> T cells in the colons of FMT-treated mice, respectively [72,73]. Our findings suggest that, in addition to glatiramer acetate, which is known to stimulate functional Bregs [74], IgG1 Fc N-glycopeptides with bisected and non-bisected monogalactosylated and monosialylated glycans without core fucose may help induce this Breg tolerogenic phenotype, whereas certain taxa such as *Acidaminococcaceae*, *Phascolarctobacterium* and *Fastidiosipila*, in addition to the fecal bile acids 3-alpha-Hydroxy-7 ketolithocholic acid and chenodeoxycholic acid, may negatively regulate Bregs.

Intriguingly, we observed a higher relative abundance of *Enterobacteriaceae* and *Enterococcaceae* in the non-responder, a finding which may indicate that such pro-inflammatory bacterial taxa, which are more abundant in older adults [75], may play a part in driving biological aging. Consistent with these findings, patients with progeria and progeroid mice display an abundance of Proteobacteria. FMT could reverse this effect, restoring secondary BA biosynthesis and enhanced health and lifespan in progeroid mice [76]. Moreover, the defective germinal center response in aged mice can be rescued by replenishing the gut microbiome of aged mice with that of a younger animal [77].

Investigation of specific temporal clusters in the FMT non-responder also showed that *Faecalibacterium*, *Fusobacterium* and *Klebsiella* increased in relative abundance at the time the non-responder temporarily showed signs of clinical improvement. In keeping with our observation, recent research has shown that specific bacterial genera strongly associate with systemic immune cell dynamics, and that gut microbial taxa, including *Faecalibacterium* as well as *Ruminococcus* 2 and *Akkermansia*, accelerate immune reconstitution after allogeneic haematopoietic cell transplantation [78]. Moreover, members of *Faecalibacterium*, *Ruminococcus* [79] in one study, and *Akkermansia* [80] in another have been associated with better responses to anti-PD-1 immunotherapy.

The main strength of this study is in the deep longitudinal characterization of four patients that have received sequential FMT for severe or fulminant CDI. However, the small sample size of patients and multiple stool donors in this observational study preclude generalization of the present findings. The single non-responder also has multiple comorbidities, including liver cirrhosis, diabetes, and intestinal bypass for morbid obesity, which may contribute to some of the observed differences from responders. For example, one of the features distinguishing responders from the non-responder was decreased CD4:CD8 ratios, which is known to be lower in liver cirrhosis due to lower CD4 counts [81]. Therefore, definitive conclusions regarding discriminating features of treatment outcomes following FMT cannot be drawn. This unique patient population tends to have multiple comorbidities [7], contributing to the heterogeneous population included in our case series. Antibiotic refractory SFCDI fortunately occurs infrequently, contributing to our small sample size. Nonetheless, the results we observed in this study may partially explain why cirrhotic patients may have worse CDI outcomes and may provide multiple starting points to further decipher FMT mechanisms of action. Larger multi-center studies are required to confirm these observations and provide additional mechanistic insight.

Together, these cases provide novel and provocative evidence that extensive host-microbial interactions occur following sequential FMT for SFCDI and suggest a potentially important role for immunosenescence in shaping clinical outcomes. Thus, further studies defining mechanistic and functional outputs of commensals are fundamental to the advancement of next-generation biotherapeutics to treat CDI and other disease states that may be associated with immune aging.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/cells10113234/s1>, Supplementary methods, Figure S1: Gating strategy for CD4 T cell subsets, Figure S2: Toxin-expressing T and B cells in patients with severe or fulminant CDI in relation to sequential FMT, Figure S3: Fecal metataxonomic changes in relation to sequential FMT in patients with severe or fulminant CDI, Figure S4: Temporal clustering of the most abundant TCRs for FMT responders and non-responder, Figure S5: K-means integrative clustering of all measures, including TCR alpha and beta clonotypes, for FMT responders and non-responder, Table S1: Dictionary of longitudinal features assessed in the present multiomics study ( $n = 681$ ), Table S2: Convergent features identified for FMT responders and non-responder, Table S3: Divergent features identified for FMT responders and non-responder, Table S4: K-means cluster analysis without TCR repertoire data showing features within each indexed cluster, Table S5: K-means cluster analysis with TCR repertoire data showing features within each indexed cluster, Table S6: Immunosenescent T cell parameter correlations with omics features for FMT responders and non-responder, Table S7: Other selected differentiating parameters and their omics correlations for FMT responders and non-responder.

**Author Contributions:** Conceptualization: T.M.M.; methodology: T.M.M., D.H.K., L.A.R., N.A.D., E.R., R.G., M.P.-B., P.T., B.H.M., C.P.; investigation: T.M.M., D.H.K., L.A.R., B.R., C.W., K.W., D.Y.Y., L.S., J.H., M.P.-B., F.V., F.K., P.T., N.A.D., M.H., E.R., N.X., T.D., Y.L., B.H.M., J.M.B., J.A.K.M., C.P., T.O.Y., N.C., M.H.; writing—original draft: T.M.M.; writing—review & editing: T.M.M., N.A.D., E.R., D.H.K., L.A.R., B.R., C.W., K.W., G.K.-S.W., D.Y.Y., L.S., R.G., M.P.-B., F.V., G.L., P.T., B.H.M., J.M.B., J.A.K.M., J.R.M., N.X., T.D., M.W., Y.L., A.A., J.H., M.P.-B., T.O.Y., N.C., M.H., A.F., C.P.; visualization: T.M.M., N.A.D., R.G., E.R., J.H., M.P.-B., F.V., F.K., P.T., T.O.Y., N.C., M.H., B.H.M., J.M.B., J.A.K.M., A.A., M.W.; supervision: T.M.M., D.H.K., A.F., G.L., J.R.M.; project administration: T.M.M.; funding

acquisition: T.M.M., D.H.K., J.M.B., G.K.-S.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by multiple funding agencies outlined with author in brackets. University of Nottingham Research Priority Area grant and the National Institute for Health Research (NIHR) Nottingham Digestive Diseases Biomedical Research Centre at Nottingham University Hospitals NHS Trust and University of Nottingham (T.M.M.); Litwin Initiative, Crohn's and Colitis Foundation (C.P.); NTU Quality Research (QR) funds (C.P., M.H.); DFG grant 409096610003 and DFG Excellence Cluster Precision Medicine in Chronic Inflammation Exc2167 (E.R., A.F.); Alberta Health Services and University of Alberta Hospital Foundation (D.H.K.); Fidaxomicin was provided by Merck (D.H.K.); Medical Research Council (MRC) Clinical Research Training Fellowship MR/R000875/1 and NIHR Academic Clinical Lectureship CL-2019-21-002 (B.H.M.); MRC and NIHR MC\_PC\_12025 (National Phenome Centre at Imperial College London); European Structural and Investment Funds CEKOM KK.01.2.2.03.0006, IRI grant KK.01.2.1.01.0003, and Croatian National Centre of Research Excellence in Personalized Healthcare grant KK.01.1.1.01.0010 (Genos Ltd., Zagreb, Croatia).

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of University of Alberta (protocol Pro00081229, date of approval 27 November 2018).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** All data can be found in supplementary material or can be made available upon request.

**Acknowledgments:** We are grateful to Matt Emberg, Melanie Lingaya and Yirga Falcone for their technical assistance in sample preparation. We would also like to thank Christine Loscher, Izabela Marszalowska and Mark Lynch at Dublin City University, and April Roberts of Public Health England for kindly gifting purified whole surface layer proteins for *C. difficile* ribotypes 001, 002 and 027 and whole toxins TcdA (Toxin A) and TcdB (Toxin B); toxinotype 0, strain VPI 10463, ribotype 087), respectively. We are grateful to Lauren Roberts for her assistance in graphic abstract preparation.

**Conflicts of Interest:** T.M.M. has received consultancy fees from Takeda. B.H.M. has received consultancy fees from Finch Therapeutics Group. J.R.M. has received consultancy fees from Cultech Ltd. Port Talbot, UK, and Enterobiotix Ltd. Glasgow, Scotland. G.L. is founder and CEO of Genos, a private research organization that specializes in high-throughput glycomic analysis and has several patents in this field. M.P.-B. and F.V. are employees of Genos. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. The remaining authors declare that they have no competing interests.

## References

1. Gravel, D.; Miller, M.; Simor, A.; Taylor, G.; Gardam, M.; McGeer, A.; Hutchinson, J.; Moore, D.; Kelly, S.; Boyd, D.; et al. Health Care–Associated *Clostridium difficile* Infection in Adults Admitted to Acute Care Hospitals in Canada: A Canadian Nosocomial Infection Surveillance Program Study. *Clin. Infect. Dis.* **2009**, *48*, 568–576. [[CrossRef](#)]
2. Kyne, L.; Hamel, M.B.; Polavaram, R.; Kelly, C.P. Health Care Costs and Mortality Associated with Nosocomial Diarrhea Due to *Clostridium difficile*. *Clin. Infect. Dis.* **2002**, *34*, 346–353. [[CrossRef](#)]
3. McDonald, L.C.; Gerding, D.N.; Johnson, S.; Bakken, J.S.; Carroll, K.C.; E Coffin, S.; Dubberke, E.R.; Garey, K.W.; Gould, C.V.; Kelly, C.; et al. Clinical Practice Guidelines for *Clostridium difficile* Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clin. Infect. Dis.* **2018**, *66*, 987–994. [[CrossRef](#)] [[PubMed](#)]
4. Martin, J.S.; Monaghan, T.M.; Wilcox, M.H. *Clostridium difficile* infection: Epidemiology, diagnosis and understanding transmission. *Nat. Rev. Gastroenterol. Hepatol.* **2016**, *13*, 206–216. [[CrossRef](#)]
5. Monaghan, T.M. New Perspectives in *Clostridium difficile* Disease Pathogenesis. *Infect. Dis. Clin. N. Am.* **2015**, *29*, 1–11. [[CrossRef](#)]
6. Baunwall, S.M.D.; Lee, M.M.; Eriksen, M.K.; Mullish, B.H.; Marchesi, J.R.; Dahlerup, J.F.; Hvas, C.L. Faecal microbiota transplantation for recurrent *Clostridioides difficile* infection: An updated systematic review and meta-analysis. *EClinicalMedicine* **2020**, *29–30*, 100642. [[CrossRef](#)] [[PubMed](#)]

7. Fischer, M.; Sipe, B.; Cheng, Y.-W.; Phelps, E.; Rogers, N.; Sagi, S.; Bohm, M.; Xu, H.; Kassam, Z. Fecal microbiota transplant in severe and severe-complicated *Clostridium difficile*: A promising treatment approach. *Gut Microbes* **2016**, *8*, 289–302. [[CrossRef](#)] [[PubMed](#)]
8. Weingarden, A.R.; Hamilton, M.J.; Sadowsky, M.J.; Khoruts, A. Resolution of Severe *Clostridium difficile* Infection Following Sequential Fecal Microbiota Transplantation. *J. Clin. Gastroenterol.* **2013**, *47*, 735–737. [[CrossRef](#)] [[PubMed](#)]
9. Alukal, J.; Dutta, S.K.; Surapaneni, B.K.; Le, M.; Tabbaa, O.; Phillips, L.; Mattar, M.C.; Phillips, L. Safety and efficacy of fecal microbiota transplant in 9 critically ill patients with severe and complicated *Clostridium difficile* infection with impending colectomy. *J. Dig. Dis.* **2019**, *20*, 301–307. [[CrossRef](#)]
10. Ianiro, G.; Masucci, L.; Quaranta, G.; Simonelli, C.; Lopetuso, L.R.; Sanguinetti, M.; Gasbarrini, A.; Cammarota, G. Randomised clinical trial: Faecal microbiota transplantation by colonoscopy plus vancomycin for the treatment of severe refractory *Clostridium difficile* infection—single versus multiple infusions. *Aliment. Pharmacol. Ther.* **2018**, *48*, 152–159. [[CrossRef](#)] [[PubMed](#)]
11. Khoruts, A.; Staley, C.; Sadowsky, M.J. Faecal microbiota transplantation for *Clostridioides difficile*: Mechanisms and pharmacology. *Nat. Rev. Gastroenterol. Hepatol.* **2020**, *18*, 67–80. [[CrossRef](#)] [[PubMed](#)]
12. Monaghan, T.; Mullish, B.H.; Patterson, J.; Wong, G.K.; Marchesi, J.R.; Xu, H.; Jilani, T.; Kao, D. Effective fecal microbiota transplantation for recurrent *Clostridioides difficile* infection in humans is associated with increased signalling in the bile acid-farnesoid X receptor-fibroblast growth factor pathway. *Gut Microbes* **2018**, *10*, 142–148. [[CrossRef](#)] [[PubMed](#)]
13. Weingarden, A.R.; Chen, C.; Bobr, A.; Yao, D.; Lu, Y.; Nelson, V.M.; Sadowsky, M.; Khoruts, A. Microbiota transplantation restores normal fecal bile acid composition in recurrent *Clostridium difficile* infection. *Am. J. Physiol. Liver Physiol.* **2014**, *306*, G310–G319. [[CrossRef](#)] [[PubMed](#)]
14. Mullish, B.H.; McDonald, J.A.K.; Pechlivanis, A.; Allegretti, J.R.; Kao, D.; Barker, G.F.; Kapila, D.; Petrof, E.O.; Joyce, S.A.; Gahan, C.; et al. Microbial bile salt hydrolases mediate the efficacy of faecal microbiota transplant in the treatment of recurrent *Clostridioides difficile* infection. *Gut* **2019**, *68*, 1791–1800. [[CrossRef](#)]
15. Allegretti, J.R.; Kearney, S.M.; Li, N.; Bogart, E.; Bullock, K.; Gerber, G.K.; Bry, L.; Clish, C.; Alm, E.J.; Korzenik, J. Recurrent-*Clostridium difficile* infection associates with distinct bile acid and microbiome profiles. *Aliment. Pharmacol. Ther.* **2016**, *43*, 1142–1153. [[CrossRef](#)] [[PubMed](#)]
16. Campbell, C.; McKenney, P.T.; Konstantinovskiy, D.; Isaeva, O.; Schizas, M.; Verter, J.; Mai, C.; Jin, W.-B.; Guo, C.-J.; Violante, S.; et al. Bacterial metabolism of bile acids promotes generation of peripheral regulatory T cells. *Nature* **2020**, *581*, 475–479. [[CrossRef](#)] [[PubMed](#)]
17. Seekatz, A.M.; Theriot, C.M.; Rao, K.; Chang, Y.-M.; Freeman, A.E.; Kao, J.Y.; Young, V.B. Restoration of short chain fatty acid and bile acid metabolism following fecal microbiota transplantation in patients with recurrent *Clostridium difficile* infection. *Anaerobe* **2018**, *53*, 64–73. [[CrossRef](#)] [[PubMed](#)]
18. McDonald, J.; Mullish, B.; Pechlivanis, A.; Liu, Z.; Brignardello, J.; Kao, D.; Holmes, E.; Li, J.; Clarke, T.B.; Thursz, M.; et al. Inhibiting Growth of *Clostridioides difficile* by Restoring Valerate, Produced by the Intestinal Microbiota. *Gastroenterology* **2018**, *155*, 1495–1507.e15. [[CrossRef](#)]
19. Smith, P.M.; Howitt, M.R.; Panikov, N.; Michaud, M.; Gallini, C.A.; Bohlooly-Y, M.; Glickman, J.N.; Garrett, W.S. The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis. *Science* **2013**, *341*, 569–573. [[CrossRef](#)]
20. Monaghan, T.M.; Pučić-Baković, M.; Vučković, F.; Lee, C.; Kao, D.; Wójcik, I.; Kliček, F.; Polytarchou, C.; Roach, B.; Louie, T.; et al. Decreased Complexity of Serum N-glycan Structures Associates with Successful Fecal Microbiota Transplantation for Recurrent *Clostridioides difficile* Infection. *Gastroenterology* **2019**, *157*, 1676–1678.e3. [[CrossRef](#)]
21. Abhyankar, M.M.; Ma, J.Z.; Scully, K.W.; Nafziger, A.J.; Frisbee, A.L.; Saleh, M.M.; Madden, G.R.; Hays, A.R.; Poulter, M.; Petri, W.A. Immune Profiling To Predict Outcome of *Clostridioides difficile* Infection. *mBio* **2020**, *11*, e00905-20. [[CrossRef](#)] [[PubMed](#)]
22. Rees, W.; Steiner, T.S. Adaptive immune response to *Clostridium difficile* infection: A perspective for prevention and therapy. *Eur. J. Immunol.* **2018**, *48*, 398–406. [[CrossRef](#)]
23. Monaghan, T.; Robins, A.; Knox, A.; Sewell, H.F.; Mahida, Y.R. Circulating Antibody and Memory B-Cell Responses to *C. difficile* Toxins A and B in Patients with *C. difficile*-Associated Diarrhoea, Inflammatory Bowel Disease and Cystic Fibrosis. *PLoS ONE* **2013**, *8*, e74452. [[CrossRef](#)]
24. Devera, T.S.; Lang, G.A.; Lanis, J.M.; Rampuria, P.; Gilmore, C.L.; James, J.A.; Ballard, J.D.; Lang, M.L. Memory B Cells Encode Neutralizing Antibody Specific for Toxin B from the *Clostridium difficile* Strains VPI 10463 and NAP1/BI/027 but with Superior Neutralization of VPI 10463 Toxin B. *Infect. Immun.* **2016**, *84*, 194–204. [[CrossRef](#)] [[PubMed](#)]
25. Amani, S.A.; Shadid, T.; Ballard, J.D.; Lang, M.L. *Clostridioides difficile* Infection Induces an Inferior IgG Response to That Induced by Immunization and Is Associated with a Lack of T Follicular Helper Cell and Memory B Cell Expansion. *Infect. Immun.* **2020**, *88*, e00829-19. [[CrossRef](#)]
26. Kao, D.; Roach, B.; Silva, M.; Beck, P.; Rioux, K.; Kaplan, G.G.; Chang, H.-J.; Coward, S.; Goodman, K.J.; Xu, H.; et al. Effect of Oral Capsule- vs. Colonoscopy-Delivered Fecal Microbiota Transplantation on Recurrent *Clostridium difficile* Infection: A Randomized Clinical Trial. *JAMA* **2017**, *318*, 1985–1993. [[CrossRef](#)] [[PubMed](#)]
27. Mullish, B.H.; Pechlivanis, A.; Barker, G.F.; Thursz, M.R.; Marchesi, J.; McDonald, J.A. Functional microbiomics: Evaluation of gut microbiota-bile acid metabolism interactions in health and disease. *Methods* **2018**, *149*, 49–58. [[CrossRef](#)]
28. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **2016**, *13*, 581–583. [[CrossRef](#)]



29. Akmačić, I.T.; Ugrina, I.; Štambuk, J.; Gudelj, I.; Vučković, F.; Lauc, G.; Pučić-Baković, M. High-throughput glycomics: Optimization of sample preparation. *Biochemistry* **2015**, *80*, 934–942. [[CrossRef](#)]
30. Šimurina, M.; de Haan, N.; Vučković, F.; Kennedy, N.A.; Štambuk, J.; Falck, D.; Trbojević-Akmačić, I.; Clerc, F.; Razdorov, G.; Khon, A.; et al. Glycosylation of Immunoglobulin G Associates With Clinical Features of Inflammatory Bowel Diseases. *Gastroenterology* **2018**, *154*, 1320–1333.e10. [[CrossRef](#)]
31. Monaghan, T.M.; Seekatz, A.M.; Markham, N.O.; Yau, T.O.; Hatzia Apostolou, M.; Jilani, T.; Christodoulou, N.; Roach, B.; Birli, E.; Pomenya, O.; et al. Fecal Microbiota Transplantation for Recurrent *Clostridioides difficile* Infection Associates With Functional Alterations in Circulating microRNAs. *Gastroenterology* **2021**, *161*, 255–270.e4. [[CrossRef](#)]
32. Negm, O.H.; Hamed, M.R.; Dilnot, E.M.; Shone, C.C.; Marszalowska, I.; Lynch, M.; Loscher, C.E.; Edwards, L.; Tighe, P.; Wilcox, M.H.; et al. Profiling Humoral Immune Responses to *Clostridium difficile*-Specific Antigens by Protein Microarray Analysis. *Clin. Vaccine Immunol.* **2015**, *22*, 1033–1039. [[CrossRef](#)] [[PubMed](#)]
33. Negm, O.H.; MacKenzie, B.; Hamed, M.R.; Ahmad, O.A.J.; Shone, C.C.; Humphreys, D.P.; Acharya, K.R.; Loscher, C.E.; Marszalowska, I.; Lynch, M.; et al. Protective antibodies against *Clostridium difficile* are present in intravenous immunoglobulin and are retained in humans following its administration. *Clin. Exp. Immunol.* **2017**, *188*, 437–443. [[CrossRef](#)] [[PubMed](#)]
34. Monaghan, T.M.; Negm, O.H.; MacKenzie, B.; Hamed, M.R.; Shone, C.C.; Humphreys, D.P.; Acharya, K.R.; Wilcox, M.H. High prevalence of subclass-specific binding and neutralizing antibodies against *Clostridium difficile* toxins in adult cystic fibrosis sera: Possible mode of immunoprotection against symptomatic *C. difficile* infection. *Clin. Exp. Gastroenterol.* **2017**, *10*, 169–175. [[CrossRef](#)]
35. Duggal, N.A.; Pollock, R.D.; Lazarus, N.R.; Harridge, S.; Lord, J.M. Major features of immunosenescence, including reduced thymic output, are ameliorated by high levels of physical activity in adulthood. *Aging Cell* **2018**, *17*, e12750. [[CrossRef](#)]
36. Duggal, N.A.; Upton, J.; Phillips, A.C.; Sapey, E.; Lord, J.M. An age-related numerical and functional deficit in CD19+CD24hiCD38hiB cells is associated with an increase in systemic autoimmunity. *Aging Cell* **2013**, *12*, 873–881. [[CrossRef](#)]
37. Pogorelyy, M.V.; Elhanati, Y.; Marcou, Q.; Sycheva, A.L.; Komech, E.A.; Nazarov, V.I.; Britanova, O.V.; Chudakov, D.M.; Mamedov, I.Z.; Lebedev, Y.B.; et al. Persisting fetal clonotypes influence the structure and overlap of adult human T cell receptor repertoires. *PLoS Comput. Biol.* **2017**, *13*, e1005572. [[CrossRef](#)] [[PubMed](#)]
38. Shugay, M.; Britanova, O.V.; Merzlyak, E.; Turchaninova, M.; Mamedov, I.Z.; Tuganbaev, T.R.; Bolotin, D.; Staroverov, D.; Putintseva, E.; Plevova, K.; et al. Towards error-free profiling of immune repertoires. *Nat. Methods* **2014**, *11*, 653–655. [[CrossRef](#)]
39. Bolotin, D.; Poslavsky, S.; Mitrophanov, I.; Shugay, M.; Mamedov, I.Z.; Putintseva, E.; Chudakov, D.M. MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods* **2015**, *12*, 380–381. [[CrossRef](#)]
40. Oksanen, J.B.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; Stevens, M.H.H.; et al. R Package Version 2.5-6. Vegan: Community Ecology Package. 2019. Available online: <http://cran.r-project.org> (accessed on 3 June 2021).
41. Kumar, L.; Futschik, M.E. Mfuzz: A software package for soft clustering of microarray data. *Bioinformatics* **2007**, *2*, 5–7. [[CrossRef](#)]
42. Müllner, D. Modern hierarchical, agglomerative clustering algorithms. *arXiv* **2011**, arXiv:1109.2378.
43. Chong, J.; Wishart, D.S.; Xia, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr. Protoc. Bioinform.* **2019**, *68*, e86. [[CrossRef](#)] [[PubMed](#)]
44. Seyda, M.; Quante, M.; Uehara, H.; Slegtenhorst, B.R.; Elkhali, A.; Tullius, S.G. Immunosenescence in renal transplantation: A changing balance of innate and adaptive immunity. *Curr. Opin. Organ Transplant.* **2015**, *20*, 417–423. [[CrossRef](#)]
45. Aiello, A.; Farzaneh, F.; Candore, G.; Caruso, C.; Davinelli, S.; Gambino, C.M.; Ligotti, M.E.; Zareian, N.; Accardi, G. Immunosenescence and Its Hallmarks: How to Oppose Aging Strategically? A Review of Potential Options for Therapeutic Intervention. *Front. Immunol.* **2019**, *10*, 2247. [[CrossRef](#)] [[PubMed](#)]
46. Sun, Y.; Coppé, J.-P.; Lam, E. Cellular Senescence: The Sought or the Unwanted? *Trends Mol. Med.* **2018**, *24*, 871–885. [[CrossRef](#)]
47. Larochette, V.; Miot, C.; Poli, C.; Beaumont, E.; Roingeard, P.; Fickenscher, H.; Jeannin, P.; Delneste, Y. IL-26, a Cytokine With Roles in Extracellular DNA-Induced Inflammation and Microbial Defense. *Front. Immunol.* **2019**, *10*, 204. [[CrossRef](#)] [[PubMed](#)]
48. Nakagawa, H.; Sasagawa, S.; Itoh, K. Sodium butyrate induces senescence and inhibits the invasiveness of glioblastoma cells. *Oncol. Lett.* **2017**, *15*, 1495–1502. [[CrossRef](#)] [[PubMed](#)]
49. Li, H.; Luo, Y.-F.; Wang, Y.-S.; Xiao, Y.-L.; Cai, H.-R.; Xie, C.-M. *Pseudomonas aeruginosa* induces cellular senescence in lung tissue at the early stage of two-hit septic mice. *Pathog. Dis.* **2018**, *76*, ftz001. [[CrossRef](#)]
50. Koch, S.; Larbi, A.; Özcelik, D.; Solana, R.; Gouttefangeas, C.; Attig, S.; Wikby, A.; Strindhall, J.; Franceschi, C.; Pawelec, G. Cytomegalovirus Infection: A Driving Force in Human T Cell Immunosenescence. *Ann. N. Y. Acad. Sci.* **2007**, *1114*, 23–35. [[CrossRef](#)]
51. Koide, N.; Morikawa, A.; Ito, H.; Sugiyama, T.; Hassan, F.; Islam, S.; Tumurkhuu, G.; Mori, I.; Yoshida, T.; Yokochi, T. Defective responsiveness of CD5+ B1 cells to lipopolysaccharide in cytokine production. *J. Endotoxin Res.* **2006**, *12*, 346–351. [[CrossRef](#)]
52. Heath, J.J.; Grant, M.D. The Immune Response Against Human Cytomegalovirus Links Cellular to Systemic Senescence. *Cells* **2020**, *9*, 766. [[CrossRef](#)] [[PubMed](#)]
53. Couzi, L.; Pitard, V.; Moreau, J.F.; Merville, P.; Dechanet-Merville, J. Direct and Indirect Effects of Cytomegalovirus-Induced gammadelta T Cells after Kidney Transplantation. *Front. Immunol.* **2015**, *6*, 3. [[CrossRef](#)] [[PubMed](#)]
54. Egli, A.; Humar, A.; Kumar, D. State-of-the-Art Monitoring of Cytomegalovirus-Specific Cell-Mediated Immunity After Organ Transplant: A Primer for the Clinician. *Clin. Infect. Dis.* **2012**, *55*, 1678–1689. [[CrossRef](#)]

55. Claesson, M.J.; Jeffery, I.B.; Conde, S.; Power, S.E.; O'Connor, E.M.; Cusack, S.; Harris, H.M.B.; Coakley, M.; Lakshminarayanan, B.; O'Sullivan, O.; et al. Gut microbiota composition correlates with diet and health in the elderly. *Nature* **2012**, *488*, 178–184. [[CrossRef](#)] [[PubMed](#)]
56. Thanissery, R.; Winston, J.A.; Theriot, C.M. Inhibition of spore germination, growth, and toxin activity of clinically relevant *C. difficile* strains by gut microbiota derived secondary bile acids. *Anaerobe* **2017**, *45*, 86–100. [[CrossRef](#)]
57. Tam, J.; Icho, S.; Utama, E.; Orrell, K.E.; Gómez-Biagi, R.F.; Theriot, C.M.; Kroh, H.K.; Rutherford, S.A.; Lacy, D.B.; Melnyk, R.A. Intestinal bile acids directly modulate the structure and function of *C. difficile* TcdB toxin. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 6792–6800. [[CrossRef](#)] [[PubMed](#)]
58. Weingarden, A.R.; Dosa, P.; Dewinter, E.; Steer, C.J.; Shaughnessy, M.K.; Johnson, J.R.; Khoruts, A.; Sadowsky, M.J. Changes in Colonic Bile Acid Composition following Fecal Microbiota Transplantation Are Sufficient to Control *Clostridium difficile* Germination and Growth. *PLoS ONE* **2016**, *11*, e0147210. [[CrossRef](#)]
59. Martinez-Gili, L.; McDonald, J.A.K.; Liu, Z.; Kao, D.; Allegretti, J.R.; Monaghan, T.M.; Barker, G.F.; Blanco, J.M.; Williams, H.R.T.; Holmes, E.; et al. Understanding the mechanisms of efficacy of fecal microbiota transplant in treating recurrent *Clostridioides difficile* infection and beyond: The contribution of gut microbial-derived metabolites. *Gut Microbes* **2020**, *12*, 1810531. [[CrossRef](#)]
60. Sinha, S.R.; Haileselassie, Y.; Nguyen, L.P.; Tropini, C.; Wang, M.; Becker, L.S.; Sim, D.; Jarr, K.; Spear, E.T.; Singh, G.; et al. Dysbiosis-Induced Secondary Bile Acid Deficiency Promotes Intestinal Inflammation. *Cell Host Microbe* **2020**, *27*, 659–670.e5. [[CrossRef](#)]
61. Glaben, R.; Batra, A.; Fedke, I.; Zeitz, M.; Lehr, H.A.; Leoni, F.; Mascagni, P.; Fantuzzi, G.; Dinarello, C.A.; Siegmund, B. Histone Hyperacetylation Is Associated with Amelioration of Experimental Colitis in Mice. *J. Immunol.* **2006**, *176*, 5015–5022. [[CrossRef](#)]
62. Gudelj, I.; Lauc, G.; Pezer, M. Immunoglobulin G glycosylation in aging and diseases. *Cell. Immunol.* **2018**, *333*, 65–79. [[CrossRef](#)] [[PubMed](#)]
63. Johal, S.S.; Lambert, C.; Hammond, J.; James, P.D.; Borriello, S.P.; Mahida, Y.R. Colonic IgA producing cells and macrophages are reduced in recurrent and non-recurrent *Clostridium difficile* associated diarrhoea. *J. Clin. Pathol.* **2004**, *57*, 973–979. [[CrossRef](#)] [[PubMed](#)]
64. Leslie, J.L.; Vendrov, K.C.; Jenior, M.L.; Young, V.B. The Gut Microbiota Is Associated with Clearance of *Clostridium difficile* Infection Independent of Adaptive Immunity. *mSphere* **2019**, *4*, e00698-18. [[CrossRef](#)]
65. Sutton, T.C.; Lamirande, E.W.; Bock, K.W.; Moore, I.N.; Koudstaal, W.; Rehman, M.; Weverling, G.J.; Goudsmit, J.; Subbarao, K. In Vitro Neutralization Is Not Predictive of Prophylactic Efficacy of Broadly Neutralizing Monoclonal Antibodies CR6261 and CR9114 against Lethal H2 Influenza Virus Challenge in Mice. *J. Virol.* **2017**, *91*, e01603-17. [[CrossRef](#)]
66. Bootz, A.; Karbach, A.; Spindler, J.; Kropff, B.; Reuter, N.; Sticht, H.; Winkler, T.H.; Britt, W.J.; Mach, M. Protective capacity of neutralizing and non-neutralizing antibodies against glycoprotein B of cytomegalovirus. *PLoS Pathog.* **2017**, *13*, e1006601. [[CrossRef](#)] [[PubMed](#)]
67. Steele, J.; Chen, K.; Sun, X.; Zhang, Y.; Wang, H.; Tzipori, S.; Feng, H. Systemic Dissemination of *Clostridium difficile* Toxins A and B Is Associated With Severe, Fatal Disease in Animal Models. *J. Infect. Dis.* **2011**, *205*, 384–391. [[CrossRef](#)]
68. Lawrence, S.J.; Dubberke, E.R.; Johnson, S.; Gerding, D.N. *Clostridium difficile*—Associated Disease Treatment Response Depends on Definition of Cure. *Clin. Infect. Dis.* **2007**, *45*, 1648. [[CrossRef](#)]
69. A Cohen, N.; Miller, T.; Na' Aminh, W.; Hod, K.; Adler, A.; Cohen, D.; Guzner-Gur, H.; Santo, E.; Halpern, Z.; Carmeli, Y.; et al. *Clostridium difficile* fecal toxin level is associated with disease severity and prognosis. *United Eur. Gastroenterol. J.* **2018**, *6*, 773–780. [[CrossRef](#)]
70. Mensen, A.; Oh, Y.; Becker, S.C.; Hemmati, P.G.; Jehn, C.; Westermann, J.; Szyska, M.; Göldner, H.; Dörken, B.; Scheibenbogen, C.; et al. Apoptosis Susceptibility Prolongs the Lack of Memory B Cells in Acute Leukemic Patients After Allogeneic Hematopoietic Stem Cell Transplantation. *Biol. Blood Marrow Transplant.* **2015**, *21*, 1895–1906. [[CrossRef](#)]
71. Shah, H.B.; Smith, K.; Scott, E.J.; Larabee, J.L.; James, J.A.; Ballard, J.D.; Lang, M.L. Human *C. difficile* toxin-specific memory B cell repertoires encode poorly neutralizing antibodies. *JCI Insight* **2020**, *5*, 138137. [[CrossRef](#)]
72. Beckett, J.; Hester, J.; Issa, F.; Shankar, S. Regulatory B cells in transplantation: Roadmaps to clinic. *Transpl. Int.* **2020**, *33*, 1353–1368. [[CrossRef](#)]
73. Burrello, C.; Garavaglia, F.; Cribiù, F.M.; Ercoli, G.; Lopez, G.; Troisi, J.; Colucci, A.; Guglietta, S.; Carloni, S.; Guglielmetti, S.; et al. Therapeutic faecal microbiota transplantation controls intestinal inflammation through IL10 secretion by immune cells. *Nat. Commun.* **2018**, *9*, 5184. [[CrossRef](#)] [[PubMed](#)]
74. Amrouche, K.; Pers, J.-O.; Jamin, C. Glatiramer Acetate Stimulates Regulatory B Cell Functions. *J. Immunol.* **2019**, *202*, 1970–1980. [[CrossRef](#)] [[PubMed](#)]
75. Vaiserman, A.M.; Koliada, A.; Marotta, F. Gut microbiota: A player in aging and a target for anti-aging intervention. *Ageing Res. Rev.* **2017**, *35*, 36–45. [[CrossRef](#)] [[PubMed](#)]
76. Bárcena, C.; Valdés-Mas, R.; Mayoral, P.; Garabaya, C.; Durand, S.; Rodríguez, F.; Fernández-García, M.T.; Salazar, N.; Nogacka, A.M.; Garatachea, N.; et al. Healthspan and lifespan extension by fecal microbiota transplantation into progeroid mice. *Nat. Med.* **2019**, *25*, 1234–1242. [[CrossRef](#)] [[PubMed](#)]
77. Stebbeg, M.; Silva-Cayetano, A.; Innocentin, S.; Jenkins, T.P.; Cantacessi, C.; Gilbert, C.; Linterman, M.A. Heterochronic faecal transplantation boosts gut germinal centres in aged mice. *Nat. Commun.* **2019**, *10*, 2443. [[CrossRef](#)]

78. Schluter, J.; Peled, J.; Taylor, B.P.; Markey, K.A.; Smith, J.A.; Taur, Y.; Niehus, R.; Staffas, A.; Dai, A.; Fontana, E.; et al. The gut microbiota is associated with immune cell dynamics in humans. *Nature* **2020**, *588*, 303–307. [[CrossRef](#)]
79. Gopalakrishnan, V.; Spencer, C.N.; Nezi, L.; Reuben, A.; Andrews, M.C.; Karpinets, T.V.; Prieto, P.A.; Vicente, D.; Hoffman, K.; Wei, S.C.; et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* **2018**, *359*, 97–103. [[CrossRef](#)]
80. Routy, B.; Le Chatelier, E.; DeRosa, L.; Duong, C.P.M.; Alou, M.T.; Daillère, R.; Fluckiger, A.; Messaoudene, M.; Rauber, C.; Roberti, M.P.; et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **2017**, *359*, 91–97. [[CrossRef](#)] [[PubMed](#)]
81. Feuth, T.; van Baarle, D.; van Erpecum, K.J.; Siersema, P.D.; Hoepelman, A.I.M.; Arends, J.E. CD4/CD8 ratio is a promising candidate for non-invasive measurement of liver fibrosis in chronic HCV-monoinfected patients. *Eur. J. Clin. Microbiol. Infect. Dis.* **2014**, *33*, 1113–1117. [[CrossRef](#)] [[PubMed](#)]

# **Fecal microbiota transplantation for recurrent *Clostridioides difficile* infection associates with functional alterations in circulating microRNAs**

Tanya M Monaghan<sup>#</sup>, Anna M Seekatz<sup>\*</sup>, Nicholas O Markham<sup>\*</sup>, **Tung On Yau<sup>\*</sup>**, Maria Hatzia Apostolou<sup>\*</sup>, Tahseen Jilani, Niki Christodoulou, Brandi Roach, Eleni Birli, Odette Pomenya, Thomas Louie, D Borden Lacy, Peter Kim, Christine Lee, Dina Kao<sup>#</sup>, Christos Polytarchou<sup>#</sup>

*Gastroenterology*. **2021**; 161(1):255-270.e4..

Journal URL: [gastrojournal.org/article/S0016-5085\(21\)00577-1/fulltext](https://gastrojournal.org/article/S0016-5085(21)00577-1/fulltext)

ScienceDirect:

[sciencedirect.com/science/article/abs/pii/S0016508521005771](https://sciencedirect.com/science/article/abs/pii/S0016508521005771)

DOI: [10.1053/j.gastro.2021.03.050](https://doi.org/10.1053/j.gastro.2021.03.050)

PMID: [33844988](https://pubmed.ncbi.nlm.nih.gov/33844988/)

31<sup>st</sup> July, 2021

To Whom It May Concern,

**Statement of Joint Authorship**

**Title of publication:** Fecal microbiota transplantation for recurrent *Clostridioides difficile* infection associates with functional alterations in circulating microRNAs

**Journal:** Gastroenterology

**Publication date:** 9<sup>th</sup> April, 2021

**Issue:** Volume 161, Issue 81, P255-270.E1

**PubMed ID:** 33921348

**Authors:** Tanya M Monaghan\*, Anna M Seekatz<sup>#</sup>, Nicholas O Markham<sup>#</sup>, **Tung On Yau**<sup>#</sup>, Maria Hatzia Apostolou<sup>#</sup>, Tahseen Jilani, Niki Christodoulou, Brandi Roach, Eleni Birli, Odette Pomenya, Thomas Louie, D Borden Lacy, Peter Kim, Christine Lee, Dina Kao\*, Christos Polytarchou\*

\* co-corresponding authors; <sup>#</sup>equal contribution

I hereby confirm that Mr. Tung On YAU is one of the co-first authors in the above publication. He was the key contributor to data management, perform experiments, data acquisition and collection. This included colonic cell culture, RNA extraction on sera samples, microRNA profiling on NanoString nCounter platform, qPCR, caspase 3/7 activity assay and luciferase reporter assays to evaluate transcriptional activity on target genes. He also involved in communication between institutions, sample management, and data visualisation.

Yours sincerely,



Christos Polytarchou, PhD  
Associate Professor  
John van Geest Cancer Research Centre  
School of Science and Technology  
Nottingham Trent University  
Nottingham  
NG11 8NS  
Email: [christos.polytarchou@ntu.ac.uk](mailto:christos.polytarchou@ntu.ac.uk)

Nottingham Trent University  
Burton Street, Nottingham, NG1 4BU, UK  
Tel. +44 (0)115 941 8418  
[www.ntu.ac.uk/vangeest](http://www.ntu.ac.uk/vangeest)





**University of  
Nottingham**  
UK | CHINA | MALAYSIA

Dr Tanya Monaghan  
Faculty of Medicine & Health Sciences  
Room W/E 1381 NDDC  
Queen's Medical Centre  
Nottingham  
NG7 2UH  
UK

[Tanya.Monaghan@nottingham.ac.uk](mailto:Tanya.Monaghan@nottingham.ac.uk)  
+44(0)115 9249924 x60589

April 12<sup>th</sup>, 2022

Re: Full Publications

To Whom It May Concern:

I am writing to confirm that Tung On (Payton) Yau, previously a Research Fellow in the laboratory of Dr Christos Polytarchou at Nottingham Trent University (currently serving as an Hourly-Paid Lecturer at Nottingham Trent University and a Post-doctoral Researcher at Scotland's Rural College), collaborated with my team on projects relating to microRNAs, microbiota, public health and *Clostridioides difficile* infection from the year 2017 to 2021. He was a significant contributor, with a role in literature search and acquisition, experimental design, clinical sample management, and performing experiments. These experiments include cell culture, RNA extraction from serum and peripheral blood mononuclear cells, high-throughput Nanostring miRNA analysis, Bio-Plex cytokine assays, and luciferase reporter assays for the identification of miRNA targets. He was subsequently involved in data analysis and visual representations of the data for publications.

He is a co-author of three published publications:

1. Fecal microbiota transplantation for recurrent *Clostridioides difficile* infection associates with functional alterations in circulating microRNAs. **Gastroenterology** **161** (1), 255-270
2. Multiomics profiling reveals signatures of dysmetabolism in urban populations in Central India. **Microorganisms** **9** (7), 1485
3. A multi-factorial observational study on sequential fecal microbiota transplant in patients with medically refractory *Clostridioides difficile* infection. **Cells** **10** (11), 3234

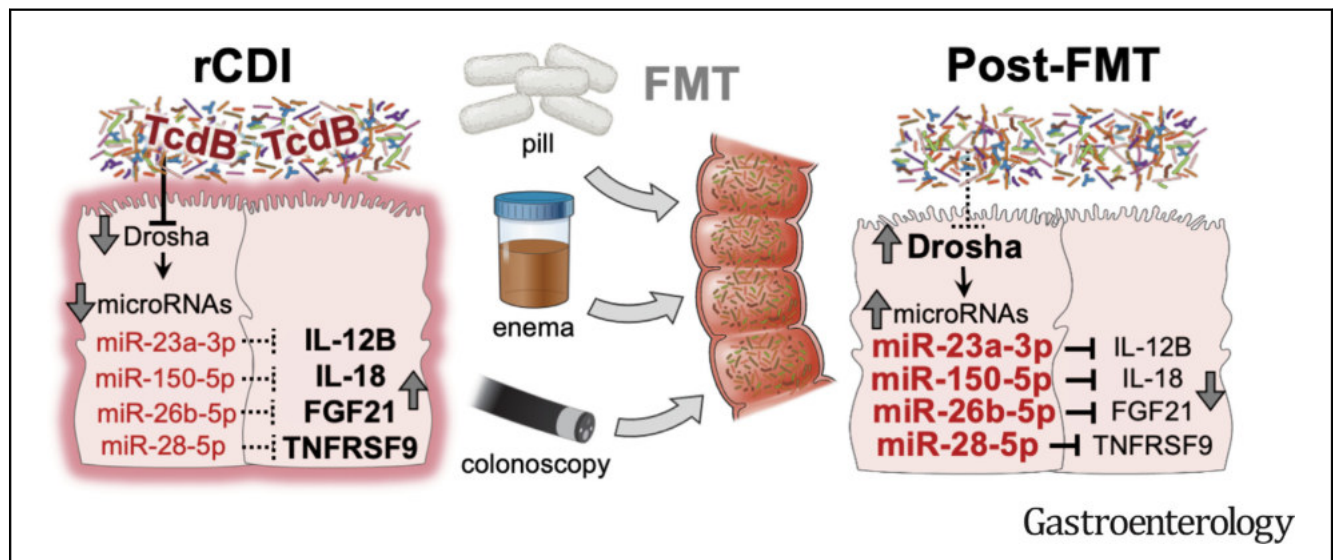
Yours Sincerely,



# Fecal Microbiota Transplantation for Recurrent *Clostridioides difficile* Infection Associates With Functional Alterations in Circulating microRNAs

Tanya M. Monaghan,<sup>1,2</sup> Anna M. Seekatz,<sup>3</sup> Nicholas O. Markham,<sup>4,5</sup> Tung On Yau,<sup>6</sup> Maria Hatzia Apostolou,<sup>6</sup> Tahseen Jilani,<sup>7</sup> Niki Christodoulou,<sup>6</sup> Brandi Roach,<sup>8</sup> Eleni Birli,<sup>6</sup> Odette Pomenya,<sup>6</sup> Thomas Louie,<sup>9</sup> D. Borden Lacy,<sup>10,11</sup> Peter Kim,<sup>12</sup> Christine Lee,<sup>13,14</sup> Dina Kao,<sup>8</sup> and Christos Polytaichou<sup>6</sup>

<sup>1</sup>National Institute for Health Research Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, United Kingdom; <sup>2</sup>Nottingham Digestive Diseases Centre, School of Medicine, University of Nottingham, Nottingham, United Kingdom; <sup>3</sup>Department of Biological Sciences, Clemson University, Clemson, South Carolina, USA; <sup>4</sup>Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA; <sup>5</sup>Epithelial Biology Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA; <sup>6</sup>Department of Biosciences, John van Geest Cancer Research Centre, Centre for Health Aging and Understanding Disease, School of Science and Technology, Nottingham Trent University, Nottingham, United Kingdom; <sup>7</sup>Advanced Data Analysis Centre, School of Computer Science, University of Nottingham, Nottingham, United Kingdom; <sup>8</sup>Department of Medicine, University of Alberta, Edmonton, Alberta, Canada; <sup>9</sup>Department of Microbiology and Infectious Diseases, University of Calgary, Calgary, Alberta, Canada; <sup>10</sup>Department of Pathology, Microbiology, and Immunology, Vanderbilt University School of Medicine, Nashville, Tennessee, USA; <sup>11</sup>Veterans Affairs Tennessee Valley Healthcare System, Nashville, Tennessee, USA; <sup>12</sup>Department of Mathematics and Statistics, University of Guelph, Ontario, Canada; <sup>13</sup>Vancouver Island Health Authority, Victoria, British Columbia, Canada; and <sup>14</sup>Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada



**BACKGROUND AND AIMS:** The molecular mechanisms underlying successful fecal microbiota transplantation (FMT) for recurrent *Clostridioides difficile* infection (rCDI) remain poorly understood. The primary objective of this study was to characterize alterations in microRNAs (miRs) following FMT for rCDI. **METHODS:** Sera from 2 prospective multicenter randomized controlled trials were analyzed for miRNA levels with the use of the Nanostring nCounter platform and quantitative reverse-transcription (RT) polymerase chain reaction (PCR). In addition, rCDI-FMT and toxin-treated animals and ex vivo human colonoids were used to compare intestinal tissue and circulating miRs. miR inflammatory gene targets in colonic

epithelial and peripheral blood mononuclear cells were evaluated by quantitative PCR (qPCR) and 3'UTR reporter assays. Colonic epithelial cells were used for mechanistic, cytoskeleton, cell growth, and apoptosis studies. **RESULTS:** miRNA profiling revealed up-regulation of 64 circulating miRs 4 and 12 weeks after FMT compared with screening, of which the top 6 were validated in the discovery cohort by means of RT-qPCR. In a murine model of relapsing-CDI, RT-qPCR analyses of sera and cecal RNA extracts demonstrated suppression of these miRs, an effect reversed by FMT. In mouse colon and human colonoids, *C. difficile* toxin B (TcdB) mediated the suppressive effects of CDI on miRs. CDI dysregulated DROSHA, an effect reversed by FMT.

Correlation analyses, qPCR, and 3'UTR reporter assays revealed that miR-23a, miR-150, miR-26b, and miR-28 target directly the 3'UTRs of IL12B, IL18, FGF21, and TNFRSF9, respectively. miR-23a and miR-150 demonstrated cytoprotective effects against TcdB. **CONCLUSIONS:** These results provide novel and provocative evidence that modulation of the gut microbiome via FMT induces alterations in circulating and intestinal tissue miRs. These findings contribute to a greater understanding of the molecular mechanisms underlying FMT and identify new potential targets for therapeutic intervention in rCDI.

**Keywords:** Fecal Transplantation; *C difficile*; microRNA; DROSHA.

**F**ecal microbiota transplantation (FMT) is a well established treatment for recurrent *Clostridioides difficile* infection (rCDI). Accumulating evidence also supports FMT as a potential treatment for other disorders associated with intestinal dysbiosis, including inflammatory bowel diseases, cancer, metabolic syndrome, and neuropsychiatric disorders.<sup>1</sup>

Despite the effectiveness of FMT in rCDI, its mechanisms of action remain poorly explored. Current evidence suggests the success of FMT may be attributed in part to the reconstitution of intestinal microbiota, restoration of secondary bile acid metabolism, and modulation of immune-mediated inflammatory responses.<sup>2,3</sup> We have previously reported that effective FMT for rCDI is associated with activation of the bile acid–farnesoid X receptor–fibroblast growth factor (FGF) pathway and decreased serum C-X-C motif chemokine 11 (CXCL11), interleukin-18 (IL18), tumor necrosis factor (TNF)-related activation-induced cytokine, IL12B, CXCL6, and *tnf* receptor superfamily member 9 (TNFRSF9).<sup>4</sup> The gut microbiota can modify host cell responses to stimuli (eg, metabolites) through alterations in the host epigenome and, ultimately, gene expression.<sup>5</sup> MicroRNAs (miRNAs) are thought to be one way in which the gut microbiota communicates with the human host. These short noncoding RNA molecules (containing ~22 nucleotides) are expressed as individual genes or as parts of longer transcripts and are processed by machinery involving DROSHA and DICER nucleases, which generate the mature miRNAs. Mature miRNAs are loaded on Argonaute AGO-containing complexes and bind to complementary sequences in the 3'-untranslated region (3'UTR) of messenger RNAs (mRNAs), resulting in transcript degradation and translational suppression of target genes.<sup>6</sup> Bacterial pathogens clearly alter host miRNA expression,<sup>7</sup> but less is known regarding the effect of commensal bacteria on the host miRNAome. A recent study has demonstrated how host fecal miRNAs, normal components of feces, can enter certain bacteria (eg, *Fusobacterium nucleatum* and *Escherichia coli*) and regulate bacterial gene transcription and growth.<sup>8</sup> However, the alterations in circulating miRNAs of rCDI patients undergoing FMT and the functional effects they may exert on downstream targets remain unknown.

Herein, we characterize the impact of FMT on circulating miRNA signatures to better understand

## WHAT YOU NEED TO KNOW

### BACKGROUND AND CONTEXT

Fecal microbiota transplantation (FMT) is highly effective at preventing recurrent *Clostridioides difficile* infection (rCDI). However, the mechanisms of action remain largely unknown. MicroRNAs (miRNAs), short noncoding RNA sequences that bind to complementary sequences of mRNA and can regulate gene expression, may be a potential mechanism by which commensal microbiota communicate with the human host.

### NEW FINDINGS

We identified several significant alterations in circulating miRNAs following successful FMT treatment in 2 independent rCDI patient cohorts. miRNA signatures were validated in animal models and human colonoids. We further demonstrated that FMT-regulated miRNAs regulate cell properties and target IL12B, IL18, FGF21, and TNFRSF9, integral in pathways linking to inflammation, autoimmunity, and cancer.

### LIMITATIONS

Deeper characterization of the epitranscriptome in FMT is required.

### IMPACT

These results describe a new mechanism of action of FMT against rCDI and provide potential new therapeutic targets for conditions associated with intestinal dysbiosis.

immunological mechanisms relevant to FMT in the treatment of rCDI. Our findings suggest a conserved mechanism involved in regulating host miRNAs by *C difficile* and identify new miRNA inflammatory targets in response to FMT.

## Materials and Methods

### Participants' Clinical Data, Sample Collection, and Storage

Randomly selected rCDI subjects participating in 2 clinical trials (capsule- vs colonoscopy-delivered FMT<sup>9</sup> and fresh vs frozen enema-delivered FMT<sup>10</sup>) comprised the discovery and replication cohorts, respectively. Blood samples were collected from October 2014 to December 2016 (discovery cohort) and from July 2012 to September 2014 (replication cohort) and stored at  $-80^{\circ}\text{C}$ . Only sera with sufficient volume were selected for miRNA analysis. Healthy control subjects ( $n = 42$ ; mean  $\pm$  SD age  $53.3 \pm 20.7$  years; 30 women [71.4%]) were defined as asymptomatic adults undergoing screening colonoscopy recruited in Edmonton, Canada. Clinical and demographic information were collected from medical records. Participant

**Abbreviations used in this paper:** AGO, argonaute; FGF, fibroblast growth factor; FMT, fecal microbiota transplantation; IL, interleukin; pri-miRNA, primary microRNA transcript; rCDI, recurrent *Clostridioides difficile* infection; Tcd, *C difficile* toxin; TNF, tumor necrosis factor; TNFRSF9, TNF receptor superfamily member 9.

 Most current article

© 2021 by the AGA Institute  
0016-5085/\$36.00

<https://doi.org/10.1053/j.gastro.2021.03.050>



baseline characteristics are presented in Table 1. All participants provided written informed consent under the approvals granted by the Research Ethics Boards of the University of Alberta (Pro 1994 and 49006), St. Joseph’s Healthcare (11-3622), and Hamilton Health Sciences (12-505) in Canada.

**Mouse Model of rCDI**

A mouse model of rCDI was used to assess miRNA, mRNA, and protein levels following infection and treatment with FMT, as described previously.<sup>11</sup> Animal work was approved by the Clemson University Institutional Animal Care and Use Committee (IACUC). Additional details are available in the Supplementary Methods. Briefly, 6–8-week-old C57BL/6 mice were given 0.5 mg/mL cefoperazone (MP Biochemicals; cat. no. 199695) in sterile drinking water (Gibco Laboratories; cat. no. 15230) ad libitum for 5 days (n = 42). Two days after cessation of antibiotics, the mice received 10<sup>3</sup> *C difficile* strain 630 spores resuspended in 1 mL sterile water, prepped as described

previously (day 0; n = 36).<sup>12</sup> A subset of mice were killed 4 days after infection (dpi) to assess miRNA during acute infection (n = 9). The remainder of the mice received 0.4 mg/mL vancomycin (Sigma; cat no. V2002) starting at 4 dpi for 5 days (4–9 dpi) ad libitum in sterile drinking water (n = 27). At 11 dpi, FMT prepped from untreated (n = 8) healthy age-matched mice (mFMT) was administered via oral gavage to a group of mice. Each mouse received 100 μL FMT material diluted in phosphate-buffered saline solution (PBS) (~0.2 g fresh fecal material in 1.5 mL prerduced PBS, homogenized via mixing and gravity filtering). One group of mice received all antibiotics and mFMT but no *C difficile* inoculum (handling and experimental control; n = 6). The remainder of infected mice (n = 12) did not receive FMT (noFMT). Fecal sampling was conducted throughout the experiment, and end point cecal sampling to assess *C difficile* load was determined by plating 20 μL of content from individual samples in 1:10 PBS and serially diluted on taurocholate cycloserine cefoxitin fructose agar under anaerobic conditions (Coy Laboratory Products, Grass Lake,

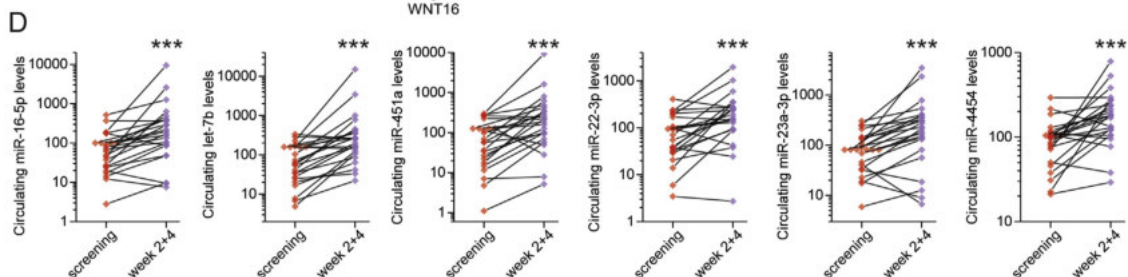
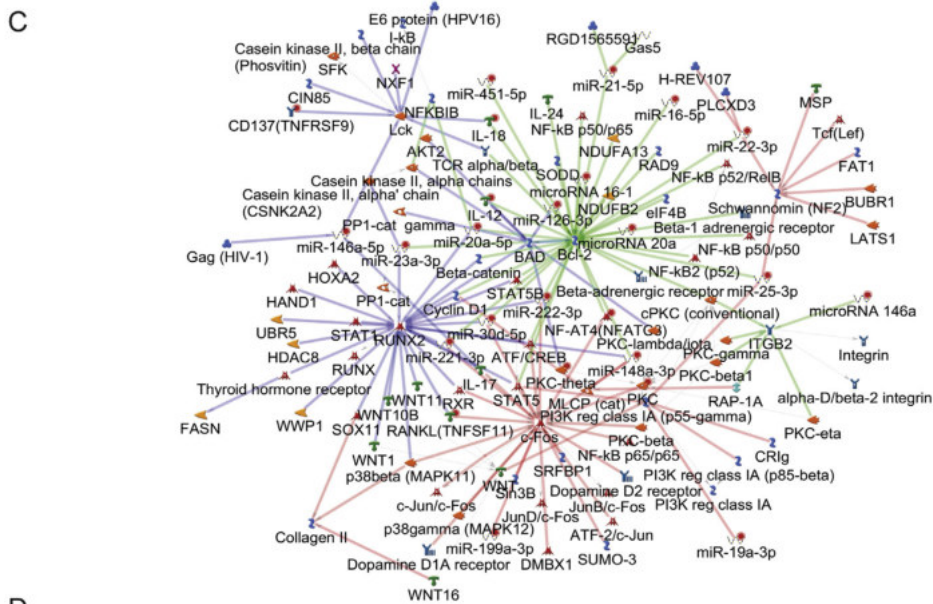
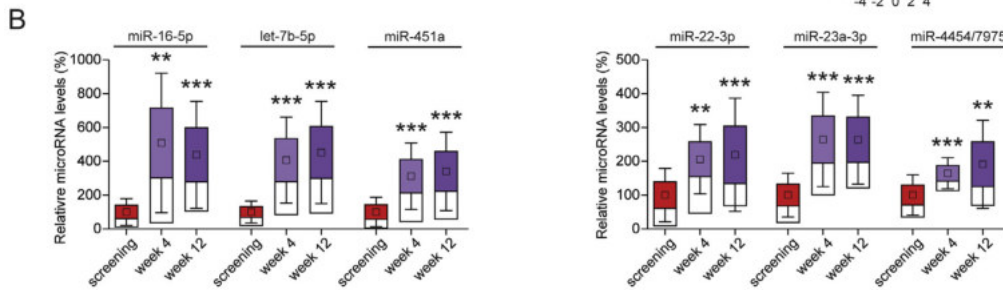
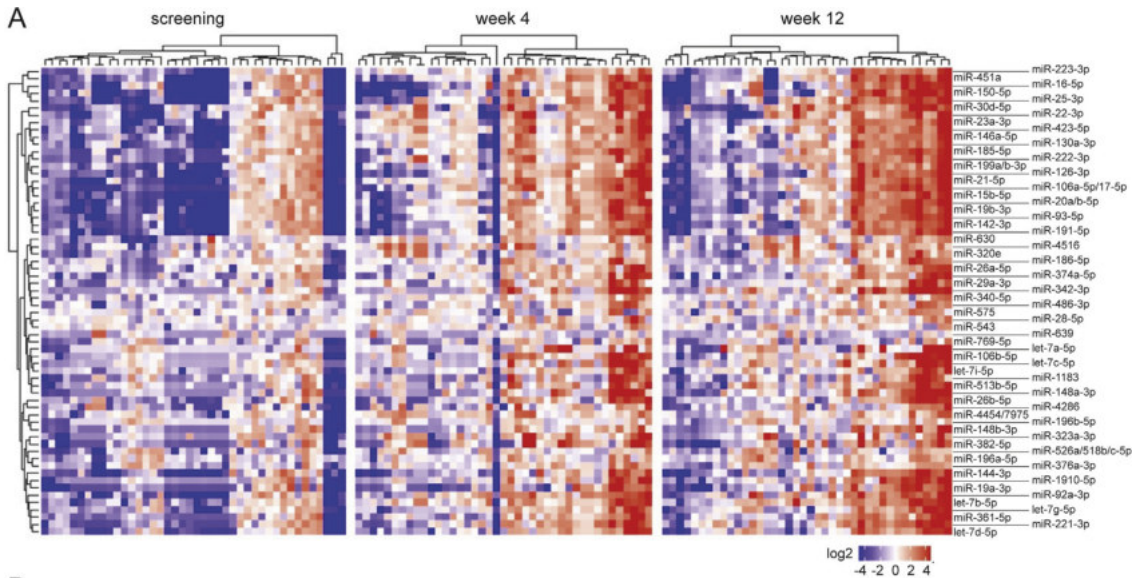
**Table 1.** Participant Baseline Characteristics for the Discovery and Replication Cohorts

	Discovery cohort—NCT02254811			Replication cohort—NCT01398969		
	Capsule (n = 25)	Colonoscopy (n = 17)	P value	Fresh (n = 11)	Frozen (n = 13)	P value
Age, y	59.0 ± 19.7)	56.4 ± 8.7	.6624	74.3 ± 14.1	71.6 ± 18.6	.7363
Female	21 (84.0%)	9 (52.9%)	.0659	4 (36.4%)	9 (69.2%)	.2305
Charlson comorbidity index	3 (2–5)	3 (0–4)	.4353			
Immunosuppressed patients	2 (8.0%)	3 (17.6%)	.6439	3 (27.3%)	1 (7.7%)	.4637
Use of immune modulator						
Corticosteroid	0 (0%)	1 (5.9%)	.8443			
Immunosuppressant	1 (4.0%)	1 (5.9%)	1	4 (36.4%)	3 (23.1%)	.7926
Biologic	2 (8.0%)	1 (5.9%)	1	5 (45.5%)	4 (30.8%)	.751
Body mass index, kg/m <sup>2</sup>	25.3 ± 6.6	27.0 ± 4.1	.3113			
Inpatient status at screening	3 (12%)	0 (0%)	.3833	7 (63.6%)	6 (46.2%)	.6561
PPI use prior to FMT	4 (16%)	1 (5.9%)	.6111	5 (45.5%)	5 (38.5%)	1
No. of rCDI episodes before FMT; median	3 (3–4)	3 (3–4)	.3893	3 (2.5–4.0)	3 (3–4)	.4641
Duration of rCDI before FMT	65 (49–96)	70 (57–135)	.2648			
No. of CDI-related hospital admissions	1 (0–2)	0 (0–0)	.0057			
Inflammatory bowel disease						
Ulcerative colitis	2 (8.3%)	3 (1.8%)	.6439			
Crohn’s disease	2 (8.3%)	1 (5.9%)	1	2 (18.2%)	2 (15.4%)	1
Hemoglobin, g/dL	137 (129–144)	138 (132–144)	.6079			
WBC, 10 <sup>9</sup> /L	8 (6.8–8.6)	6.5 (5.1–6.9)	.0143	15.30 (8.35–21.70)	8.9 (6.7–12.6)	.2212
Albumin, g/L	40.5 (38.8–43.3)	40 (38–42)	.5857	31 (28–34)	30 (24–34)	.8557
CRP, mg/L	1.9 (0.9–5.8)	6.2 (1.3–10.2)	.244			
Creatinine (mg/dL)	70 (58.75–76)	70 (59–84)	.7404	88 (69.5–127)	85 (64–128)	.6849

Values are mean ± SD, n (%), or median (range).

CDI, *Clostridioides difficile* infection; CRP, C-reactive protein; FMT, fecal microbiota transplantation; PPI, proton pump inhibitor; rCDI, recurrent *Clostridioides difficile* infection; WBC, white blood count.

BASIC AND TRANSLATIONAL AT



MI, USA). The colony-forming units (CFU)/mL content was determined after overnight incubation at 37°C. Mice were killed at 21 dpi, and cecal contents, tissue, and serum were flash frozen in liquid nitrogen and kept at -80°C for downstream analyses.

### Mouse Toxin Challenge Model

The animal protocol was approved by the Vanderbilt University Medical Center IACUC; 6–8-week-old C57BL/6 mice (Jackson Labs) were observed from arrival to ensure normal health. Mice were separated into 3 groups ( $n = 6$  per group) to receive intrarectal instillations of either purified recombinant whole *C difficile* toxin B(TcdB), TcdA and TcdB, or Hank's Balanced Salt Solution (HBSS) vehicle control, as described elsewhere.<sup>13</sup> Further details on the purification of the toxins are described in the [Supplementary methods](#). Toxins were derived from the VPI 10463 *C difficile* reference strain and prepared as 15  $\mu\text{g}$  in a total volume of 100  $\mu\text{L}$  per instillation. Mice were anesthetized with isoflurane and confirmed to be sedated by toe pinch. One mL HBSS was instilled intrarectally to evacuate stools with a flexible plastic gavage applicator (20 G  $\times$  30 mm; [gavageneedle.com](#)). Instillation was performed over 30 seconds while lightly pinching closed the anus, which was held for an additional 30 seconds as previously described.<sup>13</sup> Mice were returned to cages to recover. After 2–5 hours, mice were killed by CO<sub>2</sub> inhalation and cervical dislocation. Whole blood was extracted via cardiac puncture and allowed to clot in RNase-free microcentrifuge tubes for 15 minutes at room temperature before centrifugation for 15 minutes at 1500g at 4°C. Serum was transferred to a fresh tube and flash frozen in liquid nitrogen. The colon was isolated and dissected from surrounding visceral tissue. The whole colon was washed in chilled sterile 1 $\times$  PBS before portions of the middle and distal colon were combined and flash frozen for protein and miRNA analysis.

### Treatment of Human Colonoids With *C difficile* Toxins

Deidentified human colon tissue was obtained through the Cooperative Human Tissue Network. On the day of colon resection surgery, normal marginal colon mucosa was resected and placed into Dulbecco's Modified Eagle Medium (DMEM) at 4°C. Within hours, the tissue was prepared into normal human colonoids with Intesticult Organoid Growth Medium (Stem Cell Technologies) according to the manufacturer's protocol ([https://cdn.stemcell.com/media/files/pis/DX21423-PIS\\_1\\_0\\_0.pdf](https://cdn.stemcell.com/media/files/pis/DX21423-PIS_1_0_0.pdf)). Colonoids were suspended in Matrigel matrix (Corning) with standard growth factor concentration for maintenance and

passage. Matrigel with reduced growth factors was used for suspending colonoids in the final passage for the experiment. Eight hours before toxin exposure, organoids were serum starved with DMEM and no growth serum. Colonoids were exposed to TcdA (10 pmol/L), TcdB (10 pmol/L), or DMEM vehicle negative control for either 30 minutes or 6 hours. Colonoid-containing Matrigel was washed once with 1 $\times$  PBS and flash frozen in liquid nitrogen for protein analysis. For miRNA analysis, colonoids were removed from Matrigel with the use of Gentle Cell Dissociation Reagent (Stem Cell Technologies) and resuspended in RNAlater (Invitrogen). Colonoids in RNAlater were stored at 4°C overnight before freezing at -80°C.

### Serum miRNA Isolation and High-Content Analysis

Human serum was isolated by centrifugation (2200g for 10 minutes at room temperature) from whole blood and snap-frozen. RNA was isolated from human and animal serum samples (200  $\mu\text{L}$ ) with the use of the miRNeasy Serum/Plasma Kit (Qiagen) according to the manufacturer's instructions. Eluted RNA from serum samples was further purified and concentrated with the use of Amicon Ultra YM-3 columns (3000 kDa MWCO; Millipore).

For high-content miRNA analysis, RNAs after hybridization reactions were processed with the use of the nCounter Prep Station and nCounter Digital Analyzer. miRNA levels ( $n = 800$ ) were analyzed with the use of the nSolver software v3.0 (Nanostring Technologies, Seattle, WA, USA). Normalization was performed with the use of all miRNAs ( $n = 110$ ) with coefficients of variation <70%.<sup>14</sup>

Additional information is available in the [Supplementary Methods](#).

### Statistical Analysis

All statistical analyses were performed in SPSS v24 and R v3.5.1. Descriptive statistics for participant characteristics at baseline were reported using mean  $\pm$  SD and percentages. All data are expressed as mean and SEM. Systematic within-week changes for each miRNA were examined using a nonparametric longitudinal method (nparLD in R package) followed by Wilcoxon signed rank test for pairwise comparisons. The association between the metavariables and miRNAs was assessed by means of Spearman correlation. Heatmaps for capsule and colonoscopy combined and separately for capsule and colonoscopy delivered FMT were generated in the R package ComplexHeatmap2.<sup>15</sup> Normality was checked by means of the Kolmogorov-Smirnov test. Graph generation, fold changes, and

**Figure 1.** Fecal microbiota transplantation (FMT) in patients with *Clostridioides difficile* infection regulates the levels of circulating microRNAs (miRNAs). (A) Heatmap representation of the significantly up-regulated circulating miRNAs 4 and 12 weeks after FMT treatment compared with the screening time point ( $n = 42$ ), as assessed with the nCounter Nanostring platform. (B) Representative box plots depicting the miRNAs with highest levels of detection. Box plots denote mean % change  $\pm$  SEM, inner boxes represent mean, and error bars represent 95% confidence intervals (CIs). Statistical significance of FMT effect on circulating miRNAs was determined by nonparametric longitudinal method followed by Wilcoxon signed-rank test for pairwise comparisons. \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . (C) The overlapping top 3 miRNA-regulated pathways, as assessed with the Metacore network analysis software. (D) Validation of the top 6 up-regulated miRNAs in the replication cohort 2 and 4 weeks after FMT as assessed with reverse-transcription quantitative polymerase chain reaction (RT-qPCR). Fold changes and statistical significance in circulating levels of miRNAs were determined with OriginPro and Wilcoxon matched-pairs signed rank test. \*\*\* $P < 0.001$ .

statistical significance in levels of circulating miRNAs were assessed with the use of quantitative polymerase chain reaction (qPCR) by means of OriginPro and Wilcoxon matched-pairs signed rank test. Significance differences were considered when  $*P < 0.05$ ;  $**P < 0.01$ ;  $***P < 0.001$ .

## Results

### FMT Modifies Circulating miRNAs in rCDI

#### Patients

For miRNA profiling, we first analyzed serum samples derived from the discovery cohort.<sup>9</sup> Table 1 describes patient baseline characteristics. Sera obtained at screening and 4 and 12 weeks after FMT from 42 participants who achieved a clinical cure after FMT were subjected to miRNA analysis using the Nanostring nCounter platform. miRNA profiling of 126 samples revealed the significant up-regulation of 64 circulating miRNAs 4 and 12 weeks after FMT compared with screening (Figure 1A). The miRNAs with the highest levels detected are depicted in Figure 1B. Similar changes in miRNA levels were detected in recipients of either by-capsule or colonoscopic FMT (Supplementary Figure 1). Pathway enrichment analysis identified overlaps between the top 3 miRNA-regulated pathways, B-cell lymphoma 2, Runt-related transcription factor 2 (linked to nuclear factor  $\kappa$ B inhibitor  $\beta$ ), and phosphoinositide 3-kinase (regulatory class 1A p55- $\gamma$ ), linking inflammatory signaling to immune cell survival and differentiation (Figure 1C). Pathway analysis also uncovered commonalities with other diseases (eg, inflammatory bowel disease and multiple sclerosis) and cell functions such as apoptosis (Supplementary Figures 2 and 3).

We next sought to validate our discovery cohort results. For the replication cohort, we assessed 24 patients at 3 time points: screening and 2 and 4 weeks after FMT.<sup>10</sup> Table 1 describes patient baseline characteristics. The top 6 up-regulated miRNAs from the discovery cohort analysis (Figure 1B) were selected for reverse-transcription (RT) qPCR validation in our replication cohort. Individual miRNAs were found to be up-regulated in 78%–94% of samples analyzed, with the average changes ranging from 3- to 12-fold (Figure 1D).

### FMT Modifies Intestinal Tissue miRNAs in a Mouse Model of rCDI

The concerted increase of circulating miRNAs by FMT suggests that *C difficile*-associated dysbiosis may regulate miRNAs through a conserved mechanism. We tested this hypothesis with the use of a mouse model of rCDI.<sup>8</sup> Animals pretreated with cefoperazone received  $10^3$  *C difficile* spores, and at 4 dpi were exposed to vancomycin. At 11 dpi, a group of mice received fresh FMT derived from healthy mice. Sera collected 21 dpi were subjected to RT-qPCR. Parallel to what we found in rCDI patients, the same top 5 miRNAs (miR-4454 has not been characterized in mice) and in addition 2 potentially functional miRNAs were up-regulated after FMT. *C difficile* recurrence (Supplementary Figure 4) resulted in down-regulation of the tested circulating miRNAs in mice

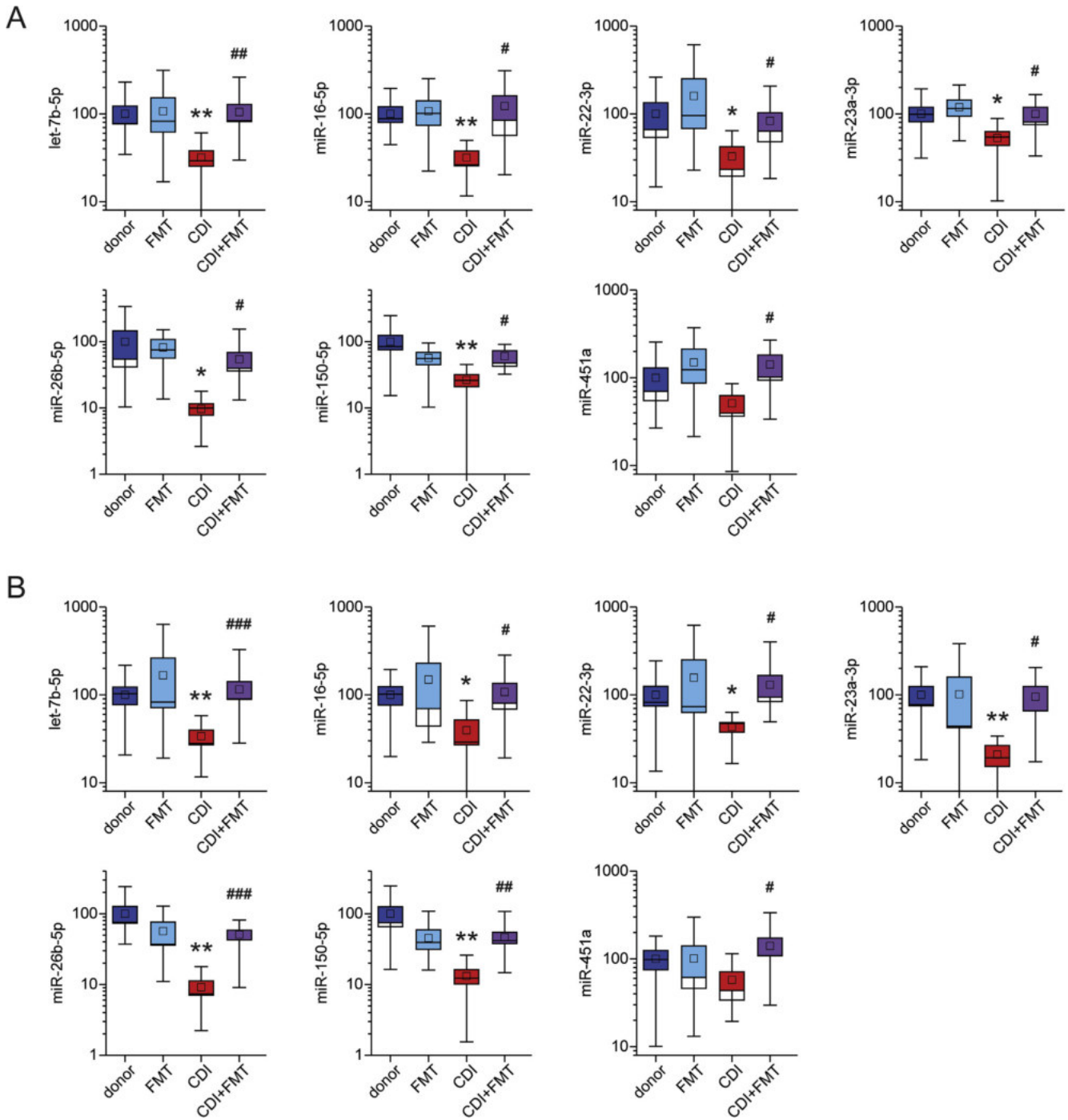
(Figure 2A) concomitant with a decrease in animal body weight (Supplementary Figure 5). To examine the time dependence of miRNA regulation by CDI, we assessed miRNA levels during the acute phase of infection (4 dpi). RT-qPCR analysis revealed that the suppression of circulating host miRNAs may be an early sign of CDI (Supplementary Figure 6). Importantly, the inhibitory effects of rCDI on miRNA levels was reversed 10 days after FMT (Figure 2A).

We next tested whether the circulating miRNA changes reflect the effects of rCDI on colonic tissue. RNA extracts from the ceca were analyzed by means of RT-qPCR. Our results showed that tissue miRNA levels altered similarly to circulating miRNAs (Figure 2B). Comparisons of the changes of individual miRNA levels in matched tissue and serum samples, from all animals, showed significant and positive correlation between circulating and ceca-expressed miRNAs (Supplementary Figure 7). This supports the notion that alterations in circulating miRNAs may be originating from colonic tissues. Furthermore, we found even stronger positive correlation between matched colonic and circulating miRNAs derived from an 84-year-old male patient with fulminant CDI treated with FMT by colonoscopy (Supplementary Figure 8). In mice, tissue miRNA levels down-regulated by rCDI reached statistical significance in the early phase of infection (Supplementary Figure 9) and FMT up-regulated the tissue miRNAs, coinciding with the reduction of *C difficile* load (Supplementary Figure 4) and the recovery of animal body weight (Supplementary Figure 5).

### Toxin B Suppresses miRNAs in the Intestinal Mucosa

*C difficile* pathogenicity is primarily mediated by exotoxins that induce cell death. We investigated the effects of purified TcdA and TcdB on host miRNA regulation. Mice were treated by intrarectal instillation with a combination of TcdB and TcdA, TcdB alone (15  $\mu$ g), or vehicle (HBSS), and colonic tissues and sera were collected 2–5 hours after instillation. RT-qPCR analysis of serum RNA extracts showed that the toxins had no effect on circulating miRNAs (Supplementary Figure 10). However, expression of the same miRNAs was suppressed in colonic tissues of *C difficile*-infected mice. Notably, TcdB alone suppressed miRNA levels, an effect enhanced by its combination with TcdA (Figure 3A). The discrepancy between circulating and tissue miRNAs in this model suggests that the impact of the toxins is lumenally confined and may not lead to changes in circulating miRNA levels owing to local and time-restricted exposure of animals to toxins at early time points (2–5 hours).

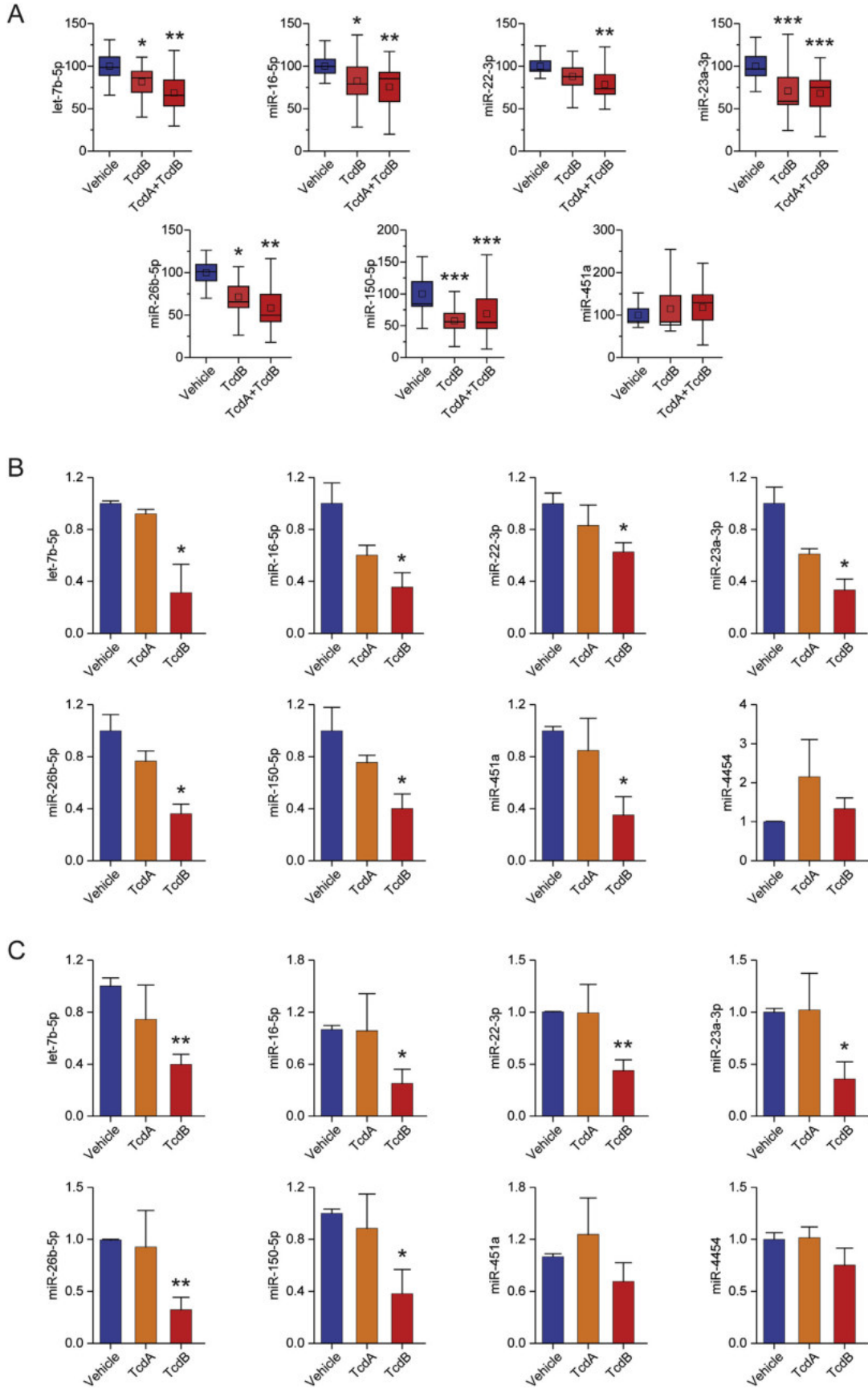
The above findings suggest the effect of CDI on miRNA regulation is mediated by the direct effect of TcdA and/or TcdB on colonic epithelial cells and the observed alterations in circulating miRNAs in FMT-treated patients may result from changes at the tissue level. To test this hypothesis, we exposed normal human colon organoids (colonoids) to TcdA or TcdB and analyzed miRNA expression. Colonoids from



**Figure 2.** FMT reverses the effects of *C difficile* on circulating and tissue miRNAs in a mouse model of recurrent *C difficile* infection (CDI). Box plots depicting the changes in miRNA levels in (A) sera and (B) ceca from animals treated with FMT, infected with *C difficile* (CDI), and infected with *C difficile* and treated with FMT (CDI+FMT) compared with FMT donors. Box plots denote mean % change ± SEM, inner boxes represent mean, and error bars represent 95% CIs. miRNA levels were assessed with RT-qPCR and normalized against (A) RNU1A1 and cel-miR-39 (spike-in) or (B) RNU5G and 5S rRNA, and compared with control (donor) samples set as 100%. Statistical significance was determined by Student *t* test. Compared with donor: \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001. Compared with CDI: #*P* < 0.05; ##*P* < 0.01; ###*P* < 0.001. Other abbreviations as in Figure 1.

freshly isolated human colon mucosa were treated with TcdA (10 pmol/L), TcdB (10 pmol/L), or vehicle (DMEM), and cell extracts were obtained 30 minutes and 6 hours later. RT-qPCR analysis showed that 30 minutes after

exposure to TcdB, all miRNAs studied were reduced (Figure 3B), an effect sustained for 6 hours (Figure 3C). These results show that miRNA suppression in CDI is attributed to the activity of TcdB and further support the



hypothesis that FMT-mediated up-regulation of circulating miRNAs is driven, at least in part, at the epithelial level.

### FMT Counteracts *C difficile* Effects on miRNA Biogenesis

The evidence collectively show that the miRNAs studied adhere to a common mode of regulation, although they are encoded by different genes and controlled by different mechanisms at the level of transcription. In fact, the miRNAs studied are located in different chromosomes, are intragenic (exonic or intronic) or intergenic, and their expression is under the control of distinct regulatory elements. Therefore, their regulation may be the result of a universal mechanism. Following transcription, miRNA maturation is a process shared by most miRNAs, and their concerted up-regulation by FMT suggests that miRNA processing may be affected by *C difficile*-associated dysbiosis. Therefore, we analyzed the effects of *C difficile* and FMT on the expression of enzymes playing a central role in miRNA biogenesis (DROSHA, DICER1, AGO2). RT-qPCR analysis of cecal RNA extracts from rCDI mice suggested that CDI suppresses *Drosha* expression by 50% (Figure 4A, left), with minor changes on *Dicer1* and *Ago2* mRNA levels (Supplementary Figure 11). A more pronounced effect (>80% decrease) on *Drosha* expression was evident at the protein level (Figure 4A, right). Importantly, the effects on both *Drosha* mRNA and protein were reversed by FMT (Figure 4A). Similarly, in colonic tissues of mice treated with exotoxins, *Drosha* mRNA levels dropped by 40% (Figure 4B). In human colonoids, TcdB showed small but significant effects on DROSHA mRNA (Figure 4C, left), while Western blot analysis revealed a robust decrease (>60%) in protein levels (Figure 4C, right).

To investigate the role of *drosha* suppression in mediating the CDI effects on the regulation of miRNA levels, we knocked down *drosha* in colonic epithelial cells (NCM356) by means of small interfering RNA. We verified that the knockdown of *drosha* mimics the effects of TcdB on *drosha* protein levels in these cells (Figure 4D). RNA extracts from these cells were subjected to RT-qPCR for the levels of the primary transcripts (pri-miRNAs) and the mature forms of 3 different miRNAs, known to be transcriptionally regulated by distinct mechanisms.<sup>16–18</sup> The results showed that on DROSHA inhibition, the levels of the pri-miRNAs (Figure 4E, top) are increased with the concurrent and significant decrease in mature miRNA levels (Supplementary Figure 12). We then measured the levels of these pri-miRNAs in colonoids treated with TcdB. In the same line,

we found that although the mature miRNAs are suppressed by TcdB (Figure 3), the levels of the respective pri-miRNAs are increased (Figure 4E, bottom).

Combined, these data suggest that *drosha* expression is decreased in response to CDI, a phenomenon regulated at both the transcriptional and post-transcriptional level. Furthermore, they attribute the concerted changes in miRNA levels to the dysregulation of miRNA biogenesis machinery (proposed model shown in Figure 4E) by rCDI and its recovery by FMT treatment.

### FMT-Regulated miRNAs Possess Functional Properties

We next investigated the downstream effects of FMT-regulated circulating miRNAs in our rCDI patient cohorts compared with healthy control subjects. We assessed miRNAs predicted to target specific chemokines and cytokines, which we previously found to be down-regulated by FMT.<sup>4</sup> Based on TargetScan prediction analyses we found that the levels of 6 miRNAs in FMT-treated rCDI patients inversely correlate with the serum levels of FGF21, IL12B, IL18, and TNFRSF9 proteins (Supplementary Figure 13). miR-23a, miR-26b and miR-130a are predicted to target the 3'UTR of FGF21 mRNA, miR-23a is predicted to target the 3'UTR of IL12B, and miR-150 the 3'UTR of IL18, whereas miR-20a and miR-28 are the 3'UTR of TNFRSF9. Overexpression of these miRNAs in colonic epithelial cells showed that miR-26b, miR-23a, miR-150, and miR-130a suppress the mRNA levels of FGF21, IL12B, IL18, and TNFRSF9, respectively (Figure 5A). These findings were further validated by 3'UTR reporter assays, where mutations in their target sequences within the 3'UTRs reversed their suppressive effects (Figure 5B), suggesting that the 4 miRNAs directly target these mRNAs.

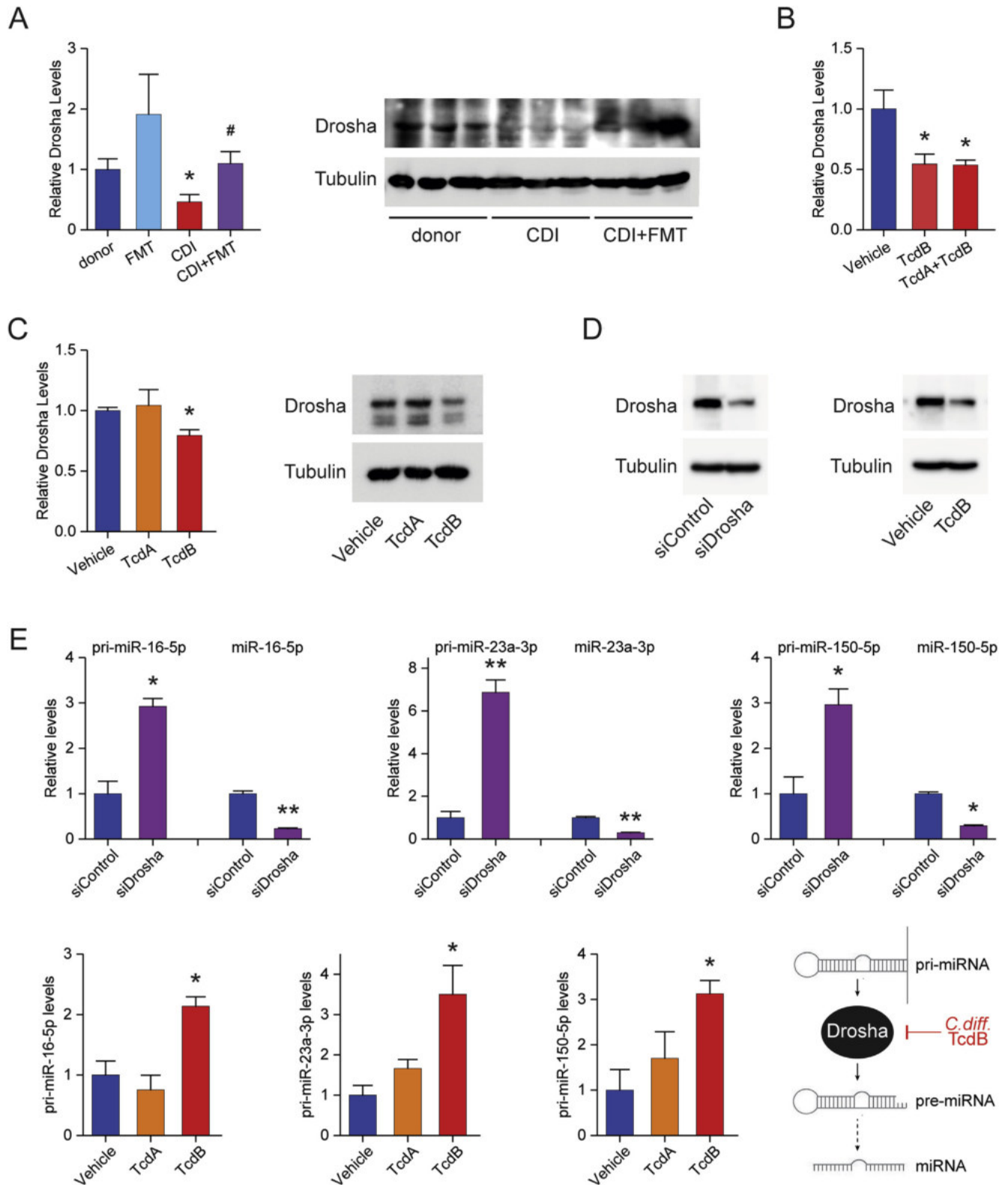
In addition to the changes observed for these miRNAs in the mouse model of rCDI human colonoids (Figure 2), in mouse colonic tissues treated with TcdB (Figure 3), and in FMT-treated rCDI patients (Figure 1A), we found that rCDI decreases the levels of these 4 miRNAs in sera compared with healthy control samples (Figure 5C). The use of miRNA inhibitors against miR-26b and miR-150, in peripheral blood mononuclear cells derived 4 weeks after FMT treatment, partly (FGF21) or completely (IL18) reversed the effect of FMT on their expression, suggesting a functional role for these 2 miRNAs in FMT therapy (Supplementary Figure 14). Pretreatment of colonic epithelial cells with these cytokines revealed that though FGF21, IL12B, and TNFRSF9 had no effects (Supplementary Figure 15), IL18 sensitizes cells to

**Figure 3.** Down-regulation of miRNAs in mouse colonic tissues and human colonoids treated with *C difficile* toxins (Tcds). (A) Box plots depicting the changes in miRNA levels in colonic tissues from animals treated with TcdB and combination of TcdA with TcdB, compared with HBSS-treated control animals (Vehicle). Box plots denote mean % change  $\pm$  SEM, inner boxes represent mean, and error bars represent 95% CIs. miRNA levels were assessed with RT-qPCR, normalized against RNU5G and 5S rRNA, and compared with control samples (Vehicle) set as 100%. Changes in miRNA levels after (B) 30 minutes and (C) 6 hours in colonoids treated with TcdA or TcdB, compared with DMEM-treated control samples (Vehicle). miRNA levels assessed with RT-qPCR were normalized against RNU1A1 and 5S rRNA and expressed as mean  $\pm$  SEM compared with control samples set as 1. Statistical significance was determined by Student *t* test. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001. HBSS, Hank's Balanced Salt Solution; other abbreviations as in Figure 1.

TcdB but not TcdA (Figure 5D), an effect validated in a second mucosal cell line (Figure 5E). The identified miRNA-target interactions provide novel links between inflammation and metabolism (Supplementary Figure 16).

*FMT-Regulated miRNAs Modulate Susceptibility to C difficile Toxin Effects*

We next examined whether miRNAs up-regulated by FMT can affect the colonic epithelial response to *C difficile*





toxins. First, we tested the effect of miRNA up-regulation, alone or in combination with TcdB, on cell survival. The results showed that individual miRNAs had minor effects on TcdB cytotoxicity. However, combination of miR-23a-3p and miR-150-5p significantly increased cell survival (Figure 6A). A live-cell analysis assay verified the above findings (Supplementary Figure 17) and suggested that miR-23a-3p and miR-150-5p alone (Supplementary Figure 18) or in combination confer a growth advantage to cells treated with TcdB (Figure 6B). In accordance, the viability of a second mucosal cell line exposed to TcdB was significantly increased by the combined overexpression of miR-23a-3p and miR-150-5p (Figure 6C).

Although miR-23a-3p and miR-150-5p had minor effect on the early TcdB-mediated cytotoxicity, they promoted cell survival (Supplementary Figure 19). Using a real-time caspase-3/7 detection assay, we found that TcdB induces cell apoptosis in a dose-dependent manner (Supplementary Figure 20); miR-23a-3p and miR-150-5p had additive effects in reducing susceptibility to TcdB-induced apoptosis over time (Figure 6D). A major cytopathic effect attributed to TcdB is the induction of cell morphologic and cytoskeleton changes, which can be visualized by staining colonic epithelial cells with fluorescence-conjugated phalloidin. Using this method, we observed major morphologic changes to epithelial cells, such as loss of actin stress fibers, spindle-like formation, and cell rounding, when exposed to both TcdA and TcdB. Quantification of the ratio of rounded and spindle-like cells revealed that TcdB-induced cytopathic effects were significantly counteracted by the overexpression of miR-23a-3p and miR-150-5p (Figure 6E and Supplementary Figure 21).

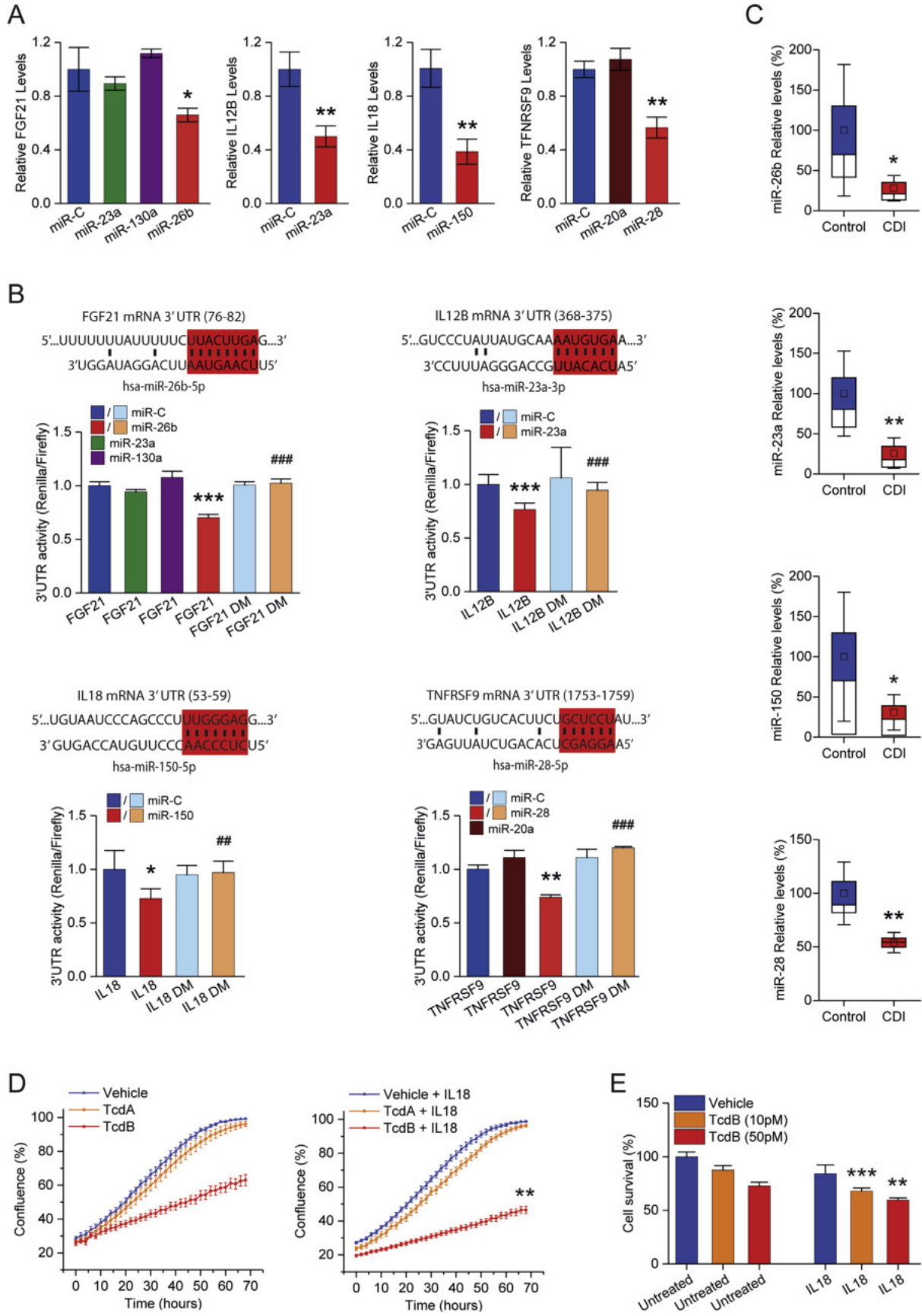
## Discussion

This is the first study to examine the effects of FMT for rCDI on miRNA signatures. Our findings in a clinical rCDI cohort demonstrate the concerted regulation of miRNA expression by FMT. Validating these observations in a mouse model of rCDI and organoids, we provide evidence that miRNA processing in colonic epithelial cells is directly altered by *C difficile* toxins and may be affected by *C difficile*-associated dysbiosis. Conditional knockout of the miRNA-processing enzyme *Dicer* in murine intestinal

epithelial cells, which secrete fecal miRNAs, has been shown to modulate the gut microbiota and exacerbate colitis. This phenotype can be rescued via wild-type fecal transplantation.<sup>8</sup> From this observation, we hypothesized that colitis due to CDI can suppress circulating miRNAs, which can be restored by FMT. In support of our hypothesis, predominant suppression of miRNAs was observed in 2 independent cohorts of rCDI patients but was differentiated from observed miRNA suppression in other colitis patient cohorts, such as in patients with chronic inflammatory bowel diseases (ulcerative colitis<sup>14</sup> and Crohn's disease<sup>19</sup>) and associated colorectal cancer.<sup>20</sup> For example, miR-30 family members are up-regulated in both ulcerative colitis and Crohn's disease, but suppressed in rCDI, although miR-23a-3p is up-regulated only in ulcerative colitis and down-regulated in rCDI. These differences may be leveraged for future diagnostic purposes and could help identify patients with inflammatory bowel disease who can benefit from biotherapeutic products such as FMT, which currently demonstrate variable efficacy compared with patients with CDI.

The miRNA signature in the mouse models suggest that changes in circulating miRNAs reflect alterations in the intestinal epithelial cells. Importantly, the effects of *C difficile* infection on miRNA levels are reversed in both sera and tissues after FMT, paralleling the post-FMT serum miRNA changes observed in patients with CDI after FMT. TcdA and TcdB have been causally linked to the pathogenetic mechanisms of CDI.<sup>21</sup> At the cellular level, toxins induce cytoskeleton reorganization and tight junction disruption resulting in cell rounding and cell death.<sup>22</sup> Here, we show that TcdB suppresses key inflammation-related miRNAs in murine intestinal tissues and human colonoids. Moreover, we identify 2 miRNAs up-regulated by FMT, miR-23a and miR-150, which exert cytoprotective effects against TcdB. Interestingly, we found that IL18, previously shown to contribute to mucosal damage,<sup>23</sup> sensitizes colonic epithelial cells to TcdB. In addition, we show that IL18 is a direct target of miR-150, suggesting a new mechanism by which FMT may counteract *C difficile*-induced epithelial disruption. Additional miRNA-regulated cytokines may be involved in the regulation of anti-inflammatory effects of FMT. We found that

**Figure 4.** *C difficile* infection suppresses DROSHA expression and miRNA processing. (A) Drosha mRNA and protein levels in ceca from animals treated with FMT, infected with *C difficile* (CDI), infected with *C difficile* and treated with FMT (CDI+FMT), compared with FMT donors. (B) Drosha mRNA levels in colonic tissues from animals treated with TcdB or combination of TcdA with TcdB, compared with control samples (Vehicle). (C) Drosha mRNA and protein levels in colonoids treated with TcdA or TcdB, compared with DMEM-treated control subjects (Vehicle). (D) DROSHA protein levels in NCM356 colonic epithelial cells on knockdown of DROSHA by means of short interfering (si) RNA (left) and treated with TcdB (right), compared with nontargeting siRNA (siControl) and DMEM-treated control samples (Vehicle), respectively. (E) Primary (pri) miRNA transcript levels in NCM356 cells on knockdown of DROSHA (top) and in colonoids treated with TcdA or TcdB (bottom), compared with nontargeting siRNA (siControl) and DMEM-treated control samples (Vehicle), respectively. mRNA and pri-miRNA levels assessed with RT-qPCR were normalized against  $\beta$ -Actin and GAPDH and expressed as mean  $\pm$  SEM compared with control samples set as 1. DROSHA protein levels were assessed with immunoblot analysis. alpha-Tubulin was used as the loading control. Bottom right: schematic representation of the proposed model of miRNA regulation by CDI. Statistical significance was determined by Student *t* test. \**P* < 0.05; \*\**P* < 0.01; compared with CDI: #*P* < 0.01. DMEM, Dulbecco's Modified Eagle Medium; other abbreviations as in Figure 1.



miR-23a targets IL12B, an essential activator of T<sub>H</sub>1-cell development, associated with CDI recurrence.<sup>24</sup> Collectively, these data support a role for miRNAs suppressed by toxins as a new pathogenic mechanism for CDI. We propose that the restoration of these miRNAs by FMT contributes to the protection of epithelial barrier integrity. Changes in circulating miRNAs may also contribute to extracolonic manifestations of CDI including cardiac, renal, and neurologic impairment.<sup>22</sup>

Our analyses of the miRNA biogenesis machinery illustrate that down-regulation of miRNAs is likely through the suppression of DROSHA by CDI/TcdB and is restored following FMT. Inhibition of DROSHA results in defective miRNA processing with accumulation of pri-miRNAs. Our findings suggest the biphasic regulation of the microprocessor by CDI. The temporal effect of TcdB suggests that early miRNA regulation is attributed to a nontranscriptional mechanism. In fact, DROSHA protein is reduced in response to TcdB before its mRNA levels are suppressed. Different types of stress have been associated with the stability of DROSHA protein. Under oxidative stress, DROSHA is phosphorylated by p38 MAPK at the N-terminus. This results in disruption of its interaction with DGSCR8, relocation to the cytoplasm, and protein degradation.<sup>25</sup> Indeed, the cytotoxic effect of TcdB has been shown to depend on assembly of the host epithelial cell NADPH oxidase complex and the production of reactive oxygen species.<sup>26</sup> Other metabolic inputs may be involved. The mammalian Target of Rapamycin nutrient/amino acid sensor activates MDM2, which catalyzes DROSHA ubiquitination, marking it for degradation.<sup>27</sup> The long-term suppression of both DROSHA mRNA and protein may involve a gene transcription regulatory mechanism. The transcription factor c-MYC activates *DROSHA* gene and up-regulates DROSHA protein,<sup>28</sup> and is under the control of the WNT/ $\beta$ -catenin pathway, which is suppressed by both toxins.<sup>29,30</sup>

The regulation of miRNA levels by *C difficile* may not rely solely on the effects of toxins on mucosal cells. miRNA suppression, mediated by surface layer proteins of specific *C difficile* ribotypes, has been proposed to attenuate the host's immune response, resulting in a

more persistent infection in mice.<sup>31</sup> Interestingly, the suppression of circulating miRNAs on depletion of regulatory T cells (Tregs), reported in mouse models of autoimmune diseases,<sup>32</sup> suggests potential links between Treg function, CDI, miRNAs, and FMT. Recent experimental evidence has linked the effectiveness of FMT for colitis with the induction of IL10 and Transforming Growth Factor  $\beta$ , cytokines critical for Treg accumulation in the intestine.<sup>33</sup> Whether FMT-directed immunosuppression aids in the recovery from *C difficile* requires further investigation.<sup>3</sup>

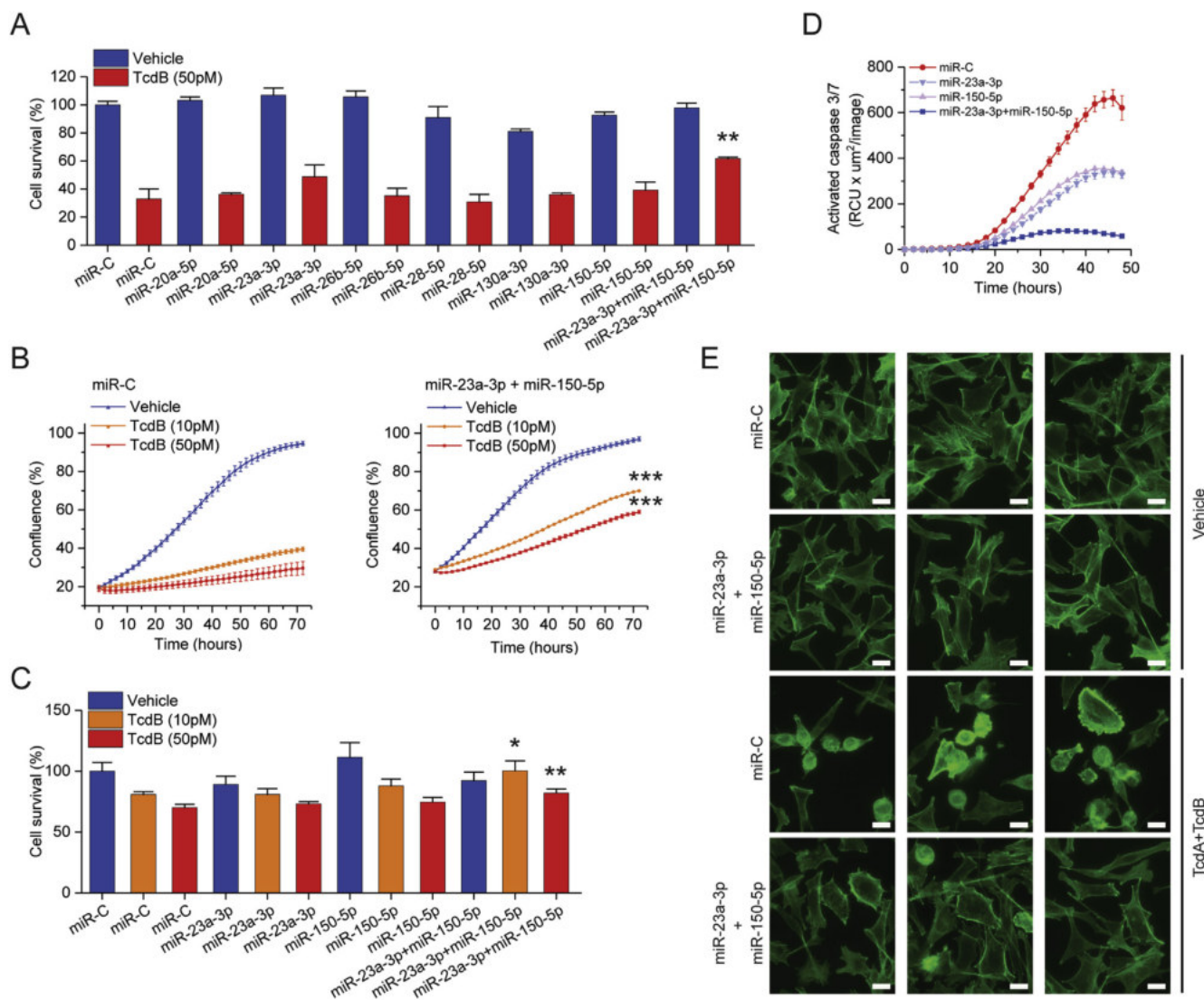
Intriguingly, in *C difficile*-naïve animals treated with FMT we observed a trend toward higher Drosha and Dicer mRNA levels and alterations in the miRNA profiles. These control animals were conditioned with cefoperazone (an antibiotic against both gram-positive cocci and gram-negative bacteria) to facilitate *C difficile* colonization. This observation proposes that host-microbe interactions regulate the host miRNA biogenesis machinery, which in turn may affect directly or indirectly, through the immune response, gut microbiota composition.

## Conclusion

We here reported changes in circulating and colonic miRNAs in the context of FMT for rCDI, validating our observations across 2 independent randomized trials. Using 2 different animal models and human colonoids and colonic epithelial cells, we substantiated that *C difficile* highjacks miRNA biogenesis and showed how miRNA restoration contributes to the therapeutic effects of FMT. Our findings strongly support the need for further mapping of the epitranscriptomic changes associated with FMT. While our data in rCDI mice suggest that TcdB affects Drosha rather than Dicer1 or Ago2 expression, studies with conditional knockout mice of the miRNA-processing machinery, such as *Dicer* or *Drosha*, may reveal additional molecular mechanisms or metabolic pathways affected by CDI-induced colitis.

Together, these results provide novel and provocative evidence that modulation of the gut microbiome via FMT

**Figure 5.** Effects of FMT-regulated miRNAs on circulating proteins in patients with CDI. (A) Effects of miR-26b-5p, miR-23a-3p, miR-150-5p, and miR-28-5p overexpression on the levels of FGF21, IL12B, IL18, and TNFRSF9 mRNAs, respectively, in colonic epithelial cells. Gene expression data normalized against  $\beta$ -Actin and GAPDH are expressed as mean  $\pm$  SEM compared with miR-C-transfected cells set as 1. (B) Effects of miR-26b-5p, miR-23a-3p, miR-150-5p, and miR-28-5p on the activity of FGF21, IL12B, IL18, and TNFRSF9 mRNA 3'UTRs, as assessed with luciferase reporter assays. 3'UTR sequences were cloned in a reporter vector downstream of the *Renilla Luciferase* gene. The reporter vector was transfected in colonic epithelial cells and luciferase activity was measured 24 hours after the overexpression of the respective miRNAs in the same cells. Direct targeting of the 3'UTR by the miRNA was validated with assays using deletion mutants (DMs) of the respective miRNA target sequences. *Renilla Luciferase* activity was normalized against the activity of the *Firefly Luciferase* gene expressed by the same vector. miR-C (cel-miR-39-3p), a nontargeting miRNA, was used as negative control. (C) CDI effect on the levels of circulating miR-26b-5p, miR-23a-3p, miR-150-5p, and miR-28-5p in patients. The levels of serum miRNAs in *C difficile* patients compared with healthy control subjects (n = 42), as assessed with the nCounter Nanostring platform. Box plots denote mean % change  $\pm$  SEM, inner boxes represent mean, and error bars represent 95% CIs. (D) Effects of IL18 on toxin-mediated cell growth inhibition. NCM356 cell growth was monitored in real time as % of confluence (IncuCyte). (E) Effects of *il18* on TcdB-mediated cell growth inhibition. NCM460 cell survival was assessed by measuring metabolically active cells (Cell-Titer Glo) and expressed as mean  $\pm$  SEM compared with untreated cells set as 100%. Statistical significance was determined by Student *t* test. Compared with miR-C (A and B), healthy control samples (C), and TcdB alone (D and E): \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001. Compared with wild-type 3'UTRs: ##*P* < 0.01; ###*P* < 0.001. UTR, untranslated region; other abbreviations as in Figures 1 and 2.



**Figure 6.** Functional effects of FMT-regulated miRNAs on colonic epithelial cells. (A) Effects of FMT-regulated miRNA overexpression on TcdB-mediated cell growth inhibition. NCM356 cell survival was assessed by measuring metabolically active cells (Cell-Titer Glo) and expressed as mean  $\pm$  SEM compared with miR-C transfected cells set as 100%. (B) Effects of miR-23a-3p and miR-150-5p overexpression on TcdB-mediated NCM356 cell growth inhibition. Cells were monitored in real time as % of confluence (IncuCyte). (C) Effects of miR-23a-3p and miR-150-5p overexpression on TcdB-mediated NCM460 cell growth inhibition. Cell survival was assessed by measuring metabolically active cells (Cell-Titer Glo) and expressed as mean  $\pm$  SEM compared with miR-C transfected cells set as 100%. (D) Effects of miR-23a-3p and miR-150-5p overexpression on TcdB-induced cell apoptosis. NCM356 cell apoptosis was monitored in real time as activated caspase3/7 fluorescence (IncuCyte). (E) Effects of miR-23a-3p and miR-150-5p overexpression on TcdB-mediated cytoskeleton rearrangements. NCM356 cytoskeleton organization was studied by fluorescence microscopy after phalloidin staining. miR-C (cel-miR-39-3p), a nontargeting miRNA, was used as negative control. Statistical significance was determined by Student *t* test. \**P* < 0.05; \*\**P* < 0.01. Abbreviations as in Figures 1 and 2

induces changes in circulating miRNAs, and that a subset of these miRNAs down-regulates inflammatory protein expression and protects epithelial cells. These findings add new insight into the molecular mechanisms underlying *C difficile* pathogenesis and FMT and identify new potential targets for therapeutic intervention.

## Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at [www.gastrojournal.org](http://www.gastrojournal.org), and at <https://doi.org/10.1053/j.gastro.2021.03.050>.

## References

1. d'Haens GR, Jobin C. Fecal microbial transplantation for diseases beyond recurrent *Clostridium difficile* infection. *Gastroenterology* 2019;157(3):624–636.
2. Khoruts A, Sadowsky MJ. Understanding the mechanisms of action of faecal microbiota transplantation. *Nat Rev Gastroenterol Hepatol* 2016;13(9):508–516.
3. Petri WA Jr, Frisbee AL. Considering the immune system during fecal microbiota transplantation for *Clostridioides difficile* infection. *Gastroenterology* 2020;26(5):496–506.
4. Monaghan T, Mullish BH, Patterson J, et al. Effective fecal microbiota transplantation for recurrent *Clostridioides difficile* infection in humans is associated with increased signalling in the bile acid–farnesoid X receptor–fibroblast growth factor pathway. *Gut Microbes* 2019;10(2):142–148.
5. Qin Y, Wade PA. Crosstalk between the microbiome and epigenome: messages from bugs. *J Biochem* 2018;163(2):105–112.
6. Gebert LFR, MacRae IJ. Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol* 2019;20:21–37.
7. Aguilar C, Mano M, Eulalio A. MicroRNAs at the host–bacteria interface: host defense or bacterial offense. *Trends Microbiol* 2019;27(3):206–218.
8. Liu S, da Cunha AP, Rezende RM, et al. The host shapes the gut microbiota via fecal microRNA. *Cell Host Microbe* 2016;19(1):32–43.
9. Kao D, Roach B, Silva M, et al. Effect of oral capsule–vs colonoscopy–delivered fecal microbiota transplantation on recurrent *Clostridium difficile* infection: a randomised clinical trial. *JAMA* 2017;318(20):1985–1993.
10. Lee C, Steiner T, Petrof EO, et al. Frozen vs fresh fecal microbiota transplantation and clinical resolution of diarrhea in patients with recurrent *Clostridium difficile* infection: a randomized clinical trial. *JAMA* 2016;315(2):142–149.
11. Seekatz AM, Theriot CM, Molloy CT, et al. Fecal microbiota transplantation eliminates *Clostridium difficile* in a murine model of relapsing disease. *Infect Immun* 2015;83(10):3838–3846.
12. Theriot CM, Koumpouras CC, Carlson PE, et al. Cefoperazone-treated mice as an experimental platform to assess differential virulence of *Clostridium difficile* strains. *Gut Microbes* 2011;2(6):326–334.
13. Hirota SA, Iablokov V, Tulk SE, et al. Intrarectal instillation of *Clostridium difficile* toxin A triggers colonic inflammation and tissue damage: development of a novel and efficient mouse model of *Clostridium difficile* toxin exposure. *Infect Immun* 2012;80(12):4474–4484.
14. Polyarchou C, Oikonomopoulos A, Mahurkar S, et al. Assessment of circulating microRNAs for the diagnosis and disease activity evaluation in patients with ulcerative colitis by using the nanostring technology. *Inflamm Bowel Dis* 2015;21(11):2533–9253.
15. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;32(18):2847–2849.
16. Ofir M, Hacohen D, Ginsberg D. miR-15 and miR-16 are direct transcriptional targets of E2F1 that limit E2F-induced proliferation by targeting cyclin E. *Mol Cancer Res* 2011;9(4):440–447.
17. Hassan MQ, Gordon JA, Beloti MM, et al. A network connecting Runx2, SATB2, and the miR-23a~27a~24-2 cluster regulates the osteoblast differentiation program. *Proc Natl Acad Sci U S A* 2010;107(46):19879–19884.
18. Hu T, Chong Y, Cai B, et al. DNA methyltransferase 1–mediated CpG methylation of the miR-150-5p promoter contributes to fibroblast growth factor receptor 1–driven leukemogenesis. *J Biol Chem* 2019;294(48):18122–18130.
19. Oikonomopoulos A, Polyarchou C, Joshi S, et al. Identification of circulating microRNA signatures in Crohn's disease using the nanostring ncounter technology. *Inflamm Bowel Dis* 2016;22(9):2063–2069.
20. Chen G, Feng Y, Li X, et al. Post-transcriptional gene regulation in colitis associated cancer. *Front Genet* 2019;10:585.
21. Chandrasekaran R, Lacy DB. The role of toxins in *Clostridium difficile* infection. *FEMS Microbiol Rev* 2017;41(6):723–750.
22. Di Bella S, Ascenzi P, Siarakas S, et al. *Clostridium difficile* toxins A and B: insights into pathogenic properties and extraintestinal effects. *Toxins (Basel)* 2016;8(5):134.
23. Nowarski R, Jackson R, Gagliani N, et al. Epithelial IL-18 equilibrium controls barrier function in colitis. *Cell* 2015;163:1444–1456.
24. Yacyshyn MB, Reddy TN, Plageman LR, et al. *Clostridium difficile* recurrence is characterized by pro-inflammatory peripheral blood mononuclear cell (PBMC) phenotype. *J Med Microbiol* 2014 Oct;63(Pt 10):1260–1273.
25. Yang Q, Li W, She H, et al. Stress induces p38 MAPK–mediated phosphorylation and inhibition of drosha-dependent cell survival. *Mol Cell* 2015;57(4):721–734.
26. Ma Farrow, Chumblor NM, Lapierre LA, et al. *Clostridium difficile* toxin B–induced necrosis is mediated by the host epithelial cell NADPH oxidase complex. *Proc Natl Acad Sci U S A* 2013;110(46):18674–18679.
27. Ye P, Liu Y, Chen C, et al. An mTORC1–Mdm2–drosha axis for miRNA biogenesis in response to glucose- and amino acid-deprivation. *Mol Cell* 2015;57(4):708–720.
28. Wang X, Zhao X, Gao P, et al. c-Myc modulates microRNA processing via the transcriptional regulation of drosha. *Sci Rep* 2013;3:1942.
29. Bezerra Lima B, Faria Fonseca B, da Graça Amado N, Moreira Lima D, Albuquerque Ribeiro R, Garcia Abreu J, de Castro Brito GA. *Clostridium difficile* toxin A attenuates Wnt/β-catenin signaling in intestinal epithelial cells. *Infect Immun* 2014;82(7):2680–2687.
30. Chen P, Tao L, Wang T, et al. Structural basis for recognition of frizzled proteins by *Clostridium difficile* toxin B. *Science* 2018;360(6389):664–669.
31. Kennedy KF. The effects of surface layer proteins isolated from *Clostridium difficile* on TLR4 signalling. PhD thesis. Dublin City University, March 2016.
32. Jin F, Hu H, Xu M, et al. Serum microRNA profiles serve as novel biomarkers for autoimmune diseases. *Front Immunol* 2018;9:2381.

33. Burrell C, Garavaglia F, Cribru FM, Ercoli G, et al. Therapeutic faecal microbiota transplantation controls intestinal inflammation through IL10 secretion by immune cells. *Nat Comm* 2018;9:5184.

Received July 18, 2020. Accepted March 23, 2021.

#### Correspondence

Address correspondence to: Tanya M. Monaghan, BSc, BM, PhD, MRCP, NIHR Nottingham Digestive Diseases Biomedical Research Centre, W/E 1381, E Floor, West Block, Queen's Medical Centre Campus, Derby Road, Nottingham NG7 2UH, United Kingdom. e-mail: [tanya.monaghan@nottingham.ac.uk](mailto:tanya.monaghan@nottingham.ac.uk); Christos Polytaichou, PhD, John van Geest Cancer Research Centre, School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, G11 8NS, United Kingdom. e-mail: [christos.polytaichou@ntu.ac.uk](mailto:christos.polytaichou@ntu.ac.uk); or Dina Kao, MD, Division of Gastroenterology, Department of Medicine, University of Alberta, 8540 112th Street, Edmonton, Alberta T6G 2X8, Canada. e-mail: [dkao@ualberta.ca](mailto:dkao@ualberta.ca).

#### Acknowledgments

The authors are grateful to the participants that have made this research possible. The authors are grateful to Matt Emberg, Melanie Lingaya, and Yirga Falcone for their technical assistance in sample preparation. Authors Anna M. Seekatz, Nicholas O. Markham, Tung On Yau, and Maria Hatzia Apostolou contributed equally.

#### CRedit Authorship Contributions

Tanya M. Monaghan, BSc, BM, PhD, MRCP (Conceptualization: Lead; Data curation: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Methodology: Equal; Project administration: Lead; Resources: Equal; Software: Supporting; Supervision: Equal; Validation: Equal; Writing – original draft: Equal; Writing – review & editing: Lead); Anna M. Seekatz, PhD (Data curation: Equal; Methodology: Supporting; Resources: Supporting; Validation: Supporting; Writing – review & editing: Supporting); Nicholas O. Markham, MD, PhD (Data curation: Supporting; Methodology: Supporting; Validation: Supporting; Writing – review & editing:

Supporting); Tung On Yau, BSc, MSc (Data curation: Supporting; Project administration: Supporting); Maria Hatzia Apostolou, BSc, MSc, PhD (Data curation: Supporting; Methodology: Supporting; Resources: Supporting); Tahseen Jilani, PhD (Formal analysis: Supporting); Niki Christodoulou, BSc, MSc (Data curation: Supporting); Brandi Roach, RN (Data curation: Supporting; Project administration: Supporting); Eleni Biri, BSc, MSc (Data curation: Supporting); Odette Pomenya, BSc, MSc (Data curation: Supporting); Thomas Louie, MD (Investigation: Supporting; Writing – review & editing: Supporting); D. Borden Lacy, PhD (Data curation: Supporting; Investigation: Supporting; Resources: Supporting; Writing – review & editing: Supporting); Peter Kim, PhD (Investigation: Supporting; Writing – review & editing: Supporting); Christine Lee, MD (Investigation: Supporting; Validation: Supporting; Writing – review & editing: Supporting); Dina Kao, MD (Funding acquisition: Supporting; Investigation: Supporting; Project administration: Equal; Supervision: Equal; Writing – review & editing: Equal); Christos Polytaichou, PhD (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Methodology: Equal; Project administration: Supporting; Resources: Supporting; Software: Supporting; Supervision: Equal; Validation: Supporting; Visualization: Equal; Writing – review & editing: Supporting).

#### Conflicts of interest

Tanya M. Monaghan is a consultant advisor for Takeda. The authors declare no conflicts.

#### Funding

This work was supported by a University of Nottingham Research Priority Area grant to Tanya M. Monaghan and supplemented by the National Institute for Health Research Nottingham Digestive Diseases Biomedical Research Centre, based at Nottingham University Hospitals NHS Trust and University of Nottingham, Litwin Initiative at the Crohn's and Colitis Foundation to Christos Polytaichou, Nottingham Trent University Quality Research funds to Christos Polytaichou and Maria Hatzia Apostolou, Micheal Smith Health Research Foundation to Christine Lee, National Institute for Diabetes and Digestive and Kidney Diseases grant K01-DK111794 to Anna M. Seekatz, and research funding provided by Alberta Health Services and University of Alberta Hospital Foundation to Dina Kao. The funders were not involved in study design, writing the report or decision for publication.

# Integrative Identification of Epstein-Barr Virus-Associated Mutations and Epigenetic Alterations in Gastric Cancer

Qiaoyi Liang\*, Xiaotian Yao\*, Senwei Tang, Jingwan Zhang, **Tung On Yau**, Xiaoxing Li, Ceen-Ming Tang, Wei Kang, Raymond Wm M. Lung, Jing Woei Li, Ting Fung Chan, Rui Xing, Youyong Lu, Kwok Wai Lo, Nathalie Wong, Ka Fai To, Chang Yu, Francis Kl L. Chan, Joseph J Y Sung, Jun Yu

*Gastroenterology*. 2014;147(6):1350-62.e4.

Journal URL: [gastrojournal.org/article/S0016-5085\(14\)01075-0/](http://gastrojournal.org/article/S0016-5085(14)01075-0/)

ScienceDirect:

[sciencedirect.com/science/article/pii/S0016508514010750](http://sciencedirect.com/science/article/pii/S0016508514010750)

DOI: [10.1053/j.gastro.2014.08.036](https://doi.org/10.1053/j.gastro.2014.08.036)

PMID: [25173755](https://pubmed.ncbi.nlm.nih.gov/25173755/)

PMCID: [PMC4047330](https://pubmed.ncbi.nlm.nih.gov/pmc/articles/PMC4047330/)



3<sup>rd</sup> November, 2015

To Whom It May Concern,

**Statement of Joint Authorship**

**Title of publication:** Integrative Identification of Epstein–Barr Virus–Associated Mutations and Epigenetic Alterations in Gastric Cancer

**Journal:** Gastroenterology

**Publication date:** 31<sup>st</sup> December 2014

**Issue:** Volume 147, Issue 6, Pages 1350 – 1362

**PubMed ID:** 24876747

**Authors:** Qiaoyi Liang, Xiaotian Yao, Senwei Tang, Jingwan Zhang, Tung On Yau, Xiaoxing Li, Ceen-Ming Tang, Wei Kang, Raymond WM Lung, Jing Woei Li, Ting Fung Chan, Rui Xing, Youyong Lu, Kwok Wai Lo, Nathalie Wong, Ka Fai To, Chang Yu, Francis KL Chan, Joseph JY Sung, Jun Yu

I hereby confirm that Mr. Tung On YAU is a co-author in the above publication. He was a significant contributor to this project, assisting with experimental design, and data collection, as well as carrying out experiments. This included analysis of mutations and methylation states through a combination of acquisition of human gastric cancer samples, gastric cancer cell cultures, DNA and RNA extraction, DNA methylation bisulfite conversion, semi-quantitative and quantitative Real-Time PCR, as well as conventional Sanger sequencing. He also contributed to the revision of the final manuscript.

Yours sincerely,

Qiaoyi Liang

Research Assistant Professor

Institute of Digestive Disease

Prince of Wales Hospital

The Chinese University of Hong Kong, Hong Kong

Tel: 852-3763 6103; Fax: 852-2144 5330

Email: jessieQY@cuhk.edu.hk

Jun Yu, MD, PhD

Professor

Institute of Digestive Disease

Prince of Wales Hospital

The Chinese University of Hong Kong, Hong Kong

Tel: 852-3763 6099; Fax: 852-2144 5330

Email: junyu@cuhk.edu.hk



# BASIC AND TRANSLATIONAL—ALIMENTARY TRACT

## Integrative Identification of Epstein–Barr Virus–Associated Mutations and Epigenetic Alterations in Gastric Cancer



Qiaoyi Liang,<sup>1,2,\*</sup> Xiaotian Yao,<sup>3,\*</sup> Senwei Tang,<sup>3</sup> Jingwan Zhang,<sup>1,2</sup> Tung On Yau,<sup>1,2</sup> Xiaoxing Li,<sup>1,2</sup> Ceen-Ming Tang,<sup>4</sup> Wei Kang,<sup>5</sup> Raymond W. M. Lung,<sup>5</sup> Jing Woei Li,<sup>6</sup> Ting Fung Chan,<sup>6</sup> Rui Xing,<sup>7</sup> Youyong Lu,<sup>7</sup> Kwok Wai Lo,<sup>5</sup> Nathalie Wong,<sup>5</sup> Ka Fai To,<sup>5</sup> Chang Yu,<sup>3</sup> Francis K. L. Chan,<sup>1,2</sup> Joseph J. Y. Sung,<sup>1,2</sup> and Jun Yu<sup>1,2</sup>

<sup>1</sup>Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, <sup>2</sup>Department of Anatomical and Cellular Pathology, <sup>3</sup>School of Life Sciences, Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong, Hong Kong, China; <sup>4</sup>The Chinese University of Hong Kong Shenzhen Research Institute, Shenzhen, China; <sup>5</sup>BGI-Shenzhen, Shenzhen, China; <sup>6</sup>Department of Pharmacology, University of Oxford, Oxford, United Kingdom; <sup>7</sup>Laboratory of Molecular Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital/Institute, Beijing, China

**BACKGROUND & AIMS:** The mechanisms by which Epstein–Barr virus (EBV) contributes to the development of gastric cancer are unclear. We investigated EBV-associated genomic and epigenomic variations in gastric cancer cells and tumors. **METHODS:** We performed whole-genome, transcriptome, and epigenome sequence analyses of a gastric adenocarcinoma cell line (AGS cells), before and after EBV infection. We then looked for alterations in gastric tumor samples, with (n = 34) or without (n = 100) EBV infection, collected from patients at the Prince of Wales Hospital, Chinese University of Hong Kong (from 1998 through 2004), or the First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China (from 1999 through 2006). **RESULTS:** Transcriptome analysis showed that infected cells expressed 9 EBV genes previously detected in EBV-associated gastric tumors and 71 EBV genes not previously reported in gastric tumors. Ten viral genes that had not been reported previously in gastric cancer but were expressed most highly in EBV-infected cells also were expressed in primary EBV-positive gastric tumors. Whole-genome sequence analysis identified 45 EBV-associated nonsynonymous mutations. These mutations, in genes such as *AKT2*, *CCNA1*, *MAP3K4*, and *TGFBR1*, were associated significantly with EBV-positive gastric tumors, compared with EBV-negative tumors. An activating mutation in *AKT2* was associated with reduced survival times of patients with EBV-positive gastric cancer ( $P = .006$ ); this mutation was found to dysregulate mitogen-activated protein kinase signaling. Integrated epigenome and transcriptome analyses identified 216 genes transcriptionally down-regulated by EBV-associated hypermethylation; methylation of *ACSS1*, *FAM3B*, *IHH*, and *TRABD* increased significantly in EBV-positive tumors. Overexpression of Indian hedgehog (*IHH*) and TraB domain containing (*TRABD*) increased proliferation and colony formation of gastric cancer cells, whereas knockdown of these genes reduced these activities. We found 5 signaling pathways (axon guidance, focal adhesion formation, interactions among cytokines and receptors, mitogen-activated protein kinase signaling, and actin cytoskeleton regulation) to be affected commonly by EBV-associated genomic and epigenomic alterations. **CONCLUSIONS:** By using genomic, transcriptome, and epigenomic comparisons of EBV infected vs noninfected gastric cancer cells and tumor samples, we identified alterations in genes, gene expression, and methylation that affect

different signaling networks. These might be involved in EBV-associated gastric carcinogenesis.

**Keywords:** Genome Sequencing; Transcriptome; Methylation; *AKT2*.

Epstein–Barr virus (EBV) is a human herpes virus that infects more than 90% of the world population before adolescence. This oncogenic virus has been identified in epithelial malignancies including gastric cancer.<sup>1</sup> EBV-associated gastric cancer accounts for 8%–10% of all gastric cancer cases and is estimated to occur in more than 90,000 patients annually.<sup>2</sup> EBV-associated (EBV(+)) gastric cancer represents a distinct subtype of gastric cancer, with unique clinicopathologic features as compared with EBV-negative (EBV(-)) gastric cancer. However, the molecular genetic changes that account for the malignant behavior of EBV-associated gastric cancer remain largely unclear.

Clonal EBV is present in nearly all neoplastic cells and thus suggests a causal role in gastric carcinogenesis. In healthy individuals, EBV infection of gastric epithelial cells is a rare event. Even if EBV infects gastric epithelial cells, EBV usually is cytotoxic and induces cell death. However, once triggered, EBV infection will evolve into a persistent latent infection, which initiates progression into gastric cancer. Previous studies on EBV-associated gastric cancer by us<sup>3</sup>

\*Authors share co-first authorship.

**Abbreviations used in this paper:** AGS-hygro, AGS cells with hygro vector producing hygromycin-resistance; AP-1, activator protein-1; 5-Aza, 5-Aza-2'-deoxycytidine; CCNA1, cyclin A1; DNMT, DNA methyltransferase; ERK, extracellular signal-regulated kinase; EBV, Epstein–Barr virus; EBV(-), Epstein–Barr virus negative; EBV(+), Epstein–Barr virus positive; LMP2A, latent membrane protein 2A; MAPK, mitogen-activated protein kinase; p-AKT, phosphorylated AKT; qPCR, quantitative polymerase chain reaction; RPKM, reads per kilobase per million; RT, reverse transcription; SNV, single-nucleotide variant; indel, small insertion and deletion; SV, structural variation; TGFBR1, transforming growth factor  $\beta$  receptor 1.

© 2014 by the AGA Institute. Open access under CC BY-NC-ND license. 0016-5085

<http://dx.doi.org/10.1053/j.gastro.2014.08.036>

and others<sup>4</sup> have focused mostly on aberrant host gene methylation, which is a consequence of increased activity of DNA methyltransferases caused by EBV gene expression such as latent membrane protein 2A (LMP2A). Other studies also have investigated host genetic abnormalities including gene mutation,<sup>5</sup> microsatellite instability,<sup>6</sup> and cytogenetics<sup>7</sup> in EBV-associated gastric cancer. These findings collectively infer that EBV infection affects host cells at both epigenomic and genomic levels during gastric carcinogenesis.

However, systematic and integrative analyses concerning the impact of EBV on host cell alterations have not been performed to date. The AGS-EBV cell model with stable EBV infection has been applied successfully to study the effects of EBV infection in gastric cancer by us<sup>3,8</sup> and others.<sup>9,10</sup> Successful identification of EBV-associated methylated genes in gastric cancer using the AGS-EBV cell model highlights the feasibility of studying EBV-associated aberrations in gastric cancer using this cell model. The purpose of this study was to systematically elucidate the molecular genetic characteristics of EBV-associated gastric cancer by cataloguing the genomic and epigenomic alterations detected by whole-genome sequencing, transcriptome sequencing, and epigenome analysis in AGS-EBV cells as compared with the parental EBV-negative AGS cells, with an emphasis on identifying EBV-associated genomic/epigenomic events and aberrant molecular pathways. The identified important molecular abnormalities were verified further in primary EBV(+) gastric cancers.

## Materials and Methods

### AGS-EBV Cell Model

The AGS-EBV cell model stably infected with a recombinant EBV strain (added with a hygromycin-resistance gene for selective maintenance of EBV-positive cells during culture) was a gift from Dr Shannon C. Kenney (University of Wisconsin School of Medicine and Public Health).<sup>3</sup> The uninfected AGS cells, and AGS cells stably transfected with the empty pRI-GFP/Hygro vector producing hygromycin-resistance (AGS-hygro), were used as controls in this study.

### Human Gastric Samples

Gastric cancer samples were collected in the Prince of Wales Hospital, The Chinese University of Hong Kong from 1998 to 2004, and the First Affiliated Hospital of Sun Yat-sen University in Guangzhou from 1999 to 2006. The presence of EBV was determined by in situ hybridization analysis of EBV-encoded small RNA, and quantitative polymerase chain reaction (qPCR) examination of *BamH1 W* and *EBNA1* regions at the DNA level as described previously.<sup>1</sup> Gastric cancer samples with positive results for both in situ hybridization and qPCR examination were considered EBV-positive ( $n = 34$ ), whereas those with negative results for both were considered EBV-negative ( $n = 100$ ). Informed consent was provided by all participants, and this study was approved by both the ethics committee of the Chinese University of Hong Kong and the Clinical Research Ethics Committee of Sun Yat-sen University.

Other details and additional experimental procedures are provided in the [Supplementary Materials and Methods](#).

## Results

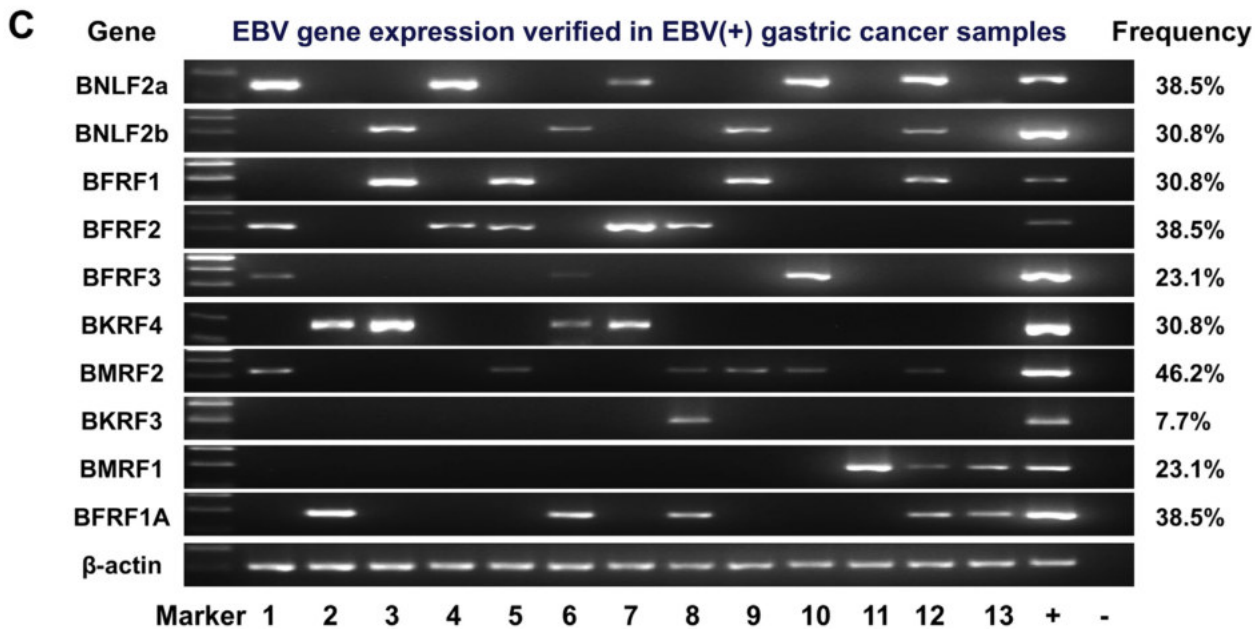
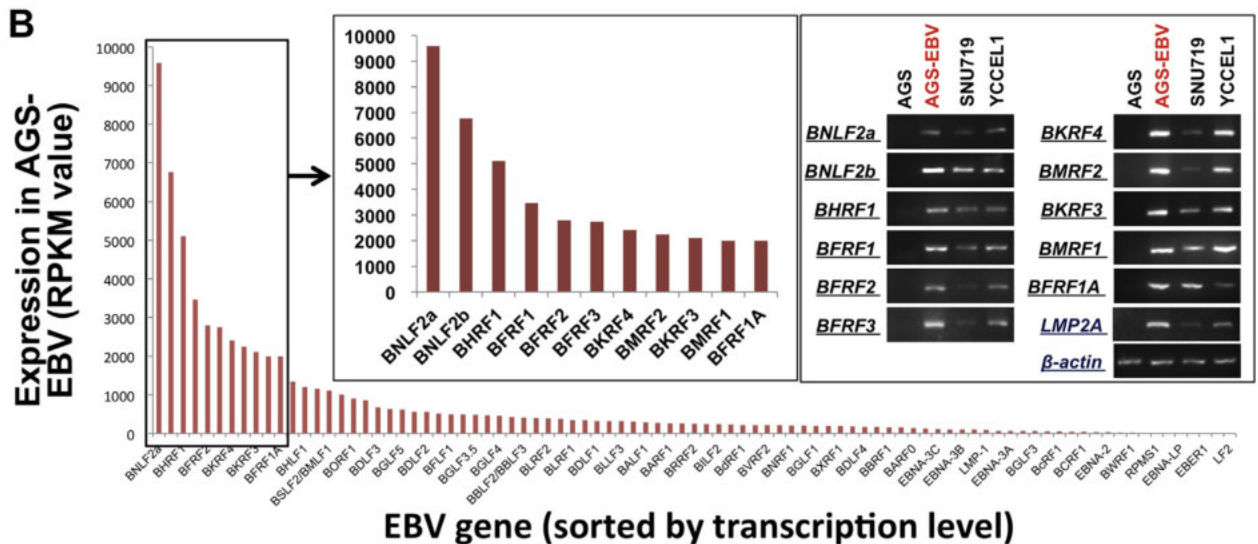
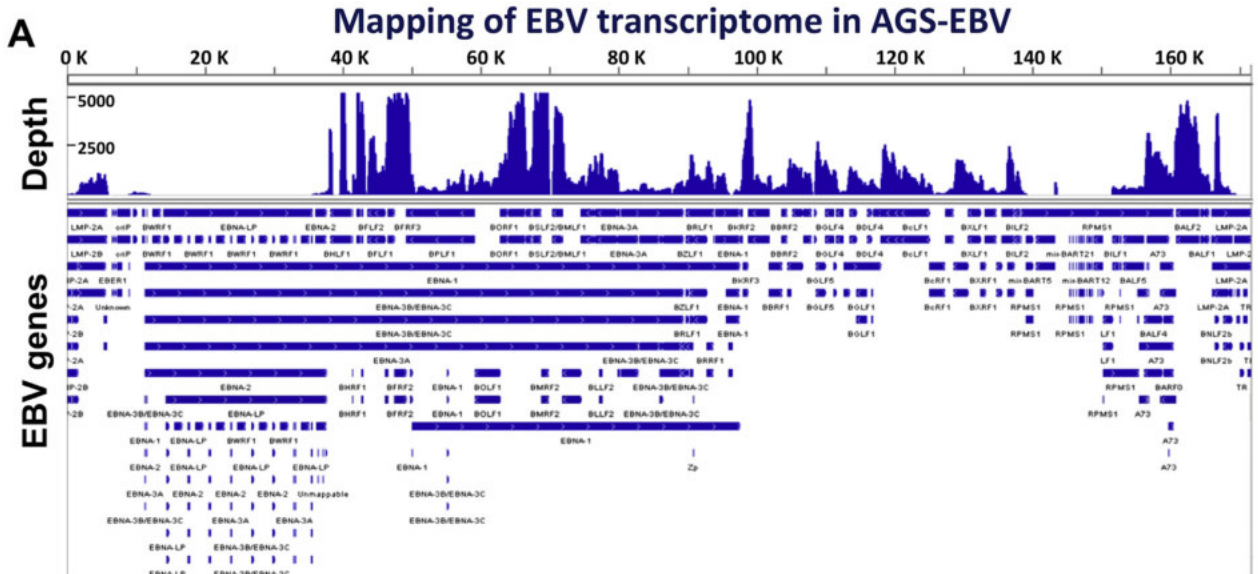
### EBV Copies and Viral Gene Expression in AGS-EBV Shown by Whole Genome and Transcriptome Sequencing

Whole-genome sequencing reads were mapped to both the human reference genome (UCSC hg19) and the EBV reference genome (NC\_007605). Whole-genome sequencing of the AGS-EBV and AGS cells showed a sequencing depth of 59-fold in AGS-EBV, and 42-fold in AGS for the human genome. A total of 91.59% and 91.57% of the whole genome region in AGS-EBV and AGS, respectively, were covered with more than 10 reads. Moreover, an 897-fold sequencing depth covering 91.38% of the whole EBV genome was obtained in AGS-EBV cells only ([Supplementary Figure 1A](#)). Therefore, approximately 15 EBV episomes in 1 AGS-EBV cell could be inferred (897-fold EBV/59-fold human = 15.2), consistent with the findings by others.<sup>11</sup>

In an attempt to uncover the EBV gene expression status in gastric cancer cells, 154.09 Mb reads of the AGS-EBV transcriptome were mapped to the EBV genome, with sequencing reads distributed across the entire EBV genome ([Figure 1A](#)). Visualization of transcriptome sequencing coverage across the EBV genome showed an EBV transcription profile in AGS-EBV cells with active regions similar to those identified in type I latency Burkitt's lymphoma cells ([Supplementary Figure 1B](#)).<sup>12</sup> Robust viral gene expression was yielded in AGS-EBV cells, with a median expression level of all genes being 255.4 reads per kilobase per million (RPKM) ([Figure 1B](#)). Transcriptome analysis of AGS-EBV identified the expression of 9 EBV genes (*BARF0*, *BARF1*, *BcLF1*, *BHRF1*, *BLLF1*, *BRLF1*, *BZLF1*, *EBNA1*, and *LMP2A*) previously detected in EBV(+) gastric tumors, and 71 EBV genes not reported previously in gastric cancer. The expression levels of these 71 genes are higher than that of *LMP2A* (27.0 RPKM), which could be well validated by reverse-transcription (RT)-PCR ([Figure 1B](#) and [Supplementary Tables 1 and 2](#)).

### EBV Gene Expression Identified in AGS-EBV Is Verified in EBV(+) Gastric Cancer Cell Lines and Primary Gastric Cancer Tissues

The top 11 EBV genes (*BNLF2a*, *BNLF2b*, *BHRF1*, *BFRF1*, *BFRF2*, *BFRF3*, *BKRF4*, *BMRF2*, *BKRF3*, *BMRF1*, and *BFRF1A*) were verified in AGS-EBV and 2 other EBV(+) gastric cancer cell lines with natural EBV infection (SNU719 and YCCEL1) by RT-PCR. The expression of all 11 genes was detected in the 3 EBV(+) gastric cancer cell lines, but not in EBV(-) AGS cells ([Figure 1B](#)). Notably, *BHRF1*, a viral oncogene detected in EBV(+) gastric cancer,<sup>13,14</sup> was the third most highly transcribed EBV gene in AGS-EBV (5103.9 RPKM). The other 10 genes have not been examined previously in primary gastric cancer, including the DNA replication or repair enzyme *BKRF3*, capsid or tegument coding genes (*BFRF1*, *BFRF3*, and *BKRF4*), the gene facilitating viral attachment to cells (*BMRF2*), 2 uncharacterized EBV genes (*BFRF2* and *BFRF1A*), and 3 lytic genes (*BNLF2a*, *BNLF2b*, and *BMRF1*). We performed immunofluorescence



BASIC AND TRANSCRIPTIONAL AT

examination using the antibody against the immediately early lytic gene *BRLF1*, and found that 2% of the AGS-EBV cells were positive for BRLF1, which are entering the lytic phase of EBV replication (Supplementary Figure 1C). The 10 EBV genes verified in AGS-EBV, SNU719, and YCCEL1 were verified further in primary EBV(+) gastric cancer tissues with a positive detection rate between 7.7% and 46.2% by RT-PCR (Figure 1C). Expression of EBV genes may contribute to EBV-associated gastric carcinogenesis.

### EBV-Associated Host Genomic Alteration Landscape Identified by Whole-Genome Sequencing

We compared the whole genome sequences of AGS-EBV and AGS to identify EBV-caused host genomic alterations, including single-nucleotide variants (SNVs)/point mutations, small insertions and deletions (indels), and structural variations (SVs) (Supplementary Tables 3–8). A total of 139 EBV-associated SNVs covering 131 genes were identified to be of interest, including 45 nonsynonymous SNVs (affecting 44 genes), and 94 SNVs located at important regulatory regions (splice sites, 5- or 3-untranslated regions, and promoter regions; affecting 87 genes). We also found 56 indels covering 54 genes in AGS-EBV and 48 AGS-EBV-specific SV events affecting 24 genes and other nongene regions.

Seven randomly selected SNVs in 6 genes (*AKT2*, *CCNA1*, *TGFBR1*, *ACVR1C*, *MAP3K4*, and *NRXN1*) were well validated in AGS-EBV, but not in AGS or AGS-hygro by PCR followed by Sanger sequencing (Supplementary Figure 2A). Among them, *AKT2*, the putative oncogene documented with important functions in the cancer pathway of mitogen-activated protein kinase (MAPK) signaling, harbors 2 EBV-associated nonsynonymous SNVs. Two randomly selected indels (*FAM35A* and *ADAMTS12*) and 4 randomly selected SVs (*GGT7-IRS1*, *KMD3A-KMD3A*, *SMAD5-STXBP5*, and *NA-KDM3B*) also were well validated in AGS-EBV by PCR followed by Sanger sequencing, but were not detected in AGS or AGS-hygro cells (Supplementary Figure 2B and C).

### Mutation Validation Refined Mutated Genes in Primary EBV(+) Gastric Cancers

By comparing the 45 EBV-associated nonsynonymous host SNVs/point mutations (covering 44 genes) (Figure 2A) identified in AGS-EBV with the Catalogue of Somatic Mutations in Cancer database, which collects somatic mutations in human cancers, we found that all 44 genes had been recorded, but none of the 45 mutation sites had been documented (Supplementary Table 4), inferring the novelty and potential importance of these mutations caused by EBV infection.

To clarify if the EBV-associated mutations in AGS-EBV also occurred in primary EBV(+) gastric cancers, we performed Sanger sequencing to compare the prevalence of mutations in *AKT2*, *CCNA1*, *TGFBR1*, *ACVR1C*, and *MAP3K4* between EBV(+) and EBV(-) gastric cancer samples. These genes were chosen for validation because they are functionally important in human cancers<sup>15–25</sup> and their EBV-associated mutations were located within conserved domains (Supplementary Figure 3). When compared with EBV(-) gastric cancers, somatic mutations occurred significantly more frequently in EBV(+) gastric cancers in *AKT2* (38.2% vs 3%;  $P < .0001$ ), *CCNA1* (25% vs 4%;  $P = .004$ ), *MAP3K4* (20.8% vs 4%;  $P = .013$ ), and *TGFBR1* (25% vs 8%;  $P = .029$ ) (Figure 2B and Supplementary Figures 4–7).

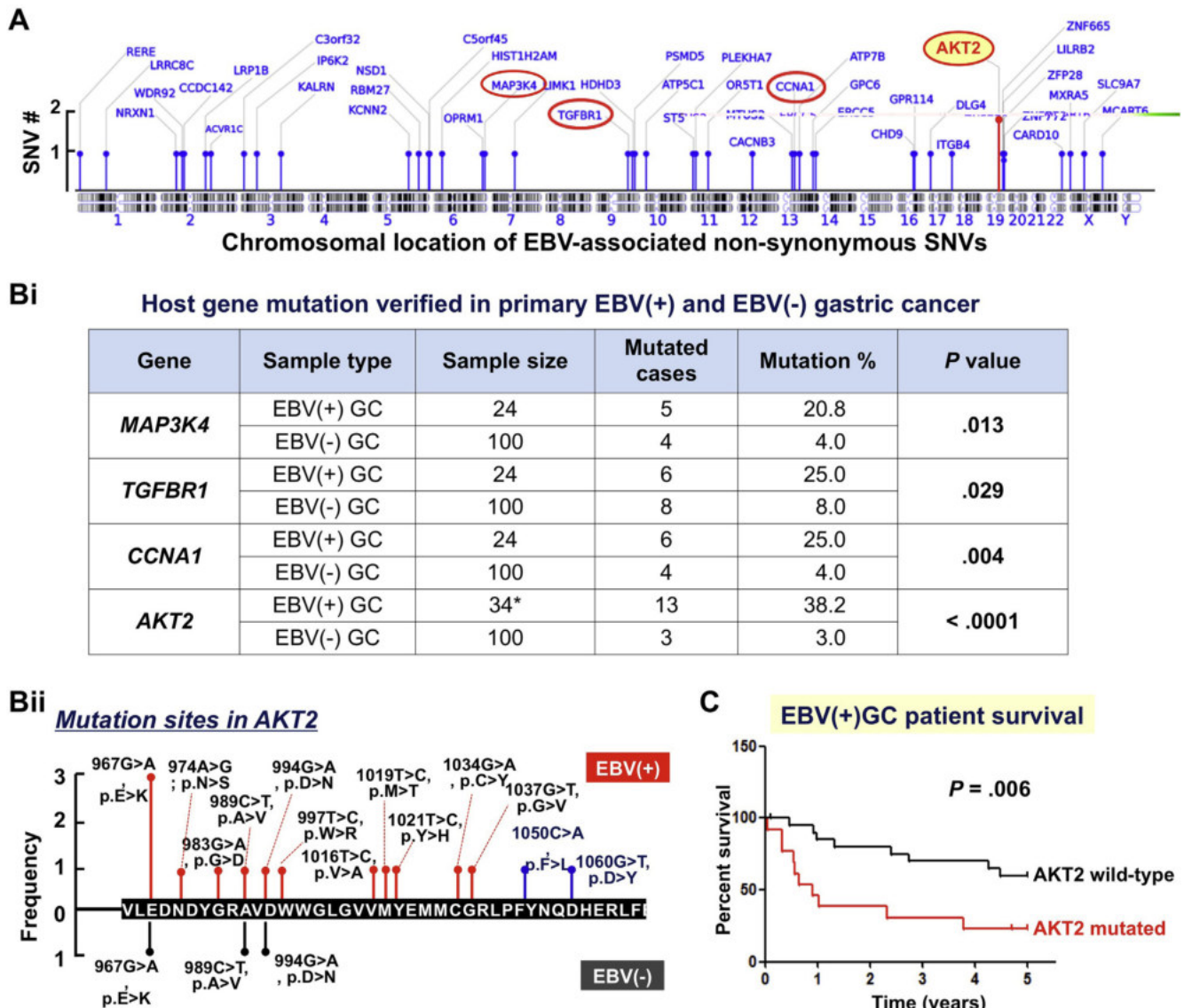
### Mutations in AKT2 Are Associated With Reduced Survival Times of Patients With EBV(+) Gastric Cancer

We further evaluated the clinical implication of mutations in the putative oncogene *AKT2*, which is the only gene harboring 2 EBV-associated nonsynonymous mutations in AGS-EBV cells, and mutation in which the most significant association with primary EBV(+) gastric cancer was shown. In the examined cohort of 34 EBV(+) gastric cancers with known follow-up data, the mutation frequency of *AKT2* was 38.2% (13 of 34) (Supplementary Tables 9 and 10). Interestingly, as shown in the Kaplan-Meier survival curves (Figure 2C), EBV(+) gastric cancer patients with an *AKT2* mutation had significantly reduced survival times (median, 3.27 y) than those with wild-type *AKT2* (median, 4.72 y;  $P = .006$ , log-rank test).

### Integrative Analysis of Epigenome and Transcriptome Showed Genes Dysregulated by Aberrant Methylation

To systematically identify genes directly dysregulated by epigenetic alterations induced by EBV infection, transcriptome of AGS-EBV, and AGS were analyzed integratively with the epigenome data. Integrated analysis showed that 216 genes were hypermethylated and transcriptionally down-regulated in AGS-EBV relative to AGS cells, whereas only 46 genes were demethylated and transcriptionally up-regulated in AGS-EBV (Figure 3A and Supplementary Table 11). Six randomly selected genes (*ACSS1*, *FAM3B*, *IHH*, *NEK9*, *SLC7A8*, and *TRABD*) were confirmed to be down-regulated significantly in AGS-EBV compared with AGS and AGS-hygro cells by semiquantitative RT-PCR and quantitative RT-PCR (Figure 3B). Down-regulation of these genes could be restored successfully in AGS-EBV cells by demethylation treatment using 5-Aza-2'-deoxycytidine (5-Aza) (Figure 3B).

**Figure 1.** EBV viral gene expression in AGS-EBV and in EBV(+) gastric cancer tissues. (A) Visualization of transcriptome sequencing coverage across the EBV reference genome (NC\_007605). The vertical axis shows the sequencing depth of mapping reads to each location of the EBV genome. The upper horizontal axis denotes the location of the EBV genome, and the EBV genes at each site are annotated at the lower panel using the Integrative Genomics Viewer (v2.1; Broad Institute, Cambridge, MA). (B) RPKM values for EBV gene expression levels. Expression of the top 11 genes was evaluated further in AGS and 3 EBV-positive gastric cancer cells (AGS-EBV, SNU719, and YCCEL1) by RT-PCR. (C) Expression of 10 EBV genes, unreported in EBV(+) gastric cancer previously, was evaluated in 13 primary EBV(+) gastric cancer samples by RT-PCR.



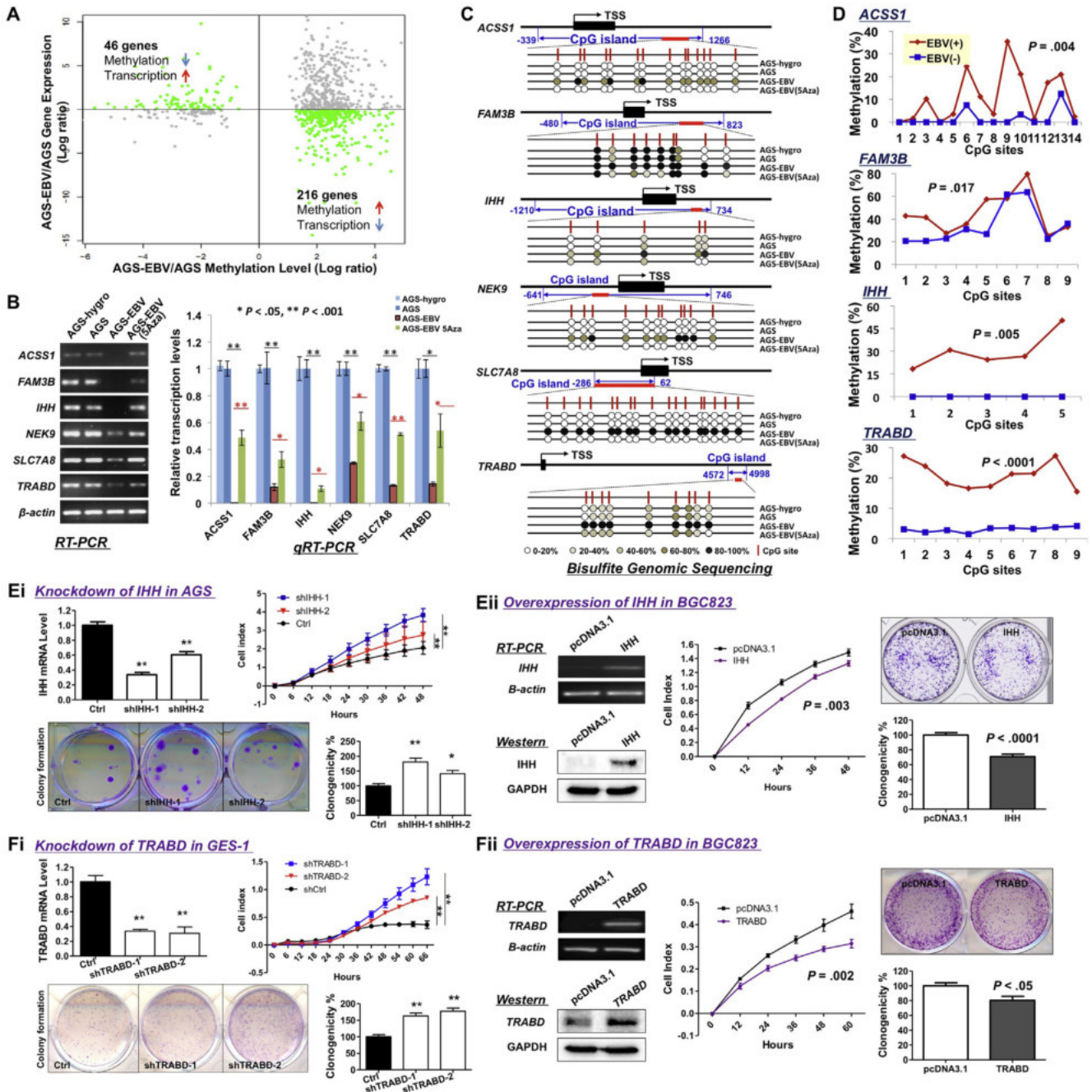
**Figure 2.** EBV-associated host gene mutations in primary gastric cancers. (A) Chromosomal location of 45 AGS-EBV-specific nonsynonymous SNVs covering 44 genes. The 5 genes screened further in primary gastric cancer samples are denoted. (B) Comparison of mutation frequencies in EBV(+) and EBV(-) gastric cancer samples by the chi-square or Fisher exact tests. \*Sample size later was increased for AKT2 genotype examination. (Bii) Schematic illustration of mutations found in EBV(+) and EBV(-) gastric cancer samples are denoted with red arrows and black arrows in the upper lanes and lower lanes, respectively. The mutation sites identified in AGS-EBV cells also are shown with blue arrows in the upper lane. Nucleotide and amino acid changes are indicated for each mutation site. The frequencies (number of cases) of each mutation are indicated by the height of the arrow. (C) Kaplan-Meier survival curves show that EBV(+) gastric cancer patients with mutated AKT2 had reduced survival times compared with patients with wild-type AKT2. This difference was statistically significant based on the log-rank test ( $P = .006$ ). GC, gastric cancer.

Higher methylation levels of these genes in AGS-EBV as compared with AGS and AGS-hygro cells were confirmed by bisulfite genomic sequencing, and the methylation levels were decreased successfully by 5-Aza treatment (Figure 3C). We have shown that DNA methyltransferase 3b (DNMT3b) was up-regulated in AGS-EBV compared with AGS cells.<sup>3</sup> There were no differences in messenger RNA expression; nuclear protein expression of DNMT1, DNMT3a, and DNMT3b; and the activity of DNMT3b between uninfected AGS and the vector-transfected, hygromycin-resistant AGS cells (Supplementary Figure 8). These findings suggest that

EBV infection causes a genome-wide aberrant methylation composed mainly of promoter/CpG island hypermethylation, which directly lead to gene transcriptional down-regulation.

*Validation Confirmed Genes Methylated Preferentially in Primary EBV(+) Gastric Cancers*

To clarify if aberrant methylation caused by EBV infection in AGS-EBV cells also occurred in primary gastric cancers, promoter methylation statuses of ACSS1, FAM3B, IHH, and TRABD were examined in EBV(+) and EBV(-)



**Figure 3.** Identification and validation of aberrantly methylated genes involved in EBV-associated gastric cancer. (A) Dot plots of genes of differential transcription and methylation between AGS and AGS-EBV cells shown by methylated DNA immunoprecipitation microarray chip and transcriptome sequencing, respectively. (B) Transcriptional down-regulation of the selected genes in AGS-EBV and restoration by demethylation treatment using 5-Aza were shown. (C) Bisulfite genomic sequencing confirmed that selected candidates were methylated at significantly higher levels in AGS-EBV cells than in AGS cells, and could be demethylated by 5-Aza treatment. The location of CpG island and bisulfite genomic sequencing target regions was shown for each gene, with *black bars* denoting the first exons. TSS, transcription start site. (D) Comparison of promoter methylation of 4 EBV-associated methylated genes in 20 EBV(+) and 20 EBV(-) primary gastric cancers by bisulfite genomic sequencing. Average methylation levels at each site for EBV(+) and EBV(-) samples are shown. *P* values were obtained by paired *t* tests. Knock-down of (Ei) IHH and (Fi) TRABD expression was performed by transfecting short hairpin RNAs (shRNAs) specifically targeting 2 different regions of each gene in cells of high expression, AGS and GES-1, respectively. Overexpression of (Eii) IHH and (Fii) TRABD in BGC823 cells was achieved by stable transfection of expression vectors. Cell growth and colony formation ability were compared correspondingly.

BASIC AND  
TRANSLATIONAL AT

gastric cancers using bisulfite genomic sequencing. Significantly higher methylation levels were observed in EBV(+) gastric cancers as compared with EBV(-) gastric cancers in *ACSS1* (12.6% vs 2.0%;  $P = .004$ ), *FAM3B* (44.6% vs 34.0%;  $P = .017$ ), *IHH* (30.1% vs 0.0%;  $P = .005$ ), and *TRABD* (20.9% vs 3.0%;  $P = .000$ ) (Figure 3D).

### EBV-Associated Methylated Genes Possess Tumor-Suppressive Potentials

We further investigated the function of 2 genes methylated in EBV(+) gastric cancers (*IHH* and *TRABD*). Gene knock-down or ectopic expression was obtained by stable transfection of specific short hairpin RNA or open reading frame-expressing vectors in cells with high or low endogenous expression of the corresponding gene. Knock-down of *IHH* by short hairpin RNA transfection in AGS cells significantly increased cell growth and colony formation ability compared with the control cells, whereas overexpression of *IHH* in the silenced cell line BGC823 significantly inhibited cell growth and colony formation (Figure 3E). Similarly, knock-down of *TRABD* significantly increased cell growth and the colony formation ability of GES-1 cells, whereas overexpression of *TRABD* in BGC823 cells significantly inhibited cell growth and colony formation (Figure 3F). These results show that *IHH* and *TRABD* possess potential tumor-suppressive properties and their down-regulation by hypermethylation may play roles in EBV-associated gastric carcinogenesis.

### Host Genetic and Epigenetic Alterations Caused by EBV Infection Commonly Involve Five Intercorrelated Cancer Pathways

To investigate the dysregulated pathways by EBV infection-induced host genomic and epigenomic changes, enrichment analysis for Kyoto Encyclopedia of Genes and Genomes pathways was conducted using 205 genes with genetic alterations and 262 genes with aberrant methylation-mediated transcriptional changes, respectively (Figure 4A). Genetically changed genes were found to be enriched in 13 pathways, whereas epigenetically changed genes were enriched in 15 pathways (with  $\geq 4$  genes involved in each pathway; adjusted  $P < .05$ ). Notably, hypermethylated genes were found to be enriched in only 10 pathways ( $\geq 4$  genes;  $P < .05$ ). Eight pathways were dysregulated significantly by both genetic and epigenetic changes. Interestingly, these 8 pathways also were dysregulated significantly by hypermethylation only (Figure 4B and Supplementary Table 12). Because pathways in cancer and metabolic pathways can be hit easily by enrichment analysis, and all altered genes in the colorectal cancer pathway are included in pathways in cancer, we paid attention to the remaining 5 important affected pathways, including axon guidance, focal adhesion, cytokine-cytokine receptor interaction, MAPK signaling, and regulation of actin cytoskeleton.

Diagrams showing genetically or epigenetically altered genes in the 5 core pathways are shown in Figure 5. Remarkably, these 5 pathways are intercorrelated. The axon guidance pathway correlates with cytokine-cytokine receptor interaction, regulation of actin cytoskeleton, and

MAPK signaling pathways; focal adhesion also correlates with cytokine-cytokine receptor interaction, regulation of actin cytoskeleton, and MAPK signaling pathways (Supplementary Figure 9). Importantly, the putative oncogene *AKT2*, mutation of which was found to be associated with reduced survival times of patients with EBV(+) gastric cancer, is involved in 2 of the 5 core pathways (focal adhesion and MAPK signaling) (Figure 5). Collectively, axon guidance, focal adhesion, cytokine-cytokine receptor interaction, MAPK signaling, and regulation of actin cytoskeleton pathways are the core pathways dysregulated during EBV-associated gastric carcinogenesis.

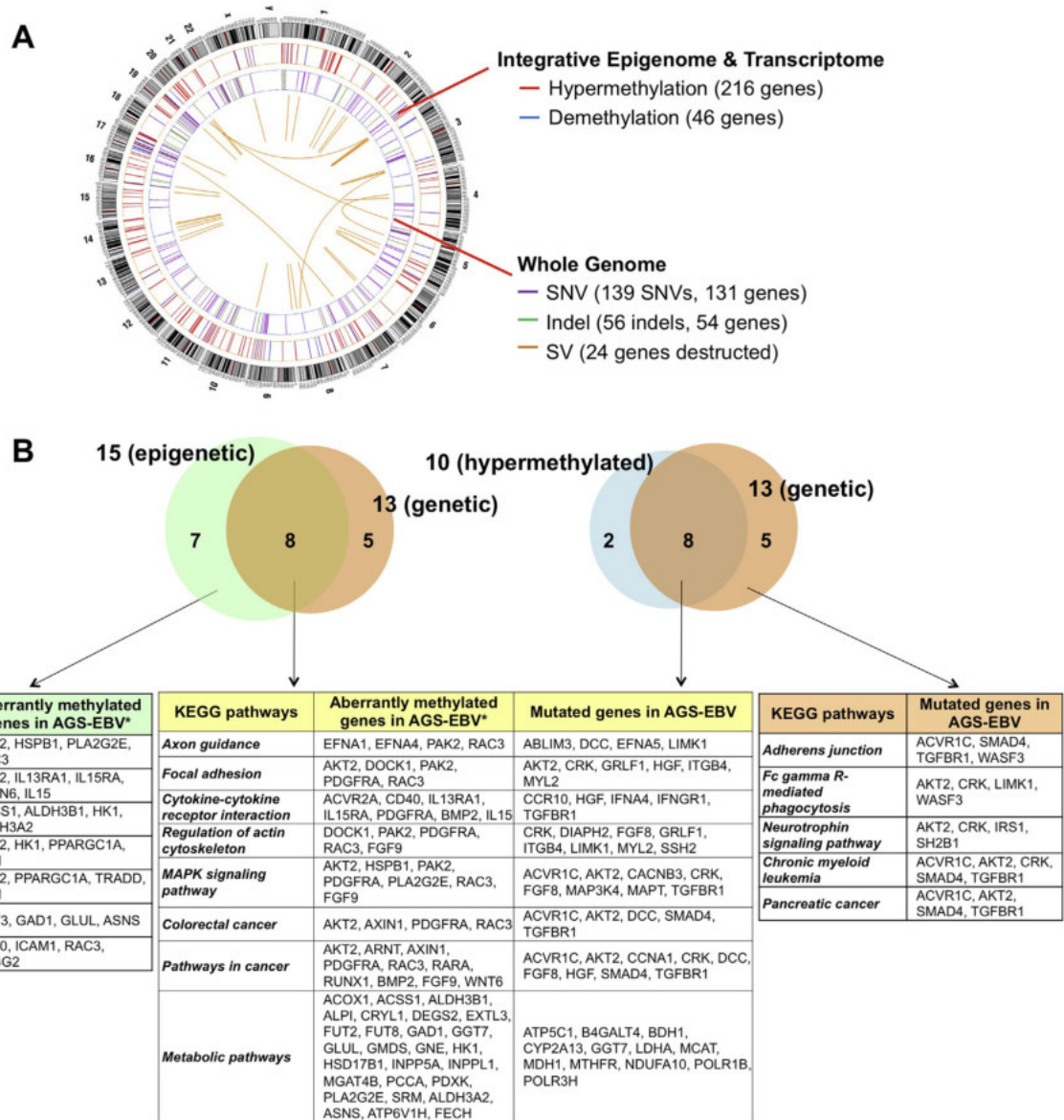
### AKT2 Was Activated by Mutation and Participated in Dysregulating MAPK Signaling

We investigated the effects of *AKT2* mutation on *AKT2* activity through assessing *AKT2* phosphorylation by Western blot and total *AKT* kinase activity by activity assays. Our results showed that the phosphorylated *AKT2* (p-*AKT2*) level was significantly higher in AGS-EBV as compared with AGS, and in mutant *AKT2*-transfected AGS than in wild-type *AKT2*-transfected AGS cells (Figure 6A). In concordance with enhanced p-*AKT2*, total *AKT* kinase activity was increased significantly in mutant *AKT2*-carrying AGS-EBV compared with AGS, and in mutant *AKT2*-carrying AGS compared with wild-type *AKT2*-overexpressed AGS (Figure 6A).

Activator protein-1 (AP-1) and extracellular signal-regulated kinase (ERK) are pivotal mediators in MAPK signaling involving *AKT2*. We evaluated the effects of *AKT2* mutation on the activities of AP-1 and ERK by promoter luciferase activity assays using promoter reporters containing AP-1 and serum response element (SRE) binding elements, respectively. Results showed that both AP-1 and ERK activities were increased significantly in mutant *AKT2*-carrying cells compared with wild-type *AKT2*-carrying cells (Figure 6B). To further confirm the role of *AKT2* mutation on AP-1 and ERK activity, mutant and wild-type *AKT2* were expressed ectopically in the immortalized normal gastric epithelial cell line GES-1 with low endogenous *AKT2* expression. Again, a higher p-*AKT2* level, increased total *AKT* kinase activity, and promoted AP-1 and ERK activities were detected in mutant *AKT2*-transfected GES-1 cells compared with wild-type *AKT2*-transfected GES-1 cells (Figure 6C and D). Moreover, mutant *AKT2* was found to promote cell growth and colony formation ability of GES-1 cells as compared with wild-type *AKT2*. These results imply that *AKT2* was activated by mutation and participated in dysregulating MAPK signaling.

## Discussion

The AGS-EBV cell model, a gastric epithelial cell model with stable EBV infection, has been applied successfully to study the effect of EBV infection on host gene transcription and methylation.<sup>3,8-10</sup> This cell model also has facilitated our integrative genome-wide scan for alterations in EBV-associated gastric cancer in this study by comparison with its parental AGS cells.



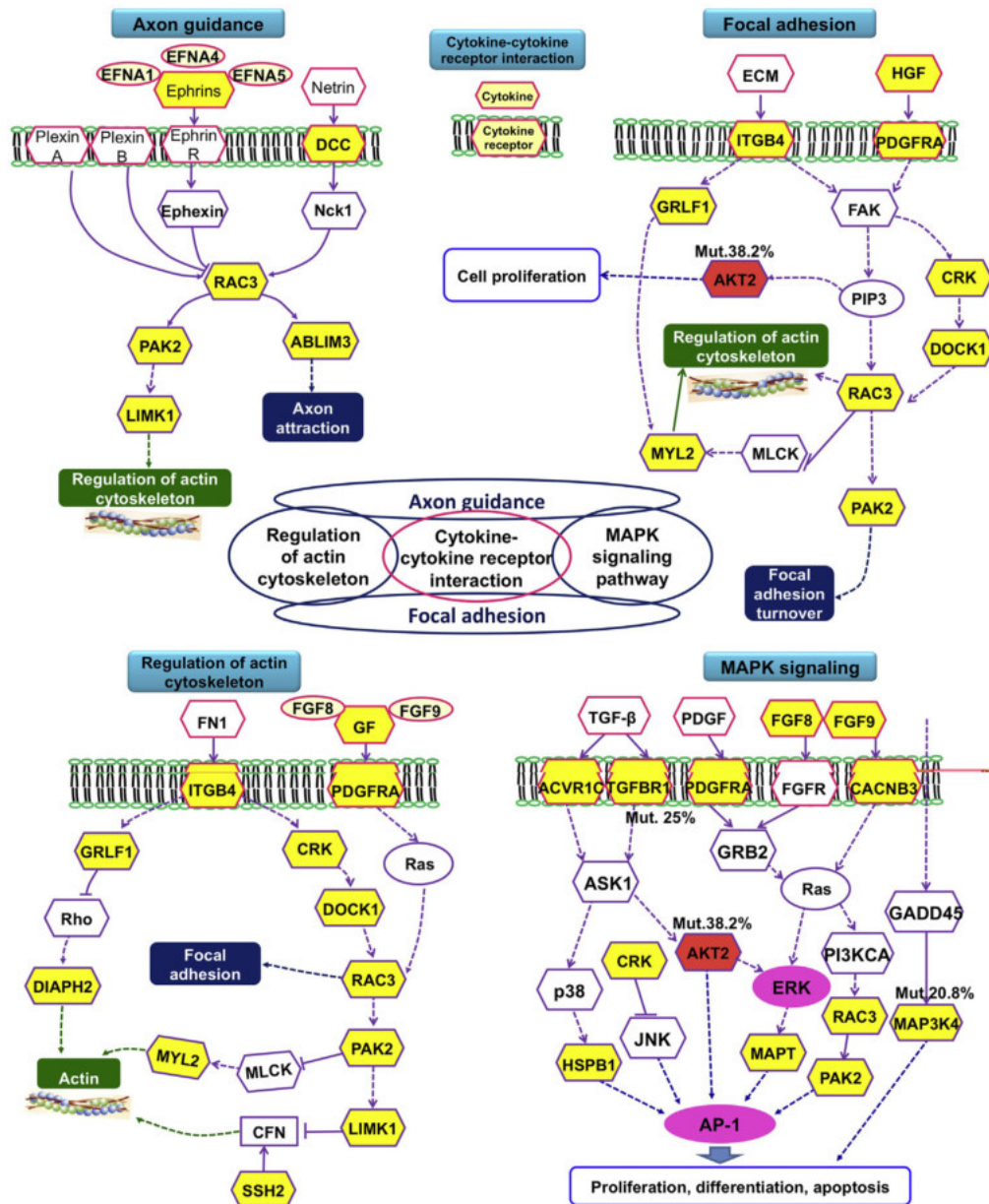
**Figure 4.** Pathways dysregulated by EBV-associated host genomic and epigenomic alterations. (A) Circos illustration of EBV-associated epigenetic and genetic changes identified in AGS-EBV as compared with AGS cells. (B) Enrichment of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with genetically or epigenetically changed genes. Pathways with 4 or more genes involved and a *P* value less than .05 are considered as significantly affected. Fc gamma R, Fc gamma receptor; Jak-STAT, Janus kinase/signal transducers and activators of transcription; VEGF, vascular endothelial growth factor.

Transcriptome sequencing showed 9 well-documented EBV genes (*BARF0*, *BHRF1*, *BcLF1*, *BHRF1*, *BLLF1*, *BRLF1*, *BZLF1*, *EBNA1*, and *LMP2A*) in EBV-associated gastric cancer,<sup>14,26–29</sup> and, notably, 71 EBV genes unreported in gastric cancer. Importantly, the 10 top unreported EBV genes all were verified in other EBV(+) gastric cancer cells and primary EBV(+) gastric cancer samples (Figure 1B and C). The frequencies of these EBV genes in EBV(+) gastric cancers all were significant except the one for the *BKRF3* gene (7.7%) when compared with those in EBV(-) gastric cancers (0%; *n* = 20, chi-square test). Expression of previously unreported EBV genes may be involved in EBV-associated gastric cancer. Expression of EBV genes with potential oncogenic function has been reported in EBV-associated gastric

carcinogenesis, including *BARF1*,<sup>29</sup> *BHRF1*,<sup>13,14</sup> and *RPMS1* (encoding BART's microRNAs).<sup>30</sup> Expression of the latent gene *LMP2A* has been reported to up-regulate survivin, contributing to the survival advantage of EBV-associated gastric cancer cells,<sup>31</sup> and activate cellular DNMT3b, causing the genome-wide aberrant methylation of host cells.<sup>3</sup> EBV resides in the host cell nucleus as an episome during latency infection and the EBV genome is too large (approximately 170 kb) to be integrated into the host genome. Therefore, EBV might induce host genetic and epigenetic variants through executing its repertoire of gene expression programs, subsequently contributing to the unique pathobiology of virus-associated gastric cancer. Identification of the previously unreported EBV genes in

BASIC AND TRANSLATIONAL AT





**Figure 5.** Diagram illustration of the 5 core pathways commonly dysregulated by epigenetic and genetic changes. Genetically or epigenetically altered genes in the 5 core pathways are denoted with a yellow background. Mutation frequencies of *AKT2*, *TGFBR1*, and *MAP3K4* in primary EBV(+) gastric cancers are indicated. ECM, extracellular matrix; PIP3, phosphatidylinositol 3,4,5-trisphosphate. The other abbreviations listed are gene names.

BASIC AND  
TRANSLATIONAL AT

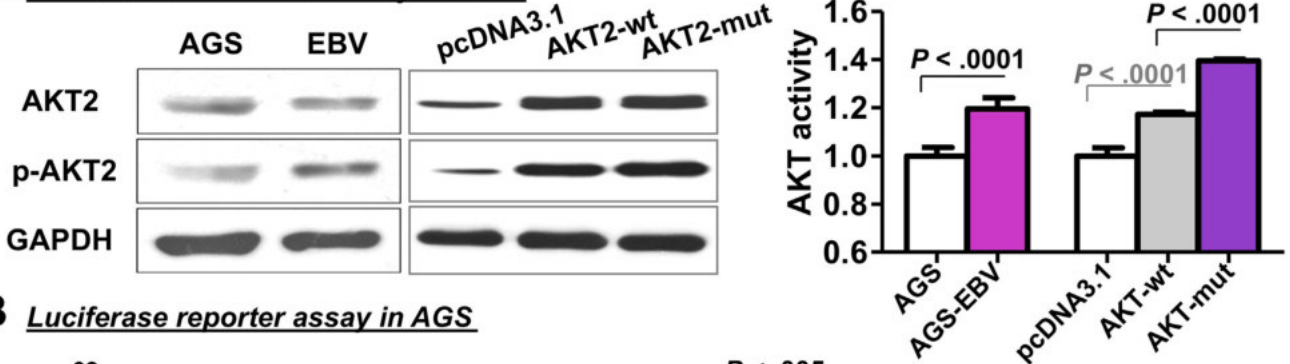
this study will add new insight into the role of EBV infection in contributing to this subtype of gastric carcinogenesis.

By analyzing the epigenome data integratively with transcriptome data in this study, we identified 216 genes transcriptionally down-regulated by EBV-caused hypermethylation and 46 genes transcriptionally up-regulated by demethylation. Genes with inconsistent changes in methylation and transcription might be the result of involvement of other regulatory mechanisms such as microRNAs and transcription factors.<sup>10,32</sup> Further validation has confirmed that promoter methylation levels of *ACSS1*, *FAM3B*, *IHH*, and *TRABD* were significantly higher in primary EBV(+) than in EBV(-) gastric cancers, with tumor-suppressive potential shown by gain-of-function and loss-of-function experiments

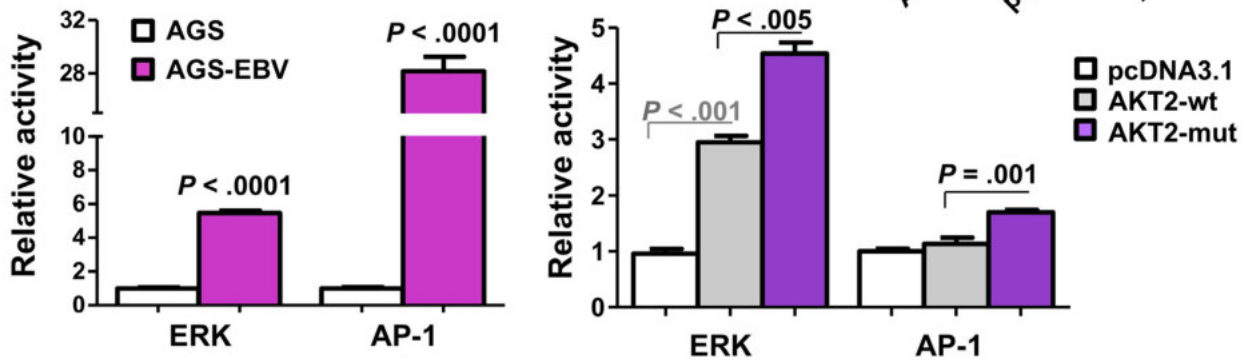
*in vitro* (Figure 3). Previous reports from us and others have shown that promoter methylation of *SSTR1*, *RECB*, *p14*, *p15*, *p16*, *p73*, *APC*, *E-cadherin*, and *PTEN* are associated with EBV-associated gastric cancer.<sup>3,8,33-35</sup> These results suggest that EBV infection causes hypermethylation of a specific group of genes, and silencing of these genes may favor malignant transformation of gastric epithelial cells during development of this unique subtype of gastric cancer.

Whole-genome sequencing of the AGS-EBV and AGS cells identified EBV infection-associated genetic alterations affecting 205 host genes. Among the 44 genes harboring amino acid-changing mutations, we confirmed that mutations of *AKT2*, *CCNA1*, *MAP3K4*, and *TGFBR1* were associated significantly with EBV(+) gastric cancers (Figure 2B). No mutations in these genes were detected in the

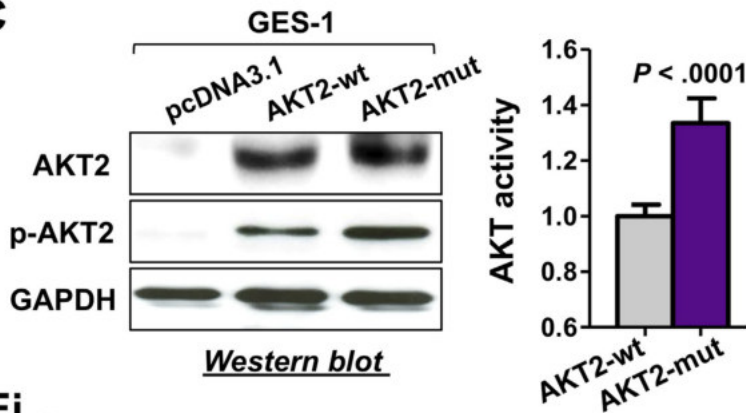
**A** *Western blot & AKT activity in AGS*



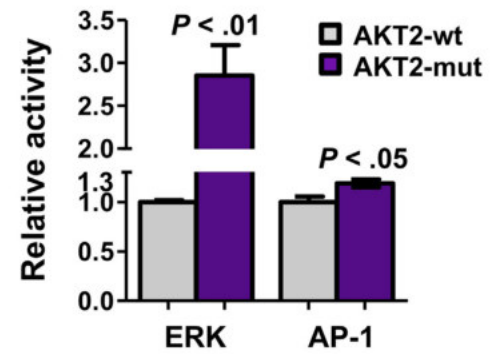
**B** *Luciferase reporter assay in AGS*



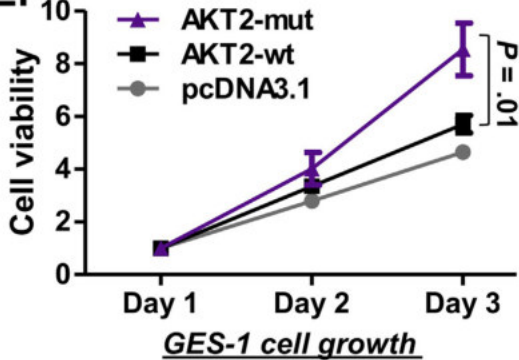
**C**



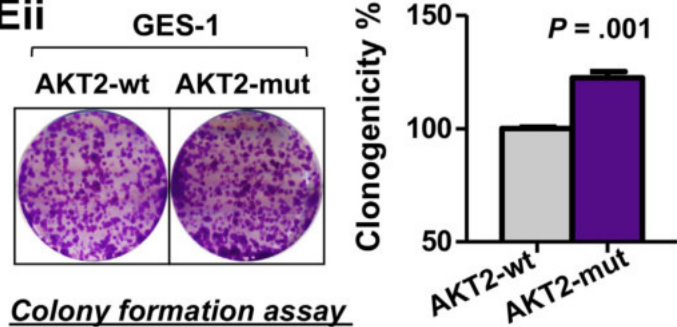
**D** *Luciferase reporter assay in GES-1*



**Ei**



**Eii**

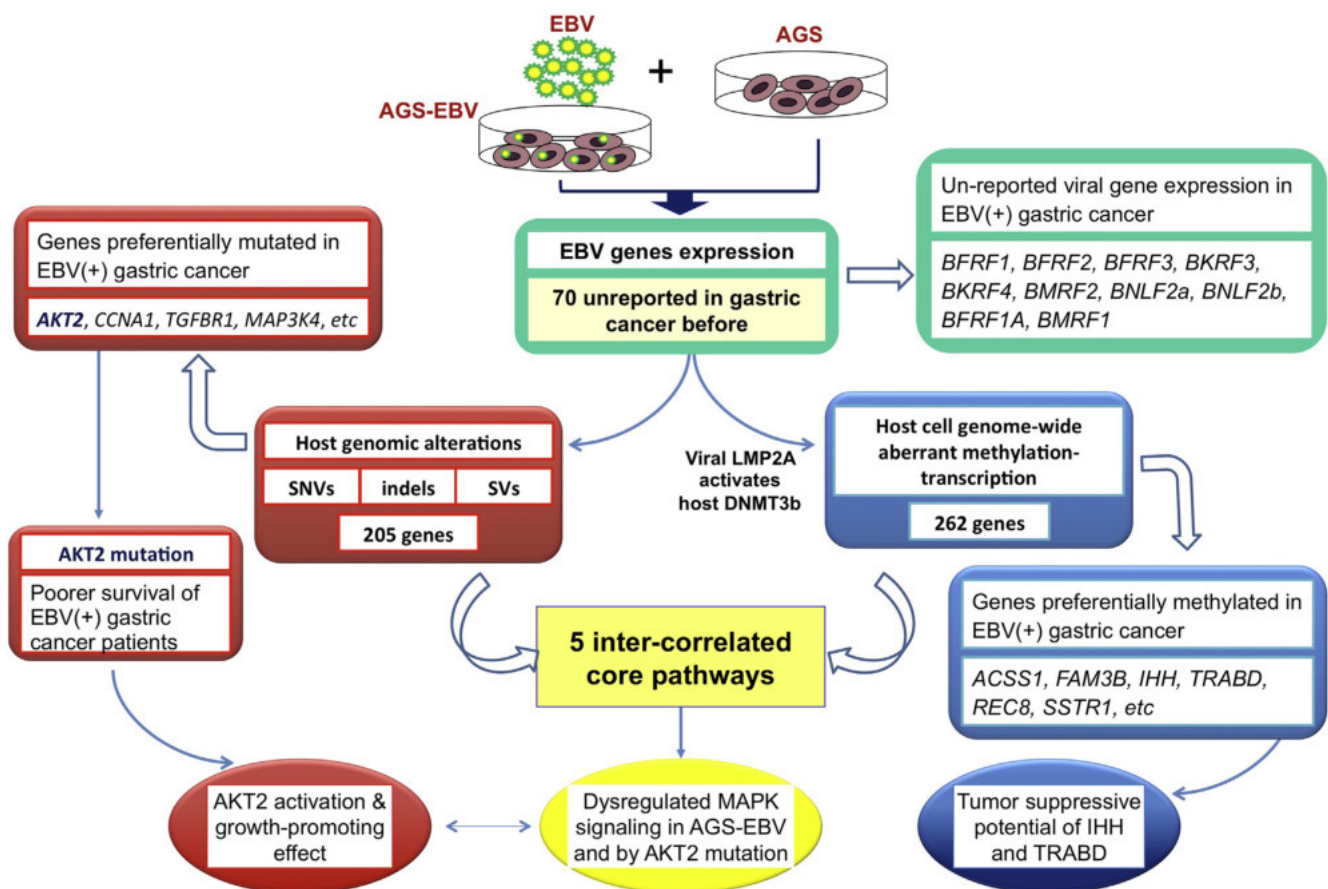


**Figure 6.** AKT2 was activated by mutation and functioned in dysregulating MAPK signaling. (A) Total and phosphorylated AKT2 protein levels were assessed by Western blot. Total AKT kinase activity was evaluated by AKT kinase activity assay. (B) Luciferase reporter assays indicated that the important MAPK signaling mediators (AP-1 and ERK) were both activated in mutant AKT2-carrying cells as compared with wild-type AKT2-carrying cells. (C) After ectopic expression in GES-1 cells, mutant AKT2 was phosphorylated at a higher level, and accompanied by increased total AKT kinase activity as compared with wild-type AKT2. (D) Mutant AKT2 activated AP-1 and ERK signaling as compared with wild-type AKT2. (E) Mutant AKT2 promoted cell viability and colony formation ability as compared with wild-type AKT2. GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

corresponding nontumor tissues or in 30 noncancerous stomach samples (data not shown).

*AKT2* is a putative oncogene encoding a protein that participates in important cancer pathways such as MAPK signaling. Mutation of *AKT2* has been investigated in human cancers,<sup>15,16</sup> but not in EBV-associated gastric cancer. Cyclin A1 (*CCNA1*) belongs to the cyclin family, and primarily functions in the control of the germline meiotic cell cycle. Previous studies have shown that *CCNA1* play different roles in virus-related and non-virus-related malignancies.<sup>17–20</sup> However, mutation of *CCNA1* has never been reported. *CCNA1* mutations in EBV(+) gastric cancer as identified by us might suggest another mechanism of the role of *CCNA1* in human malignancies. Transforming growth factor- $\beta$ -receptor 1 (*TGFBR1*) is a serine/threonine protein kinase and receptor for *TGF- $\beta$* . Mutations in *TGFBR1* have been found in skin and colorectum cancers.<sup>21,22</sup> *MAP3K4* functions as a major mediator of environmental stressors that activate the p38 MAPK pathway,<sup>23</sup> and its mutation has been reported in endometrial cancer.<sup>24</sup> Recognizing the functional importance of these genes in human cancers, mutations of these genes caused by EBV infection may contribute at least in part to the pathogenesis of EBV-associated gastric cancer.

Finally, 5 intercorrelated core pathways (axon guidance, focal adhesion, cytokine-cytokine receptor interaction, MAPK signaling, and regulation of actin cytoskeleton) were found to be commonly enriched with genetically and epigenetically changed genes caused by EBV infection. In addition to the several epigenetically or genetically changed up-stream and down-stream targets of focal adhesion kinase in the focal adhesion pathway we identified (Figure 5), focal adhesion kinase phosphorylation has been reported to be increased by EBV infection and the subsequently increased cell motility in AGS cells.<sup>36</sup> This finding further supports the importance of the focal adhesion pathway in EBV-associated gastric cancer. Promoted anchorage-independent growth of EBV-infected AGS in soft agar, a hallmark phenotype of cellular transformation, has been reported by others.<sup>10</sup> We also have observed a more undifferentiated morphology of AGS-EBV as compared with AGS when both cells were cultured in the same F12 medium (not shown). These phenotype changes might be associated with the focal adhesion pathway. Although the other 4 pathways have never been reported in EBV-associated cancer, 3 of them (cytokine-cytokine receptor interaction, MAPK signaling, and regulation of actin cytoskeleton) have been reported to be affected by EBV



**Figure 7.** Study summary. EBV gene expression profile, EBV-associated host genomic and epigenomic alterations were identified in a cell model and validated further in primary EBV(+) gastric cancers. Our previous study showed that EBV viral gene *LMP2A* expression activates DNMT3b and causes genome-wide aberrant methylation in AGS-EBV. Five core pathways were found to be enriched significantly with genetically and epigenetically altered genes caused by EBV infection. The functional importance of selected methylated and mutated genes as well as the pathway was shown.

infection in lymphoblastoid cell lines and in primary B cells,<sup>37,38</sup> suggesting common dysregulation of these pathways by EBV infection in different cell types during disease initiation. Dysregulation of the 5 core pathways through both genetic and epigenetic modulation of host genes by EBV infection may play important roles during this subtype of gastric carcinogenesis.

Among all the host genes altered by EBV infection identified in this study, *AKT2* is the most notable one. *AKT2* was the only gene (of 44 genes) harboring 2 nonsynonymous point mutations identified in AGS-EBV cells. *AKT2* mutation was also the highest in frequency and associated most significantly with primary EBV(+) gastric cancer as compared with EBV(-) gastric cancer. Importantly, we further confirmed that mutations in *AKT2* were associated with reduced survival in EBV(+) gastric cancer patients. Interestingly, *AKT2* is also the only gene involved in 2 of the 5 core pathways (focal adhesion and MAPK signaling). The mutant form of *AKT2* identified in AGS-EBV possessed higher kinase activity, increased activities of the important mediators of the MAPK signaling pathway (AP-1 and ERK), and exerted a promoting effect on cell growth as compared with wild-type *AKT2* (Figure 6). All these findings emphasize the importance of *AKT2* in connection with EBV(+) gastric cancer.

In summary, as shown in Figure 7, this study systematically showed the EBV-associated genomic and epigenomic alterations in gastric cancer. Expression of EBV genes in gastric cancer was shown by transcriptome analysis of the EBV-infected cell model and further confirmed in EBV(+) primary gastric cancers. Whole-genome sequencing showed EBV-associated host mutations in genes such as *AKT2*, *CCNA1*, *MAP3K4*, and *TGFBR1*, and mutations in *AKT2* are associated with reduced survival times of patients with EBV(+) gastric cancer. Epigenome analysis uncovered hypermethylation of genes including *ACSS1*, *IHH*, *FAM3B*, and *TRABD* through EBV infection. Five core pathways were shown to be dysregulated by EBV-associated host genomic and epigenomic aberrations in gastric cancer. Moreover, the functional importance of selected genes (*IHH*, *TRABD*, and *AKT2*) and pathway (MAPK) were shown further. These findings provide a systematic view of EBV-associated host genomic and epigenomic abnormalities and signaling networks that may govern the pathogenesis of EBV-associated gastric cancer.

## Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at [www.gastrojournal.org](http://www.gastrojournal.org), and at <http://dx.doi.org/10.1053/j.gastro.2014.08.036>.

## References

1. Zhao J, Jin H, Cheung KF, et al. Zinc finger E-box binding factor 1 plays a central role in regulating Epstein-Barr virus (EBV) latent-lytic switch and acts as a therapeutic target in EBV-associated gastric cancer. *Cancer* 2012; 118:924–936.
2. Fukayama M, Hino R, Uozaki H. Epstein-Barr virus and gastric carcinoma: virus-host interactions leading to carcinoma. *Cancer Sci* 2008;99:1726–1733.
3. Zhao J, Liang Q, Cheung KF, et al. Genome-wide identification of Epstein-Barr virus-driven promoter methylation profiles of human genes in gastric cancer cells. *Cancer* 2013;119:304–312.
4. Hino R, Uozaki H, Murakami N, et al. Activation of DNA methyltransferase 1 by EBV latent membrane protein 2A leads to promoter hypermethylation of PTEN gene in gastric carcinoma. *Cancer Res* 2009; 69:2766–2774.
5. Wang K, Kan J, Yuen ST, et al. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet* 2011;43:1219–1223.
6. Chang MS, Lee HS, Kim HS, et al. Epstein-Barr virus and microsatellite instability in gastric carcinogenesis. *J Pathol* 2003;199:447–452.
7. Chan WY, Liu Y, Li CY, et al. Recurrent genomic aberrations in gastric carcinomas associated with Helicobacter pylori and Epstein-Barr virus. *Diagn Mol Pathol* 2002;11:127–134.
8. Zhao J, Liang Q, Cheung KF, et al. Somatostatin receptor 1, a novel EBV-associated CpG hypermethylated gene, contributes to the pathogenesis of EBV-associated gastric cancer. *Br J Cancer* 2013;108:2557–2564.
9. Ryan JL, Jones RJ, Kenney SC, et al. Epstein-Barr virus-specific methylation of human genes in gastric cancer cells. *Infect Agent Cancer* 2010;5:27.
10. Marquitz AR, Mathur A, Shair KH, et al. Infection of Epstein-Barr virus in a gastric carcinoma cell line induces anchorage independence and global changes in gene expression. *Proc Natl Acad Sci U S A* 2012;109: 9593–9598.
11. Shannon-Lowe C, Adland E, Bell AI, et al. Features distinguishing Epstein-Barr virus infections of epithelial cells and B cells: viral genome expression, genome maintenance, and genome amplification. *J Virol* 2009; 83:7749–7760.
12. Lin Z, Xu G, Deng N, et al. Quantitative and qualitative RNA-Seq-based evaluation of Epstein-Barr virus transcription in type I latency Burkitt's lymphoma cells. *J Virol* 2010;84:13053–13058.
13. Dawson CW, Dawson J, Jones R, et al. Functional differences between BHRF1, the Epstein-Barr virus-encoded Bcl-2 homologue, and Bcl-2 in human epithelial cells. *J Virol* 1998;72:9016–9024.
14. Luo B, Wang Y, Wang XF, et al. Correlation of Epstein-Barr virus and its encoded proteins with Helicobacter pylori and expression of c-met and c-myc in gastric carcinoma. *World J Gastroenterol* 2006;12:1842–1848.
15. Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012;486:400–404.
16. Soung YH, Lee JW, Nam SW, et al. Mutational analysis of AKT1, AKT2 and AKT3 genes in common human carcinomas. *Oncology* 2006;70:285–289.
17. Yanatatsanejit P, Chalermchai T, Kerekhanjanarong V, et al. Promoter hypermethylation of CCNA1, RARRES1,

- and HRASLS3 in nasopharyngeal carcinoma. *Oral Oncol* 2008;44:400–406.
18. Sartor MA, Dolinoy DC, Jones TR, et al. Genome-wide methylation and expression differences in HPV(+) and HPV(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. *Epigenetics* 2011;6:777–787.
  19. Wegiel B, Bjartell A, Tuomela J, et al. Multiple cellular mechanisms related to cyclinA1 in prostate cancer invasion and metastasis. *J Natl Cancer Inst* 2008;100:1022–1036.
  20. Coletta RD, Christensen K, Reichenberger KJ, et al. The Six1 homeoprotein stimulates tumorigenesis by reactivation of cyclin A1. *Proc Natl Acad Sci U S A* 2004;101:6478–6483.
  21. **Arnault JP, Mateus C**, Escudier B, et al. Skin tumors induced by sorafenib; paradoxical RAS-RAF pathway activation and oncogenic mutations of HRAS, TP53, and TGFBR1. *Clin Cancer Res* 2012;18:263–272.
  22. Carvajal-Carmona LG, Churchman M, Bonilla C, et al. Comprehensive assessment of variation at the transforming growth factor beta type 1 receptor locus and colorectal cancer predisposition. *Proc Natl Acad Sci U S A* 2010;107:7858–7862.
  23. Takekawa M, Posas F, Saito H. A human homolog of the yeast Ssk2/Ssk22 MAP kinase kinase kinases, MTK1, mediates stress-induced activation of the p38 and JNK pathways. *EMBO J* 1997;16:4973–4982.
  24. **Le Gallo M, O'Hara AJ**, Rudd ML, et al. Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nat Genet* 2012;44:1310–1315.
  25. Xu G, Zhou H, Wang Q, et al. Activin receptor-like kinase 7 induces apoptosis through up-regulation of Bax and down-regulation of Xiap in normal and malignant ovarian epithelial cell lines. *Mol Cancer Res* 2006;4:235–246.
  26. Shukla SK, Prasad KN, Tripathi A, et al. Expression profile of latent and lytic transcripts of Epstein-Barr virus in patients with gastroduodenal diseases: a study from northern India. *J Med Virol* 2012;84:1289–1297.
  27. Hoshikawa Y, Satoh Y, Murakami M, et al. Evidence of lytic infection of Epstein-Barr virus (EBV) in EBV-positive gastric carcinoma. *J Med Virol* 2002;66:351–359.
  28. Qiu J, Cosmopoulos K, Pegtel M, et al. A novel persistence associated EBV miRNA expression profile is disrupted in neoplasia. *PLoS Pathog* 2011;7:e1002193.
  29. zur Hausen A, Brink AA, Craanen ME, et al. Unique transcription pattern of Epstein-Barr virus (EBV) in EBV-carrying gastric adenocarcinomas: expression of the transforming BARF1 gene. *Cancer Res* 2000;60:2745–2748.
  30. **Kim do N, Chae HS**, Oh ST, et al. Expression of viral microRNAs in Epstein-Barr virus-associated gastric carcinoma. *J Virol* 2007;81:1033–1036.
  31. Hino R, Uozaki H, Inoue Y, et al. Survival advantage of EBV-associated gastric carcinoma: survivin up-regulation by viral latent membrane protein 2A. *Cancer Res* 2008;68:1427–1435.
  32. Chang LS, Wang JT, Doong SL, et al. Epstein-Barr virus BGLF4 kinase downregulates NF-kappaB transactivation through phosphorylation of coactivator UXT. *J Virol* 2012;86:12176–12186.
  33. Ushiku T, Chong JM, Uozaki H, et al. p73 gene promoter methylation in Epstein-Barr virus-associated gastric carcinoma. *Int J Cancer* 2007;120:60–66.
  34. Geddert H, Zur Hausen A, Gabbert HE, et al. EBV-infection in cardiac and non-cardiac gastric adenocarcinomas is associated with promoter methylation of p16, p14 and APC, but not hMLH1. *Cell Oncol (Dordr)* 2011;34:209–214.
  35. Sudo M, Chong JM, Sakuma K, et al. Promoter hypermethylation of E-cadherin and its abnormal expression in Epstein-Barr virus-associated gastric carcinoma. *Int J Cancer* 2004;109:194–199.
  36. Kassis J, Maeda A, Teramoto N, et al. EBV-expressing AGS gastric carcinoma cell sublines present increased motility and invasiveness. *Int J Cancer* 2002;99:644–651.
  37. **Zhao B, Mar JC, Maruo S**, et al. Epstein-Barr virus nuclear antigen 3C regulated genes in lymphoblastoid cell lines. *Proc Natl Acad Sci U S A* 2011;108:337–342.
  38. **Price AM, Tourigny JP**, Forte E, et al. Analysis of Epstein-Barr virus-regulated host gene expression changes through primary B-cell outgrowth reveals delayed kinetics of latent membrane protein 1-mediated NF-kappaB activation. *J Virol* 2012;86:11096–11106.

---

Author names in bold designate shared co-first authorship.

Received September 2, 2013. Accepted August 23, 2014.

#### Reprint requests

Address requests for reprints to: Jun Yu, MD, PhD, Department of Medicine and Therapeutics, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, New Territory, Hong Kong. e-mail: junyu@cuhk.edu.hk; fax: (852) 21445330.

#### Acknowledgments

Sequencing data deposition: all sequencing data from this study have been deposited in the NCBI Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra>); accession number: SRA067982.

#### Conflicts of interest

The authors disclose no conflicts.

#### Funding

This project was supported by research funds from the China 863 Program (2012AA02A203), the National Natural Science Foundation of China (81272304), the China 973 Program (2010CB529305), the Theme-based Research Scheme of the Hong Kong Research Grants Council (T12-403-11), Shenzhen Municipal Science and Technology R&D fund (JCYJ 20120619152326450), and Shenzhen Virtual University Park Support Scheme to CUHK Shenzhen Research Institute.

# **APPENDIX 1**

## **PUBLISHED PEER-REVIEWED ARTICLES REVERENT TO THE THESIS**

Title	Authors	Journal name	IF	CS	No of Cit.	Contribution towards publication	Access domains / identifier
<b>EPIGENETICS</b>							
Akt3 induces oxidative stress and DNA damage by activating the NADPH oxidase via phosphorylation of p47phox	Christos Polytarchou*, Maria Hatzia Apostolou*, <b>Tung On Yau</b> , Niki Christodoulou, Philip W Hinds, Filippos Kottakis, Ioannis Sanidas, Philip N Tsihchlis	<i>Proceedings of the National Academy of Sciences of the United States of America.</i> <b>2020</b> ; 117 (46), 28806-28815.	12.779	16.2	9	<ul style="list-style-type: none"> <li>• Co-author (30%)</li> <li>• Performed wet lab-based experiments</li> </ul>	Journal URL: <a href="https://pnas.org/content/117/46/28806">pnas.org/content/117/46/28806</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/33139577/">33139577</a> DOI: <a href="https://doi.org/10.1073/pnas.2017830117">10.1073/pnas.2017830117</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC7682348/">PMC7682348</a>
<b>HEPATIC DISEASES</b>							
CXCL10 plays a key role as an inflammatory mediator and a non-invasive biomarker of non-alcoholic steatohepatitis	Xiang Zhang, Jiayun Shen, Kwan Man, Eagle S H Chu, <b>Tung On Yau</b> , Joanne C Y Sung, Minnie Y Y Go, Jun Deng, Liwei Lu, Vincent W S Wong, Joseph J Y Sung, Geoffrey Farrell, Jun Yu	<i>Journal of Hepatology.</i> <b>2014</b> ;61 (6), 1365-1375.	30.083	39.2	161	<ul style="list-style-type: none"> <li>• Co-author (20%)</li> <li>• Conduct animal experiments and for histological examination</li> <li>• Hepatic cell culture</li> <li>• Manuscript revision</li> </ul>	Journal URL: <a href="https://journal-of-hepatology.eu/article/S0168-8278(14)00475-9">journal-of-hepatology.eu/article/S0168-8278(14)00475-9</a> ScienceDirect: <a href="https://sciencedirect.com/science/article/pii/S0168827814004759">sciencedirect.com/science/article/pii/S0168827814004759</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/25048951/">25048951</a> DOI: <a href="https://doi.org/10.1016/j.jhep.2014.07.006">10.1016/j.jhep.2014.07.006</a>
Management of chronic hepatitis B infection: Current treatment guidelines, challenges, and new developments	Ceen-Ming Tang, <b>Tung On Yau</b> , Jun Yu	<i>World Journal of Gastroenterology.</i> <b>2014</b> ;20 (20), 6262-6278.	5.374	8.1	169	<ul style="list-style-type: none"> <li>• Co-author (45%)</li> <li>• Design the search strategy</li> <li>• Wrote the manuscript in multiple sub-sections</li> </ul>	Journal URL: <a href="http://wjnet.com/1007-9327/full/v20/i20/6262.htm">wjnet.com/1007-9327/full/v20/i20/6262.htm</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/24876747/">24876747</a> DOI: <a href="https://doi.org/10.3748/wjg.v20.i20.6262">10.3748/wjg.v20.i20.6262</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC4033464/">PMC4033464</a>

IF, impact factor (2021); CS, CiteScore (2021); Cit., citation, based on Google Scholar record dated August 08, 2022



Title	Authors	Journal name	IF	CS	No of Cit.	Contribution towards publication	Access domains / Identifier
<b>MICROBIOLOGY AND GENOMICS</b>							
Full-Length Genome of an <i>Ogataea polymorpha</i> Strain CBS4732 <i>ura3Δ</i> Reveals Large Duplicated Segments in Subtelomeric Regions	Jia Chang, Jinlong Bei, Qi Shao, Hemu Wang, Huan Fan, <b>Tung On Yau</b> , Wenjun Bu, Jishou Ruan, Dongshang Wei, Shan Gao	<i>Frontiers in Microbiology</i> . <b>2022</b> ;13, 855666.	6.064	8.2	1	<ul style="list-style-type: none"> <li>• Co-author (10%)</li> <li>• Data analysis</li> </ul>	Journal URL: <a href="https://www.frontiersin.org/articles/10.3389/fmicb.2022.855666">frontiersin.org/articles/10.3389/fmicb.2022.855666</a> DOI: <a href="https://doi.org/10.3389/fmicb.2022.855666">10.3389/fmicb.2022.855666</a> PMID: <a href="https://pubmed.ncbi.nlm.nih.gov/35464988/">35464988</a> PMCID: <a href="https://pubmed.ncbi.nlm.nih.gov/PMC9019687/">PMC9019687</a>

IF, impact factor (2021); CS, CiteScore (2021); Cit., citation, based on Google Scholar record dated August 08, 2022





# Akt3 induces oxidative stress and DNA damage by activating the NADPH oxidase via phosphorylation of p47<sup>phox</sup>

Christos Polytarchou<sup>a,b,c,d,1,2</sup>, Maria Hatzia Apostolou<sup>b,c,d,1</sup>, Tung On Yau<sup>b,c</sup> , Niki Christodoulou<sup>b,c</sup>, Philip W. Hinds<sup>d,e</sup>, Filippos Kottakis<sup>d</sup>, Ioannis Sanidas<sup>d,f</sup> , and Philip N. Tsichlis<sup>a,d,2</sup>

<sup>a</sup>Department of Cancer Biology and Genetics, The Ohio State University Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210; <sup>b</sup>Department of Biosciences, John van Geest Cancer Research Centre, School of Science and Technology, Nottingham Trent University, NG11 8NS Nottingham, United Kingdom; <sup>c</sup>Centre for Health, Aging and Understanding Disease, Nottingham Trent University, NG11 8NS Nottingham, United Kingdom; <sup>d</sup>Molecular Oncology Research Institute, Tufts Medical Center, Boston, MA 02111; <sup>e</sup>Department of Developmental, Molecular and Chemical Biology, Tufts Cancer Center, Tufts University School of Medicine, Boston, MA 02111; and <sup>f</sup>Department of Medicine, Massachusetts General Hospital Cancer Center and Harvard Medical School, Charlestown, MA 02129

Edited by Peter K. Vogt, Scripps Research Institute, La Jolla, CA, and approved September 25, 2020 (received for review September 18, 2020)

**Akt activation up-regulates the intracellular levels of reactive oxygen species (ROS) by inhibiting ROS scavenging. Of the Akt isoforms, Akt3 has also been shown to up-regulate ROS by promoting mitochondrial biogenesis. Here, we employ a set of isogenic cell lines that express different Akt isoforms, to show that the most robust inducer of ROS is Akt3. As a result, Akt3-expressing cells activate the DNA damage response pathway, express high levels of p53 and its direct transcriptional target miR-34, and exhibit a proliferation defect, which is rescued by the antioxidant *N*-acetylcysteine. The importance of the DNA damage response in the inhibition of cell proliferation by Akt3 was confirmed by Akt3 overexpression in *p53*<sup>-/-</sup> and *INK4a*<sup>-/-</sup>/*Arf*<sup>-/-</sup> mouse embryonic fibroblasts (MEFs), which failed to inhibit cell proliferation, despite the induction of high levels of ROS. The induction of ROS by Akt3 is due to the phosphorylation of the NADPH oxidase subunit p47<sup>phox</sup>, which results in NADPH oxidase activation. Expression of Akt3 in *p47*<sup>phox</sup><sup>-/-</sup> MEFs failed to induce ROS and to inhibit cell proliferation. Notably, the proliferation defect was rescued by wild-type p47<sup>phox</sup>, but not by the phosphorylation site mutant of p47<sup>phox</sup>. In agreement with these observations, Akt3 up-regulates p53 in human cancer cell lines, and the expression of Akt3 positively correlates with the levels of p53 in a variety of human tumors. More important, Akt3 alterations correlate with a higher frequency of mutation of p53, suggesting that tumor cells may adapt to high levels of Akt3, by inactivating the DNA damage response.**

Akt isoforms | NADPH oxidase | oxidative stress | DNA damage | cancer

The utilization of oxygen by all aerobic organisms gives rise to a heterogeneous group of molecules and radicals with distinct chemical properties, levels of reactivity, and biological impact, which are collectively known as reactive oxygen species (ROS). Under this term we include oxygen anions and radicals [superoxide (O<sub>2</sub><sup>-</sup>) and hydroxyl radical (OH·)], singlet oxygen (<sup>1</sup>O<sub>2</sub>), hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>), nitric oxide (NO·), peroxynitrite anions (ONOO<sup>-</sup>), and various peroxides (ROOR') and hydroperoxides (ROOH) (1, 2). Reactive oxygen species are generated by multiple mechanisms. Prominent among them are: 1) The escape of electrons during oxidative phosphorylation in the mitochondria. At least 2% of the electrons traveling along the respiratory chain in mitochondria escape and target oxygen to form superoxide (3, 4). 2) The enzymatic activation of the NOX family of NADPH oxidases. The multisubunit enzyme NADPH oxidase is most abundant in leukocytes, where it becomes activated when the cells engulf invading microorganisms. The activation of this enzyme promotes the rapid generation of superoxide whose sharp increase in these cells is known as the “respiratory burst” (5–7). Enzymes related to the leukocyte-specific NADPH oxidase are also present in other cell types (8). The preceding mechanisms produce ROS via processes endogenous to the cell (*SI Appendix, Fig. S1*). In addition to

the endogenous sources of ROS, there are exogenous sources such as environmental pollutants (4, 9).

NOX2, the first characterized NOX isoform, is an enzyme complex that mediates the transfer of electrons to molecular oxygen through the action of flavocytochrome b558, which is composed of two proteins gp91<sup>phox</sup> and p22<sup>phox</sup> and is located at the plasma membrane. Activation of the enzyme is induced via its association with a trimeric cytosolic protein complex, composed of p40<sup>phox</sup>, p47<sup>phox</sup>, and p67<sup>phox</sup>. Upon stimulation by external signals, p47<sup>phox</sup> undergoes phosphorylation and the trimeric complex translocates to the membrane to form the active oxidase by binding to b558 (10). The activation of the complex also requires two guanine nucleotide-binding proteins, cytosolic Rac2 and membrane-bound Rap1A (5, 11, 12). There are seven isoforms of gp91<sup>phox</sup>, the catalytic transmembrane subunit of NOX. Based on this, the NOX family of oxidases consists of seven members, with each member exhibiting a distinct tissue distribution.

Signals originating in membrane receptors have been shown to induce ROS in a variety of cell types (13–15). ROS produced in

## Significance

Although the three Akt isoforms share mechanisms of activation and exhibit overlaps in their downstream signaling pathways, significant differences are now being uncovered. Here, we show that among Akt isoforms, Akt3 preferably phosphorylates p47<sup>phox</sup> and activates NADPH oxidase, resulting in robust induction of reactive oxygen species (ROS). Akt3-induced ROS activate the DNA damage response and upregulate p53 expression. Consequently, the proliferation rate of Akt3-expressing cells is reduced, an effect reversed by p53 loss. In cancer, Akt3 expression correlates with the abundance of p53; however, tumors can adapt to high Akt3 activity by inactivating the DNA damage response. These findings reveal differences in the regulation of ROS by Akt isoforms, which may be exploited therapeutically for Akt3-driven cancers.

Author contributions: C.P. and P.N.T. designed research; C.P., M.H., T.O.Y., N.C., F.K., and I.S. performed research; P.W.H. and P.N.T. contributed new reagents/analytic tools; C.P., M.H., and I.S. analyzed data; and C.P., M.H., and P.N.T. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

<sup>1</sup>C.P. and M.H. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: christos.polytarchou@ntu.ac.uk or Philip.Tsichlis@osumc.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2017830117/-DCSupplemental>.

First published November 2, 2020.

response to such signals function as second messengers by inducing reversible oxidation of a number of signaling molecules, including peroxiredoxins (16, 17) and various phosphatases, such as PTP-1B, SHP-2, and the tumor suppressor PTEN (18–20). Most relevant as a second messenger among ROS is  $H_2O_2$ , which is relatively stable and diffuses easily across biological membranes (21, 22). A critical target of ROS-producing signals originating in membrane receptors is the NADPH oxidase. Thus, ROS-producing insulin signals in adipocytes depend on NOX-4, a homolog of gp91<sup>Phox</sup>, the catalytic subunit of NADPH oxidase (23). Although the function of some ROS may be physiologically important, persistence of ROS is generally harmful (3, 21). Their toxicity is due to irreversible oxidation of a variety of macromolecules such as DNA, lipids, carbohydrates, and proteins. To prevent ROS toxicity, cells employ a variety of detoxification mechanisms, which collectively are responsible for the very short half-life of ROS. The main mechanisms of ROS detoxification include: 1) ROS-induced oxidation of small antioxidant molecules; and 2) enzymatic conversion of ROS to less reactive species. Examples include the conversion of superoxide to  $H_2O_2$  by superoxide dismutase, the conversion of  $H_2O_2$  to  $H_2O$  and oxygen by catalase, and the action of peroxiredoxins. The overall levels of ROS in cells are determined by the balance between the rate of production and the rate of conversion (3, 17, 24, 25).

Akt1, also known as protein kinase B $\alpha$  (PKB $\alpha$ ), is the founding member of a protein kinase family composed of three members, Akt1, Akt2, and Akt3. Akt family members regulate a diverse array of cellular functions and play important roles in most types of human cancer. Akt activation depends on PtdIns-3,4,5- $P_3$ , and to a lesser extent on PtdIns-3,4- $P_2$ , both of which are products of phosphoinositide 3-kinase (26, 27). The interaction of PtdIns-3,4,5- $P_3$  with the PH domain of Akt, promotes the translocation of Akt to the plasma membrane where it undergoes phosphorylation at two sites, one in the activation loop and one in the carboxyl-terminal tail (26–28). Phosphorylated Akt may translocate from the plasma membrane to the cytosol or the nucleus. Activated Akt ultimately undergoes dephosphorylation by phosphatases and returns to the inactive state (29).

Recently, we and others showed that despite their sequence similarity, Akt isoforms exhibit dramatic signaling differences (30–35). Data in this study show that although all Akt isoforms up-regulate the cellular levels of ROS, it is Akt3, which exhibits maximum activity. Akt2-expressing cells are also characterized by relatively high levels of superoxide. However, the levels of  $H_2O_2$  in these cells are very low, perhaps because they express low levels of superoxide dismutase 1 and 2 (SOD1 and SOD2). Akt3 promotes the accumulation of superoxide and  $H_2O_2$  by preferentially activating the NADPH oxidase through p47<sup>phox</sup>. The high levels of ROS in Akt3-expressing cells induce DNA damage. This activates the DNA damage response, which inhibits cell proliferation. The levels of Akt3 in several types of human cancer correlate with the levels of p53. More important, *Akt3* amplification correlates positively with the mutation of *p53*. Our data suggest that tumor cells may adapt to Akt3 expression by inactivating the DNA damage response pathways.

## Experimental Procedures

**Expression Constructs.** Expression constructs of Myc-Akt1, Myc-Akt2, and Myc-Akt3 in the retroviral pBabe vectors were described earlier (30, 32, 33). Human p47<sup>phox</sup> was cloned in the vector pBabe-bleo. Ser and Thr to Ala mutations were introduced into this vector by the QuikChange II Site-Directed Mutagenesis Kit (Agilent, Inc.).

**Cells and Culture Conditions.** Akt1-, Akt2-, and Akt3-expressing and *Akt1/Akt2/Akt3* triple knockout (TKO) lung fibroblasts have been described previously (30, 32, 33). Briefly, lung fibroblasts cultured from an *Akt1<sup>fl/fl</sup>/Akt2<sup>-/-</sup>/Akt3<sup>-/-</sup>* mouse were spontaneously immortalized via a 3T3-type protocol. The immortalized cells were transduced with retroviral constructs of Myc-Akt1, Myc-Akt2, or Myc-Akt3 in the retroviral vector pBabe-puro, neo, bleo,

or GFP. Subsequently, they were transduced with a MigR1-Cre-hygro construct to ablate the floxed *Akt1* allele. This gave rise to isogenic Akt1-, Akt2-, or Akt3-expressing cells. To generate the TKO cells, the *Akt1<sup>fl/fl</sup>/Akt2<sup>-/-</sup>/Akt3<sup>-/-</sup>* lung fibroblasts were transduced with the MigR1-Cre-hygro construct and the transduced cells were selected with hygromycin. Cells were used soon after selection, because they do not proliferate but remain viable for about a week (33). The Akt1-, Akt2-, Akt3-, and Akt1/2/3-expressing cells and the TKO lung fibroblasts were cultured in DMEM (Dulbecco's Modified Eagle Medium) supplemented with 10% FBS (fetal bovine serum) and antibiotics. Selected cultures were treated with 5 mM of the antioxidant *N*-acetylcysteine (NAC, A7250, Sigma-Aldrich).

*p47<sup>phox</sup>-/-* mice were previously described (36). Mouse embryonic fibroblasts (MEFs) derived from these and wild-type C57BL/6 mice were cultured and spontaneously immortalized by serial passaging. The immortalized *p47<sup>phox</sup>-/-* MEFs were transduced with the wild-type or mutant *p47<sup>phox</sup>* constructs. *P53<sup>-/-</sup>* (37) and *INK4a/Arf<sup>-/-</sup>* (38) MEFs were previously described. Primary and spontaneously immortalized MEFs were cultured in DMEM supplemented with 10% FBS and antibiotics. The breast cancer cell lines MDA-MB-231, and T-47D (ATCC) were cultured in RPMI 1640 supplemented with 10% FBS and antibiotics. Melanoma cell lines SKMEL2, LoxMVII, and UACC62 were cultured in DMEM supplemented with 10% FBS and antibiotics.

**Quantitative RT-PCR.** qRT-PCR was used to measure the levels of expression of the microRNAs miR34a, miR34b, and miR34c and the SOD isoforms, SOD1, SOD2, and SOD3. For the measurement of the microRNA levels, we used miRCURY LNA Universal RT and LNA PCR primer sets for miR-34a, miR-34b, miR-34c (204486, 204005, and 204407 respectively, Exiqon). For the measurement of the SOD levels, we used the following primer pairs: CCAGCATGGTCCACGTCCAT and CGCCGGCCACCATGTTTCTT (SOD1), TCGCTACAGATTGCTGCTGCTCT and AAGCGTGCTCCACACGTCA (SOD2), and CTCTAGCTGGGTGCTGCCCTGA and GCCGCCAGTAGCAAGCCGTAG (SOD3). Gene expression levels were normalized against U6, and b-actin and GAPDH for microRNAs and SODs, respectively.

**Intracellular Levels of ROS, Protein and Glutathione Oxidation, and Antioxidant Activity.** Cells were treated with 5-(and-6)-carboxy-2',7'-dichlorofluorescein diacetate (carboxy-DCFDA) (10  $\mu$ M), dihydroethidium (DHE) (5  $\mu$ M), or Mitotracker Red CM-H2Xros (MitoROS) (Invitrogen) (2  $\mu$ M), in DMEM without phenol red. ROS-induced fluorescence in these cells was measured by flow cytometry. Global protein oxidation, reduced glutathione/glutathione disulfide (GSH/GSSG) ratios, and total cell antioxidant activity were measured with the OxyBlot Protein Oxidation Detection Kit (S7150), Glutathione Assay Kit (CS0260), and Antioxidant Assay Kit (CS0790, Sigma), respectively, according to the instructions of the manufacturer.

**Single-Cell Gel Electrophoresis Assay.** DNA damage was measured using the CometAssay (Trevigen), as previously described (39). Akt1- and Akt3-expressing lung fibroblasts were suspended in molten low-melting-point agarose and spread onto the CometSlide. Cells were lysed, and following alkaline gel electrophoresis, they were stained with SyBrGreen I. Fluorescent SyBrGreen I-bound DNA was photographed using a Nikon Eclipse 80i microscope with a 20x objective and a spot charge-coupled device camera (Diagnostic Instruments). To measure DNA damage, the comet tail intensity and length were quantified using Comet Assay IV software (Perceptive Instruments).

**TUNEL Assay.** Apoptosis was measured using a terminal deoxynucleotidyl-transferase biotin-dUTP nick end labeling (TUNEL) kit (Roche) (39).

**Antibodies, Immunoprecipitation, and Western Blotting.** Akt1-, Akt2-, and Akt3-specific antibodies, as well as antibodies against p53, flag-tag, phosphorylated Akt substrate (RXXS/T), and phosphorylated histone H2AX on Ser139, were purchased from Cell Signaling Technology. Cells were lysed using radioimmunoprecipitation assay (RIPA) cell lysis buffer (Cell Signaling Technology) supplemented with protease and phosphatase inhibitors (Roche). Immunoprecipitation (IP) assays were performed with mouse DYKDDDDK Tag (9A3) antibody and Protein G magnetic beads and Western blots of electrophoresed immunoprecipitates were probed with rabbit Phospho-Akt substrate and TrueBlot secondary antibodies (Rockland, Inc.). Western blots of electrophoresed cell lysates were probed with the indicated antibodies, following standard procedures.

## Results

**Triple Akt Knockout Lung Fibroblasts, Rescued by Akt3, Grow Slower than the Akt1- or Akt2-Rescued Cells.** Lung fibroblasts and kidney-derived cells from *Akt1<sup>fl/fl</sup>/Akt2<sup>-/-</sup>/Akt3<sup>-/-</sup>* mice were spontaneously immortalized, as previously described (32, 33). The immortalized cells were transduced with retroviral constructs of Akt1, Akt2, or Akt3, or with the empty retroviral vector. Knocking out the floxed Akt1 allele in these cells with Cre, gave rise to otherwise identical cell lines that express one Akt isoform at a time, or they are *Akt*-null. The physiological relevance of exogenous Akt expression was confirmed by experiments comparing the rate of proliferation of Akt1-rescued cells with the rate of proliferation of *Akt1<sup>fl/fl</sup>/Akt2<sup>-/-</sup>/Akt3<sup>-/-</sup>* and *Akt*-null cells. These experiments showed that cells expressing endogenous and exogenous Akt1 exhibit similar rates of proliferation (*SI Appendix*, Fig. S2). Monitoring the expression of the three Akt isoforms in serially passaged cells by Western blotting revealed that although all Akt isoforms were expressed well in early passage cells, Akt3 expression decreased progressively in subsequent passages (Fig. 1A). This observation suggested that cells expressing high Akt3 levels may proliferate slower and may therefore be counterselected over time. To address this hypothesis, Akt1-, Akt2-, Akt3-, and Akt1/2/3-expressing cells were plated at equal densities and their proliferation rates were monitored by counting them daily for 3 d. The results confirmed that the proliferation of Akt3-expressing cells lagged behind the proliferation of cells expressing Akt1, Akt2, or Akt1/2/3 (Fig. 1B). The slow proliferation of Akt3-rescued cells did not depend on the cell type, as it was also observed in Akt3-rescued kidney-derived cell lines (*SI Appendix*, Fig. S3).

**Akt3 Expression Activates the DNA Damage Response.** Our earlier studies showing that spontaneously immortalized lung fibroblasts from triple *Akt* knockout mice engineered to express Akt1, Akt2, or Akt3, exhibit dramatic differences in microRNA profiles (33), provided a potential explanation for the inhibitory effects of Akt3 on cell proliferation and survival. One microRNA family that is induced upon IGF1 stimulation in Akt3-, and to a lesser extent in Akt2-expressing cells is the miR-34 family (Fig. 2A). The real time RT-PCR data in Fig. 2B confirmed the results of the microarray analysis. Given that the miR-34 microRNA family is a direct transcriptional target of p53 (40), we examined the expression of p53 in these cells. The results in Fig. 2C revealed that p53 is also up-regulated by Akt3.

Given that p53 and the miR-34 microRNAs inhibit cell proliferation (40–42), we hypothesized that the inhibitory effects of Akt3 on cell proliferation may be mediated by p53 and its

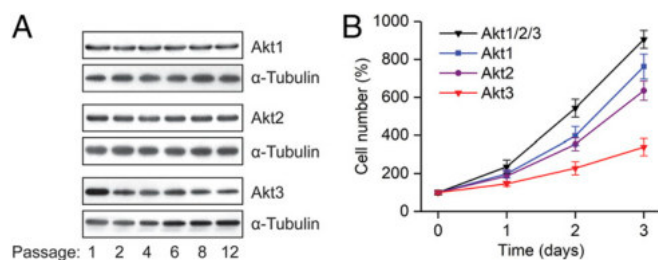
transcriptional target, miR-34. The data in Fig. 2D and E show that high Akt3 expression coincides with high levels of p53 and low cell proliferation rates and they support the hypothesis that p53 and miR-34 are the mediators of the Akt3-induced inhibition of cellular proliferation.

p53 is induced by DNA damage (43, 44). The high levels of p53, in Akt3-expressing cells, therefore, suggested that Akt3 expression may promote DNA damage and the activation of the DNA damage response. To address this question, we examined the phosphorylation of histone variant H2A.X ( $\gamma$ -H2A.X) (45, 46) in passaged Akt3-rescued cells expressing progressively lower levels of Akt3 and p53. The results showed that the phosphorylation of H2A.X, an established marker of DNA damage, decreases with each passage, in parallel with the Akt3 and p53 (Fig. 2F). In agreement with this observation, DNA damage was more pronounced in Akt3-expressing cells, as determined by the Comet assay. The same cells were also characterized by a higher apoptotic index, as determined by the TUNEL assay (*SI Appendix*, Fig. S4). These data suggest that Akt3 expression is indeed associated with DNA damage and with the activation of the DNA damage response.

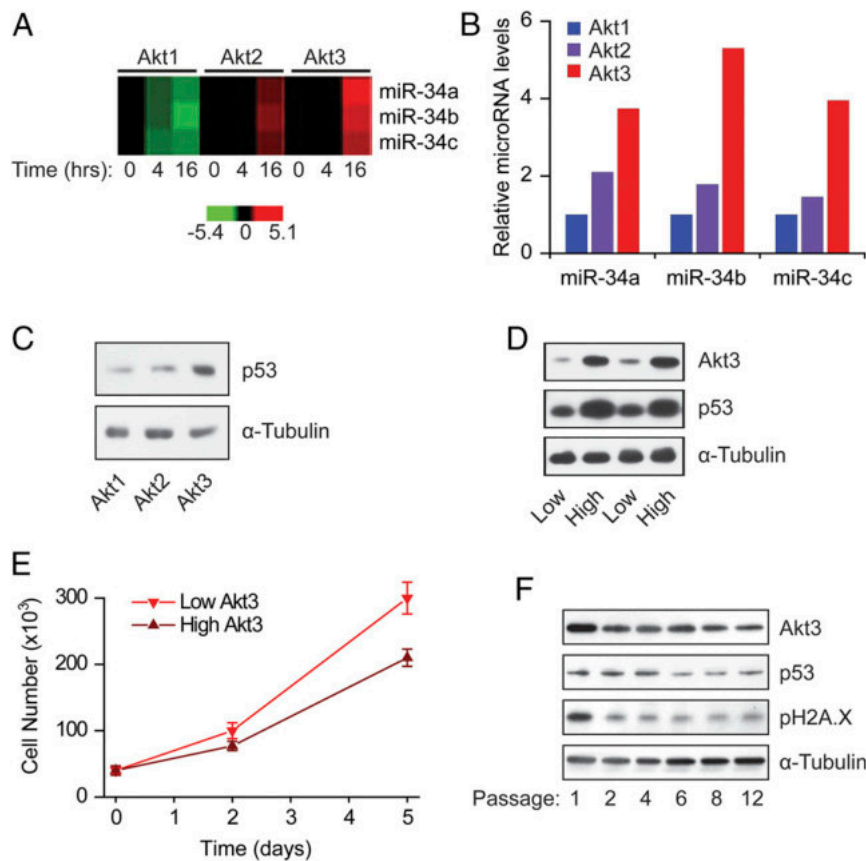
**Akt Isoforms Differentially Regulate the Generation of ROS.** Given that Akt-transduced signals have been linked to the up-regulation of the cellular levels of ROS (44, 47–51), we hypothesized that the DNA damage and the induction of p53 and miR-34 in Akt3-expressing cells may be caused by reactive oxygen species induced differentially by Akt isoforms. To address this hypothesis, the triple *Akt* knockout lung fibroblasts and their derivatives expressing Akt1, Akt2, or Akt3, were loaded with DCFDA or DHE and analyzed by flow cytometry. DCFDA fluoresces upon interaction with a variety of reactive oxygen species (primarily H<sub>2</sub>O<sub>2</sub>), while DHE exhibits significant specificity for superoxide (52). The results in Fig. 3A show that the highest levels of DCFDA-specific ROS (H<sub>2</sub>O<sub>2</sub>) were detected in Akt3-expressing cells and the lowest in Akt2-expressing cells. The Akt1-expressing cells harbored intermediate levels of H<sub>2</sub>O<sub>2</sub>. The results in Fig. 3B show that the highest levels of DHE-specific ROS (superoxide) were detected again in Akt3-expressing cells. Akt2-expressing cells, however, which harbored low levels of H<sub>2</sub>O<sub>2</sub> (Fig. 3A), had increased superoxide levels, slightly lower than those detected in Akt3-expressing cells. The levels of superoxide in Akt1-expressing cells were low, relative to its levels in Akt2- and Akt3-expressing cells. Fig. 3C shows that platelet-derived growth factor (PDGF) and serum stimulation of serum-starved cells induce DCFDA-detectable ROS (H<sub>2</sub>O<sub>2</sub>). In agreement with data in Fig. 3A, H<sub>2</sub>O<sub>2</sub> induction was again highest in Akt3- and lowest in Akt2-expressing cells.

The relationship of DCFDA- and DHE-detectable ROS (H<sub>2</sub>O<sub>2</sub> and superoxide, respectively) in Akt-expressing cells may be due to differences in the level and activity of SOD, the enzyme that converts superoxide into H<sub>2</sub>O<sub>2</sub>. There are three isoforms of SOD: SOD1 or Cu/ZnSOD, which is localized in the nucleus and the cytoplasm; SOD2, or MnSOD, which is localized in the mitochondria; and SOD3, which is membrane bound (53). Based on the preceding data, we hypothesized that the expression of SODs may be lower in Akt2-rescued cells. Quantitative RT-PCR confirmed the prediction by showing that Akt2-rescued cells indeed express lower levels of the two main isoforms of this enzyme, SOD1 and SOD2 (Fig. 3D).

Reactive oxygen species modulate the cellular phenotype in part by oxidizing glutathione and cellular proteins. We therefore examined the GSH/GSSG ratio and the abundance of oxidized proteins in cellular extracts of triple *Akt* knockout lung fibroblasts and their derivatives expressing Akt1, Akt2, or Akt3. To determine the GSH/GSSG ratio, we measured the levels of reduced and oxidized glutathione in deproteinized extracts derived from equal numbers of these cells. The results showed that, following serum starvation for 6 h, the ratio was similar in all of the cell lines.



**Fig. 1.** Triple *Akt* knockout lung fibroblasts engineered to express Akt3 proliferate more slowly than isogenic lung fibroblast lines expressing Akt1 or Akt2. (A) Whereas the expression of Akt1 and Akt2 remains stable, Akt3 expression declines with cell passaging. Western blots of cell lysates of passaged *Akt*-null lung fibroblasts engineered to express Akt1, Akt2, or Akt3 were probed with Akt isoform-specific antibodies. Tubulin was used as the loading control. (B) Growth curves of the same cells show that the Akt3-expressing cells grow more slowly, while Akt2-expressing cells grow at an intermediate rate. Akt1/2/3: *Akt* null cells engineered to express all three isoforms.



**Fig. 2.** Akt3 expression activates the DNA damage response. (A) Heatmap showing that treatment with IGF1 induces miR-34a, miR-34b, and miR-34c in Akt3-expressing, and to a lesser extent, in Akt2-expressing cells. (B) Quantitative RT-PCR comparing the expression of miR-34a, miR-34b, and miR-34c in Akt1-, Akt2-, and Akt3-expressing cells growing in serum-containing media. (C) Western blot showing the expression of p53 in cell lysates of Akt1-, Akt2-, and Akt3-expressing cells growing in serum-supplemented media. Tubulin was used as the loading control. (D) Western blot showing the expression of p53 in independent cultures of lung fibroblasts expressing high and low levels of Akt3 (two independent cultures of each). Tubulin was used as the loading control. (E) Growth curves of cells expressing high and low levels of Akt3. (F) Serial passage of Akt3-expressing cells selects for cells expressing progressively lower levels of Akt3. Western blots show that Akt3 expression correlates with the expression of p53 and with the level of H2A.X phosphorylation on Ser139. Tubulin was used as the loading control.

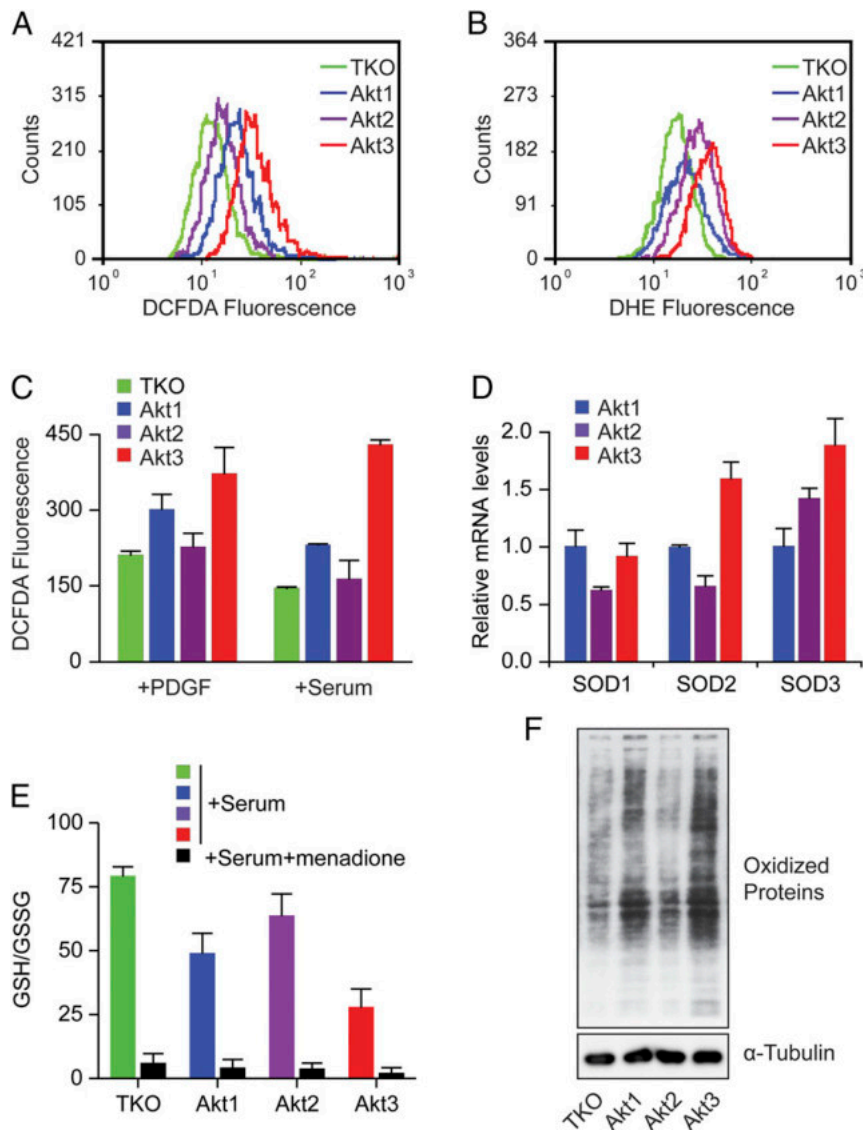
However, in cells growing in serum-supplemented media, the GSH/GSSG ratio was the lowest in cells expressing Akt3 (*SI Appendix, Fig. S5*), as expected. In agreement with these data, the abundance of oxidized proteins, detected by probing immunoblots of whole cell extracts with an antibody that recognizes carbonyl groups in cellular proteins, was also the highest in the Akt3-, followed by the Akt1-rescued cells (Fig. 3F).

**The Activation of the DNA Damage Response in Akt3-Rescued Cells Is ROS Dependent.** The preceding data (Figs. 1–3) suggested that the slow growth of Akt3-expressing cells is due to the accumulation of ROS, which induces DNA damage and consequently activates the p53 response pathway. To address this hypothesis, we examined the effects of the ROS scavenger NAC on the proliferation of Akt1-, Akt2-, and Akt3-expressing cells. The results revealed that NAC stimulates cellular proliferation in the Akt3-rescued, but not in the Akt1- and Akt2-rescued cells (Fig. 4A, B, and C). Moreover, NAC suppresses the expression of p53 (Fig. 4D) and miR-34 (Fig. 4E) in Akt3-expressing cells. These data support the proposed hypothesis.

**The Akt3-Mediated Suppression of Cellular Proliferation via ROS Is p53 and INK4a/Arf Dependent.** To determine whether the induction of p53 by ROS in Akt3-expressing cells is required for the Akt3-mediated inhibition of cellular proliferation, wild-type MEFs and *p53*<sup>-/-</sup> MEFs were transduced with retroviral constructs of

Akt1 or Akt3. Western blot analysis verified that Akt3 overexpression induces the expression of p53 (*SI Appendix, Fig. S6*), while cell growth assays revealed that Akt3 overexpression inhibits the proliferation of wild-type MEFs (Fig. 5A), but does not affect the proliferation of *p53*<sup>-/-</sup> cells (Fig. 5B). Measurements of DHE- and DCFDA-detectable ROS (superoxide and H<sub>2</sub>O<sub>2</sub>, respectively) by flow cytometry confirmed that both are selectively induced by Akt3 in both wild-type and *p53*<sup>-/-</sup> cells (Fig. 5C and D). Based on these data we conclude that the inhibition of cellular proliferation by Akt3-induced ROS is indeed p53 dependent. To investigate the involvement of p53 in miR-34 regulation downstream of Akt3, we examined the effects of Akt3 overexpression on the levels of miR-34 family members in wild-type and *p53*<sup>-/-</sup> MEFs. qRT-PCR analysis revealed that the induction of miR-34a, b and c, depends at least in part on p53 (*SI Appendix, Fig. S7*). To assess the role of miR-34 in Akt3-mediated suppression of cell proliferation, we inhibited miR-34 activity by antisense oligos. Inhibition of miR-34 increased the rate of proliferation of Akt3-rescued cells by 15% with no effects on Akt1-expressing cells (*SI Appendix, Fig. S8*), suggesting that miR-34 plays a minor role, relative to other p53-dependent mechanisms.

In addition to p53, DNA damage activates the *INK4a/ARF* locus, which encodes the p16<sup>INK4a</sup>, p15<sup>INK4b</sup>, and p19<sup>Arf</sup> proteins (p14<sup>Arf</sup> in humans) (54, 55). The cyclin-dependent kinase inhibitors p16<sup>INK4a</sup> and p15<sup>INK4b</sup> regulate the phosphorylation of the retinoblastoma (Rb) family members, in cells traversing the

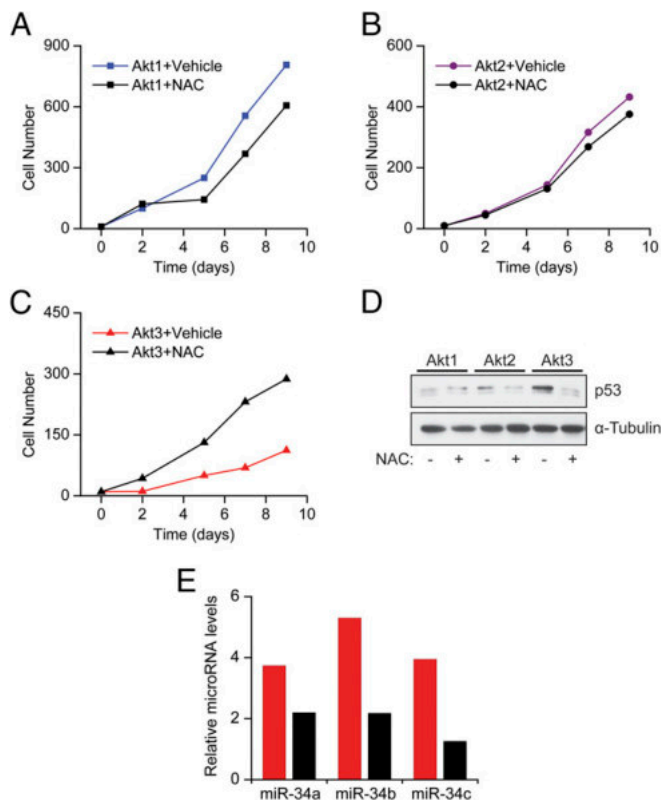


**Fig. 3.** ROS are differentially regulated by the three Akt isoforms. (A) Flow cytometry of TKO, Akt1-, Akt2-, and Akt3-expressing cells growing in serum-supplemented media and treated with DCFDA. (B) Flow cytometry of TKO, Akt1-, Akt2-, and Akt3-expressing cells growing in complete serum-supplemented media and treated with DHE. (C) DCFDA-detectable fluorescence in TKO, Akt1-, Akt2-, and Akt3-expressing cells, 10 min after treatment with PDGF or serum. (D) Expression of SOD1, SOD2, and SOD3 was determined by qRT-PCR in Akt1-, Akt2-, and Akt3-expressing cells growing in complete serum-supplemented media. (E) Ratio of reduced-to-oxidized glutathione (GSH/GSSG) in TKO, Akt1-, Akt2-, and Akt3-expressing cells growing in complete serum-supplemented media was determined by spectrophotometric (412 nm) measurement of 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) reduction to 2-nitro-5-thiobenzoic acid (TNB) in deproteinized cell lysates and are expressed in nanomoles per milliliter (nmol/mL). Cells pretreated with oxidative menadione (40  $\mu$ M) to increase GSSG, were used as positive controls. (F) Western blot of cell lysates from TKO, Akt1-, Akt2-, and Akt3-expressing lung fibroblasts, probed with an antibody recognizing carbonyl groups on protein side chains. Such groups reflect the oxidation status of proteins. Tubulin was used as the loading control. TKO: Akt-null (triple knockout) cells.

G1 phase of the cell cycle, while p19<sup>Arf</sup> regulates the abundance of p53. Based on these considerations, we examined whether the Akt3-mediated inhibition of cell proliferation depends also on the INK4/ARF locus. To this end, MEFs in which both INK4 and ARF were ablated (*INK4a*<sup>-/-</sup>/*Arf*<sup>-/-</sup>), and MEFs in which ARF was selectively ablated (*Arf*<sup>-/-</sup>), were transduced with retroviral constructs of Akt1 or Akt3. The results confirmed that the expression of Akt3 does not suppress the proliferation of these cells (Fig. 5 E and F). We conclude that the inhibition of cellular proliferation by Akt3 depends not only on p53, but also on the *INK4a/ARF* locus. Collectively, our data indicate that the ROS-mediated inhibition of cellular proliferation in Akt3-expressing cells depends on the activation of the DNA damage response.

**Akt3 Inhibits Cell Proliferation by Promoting ROS Production via Phosphorylation of p47<sup>phox</sup> and Activation of the NADPH Oxidase.** The levels of ROS in cells expressing different Akt isoforms may be due to the differential regulation of ROS production or ROS detoxification by Akt1, Akt2, and Akt3 (SI Appendix, Fig. S1). To address this question, we employed an assay that measures the overall antioxidant activity in cell lysates. The results (SI Appendix, Fig. S9) revealed that Akt3-expressing cells exhibit high, rather than low antioxidant activity, suggesting that Akt3 may up-regulate ROS by promoting ROS production rather than by inhibiting ROS detoxification and that feedback mechanisms up-regulate antioxidant activity to detoxify the increased levels of ROS.

Akt3 may stimulate ROS production by one of two mechanisms: 1) It may stimulate mitochondrial biogenesis and mitochondrial



**Fig. 4.** The slow proliferation of Akt3-expressing cells is ROS dependent. (A–C) Growth curves of Akt1-, Akt2-, and Akt3-expressing cells before and after treatment with the antioxidant NAC. The antioxidant was renewed every 3 d. (D) Western blot showing the expression of p53 in Akt1-, Akt2-, and Akt3-expressing cells before and after treatment with NAC. Tubulin was used as the loading control. (E) Quantitative RT-PCR comparing the expression of miR-34a, miR-34b, and miR-34c in Akt3-expressing cells before and after treatment with NAC.

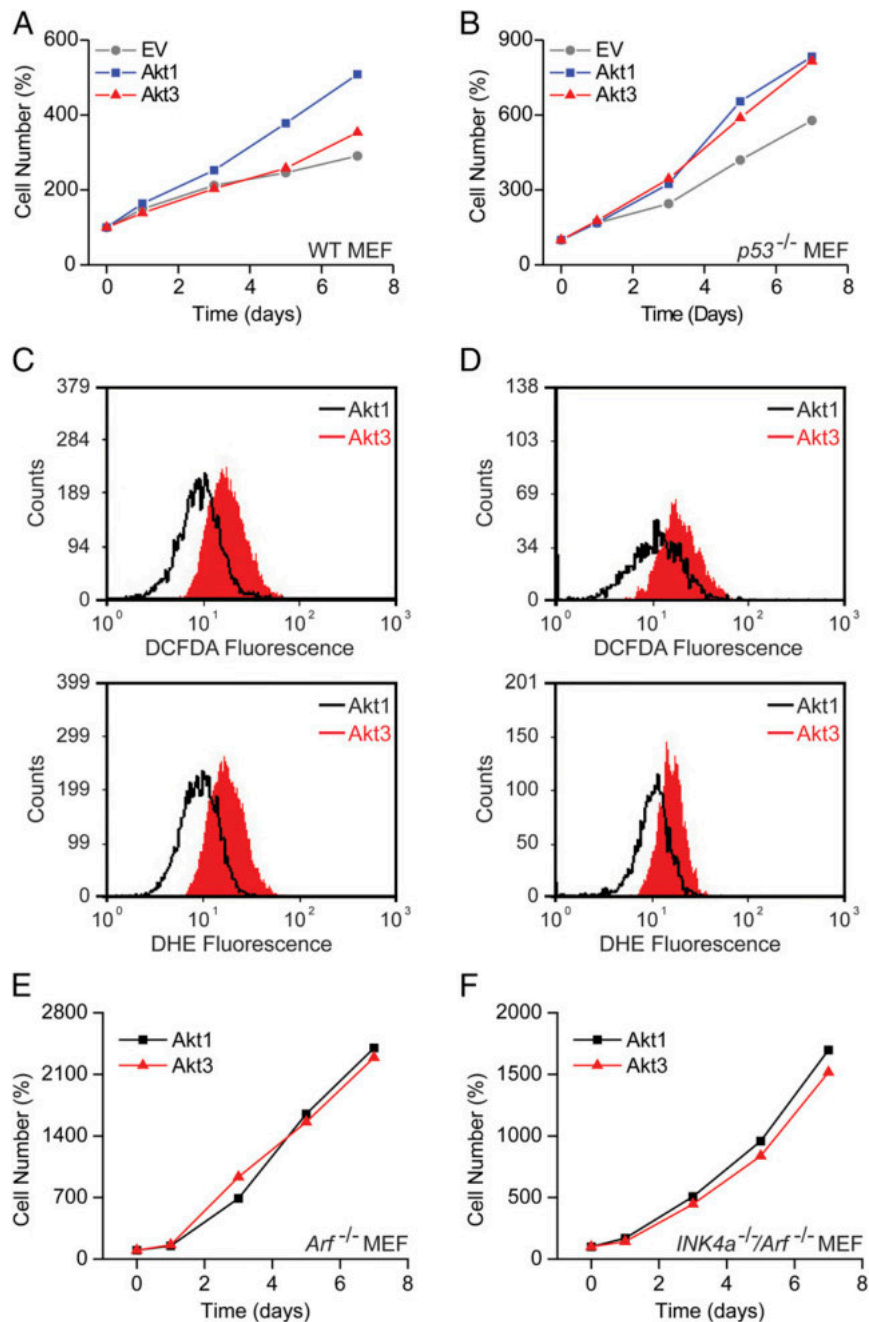
oxidative phosphorylation. This mechanism appears to play an important role in Akt3-mediated ROS production in VEGF-stimulated endothelial cells (56). 2) It may activate the NADPH oxidase by phosphorylating components of the enzyme complex. Phosphorylation of p47<sup>phox</sup>, one of the regulatory components of the NADPH oxidase, at Ser304 and Ser328, had indeed been observed in earlier studies addressing the effects of Akt on the activity of purified NADPH oxidase *in vitro* (51, 57, 58). Experiments using the ROS-activated fluorescent dye mitotracker, which monitors ROS levels in the mitochondria, revealed that all Akt isoforms up-regulate mitochondrial ROS with similar efficiencies (SI Appendix, Fig. S10). Moreover, parallel experiments revealed that the induction of DCFDA-detectable ROS (H<sub>2</sub>O<sub>2</sub>), which is more robust in Akt3-expressing cells, is completely blocked by the NADPH oxidase inhibitors 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF) and diphenyleioidonium (DPI) (SI Appendix, Figs. S11 and S12). These observations combined suggested that ROS are produced in response to growth factor signals that are transduced by Akt (primarily Akt3), and activate the NADPH oxidase. To test this hypothesis, we expressed the Flag-tagged p47<sup>phox</sup> in lung fibroblasts expressing one Akt isoform at a time. Cell lysates were immunoprecipitated by anti-Flag (p47<sup>phox</sup>), and Western blots were probed with the phosphorylated Akt substrate (RXXS\*/T\*) antibody. This approach demonstrated that Akt3 phosphorylates p47<sup>phox</sup> more efficiently than Akt1 or Akt2 (Fig. 6A). Importantly, in Akt3-expressing cells, p47<sup>phox</sup> phosphorylation was lost by treatment with the Akt inhibitor MK2206 (SI Appendix, Fig. S13).

To link p47<sup>phox</sup> phosphorylation and NADPH activity, spontaneously immortalized p47<sup>phox</sup><sup>-/-</sup> MEFs were transduced with Akt1 or Akt3 retroviral constructs. Transduced cells were subsequently compared to determine whether they differ in p53 expression, cell proliferation rates, and in DCFDA- and DHE-detectable ROS levels. The results showed that both the rates of proliferation (Fig. 6B) and p53 protein levels (Fig. 6C) were similar in Akt1- and Akt3-expressing p47<sup>phox</sup><sup>-/-</sup> MEFs, and accordingly, the ability of Akt3 to induce miR-34 was impaired (SI Appendix, Fig. S14). Importantly, the differences between Akt1- and Akt3-induced ROS observed in wild-type MEFs (Fig. 5C) were diminished in p47<sup>phox</sup><sup>-/-</sup> MEFs (Fig. 6D and E). We conclude that Akt3 induces ROS and subsequently p53 expression by activating the NADPH oxidase.

To determine whether NADPH oxidase is preferentially regulated by Akt3, we transduced p47<sup>phox</sup><sup>-/-</sup> MEFs with wild-type p47<sup>phox</sup> and retroviral constructs of Akt1 or Akt3. Control p47<sup>phox</sup>-rescued p47<sup>phox</sup><sup>-/-</sup> cells were transduced with the empty retroviral vector (pBabe puro). DHE (superoxide) fluorescence monitored by flow cytometry was significantly more robust in cells transduced with Akt3 than in cells transduced with Akt1 or with the empty vector (Fig. 6F). To determine whether Akt3 enhances ROS production by phosphorylating p47<sup>phox</sup>, we mutated both phosphorylation motifs of p47<sup>phox</sup> (Fig. 6G). Spontaneously immortalized MEFs from p47<sup>phox</sup><sup>-/-</sup> mice were transduced with retroviral constructs of Flag-tagged wild-type p47<sup>phox</sup> or the Flag-tagged phosphorylation site mutants [p47<sup>phox</sup>S304A, p47<sup>phox</sup>S328A, or p47<sup>phox</sup>S304A/S328A (DMp47phox)]. The same cells were superinfected with retroviral constructs of Akt1 or Akt3. Transduced cells were analyzed for p47<sup>phox</sup> phosphorylation by probing anti-Flag immunoprecipitates with the Akt phosphosubstrate antibody. The same cells were treated with DHE and they were analyzed by flow cytometry. The results confirmed that the Ser304 and Ser328 sites were both phosphorylated by Akt (primarily Akt3) *in vivo*. Notably, although the phosphorylation of the single mutants is higher in the Akt3 than in the Akt1-expressing cells, the background phosphorylation of the double mutant by Akt1 and Akt3 is equal (Fig. 6H). They also confirmed that, whereas Akt3 significantly up-regulates superoxide in p47<sup>phox</sup><sup>-/-</sup> MEFs rescued with the wild-type p47<sup>phox</sup>, neither Akt3 nor Akt1 up-regulates superoxide in MEFs rescued with the phosphorylation site mutant of p47<sup>phox</sup> (Fig. 6I). We conclude that ROS induction by Akt3 depends on the NADPH oxidase which is preferentially activated by Akt3 via phosphorylation of p47<sup>phox</sup> at Ser304 and Ser328.

**Akt3 Regulates the Expression of p53, and Cancer Cells Adapt to the Akt3-p53 Axis.** The link between Akt3 and p53 was also observed in human cancer cell lines. Comparison of the p53 levels in three melanoma cell lines, of which one (SKMEL2) expresses high levels of Akt3 and two (LoxMVII and UACC62) express low levels of Akt3, revealed that the levels of Akt3 correlate with the levels of p53 (SI Appendix, Fig. S15A). More important, knockdown of Akt3 in SKMEL2 cells suppressed p53 levels (SI Appendix, Fig. S15B), suggesting that p53 expression is under the control of Akt3. The above observation was validated in breast cancer cell lines expressing all Akt isoforms (T-47D, MDA-MB-231). Whereas, silencing of Akt3 in both cell lines resulted in suppression of p53, knockdown of Akt1 or Akt2 had minimal effects (Fig. 7A and B).

Akt3 is highly expressed in several tumor types (68–70). Interrogation of cancer genomics datasets (TCGA: The Cancer Genome Atlas) through the cBioportal website (71, 72), revealed that the *Akt3* gene is genetically altered, most commonly amplified, in numerous types of human cancer (SI Appendix, Fig. S16). Given the antiproliferative effects of Akt3, the selection of tumor cells with an amplified and/or overexpressed *Akt3* gene is a paradox. To address this paradox, we hypothesized that cells adapt to the overexpression of Akt3 by inactivating the DNA damage response pathway. To address this hypothesis, we employed data



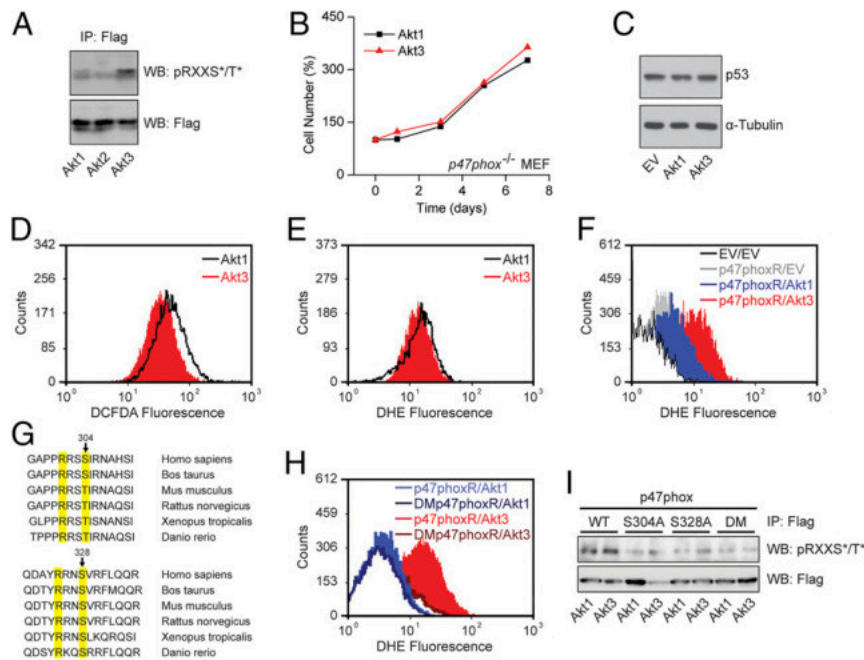
**Fig. 5.** The growth inhibitory effects of Akt3 are p53 and INK4A/ARF dependent. (A and B) Growth curves of wild-type (A) and  $p53^{-/-}$  (B) MEFs transfected with Akt1, Akt3, or the empty vector, and growing in complete serum-supplemented media. (C and D) DCFDA- and DHE-detectable ROS were measured by flow cytometry in the Akt1- and Akt3-expressing wild-type (C) and  $p53^{-/-}$  (D) MEFs. (E and F) Growth curves of  $Arf^{-/-}$  (E) or  $INK4a^{-/-}/Arf^{-/-}$  (F) MEFs, expressing Akt1 or Akt3 and growing in serum-supplemented media.

extracted from the Oncomine database to examine the correlation between Akt3 and p53 expression. These analyses revealed that in three types of human cancer [melanomas (Fig. 7C), breast (Fig. 7D), and lung carcinomas (Fig. 7E)], the expression of Akt3 exhibits an excellent positive correlation with the expression of p53. Moreover, the frequency of p53 mutation is significantly higher in tumors with Akt3 amplification than in tumors without Akt3 alterations (Fig. 7F).

### Discussion

Data presented in this report address the role of individual Akt isoforms in the generation of ROS. The initial experiments were

carried out in immortalized lung fibroblasts, which were engineered to express one Akt isoform at a time, but were otherwise identical (30–33). A screen comparing these cells with each other and with isogenic Akt-null cells for the abundance of ROS showed that although all Akt isoforms contribute to ROS generation, it is Akt3 that is the most robust ROS inducer. These studies were initiated because of the observation that Akt3-expressing cells grow significantly slower than Akt1- or Akt2-expressing isogenic cells and their slow growth is associated with high levels of p53 and high levels of the microRNAs miR-34a, miR-34b, and miR-34c, which are direct transcriptional targets of p53. The elevation of p53 was due to the activation of



**Fig. 6.** The slow proliferation of Akt3-expressing cells is caused by ROS, induced by Akt3, via  $p47^{phox}$  phosphorylation, and activation of the NADPH oxidase. (A) Phosphorylation of  $p47^{phox}$  in Akt1-, Akt2-, and Akt3-expressing cells growing in serum-supplemented media. Akt1-, Akt2-, and Akt3-expressing lung fibroblasts were transduced with wild-type Flag- $p47^{phox}$ .  $p47^{phox}$  was immunoprecipitated from cell lysates with an anti-Flag antibody. A Western blot of the immunoprecipitates was probed with an Akt phosphosubstrate antibody (RXXS\*/T\*). The same immunoprecipitates were probed with the anti-Flag antibody (loading control). (B) Growth curves of  $p47^{phox-/-}$  MEFs expressing Akt1 or Akt3 and growing in complete serum-supplemented media. (C) Western blot showing the expression of p53 in cell lysates of Akt1- and Akt3-expressing  $p47^{phox-/-}$  MEFs growing in serum-supplemented media. Tubulin was used as the loading control. (D and E) DCFDA- and DHE-detectable ROS in the Akt1- and Akt3-expressing  $p47^{phox-/-}$  MEFs were measured by flow cytometry. (F) DHE-detectable ROS levels were measured by flow cytometry, in  $p47^{phox-/-}$  MEFs and their derivatives, transduced with the empty vector (pBabe-neo) or with wild-type  $p47^{phox}$  ( $p47^{phoxR}$ ). The wild-type  $p47^{phox}$ -rescued cells were also transduced with Akt1 or Akt3 retroviral constructs or with the empty vector (EV). (G) Conservation of the Akt phosphorylation motifs, RXXS/T, on  $p47^{phox}$  (Ser304 and Ser328). (H) Phosphorylation of wild type and  $p47^{phox}$  mutants in Akt1- and Akt3-expressing cells growing in serum-supplemented media. Akt1- and Akt3-expressing lung fibroblasts were transduced with wild-type Flag- $p47^{phox}$  (WT) or its mutants Flag- $p47^{phoxS304A}$ , Flag- $p47^{phoxS328A}$ , or Flag- $p47^{phoxS304A/S328A}$  double mutant (DM). Cell lysates were immunoprecipitated with the anti-Flag antibody ( $p47^{phox}$ ), and Western blots of the immunoprecipitates were probed with the Akt phosphosubstrate antibody (RXXS\*/T\*). Probing immunoprecipitates with anti-Flag antibody were used as the loading control. (I) DHE-detectable ROS, measured by flow cytometry, in  $p47^{phox-/-}$  MEFs and their derivatives, transduced with the wild-type  $p47^{phox}$  ( $p47^{phoxR}$ ) or the double phosphorylation site mutant of  $p47^{phox}$  (DM $p47^{phoxR}$ ). Both the wild type and the mutant  $p47^{phox}$ -rescued cells were also transduced with Akt1 or Akt3 retroviral constructs. EV/EV, cells transduced with both empty vectors; R, rescued.

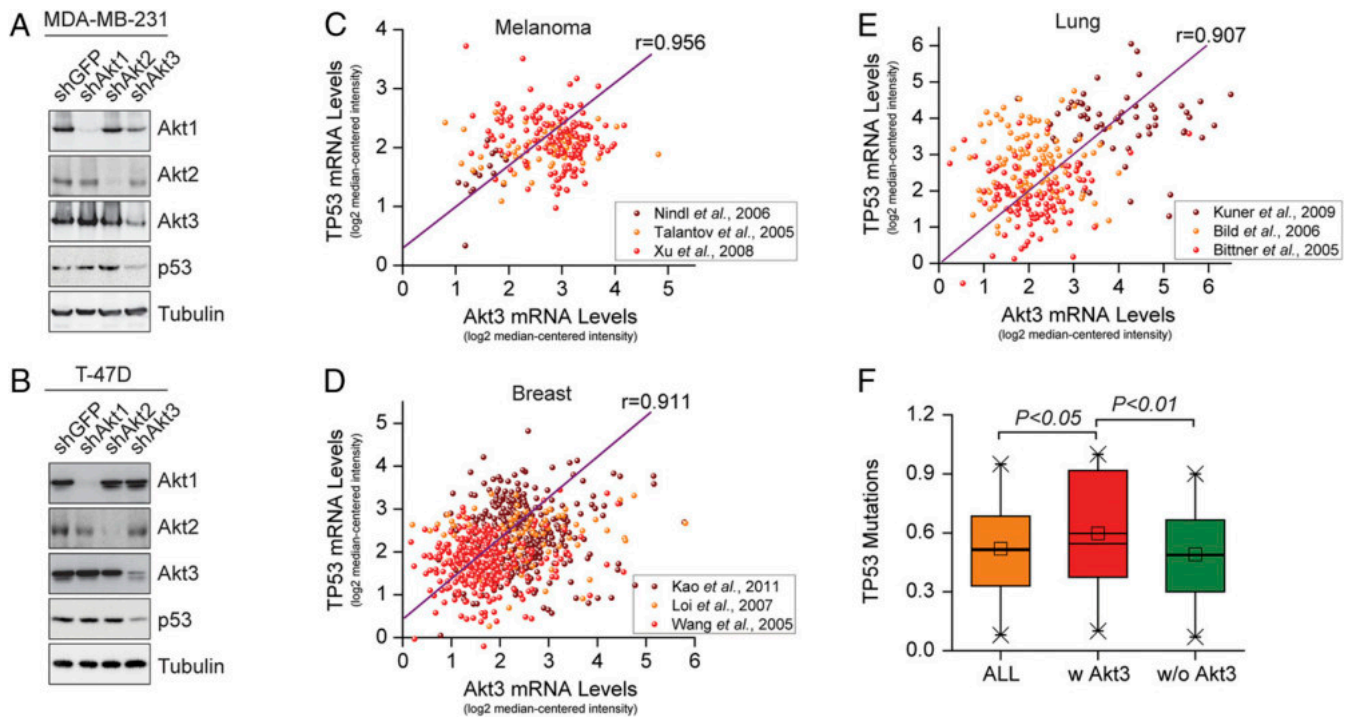
the DNA damage response by reactive oxygen species, whose levels were significantly higher in the Akt3-expressing cells. The importance of ROS in the activation of the DNA damage response in the Akt3-expressing cells was confirmed both by pharmacologic and genetic experiments. Specifically, we showed that the antioxidant NAC reduced the levels of p53 and promoted cell proliferation selectively in Akt3-expressing cells. Moreover, Akt3 expression failed to inhibit cell proliferation when expressed in  $p53^{-/-}$ ,  $INK4a^{-/-}/Arf^{-/-}$ , and  $Arf^{-/-}$  cells, despite the fact that the loss of these genes had no effect on the up-regulation of ROS by Akt3. Further studies revealed that the robust up-regulation of ROS by Akt3, as opposed to Akt1 and Akt2, is due to differences between Akt isoforms in the generation rather than the clearance of ROS.

The induction of ROS by Akt has been described previously (48–51). However, most of the earlier studies do not distinguish between Akt isoforms and they focus primarily on the role of Akt in the regulation of ROS detoxification mechanisms. Thus, it has been observed that Akt phosphorylates and inactivates the FOXO family of transcription factors, which normally promote the expression of the mitochondrial enzyme MnSOD (SOD2) (73, 74). ROS induction by Akt and other oncogenes also promotes the expression of FOXM1, another member of the forkhead family of transcription factors, which induces the expression of several ROS detoxifying enzymes, including MnSOD, catalase, and PRDX3 (75). Importantly, Akt also stabilizes NRF2 and promotes the

expression of regulators of glutathione synthesis, resulting in glutathione up-regulation (76). These Akt-regulated activities may be important for the overall regulation of ROS by Akt; however, they do not seem to contribute to the differential induction of ROS by Akt3. First, Akt3 is a more robust inducer of ROS than the other Akt isoforms, and it functions by stimulating ROS generation rather than by inhibiting ROS detoxification. Second, the expression of SOD2 and SOD3 is higher in Akt3-expressing, relative to the Akt1- and Akt2-expressing cells, while the expression of SOD1 is equivalent in Akt3- and Akt1-expressing cells. Significantly, SOD1 and SOD2 expression is higher in Akt1- than in Akt2-expressing cells and this correlates with the levels of  $H_2O_2$  and superoxide in the two cell types, with the levels of  $H_2O_2$  (DCFDA-detectable ROS) being higher in Akt1-expressing cells and the levels of superoxide (DHE-detectable ROS) being higher in Akt2-expressing cells. We have not addressed the mechanism of this difference. However, we have observed that Akt2-expressing cells tend to express higher levels of p53 and microRNAs of the miR-34 family and they proliferate slower than the Akt1-expressing cells.

In the experiments presented here, the high levels of ROS in the Akt3-expressing cells were induced by growth factor stimulation, and they were not associated with the mitochondria. More detailed analyses showed that the Akt3-dependent stimulation of ROS production was due to the phosphorylation of the NADPH oxidase subunit  $p47^{phox}$ , which results in NADPH activation. Previous studies had shown that NADPH oxidase can be activated





**Fig. 7.** The expression of Akt3 in multiple human tumor types correlates with the expression of p53. Mechanisms of tumor cell adaptation to Akt3 expression. (A and B) Akt1, Akt2, or Akt3 were knocked down in the high Akt3-expressing breast cancer cell lines MDA-MB-231 (A) and T-47D (B). Western blots of lysates derived from these and control cells (shGFP) were probed with antibodies against Akt1, Akt2, Akt3, p53, and tubulin (loading control). (C–E) Analysis of data obtained from OncoPrint, shows that the abundance of p53 mRNA (probe ID number 201746\_at) correlates with the abundance of Akt3 mRNA (probe ID number 212607\_at) in human melanomas (59–61) (C), breast (62–64) (D), and lung (65–67) (E) carcinomas. (F) Akt3 amplification or mutation in human tumors correlates with the deletion or mutation of p53. Analysis of data was obtained from the cBioportal for Cancer Genomics. Data compiled from 36 different studies with Akt3 alterations ranging from 0.3 to 34% were plotted to show the frequency of p53 genetic alterations in tumors with or without alterations in Akt3 (statistical comparisons with the paired sample *t* test).

via the Akt-mediated phosphorylation of p47<sup>phox</sup> at two conserved sites (51, 57) (Fig. 6). However, these data were obtained from biochemical *in vitro* experiments, which were carried out using purified proteins. In addition, they did not distinguish between Akt isoforms. The experiments presented here provide *in vivo* confirmation of the *in vitro* biochemical data. In addition, they address the roles of the three Akt isoforms and they explore the biological consequences of p47<sup>phox</sup> phosphorylation in live cells. An earlier study focusing on the role of Akt3 in ROS induction in endothelial cells had shown that Akt3 promotes the generation of ROS by stimulating mitochondrial biogenesis (56). The data presented in this report did not directly address the role of Akt3 in mitochondrial biogenesis. However, the earlier study failed to show a disproportionate increase in mitochondrial ROS following Akt3 activation. It is possible that the difference between this and the earlier study showing the effects of Akt3 on mitochondrial biogenesis is in the types of cells used in these studies.

The high expression of p53 in Akt3-expressing cells suppresses cell proliferation. The overexpression of Akt3 in human melanomas and other types of human tumors therefore is puzzling, because it would be expected to inhibit the proliferation of the tumor cells. We therefore hypothesized that these tumors may have undergone genetic or epigenetic changes that allow them to tolerate the high Akt3 levels. These changes may result in the deregulation of the DNA damage response pathways. Data extracted from the OncoPrint and TCGA databases support this conclusion by showing that in several types of cancer, Akt3 levels positively

correlate with the levels of p53 and that p53 mutations are more common in tumors with a genetically altered (most commonly amplified) Akt3, than in tumors with a normal Akt3 gene. Cells with impaired DNA damage response and high levels of ROS would be expected to exhibit significant genetic instability, which may contribute to the aggressiveness of these tumors. Akt3 activity may result in loss of the p53 “brake” but due to high levels of ROS and/or DNA damage maintain high ATM and Rad3-related serine/threonine kinase (ATR)/checkpoint kinase 1 (CHK1) activity (77, 78). Based on the promising data from ongoing clinical trials with ATR/CHK1 inhibitors (79), we propose that these inhibitors could prove effective against tumors with high Akt3 activity.

In summary, although all Akt isoforms promote the up-regulation of ROS, Akt3 is the Akt isoform most efficiently up-regulating ROS. ROS up-regulation by Akt3 is due primarily to the phosphorylation of p47<sup>phox</sup>, which activates the NADPH oxidase. ROS induction induces the DNA damage response and inhibits cell proliferation. Therefore, tumor cells expressing high levels of Akt3 adapt to grow by inactivating the DNA damage response, which allows them to bypass the inhibitory effects of ROS on cell proliferation.

**Data Availability.** All study data are included in the article and supporting information.

**ACKNOWLEDGMENTS.** This work was supported by NIH grant R01 CA186729 to P.N.T. and NTU quality research (QR) funds to C.P. and M.H.

1. R. Radi, Oxygen radicals, nitric oxide, and peroxynitrite: Redox pathways in molecular medicine. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5839–5848 (2018).
2. D. P. Jones, Radical-free biology of oxidative stress. *Am. J. Physiol. Cell Physiol.* **295**, C849–C868 (2008).

3. S. S. Sabharwal, P. T. Schumacker, Mitochondrial ROS in cancer: Initiators, amplifiers or an achilles’ heel? *Nat. Rev. Cancer* **14**, 709–721 (2014).
4. J. Yan, J. Jiang, L. He, L. Chen, Mitochondrial superoxide/hydrogen peroxide: An emerging therapeutic target for metabolic diseases. *Free Radic. Biol. Med.* **152**, 33–42 (2020).

5. A. Panday, M. K. Sahoo, D. Osorio, S. Batra, NADPH oxidases: An overview from structure to innate immunity-associated pathologies. *Cell. Mol. Immunol.* **12**, 5–23 (2015).
6. J. El-Benna *et al.*, Priming of the neutrophil respiratory burst: Role in host defense and inflammation. *Immunol. Rev.* **273**, 180–193 (2016).
7. A. M. Franchini, D. Hunt, J. A. Melendez, J. R. Drake, Fc $\gamma$ R-driven release of IL-6 by macrophages requires NOX2-dependent production of reactive oxygen species. *J. Biol. Chem.* **288**, 25098–25108 (2013).
8. H. Buvelot, V. Jaquet, K. H. Krause, Mammalian NADPH oxidases. *Methods Mol. Biol.* **1982**, 17–36 (2019).
9. I. S. Mudway, F. J. Kelly, S. T. Holgate, Oxidative stress in air pollution research. *Free Radic. Biol. Med.* **151**, 2–6 (2020).
10. R. P. Brandes, N. Weissmann, K. Schröder, Nox family NADPH oxidases: Molecular mechanisms of activation. *Free Radic. Biol. Med.* **76**, 208–226 (2014).
11. Y. Zhang, P. Murugesan, K. Huang, H. Cai, NADPH oxidases and oxidase crosstalk in cardiovascular diseases: Novel therapeutic targets. *Nat. Rev. Cardiol.* **17**, 170–194 (2020).
12. M. Takahashi *et al.*, Protein kinase A-dependent phosphorylation of Rap1 regulates its membrane localization and cell migration. *J. Biol. Chem.* **288**, 27712–27723 (2013).
13. S. J. Forrester, D. S. Kikuchi, M. S. Hernandez, Q. Xu, K. K. Griendling, Reactive oxygen species in metabolic and inflammatory signaling. *Circ. Res.* **122**, 877–902 (2018).
14. D. A. Mogilenko *et al.*, Metabolic and innate immune cues merge into a specific inflammatory response via the UPR. *Cell* **177**, 1201–1216.e19 (2019).
15. G. Martel-Gallegos *et al.*, Oxidative stress induced by P2X7 receptor stimulation in murine macrophages is mediated by c-Src/Pyk2 and ERK1/2. *Biochim. Biophys. Acta* **1830**, 4650–4659 (2013).
16. N. T. Moldogazieva, S. V. Lutsenko, A. A. Terentiev, Reactive oxygen and nitrogen species-induced protein modifications: Implication in carcinogenesis and anticancer therapy. *Cancer Res.* **78**, 6040–6047 (2018).
17. H. Sies, D. P. Jones, Reactive oxygen species (ROS) as pleiotropic physiological signalling agents. *Nat. Rev. Mol. Cell Biol.* **21**, 363–383 (2020).
18. N. Krishnan *et al.*, Harnessing insulin- and leptin-induced oxidation of PTP1B for therapeutic development. *Nat. Commun.* **9**, 283 (2018).
19. Y. Heun *et al.*, Inactivation of the tyrosine phosphatase SHP-2 drives vascular dysfunction in Sepsis. *EBioMedicine* **42**, 120–132 (2019).
20. J. Kwon *et al.*, Reversible oxidation and inactivation of the tumor suppressor PTEN in cells stimulated with peptide growth factors. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16419–16424 (2004).
21. H. Kong, N. S. Chandel, Regulation of redox balance in cancer and T cells. *J. Biol. Chem.* **293**, 7499–7507 (2018).
22. C. Polyarchou, M. HatziaPOSTOLOU, E. Papadimitriou, Hydrogen peroxide stimulates proliferation and migration of human prostate cancer cells through activation of activator protein-1 and up-regulation of the heparin affinin regulatory peptide gene. *J. Biol. Chem.* **280**, 40428–40435 (2005).
23. C. Y. Han *et al.*, NADPH oxidase-derived reactive oxygen species increases expression of monocyte chemotactic factor genes in cultured adipocytes. *J. Biol. Chem.* **287**, 10379–10393 (2012).
24. V. I. Lushchak, Free radicals, reactive oxygen species, oxidative stress and its classification. *Chem. Biol. Interact.* **224**, 164–175 (2014).
25. B. Morgan *et al.*, Multiple glutathione disulfide removal pathways mediate cytosolic redox homeostasis. *Nat. Chem. Biol.* **9**, 119–125 (2013).
26. T. F. Franke *et al.*, The protein kinase encoded by the Akt proto-oncogene is a target of the PDGF-activated phosphatidylinositol 3-kinase. *Cell* **81**, 727–736 (1995).
27. B. D. Manning, L. C. Cantley, AKT/PKB signaling: Navigating downstream. *Cell* **129**, 1261–1274 (2007).
28. A. Bellacosa, J. R. Testa, S. P. Staal, P. N. Tsichlis, A retroviral oncogene, akt, encoding a serine-threonine kinase containing an SH2-like region. *Science* **254**, 274–277 (1991).
29. L. C. Cantley, B. G. Neel, New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4240–4245 (1999).
30. I. Sanidas *et al.*, Phosphoproteomics screen reveals akt isoform-specific signals linking RNA processing to lung cancer. *Mol. Cell* **53**, 577–590 (2014).
31. S. A. Ezell *et al.*, The protein kinase Akt1 regulates the interferon response through phosphorylation of the transcriptional repressor EMSY. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E613–E621 (2012).
32. C. Polyarchou *et al.*, Akt2 regulates all Akt isoforms and promotes resistance to hypoxia through induction of miR-21 upon oxygen deprivation. *Cancer Res.* **71**, 4720–4731 (2011).
33. D. Iliopoulos *et al.*, MicroRNAs differentially regulated by Akt isoforms control EMT and stem cell renewal in cancer cells. *Sci. Signal.* **2**, ra62 (2009).
34. W. S. Chen *et al.*, Growth retardation and increased apoptosis in mice with homozygous disruption of the Akt1 gene. *Genes Dev.* **15**, 2203–2208 (2001).
35. H. Cho, J. L. Thorvaldsen, Q. Chu, F. Feng, M. V. Birnbaum, Akt1/PKBalpha is required for normal growth but dispensable for maintenance of glucose homeostasis in mice. *J. Biol. Chem.* **276**, 38349–38352 (2001).
36. S. H. Jackson, J. I. Gallin, S. M. Holland, The p47phox mouse knock-out model of chronic granulomatous disease. *J. Exp. Med.* **182**, 751–758 (1995).
37. L. A. Donehower *et al.*, Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
38. M. Serrano *et al.*, Role of the INK4a locus in tumor suppression and cell mortality. *Cell* **85**, 27–37 (1996).
39. C. Polyarchou, R. Pfau, M. HatziaPOSTOLOU, P. N. Tsichlis, The JmjC domain histone demethylase Ndy1 regulates redox homeostasis and protects cells from oxidative stress. *Mol. Cell. Biol.* **28**, 7451–7464 (2008).
40. L. He *et al.*, A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130–1134 (2007).
41. A. J. Levine, W. Hu, Z. Feng, The P53 pathway: What questions remain to be explored? *Cell Death Differ.* **13**, 1027–1036 (2006).
42. L. J. Ko, C. Prives, p53: Puzzle and paradigm. *Genes Dev.* **10**, 1054–1072 (1996).
43. D. P. Lane, Cancer, p53, guardian of the genome. *Nature* **358**, 15–16 (1992).
44. K. H. Vousden, D. P. Lane, p53 in health and disease. *Nat. Rev. Mol. Cell Biol.* **8**, 275–283 (2007).
45. J. Yuan, R. Adamski, J. Chen, Focus on histone variant H2AX: To be or not to be. *FEBS Lett.* **584**, 3717–3724 (2010).
46. S. Burma, B. P. Chen, M. Murphy, A. Kurimasa, D. J. Chen, ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *J. Biol. Chem.* **276**, 42462–42467 (2001).
47. A. V. Budanov, The role of tumor suppressor p53 in the antioxidant defense and metabolism. *Subcell. Biochem.* **85**, 337–358 (2014).
48. P. V. Usatyuk *et al.*, Role of c-Met/phosphatidylinositol 3-kinase (PI3K)/Akt signaling in hepatocyte growth factor (HGF)-mediated lamellipodia formation, reactive oxygen species (ROS) generation, and motility of lung endothelial cells. *J. Biol. Chem.* **289**, 13476–13491 (2014).
49. S. Chatterjee *et al.*, Membrane depolarization is the trigger for PI3K/Akt activation and leads to the generation of ROS. *Am. J. Physiol. Heart Circ. Physiol.* **302**, H105–H114 (2012).
50. B. Govindarajan *et al.*, Overexpression of Akt converts radial growth melanoma to vertical growth melanoma. *J. Clin. Invest.* **117**, 719–729 (2007).
51. Q. Chen *et al.*, Akt phosphorylates p47phox and mediates respiratory burst activity in human neutrophils. *J. Immunol.* **170**, 5302–5308 (2003).
52. P. Wardman, Fluorescent and luminescent probes for measurement of oxidative and nitrosative species in cells and tissues: Progress, pitfalls, and prospects. *Free Radic. Biol. Med.* **43**, 995–1022 (2007).
53. I. N. Zelko, T. J. Mariani, R. J. Folz, Superoxide dismutase multigene family: A comparison of the CuZn-SOD (SOD1), Mn-SOD (SOD2), and EC-SOD (SOD3) gene structures, evolution, and expression. *Free Radic. Biol. Med.* **33**, 337–349 (2002).
54. T. Sperka, J. Wang, K. L. Rudolph, DNA damage checkpoints in stem cells, ageing and cancer. *Nat. Rev. Mol. Cell Biol.* **13**, 579–590 (2012).
55. W. Y. Kim, N. E. Sharpless, The regulation of INK4/ARF in cancer and aging. *Cell* **127**, 265–275 (2006).
56. Y. Wang *et al.*, Regulation of VEGF-induced endothelial cell migration by mitochondrial reactive oxygen species. *Am. J. Physiol. Cell Physiol.* **301**, C695–C704 (2011).
57. C. R. Hoyal *et al.*, Modulation of p47PHOX activity by site-specific phosphorylation: Akt-dependent activation of the NADPH oxidase. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5130–5135 (2003).
58. J. Chen, H. Tang, N. Hay, J. Xu, R. D. Ye, Akt isoforms differentially regulate neutrophil functions. *Blood* **115**, 4237–4246 (2010).
59. L. Xu *et al.*, Gene expression changes in an animal melanoma model correlate with aggressiveness of human melanoma metastases. *Mol. Cancer Res.* **6**, 760–769 (2008).
60. D. Talantov *et al.*, Novel genes associated with malignant melanoma but not benign melanocytic lesions. *Clin. Cancer Res.* **11**, 7234–7242 (2005).
61. I. Nindl *et al.*, Identification of differentially expressed genes in cutaneous squamous cell carcinoma by microarray expression profiling. *Mol. Cancer* **5**, 30 (2006).
62. K. J. Kao, K. M. Chang, H. C. Hsu, A. T. Huang, Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC Cancer* **11**, 143 (2011).
63. S. Loi *et al.*, Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J. Clin. Oncol.* **25**, 1239–1246 (2007).
64. Y. Wang *et al.*, Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**, 671–679 (2005).
65. R. Kuner *et al.*, Global gene expression analysis reveals specific patterns of cell junctions in non-small cell lung cancer subtypes. *Lung Cancer* **63**, 32–38 (2009).
66. A. H. Bild *et al.*, Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357 (2006).
67. M. Bittner *et al.*, "Expression Project for Oncology (expO)." NCBI Gene Expression Omnibus. <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2109>. Accessed 18 October 2020.
68. Y. R. Chin *et al.*, Targeting Akt3 signaling in triple-negative breast cancer. *Cancer Res.* **74**, 964–973 (2014).
69. S. Banerji *et al.*, Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486**, 405–409 (2012).
70. J. M. Stahl *et al.*, Deregulated Akt3 activity promotes development of malignant melanoma. *Cancer Res.* **64**, 7002–7010 (2004).
71. J. Gao *et al.*, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, p11 (2013).
72. E. Cerami *et al.*, The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
73. J. Guo, Z. Gertsberg, N. Ozgen, S. F. Steinberg, p66Shc links alpha1-adrenergic receptors to a reactive oxygen species-dependent AKT-FOXO3A phosphorylation pathway in cardiomyocytes. *Circ. Res.* **104**, 660–669 (2009).
74. D. R. Calnan, A. Brunet, The FoxO code. *Oncogene* **27**, 2276–2288 (2008).
75. H. J. Park *et al.*, FoxM1, a critical regulator of oxidative stress during oncogenesis. *EMBO J.* **28**, 2908–2918 (2009).
76. E. C. Lien *et al.*, Glutathione biosynthesis is a metabolic vulnerability in PI(3)K/Akt-driven breast cancer. *Nat. Cell Biol.* **18**, 572–578 (2016).
77. T. Katsube *et al.*, Most hydrogen peroxide-induced histone H2AX phosphorylation is mediated by ATR and is not dependent on DNA double-strand breaks. *J. Biochem.* **156**, 85–95 (2014).
78. J. Willis, Y. Patel, B. L. Lentz, S. Yan, APE2 is required for ATR-Chk1 checkpoint activation in response to oxidative stress. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10592–10597 (2013).
79. E. Lecona, O. Fernandez-Capitillo, Targeting ATR in cancer. *Nat. Rev. Cancer* **18**, 586–595 (2018).



# CXCL10 plays a key role as an inflammatory mediator and a non-invasive biomarker of non-alcoholic steatohepatitis

Xiang Zhang<sup>1,2</sup>, Jiayun Shen<sup>1,2</sup>, Kwan Man<sup>3</sup>, Eagle S.H. Chu<sup>1,2</sup>, Tung On Yau<sup>1,2</sup>, Joanne C.Y. Sung<sup>1</sup>, Minnie Y.Y. Go<sup>1</sup>, Jun Deng<sup>4</sup>, Liwei Lu<sup>4</sup>, Vincent W.S. Wong<sup>1</sup>, Joseph J.Y. Sung<sup>1,2</sup>, Geoffrey Farrell<sup>5</sup>, Jun Yu<sup>1,2,\*</sup>

<sup>1</sup>Institute of Digestive Disease and The Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China; <sup>2</sup>Gastrointestinal Cancer Biology and Therapeutics Laboratory, Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China; <sup>3</sup>Department of Surgery, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China; <sup>4</sup>Department of Pathology and Center of Infection and Immunology, The University of Hong Kong, Hong Kong, China; <sup>5</sup>Australian National University Medical School at The Canberra Hospital, Canberra, Australia

**Background & Aims:** Perpetuate liver inflammation is crucial in the pathogenesis of non-alcoholic steatohepatitis (NASH). Expression of CXCL10, a pro-inflammatory cytokine, correlates positively with obesity and type 2 diabetes. Whether CXCL10 plays a role in NASH was unknown. We aimed to investigate the functional and clinical impact of CXCL10 in NASH.

**Methods:** *Cxcl10* gene-deleted (*Cxcl10*<sup>-/-</sup>) and C57BL/6 wild type (WT) mice were fed a methionine- and choline-deficient (MCD) diet for 4 or 8 weeks. In other experiments, we injected neutralizing anti-CXCL10 mAb into MCD-fed WT mice. Human serum was obtained from 147 patients with biopsy-proven non-alcoholic fatty liver disease and 73 control subjects.

**Results:** WT mice, fed the MCD diet, developed steatohepatitis with higher hepatic CXCL10 expression. *Cxcl10*<sup>-/-</sup> mice were refractory to MCD-induced steatohepatitis. We further revealed that CXCL10 was associated with the induction of important pro-inflammatory cytokines (TNF- $\alpha$ , IL-1 $\beta$ , and MCP-1) and

activation of the NF- $\kappa$ B pathway. CXCL10 was linked to steatosis through upregulation of the lipogenic factors SREBP-1c and LXR, and also to oxidative stress (upregulation of CYP2E1 and C/EBP $\beta$ ). Blockade of CXCL10 protected against hepatocyte injury *in vitro* and against steatohepatitis development in mice. We further investigated the clinical impact of CXCL10 and found circulating and hepatic CXCL10 levels were significantly higher in human NASH. Importantly, the circulating CXCL10 level was correlated with the degree of lobular inflammation and was an independent risk factor for NASH patients.

**Conclusions:** We demonstrate for the first time that CXCL10 plays a pivotal role in the pathogenesis of experimental steatohepatitis. CXCL10 maybe a potential non-invasive biomarker for NASH patients.

© 2014 European Association for the Study of the Liver. Published by Elsevier B.V. Open access under [CC BY-NC-ND license](#).

**Keywords:** Non-alcoholic fatty liver disease; Inflammation; Chemokine; Animal model; Biomarker.

Received 16 January 2014; received in revised form 14 June 2014; accepted 6 July 2014; available online 15 July 2014

\* Corresponding author. Address: Institute of Digestive Disease, Department of Medicine and Therapeutics, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China. Tel.: +852 37636099; fax: +852 21445330.

E-mail address: [junyu@cuhk.edu.hk](mailto:junyu@cuhk.edu.hk) (J. Yu).

**Abbreviations:** NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; CXCL, CXC chemokine ligand; TLR, toll-like receptor; NF- $\kappa$ B, nuclear factor- $\kappa$ B; WT, wild type; MCD, methionine- and choline-deficient; ALT, alanine aminotransferase; TUNEL, terminal deoxynucleotidyl transferase dUTP nick end labelling; FACS, fluorescence activated cell sorting; ROC, receiver operating characteristic; IQR, interquartile range; TBARS, thiobarbituric acid reactive substances; TNF- $\alpha$ , tumor necrosis factor- $\alpha$ ; IL, interleukin; MCP-1, monocyte chemoattractant protein 1; C/EBP $\beta$ , CCAAT/enhancer binding protein beta; COX-2, cyclooxygenase-2; ICAM-1, intercellular adhesion molecule-1; LXR, liver X receptors; SREBP-1c, sterol regulatory element binding protein isoform 1c; ChREBP, carbohydrate response element binding protein; SCD-1, stearoyl-CoA desaturase isoform-1; Cyp, cytochrome P450; mAb, monoclonal antibodies; BMI, body mass index; AUROC, area under the receiver operating characteristic curve; LDL-c, low density lipoprotein-cholesterol; HbA1c, glycated haemoglobin; ACK, ammonium chloride potassium.

## Introduction

Non-alcoholic fatty liver disease (NAFLD) has become increasingly important worldwide due to changes in lifestyle and resultant over-nutrition [1]. Non-alcoholic steatohepatitis (NASH) is a severe form of NAFLD, characterized by necroinflammation and lipid accumulation [2,3]. Little is known about the factors responsible for the transition from benign steatosis to steatohepatitis in NAFLD/NASH. As a consequence, apart from addressing lifestyle issues, there are few effective interventions to treat patients with NASH. The present concept about NASH pathogenesis is that increased levels of toxic lipids, such as free fatty acids or free cholesterol provide initiating and propagating mechanism for hepatocellular injury and resultant inflammation. Inflammation may result from oxidative stress and pro-inflammatory chemokines and cytokines, which perpetuate liver injury and lead to fibrosis [4]. Identification of the pro-inflammatory cytokines, which are associated with lipotoxicity, may improve our understanding of



## Research Article

the pathogenesis of NASH, enabling the development of novel pharmacological treatments.

One particularly important pro-inflammatory cytokine associated with lipotoxicity is the CXC motif chemokine ligand 10 (CXCL10), which recruits inflammatory cells to the site of tissue damage [5,6]. CXCL10 has been implicated in the pathogenesis of hepatitis C virus infection through interactions with the toll-like receptor (TLR) 2 [7], and in hepatitis B virus-infection through the nuclear factor- $\kappa$ B (NF- $\kappa$ B) pathway [8]. In various types of liver injury, CXCL10 is secreted by hepatocytes in areas of lobular inflammation [9,10] and neutralization of CXCL10 accelerates liver regeneration [11]. These data indicate a potential role for CXCL10 in the development of intrahepatic inflammation. Moreover, CXCL10 is upregulated in NASH patients [12] and correlates positively with the incidence of obesity and type 2 diabetes [13,14]. These findings suggest that CXCL10 could be a pivotal molecule that facilitates transition from benign steatosis to progressively hepatocellular damage and inflammation in steatohepatitis.

We have recently reported that the anti-oxidant enzyme heme oxygenase-1 protects against development of experimental steatohepatitis in association with reduced production of CXCL10 [15]. In the present study, we first investigated the functional role of CXCL10 in the development of steatohepatitis using *Cxcl10* gene-deleted mice, and further explored the molecular mechanisms by which CXCL10 exerts its effects on inflammation, steatosis, oxidative stress and apoptosis. We demonstrated by *in vitro* and *in vivo* approaches that blockade of CXCL10 (neutralizing anti-CXCL10 mAb) protected against steatohepatitis. In particular, we tested the clinical impact of CXCL10 in 147 patients with biopsy-proven NAFLD and 73 control subjects and demonstrated that circulating CXCL10 is an independent risk factor for patients with NASH.

### Materials and methods

#### Animals and treatments

Age-matched male *Cxcl10* knock out (*Cxcl10*<sup>-/-</sup>) and C57BL/6 wild type (WT) mice (from Dr. Andrew D. Luster, Harvard Medical School) were fed either a methionine- and choline-deficient (MCD) diet or a control diet for 4 weeks to establish steatohepatitis, or for 8 weeks to establish fibrosing steatohepatitis [15,16].

For CXCL10 neutralization experiments, male C57BL/6 WT mice were given CXCL10-specific anti-CXCL10 mAb (R&D System, Minneapolis, MN) by intraperitoneal injection (50  $\mu$ g in 200  $\mu$ L PBS per mouse) at 12 h before MCD diet, and then the injection was repeated every 2 days for 5 cycles [10,14]. Mice were also given an isotype-matched rat IgG2A mAb (R&D System) at the same time as the controls. In a separate experiment, anti-CXCL10 mAb or control mAb were supplemented for 10 days under MCD diet after induction of steatohepatitis in mice fed the MCD diet for 3 weeks. All animals received humane care and all animal studies were performed in accordance with guidelines approved by the Animal Experimentation Ethics Committee of the Chinese University of Hong Kong.

Mice were sacrificed as previously described [17]. Biochemical determination of serum alanine aminotransferase (ALT) levels, triglycerides and lipid peroxidation rates were performed. Liver histology, liver collagen content analysis, cytokine profiling assay, cDNA expression array, nuclear DNA binding activity assay, terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) assay, fluorescence activated cell sorting (FACS) analysis, qPCR and western blot were performed.

#### Subjects and human sample collection

Serum samples were collected from 147 patients with biopsy-proven NAFLD and 73 healthy subjects as previously described [18,19]. Percutaneous liver biopsy specimens were collected from 11 patients with NASH, 11 patients with

simple steatosis and 15 healthy controls in the Prince of Wales Hospital and the Queen Mary Hospital, Hong Kong. All subjects had given written informed consent and the study protocol was approved by the Clinical Research Ethics Committee of the Chinese University of Hong Kong and the University of Hong Kong.

#### Statistical analysis

Differences between two groups were compared by the Mann-Whitney *U* test or Student's *t* test. Multiple group comparisons were made by the Kruskal-Wallis test or one-way ANOVA. Spearman's correlation coefficient was used to estimate the association of serum CXCL10 levels and several factors of interest, while multiple linear regression was used to determine the independent factors associated with levels of CXCL10. Multiple logistic regression was performed to identify the independent risk factors of NASH. A receiver operating characteristic (ROC) curve analysis was conducted to assess the performance of CXCL10 in the prediction of NAFLD/NASH. All statistical tests were performed using SPSS or GraphPad Software. Data were expressed as mean  $\pm$  standard deviation or median (interquartile range [IQR]) and considered significant at *p* < 0.05.

Additional experimental procedures are provided in the [Supplementary Materials and methods section](#).

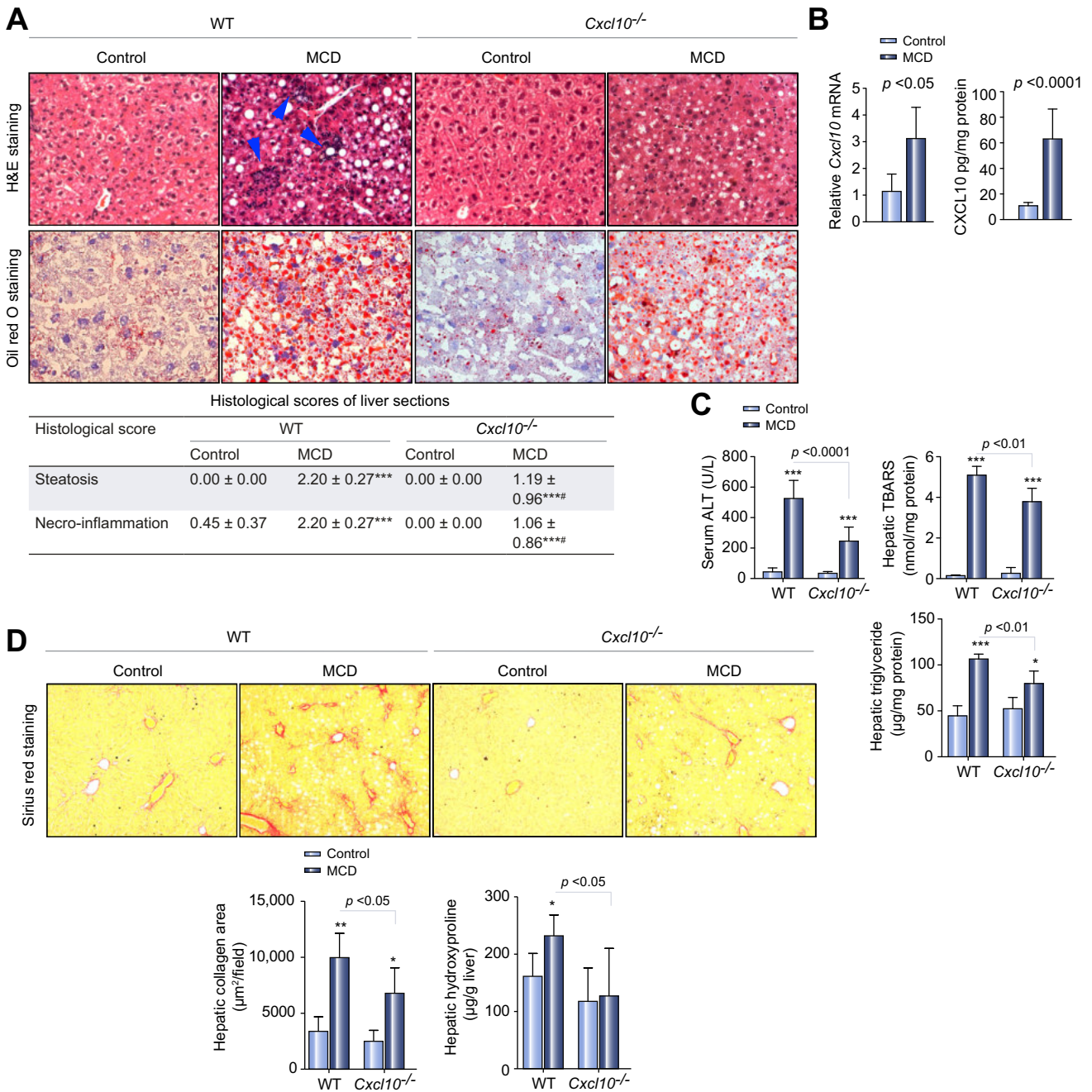
### Results

#### Hepatic CXCL10 expression is upregulated in experimental steatohepatitis and is required for its development

To elucidate the role of CXCL10 in the development of steatohepatitis, *Cxcl10*<sup>-/-</sup> and WT mice were fed control or MCD diets for 4 weeks. MCD-fed WT mice developed steatosis, ballooning hepatocytes, scattered lobular inflammatory cell infiltration, and inflammatory foci (Fig. 1A), consistent with steatohepatitis. This was associated with increased hepatic CXCL10 mRNA and protein levels compared with mice fed a control diet (Fig. 1B), which showed normal liver histology (Fig. 1A). Conversely, MCD-fed *Cxcl10*<sup>-/-</sup> mice showed significant less steatosis (*p* < 0.01) and reduced inflammatory cell infiltration (*p* < 0.01), as indicated by steatosis and necroinflammatory scores (Fig. 1A). Consistent with the histologic findings, measurement of serum ALT (*p* < 0.0001), hepatic lipid peroxide by the thiobarbituric acid reactive substances (TBARS) assay (*p* < 0.01) and hepatic triglyceride contents (*p* < 0.01) revealed that loss of CXCL10 protected mice from MCD diet-induced liver injury (Fig. 1C). The decreased lipid accumulation in MCD-fed *Cxcl10*<sup>-/-</sup> mice was confirmed by Oil red O staining (Fig. 1A). Taken together, these data suggest that CXCL10 contributes to the development of steatohepatitis.

#### CXCL10 is required for hepatic nutritional fibrosis

To examine whether CXCL10 plays a role in hepatic nutritional fibrosis, *Cxcl10*<sup>-/-</sup> mice and WT mice were fed with control or MCD diet for 8 weeks. Intraparenchymal pericellular fibrosis developed from steatohepatitis in WT mice fed with MCD for 8 weeks as shown by Sirius Red staining (Fig. 1D), whilst, MCD-fed *Cxcl10*<sup>-/-</sup> mice showed impressively reduced amounts of collagen fibres (Fig. 1D). Morphometric analysis yielded concordant results where the Sirius Red-stained collagen areas were significantly reduced in MCD-fed *Cxcl10*<sup>-/-</sup> mice compared to MCD-fed WT mice (*p* < 0.05). Moreover, quantitation of collagen by measuring hepatic hydroxyproline content supported the improvement of liver fibrosis by CXCL10 deficiency (Fig. 1D).



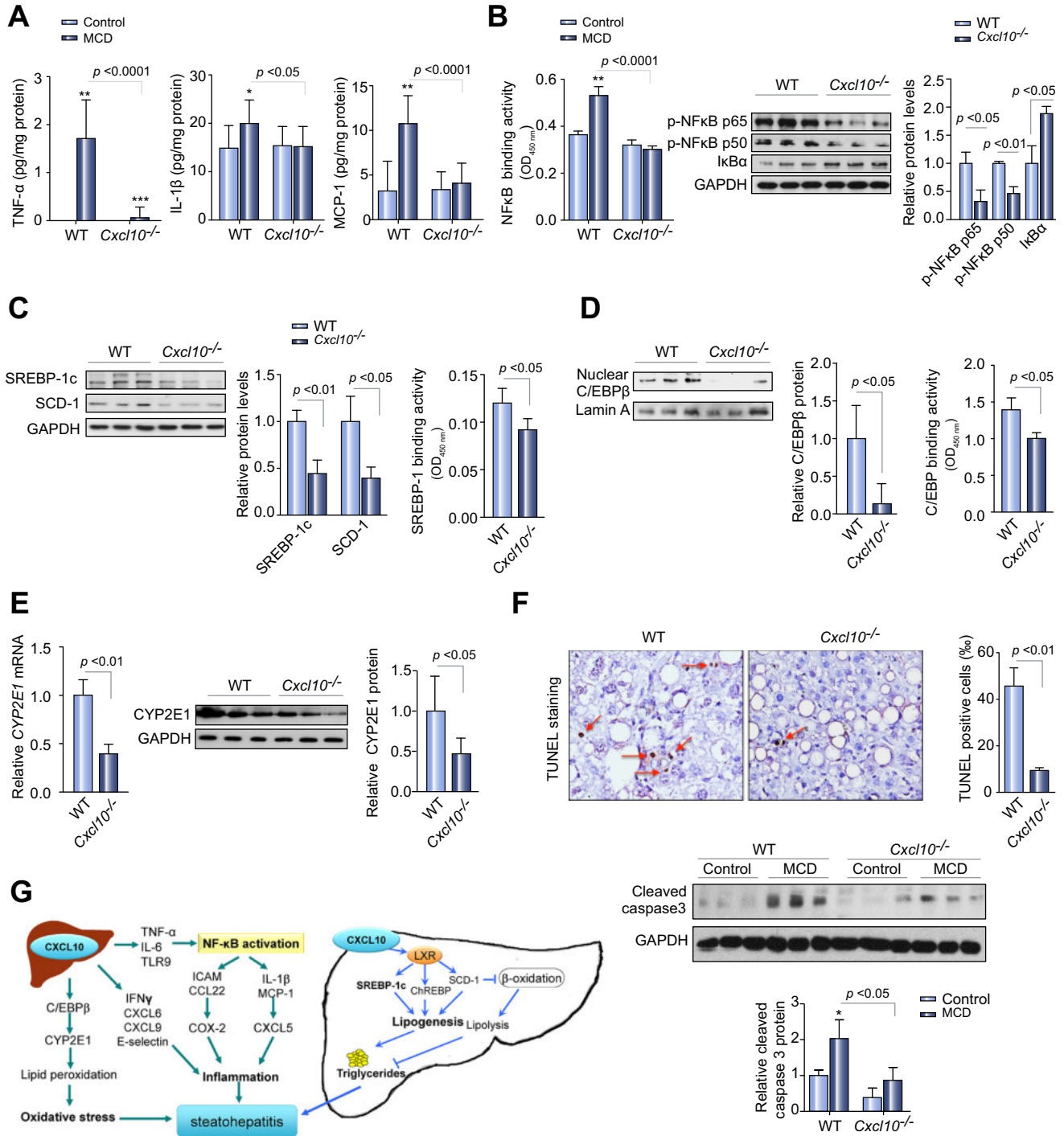
**Fig. 1. Deficiency of CXCL10 attenuates experimental steatohepatitis.** (A) Representative H&E staining (arrows, inflammatory cells) and Oil red O staining from 4-week liver sections of Cxcl10<sup>-/-</sup> and WT mice fed a control or MCD diet. (B) Hepatic CXCL10 mRNA and protein levels in liver tissues of WT mice. (C) Serum ALT, total hepatic lipid peroxide and liver triglyceride content in WT and Cxcl10<sup>-/-</sup> mice fed control or MCD diet for 4 weeks. (D) Collagen deposition by Sirius Red staining and hydroxyproline content of liver sections in mice fed a control or MCD diet for 8 weeks. Data are mean ± SD, n = 5–8/group. \*p < 0.05, \*\*p < 0.001, \*\*\*p < 0.0001 vs. same genotype mice fed control diet. \*\*\*\*p < 0.01 vs. WT mice fed MCD diet.

*CXCL10 induces hepatic chemokines, cytokines and other proinflammatory molecules*

We next determined the mechanisms of CXCL10 in regulating hepatic inflammation by analysing chemokines and cytokines involved in inflammation and cell recruitment. In keeping with the improved liver histology and reduction of liver injury, loss

of CXCL10 significantly reduced the production of key pro-inflammatory chemokines and cytokines such as tumor necrosis factor-α (TNF-α), interleukin (IL)-1β, and monocyte chemoattractant protein-1 (MCP-1), as indicated initially by a cytokine profiling assay (Fig. 2A) and confirmed by qRT-PCR (Supplementary Fig. 1A–C). We then conducted a cDNA expression assay to identify molecules involved in CXCL10-mediated pathogenesis of

# Research Article



**Fig. 2. CXCL10 induces steatohepatitis through hepatic inflammatory molecules, lipogenic factors and oxidative stress.** (A) Hepatic TNF- $\alpha$ , IL-1 $\beta$ , and MCP-1 protein levels in mice fed control or MCD diet. (B) NF- $\kappa$ B nuclear binding activity and protein levels of phosphorylated NF- $\kappa$ B subunits p65, p50 and NF- $\kappa$ B suppressor I $\kappa$ B $\alpha$ , (C) protein levels of SREBP-1c, SCD1 and nuclear SREBP-1c DNA binding activity, (D) protein levels of C/EBP $\beta$  and nuclear C/EBP DNA binding activity, (E) hepatic CYP2E1 mRNA and protein expression in MCD-fed mice. (F) TUNEL positive cells per 1000 cells and cleaved caspase 3 expression in liver tissues for the mechanisms of CXCL10 in the promotion of steatohepatitis. \* $p < 0.05$ , \*\* $p < 0.0001$  vs. same genotype mice fed control diet.

steatohepatitis using an additional panel of inflammatory response factors and comparing their expression in MCD-fed *Cxcl10*<sup>-/-</sup> and WT mouse livers. Loss of CXCL10 was associated with substantially increased expression of CXCL6 (63.2-fold),

CXCL9 (27.7-fold), E-selectin (*SELE*), IFN- $\gamma$ , oxidative stress-associated transcription factor CCAAT/enhancer binding protein beta (*C/EBP $\beta$* ), and NF- $\kappa$ B signalling components including IL-6 (24.1-fold), CCL22 (13.2-fold), TLR9, CXCL5 and cyclooxygenase-2

**Table 1. The effect of CXCL10 on gene expression profiles of inflammatory response in mice liver tissues.**

Gene name	Fold change <sup>†</sup>	Gene function
<i>CXCL6</i>	63.2	Pro-inflammatory cytokine
<i>CXCL9</i>	27.7	Inflammatory response
<i>E-Selectin</i>	7.4	Inflammatory response
<i>IFN-γ</i>	5.3	Pro-inflammatory cytokine
<i>C/EBPβ</i>	4.2	Oxidative stress
<i>IL-6</i>	24.1	Acute-phase response
<i>CCL22</i>	13.2	Inflammatory response
<i>TLR9</i>	4.2	Inflammatory response
<i>CXCL5</i>	2.8	Inflammatory response
<i>COX-2</i>	2.7	Acute-phase response

<sup>†</sup>MCD-fed WT mice vs. MCD-fed *Cxcl10*<sup>-/-</sup> mice.

(*COX-2*) (Table 1). These data support that *CXCL10* plays a critical role in liver inflammation.

#### *CXCL10* activates NF-κB

Given the crucial role of NF-κB signalling in the pathogenesis of steatohepatitis [20,21], we examined whether *CXCL10* played any role in modulation of this pathway in steatohepatitis. NF-κB nuclear binding activity was increased in MCD-fed WT mice compared with control diet ( $p < 0.0001$ ), but not in MCD-fed *Cxcl10*<sup>-/-</sup> mice (Fig. 2B). This was confirmed by enhanced levels of phosphorylated NF-κB subunits p65 and p50, decreased cytosolic NF-κB suppressor IκBα (Fig. 2B) and upregulation of NF-κB downstream factor intercellular adhesion molecule-1 (ICAM-1) in MCD-fed WT mice compared to corresponding *Cxcl10*<sup>-/-</sup> mice (Supplementary Fig. 1D). Our findings indicate that *CXCL10* employs NF-κB signalling to mediate inflammation in steatohepatitis.

#### *CXCL10* contributes to hepatic steatosis by inducing lipogenic genes

To seek an explanation for the reason why deletion of *CXCL10* caused less steatosis, we assessed hepatic expression of lipogenic regulators and genes, including liver X receptor (*LXR*)α, *LXR*β, sterol regulatory element binding protein isoform 1c (*SREBP-1c*), carbohydrate response element binding protein (*ChREBP*) and stearoyl-CoA desaturase isoform-1 (*SCD-1*) as well as genes and regulators of hepatic fatty acid oxidation, such as adiponectin, peroxisome proliferator-activated receptor-alpha and its downstream target molecules, acyl-CoA oxidase, long-chain acyl-CoA dehydrogenase, cytochrome P450 (*CYP*) 4a10 and 4a14. Compared to MCD-fed WT mice, MCD-fed *Cxcl10*<sup>-/-</sup> mice showed significantly lower mRNA expression of *LXR*α, *LXR*β, *SREBP-1c*, *ChREBP*, and *SCD-1* (Table 2). Western blot confirmed the down-regulation of *SREBP-1c* and *SCD-1* protein expression in liver lysates (Fig. 2C). Concomitantly, the nuclear DNA-binding activity of *SREBP-1c* was decreased in *Cxcl10*<sup>-/-</sup> mice (Fig. 2C). Expression of lipolytic genes, regulating fatty acid oxidation, was similar between *Cxcl10*<sup>-/-</sup> and WT mice fed MCD (Table 2). These findings indicate that *CXCL10* either directly or indirectly (such as via MCP-1) can influence hepatic lipogenesis, thereby contributing to steatosis as well as its inflammatory consequences.

#### *CXCL10* contributes to oxidative stress through *CYP2E1* and *C/EBPβ*

In addition to the accumulation of hepatic triglycerides in response to choline deficiency and lipogenesis [22,23], and impaired antioxidant defences in response to methionine deficiency, induction of *CYP2E1* (or *CYP4A*) [24] and *C/EBPβ* [17] may induce oxidative stress in the MCD model of steatohepatitis. To establish whether the latter factors contributed to the protection against steatohepatitis afforded by *CXCL10* deletion, we evaluated the levels of *C/EBPβ* and *CYP2E1*. Both *C/EBPβ* and *CYP2E1* mRNA and protein expression were significantly less in MCD-fed *Cxcl10*<sup>-/-</sup> compared to corresponding WT mice (Fig. 2D and E, Table 1). Nuclear *C/EBP* DNA-binding activity was also decreased in *Cxcl10*<sup>-/-</sup> mice (Fig. 2D). Thus, *CXCL10* could contribute to hepatic oxidative stress in steatohepatitis by regulating *C/EBP* and its downstream target *CYP2E1*.

#### *CXCL10* contributes to hepatic apoptosis

As the elevation of apoptotic cell death is closely associated with the severity of NASH [25], we assessed the role of *CXCL10* in regulating hepatic apoptosis in steatohepatitis by TUNEL assay. We found that TUNEL-positive cells were significantly less in *Cxcl10*<sup>-/-</sup> mice compared to WT mice fed MCD (0.94% vs. 4.56%,  $p < 0.01$ ) (Fig. 2F). Consistent with the impaired apoptosis, the protein expression of the active form of the apoptosis regulator caspase-3 was significantly downregulated in *Cxcl10*<sup>-/-</sup> mice compared with WT mice (Fig. 2F).

#### Phenotypic analysis of immune cells in the spleen and peripheral blood of *Cxcl10*<sup>-/-</sup> and WT mice

In order to investigate the major immune cell populations in *CXCL10*<sup>-/-</sup> and WT mice, we performed FACS analysis in the spleen and peripheral blood. Consistent with a previous report of *Cxcl10*<sup>-/-</sup> mice [26], the frequencies of B cells (*CD19*<sup>+</sup>*CD3*<sup>-</sup>), T cells (*CD3*<sup>+</sup>*CD19*<sup>-</sup>), NK cells (*NK1.1*<sup>+</sup>*CD3*<sup>-</sup>), NKT cells (*NK1.1*<sup>+</sup>*CD3*<sup>+</sup>), macrophages (*CD11b*<sup>+</sup>*F4/80*<sup>+</sup>) and neutrophils (*CD11b*<sup>+</sup>*Ly6G*<sup>+</sup>) in the spleen and peripheral blood of *Cxcl10*<sup>-/-</sup> mice were not changed compared to WT mice under MCD or control diet (Supplementary Figs. 2 and 3). Only a slight reduction of *CD8*<sup>+</sup> T lymphocytes (*CD8*<sup>+</sup>*CD4*<sup>-</sup>) was observed in the spleen of MCD-fed *Cxcl10*<sup>-/-</sup> mice compared to WT mice fed with the same diet ( $34.3 \pm 2.26$  vs.  $38.5 \pm 1.34$ ) (Supplementary Fig. 2).

#### Inactivation of *CXCL10* by anti-*CXCL10* mAb antagonizes MCD-induced steatohepatitis

The above results indicate essential and multiple roles of *CXCL10* in steatohepatitis pathogenesis. If this is the case, specific *CXCL10* inhibition should dampen or abrogate the development of this type of liver pathology. To test this, we first examined the functional effect of an anti-*CXCL10* monoclonal antibody (mAb) on steatosis and injury to hepatocyte-derived cells *in vitro*. As reported [15,17], incubation of the immortalized murine hepatocyte cell line AML-12 with MCD medium for 24 h increased medium ALT, cellular triglyceride and oxidative stress, detected by TBARS and lipid hydroperoxide assays (Supplementary Fig. 4). Conversely, anti-*CXCL10* mAb added to AML-12 cells incubated in MCD medium significantly reduced medium ALT, cellular triglycerides, cellular TBARS and lipid hydroperoxide levels

## Research Article

**Table 2. Hepatic mRNA expression of genes involved in fatty acid regulation in *Cxcl10*<sup>-/-</sup> mice.**

Gene	WT mice		<i>Cxcl10</i> <sup>-/-</sup> mice	
	Control	MCD	Control	MCD
<b>Lipogenic genes</b>				
<i>LXRα</i>	1.05 ± 0.33	1.52 ± 0.31**	0.99 ± 0.12	0.97 ± 0.23##
<i>LXRβ</i>	1.05 ± 0.30	1.91 ± 0.68**	1.28 ± 0.31	1.30 ± 0.34#
<i>SREBP-1c</i>	1.01 ± 0.16	0.63 ± 0.30*	0.53 ± 0.23	0.30 ± 0.11#
<i>ChREBP</i>	1.01 ± 0.18	0.66 ± 0.26*	1.25 ± 0.31	0.32 ± 0.08***#
<i>SCD-1</i>	1.035 ± 0.33	0.012 ± 0.006***	0.570 ± 0.213	0.006 ± 0.004***#
<b>Lipolytic genes</b>				
<i>Adiponectin</i>	1.44 ± 1.09	0.30 ± 0.29	2.15 ± 0.97	1.08 ± 0.91
<i>PPARα</i>	1.01 ± 0.14	0.69 ± 0.23*	1.17 ± 0.14	0.59 ± 0.09***
<i>ACO</i>	1.04 ± 0.35	0.28 ± 0.07**	1.27 ± 0.49	0.29 ± 0.06
<i>LCAD</i>	1.04 ± 0.31	1.19 ± 0.53	0.74 ± 0.29	0.71 ± 0.22
<i>CYP4A10</i>	0.91 ± 0.55	2.79 ± 1.20*	1.30 ± 0.68	1.46 ± 0.61
<i>CYP4A14</i>	2.03 ± 2.45	79.5 ± 34.2***	1.69 ± 1.16	50.2 ± 25.4**

Specific mRNA expression values were normalized to the expression of GAPDH. Data are mean ± SD, n = 5–8/group. \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.0001 compared with corresponding mice fed control diet. #*p* < 0.05, ##*p* < 0.0001 compared with WT mice fed the MCD diet.

compared to AML-12 cells in MCD medium exposed to control IgG2A mAb (Supplementary Fig. 4).

We next examined whether administration of the anti-CXCL10 mAb by intraperitoneal injection could prevent MCD-induced steatohepatitis *in vivo*. Administration of the anti-CXCL10 mAb to MCD-fed WT mice reduced steatosis and inflammatory cell infiltration (Fig. 3A), with concordant reduction of serum ALT, hepatic triglyceride and lipid hydroperoxide levels (Fig. 3B) compared to MCD-fed mice administered control mAb. Likewise, CXCL10 neutralization suppressed NF-κB binding activity (*p* < 0.01), and reduced the expression of phosphorylated NF-κB subunits p65 and p50 (*p* < 0.05) and *ICAM-1* mRNA (*p* < 0.05) (Fig. 3C). Moreover, blocking CXCL10 significantly decreased the levels of CYP2E1 (*p* < 0.01) and *SREBP-1c* (*p* < 0.05) (Fig. 3D).

After confirming a preventive effect on steatohepatitis, we further examined whether CXCL10 neutralization could treat steatohepatitis after it has been established. After induction of steatohepatitis in mice fed the MCD diet for 3 weeks, anti-CXCL10 mAb or control mAb was supplemented for 10 days under MCD diet. Histological analysis of livers by H&E and Oil red O staining showed significantly reduced lipid accumulation and inflammatory cell infiltration in MCD-fed mice treated with anti-CXCL10 mAb (Fig. 3E). Anti-CXCL10 mAb treatment in MCD-fed mice also significantly decreased hepatic triglyceride and lipid peroxide levels compared to MCD-fed mice administered with control mAb (Fig. 3F). Moreover, CXCL10 neutralization suppressed hepatic *TNF-α* (*p* < 0.05) and *ICAM-1* (*p* < 0.05) mRNA expression (Fig. 3G). These data added further weight to the effects of CXCL10 in mediating inflammation, oxidative stress and steatosis in the evolution of steatohepatitis.

*CXCL10 is associated with lobular inflammation and acts as an independent risk factor of human NASH*

Since the MCD model reflects pathologically severe steatohepatitis with choline and amino acid nutritional deficiency and a context of “lipid trapping” in the liver with severe oxidative stress, it remains important to establish whether human NASH related to

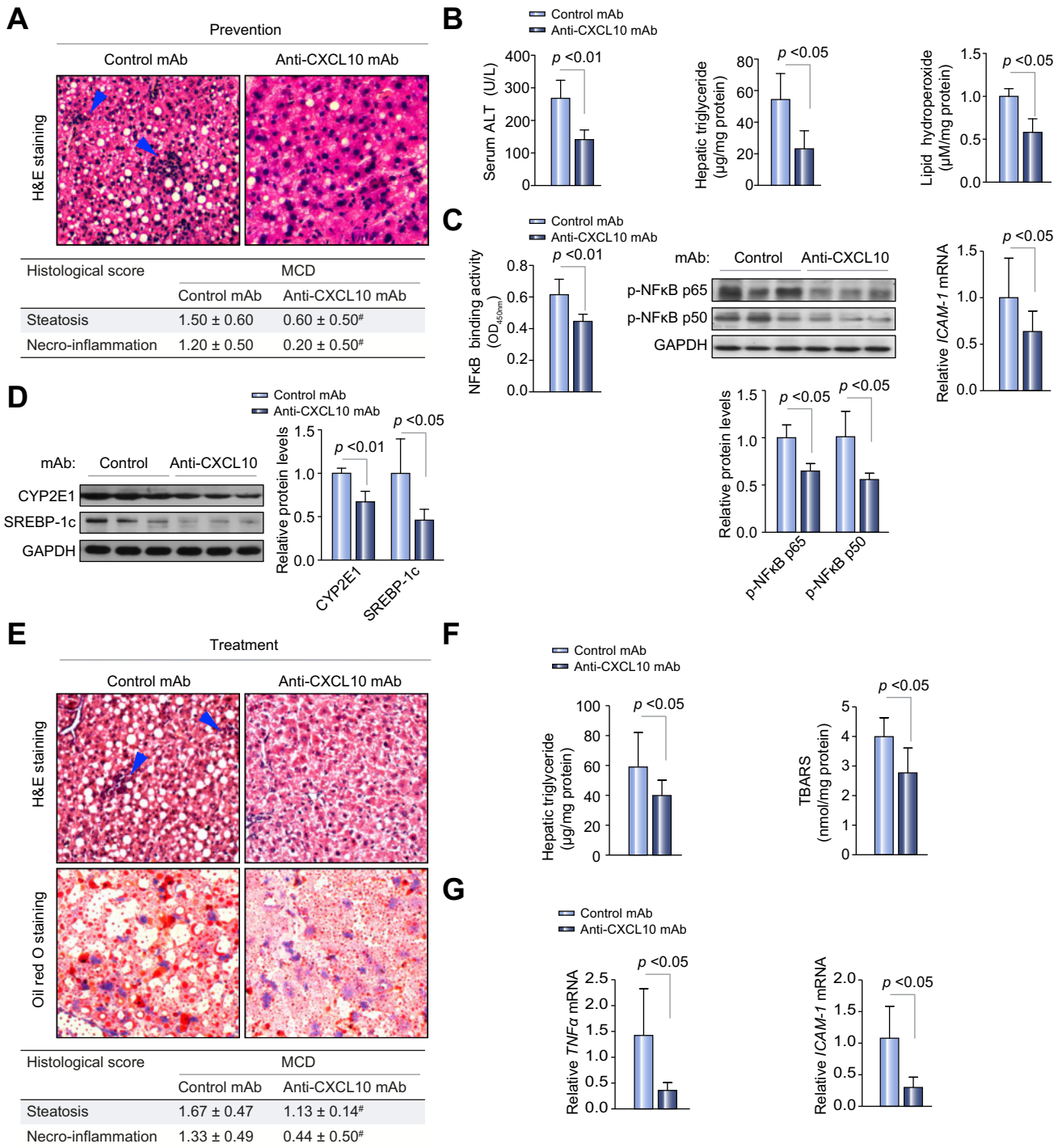
over-nutrition is also associated with increased liver expression and circulating levels of CXCL10. To this end, we assayed *CXCL10* mRNA in liver biopsy from 15 control subjects and 22 NAFLD patients (11 simple steatosis patients and 11 human NASH patients). The results showed that hepatic *CXCL10* mRNA levels were significantly higher in primary NASH tissue compared to simple steatosis (*p* < 0.05) and normal controls (*p* < 0.001) (Fig. 4A), inferring that hepatic CXCL10 production is prominent in patients with NASH.

We next ascertained the clinical impact of CXCL10 in NASH patients. We enrolled a well-established prospective cohort of 73 control subjects without fatty liver measured by proton-magnetic resonance spectroscopy and 147 age and gender matched biopsy-proven NAFLD patients, 69 of whom were diagnosed as NASH [18,19]. We found that serum CXCL10 was significantly increased in a stepwise fashion from control subjects (111 [IQR: 98–146] pg/ml), patients with simple steatosis (170 [133–225] pg/ml) to patients with NASH (248 [154–310] pg/ml) (Fig. 4B, all *p* < 0.0001). In NAFLD patients (simple steatosis and NASH), CXCL10 was significantly and positively correlated with lobular inflammation ( $\rho$ : 0.26, *p* = 0.002) and hepatocyte ballooning degeneration ( $\rho$ : 0.24, *p* = 0.004), which are two major histological features of NASH (Table 3). Multivariable linear regression analysis also demonstrated that the serum CXCL10 level was positively associated with lobular inflammation (Beta: 47.9; 95% CI: 15.0–80.8; *p* = 0.005) and ballooning (Beta: 51.1; 95% CI: 20.0–82.1; *p* = 0.001) independent of metabolic syndrome, body mass index (BMI), ALT, triglyceride, fasting glucose and cholesterol. Moreover, we performed a multivariate logistic regression analysis on these subjects and identified that CXCL10 was an independent risk factor for NASH in NAFLD patients (OR: 1.008, 95% CI: 1.004–1.013, *p* < 0.001) (Table 4, with factors included in the regression model listed).

*CXCL10 is a potential biomarker for the clinical diagnosis of NASH*

To evaluate the utility of CXCL10 as a biomarker in the diagnosis of NAFLD and NASH, a ROC curve was constructed. CXCL10



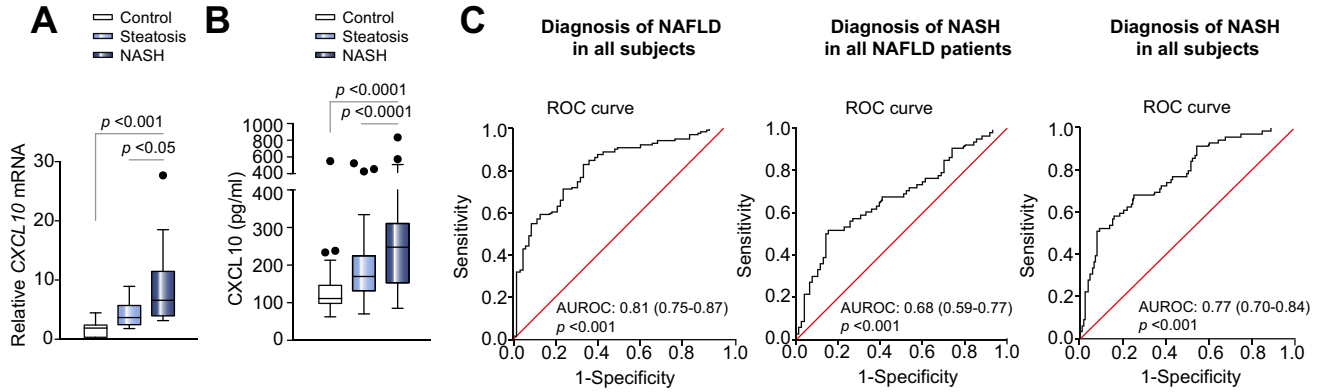


**Fig. 3. CXCL10 neutralization protects against steatohepatitis in vivo.** (A) Representative H&E staining, (B) serum ALT, hepatic triglyceride, lipid hydroperoxide, (C) NF-κB binding activity, phospho NF-κB p65, p50, ICAM-1 levels, (D) CYP2E1 and SREBP-1c expression in mice administrated with anti-CXCL10 or control mAb at 12 h before feeding MCD. (E) Liver sections with H&E staining and Oil red O staining, respectively, (F) hepatic triglyceride and lipid peroxidation products (TBARS), (G) *TNF-α* and *ICAM-1* mRNA expression from mice injected with anti-CXCL10 or control mAb at 3 weeks after MCD feeding. <sup>#</sup>*p* < 0.05 vs. mice treated with control mAb. Data are mean ± SD, n = 5/group.

exhibited a high overall accuracy in discriminating NAFLD from control subjects with the area under the receiver operating characteristic curve (AUROC) of 0.81 (95% CI: 0.75–0.87) (Fig. 4C). In

NAFLD patients, CXCL10 had a moderate accuracy with the AUROC of 0.68 (95% CI: 0.59–0.77) in discriminating NASH from simple steatosis (Fig. 4C). If control subjects were also added to

# Research Article



**Fig. 4. CXCL10 in control and NAFLD patients.** (A) Hepatic human CXCL10 mRNA levels; (B) Serum CXCL10 protein levels; (C) Receiver-operating characteristics curves of CXCL10 in diagnosing NAFLD in all subjects, NASH in NAFLD patients and NASH in all subjects.

**Table 3. Correlations with CXCL10 in NAFLD patients.**

	CXCL10	
	rho	p value <sup>‡</sup>
Age	0.16	0.062
BMI	0.07	0.408
Total cholesterol	0.09	0.304
Triglyceride	0.09	0.286
Steatosis	0.15	0.070
Lobular inflammation	0.26	0.002
Ballooning	0.24	0.004
Fibrosis	0.25	0.002

<sup>‡</sup>p value corresponds to  $H_0: \rho = 0$ .

**Table 4. Multivariable analysis for independent risk factors for NASH in NAFLD patients.**

	OR	95% CI	p value
CXCL10	1.008 <sup>§</sup>	1.004-1.013	<0.001
Metabolic syndrome	3.083	1.203-7.903	0.019

Variables entered in the regression model: CXCL10, gender, age, body mass index (BMI), metabolic syndrome, alanine aminotransferase (ALT), fasting glucose, triglyceride, low density lipoprotein-cholesterol (LDL-c), glycated haemoglobin (HbA1c).

<sup>§</sup>For every 1 unit increase of CXCL10 level.

the analysis, the AUROC of diagnosing NASH increased to 0.77 (95% CI: 0.70–0.84) (Fig. 4C). Thus, CXCL10 can be a novel biomarker for the clinical diagnosis of NAFLD and NASH.

## Discussion

The first novel finding in these studies is that *Cxcl10*<sup>-/-</sup> mice administrated with MCD diet showed significantly attenuated steatohepatitis compared with WT mice fed the same diet; these findings were corroborated by improved liver histology, lowered serum ALT, and hepatic triglyceride content. Moreover, CXCL10 deletion was associated with a significant reduction of intrahepatic oxidative stress, as indicated by decreased lipid peroxide levels. This change was clearly associated with the attenuation

of hepatic inflammation. In addition, CXCL10 deficiency confers protection from hepatic nutritional fibrosis. Our data provide the first evidence that CXCL10 may contribute to lipogenesis, thereby influencing steatosis and possibly lipotoxicity, as well as hepatocellular injury and perpetuation of liver inflammation in steatohepatitis, at least in the MCD model.

The molecular mechanisms by which CXCL10 exerts its broad range of functions in steatohepatitis were subsequently studied. As a key pro-inflammatory cytokine, CXCL10 often amplifies the effects of other cytokines [5]. We therefore evaluated the effect of CXCL10 on other potential cytokines in steatohepatitis and showed that CXCL10 was associated with induction of TNF- $\alpha$ , IL-1 $\beta$ , and MCP-1. TNF- $\alpha$  is a key inflammatory factor involved in the development of human NASH [27] and experimental steatohepatitis [28]. TNF- $\alpha$  can activate neutrophils, cause insulin resistance and promote NASH development. TNF- $\alpha$  and IL-1 $\beta$  are able to induce MCP-1 *in vitro*, suggesting that these cytokines are functionally related [29]. MCP-1 is also an important molecule in NASH as it may bridge inflammatory responses with the induction of insulin resistance [30]. Moreover, MCP-1 can stimulate lipogenesis to promote steatosis in the liver, allowing inflammation to exacerbate steatosis [4]. This suggests that CXCL10 induces cytokine expression, leading to the development of steatohepatitis.

We further characterized the inflammatory factors, regulated by CXCL10 in steatohepatitis, by a cDNA array covering 84 well-known inflammatory genes. Our results show that pro-inflammatory factors, including IFN- $\gamma$ , TLR9, CXCL9, IL-6, SELE, CXCL6, CCL22, CXCL5, and COX-2, were significantly higher in WT mice than in *Cxcl10*<sup>-/-</sup> mice fed with MCD. Each of these molecules could amplify the inflammatory recruitment in steatohepatitis. To be specific, IFN- $\gamma$  is a major inducer of CXCL10 related to NASH pathogenesis [31]; TLR9 activates IFN regulatory factors that induce production of IL-1 $\beta$ , leading to NASH development in mouse model [32]; CXCL9, induced by IFN- $\gamma$ , is increased in the livers of patients with NASH [33], while IL-6 is a key inflammatory factor involved in NASH development [34]. Serum levels of E-selectin (SELE) are also higher in patients with NASH similar to those of IL-6 [35]. CXCL6 is associated with the severity of hepatic inflammation in NAFLD patients and it can be used for predicting NASH progression [36]. Similarly, CCL22 and CXCL5, two small chemokines, are serum markers for NASH and the related obesity and metabolic syndrome [37]. Finally, COX-2,

another pro-inflammatory mediator, plays an important role in metabolic forms of steatohepatitis as reported earlier by us [28]. In addition to IFN- $\gamma$  and TLR9, a close correlation between CXCL10 and TNF- $\alpha$ , MCP-1, IL-6 has been well documented [38,39]. Collectively, these data suggested that induction of pro-inflammatory cytokines and key inflammatory factors by CXCL10 is part of a mechanism for the inflammatory recruitment in MCD-induced steatohepatitis.

An important observation in the present study is that the majority of the above cytokines and inflammatory factors are regulated by NF- $\kappa$ B signalling (Fig. 2G). TNF- $\alpha$  is a potent activator of NF- $\kappa$ B, and in turn activated NF- $\kappa$ B induces TNF- $\alpha$  expression [40]. TLR9 and IL-6 can also activate NF- $\kappa$ B [32], while ICAM-1, CCL22, COX-2, IL-1 $\beta$ , MCP-1, and CXCL5 are downstream effectors of NF- $\kappa$ B activation (Fig. 2G) [20]. These data, when combined with the previous finding that NF- $\kappa$ B is a key regulator of early hepatic inflammatory recruitment and liver injury in NASH [21], implicate a collaborative interaction of CXCL10 and NF- $\kappa$ B to promote steatohepatitis. Therefore, CXCL10 may act as a lipotoxic molecule that activates NF- $\kappa$ B and its downstream inflammatory effectors to induce hepatocyte apoptosis and liver injury, leading to the progression of steatohepatitis.

The underlying causes of hepatic triglyceride accumulation in steatosis include enhanced uptake and synthesis of fatty acids, and inhibition of fatty acid oxidation. In our experimental steatohepatitis model, knockout of *CXCL10* significantly reduced hepatic triglyceride content and steatosis (Fig. 1A and C). This reduction was associated with reduced activity of SREBP-1c and downregulation of SREBP-1c, ChREBP, LXRs, and SCD-1 (Fig. 2 and Table 2), which are involved in *de novo* fatty acid synthesis. In addition to SREBP-1c and ChREBP, LXR is a major transcriptional activator for lipogenesis [41]; it modulates the expression of SREBP-1c through directly binding to the promoter of *SREBP-1c*. LXR also induces the transcription of the lipogenic genes *SCD-1* and *ChREBP* [41]. Thus, the likely pathways by which CXCL10 promotes hepatic steatosis include the upregulation of key fatty acid synthesis genes that promote fatty acid synthesis (Fig. 2G).

It is of general agreement that oxidative stress facilitates the advancement of steatosis to steatohepatitis. Among the common mediators of oxidative stress [42], CYP2E1 is an oxido-reductase that can promote NASH development by inducing oxidative/nitrosative stress, protein modifications, inflammation and insulin resistance [43]. Consistent with this, we confirmed earlier findings [17,24] that CYP2E1 expression is upregulated in MCD-induced steatohepatitis. Importantly, we showed that deletion of *CXCL10* completely abolished the MCD-dependent stimulation of CYP2E1, and significantly reduced the expression of its transcriptional activator C/EBP $\beta$  (Fig. 2D–E). These data were in accordance with the lower level of CYP2E1 in MCD-fed *Cebpb*<sup>-/-</sup> mice compared with MCD-fed WT mice [41], and demonstrate that CXCL10 can act upstream of C/EBP $\beta$  and CYP2E1 to modulate oxidative stress.

If CXCL10 plays a key part in the pathogenesis of steatohepatitis, it would be important to establish that its functional blockade ameliorates the severity of steatohepatitis. To test this, we used anti-CXCL10 mAb to neutralize CXCL10 *in vitro*. Such neutralization caused a dose-dependent decrease in triglyceride secretion and ALT release, together with a concomitant suppression of cellular oxidative stress in AML12 hepatocytes (Supplementary Fig. 4). Moreover, anti-CXCL10 mAb ameliorated the

severity of fatty liver disease in MCD-fed mice. In the present work, CXCL10 neutralization using anti-CXCL10 mAb in mice showed significant improvements in the prevention and regression of steatohepatitis (Fig. 3). These effects were associated with reduced hepatic triglyceride and lipid peroxide levels (Fig. 3). Thus, CXCL10 is a potential target for the prevention and treatment of steatohepatitis.

These mechanistic findings of CXCL10 in the evolution of experimental steatohepatitis encouraged us to explore the clinical impact of CXCL10 in patients with NAFLD and NASH. We first demonstrated that CXCL10 was significantly upregulated both in liver and serum samples of NASH patients. Moreover, the circulating level of CXCL10 in NASH patients was associated with lobular inflammation, which is supported by a previous study that showed that increased CXCL10 levels were correlated with the degree of chronic liver inflammatory damage caused by hepatitis C virus infection [9]. Early identification of patients with NASH may allow intervention that may alter the course of the disease. Currently, liver biopsy remains the standard method for the diagnosis of NASH and differentiation from simple steatosis. However, biopsy is an invasive diagnostic procedure that has been associated with sampling error and observer variability. Thus, the development of a non-invasive test is paramount to the management of NASH. To date, there are no reliable serologic tests for the identification of NASH. Identification of such biomarker would aid clinicians in the identification of patients with NASH, and allow for non-invasive frequent monitoring of disease progression and response to therapy. Building on the significantly elevated CXCL10 level in NASH patients, we tested the clinical utility of CXCL10 as a serologic biomarker for the diagnosis of NASH. Based on a multivariate Cox regression analyses in a study cohort of 147 NAFLD patients and 73 control subjects, CXCL10 was revealed to be a novel risk factor of NASH independent of metabolic syndrome, ALT, diabetes and triglycerides (Table 4). Moreover, the AUROC indicated an overall accuracy of 81% to diagnose NAFLD and an accuracy of 77% to diagnose NASH, suggesting that circulating CXCL10 production could be regarded as a valuable new diagnostic factor for NAFLD and NASH. However, it should be noted that a few prediction models such as the NAFLD fibrosis score have also been developed to predict advanced fibrosis [44,45]. These scores are comprised of predicting factors of fibrosis such as age, BMI and metabolic factors. While it is interesting that CXCL10 may serve as a marker of NASH, the finding warrants independent validation. Furthermore, it would also be important to explore its role in conjunction with other predicting factors to improve the diagnosis.

The MCD diet model is a classic and widely adopted dietary model for studying NASH. It can induce hepatic steatohepatitis with inflammation, oxidative stress, mitochondrial DNA damage, apoptosis and fibrosis [46]. Therefore, it is considered as one of the best-established models for studying NASH-associated inflammation, oxidative stress and fibrosis. However, it does not fully manifest all human NASH features. Mice fed with MCD diet lose weight instead of being obese and lack insulin resistance [47]. In the future, high-fat and high-fructose model (also termed as American Lifestyle-Induced Obesity Syndrome [ALIOS]), which may result in an obese animal with severe steatosis, inflammation, oxidative stress and insulin resistance at 16 weeks [46,48], could be used to support our findings.

In conclusion, these observations and interventions demonstrate for the first time that CXCL10 plays an essential role in

## Research Article

the development of steatohepatitis in the context of fatty liver disease. Further, the mechanism of this effect is through regulation of lipogenesis and oxidative stress either directly or indirectly via pathway modulation and pro-inflammatory signalling, altering the expression of other key chemokines, cytokines and pro-inflammatory molecules. Circulating CXCL10 may be a potential biomarker for patients with NAFLD and NASH.

### Financial support

The project was supported by the Collaborative Research Fund (HKU3/CRF11R, CUHK3/CRF/12R) of the Research Grant Council Hong Kong; National Basic Research Program of China (973 Program, 2013CB531401), the Theme-based Research Scheme of the Hong Kong Research Grants Council (T12-403-11), the CUHK Focused Investments Scheme B, Shenzhen Municipal Science and Technology R & D funding (JCYJ20120619152326450), Shenzhen Virtual University Park Support Scheme to CUHK Shenzhen Research Institute and the Australian National Health and Medical Research Council project grants 585411 and 104288.

### Conflict of interest

The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

### Authors' contributions

XZ was involved in study design, experiments conduct, and data analysis; JS, ESHC, TOY, JCYS, MYYG, and JD performed the research; KM, LL, and VWSW provided material support; JJYS and GF commented on the study and revised the paper. JY designed, supervised the study and wrote the paper.

### Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jhep.2014.07.006>.

### References

- [1] Farrell GC, Wong VW, Chitturi S. NAFLD in Asia – As common and important as in the West. *Nat Rev Gastroenterol Hepatol* 2013;10:307–318.
- [2] Cohen JC, Horton JD, Hobbs HH. Human fatty liver disease: old questions and new insights. *Science* 2011;332:1519–1523.
- [3] Farrell GC, Larter CZ. Nonalcoholic fatty liver disease: from steatosis to cirrhosis. *Hepatology* 2006;43:S99–S112.
- [4] Farrell GC, van Rooyen D, Gan L, Chitturi S. NASH is an inflammatory disorder: pathogenic, prognostic, and therapeutic implications. *Gut Liver* 2012;6:149–171.
- [5] Neville LF, Mathiak G, Bagasra O. The immunobiology of interferon-gamma inducible protein 10 kD (IP-10): a novel, pleiotropic member of the C-X-C chemokine superfamily. *Cytokine Growth Factor Rev* 1997;8:207–219.
- [6] Luster AD, Unkeless JC, Ravetch JV. Gamma-interferon transcriptionally regulates an early-response gene containing homology to platelet proteins. *Nature* 1985;315:672–676.

- [7] Abe T, Fukuhara T, Wen X, Ninomiya A, Moriishi K, Maehara Y, et al. CD44 participates in IP-10 induction in cells in which hepatitis C virus RNA is replicating, through an interaction with Toll-like receptor 2 and hyaluronan. *J Virol* 2012;86:6159–6170.
- [8] Zhou Y, Wang S, Ma JW, Lei Z, Zhu HF, Lei P, et al. Hepatitis B virus protein X-induced expression of the CXC chemokine IP-10 is mediated through activation of NF-kappaB and increases migration of leukocytes. *J Biol Chem* 2010;285:12159–12168.
- [9] Harvey CE, Post JJ, Palladinetti P, Freeman AJ, Ffrench RA, Kumar RK, et al. Expression of the chemokine IP-10 (CXCL10) by hepatocytes in chronic hepatitis C virus infection correlates with histological severity and lobular inflammation. *J Leukoc Biol* 2003;74:360–369.
- [10] Hintermann E, Bayer M, Pfeilschifter JM, Luster AD, Christen U. CXCL10 promotes liver fibrosis by prevention of NK cell mediated hepatic stellate cell inactivation. *J Autoimmun* 2010;35:424–435.
- [11] Yoneyama H, Kai Y, Koyama J, Suzuki K, Kawachi H, Narumi S, et al. Neutralization of CXCL10 accelerates liver regeneration in carbon tetrachloride-induced acute liver injury. *Med Mol Morphol* 2007;40:191–197.
- [12] Bertola A, Bonnafous S, Anty R, Patouraux S, Saint-Paul MC, Iannelli A, et al. Hepatic expression patterns of inflammatory and immune response genes associated with obesity and NASH in morbidly obese patients. *PLoS One* 2010;5:e13577.
- [13] Schulthess FT, Paroni F, Sauter NS, Shu L, Ribaux P, Haataja L, et al. CXCL10 impairs beta cell function and viability in diabetes through TLR4 signaling. *Cell Metab* 2009;9:125–139.
- [14] Morimoto J, Yoneyama H, Shimada A, Shigihara T, Yamada S, Oikawa Y, et al. CXC chemokine ligand 10 neutralization suppresses the occurrence of diabetes in nonobese diabetic mice through enhanced beta cell proliferation without affecting insulinitis. *J Immunol* 2004;173:7017–7024.
- [15] Yu J, Chu ES, Wang R, Wang S, Wu CW, Wong VW, et al. Heme oxygenase-1 protects against steatohepatitis in both cultured hepatocytes and mice. *Gastroenterology* 2010;138:694–704.
- [16] Yu J, Zhang S, Chu ES, Go MY, Lau RH, Zhao J, et al. Peroxisome proliferator-activated receptors gamma reverses hepatic nutritional fibrosis in mice and suppresses activation of hepatic stellate cells in vitro. *Int J Biochem Cell Biol* 2010;42:948–957.
- [17] Shen B, Yu J, Wang S, Chu ES, Wong VW, Zhou X, et al. *Phyllanthus urinaria* ameliorates the severity of nutritional steatohepatitis both in vitro and in vivo. *Hepatology* 2008;47:473–483.
- [18] Shen J, Chan HL, Wong GL, Chan AW, Choi PC, Chan HY, et al. Assessment of non-alcoholic fatty liver disease using serum total cell death and apoptosis markers. *Aliment Pharmacol Ther* 2012;36:1057–1066.
- [19] Shen J, Chan HL, Wong GL, Choi PC, Chan AW, Chan HY, et al. Non-invasive diagnosis of non-alcoholic steatohepatitis by combined serum biomarkers. *J Hepatol* 2012;56:1363–1370.
- [20] Baker RG, Hayden MS, Ghosh S. NF-kappaB, inflammation, and metabolic disease. *Cell Metab* 2011;13:11–22.
- [21] Dela Pena A, Leclercq I, Field J, George J, Jones B, Farrell G. NF-kappaB activation, rather than TNF, mediates hepatic inflammation in a murine dietary model of steatohepatitis. *Gastroenterology* 2005;129:1663–1674.
- [22] Rizki G, Arnaboldi L, Gabrielli B, Yan J, Lee GS, Ng RK, et al. Mice fed a lipogenic methionine-choline-deficient diet develop hypermetabolism coincident with hepatic suppression of SCD-1. *J Lipid Res* 2006;47:2280–2290.
- [23] Larter CZ, Yeh MM, Haigh WG, Williams J, Brown S, Bell-Anderson KS, et al. Hepatic free fatty acids accumulate in experimental steatohepatitis: role of adaptive pathways. *J Hepatol* 2008;48:638–647.
- [24] Leclercq IA, Farrell GC, Field J, Bell DR, Gonzalez FJ, Robertson GR. CYP2E1 and CYP4A as microsomal catalysts of lipid peroxides in murine nonalcoholic steatohepatitis. *J Clin Invest* 2000;105:1067–1075.
- [25] Feldstein AE, Canbay A, Angulo P, Taniai M, Burgart LJ, Lindor KD, et al. Hepatocyte apoptosis and fas expression are prominent features of human nonalcoholic steatohepatitis. *Gastroenterology* 2003;125:437–443.
- [26] Dufour JH, Dziejman M, Liu MT, Leung JH, Lane TE, Luster AD. IFN-gamma-inducible protein 10 (IP-10; CXCL10)-deficient mice reveal a role for IP-10 in effector T cell generation and trafficking. *J Immunol* 2002;168:3195–3204.
- [27] Kugelmas M, Hill DB, Vivian B, Marsano L, McClain CJ. Cytokines and NASH: a pilot study of the effects of lifestyle modification and vitamin E. *Hepatology* 2003;38:413–419.
- [28] Yu J, Ip E, Dela Pena A, Hou JY, Sessa J, Pera N, et al. COX-2 induction in mice with experimental nutritional steatohepatitis: role as pro-inflammatory mediator. *Hepatology* 2006;43:826–836.
- [29] Petrasek J, Bala S, Csak T, Lippai D, Kodys K, Menashy V, et al. IL-1 receptor antagonist ameliorates inflammation-dependent alcoholic steatohepatitis in mice. *J Clin Invest* 2012;122:3476–3489.

- [30] Maher JJ, Leon P, Ryan JC. Beyond insulin resistance: innate immunity in nonalcoholic steatohepatitis. *Hepatology* 2008;48:670–678.
- [31] Baranova A, Schlauch K, Elariny H, Jarrar M, Bennett C, Nugent C, et al. Gene expression patterns in hepatic tissue and visceral adipose tissue of patients with non-alcoholic fatty liver disease. *Obes Surg* 2007;17:1111–1118.
- [32] Miura K, Kodama Y, Inokuchi S, Schnabl B, Aoyama T, Ohnishi H, et al. Toll-like receptor 9 promotes steatohepatitis by induction of interleukin-1beta in mice. *Gastroenterology* 2010;139:323–334.
- [33] Wasmuth HE, Lammert F, Zaldivar MM, Weiskirchen R, Hellerbrand C, Scholten D, et al. Antifibrotic effects of CXCL9 and its receptor CXCR3 in livers of mice and humans. *Gastroenterology* 2009;137:309–319.
- [34] Wieckowska A, Papouchado BG, Li Z, Lopez R, Zein NN, Feldstein AE. Increased hepatic and circulating interleukin-6 levels in human nonalcoholic steatohepatitis. *Am J Gastroenterol* 2008;103:1372–1379.
- [35] Musso G, Gambino R, Bo S, Uberti B, Biroli G, Pagano G, et al. Should nonalcoholic fatty liver disease be included in the definition of metabolic syndrome? A cross-sectional comparison with Adult Treatment Panel III criteria in nonobese nondiabetic subjects. *Diabetes Care* 2008;31:562–568.
- [36] Mehta R, Birerdinc A, Neupane A, Shamsaddini A, Afendy A, Elariny H, et al. Expression of inflammation-related genes is altered in gastric tissue of patients with advanced stages of NAFLD. *Mediators Inflamm* 2013;2013:684237.
- [37] Hammerich L, Heymann F, Tacke F. Role of IL-17 and Th17 cells in liver diseases. *Clin Dev Immunol* 2011;2011:345803.
- [38] Xu W, Joo H, Clayton S, Dullaers M, Herve MC, Blankenship D, et al. Macrophages induce differentiation of plasma cells through CXCL10/IP-10. *J Exp Med* 2012;209:S1811–S1812.
- [39] Tosello-Trampont AC, Landes SG, Nguyen V, Novobrantseva TI, Hahn YS. Kupffer cells trigger nonalcoholic steatohepatitis development in diet-induced mouse model through tumor necrosis factor-alpha production. *J Biol Chem* 2012;287:40161–40172.
- [40] Tak PP, Firestein GS. NF-kappaB: a key role in inflammatory diseases. *J Clin Invest* 2001;107:7–11.
- [41] Jump DB, Botolin D, Wang Y, Xu J, Christian B, Demeure O. Fatty acid regulation of hepatic gene transcription. *J Nutr* 2005;135:2503–2506.
- [42] Abdelmegeed MA, Banerjee A, Yoo SH, Jang S, Gonzalez FJ, Song BJ. Critical role of cytochrome P450 2E1 (CYP2E1) in the development of high fat-induced non-alcoholic steatohepatitis. *J Hepatol* 2012;57:860–866.
- [43] Rahman SM, Schroeder-Gloeckler JM, Janssen RC, Jiang H, Qadri I, Maclean KN, et al. CCAAT/enhancing binding protein beta deletion in mice attenuates inflammation, endoplasmic reticulum stress, and lipid accumulation in diet-induced nonalcoholic steatohepatitis. *Hepatology* 2007;45:1108–1117.
- [44] Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* 2007;45:846–854.
- [45] Wong VW, Wong GL, Chim AM, Tse AM, Tsang SW, Hui AY, et al. Validation of the NAFLD fibrosis score in a Chinese population with low prevalence of advanced fibrosis. *Am J Gastroenterol* 2008;103:1682–1688.
- [46] Kohli R, Feldstein AE. NASH animal models: are we there yet? *J Hepatol* 2011;55:941–943.
- [47] Rinella ME, Green RM. The methionine-choline deficient dietary model of steatohepatitis does not exhibit insulin resistance. *J Hepatol* 2004;40:47–51.
- [48] Tetri LH, Basaranoglu M, Brunt EM, Yerian LM, Neuschwander-Tetri BA. Severe NAFLD with hepatic necroinflammatory changes in mice fed trans fats and a high-fructose corn syrup equivalent. *Am J Physiol Gastrointest Liver Physiol* 2008;295:G987–G995.

WJG 20<sup>th</sup> Anniversary Special Issues (9): Hepatitis B virus

## Management of chronic hepatitis B infection: Current treatment guidelines, challenges, and new developments

Ceen-Ming Tang, Tung On Yau, Jun Yu

Ceen-Ming Tang, Tung On Yau, Jun Yu, Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, LKS Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China  
Tung On Yau, Jun Yu, Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen 518057, Guangdong Province, China

Ceen-Ming Tang, Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, United Kingdom

Author contributions: Tang CM wrote this review; Yau TO revised the article; and Yu J revised the article and supervised the work.

Supported by Collaborative Research Fund (CUHK3/CRF/12R; HKU3/CRF11R) of the Research Grant Council Hong Kong; National Basic Research Program of China, 973 Program, No. 2013CB531401; CUHK Focused Investments Scheme B to HY Lan; and Theme-based Research Scheme of the Hong Kong Research Grants Council, No. T12-403-11

Correspondence to: Jun Yu, MD, PhD, Professor, Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, LKS Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China. [junyu@cuhk.edu.hk](mailto:junyu@cuhk.edu.hk)

Telephone: +852-3-7636099 Fax: +852-2-1445330

Received: October 21, 2013 Revised: November 24, 2013

Accepted: January 19, 2014

Published online: May 28, 2014

### Abstract

Chronic hepatitis B (CHB) virus infection is a global public health problem, affecting more than 400 million people worldwide. The clinical spectrum is wide, ranging from a subclinical inactive carrier state, to progressive chronic hepatitis, cirrhosis, decompensation, and hepatocellular carcinoma. However, complications of hepatitis B virus (HBV)-related chronic liver disease may be reduced by viral suppression. Current international guidelines recommend first-line treatment of CHB infection with pegylated interferon, entecavir, or tenofovir, but the optimal treatment for an individual

patient is controversial. The indications for treatment are contentious, and increasing evidence suggests that HBV genotyping, as well as serial on-treatment measurements of hepatitis B surface antigen and HBV DNA kinetics should be used to predict antiviral treatment response. The likelihood of achieving a sustained virological response is also increased by extending treatment duration, and using combination therapy. Hence the paradigm for treatment of CHB is constantly evolving. This article summarizes the different indications for treatment, and systematically reviews the evidence for the efficacy of various antiviral agents. It further discusses the shortcomings of current guidelines, use of rescue therapy in drug-resistant strains of HBV, and highlights the promising clinical trials for emerging therapies in the pipeline. This concise overview presents an updated practical approach to guide the clinical management of CHB.

© 2014 Baishideng Publishing Group Inc. All rights reserved.

**Key words:** Chronic hepatitis B virus infection; National institute for health and care excellence; Treatment guidelines; Interferon; Pegylated interferon; Nucleos(t)ide analogues; Antiviral resistance; Rescue therapy; Clinical trials

**Core tip:** This article summarizes the different indications for treatment, and systematically reviews the evidence for the efficacy of various antiviral agents. It further discusses the shortcomings of current guidelines, use of rescue therapy in drug-resistant strains of hepatitis B virus, and highlights the promising clinical trials for emerging therapies in the pipeline. This concise overview presents an updated practical approach to guide the clinical management of chronic hepatitis B.

Tang CM, Yau TO, Yu J. Management of chronic hepatitis B infection: Current treatment guidelines, challenges, and new

developments. *World J Gastroenterol* 2014; 20(20): 6262-6278 Available from: URL: <http://www.wjgnet.com/1007-9327/full/v20/i20/6262.htm> DOI: <http://dx.doi.org/10.3748/wjg.v20.i20.6262>

## INTRODUCTION

An estimated 400 million people worldwide have chronic hepatitis B virus (HBV) infection, and more than 750000 deaths are attributed annually to HBV-related complications<sup>[1,2]</sup>. HBV carriers are not only predisposed to developing liver cirrhosis and hepatic decompensation, but also have a 100-fold increased risk of developing hepatocellular carcinoma (HCC)<sup>[3,4]</sup>. Hence early diagnosis and treatment of chronic hepatitis B (CHB) infection is crucial for reducing morbidity and mortality.

Management is guided by recommendations from the American Association for the Society of Liver Disease<sup>[5]</sup>, Asian Pacific Association for the Study of Liver (APASL)<sup>[6]</sup>, European Association for the Study of Liver<sup>[7]</sup>, and the National Institute for Health and Care Excellence (NICE)<sup>[8]</sup>. Broadly, there are two different treatment strategies for patients with CHB infection: therapies of finite duration using immunomodulators such as standard or pegylated interferon- $\alpha$ , as well as long-term treatment with the nucleos(t)ide analogues lamivudine, adefovir dipivoxil, entecavir, telbivudine, or tenofovir. Increasing rates of resistance to antiviral therapy however necessitates consideration of combination therapy, and research into novel treatments. The objective of this review is to provide an update on major advances in the field, addressing controversial areas of uncertainty to aid clinicians in selecting an appropriate therapeutic strategy.

## LIFE CYCLE AND NATURAL HISTORY OF HBV

HBV is a small, partially double-stranded DNA virus that belongs to the family *Hepadnaviridae*<sup>[9]</sup>. The virion is comprised of a core particle containing the viral genome, nucleocapsid protein and polymerase, as well as a lipoprotein envelope composed of viral antigens. Broadly, HBV is classified into four serotypes (adr, adw, ayr and ayw) based on antigenic determinants of the hepatitis B surface antigen (HBsAg), and eight genotypes (A to H) based on its nucleotide sequence. The genotypes have distinct geographic distributions, and an increasing body of evidence suggests it may also influence disease severity and response to treatment (Table 1)<sup>[10]</sup>.

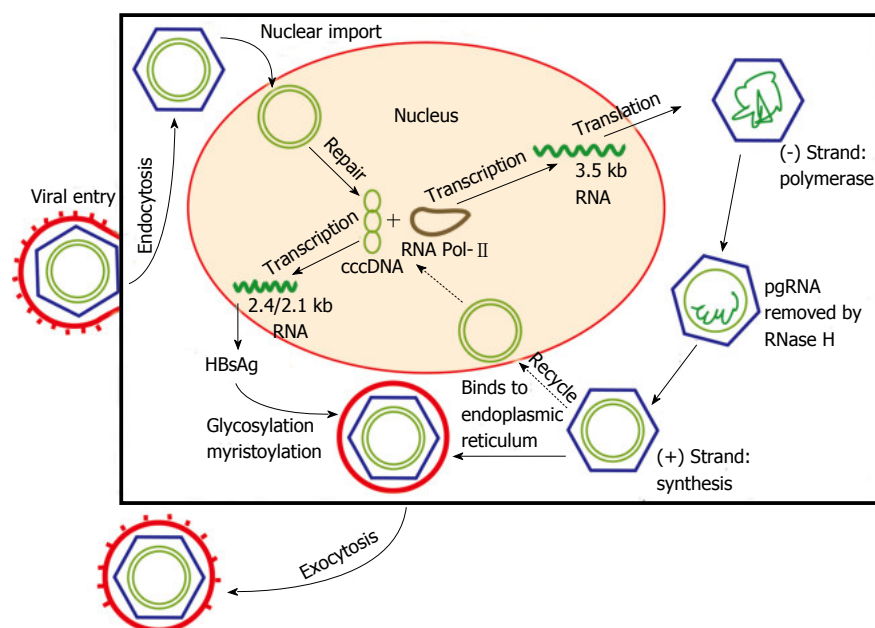
The replication cycle of HBV begins with viral entry into hepatocytes, mediated by the binding of the pre-S1 region on the virion envelope to the cellular sodium taurocholate cotransporting polypeptide<sup>[11]</sup>. The virion is then uncoated, and transported into the nucleus. From a drug discovery point of view, two key events occur. One is the formation of covalently closed circular DNA (cccDNA) through covalent ligation<sup>[12]</sup>. This DNA inter-

**Table 1** Geographic distribution of hepatitis B virus genotypes

Genotypes	Geographic distribution	Tendency of chronicity	Clinical outcome
A	Europe, United States	Higher	Better
B	Eastern Asia	Lower	Better
C	Eastern Asia	Higher	Worse
D	Southern Europe, North Africa, Middle East, Indian Sub-Continent	Lower	Worse
E	Sub-Saharan Africa	-	-
F	South America	-	-
G	Europe, United States	-	-
H	Central America	-	-

mediate is responsible for viral persistence, and is highly resistant to antiviral therapy. The second key event is viral genome replication by reverse transcription *via* pre-genomic RNA. The reverse transcriptase is inherently error prone, and is ultimately responsible for the emergence of nucleos(t)ide-resistant HBV quasiespecies. The mature nucleocapsids may subsequently be recycled into the nucleus to mediate viral persistence, or secreted through exocytosis as Dane particles to infect other hepatocytes (Figure 1)<sup>[13]</sup>. A greater understanding of the viral cycle of HBV will enable new therapeutic strategies.

The natural history of CHB infection is dynamic, involving a complex interplay between the virus and the host immune system. Broadly, there are four stages of variable duration. The initial phase of infection is characterized by immune tolerance. Consequently, serum alanine transaminase (ALT) is normal and liver disease is minimal despite a high level of HBV DNA replication. Individuals, however, are e-antigen (HBeAg) positive and highly infectious. This phase is short when infection is acquired as an adult, but may persist for decades in patients infected perinatally. Nevertheless, tolerance is eventually lost. In the following immune clearance phase, the lysis of infected hepatocytes causes hepatitis, as evidenced by liver necroinflammation and fibrosis, as well as elevations in serum HBV DNA and ALT. The annual rate of HBeAg seroconversion and clearance is between 10% to 20%, and is dependent on factors including HBV genotype, and individuals' age at acute infection<sup>[14,15]</sup>. Where 80% to 90% of infants infected will develop chronic infections, less than 5% of otherwise healthy adults who are infected will fail to spontaneously resolve an acute infection<sup>[16]</sup>. Since repeated exacerbations may occur before viral clearance, the cumulative risk of developing cirrhosis and HCC is increased. Following seroconversion, there is a decrease in viral replication, and remission of inflammation as evidenced by normalization of serum ALT. In contrast to the inactive carrier state, individuals with HBeAg-negative CHB continue to have moderate levels of HBV replication, and active liver disease. This stage may develop immediately after seroconversion, or following several years in the inactive carrier state. It is important to distinguish between inactive carriers and individuals with HBeAg-negative CHB, because the former has a favourable long-term outcome, whilst the latter is



**Figure 1 Hepatitis B viral replication cycle.** The hepatitis B virus virion enters the hepatocyte via endocytosis. Viral nucleocapsids are uncoated and transported into the nucleus, where viral DNA is transformed into covalently closed circular DNA (cccDNA). Replication subsequently occurs through reverse transcription. The mature nucleocapsids are responsible for mediating viral persistence, and may be released to infect neighbouring hepatocytes. HBsAg: Hepatitis B surface antigen.

less responsive to treatment and associated with progressive liver disease<sup>[17]</sup>. Consequently, an understanding of the phases in chronic HBV infection is needed to risk stratify patients, and identify those whom would benefit from treatment.

## IMMUNE SYSTEM IN PATHOGENESIS OF INFECTION

In patients with acute self-limiting HBV infection, the production of interferon gamma (IFN- $\gamma$ ) triggers the activation of natural killer cells, and induction of a robust T cell response. HBV-infected hepatocytes are subsequently cleared by CD8<sup>+</sup> T cells through cytolytic, and non-cytolytic mechanisms<sup>[18]</sup>. Conversely, a weak and transient virus-specific T cell response is observed in patients with chronic HBV infection<sup>[19]</sup>. Potential causes of intrinsic T cell defects include up-regulation of Bcl2-interacting mediator (BIM), which causes deletion of HBV-specific CD8<sup>+</sup> T cells<sup>[20]</sup>. T cell tolerance and exhaustion is perpetuated by an excess of co-inhibitory signals, including CTLA-4 and PD-L1<sup>[21]</sup>, with the activity of remaining CD8<sup>+</sup> T cells further suppressed by high levels of HBV DNA, microRNA-146a, and immunosuppressive cytokines such as interleukin-10<sup>[22,23]</sup>. More controversially, a tolerogenic liver environment, excessive immunosuppression by regulatory T cells<sup>[24]</sup>, and dysfunction of dendritic cells<sup>[25]</sup> have been implicated in the impaired host immune response.

However, repeated attempts by the host immune system to control the infection causes hepatic injury. Natural killer cells may contribute to hepatocellular inflammation through expression of TNF-related apoptosis-inducing ligand<sup>[26]</sup>, and induce apoptosis through the release of granzymes and perforins<sup>[27]</sup>. Recent research has also identified dense non-antigen specific T cell infiltration as a cause of hepatocyte lysis<sup>[28]</sup>. Critically, HBV is not directly cytopathic. Hence modulation of the host immune

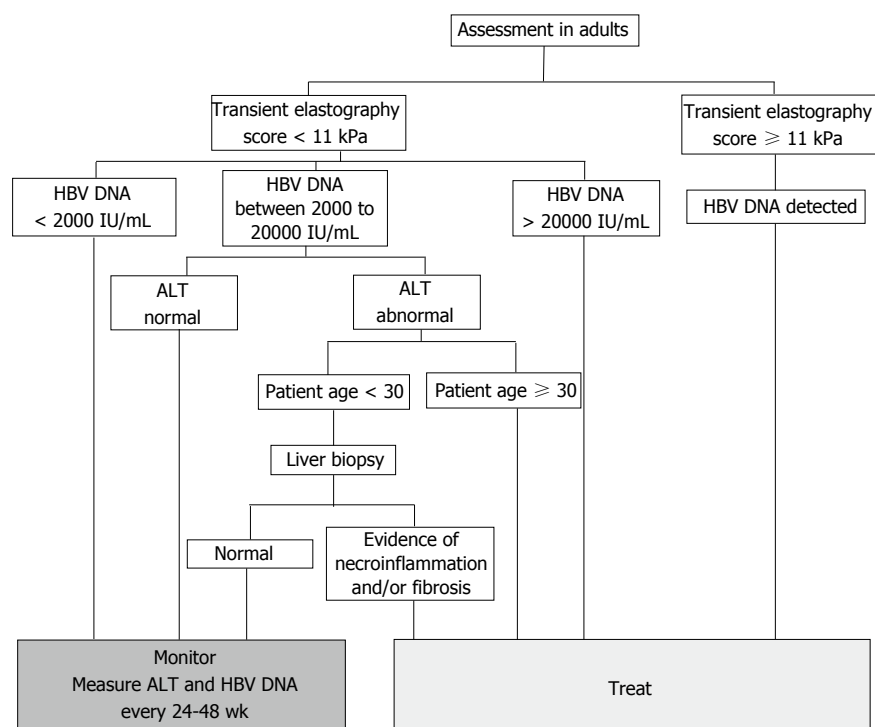
response may enable viral clearance in chronic HBV infection.

## TREATMENT CRITERIA - WHO SHOULD WE TREAT?

The primary goal of CHB treatment is to reduce the risk of developing chronic liver disease associated complications. Longitudinal studies in large cohorts of chronically infected patients have revealed a 15% to 40% cumulative lifetime risk of developing cirrhosis, and amongst patients with established cirrhosis, an annual incidence rate of HCC between 2% to 5%<sup>[4]</sup>. Consequently, every patient would be a candidate for therapy if the virus could be eradicated. However, current therapeutic options only achieve a functional cure through viral suppression, and are associated with numerous adverse effects. Hence the decision on when to initiate treatment is controversial.

The indications for treatment are based on a combination of three criteria: levels of serum HBV DNA, serum ALT, and the severity of liver disease<sup>[7]</sup>. There is no significant distinction made between HBeAg positive and HBeAg negative infection. Current guidelines from NICE in the United Kingdom recommend the use of transient elastography as an initial test to assess the severity of liver disease, and need for treatment (Figure 2)<sup>[8]</sup>. For adults with a transient elastography score  $\geq 11$ , there is a high likelihood of hepatic fibrosis, and treatment should be commenced irrespective of viral load to prevent further deterioration of liver function<sup>[29]</sup>. Similarly, all patients with a HBV DNA level > 20000 IU/mL are offered treatment due to a strong correlation between high viremia, cirrhosis, and HCC<sup>[30,31]</sup>. Cohort studies have also demonstrated a poorer prognosis in patients who had a prolonged immune clearance phase, with early clearance of HBeAg bringing about a 2.2-fold decrease in mortality<sup>[32]</sup>. Hence treatment is recommended for most





**Figure 2** National Institute for Health and Care Excellence algorithm for initiation of treatment in chronic hepatitis B infection.

Current indications for treatment are based on a combination of levels of serum hepatitis B virus (HBV) DNA, serum alanine transaminase (ALT), and the severity of liver disease. Specifically, patients with a transient elastography score  $\geq 11$  kPa are likely to have cirrhosis and confirmation by liver biopsy is not needed. Abnormal ALT, measured by two consecutive tests conducted 3 mo apart, is defined as  $\geq 30$  IU/mL in males, and  $\geq 19$  IU/mL in females.

individuals with a high HBV DNA and an elevated serum ALT.

The exception is for patients under 30 years of age with a normal liver biopsy. Whilst nucleos(t)ide therapy accelerates seroconversion, the risk of HBV reactivation is higher after nucleos(t)ide analogue induced seroconversion, as evidenced by an annual incidence of 12.0% compared to 2.9% following spontaneous seroconversion ( $P = 0.004$ )<sup>[33]</sup>. Hence close monitoring of young, non-cirrhotic, and compensated HBeAg positive patients may be more appropriate than immediate nucleos(t)ide analogue therapy. Treatment is also not recommended for patients who are in the immunotolerant (HBV DNA  $< 20000$  IU/mL, normal serum ALT) or inactive carrier phase (HBV DNA  $< 2000$  IU/mL, serum ALT normal) due to minimal liver disease<sup>[34]</sup>. Instead, patients should be monitored every 6 mo to diagnose a break in immune tolerance or reactivation of viral replication.

In the future, NICE guidelines should be revised to include levels of HBsAg as an indication for, and predictor of treatment response. In HBeAg negative patients with a low viral load ( $< 2000$  IU/mL), multivariate analyses have showed that high levels of HBsAg ( $\geq 1000$  IU/mL) increase the likelihood of developing cirrhosis and HCC<sup>[35,36]</sup>. Significantly, the adjusted hazard ratio for HCC in patients with levels of HBsAg  $\geq 1000$  IU/mL *vs*  $< 1000$  IU/mL was 13.7<sup>[36]</sup>. Thus HBsAg is an independent predictor of outcome in patients with a low viral load. Emerging evidence also suggests increased mortality from liver disease in patients with ALT levels on the upper limit of normal<sup>[37]</sup>. Consequently, further research is needed to improve risk stratification, and identification of patients who would benefit from treatment.

## TREATMENT OPTIONS - WHAT WEAPONS ARE IN OUR ARSENAL?

There are currently seven drugs approved for the treatment of CHB infection in Europe and the United States (Table 2). Broadly, nucleos(t)ide analogues have few side effects, and effectively suppress HBV DNA levels to cause clinical and histological improvements. However, long-term treatment is required to maintain virological control. Conversely, pegylated interferon (PEG-IFN) has a finite duration of treatment and is more likely to produce a sustained virological response. Its use, however, is limited by high costs and numerous associated side effects. Hence treatment must be optimized for each individual.

### PEG-IFN

IFNs are cytokines which interfere with viral replication in host cells by inhibiting viral DNA synthesis, and enhancing the cellular immune response against HBV-infected hepatocytes. Its half-life and drug efficacy may be improved through pegylation. A study of a 24-wk course of weekly PEG-IFN- $\alpha$ -2a showed a higher combined response rate (HBeAg loss, HBV DNA suppression, and ALT normalization) than that achieved by conventional IFN- $\alpha$ -2a (24% *vs* 12%,  $P = 0.036$ ), with no significant difference with respect to frequency and severity of adverse events<sup>[38]</sup>. Hence standard IFN is no longer used where PEG-IFN is available.

PEG-IFN not only offers a finite duration of therapy, but is also superior to lamivudine on the basis of HBeAg seroconversion, HBV DNA suppression, and HBsAg seroconversion in both HBeAg-positive and HBeAg-

**Table 2 Comparison of antiviral agents for chronic hepatitis B**

Antiviral agents	Immunomodulators			Nucleos(t)ide analogues				
	IFN- $\alpha$	PEG-IFN- $\alpha$	Thymosin	Lamivudine	Adefovir	Entecavir	Telbivudine	Tenofovir
Route	SC	SC	Oral	Oral	Oral	Oral	Oral	Oral
Dose	5-10 MIU <i>tiw</i>	180 $\mu$ g <i>qw</i>	1.6 mg <i>biw</i>	100 mg <i>od</i>	10 mg <i>od</i>	0.5-1 mg <i>od</i>	600 mg <i>od</i>	300 mg <i>od</i>
Year approved	1992	2005	Asia only	1998	2002	2005	2006	2008
Antiviral effects								
HBV DNA	37	30	42	36-40	21	67	60	76
HBsAg clearance	++	++	N/A	-	-	+	-	-
HBsAg seroconversion	20-40	27	40	18-20	12	21	22	21
ALT normalization		39	42	62-77	48	68	77	68
Histological improvement		38	N/A	56-62	53	72	65	74
Side effects	Many	Many	Negligible	Negligible	Nephrotoxicity	Negligible	Negligible	Nephrotoxicity
Contraindications	Numerous	Numerous	Uncommon	Uncommon	Uncommon	Uncommon	Uncommon	Uncommon
Drug resistance (treatment-naïve patients)								
1 yr		None, but non-response		24	None	0	4	0
2 yr				38	3	0.2	25	0
> 5 yr				80	29	1	N/A	0
Drug resistance (LAM resistant patients)								
2 yr		None, but non-response		N/A	25	9	N/A	0
4 yr				N/A	N/A	39	N/A	0

PEG-IFN: Pegylated interferon; SC: Subcutaneous; *tiw*: Three times a week; *qw*: Once a week; *biw*: Twice a week; *od*: Once daily; ALT: Alanine transaminase; ETV: Entecavir; LAM: Lamivudine; ADV: Adefovir; TBV: Telbivudine; TDF: Tenofovir disoproxil fumarate; N/A: Not applicable.

negative CHB<sup>[39,40]</sup>. At 6-mo post-treatment, HBeAg seroconversion was observed in 32% of patients treated with PEG-IFN, compared with 19% of patients on lamivudine therapy<sup>[40,41]</sup>. HBsAg seroconversion was also achieved in 16 patients on PEG-IFN therapy, as opposed to zero in the group receiving lamivudine alone ( $P = 0.001$ )<sup>[40]</sup>. Hence PEG-IFN induces a more durable virological response<sup>[42]</sup>. However, the response to IFN is relatively low, and its use is associated with numerous adverse phenomena including depression, paraesthesia, myelosuppression, and other influenza-like symptoms such as fatigue, headaches, and weight loss. Thus HBV genotype and pre-treatment HBV DNA levels must be used to determine the likelihood of patients deriving benefit from PEG-IFN treatment.

### Entecavir

Entecavir is a guanosine analogue with potent activity against HBV through inhibition of DNA polymerase which has been evaluated in two large double-blinded phase III clinical trials involving 715 HBeAg positive and 648 HBeAg negative nucleos(t)ide-naïve patients. Entecavir administered at a dose of 0.5 mg orally once daily was shown to be superior to lamivudine and adefovir<sup>[43-45]</sup>. Specifically, 80% and 87% of entecavir-treated *vs* 39% and 79% of lamivudine-treated patients achieved undetectable HBV DNA and ALT normalization through 96 wk respectively<sup>[46]</sup>. Long term follow-up of the same cohort revealed that an additional 23% achieved HBeAg seroconversion, and 1.4% lost HBsAg after 5 years of continuous entecavir therapy<sup>[47]</sup>. More importantly, entecavir-mediated viral suppression decreased the risk of decompensation, HCC, and death during a median of 20 mo of follow-up<sup>[48]</sup>. Long-term use of entecavir is not associated with any serious adverse effects, and development of resistance is rare in nucleos(t)ide-naïve pa-

tients<sup>[49]</sup>. Hence meta-analyses recommend entecavir and tenofovir as optimal antiviral agents in treatment-naïve individuals<sup>[50]</sup>.

### Tenofovir disoproxil fumarate

Tenofovir is the newest antiviral agent licensed for the treatment of CHB. In a phase III clinical trial, tenofovir at a daily dose of 300 mg for 48 wk induced viral suppression with HBV DNA less than 400 copies/mL in 76% of HBeAg positive and 93% of HBeAg negative patients, as opposed to only 68% and 14% of patients treated with adefovir respectively<sup>[51]</sup>. 8% of HBeAg positive patients had loss of HBsAg after 3 years of tenofovir monotherapy<sup>[52]</sup>. Critically, follow-up of the same cohort after 4.5 years revealed that 87% had histological improvement, 51% had regression of fibrosis, and that 74% of patients with cirrhosis at baseline were no longer cirrhotic<sup>[53]</sup>. Virological breakthrough was infrequent, and there have been no reports of tenofovir resistance after 6 years of therapy<sup>[54]</sup>.

### Alternative treatments

Other pharmacological agents effective in CHB infection include thymosin- $\alpha_1$ , an immunomodulating agent which augments the host Th1 immune response<sup>[55]</sup>. In a meta-analysis including 353 patients from five randomized trials, the odds ratio for a virological response to thymosin- $\alpha_1$  over placebo at the end of treatment, 6 mo post-treatment, and 12 mo post-treatment were 0.56 (0.2-1.52), 1.67 (0.83-3.37), and 2.67 (1.25-5.68) respectively, with virological response increasing over time after cessation of thymosin treatment<sup>[56]</sup>. Subsequent clinical trials demonstrated that the odds of ALT normalization and negative HBV DNA at the end of follow-up (12 mo) was three-fold higher in the thymosin- $\alpha_1$  than the IFN- $\alpha$  group<sup>[57]</sup>. Hence thymosin  $\alpha_1$  is approved in 35 different

**Table 3 National Institute for Health and Care Excellence treatment guidelines**

Guidelines	HBeAg positive	HBeAg negative	Decompensated
1 <sup>st</sup> line	48-wk of PEG-IFN- $\alpha$ -2a	48-wk of PEG-IFN- $\alpha$ -2a	ETV or TDF (LAM resistance)
2 <sup>nd</sup> line	TDF or ETV (TDF contraindication)	ETV or TDF	
3 <sup>rd</sup> line	LAM + TDF or ETV + TDF (LAM resistance)	ETV or TDF	

In adults with compensated liver disease, current guidelines advise first-line treatment with 48 wk of PEG-IFN- $\alpha$ -2a. If PEG-IFN- $\alpha$ -2a is contraindicated, tenofovir or entecavir should be trialed. PEG-IFN: Pegylated interferon; HBeAg: Hepatitis B e antigen; ETV: Entecavir; LAM: Lamivudine; TDF: Tenofovir disoproxil fumarate.

countries worldwide, and is recommended by the APASL as an option in treatment naïve patients<sup>[6]</sup>.

Oxymatrine, an alkaloid extracted from *Sophora alopecuroides* L, has also been shown to effectively and safely suppress HBV replication<sup>[58]</sup>. It acts not only as an immunomodulator<sup>[59]</sup>, but also induces cytochrome P450 to enhance degradation of HBV mRNA to inhibit viral replication<sup>[58]</sup>. In a randomized double-blind placebo-controlled trial, normalization of serum ALT and HBeAg seroconversion occurred in 83% and 40% of patients respectively - an efficacy comparable to that of IFN- $\alpha$ <sup>[60]</sup>. Nevertheless, trials have been small, and larger studies are needed.

## CURRENT TREATMENT GUIDELINES - WHERE DO THEY FALL SHORT?

In adults with compensated liver disease, current guidelines from the NICE advise first-line treatment with 48 wk of PEG-IFN- $\alpha$ -2a, because it results in the highest rate of off-treatment sustained response. Following 24 wk of therapy, HBV DNA should be measured to predict the likelihood of a sustained virological response. In patients with a suboptimal response, defined as a decrease in HBV DNA of less than 2 log<sub>10</sub> IU/mL after 24 wk of therapy, and patients who do not undergo HBeAg seroconversion or a decrease in serum HBsAg after 48 wk of therapy, second line treatment with tenofovir should be offered. For HBeAg negative patients with detectable HBV DNA after 48 wk of treatment, consider switching from tenofovir to entecavir as third line treatment. For HBeAg positive patients where HBV DNA remains detectable at 96 wk with no history of lamivudine resistance, add lamivudine to tenofovir. Otherwise, consider combination therapy with entecavir and tenofovir (Table 3).

However, the paradigm of treatment for CHB is constantly evolving. Growing evidence suggests that the response rate to IFN therapy may vary amongst HBV genotypes. Studies have shown that the rates of HBeAg loss are significantly higher in patients with genotype

**Table 4 Recommendations for the use of pegylated interferon as initial antiviral therapy**

HBV genotype	General recommendations for HBeAg positive patients
A	Either high ALT ( $\geq 2 \times$ ULN) or low HBV DNA levels ( $< 9 \log_{10}$ copies/mL)
B and C	Both high ALT ( $\geq 2 \times$ ULN) and low HBV DNA levels ( $< 9 \log_{10}$ copies/mL)
D	Not recommended

HBV: Hepatitis B virus; HBeAg: Hepatitis B e antigen; ALT: Alanine transaminase; ULN: Upper limit of normal.

B compared to genotype C<sup>[61,62]</sup>. The response to IFN therapy is also higher amongst patients with genotype A as opposed to those with genotype D, in both HBeAg positive (46% *vs* 24%), and HBeAg negative CHB (59% *vs* 29%)<sup>[63]</sup>. Moreover, numerous trials have demonstrated that PEG-IFN- $\alpha$ -2b is the best therapy for achieving HBsAg clearance and a sustained virological response in patients with genotype A<sup>[64,65]</sup>. Therefore, contrary to recommendations from NICE, we believe that routine genotyping is essential for all patients in whom IFN therapy is considered. The likelihood of achieving a sustained response should be predicted based on HBV genotype, ALT, and HBV DNA levels (Table 4)<sup>[66]</sup>. We advise that patients should only receive PEG-IFN as first-line treatment if the baseline probability of sustained virological response is greater than 30%, for only in this patient subgroup do the potential benefits outweigh the costs<sup>[66]</sup>.

More recently, studies have shown that patients' IL28B genotypes are significantly associated with the likelihood of HBeAg seroconversion, and HBsAg loss. In HBeAg negative patients, a retrospective analysis found that a sustained virological response with HBsAg seroconversion was more likely in the rs12979860 CC (*vs* CT/TT) genotype carriers (29% *vs* 13%,  $P < 0.039$ )<sup>[67]</sup>. Other multi-centre studies in HBeAg positive patients have demonstrated a favourable outcome, with an adjusted odds ratio for seroconversion following PEG-IFN treatment of 3.16 for AA *vs* AA/GG genotype carriers at rs12980275<sup>[68,69]</sup>. Hence genome sequencing for IL-28 polymorphisms may provide additional information on the probability of response to IFN therapy.

On-treatment monitoring of HBsAg kinetics may further optimize PEG-IFN treatment by identifying non-responders at an early stage. Broadly, serum HBsAg levels correlate with intrahepatic cccDNA, and low cccDNA levels are predictive of a sustained virological response. Retrospective analyses have demonstrated that 57% of patients with HBsAg levels  $< 1500$  IU/mL after 12 wk of treatment achieved HBeAg seroconversion 6 mo post-treatment<sup>[70]</sup>. This is significantly higher than in patients with HBsAg 1500-20000 IU/mL or  $> 20000$  IU/mL, where HBeAg seroconversion was attained in only 32% and 16% of cases respectively ( $P < 0.0001$ )<sup>[70]</sup>. The predictive power of early HBsAg levels was further emphasized by retrospective analysis of patients enrolled in the Neptune study, where no patients with HBsAg levels  $>$

**Table 5** Comparison of *de novo* combination therapy and monotherapy

Combination therapy	Monotherapy	HBeAg seroconversion	HBV DNA	HBsAg clearance	Histological improvement	Drug resistance
LAM + ADV	LAM	=	=	N/A	=	↓
LAM + ADV	ETV	N/A	=	N/A	N/A	N/A
ETV + TDF	ETV	=	=	N/A	=	=
LAM + PEG-IFN	LAM	↑	↓	↑	↑	↓
ADV + PEG-IFN	PEG-IFN	N/A	↓	↑	N/A	N/A
ETV + PEG-IFN	ETV	↑	↓	↑	N/A	N/A

In several randomized controlled trials, combination therapy with nucleos(t)ide analogues and PEG-IFN are superior to existing management options with regards to viral load, HBeAg seroconversion, and HBsAg clearance. HBeAg: Hepatitis B e antigen; HBsAg: Hepatitis B surface antigen; PEG-IFN: Pegylated interferon; ETV: Entecavir; LAM: Lamivudine; ADV: Adefovir; TBV: Telbivudine; TDF: Tenofovir disoproxil fumarate; N/A: Not applicable.

20000 IU/mL at week 12 on-treatment achieved HBeAg seroconversion<sup>[71]</sup>. Similarly in HBeAg negative patients, the PARC study found that a decrease in HBsAg of < 2 log<sub>10</sub> IU/mL at week 12 was invariably associated with treatment failure<sup>[72]</sup>. Consequently, the cost-effectiveness of PEG-IFN therapy may be improved by using a response-guided approach based on HBsAg kinetics at 12 wk of treatment.

The recommended duration of PEG-IFN therapy should also be prolonged. Results from a recent clinical trial demonstrated that treatment extension from 48 to 96 wk improved clinical outcome, significantly increasing rates of viral suppression, ALT normalization, and HBsAg clearance<sup>[73]</sup>. Furthermore, discontinuation rates did not increase, implying that a prolonged course of treatment was equally well tolerated. Consequently, guidelines should be updated to recommend extension of PEG-IFN treatment to 96 wk.

In patients where IFN therapy is contraindicated, or where there was a suboptimal response, treatment with entecavir or tenofovir is recommended by numerous meta-analyses and international guidelines due to its high efficacy, high barrier to resistance, and favourable safety profile<sup>[50,74]</sup>. However, there is little evidence for the use of lamivudine and tenofovir combination therapy in treatment refractory patients. In a population of human immunodeficiency virus (HIV) co-infected patients, combination therapy with lamivudine and tenofovir did not increase the probability of HBV DNA suppression<sup>[75]</sup>. There is more evidence for the use of combination therapy in patients harbouring patterns of viral resistance. In a population with mixed patterns of nucleos(t)ide analogue resistance, the majority of patients treated with entecavir and tenofovir achieved undetectable HBV DNA levels<sup>[76]</sup>. However, tenofovir monotherapy is also highly effective in multi-drug resistant strains of HBV<sup>[77]</sup>. Consequently, further studies are needed to determine whether combination therapy offers any advantage over monotherapy alone.

The management guidelines differ for patients with decompensated liver disease. Notably, entecavir should be offered as first-line treatment instead of PEG-IFN, due to the high risk of serious infections and hepatic failure. PEG-IFN is also contraindicated in pregnancy due to teratogenic effects arising from inhibition of cell prolifer-

ation. Instead, the category B antiviral drugs lamivudine or tenofovir should be used to reduce vertical transmission<sup>[78]</sup>. Both topics have been extensively reviewed elsewhere.

## COMBINATION THERAPY - IS TWO BETTER THAN ONE?

Combination therapy offers the potential of achieving long-term viral suppression whilst avoiding drug resistance, and has been shown to be beneficial in patients with HIV and chronic HCV. Consequently, numerous combinations of antivirals have been trialed in the treatment of CHB (Table 5). Current guidelines recommend first-line treatment of CHB/HIV co-infection with emtricitabine and tenofovir combination therapy<sup>[79]</sup>. However, the use of combination therapy *de novo*, or in cases where there is suboptimal response or drug resistance is more controversial and subject to ongoing studies.

At present, there is no evidence supporting *de novo* combination therapy with nucleos(t)ide analogues that have a high barrier of resistance. A large randomized-controlled trial found the antiviral efficacy of entecavir monotherapy to be comparable to that of entecavir plus tenofovir combination therapy in a mixed-population of nucleos(t)ide-naïve patients<sup>[80]</sup>. Where entecavir or tenofovir is not available as first-line treatment, the combination of lamivudine and adefovir may be used with similar efficacy<sup>[81]</sup>.

The combination of nucleos(t)ide analogues with PEG-IFN therapy is a more promising strategy. In phase III clinical trials, the combination of PEG-IFN and lamivudine induced a significantly greater decline in HBV DNA, as well as a higher rate of HBeAg and HBsAg seroconversion than lamivudine monotherapy<sup>[40,82]</sup>. In addition, there was a higher rate of sustained virological response, and lower rate of resistance when lamivudine was administered together with PEG-IFN, presumably because IFN is effective against HBV resistant mutants. Evidence also demonstrates that the *de novo* use of adefovir and PEG-IFN combination therapy lead to a greater decline in HBV DNA, HBsAg seroconversion, and intrahepatic cccDNA than PEG-IFN alone<sup>[83]</sup>. More importantly, a recent multi-centered randomized controlled

**Table 6** Antiviral resistance patterns and rescue therapy

Resistance	Mutation patterns	Treatment adaptation
LAM	M204V/I + L180M M204I M204V + L180M + V173L M204I + L180I Q215S + M204I/V + M204V I169T + V173L + L180M + M204V A181T T184S + M204I/V + L180M M204S + L180M	Switch to TDF or Switch to ETV + ADV combination therapy or Add ADV if TDF or ETV unavailable
ADV	A181V/T N236T A181V/T + N236T	Switch to TDF and add ETV
ADV + LAM	A181V/T + N236T L80V/I	Switch to TDF and add ETV
TBV	M204I/V ± L180M L80I/V ± L180M A181T/V	Switch to or add TDF
ETV	L180M + M204V/I ± I169T ± M250V L180M + M204V/I ± T184G ± S202I/G	Switch to or add TDF
TDF	No known mutations	N/A

Common mutation patterns, and rescue therapies suggested by the American Association for the Study of Liver Disease, Asian Pacific Association for the Study of Liver; European Association for the Study of Liver and National Institute for Health and Care Excellence. LAM: Lamivudine; ETV: Entecavir; ADV: Adefovir; TBV: Telbivudine; TDF: Tenofovir disoproxil fumarate.

study found that the use of entecavir, a nucleos(t)ide analogue recommended for first-line use by international guidelines, in combination with PEG-IFN- $\alpha$ -2a, increased clearance of HBeAg and HBsAg<sup>[84]</sup>. Hence combination therapy with nucleos(t)ide analogues and PEG-IFN may be an option for management in treatment-naïve patients.

In cases of inadequate HBV DNA suppression (> 2000 IU/mL), combination therapy may also be used to salvage treatment. One recent study demonstrated that the addition of lamivudine for 24 wk in patients with suboptimal response to adefovir monotherapy induced clearance of HBV DNA and HBeAg in approximately 35% of patients<sup>[85]</sup>. However, the addition of adefovir to entecavir, which has an inherently high barrier of resistance, did not increase virological response<sup>[86]</sup>. Given that all nucleos(t)ide analogues share the same viral target, the advantage of lamivudine and adefovir combination therapy is likely to be a consequence of reduced drug resistance. This hypothesis is supported by a 3-year study of 145 lamivudine-resistant patients, where treatment with the combination of adefovir and lamivudine induced clearance of HBV DNA in 80% of patients<sup>[87]</sup>. Moreover, all patients remained free of virological or clinical breakthroughs, with the 1-, 2-, 3- and 4-year cumulative rates of drug resistance at 1%, 2%, 4%, and 4% respectively<sup>[87]</sup>. Thus combination therapy is useful for rescuing therapy in patients with suboptimal treatment response or drug resistance.

## DRUG RESISTANCE - THE NEED FOR RESCUE THERAPY

HBV reverse transcriptase is an error-prone enzyme - its low fidelity, and lack of proofreading activity leads to a mutation rate 10-fold higher than that of other DNA viruses<sup>[88]</sup>. This fact, together with its high rate of replication ( $10^{12}$  virions/d), may result in the selection of HBV quasispecies containing mutations conferring drug-resistance.

Currently, effective monotherapy has been made possible with entecavir and tenofovir - two agents with high efficacy and low rates of resistance. Earlier use of less potent nucleos(t)ide analogues, however, has led to the emergence of multi-drug resistant strains of HBV. A summary of common mutation patterns, and the best available rescue treatment strategies is summarized in Table 6.

Presently, lamivudine resistance poses a major problem in the management of patients with CHB infection. Approximately 80% of patients develop lamivudine resistance after 5 years of therapy<sup>[89]</sup>. One retrospective review of various rescue strategies on patients with lamivudine resistance showed that switching to adefovir and entecavir combination therapy was the most effective, with 100% normalization in ALT, suppression of viral load, and no development of resistance<sup>[90]</sup>. In contrast, a switch to entecavir monotherapy was associated with drug resistance in 9% of patients after 2 years of therapy<sup>[91]</sup>. If costs prohibit the use of entecavir, add-on adefovir or oxymatrine may be considered as an alternative<sup>[90,92]</sup>.

In patients with multi-drug resistant HBV, studies have demonstrated that tenofovir monotherapy is able to induce long-term viral suppression. A prospective multi-centre open-label study of 60 patients who previously failed serial lamivudine and adefovir switch or add-on therapy showed that 4 years of continuous tenofovir resulted in 63% (38/60) patients achieving viral suppression by intention-to-treat analysis<sup>[77]</sup>. However, in patients with HIV co-infection, HBV strains with rtL180M, rtA194T, and rtM204V mutations displayed a 10-fold increase in the IC<sub>50</sub> value for tenofovir as compared with wildtype strains<sup>[93]</sup>. An early switch to combination therapy with tenofovir and emtricitabine, or tenofovir and entecavir, has been proven to be effective at inducing viral suppression in this difficult-to-treat population<sup>[76]</sup>. Moreover, patients should be investigated for drug compliance - recent studies have suggested that up to 40% of patients may not be fully adherent to their treatment regime, significantly influencing rates of viral suppression<sup>[94]</sup>. Consequently, all patients should be monitored closely for compliance, and development of resistance during antiviral therapy.

## DEFINING TREATMENT ENDPOINTS - WHEN DO WE STOP?

The ideal treatment endpoint is clearance of cccDNA

**Table 7 Treatment endpoints in clinical use**

Treatment	Description
Biochemical	Normalization of serum ALT
Virological	
HBeAg-positive	Loss of HBeAg, Anti-HBe antibodies, serum HBV-DNA < 2000 IU/mL
HBeAg-negative	Serum HBV-DNA < 2000 IU/mL
Complete	Biochemical and virological response with loss of serum HBsAg
Histological	Decrease in necroinflammatory activity without worsening in fibrosis

ALT: Alanine transaminase; HBV: Hepatitis B virus; HBeAg: Hepatitis B e antigen; HBsAg: Hepatitis B surface antigen.

from hepatocytes, but eradication of HBV is rarely achieved with currently available treatment. Instead, various biochemical, virological, and histological markers including the suppression of HBV DNA, HBeAg seroconversion, loss of HBsAg, ALT normalization, and improvement of inflammation or fibrosis on liver biopsy are used as surrogate endpoints to measure the efficacy of antiviral therapy (Table 7).

In HBeAg positive patients without cirrhosis, current NICE guidelines suggest stopping nucleos(t)ide analogue therapy 12 mo after HBeAg seroconversion. Whilst spontaneous or treatment-induced HBeAg seroconversion is associated with improved survival, it may not be a sufficient therapeutic endpoint in Asian populations, in which risks for disease progression are strongly related to viremia independent of HBeAg status<sup>[95]</sup>. Hence complete suppression of HBV DNA may be a better endpoint.

In HBeAg negative patients without cirrhosis, the ideal treatment endpoint is considerably less clear. It is possible that nucleos(t)ide analogue therapy may be safely withdrawn after achieving complete HBV DNA suppression. However, studies have shown that 90% of patients relapse within a year of discontinuing nucleos(t)ide analogue therapy<sup>[96]</sup>. The rate of relapse may be reduced by prolonging antiviral consolidation therapy<sup>[97]</sup>, or by continuing treatment indefinitely until HBsAg seroconversion is achieved. Interferon- and nucleos(t)ide analogue therapy-induced HBsAg seroclearance is durable after treatment discontinuation, with less than 5% of patients experiencing either virological or biochemical relapse<sup>[98]</sup>. Hence current NICE guidelines recommend stopping nucleos(t)ide treatment 12 mo after achieving both undetectable HBV DNA and HBsAg seroconversion in non-cirrhotic patients. But given only 5% of HBeAg negative patients clear HBsAg after 5 years of continuous therapy, life-long treatment is often required. Further research is needed to establish a safe and realistic endpoint for CHB therapy.

## FUTURE TREATMENTS - IS A CURE ON THE HORIZON?

Current treatments for CHB infection based on nucleos(t)ide

analogues effectively suppress viral replication to induce a functional cure. However, drug resistance is common, and life-long therapy is frequently required to prevent disease recurrence. Treatment with IFN- $\alpha$  is more likely to produce a sustained virological response, but is associated with numerous side effects, and is contraindicated in many patient populations. We believe that improvements in drug delivery systems, and development of drugs with different mechanisms of action will prevent resistance and improve viral clearance.

From a drug discovery point of view, rational drug targets might include several steps of the HBV viral cycle, such as viral entry, formation of cccDNA, viral genome integration, capsid assembly, viral envelopment, and exocytosis. In particular, the clearance of cccDNA, which mediates viral persistence in hepatocytes, is critical for preventing viral reactivation and disease recurrence following cessation of antiviral therapy. Since maintenance of the cccDNA pool results from a weak HBV-specific immune response, research should also focus on the development of immunomodulators. A number of new drug delivery systems and antiviral agents are currently under development (Table 8).

### Improvements in drug delivery systems

The efficacy of existing antiviral therapies is partly dependent on their pharmacokinetics. Hence implementation of liver-targeting drug delivery systems may improve drug efficacy by overcoming the development of resistance, and tolerability by promoting selective accumulation in the liver to limit systemic side effects. Dextran is a complex polysaccharide which may be used as a carrier molecule for tissue-specific delivery of drugs. Recently, it was reported that lamivudine-dextran conjugates selectively accumulate in hepatocytes and Kupffer cells, resulting in a seven-fold higher concentration compared to controls<sup>[99]</sup>. The inhibitory effects of adefovir loaded into nanoparticles on HBsAg, HBeAg, and serum HBV DNA levels *in vitro* were also significantly enhanced<sup>[100]</sup>. Nevertheless, these drugs do not eradicate cccDNA, which perpetuates infection. Consequently, liver-targeting drug delivery systems must be used in conjunction with new anti-HBV drug candidates to improve viral clearance.

### New drugs to current targets

Emtricitabine is a nucleos(t)ide analogue approved for the treatment of HIV with clinical activity against HBV. It significantly suppresses viral replication and improves liver histology compared to placebo, but is not used as a monotherapy due to high rates of resistance<sup>[101]</sup>. Instead, dual therapy with tenofovir and emtricitabine was found to produce high rates of HBeAg seroconversion and HBsAg loss, improving clinical outcomes for patients with decompensated HBV cirrhosis<sup>[102]</sup>. It is also the treatment of choice in patients with HIV co-infection<sup>[103]</sup>. However, further studies are needed to establish whether tenofovir plus emtricitabine offers any advantage over tenofovir monotherapy in treatment-naïve patients (Table 8).

**Table 8 Emerging pipeline drugs for chronic hepatitis B virus infection**

Drug name	Mechanism of action	Status
Nucleos(t)ide analogues		
Clevudine	Inhibits DNA polymerase	Partial approval
Emtricitabine	Inhibits DNA polymerase	FDA approved for HIV
Nucleos(t)ide analogue prodrugs		
Amdoxovir	Inhibits DNA polymerase	II (for HIV)
LB80380	Inhibits DNA polymerase	II b
Famciclovir	Inhibits DNA polymerase	Abandoned
Pradefovir	Inhibits DNA polymerase	Abandoned
Tenofovir	Inhibits DNA polymerase	II / III
alafenamide (GS 7340)		
MIV-210	Inhibits DNA polymerase	Abandoned
Non-nucleos(t)ide antivirals		
NOV-205 (BAM 205)	Unknown	Approved in Russia
Myrcludex-B	Inhibits viral entry	Ib/ II a
Bay 41-4109	Inhibits viral core formation	I
GLS4	Inhibits HBV viral core assembly	Pre-clinical
Rep 9 AC	Blocks HBsAg release	II
NVR-1221	Capsid inhibitor	Pre-clinical
Immunomodulators		
Pegylated interferon lambda	Cytokine modulating innate/adaptive immune response	I
GS-9620	TLR-7 agonist	Pre-clinical
Nitazoxanide	Unknown	II
EHT899	Immune enhancer	II
Therapeutic vaccines		
HBV core antigen vaccine	Enhance T cell response	I
HBV-EPV	Immunogenic	Withdrawn
ePA-44	Immunogenic	II
HI-8 HBV	Stimulates IFN- $\gamma$ producing T cells	II
Others		
$\beta$ -thujaplicinol	Blocks viral ribonuclease H activity	Pre-clinical
ARC520	RNA interference	I
Herbal bushen formula	Down-regulate CD4 <sup>+</sup> and CD25 <sup>+</sup> T cells	TCM

HBV: Hepatitis B virus; TCM: Traditional chinese medicine; HBsAg: Hepatitis B surface antigen; HIV: Human immunodeficiency virus; FDA: Food and Drug Administration; IFN: Interferon.

LB80380 is a nucleos(t)ide analogue prodrug which has potent anti-HBV activity in treatment-naïve patients<sup>[104]</sup>, as well as *in vitro* activity against mutant strains resistant to lamivudine, adefovir, entecavir, and telbivudine<sup>[105]</sup>. Additionally, 90 mg and 150 mg of LB80380 daily were found to be non-inferior to entecavir 0.5 mg daily at 48-wk in a phase II b study of treatment-naïve CHB patients<sup>[106]</sup>. Results from a 96-wk extension study (NCT01242787) are awaited (Table 9). With its high potency, effectiveness in lamivudine-resistant patients<sup>[107]</sup>, and good safety profile, LB80380 is a potential alternative agent for the treatment of CHB.

Clevudine is another nucleos(t)ide analogue which has potent antiviral activity against HBV<sup>[108]</sup>. Initial studies demonstrated that 24 wk of clevudine at 30 mg once daily achieved statistically significant differences in viral

suppression and durability compared to placebo<sup>[109]</sup>. Clevudine was also found to be superior to lamivudine in suppressing HBV replication<sup>[110]</sup>, but its efficacy is no higher than that of entecavir<sup>[111]</sup>. Long-term use is further limited by development of drug resistance<sup>[112]</sup>, and drug-induced toxic myopathy due to depletion of mitochondrial DNA<sup>[113]</sup>. Clevudine is currently approved for treatment of CHB in South Korea and the Philippines, but given the safety concerns and availability of more effective alternatives, clevudine is unlikely to have a role in treatment where entecavir and tenofovir are available. Thus new therapeutic approaches are needed.

### Emerging targets for HBV treatment

In developing novel anti-HBV therapies, many approaches including pharmacophore-based antiviral drug design and traditional high-throughput screening have been used. A number of lead compounds with new mechanisms of action are currently being evaluated *in vitro*, in animal models of CHB, and in clinical trials (Table 8).

The discovery of compounds which directly target cccDNA has been one of the major challenges to curing CHB infection. In an infected hepatocyte, HBV replication is regulated by the acetylation or methylation of histone proteins which surround the cccDNA minichromosome. The hSirt1/2 activator MC2791 and the JMJD3 inhibitor MC3119 suppressed both HBV replication and cccDNA transcription, providing a proof of concept that epigenetic modifiers may mediate persistent cccDNA silencing<sup>[114]</sup>. More recent data presented at the 2013 International Liver Congress suggests that antibody mediated stimulation of the lymphotoxin beta receptor in cell culture models exerts a strong and dose-dependent anti-HBV effect, including cccDNA in cells where HBV infection was established. In addition, zinc finger proteins designed to target the HBV enhancer region<sup>[115]</sup>, and various small-molecule compounds including CCC-0975 and CCC-0346<sup>[116]</sup> have been identified as inhibitors of cccDNA formation *in vitro*. Other replicative intermediates may also be targeted in CHB therapy. ARC-520 is a liver-tropic cholesterol-conjugated small interfering RNA which targets conserved HBV sequences to knockdown expression of replicative intermediates. In a chimpanzee with CHB infection, a low dose of ARC-520 induced rapid and multi-log repression of viral RNA, DNA, and key viral antigens including HBsAg and HBeAg with long duration of effect<sup>[117]</sup>. Collectively, this preliminary data suggests that targeting replicative intermediates such as cccDNA may form the basis of a cure for CHB infection.

Another promising therapeutic approach is to inhibit HBV entry into the hepatocyte, as this not only prevents cell-to-cell spread in established infection, but can replace HBIg to prevent infection. Myrcludex-B is a synthetic lipopeptide derived from the HBV envelope protein which inactivates the HBV pre-S1 receptor at picomolar concentrations<sup>[118]</sup>. In immunodeficient chimeric mice reconstituted with human hepatocytes (uPA/SCID-mice), myrcludex-B not only prevented viral spreading in between hepatocytes, but also reduced levels of HBV

**Table 9** A partial list of ongoing clinical trials

Phase	Trial identifier	Design	Drugs	Enrollment	Expected end date
I	NCT01872065	Double-blind, randomized	ARC520	44	October 2013
	NCT01590641	Double-blind, randomized	GS-9620	48	September 2013
II	NCT00524173	Open-label, randomized	TDF <i>vs</i> TDF + emtricitabine	100	January 2015
	NCT01204762	Double-blind, randomized	IFN- $\lambda$ + ETV	170	July 2017
III	NCT01242787	Open-label, randomized	LB80380	115	September 2012
	NCT01595685	Open-label, randomized	TBV <i>vs</i> ETV	184	December 2014
	NCT01369199	Open-label, randomized	8 wk ETV followed by 40 wk PEG-IFN- $\alpha$ -2a + ETV	250	May 2016
IV	NCT01804387	Open-label, randomized	TBV + ADV <i>vs</i> LAM + ADV	60	May 2014
	NCT01906580	Open-label, randomized	PEG-IFN- $\alpha$ -2a <i>vs</i> ETV	105	July 2014

Full details of the clinical trials may be found on the United States National Institutes of Health. Available from: URL: <http://www.clinicaltrials.gov>. LAM: Lamivudine; ETV: Entecavir; ADV: Adefovir; TBV: Telbivudine; TDF: Tenofovir disoproxil fumarate; PEG-IFN: Pegylated interferon.

DNA, HBsAg and cccDNA compared to untreated mice<sup>[119]</sup>. Results from a phase Ib/IIa clinical trial are expected at the end of 2013. Initial data for REP 9AC', a nucleic acid-based polymer which clears serum HBsAg by blocking HBV subviral particle formation and release, is also promising. A proof-of-concept study in 8 patients demonstrated that REP 9AC' reduced serum HBV DNA, inducing HBsAg seroclearance and development of anti-HBs to sustain virological response over 24 mo of follow-up off treatment. REP 9AC' also potentiated the immunostimulatory effects of PEG-IFN- $\alpha$ , generating anti-HBs titres comparable to those seen in healthy individuals with a strong vaccine response, raising the possibility of durable immunological control<sup>[120]</sup>.

The immune system is inherently capable of controlling HBV infection, as evidenced by the 90% of adults who resolve acute HBV infection, and by the resolution of CHB following bone marrow transplantation from an immune donor<sup>[121]</sup>. An increasing emphasis is thus placed on developing immunomodulatory therapy. Immunotherapeutic strategies for CHB include exogenous administration of cytokines with antiviral activity, and stimulation of the host T cell immune response. However, the antiviral effects of therapeutic vaccines have been disappointing. S and pre-S antigen vaccines, as well as T cell specific vaccines have all failed to achieve significant viral clearance<sup>[122,123]</sup>. Greater success has been found with IFN- $\lambda$  and GS-9620. IFN- $\lambda$  has potent anti-HBV activity *in vitro* and in transgenic mice<sup>[124,125]</sup>, and was shown in a phase IIb trial to cause significantly greater suppression of HBV DNA and HBsAg than IFN- $\lambda$ <sup>[126]</sup>. Critically, it also has improved tolerability due to the limited distribution of  $\lambda$  receptors outside the liver<sup>[127]</sup>. GS-9620 is an orally bioavailable small-molecule which activates Toll-like receptor 7 signalling. Administration of GS-9620 thrice weekly for 4 wk at 1 mg/kg, followed by 2 mg/kg, resulted in a prolonged 2.2 log unit reduction in HBV DNA in all three chronically infected chimpanzees<sup>[128]</sup>. Clinical trials for both IFN- $\lambda$  and GS-9620 are currently on going (Tables 8 and 9).

## CONCLUSION

Patients with minimal disease should not be treated. Con-

versely, patients at risk of developing complications such as cirrhosis and HCC should receive antiviral therapy. In patients with predictive factors of favourable response, a finite course of PEG-IFN should be trialed as first-line treatment. However, the majority of patients still require life-long treatment with nucleos(t)ide analogues, which is associated with substantial costs and a high risk for developing antiviral resistance. In our opinion, further studies are required to identify mechanisms by which the low viral load in the persistent carrier state may be eliminated. New strategies will likely favour an immunomodulatory approach, and or involve eradication of the pool of cccDNA from infected hepatocytes.

## REFERENCES

- 1 **McMahon BJ.** Epidemiology and natural history of hepatitis B. *Semin Liver Dis* 2005; **25** Suppl 1: 3-8 [PMID: 16103976 DOI: 10.1055/s-2005-915644]
- 2 **Lozano R,** Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, Bennett D, Bhalla K, Bikbov B, Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C, Burch M, Burney P, Carapetis J, Chen H, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahodwala N, De Leo D, Degenhardt L, Delossantos A, Denenberg J, Des Jarlais DC, Dharmaratne SD, Dorsey ER, Driscoll T, Duber H, Ebel B, Erwin PJ, Espindola P, Ezzati M, Feigin V, Flaxman AD, Forouzanfar MH, Fowkes FGR, Franklin R, Fransen M, Freeman MK, Gabriel SE, Gakidou E, Gaspari F, Gillum RF, Gonzalez-Medina D, Halasa YA, Haring D, Harrison JE, Havmoeller R, Hay RJ, Hoen B, Hotez PJ, Hoy D, Jacobsen KH, James SL, Jasrasaria R, Jayaraman S, Johns N, Karthikeyan G, Kassebaum N, Keren A, Khoo J-P, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Lipnick M, Lipshultz SE, Ohno SL, Mabweijano J, MacIntyre MF, Mallinger L, March L, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM, McAnulty JH, McDermott MM, McGrath J, Mensah GA, Merriman TR, Michaud C, Miller M, Miller TR, Mock C, Mocumbi AO, Mokdad AA, Moran A, Mulholland K, Nair MN, Naldi L, Narayan KMV, Nasseri K, Norman P, O'Donnell M, Omer SB, Ortblad K, Osborne R, Ozgediz D, Pahari B, Pandian JD, Rivero AP, Padilla RP, Perez-Ruiz F, Perico N, Phillips D, Pierce K, Pope CA, Porri E, Pourmalek F, Raju M, Ranganathan D, Rehm JT, Rein



- DB, Remuzzi G, Rivara FP, Roberts T, De León FR, Rosenfeld LC, Rushton L, Sacco RL, Salomon JA, Sampson U, Sanman E, Schwebel DC, Segui-Gomez M, Shepard DS, Singh D, Singleton J, Sliwa K, Smith E, Steer A, Taylor JA, Thomas B, Tleyjeh IM, Towbin JA, Truelsen T, Undurraga EA, Venkatasubramanian N, Vijayakumar L, Vos T, Wagner GR, Wang M, Wang W, Watt K, Weinstock MA, Weintraub R, Wilkinson JD, Woolf AD, Wulf S, Yeh P-H, Yip P, Zabetian A, Zheng Z-J, Lopez AD, Murray CJL, AlMazroa MA, Memish ZA. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380**: 2095-128 [PMID: 23245604 DOI: 10.1016/S0140-6736(12)61728-0]
- 3 **Liaw YF**, Tai DI, Chu CM, Chen TJ. The development of cirrhosis in patients with chronic type B hepatitis: a prospective study. *Hepatology* 1988; **8**: 493-496 [PMID: 3371868 DOI: 10.1002/hep.1840080310]
  - 4 **Fattovich G**, Stroffolini T, Zagni I, Donato F. Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterology* 2004; **127**: S35-S50 [PMID: 15508101 DOI: 10.1053/j.gastro.2004.09.014]
  - 5 **Lok AS**, McMahon BJ. Chronic hepatitis B: update 2009. *Hepatology* 2009; **50**: 661-662 [PMID: 19714720 DOI: 10.1002/hep.23190]
  - 6 **Liaw YF**, Kao JH, Piratvisuth T, Chan HLY, Chien RN, Liu CJ, Gan E, Locarnini S, Lim SG, Han KH, Amarapurkar D, Cooksley G, Jafri W, Mohamed R, Hou JL, Chuang WL, Lesmana L, Sollano JD, Suh DJ, Omata M. Asian-Pacific consensus statement on the management of chronic hepatitis B: a 2012 update. *Hepatol Int* 2012; **3**: 531-561 [DOI: 10.1007/s12072-012-9365-4]
  - 7 **European Association For The Study Of The Liver**. EASL clinical practice guidelines: Management of chronic hepatitis B virus infection. *J Hepatol* 2012; **57**: 167-185 [PMID: 22436845 DOI: 10.1016/j.jhep.2012.02.010]
  - 8 **Sarri G**, Westby M, Birmingham S, Hill-Cawthorne G, Thomas H. Diagnosis and management of chronic hepatitis B in children, young people, and adults: summary of NICE guidance. *BMJ* 2013; **346**: f3893 [PMID: 23804177 DOI: 10.1136/bmj.f3893]
  - 9 **Robinson WS**, Lutwick LI. The virus of hepatitis, type B (first of two parts). *N Engl J Med* 1976; **295**: 1168-1175 [PMID: 62280 DOI: 10.1056/NEJM197611182952105]
  - 10 **Fung SK**, Lok AS. Hepatitis B virus genotypes: do they play a role in the outcome of HBV infection? *Hepatology* 2004; **40**: 790-792 [PMID: 15382157 DOI: 10.1002/hep.1840400407]
  - 11 **Yan H**, Zhong G, Xu G, He W, Jing Z, Gao Z, Huang Y, Qi Y, Peng B, Wang H, Fu L, Song M, Chen P, Gao W, Ren B, Sun Y, Cai T, Feng X, Sui J, Li W. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *Elife* 2012; **1**: e00049 [PMID: 23150796 DOI: 10.7554/eLife.00049]
  - 12 **Tuttleman JS**, Pourcel C, Summers J. Formation of the pool of covalently closed circular viral DNA in hepadnavirus-infected cells. *Cell* 1986; **47**: 451-460 [PMID: 3768961 DOI: 10.1016/0092-8674(86)90602-1]
  - 13 **Roingard P**, Lu SL, Sureau C, Freschlin M, Arbeille B, Essex M, Romet-Lemonne JL. Immunocytochemical and electron microscopic study of hepatitis B virus antigen and complete particle production in hepatitis B virus DNA transfected HepG2 cells. *Hepatology* 1990; **11**: 277-285 [PMID: 2407629 DOI: 10.1002/hep.1840110219]
  - 14 **Liaw YF**, Chu CM, Lin DY, Sheen IS, Yang CY, Huang MJ. Age-specific prevalence and significance of hepatitis B e antigen and antibody in chronic hepatitis B virus infection in Taiwan: a comparison among asymptomatic carriers, chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma. *J Med Virol* 1984; **13**: 385-391 [PMID: 6330293 DOI: 10.1002/jmv.1890130410]
  - 15 **Yuen MF**, Sablon E, Yuan HJ, Wong DK, Hui CK, Wong BC, Chan AO, Lai CL. Significance of hepatitis B genotype in acute exacerbation, HBeAg seroconversion, cirrhosis-related complications, and hepatocellular carcinoma. *Hepatology* 2003; **37**: 562-567 [PMID: 12601354 DOI: 10.1053/jhep.2003.50098]
  - 16 **Beasley RP**, Hwang LY, Lee GC, Lan CC, Roan CH, Huang FY, Chen CL. Prevention of perinatally transmitted hepatitis B virus infections with hepatitis B immune globulin and hepatitis B vaccine. *Lancet* 1983; **2**: 1099-1102 [PMID: 6138642 DOI: 10.1016/S0140-6736(83)90624-4]
  - 17 **Ganem D**, Prince AM. Hepatitis B virus infection--natural history and clinical consequences. *N Engl J Med* 2004; **350**: 1118-1129 [PMID: 15014185 DOI: 10.1056/NEJMra031087]
  - 18 **Guidotti LG**, Ishikawa T, Hobbs MV, Matzke B, Schreiber R, Chisari FV. Intracellular inactivation of the hepatitis B virus by cytotoxic T lymphocytes. *Immunity* 1996; **4**: 25-36 [PMID: 8574849 DOI: 10.1016/S1074-7613(00)80295-2]
  - 19 **Jung MC**, Spengler U, Schraud W, Hoffmann R, Zachoval R, Eisenburg J, Eichenlaub D, Riethmüller G, Paumgartner G, Ziegler-Heitbrock HW. Hepatitis B virus antigen-specific T-cell activation in patients with acute and chronic hepatitis B. *J Hepatol* 1991; **13**: 310-317 [PMID: 1808224 DOI: 10.1016/0168-8278(91)90074-L]
  - 20 **Lopes AR**, Kellam P, Das A, Dunn C, Kwan A, Turner J, Peppas D, Gilson RJ, Gehring A, Bertoletti A, Maini MK. Bim-mediated deletion of antigen-specific CD8 T cells in patients unable to control HBV infection. *J Clin Invest* 2008; **118**: 1835-1845 [PMID: 18398508 DOI: 10.1172/JCI33402]
  - 21 **Schurich A**, Khanna P, Lopes AR, Han KJ, Peppas D, Micco L, Nebbia G, Kennedy PT, Geretti AM, Dusheiko G, Maini MK. Role of the coinhibitory receptor cytotoxic T lymphocyte antigen-4 on apoptosis-Prone CD8 T cells in persistent hepatitis B virus infection. *Hepatology* 2011; **53**: 1494-1503 [PMID: 21360567 DOI: 10.1002/hep.24249]
  - 22 **Sobao Y**, Tomiyama H, Sugi K, Tokunaga M, Ueno T, Saito S, Fujiyama S, Morimoto M, Tanaka K, Takiguchi M. The role of hepatitis B virus-specific memory CD8 T cells in the control of viral replication. *J Hepatol* 2002; **36**: 105-115 [PMID: 11804672 DOI: 10.1016/S0168-8278(01)00264-1]
  - 23 **Wang S**, Zhang X, Ju Y, Zhao B, Yan X, Hu J, Shi L, Yang L, Ma Z, Chen L, Liu Y, Duan Z, Chen X, Meng S. MicroRNA-146a feedback suppresses T cell immune function by targeting Stat1 in patients with chronic hepatitis B. *J Immunol* 2013; **191**: 293-301 [PMID: 23698745 DOI: 10.4049/jimmunol.1202100]
  - 24 **Stoop JN**, van der Molen RG, Baan CC, van der Laan LJ, Kuipers EJ, Kusters JG, Janssen HL. Regulatory T cells contribute to the impaired immune response in patients with chronic hepatitis B virus infection. *Hepatology* 2005; **41**: 771-778 [PMID: 15791617 DOI: 10.1002/hep.20649]
  - 25 **Op den Brouw ML**, Binda RS, van Roosmalen MH, Protzer U, Janssen HL, van der Molen RG, Woltman AM. Hepatitis B virus surface antigen impairs myeloid dendritic cell function: a possible immune escape mechanism of hepatitis B virus. *Immunology* 2009; **126**: 280-289 [PMID: 18624732 DOI: 10.1111/j.1365-2567.2008.02896.x]
  - 26 **Dunn C**, Brunetto M, Reynolds G, Christophides T, Kennedy PT, Lampertico P, Das A, Lopes AR, Borrow P, Williams K, Humphreys E, Afford S, Adams DH, Bertoletti A, Maini MK. Cytokines induced during chronic hepatitis B virus infection promote a pathway for NK cell-mediated liver damage. *J Exp Med* 2007; **204**: 667-680 [PMID: 17353365 DOI: 10.1084/jem.20061287]
  - 27 **Kägi D**, Vignaux F, Ledermann B, Bürki K, Depraetere V, Nagata S, Hengartner H, Golstein P. Fas and perforin pathways as major mechanisms of T cell-mediated cytotoxicity. *Science* 1994; **265**: 528-530 [PMID: 7518614 DOI: 10.1126/science.7518614]
  - 28 **Maini MK**, Boni C, Lee CK, Larrubia JR, Reignat S, Ogg GS, King AS, Herberg J, Gilson R, Alisa A, Williams R, Vergani

- D, Naoumov NV, Ferrari C, Bertolotti A. The role of virus-specific CD8(+) cells in liver damage and viral control during persistent hepatitis B virus infection. *J Exp Med* 2000; **191**: 1269-1280 [PMID: 10770795 DOI: 10.1084/jem.191.8.1269]
- 29 **Castera L**, Fornis X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. *J Hepatol* 2008; **48**: 835-847 [PMID: 18334275 DOI: 10.1016/j.jhep.2008.02.008]
- 30 **Iloeje UH**, Yang HI, Su J, Jen CL, You SL, Chen CJ. Predicting cirrhosis risk based on the level of circulating hepatitis B viral load. *Gastroenterology* 2006; **130**: 678-686 [PMID: 16530509 DOI: 10.1053/j.gastro.2005.11.016]
- 31 **Chen CJ**, Yang HI, Su J, Jen CL, You SL, Lu SN, Huang GT, Iloeje UH. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. *JAMA* 2006; **295**: 65-73 [PMID: 16391218 DOI: 10.1001/jama.295.1.65]
- 32 **de Jongh FE**, Janssen HL, de Man RA, Hop WC, Schalm SW, van Blankenstein M. Survival and prognostic indicators in hepatitis B surface antigen-positive cirrhosis of the liver. *Gastroenterology* 1992; **103**: 1630-1635 [PMID: 1426884]
- 33 **Tseng TC**, Liu CJ, Su TH, Yang HC, Wang CC, Chen CL, Kuo SF, Liu CH, Chen PJ, Chen DS, Kao JH. Young chronic hepatitis B patients with nucleos(t)ide analogue-induced hepatitis B e antigen seroconversion have a higher risk of HBV reactivation. *J Infect Dis* 2012; **206**: 1521-1531 [PMID: 22966125 DOI: 10.1093/infdis/jis569]
- 34 **Papatheodoridis GV**, Manolakopoulos S, Liaw YF, Lok A. Follow-up and indications for liver biopsy in HBeAg-negative chronic hepatitis B virus infection with persistently normal ALT: a systematic review. *J Hepatol* 2012; **57**: 196-202 [PMID: 22450396 DOI: 10.1016/j.jhep.2011.11.030]
- 35 **Tseng TC**, Liu CJ, Yang HC, Su TH, Wang CC, Chen CL, Hsu CA, Kuo SF, Liu CH, Chen PJ, Chen DS, Kao JH. Serum hepatitis B surface antigen levels help predict disease progression in patients with low hepatitis B virus loads. *Hepatology* 2013; **57**: 441-450 [PMID: 22941922 DOI: 10.1002/hep.26041]
- 36 **Tseng TC**, Liu CJ, Yang HC, Su TH, Wang CC, Chen CL, Kuo SF, Liu CH, Chen PJ, Chen DS, Kao JH. High levels of hepatitis B surface antigen increase risk of hepatocellular carcinoma in patients with low HBV load. *Gastroenterology* 2012; **142**: 1140-1149.e3; quiz e13-14 [PMID: 22333950 DOI: 10.1053/j.gastro.2012.02.007]
- 37 **Kim HC**, Nam CM, Jee SH, Han KH, Oh DK, Suh I. Normal serum aminotransferase concentration and risk of mortality from liver diseases: prospective cohort study. *BMJ* 2004; **328**: 983 [PMID: 15028636 DOI: 10.1136/bmj.38050.593634.63]
- 38 **Cooksley WG**, Piratvisuth T, Lee SD, Mahachai V, Chao YC, Vanwadee T, Chutaputti A, Chang WY, Zahm FE, Pluck N. Peginterferon alpha-2a (40 kDa): an advance in the treatment of hepatitis B e antigen-positive chronic hepatitis B. *J Viral Hepat* 2003; **10**: 298-305 [PMID: 12823597 DOI: 10.1046/j.1365-2893.2003.00450.x]
- 39 **Marcellin P**, Lau GK, Bonino F, Farci P, Hadziyannis S, Jin R, Lu ZM, Piratvisuth T, Germanidis G, Yurdaydin C, Diago M, Gurel S, Lai MY, Button P, Pluck N. Peginterferon alfa-2a alone, lamivudine alone, and the two in combination in patients with HBeAg-negative chronic hepatitis B. *N Engl J Med* 2004; **351**: 1206-1217 [PMID: 15371578 DOI: 10.1056/NEJMoa040431]
- 40 **Lau GK**, Piratvisuth T, Luo KX, Marcellin P, Thongsawat S, Cooksley G, Gane E, Fried MW, Chow WC, Paik SW, Chang WY, Berg T, Flisiak R, McCloud P, Pluck N. Peginterferon Alfa-2a, lamivudine, and the combination for HBeAg-positive chronic hepatitis B. *N Engl J Med* 2005; **352**: 2682-2695 [PMID: 15987917 DOI: 10.1056/NEJMoa043470]
- 41 **Janssen HL**, van Zonneveld M, Senturk H, Zeuzem S, Akarca US, Cakaloglu Y, Simon C, So TM, Gerken G, de Man RA, Niesters HG, Zondervan P, Hansen B, Schalm SW. Pegylated interferon alfa-2b alone or in combination with lamivudine for HBeAg-positive chronic hepatitis B: a randomised trial. *Lancet* 2005; **365**: 123-129 [PMID: 15639293 DOI: 10.1016/S0140-6736(05)17701-0]
- 42 **van Nunen AB**, Hansen BE, Suh DJ, Löhr HF, Chemello L, Fontaine H, Heathcote J, Song BC, Janssen HL, de Man RA, Schalm SW. Durability of HBeAg seroconversion following antiviral therapy for chronic hepatitis B: relation to type of therapy and pretreatment serum hepatitis B virus DNA and alanine aminotransferase. *Gut* 2003; **52**: 420-424 [PMID: 12584227 DOI: 10.1136/gut.52.3.420]
- 43 **Lai CL**, Shouval D, Lok AS, Chang TT, Cheinquer H, Goodman Z, DeHertogh D, Wilber R, Zink RC, Cross A, Colonno R, Fernandes L. Entecavir versus lamivudine for patients with HBeAg-negative chronic hepatitis B. *N Engl J Med* 2006; **354**: 1011-1020 [PMID: 16525138 DOI: 10.1056/NEJMoa051287]
- 44 **Chang TT**, Gish RG, de Man R, Gadano A, Sollano J, Chao YC, Lok AS, Han KH, Goodman Z, Zhu J, Cross A, DeHertogh D, Wilber R, Colonno R, Apelian D. A comparison of entecavir and lamivudine for HBeAg-positive chronic hepatitis B. *N Engl J Med* 2006; **354**: 1001-1010 [PMID: 16525137 DOI: 10.1056/NEJMoa051285]
- 45 **Leung N**, Peng CY, Hann HW, Sollano J, Lao-Tan J, Hsu CW, Lesmana L, Yuen MF, Jeffers L, Sherman M, Min A, Mencarini K, Diva U, Cross A, Wilber R, Lopez-Talavera J. Early hepatitis B virus DNA reduction in hepatitis B e antigen-positive patients with chronic hepatitis B: A randomized international study of entecavir versus adefovir. *Hepatology* 2009; **49**: 72-79 [PMID: 19065670 DOI: 10.1002/hep.22658]
- 46 **Gish RG**, Lok AS, Chang TT, de Man RA, Gadano A, Sollano J, Han KH, Chao YC, Lee SD, Harris M, Yang J, Colonno R, Brett-Smith H. Entecavir therapy for up to 96 weeks in patients with HBeAg-positive chronic hepatitis B. *Gastroenterology* 2007; **133**: 1437-1444 [PMID: 17983800 DOI: 10.1053/j.gastro.2007.08.025]
- 47 **Chang TT**, Lai CL, Kew Yoon S, Lee SS, Coelho HS, Carrilho FJ, Poordad F, Halota W, Horsmans Y, Tsai N, Zhang H, Tenney DJ, Tamez R, Iloeje U. Entecavir treatment for up to 5 years in patients with hepatitis B e antigen-positive chronic hepatitis B. *Hepatology* 2010; **51**: 422-430 [PMID: 20049753 DOI: 10.1002/hep.23327]
- 48 **Zoutendijk R**, Reijnders JG, Zoulim F, Brown A, Mutimer DJ, Deterding K, Hofmann WP, Petersen J, Fasano M, Buti M, Berg T, Hansen BE, Sonneveld MJ, Wedemeyer H, Janssen HL. Virological response to entecavir is associated with a better clinical outcome in chronic hepatitis B patients with cirrhosis. *Gut* 2013; **62**: 760-765 [PMID: 22490523 DOI: 10.1136/gutjnl-2012-302024]
- 49 **Tenney DJ**, Rose RE, Baldick CJ, Pokornowski KA, Eggers BJ, Fang J, Wichroski MJ, Xu D, Yang J, Wilber RB, Colonno RJ. Long-term monitoring shows hepatitis B virus resistance to entecavir in nucleoside-naïve patients is rare through 5 years of therapy. *Hepatology* 2009; **49**: 1503-1514 [PMID: 19280622 DOI: 10.1002/hep.22841]
- 50 **Woo G**, Tomlinson G, Nishikawa Y, Kowgier M, Sherman M, Wong DK, Pham B, Ungar WJ, Einarson TR, Heathcote EJ, Krahn M. Tenofovir and entecavir are the most effective antiviral agents for chronic hepatitis B: a systematic review and Bayesian meta-analyses. *Gastroenterology* 2010; **139**: 1218-1229 [PMID: 20600036 DOI: 10.1053/j.gastro.2010.06.042]
- 51 **Marcellin P**, Heathcote EJ, Buti M, Gane E, de Man RA, Krastev Z, Germanidis G, Lee SS, Flisiak R, Kaita K, Manns M, Kotzev I, Tchernev K, Buggisch P, Weilert F, Kurdas OO, Shiffman ML, Trinh H, Washington MK, Sorbel J, Anderson J, Snow-Lampart A, Mondou E, Quinn J, Rousseau F. Tenofovir disoproxil fumarate versus adefovir dipivoxil for chronic hepatitis B. *N Engl J Med* 2008; **359**: 2442-2455 [PMID: 19052126 DOI: 10.1056/NEJMoa0802878]
- 52 **Heathcote EJ**, Marcellin P, Buti M, Gane E, De Man RA, Krastev Z, Germanidis G, Lee SS, Flisiak R, Kaita K, Manns

- M, Kotzev I, Tchernev K, Buggisch P, Weilert F, Kurdas OO, Shiffman ML, Trinh H, Gurel S, Snow-Lampart A, Borroto-Esoda K, Mondou E, Anderson J, Sorbel J, Rousseau F. Three-year efficacy and safety of tenofovir disoproxil fumarate treatment for chronic hepatitis B. *Gastroenterology* 2011; **140**: 132-143 [PMID: 20955704 DOI: 10.1053/j.gastro.2010.10.011]
- 53 **Marcellin P**, Gane E, Buti M, Afdhal N, Sievert W, Jacobson IM, Washington MK, Germanidis G, Flaherty JF, Schall RA, Bornstein JD, Kitrinis KM, Subramanian GM, McHutchison JG, Heathcote EJ. Regression of cirrhosis during treatment with tenofovir disoproxil fumarate for chronic hepatitis B: a 5-year open-label follow-up study. *Lancet* 2013; **381**: 468-475 [PMID: 23234725 DOI: 10.1016/S0140-6736(12)61425-1]
- 54 **Kitrinis KM**, Corsa A, Liu Y, Flaherty J, Snow-Lampart A, Marcellin P, Borroto-Esoda K, Miller MD. No detectable resistance to tenofovir disoproxil fumarate after 6 years of therapy in patients with chronic hepatitis B. *Hepatology* 2014; **59**: 434-442 [PMID: 23939953 DOI: 10.1002/hep.26686]
- 55 **Low TL**, Goldstein AL. Thymosins: structure, function and therapeutic applications. *Thymus* 1984; **6**: 27-42 [PMID: 6087503]
- 56 **Chan HL**, Tang JL, Tam W, Sung JJ. The efficacy of thymosin in the treatment of chronic hepatitis B virus infection: a meta-analysis. *Aliment Pharmacol Ther* 2001; **15**: 1899-1905 [PMID: 11736720 DOI: 10.1046/j.1365-2036.2001.01135.x]
- 57 **You J**, Zhuang L, Cheng HY, Yan SM, Yu L, Huang JH, Tang BZ, Huang ML, Ma YL, Chongsuvivatwong V, Sripilung H, Geater A, Qiao YW, Wu RX. Efficacy of thymosin alpha-1 and interferon alpha in treatment of chronic viral hepatitis B: a randomized controlled study. *World J Gastroenterol* 2006; **12**: 6715-6721 [PMID: 17075991]
- 58 **Chen XS**, Wang GJ, Cai X, Yu HY, Hu YP. Inhibition of hepatitis B virus by oxymatrine in vivo. *World J Gastroenterol* 2001; **7**: 49-52 [PMID: 11819732]
- 59 **Dong Y**, Xi H, Yu Y, Wang Q, Jiang K, Li L. Effects of oxymatrine on the serum levels of T helper cell 1 and 2 cytokines and the expression of the S gene in hepatitis B virus S gene transgenic mice: a study on the anti-hepatitis B virus mechanism of oxymatrine. *J Gastroenterol Hepatol* 2002; **17**: 1299-1306 [PMID: 12423275 DOI: 10.1046/j.1440-1746.2002.02885.x]
- 60 **Lu LG**, Zeng MD, Mao YM, Li JQ, Wan MB, Li CZ, Chen CW, Fu QC, Wang JY, She WM, Cai X, Ye J, Zhou XQ, Wang H, Wu SM, Tang MF, Zhu JS, Chen WX, Zhang HQ. Oxymatrine therapy for chronic hepatitis B: a randomized double-blind and placebo-controlled multi-center trial. *World J Gastroenterol* 2003; **9**: 2480-2483 [PMID: 14606080]
- 61 **Kao JH**, Wu NH, Chen PJ, Lai MY, Chen DS. Hepatitis B genotypes and the response to interferon therapy. *J Hepatol* 2000; **33**: 998-1002 [PMID: 11131465 DOI: 10.1016/S0168-8278(00)80135-X]
- 62 **Wai CT**, Chu CJ, Hussain M, Lok AS. HBV genotype B is associated with better response to interferon therapy in HBeAg(+) chronic hepatitis than genotype C. *Hepatology* 2002; **36**: 1425-1430 [PMID: 12447868 DOI: 10.1053/jhep.2002.37139]
- 63 **Erhardt A**, Blondin D, Hauck K, Sagir A, Kohnle T, Heintges T, Häussinger D. Response to interferon alfa is hepatitis B virus genotype dependent: genotype A is more sensitive to interferon than genotype D. *Gut* 2005; **54**: 1009-1013 [PMID: 15951551 DOI: 10.1136/gut.2004.060327]
- 64 **Flink HJ**, van Zonneveld M, Hansen BE, de Man RA, Schalm SW, Janssen HL. Treatment with Peg-interferon alpha-2b for HBeAg-positive chronic hepatitis B: HBeAg loss is associated with HBV genotype. *Am J Gastroenterol* 2006; **101**: 297-303 [PMID: 16454834 DOI: 10.1111/j.1572-0241.2006.00418.x]
- 65 **Buster EH**, Flink HJ, Cakaloglu Y, Simon K, Trojan J, Tabak F, So TM, Feinman SV, Mach T, Akarca US, Schutten M, Tieleman W, van Vuuren AJ, Hansen BE, Janssen HL. Sustained HBeAg and HBsAg loss after long-term follow-up of HBeAg-positive patients treated with peginterferon alpha-2b. *Gastroenterology* 2008; **135**: 459-467 [PMID: 18585385 DOI: 10.1053/j.gastro.2008.05.031]
- 66 **Buster EH**, Hansen BE, Lau GK, Piratvisuth T, Zeuzem S, Steyerberg EW, Janssen HL. Factors that predict response of patients with hepatitis B e antigen-positive chronic hepatitis B to peginterferon-alfa. *Gastroenterology* 2009; **137**: 2002-2009 [PMID: 19737568 DOI: 10.1053/j.gastro.2009.08.061]
- 67 **Lampertico P**, Viganò M, Cheroni C, Facchetti F, Invernizzi F, Valveri V, Soffredini R, Abrignani S, De Francesco R, Colombo M. IL28B polymorphisms predict interferon-related hepatitis B surface antigen seroclearance in genotype D hepatitis B e antigen-negative patients with chronic hepatitis B. *Hepatology* 2013; **57**: 890-896 [PMID: 22473858 DOI: 10.1002/hep.25749]
- 68 **Sonneveld MJ**, Wong VW, Woltman AM, Wong GL, Cakaloglu Y, Zeuzem S, Buster EH, Uitterlinden AG, Hansen BE, Chan HL, Janssen HL. Polymorphisms near IL28B and serologic response to peginterferon in HBeAg-positive patients with chronic hepatitis B. *Gastroenterology* 2012; **142**: 513-520. e1 [PMID: 22108195 DOI: 10.1053/j.gastro.2011.11.025]
- 69 **Wu X**, Xin Z, Zhu X, Pan L, Li Z, Li H, Liu Y. Evaluation of susceptibility locus for response to interferon- $\alpha$  based therapy in chronic hepatitis B patients in Chinese. *Antiviral Res* 2012; **93**: 297-300 [PMID: 22209781 DOI: 10.1016/j.antiviral.2011.12.009]
- 70 **Piratvisuth T**, Marcellin P, Popescu M, Kapprell HP, Rothe V, Lu ZM. Hepatitis B surface antigen: association with sustained response to peginterferon alfa-2a in hepatitis B e antigen-positive patients. *Hepatol Int* 2011 Jun 24; Epub ahead of print [PMID: 21701902 DOI: 10.1007/s12072-011-9280-0]
- 71 **Liaw YF**, Jia JD, Chan HL, Han KH, Tanwandee T, Chuang WL, Tan DM, Chen XY, Gane E, Piratvisuth T, Chen L, Xie Q, Sung JJ, Wat C, Bernaards C, Cui Y, Marcellin P. Shorter durations and lower doses of peginterferon alfa-2a are associated with inferior hepatitis B e antigen seroconversion rates in hepatitis B virus genotypes B or C. *Hepatology* 2011; **54**: 1591-1599 [PMID: 22045673 DOI: 10.1002/hep.24555]
- 72 **Rijckborst V**, Hansen BE, Cakaloglu Y, Ferenci P, Tabak F, Akdogan M, Simon K, Akarca US, Flisiak R, Verhey E, Van Vuuren AJ, Boucher CA, ter Borg MJ, Janssen HL. Early on-treatment prediction of response to peginterferon alfa-2a for HBeAg-negative chronic hepatitis B using HBsAg and HBV DNA levels. *Hepatology* 2010; **52**: 454-461 [PMID: 20683945 DOI: 10.1002/hep.23722]
- 73 **Lampertico P**, Viganò M, Di Costanzo GG, Sagnelli E, Fasano M, Di Marco V, Boninsegna S, Farci P, Fargion S, Giuberti T, Iannacone C, Regep L, Massetto B, Facchetti F, Colombo M. Randomised study comparing 48 and 96 weeks peginterferon  $\alpha$ -2a therapy in genotype D HBeAg-negative chronic hepatitis B. *Gut* 2013; **62**: 290-298 [PMID: 22859496 DOI: 10.1136/gutjnl-2011-301430]
- 74 **Wiens A**, Lenzi L, Venson R, Correr CJ, Rotta I, Pedroso ML, Pontarolo R. Comparative efficacy of oral nucleoside or nucleotide analog monotherapy used in chronic hepatitis B: a mixed-treatment comparison meta-analysis. *Pharmacotherapy* 2013; **33**: 144-151 [PMID: 23359454 DOI: 10.1002/phar.1188]
- 75 **Schmutz G**, Nelson M, Lutz T, Sheldon J, Bruno R, von Boehm F, Hoffmann C, Rockstroh J, Stoehr A, Wolf E, Soriano V, Berger F, Berg T, Carlebach A, Schwarze-Zander C, Schürmann D, Jaeger H, Mauss S. Combination of tenofovir and lamivudine versus tenofovir after lamivudine failure for therapy of hepatitis B in HIV-coinfection. *AIDS* 2006; **20**: 1951-1954 [PMID: 16988516 DOI: 10.1097/01.aids.0000247116.89455.5d]
- 76 **Petersen J**, Ratziu V, Buti M, Janssen HL, Brown A, Lampertico P, Schollmeyer J, Zoulim F, Wedemeyer H, Sternecker M, Berg T, Sarrazin C, Lutgehetmann M, Buggisch P. Entecavir plus tenofovir combination as rescue therapy in pre-treated chronic hepatitis B patients: an international multicenter cohort study. *J Hepatol* 2012; **56**: 520-526 [PMID: 22037226 DOI: 10.1016/j.jhep.2012.05.011]

- 10.1016/j.jhep.2011.09.018]
- 77 **Lim LY**, Patterson S, George J, Strasser SI, Lee AU, Sievert W, Nicoll AJ, Roberts SK, Desmond P V, Bowden S, Thompson AJ, Locarnini S, Angus PW. Tenofovir Rescue Therapy Achieves Long-Term Suppression of HBV Replication in Patients with Multi-Drug Resistant HBV: 4 Year Follow-Up of the TDF109 Cohort. *Hepatology* 2012; **56**: 368A
  - 78 **Guclu E**, Karabay O. Choice of drugs in the treatment of chronic hepatitis B in pregnancy. *World J Gastroenterol* 2013; **19**: 1671-1672 [PMID: 23539671 DOI: 10.3748/wjg.v19.i10.1671]
  - 79 **Dore GJ**, Cooper DA, Pozniak AL, DeJesus E, Zhong L, Miller MD, Lu B, Cheng AK. Efficacy of tenofovir disoproxil fumarate in antiretroviral therapy-naïve and -experienced patients coinfecting with HIV-1 and hepatitis B virus. *J Infect Dis* 2004; **189**: 1185-1192 [PMID: 15031786 DOI: 10.1086/380398]
  - 80 **Lok AS**, Trinh H, Carosi G, Akarca US, Gadano A, Habersetzer F, Sievert W, Wong D, Lovegren M, Cohen D, Llamoso C. Efficacy of entecavir with or without tenofovir disoproxil fumarate for nucleos(t)ide-naïve patients with chronic hepatitis B. *Gastroenterology* 2012; **143**: 619-628.e1 [PMID: 22643350 DOI: 10.1053/j.gastro.2012.05.037]
  - 81 **Wang LC**, Chen EQ, Cao J, Liu L, Zheng L, Li DJ, Xu L, Lei XZ, Liu C, Tang H. De novo combination of lamivudine and adefovir versus entecavir monotherapy for the treatment of naïve HBeAg-negative chronic hepatitis B patients. *Hepatol Int* 2011; **5**: 671-676 [PMID: 21484140 DOI: 10.1007/s12072-010-9243-x]
  - 82 **Barbaro G**, Zechini F, Pellicelli AM, Francavilla R, Scotto G, Bacca D, Bruno M, Babudieri S, Annese M, Matarazzo F, Di Stefano G, Barbarini G. Long-term efficacy of interferon alpha-2b and lamivudine in combination compared to lamivudine monotherapy in patients with chronic hepatitis B. An Italian multicenter, randomized trial. *J Hepatol* 2001; **35**: 406-411 [PMID: 11592603]
  - 83 **Wursthorn K**, Lutgehetmann M, Dandri M, Volz T, Buggisch P, Zollner B, Longerrich T, Schirmacher P, Metzler F, Zankel M, Fischer C, Currie G, Brosgart C, Petersen J. Peginterferon alpha-2b plus adefovir induce strong cccDNA decline and HBsAg reduction in patients with chronic hepatitis B. *Hepatology* 2006; **44**: 675-684 [PMID: 16941693 DOI: 10.1002/hep.21282]
  - 84 **Sonneveld MJ**, Xie Q, Zhang NP, Zhang Q, Fehmi T, Streinu-cercel A, Wang J, Idilman R, de Niet A, Diculescu M, van Vuuren AJ, Verhey E, Hansen BE, Janssen HL. Adding peginterferon alfa-2a to entecavir increases HBsAg decline and HBeAg clearance - first results from a global randomized trial (ARES study). *Hepatology* 2012; **56**: S199A
  - 85 **Wang LC**, Chen EQ, Cao J, Liu L, Wang JR, Lei BJ, Tang H. Combination of Lamivudine and adefovir therapy in HBeAg-positive chronic hepatitis B patients with poor response to adefovir monotherapy. *J Viral Hepat* 2010; **17**: 178-184 [PMID: 19656287 DOI: 10.1111/j.1365-2893.2009.01164.x]
  - 86 **Kim SS**, Cheong JY, Lee D, Lee MH, Hong SP, Kim SO, Cho SW. Adefovir-based combination therapy with entecavir or lamivudine for patients with entecavir-refractory chronic hepatitis B. *J Med Virol* 2012; **84**: 18-25 [PMID: 22028068 DOI: 10.1002/jmv.22227]
  - 87 **Lampertico P**, Viganò M, Manenti E, Iavarone M, Sablon E, Colombo M. Low resistance to adefovir combined with lamivudine: a 3-year study of 145 lamivudine-resistant hepatitis B patients. *Gastroenterology* 2007; **133**: 1445-1451 [PMID: 17983801 DOI: 10.1053/j.gastro.2007.08.079]
  - 88 **Wei X**, Peterson DL. Expression, purification, and characterization of an active RNase H domain of the hepatitis B viral polymerase. *J Biol Chem* 1996; **271**: 32617-32622 [PMID: 8955090 DOI: 10.1074/jbc.271.51.32617]
  - 89 **Lok AS**, Lai CL, Leung N, Yao GB, Cui ZY, Schiff ER, Dienstag JL, Heathcote EJ, Little NR, Griffiths DA, Gardner SD, Castiglia M. Long-term safety of lamivudine treatment in patients with chronic hepatitis B. *Gastroenterology* 2003; **125**: 1714-1722 [PMID: 14724824 DOI: 10.1053/j.gastro.2003.09.033]
  - 90 **Zhao P**, Wang C, Huang L, Xu D, Li T. Comparison of rescue strategies in lamivudine-resistant patients with chronic hepatitis B. *Antiviral Res* 2012; **96**: 100-104 [PMID: 22960601 DOI: 10.1016/j.antiviral.2012.08.008]
  - 91 **Tenney DJ**, Rose RE, Baldick CJ, Levine SM, Pokornowski KA, Walsh AW, Fang J, Yu CF, Zhang S, Mazzucco CE, Eggers B, Hsu M, Plym MJ, Poundstone P, Yang J, Colonno RJ. Two-year assessment of entecavir resistance in Lamivudine-refractory hepatitis B virus patients reveals different clinical outcomes depending on the resistance substitutions present. *Antimicrob Agents Chemother* 2007; **51**: 902-911 [PMID: 17178796 DOI: 10.1128/AAC.00833-06]
  - 92 **Wang YP**, Zhao W, Xue R, Zhou ZX, Liu F, Han YX, Ren G, Peng ZG, Cen S, Chen HS, Li YH, Jiang JD. Oxymatrine inhibits hepatitis B infection with an advantage of overcoming drug-resistance. *Antiviral Res* 2011; **89**: 227-231 [PMID: 21277330 DOI: 10.1016/j.antiviral.2011.01.005]
  - 93 **Sheldon J**, Camino N, Rodés B, Bartholomeusz A, Kuiper M, Tacke F, Núñez M, Mauss S, Lutz T, Klausen G, Locarnini S, Soriano V. Selection of hepatitis B virus polymerase mutations in HIV-coinfected patients treated with tenofovir. *Antivir Ther* 2005; **10**: 727-734 [PMID: 16218172]
  - 94 **Pol S**, Sogni P. [Treatment of chronic hepatitis B: adherence and safety]. *Gastroenterol Clin Biol* 2010; **34** Suppl 2: S142-S148 [PMID: 21095518 DOI: 10.1016/S0399-8320(10)70034-8]
  - 95 **Fattovich G**, Olivari N, Pasino M, D'Onofrio M, Martone E, Donato F. Long-term outcome of chronic hepatitis B in Caucasian patients: mortality after 25 years. *Gut* 2008; **57**: 84-90 [PMID: 17715267 DOI: 10.1136/gut.2007.128496]
  - 96 **Hadziyannis SJ**, Tassopoulos NC, Heathcote EJ, Chang TT, Kitis G, Rizzetto M, Marcellin P, Lim SG, Goodman Z, Ma J, Arterburn S, Xiong S, Currie G, Brosgart CL. Long-term therapy with adefovir dipivoxil for HBeAg-negative chronic hepatitis B. *N Engl J Med* 2005; **352**: 2673-2681 [PMID: 15987916 DOI: 10.1056/NEJMoa042957]
  - 97 **Jeng WJ**, Sheen IS, Chen YC, Hsu CW, Chien RN, Chu CM, Liaw YF. Off-therapy durability of response to entecavir therapy in hepatitis B e antigen-negative chronic hepatitis B patients. *Hepatology* 2013; **58**: 1888-1896 [PMID: 23744454 DOI: 10.1002/hep.26549]
  - 98 **Invernizzi F**, Lampertico P, Loglio A, Iavarone M, Viganò M, Facchetti F, Vezali E, Lunghi G, Colombo M. Nucleos(t)ide analogues can be safely discontinued in chronic hepatitis B patients achieving HBsAg seroclearance. *Hepatology* 2012; **56**: 368A
  - 99 **Chimalakonda KC**, Agarwal HK, Kumar A, Parang K, Mehar R. Synthesis, analysis, in vitro characterization, and in vivo disposition of a lamivudine-dextran conjugate for selective antiviral delivery to the liver. *Bioconjug Chem* 2007; **18**: 2097-2108 [PMID: 17922546 DOI: 10.1021/bc700193d]
  - 100 **Zhang XG**, Miao J, Li MW, Jiang SP, Hu FQ, Du YZ. Solid lipid nanoparticles loading adefovir dipivoxil for antiviral therapy. *J Zhejiang Univ Sci B* 2008; **9**: 506-510 [PMID: 18543406 DOI: 10.1631/jzus.B0820047]
  - 101 **Lim SG**, Ng TM, Kung N, Krastev Z, Volfova M, Husa P, Lee SS, Chan S, Shiffman ML, Washington MK, Rigney A, Anderson J, Mondou E, Snow A, Sorbel J, Guan R, Rousseau F. A double-blind placebo-controlled study of emtricitabine in chronic hepatitis B. *Arch Intern Med* 2006; **166**: 49-56 [PMID: 16401810 DOI: 10.1001/archinte.166.1.49]
  - 102 **Liaw YF**, Sheen IS, Lee CM, Akarca US, Papatheodoridis GV, Suet-Hing Wong F, Chang TT, Horban A, Wang C, Kwan P, Buti M, Prieto M, Berg T, Kitrinos K, Peschell K, Mondou E, Frederick D, Rousseau F, Schiff ER. Tenofovir disoproxil fumarate (TDF), emtricitabine/TDF, and entecavir in patients with decompensated chronic hepatitis B liver disease. *Hepatology* 2011; **53**: 62-72 [PMID: 21254162 DOI: 10.1002/hep.23952]

- 103 **Kosi L**, Reiberger T, Payer BA, Grabmeier-Pfistershammer K, Strassl R, Rieger A, Peck-Radosavljevic M. Five-year on-treatment efficacy of lamivudine-, tenofovir- and tenofovir + emtricitabine-based HAART in HBV-HIV-coinfected patients. *J Viral Hepat* 2012; **19**: 801-810 [PMID: 23043387 DOI: 10.1111/j.1365-2893.2012.01601.x]
- 104 **Yuen MF**, Kim J, Kim CR, Ngai V, Yuen JC, Min C, Kang HM, Shin BS, Yoo SD, Lai CL. A randomized placebo-controlled, dose-finding study of oral LB80380 in HBeAg-positive patients with chronic hepatitis B. *Antivir Ther* 2006; **11**: 977-983 [PMID: 17302367]
- 105 **Min C**, Kim C, Steffy K, Averett D, Locarnini S, Shaw T. The active metabolite of LB80380/ANA380, a novel nucleotide analog, exhibits activity in vitro against multiple clinically relevant hepatitis B virus mutants. *J Hepatol* 2007; **46**: S159 [DOI: 10.1016/S0168-8278(07)62013-3]
- 106 **Lai CL**, Ahn SH, Lee KS, Um SH, Cho M, Yoon SK, Lee JW, Park NH, Kweon YO, Sohn JH, Lee J, Kim JA, Han KH, Yuen MF. Phase IIIb multicentred randomised trial of besifovir (LB80380) versus entecavir in Asian patients with chronic hepatitis B. *Gut* 2014; **63**: 996-1004 [PMID: 23979965 DOI: 10.1136/gutjnl-2013-305138]
- 107 **Yuen MF**, Han KH, Um SH, Yoon SK, Kim HR, Kim J, Kim CR, Lai CL. Antiviral activity and safety of LB80380 in hepatitis B e antigen-positive chronic hepatitis B patients with lamivudine-resistant disease. *Hepatology* 2010; **51**: 767-776 [PMID: 20091678 DOI: 10.1002/hep.23462]
- 108 **Lee JS**, Park ET, Kang SS, Gu ES, Kim JS, Jang DS, Lee KS, Lee JS, Park NH, Bae CH, Baik SK, Yu BJ, Lee SH, Lee EJ, Park SI, Bae M, Shin JW, Choi JH, Gu C, Moon SK, Chun GJ, Kim JH, Kim HS, Choi SK. Clevudine demonstrates potent antiviral activity in naïve chronic hepatitis B patients. *Intervirology* 2010; **53**: 83-86 [PMID: 19955812 DOI: 10.1159/000264197]
- 109 **Yoo BC**, Kim JH, Chung YH, Lee KS, Paik SW, Ryu SH, Han BH, Han JY, Byun KS, Cho M, Lee HJ, Kim TH, Cho SH, Park JW, Um SH, Hwang SG, Kim YS, Lee YJ, Chon CY, Kim BI, Lee YS, Yang JM, Kim HC, Hwang JS, Choi SK, Kweon YO, Jeong SH, Lee MS, Choi JY, Kim DG, Kim YS, Lee HY, Yoo K, Yoo HW, Lee HS. Twenty-four-week clevudine therapy showed potent and sustained antiviral activity in HBeAg-positive chronic hepatitis B. *Hepatology* 2007; **45**: 1172-1178 [PMID: 17464992 DOI: 10.1002/hep.21629]
- 110 **Lau GK**, Leung N. Forty-eight weeks treatment with clevudine 30 mg qd versus lamivudine 100 mg qd for chronic hepatitis B infection: a double-blind randomized study. *Korean J Hepatol* 2010; **16**: 315-320 [PMID: 20924215 DOI: 10.3350/kjhep.2010.16.3.315]
- 111 **Yoon EL**, Yim HJ, Lee HJ, Lee YS, Kim JH, Jung ES, Kim JH, Seo YS, Yeon JE, Lee HS, Um SH, Byun KS. Comparison of clevudine and entecavir for treatment-naïve patients with chronic hepatitis B virus infection: two-year follow-up data. *J Clin Gastroenterol* 2011; **45**: 893-899 [PMID: 21617542 DOI: 10.1097/MCG.0b013e31821f8bdf]
- 112 **Kwon SY**, Park YK, Ahn SH, Cho ES, Choe WH, Lee CH, Kim BK, Ko SY, Choi HS, Park ES, Shin GC, Kim KH. Identification and characterization of clevudine-resistant mutants of hepatitis B virus isolated from chronic hepatitis B patients. *J Virol* 2010; **84**: 4494-4503 [PMID: 20164224 DOI: 10.1128/JVI.02066-09]
- 113 **Seok JI**, Lee DK, Lee CH, Park MS, Kim SY, Kim HS, Jo HY, Lee CH, Kim DS. Long-term therapy with clevudine for chronic hepatitis B can be associated with myopathy characterized by depletion of mitochondrial DNA. *Hepatology* 2009; **49**: 2080-2086 [PMID: 19333909 DOI: 10.1002/hep.22959]
- 114 **Pollicino T**, Belloni L, Raffa G, Pediconi N, Squadrito G, Raimondo G, Levvero M. Hepatitis B virus replication is regulated by the acetylation status of hepatitis B virus cccDNA-bound H3 and H4 histones. *Gastroenterology* 2006; **130**: 823-837 [PMID: 16530522 DOI: 10.1053/j.gastro.2006.01.001]
- 115 **Zimmerman KA**, Fischer KP, Joyce MA, Tyrrell DL. Zinc finger proteins designed to specifically target duck hepatitis B virus covalently closed circular DNA inhibit viral transcription in tissue culture. *J Virol* 2008; **82**: 8013-8021 [PMID: 18524822 DOI: 10.1128/JVI.00366-08]
- 116 **Cai D**, Mills C, Yu W, Yan R, Aldrich CE, Saputelli JR, Mason WS, Xu X, Guo JT, Block TM, Cuconati A, Guo H. Identification of disubstituted sulfonamide compounds as specific inhibitors of hepatitis B virus covalently closed circular DNA formation. *Antimicrob Agents Chemother* 2012; **56**: 4277-4288 [PMID: 22644022 DOI: 10.1128/AAC.00473-12]
- 117 **Wooddell CI**, Rozema DB, Hossbach M, John M, Hamilton HL, Chu Q, Hegge JO, Klein JJ, Wakefield DH, Oropeza CE, Deckert J, Roehl I, Jahn-Hofmann K, Hadwiger P, Vornlocher HP, McLachlan A, Lewis DL. Hepatocyte-targeted RNAi therapeutics for the treatment of chronic hepatitis B virus infection. *Mol Ther* 2013; **21**: 973-985 [PMID: 23439496 DOI: 10.1038/mt.2013.31]
- 118 **Petersen J**, Dandri M, Mier W, Lütgehetmann M, Volz T, von Weizsäcker F, Haberkorn U, Fischer L, Pollok JM, Erbes B, Seitz S, Urban S. Prevention of hepatitis B virus infection in vivo by entry inhibitors derived from the large envelope protein. *Nat Biotechnol* 2008; **26**: 335-341 [DOI: 10.1038/nbt1389]
- 119 **Volz T**, Allweiss L, Ben MBarek M, Warlich M, Lohse AW, Pollok JM, Alexandrov A, Urban S, Petersen J, Lütgehetmann M, Dandri M. The entry inhibitor Myrcludex-B efficiently blocks intrahepatic virus spreading in humanized mice previously infected with hepatitis B virus. *J Hepatol* 2013; **58**: 861-867 [PMID: 23246506 DOI: 10.1016/j.jhep.2012.12.008]
- 120 **Al-Mahtab M**, Bazinet M, Vaillant A. Establishment of a potent anti-HBsAg response and durable immunological control of viremia with short term immunotherapy after REP 9AC-induced HBsAg seroclearance in chronic HBV infection (Internet). *J Hepatol* 2013; **58**: S316 [DOI: 10.1016/S0168-8278(13)60778-3]
- 121 **Ilan Y**, Nagler A, Adler R, Tur-Kaspa R, Slavin S, Shouval D. Ablation of persistent hepatitis B by bone marrow transplantation from a hepatitis B-immune donor. *Gastroenterology* 1993; **104**: 1818-1821 [PMID: 8500741]
- 122 **Couillin I**, Pol S, Mancini M, Driss F, Bréchet C, Tiollais P, Michel ML. Specific vaccine therapy in chronic hepatitis B: induction of T cell proliferative responses specific for envelope antigens. *J Infect Dis* 1999; **180**: 15-26 [PMID: 10353856 DOI: 10.1086/314828]
- 123 **Heathcote J**, McHutchison J, Lee S, Tong M, Benner K, Minuk G, Wright T, Fikes J, Livingston B, Sette A, Chestnut R. A pilot study of the CY-1899 T-cell vaccine in subjects chronically infected with hepatitis B virus. The CY1899 T Cell Vaccine Study Group. *Hepatology* 1999; **30**: 531-536 [PMID: 10421664 DOI: 10.1002/hep.510300208]
- 124 **Hong SH**, Cho O, Kim K, Shin HJ, Kotenko SV, Park S. Effect of interferon-lambda on replication of hepatitis B virus in human hepatoma cells. *Virus Res* 2007; **126**: 245-249 [PMID: 17451832 DOI: 10.1016/j.virusres.2007.03.006]
- 125 **Nakagawa S**, Hirata Y, Kameyama T, Tokunaga Y, Nishito Y, Hirabayashi K, Yano J, Ochiya T, Tateno C, Tanaka Y, Mizokami M, Tsukiyama-Kohara K, Inoue K, Yoshida M, Takaoka A, Kohara M. Targeted induction of interferon-λ in humanized chimeric mouse liver abrogates hepatotropic virus infection. *PLoS One* 2013; **8**: e59611 [PMID: 23555725 DOI: 10.1371/journal.pone.0059611]
- 126 **Chan H**, Ahn S, Chang T, Peng C, Wong D, Coffin C, Lim S, Chen P, Janssen H, Marcellin P, Serfaty L, Zeuzem S, Hu W, Critelli L, Lopez-Talavera J, Cooney E. Peginterferon Lambda for the Treatment of HBeAg-Positive Chronic Hepatitis B: A Phase 2B Comparison with Peginterferon Alfa. APASL Liver Week 2013; Singapore, June 6-10, 2013
- 127 **Zeuzem S**, Arora S, Bacon B, Box T, Charlton M, Diago M, Dieterich D, Mur RE, Everson G, Fallon M, Ferenci P, Fli-

siak R, George J, Ghalib R, Gitlin N, Gladysz A, Gordon S, Greenbloom S, Hassanein T, Jacobson I, Jeffers L, Kowdley K, Lawitz E, Lee SS, Leggett B, Lueth S, Nelson D, Pockros P, Rodriguez-Torres M, Rustgi V, Serfaty L, Sherman M, Shiffman M, Sola R, Sulkowski M, Vargas H, Vierling J, Yoffe B, Ishak L, Fontana D, Xu D, Gray T, Horga A, Hillson J, Lopez-Talavera JC, Muir A. Peginterferon Lambda-1a (lambda) compared to peginterferon alfa-2a (alfa) in treatment-naive patients with HCV genotypes (G) 2 or 3: First SVR24 results

from EMERGE phase IIB. *J Hepatol* 2012; **56** Supple: S5-S6 [DOI: 10.1016/S0168-8278(12)60024-5]

128 **Lanford RE**, Guerra B, Chavez D, Giavedoni L, Hodara VL, Brasky KM, Fosdick A, Frey CR, Zheng J, Wolfgang G, Halcomb RL, Tumas DB. GS-9620, an oral agonist of Toll-like receptor-7, induces prolonged suppression of hepatitis B virus in chronically infected chimpanzees. *Gastroenterology* 2013; **144**: 1508-1517, 1517.e1-e10 [PMID: 23415804 DOI: 10.1053/j.gastro.2013.02.003]

**P- Reviewers:** Huerta-Franco MR, Karayiannis P, Peter LL  
**S- Editor:** Ma YJ **L- Editor:** A **E- Editor:** Wu HL





# Full-Length Genome of an *Ogataea polymorpha* Strain CBS4732 *ura3Δ* Reveals Large Duplicated Segments in Subtelomeric Regions

Jia Chang<sup>1†</sup>, Jinlong Bei<sup>2,3†</sup>, Qi Shao<sup>1</sup>, Hemu Wang<sup>4</sup>, Huan Fan<sup>5</sup>, Tung On Yau<sup>6,7</sup>, Wenjun Bu<sup>1</sup>, Jishou Ruan<sup>8</sup>, Dongsheng Wei<sup>1\*</sup> and Shan Gao<sup>1\*</sup>

<sup>1</sup> Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, College of Life Science, Nankai University, Tianjin, China, <sup>2</sup> Agro-Biological Gene Research Center, Guangdong Academy of Agricultural Sciences, Guangdong Provincial Key Laboratory for Crop Germplasm Resources Preservation and Utilization, Guangzhou, China, <sup>3</sup> Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, China, <sup>4</sup> Tianjin Hemu Health Biotechnological Co., Ltd., Tianjin, China, <sup>5</sup> Tianjin Institute of Animal Husbandry and Veterinary Research, Tianjin, China, <sup>6</sup> John Van Geest Cancer Research Centre, School of Science and Technology, Nottingham Trent University, Nottingham, United Kingdom, <sup>7</sup> Department of Rural Land Use, Scotland's Rural College, Aberdeen, United Kingdom, <sup>8</sup> School of Mathematical Sciences, Nankai University, Tianjin, China

## OPEN ACCESS

### Edited by:

Feng Gao,  
Tianjin University, China

### Reviewed by:

Fabio Mitsuo Lima,  
Centro Universitário São Camilo,  
Brazil  
Sara Hanson,  
Colorado College, United States  
Yanzhu Ji,  
Institute of Zoology (CAS), China  
Dzmitry Padhorniy,  
Stony Brook University, United States

### \*Correspondence:

Dongsheng Wei  
weidongsheng@nankai.edu.cn  
Shan Gao  
gao\_shan@mail.nankai.edu.cn

† These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 15 January 2022

Accepted: 25 February 2022

Published: 04 April 2022

### Citation:

Chang J, Bei J, Shao Q, Wang H,  
Fan H, Yau TO, Bu W, Ruan J, Wei D  
and Gao S (2022) Full-Length  
Genome of an *Ogataea polymorpha*  
Strain CBS4732 *ura3Δ* Reveals Large  
Duplicated Segments in Subtelomeric  
Regions. *Front. Microbiol.* 13:855666.  
doi: 10.3389/fmicb.2022.855666

**Background:** Currently, methylotrophic yeasts (e.g., *Pichia pastoris*, *Ogataea polymorpha*, and *Candida boindii*) are subjects of intense genomics studies in basic research and industrial applications. In the genus *Ogataea*, most research is focused on three basic *O. polymorpha* strains-CBS4732, NCYC495, and DL-1. However, the relationship between CBS4732, NCYC495, and DL-1 remains unclear, as the genomic differences between them have not been exactly determined without their high-quality complete genomes. As a nutritionally deficient mutant derived from CBS4732, the *O. polymorpha* strain CBS4732 *ura3Δ* (named HU-11) is being used for high-yield production of several important proteins or peptides. HU-11 has the same reference genome as CBS4732 (noted as HU-11/CBS4732), because the only genomic difference between them is a 5-bp insertion.

**Results:** In the present study, we have assembled the full-length genome of *O. polymorpha* HU-11/CBS4732 using high-depth PacBio and Illumina data. Long terminal repeat retrotransposons (LTR-rts), rDNA, 5' and 3' telomeric, subtelomeric, low complexity and other repeat regions were exactly determined to improve the genome quality. In brief, the main findings include complete rDNAs, complete LTR-rts, three large duplicated segments in subtelomeric regions and three structural variations between the HU-11/CBS4732 and NCYC495 genomes. These findings are very important for the assembly of full-length genomes of yeast and the correction of assembly errors in the published genomes of *Ogataea* spp. HU-11/CBS4732 is so phylogenetically close to NCYC495 that the syntenic regions cover nearly 100% of their genomes. Moreover, HU-11/CBS4732 and NCYC495 share a nucleotide identity of 99.5% through their whole genomes. CBS4732 and NCYC495 can be regarded as the same strain in basic research and industrial applications.

**Conclusion:** The present study preliminarily revealed the relationship between CBS4732, NCYC495, and DL-1. Our findings provide new opportunities for in-depth

understanding of genome evolution in methylotrophic yeasts and lay the foundations for the industrial applications of *O. polymorpha* CBS4732, NCYC495, DL-1, and their derivative strains. The full-length genome of *O. polymorpha* HU-11/CBS4732 should be included into the NCBI RefSeq database for future studies of *Ogataea* spp.

**Keywords:** methylotrophic yeast, *Hansenula polymorpha*, rDNA quadruple, genome expansion, long terminal repeat

## INTRODUCTION

Currently, methylotrophic yeasts (e.g., *Pichia pastoris*, *Hansenula polymorpha*, and *Candida boindii*) are subjects of intense genomics studies in basic research and industrial applications. However, genomic research on *Ogataea* (*Hansenula*) *polymorpha* trails behind that on *P. pastoris* (Ravin et al., 2013), although they both are widely used species of methylotrophic yeasts. In the genus *Ogataea*, most research is focused on three basic *O. polymorpha* strains—CBS4732 (synonymous to NRRL Y-5445 or ATCC34438), NCYC495 (synonymous to NRRL Y-1798, ATCC14754, or CBS1976), and DL-1 (synonymous to NRRL Y-7560 or ATCC26012). These three strains are of independent geographic and ecological origins: CBS4732 was originally isolated from soil irrigated with waste water from a distillery in Pernambuco, Brazil in 1959 (Morais and Maia, 1959); NCYC495 is identical to a strain first isolated from spoiled concentrated orange juice in Florida and initially designated as *Hansenula angusta* by Wickerham (1951); DL-1 was isolated from soil by Levine and Cooney (1973). CBS4732 and its derivatives—LR9 and RB11—have been developed as genetically engineered strains to produce many heterologous proteins, including enzymes (e.g., feed additive phytase), anticoagulants (e.g., hirudin and saratin), and an efficient vaccine against hepatitis B infection (Massoud et al., 2003). As a nutritionally deficient mutant derived from CBS4732, the *O. polymorpha* strain HU-11 (CBS4732 *ura3*Δ) (Wang et al., 2007) is being used for high-yield production of several important proteins or peptides, particularly including recombinant hepatitis B surface antigen (HBsAg) vaccine (Wang H. et al., 2016) and hirudin (Wang et al., 2011). HU-11 has the same reference genome as CBS4732 (noted as HU-11/CBS4732), as the only genomic difference between them is a 5-bp insertion caused by frame-shift mutation of its *URA3* gene, which encodes orotidine 5'-phosphate decarboxylase. Although CBS4732 and NCYC495 are classified as *O. polymorpha*, and DL-1 is reclassified as *O. parapolyomorpha* (Hanson et al., 2017), the relationship between CBS4732, NCYC495, and DL-1 remains unclear, as the genomic differences between them have not been exactly determined due to lack of their high-quality complete genomes. Thus, knowledge obtained from any of the three strains can't be used to investigate the other two strains.

To facilitate genomic research of yeasts, genome sequences have been increasingly submitted to the Genome-NCBI datasets. Among the genomes of 34 species in the *Ogataea* or *Candida*

genus (Supplementary File 1), those of NCYC495 and DL-1 have been assembled at chromosome levels. However, the other genomes have been assembled at contig or scaffold levels. Furthermore, the genome sequence of CBS4732 was not available in the Genome-NCBI datasets until this manuscript was drafted. Among the genomes of 33 *Komagataella* (*Pichia*) spp., the genome of the *P. pastoris* strain GS115 is the only genome assembled at the chromosome level. The main problem of the *Ogataea*, *Candida*, and *Pichia* genome data is their incomplete sequences and poor annotations. For example, the rDNA sequence (GenBank: FN392325) of *P. pastoris* GS115 cannot be well aligned to its genome (GenBank assembly: GCA\_001708105). Most genome sequences do not contain complete subtelomeric regions and, as a result, subtelomeres are often overlooked in comparative genomics (Brown, 2010). For example, the genome of DL-1 has been analyzed for better understanding the phylogenetics and molecular basis of *O. polymorpha* (Ravin et al., 2013); however, it does not contain complete subtelomeric regions due to assembly using short sequences. Another problem of the *Ogataea*, *Candida*, and *Pichia* genome data is that the mitochondrial (mt) genome sequence is not simultaneously released with the corresponding nuclear genome sequences. The only exception is the *O. polymorpha* DL-1 mt genome (RefSeq: NC\_014805). Therefore, more complete genome sequences of *Ogataea* spp. are being accomplished to bridge the gap between basic research and industrial applications. For example, a new project has been conducted to provide a high-quality complete genome of DL-1 (GenBank: CP080316-22) based on Nanopore technology.

In the present study, we have assembled the full-length genome of *O. polymorpha* HU-11/CBS4732 using high-depth PacBio and Illumina data, and conducted the annotation and analysis to achieve the following research goals: (1) to provide a high-quality and well-curated reference genome for future studies of *Ogataea* spp.; (2) to determine the relationship between CBS4732, NCYC495, and DL-1; and (3) to discover important genomic features (e.g., high yield) of *Ogataea* spp. for basic research (e.g., synthetic biology) and industrial applications.

## RESULTS AND DISCUSSION

### Genome Sequencing, Assembly and Annotation

One 500 bp and one 10 Kbp DNA library were prepared using fresh cells of *O. polymorpha* HU-11 and sequenced on the Illumina HiSeq X Ten and PacBio Sequel platforms,

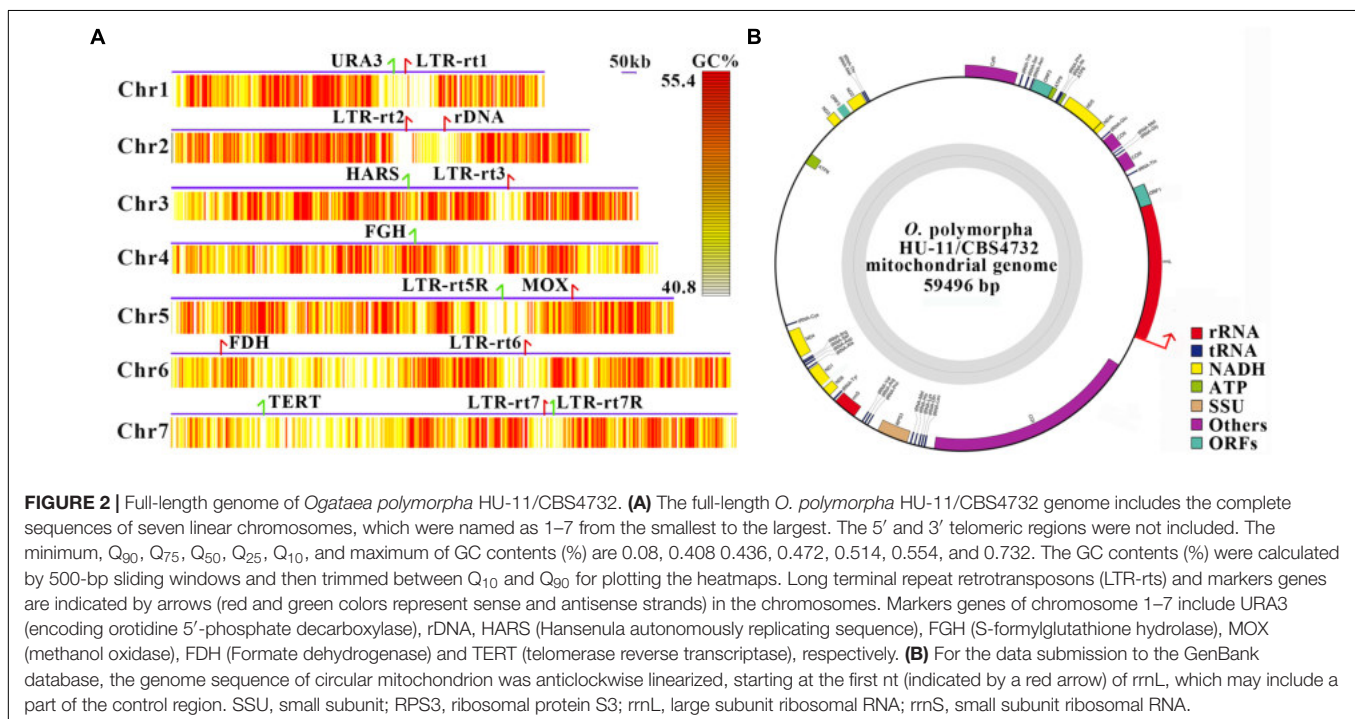
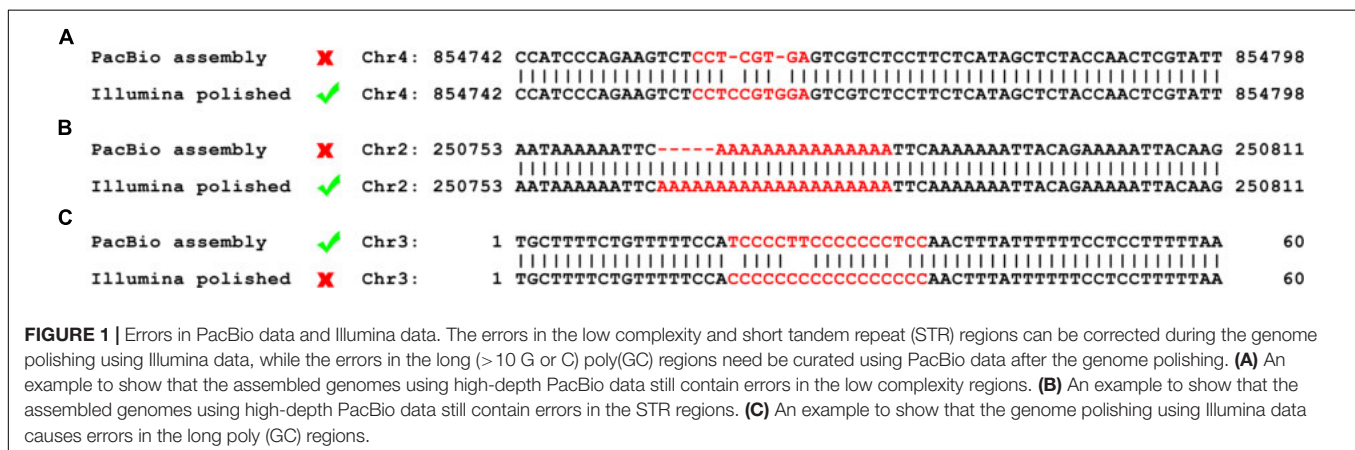
**Abbreviations:** TR, tandem repeat; STR, short tandem repeat; LTR, long terminal repeat; ORF, open reading frame; CDS, coding sequence; SV, structural variation; SNP, single nucleotide polymorphism; InDel, insertion and deletion; mt, mitochondrial; nt, nucleotide; aa, amino acid.



respectively, for *de novo* assembly of a high-quality genome. Firstly, 18,319,084,791 bp cleaned PacBio DNA-seq data were used to assemble the complete genome, except the rDNA region, with an extremely high depth of ~1800X. However, the draft genome using high-depth PacBio data still contained two types of errors in the low complexity (Figure 1A) and the short tandem repeat (STR) regions, respectively (Figure 1B). Then, 6,628,480,424 bp cleaned Illumina DNA-seq data were used to polish the complete genome of HU-11/CBS4732 to remove the two types of errors. However, Illumina DNA-seq data contained errors in the long (>10 G or C) poly(GC) regions. Following this, the poly(GC) regions, polished using PacBio subreads (Figure 1C). Finally, 5' and 3' telomeric, subtelomeric, rDNA, Long Terminal Repeat retrotransposons (LTR-rt), low complexity, and other

repeat regions were exactly determined and confirmed by human curation (see section "Materials and Methods"). The complete *O. polymorpha* HU-11/CBS4732 genome is a full-length genome, which is defined to have sequences ending at the 5' and 3' telomeric sites without gaps and ambiguous nucleotides (Xu et al., 2020).

The full-length *O. polymorpha* HU-11/CBS4732 genome includes the complete sequences of all seven chromosomes (Figure 2A), which were named as 1 to 7 from the smallest to the largest, respectively (Table 1). As the 5' and 3' telomeric regions vary in lengths, they were not included into the seven linear chromosomes of HU-11/CBS4732. Analysis of long PacBio subreads revealed that the telomeric regions at 5' and 3' ends of each chromosome consist of tandem repeats (TRs)  $[ACCCCGCC]_n$  and  $[GGCGGGT]_n$  ( $n$  is



**TABLE 1** | Genomes of three basic *Ogataea polymorpha* strains.

Chromosome	CBS4732/HU-11	NCYC495	DL-1	HU-11 Size (bp)	Marker
Chr1	CP073033	NW_017264703	NC_027865	1,000,895	URA3
Chr2	CP073034	NW_017264704	NC_027866	1,125,341	rDNA
Chr3	CP073035	NW_017264702	NC_027864	1,265,401	HARS
Chr4	CP073036	NW_017264701	NC_027863	1,315,956	FGH
Chr5	CP073037	NW_017264700	NC_027862	1,357,435	MOX
Chr6	CP073038	NW_017264698	NC_027860	1,513,391	FDH
Chr7	CP073039	NW_017264699	NC_027861	1,525,912	TERT
ChrM	CP073040	NA	NC_014805	59,496	COIII
Total (Mbp)	9.1	8.97	8.87		
GC%	47.76	47.86	47.83		
Gene <sup>#</sup>	5453*	5454*	5309*		
tRNA <sup>#</sup>	80	80	80		
rRNA <sup>#</sup>	4 × 20	4 × 6	4 × 25		

HU-11 has the same reference genome as CBS4732 (noted as HU-11/CBS4732), because the only genomic difference between them is a 5-bp insertion. <sup>#</sup>The numbers of mitochondrial genes in the LTR-rtS were not counted. \*The genome sequences of NCYC495 and DL-1 with annotations were corrected, so the numbers of protein-coding genes (>150 bp) are different from their original records. The full-length *O. polymorpha* HU-11/CBS4732 genome includes the complete sequences of all seven chromosomes, which were named as 1–7 from the smallest to the largest, respectively. As the 5' and 3' telomeric regions vary in lengths, they were not included into the seven linear chromosomes of HU-11/CBS4732. The accession numbers of NCYC495 and DL-1 were mapped to the chromosome numbers of HU-11, according to the marker genes of seven chromosomes. They are: URA3 (encoding orotidine 5'-phosphate decarboxylase), rDNA, HARS (Hansenula autonomously replicating sequence), FGH (S-formylglutathione hydrolase), MOX (methanol oxidase), FDH (Formate dehydrogenase), TERT (telomerase reverse transcriptase) and COIII (cytochrome c oxidase subunit 3).

the copy number) with average lengths of 166 bp and 168 bp (~20 copy numbers), respectively. Finally, we released the data of the nuclear genome (GenBank: CP073033-39) with a summed sequence length of 9.1 Mbp and the mt genome (GenBank: CP073040) with a sequence length of 59,496 bp (Table 1). For the submission to the GenBank database, the sequence of circular mt genome (Figure 2B) was anticlockwise linearized, starting at the first nt of large subunit ribosomal RNA (rrnL). The complete genome sequence of the *O. polymorpha* strain CBS4732 *ura3*Δ (named HU-11) is available at the NCBI GenBank database, which need be included into the NCBI RefSeq database to facilitate future studies on *O. polymorpha* CBS4732 and its derivatives- LR9, RB11, and HU-11.

The HU-11/CBS4732 (nuclear) genome has a summed length of 9.1 Mbp that is close to the estimated length of the *O. polymorpha* DL-1 genome (Ravin et al., 2013), while the NCYC495 (RefSeq: NW\_017264698-704) and DL-1 genomes (RefSeq: NC\_027860-66) have shorter lengths of 8.97 and 8.87 Mbp, respectively (Table 1), as both of them are incomplete and have many errors at 5' and 3' ends of their chromosomes (Supplementary File 1). The GC contents of the HU-11, NCYC495, and DL-1 genomes are very close (~48%). Syntenic comparison (see section "Materials and Methods") revealed that *O. polymorpha* HU-11/CBS4732 is so phylogenetically close to NCYC495 that the syntenic regions cover nearly 100% of their genomes, however, HU-11/CBS4732 and NCYC495 are significantly distinct from DL-1. Then, we discovered large duplicated segments (LDSs) in the subtelomeric regions and exactly determined all the structural variations (SVs) between HU-11/CBS4732 and NCYC495 (Detailed later). We improved the annotations of NCYC495 protein-coding genes

(>150 bp) using a high quality RNA-seq data of NCYC495 (NCBI SRA: SRP124832), and then HU-11/CBS4732 and DL-1 genes by gene ID mapping (Supplementary File 2). As a result, we updated the annotations (Table 1) of: (1) 5,453 protein-coding genes of HU-11/CBS4732, including 5,021 single-exon genes, and 432 multi-exon genes; (2) 5,454 protein-coding genes of NCYC495, including 5,022 single-exon genes, and 432 multi-exon genes; (3) 5,309 protein-coding genes of DL-1, including 4,843 single-exon genes, and 464 multi-exon genes; and (4) 80 identical tRNA genes of HU-11/CBS4732, NCYC495, and DL-1. For 432 multi-exon genes of HU-11/CBS4732, only the longest splicing isoforms of them were annotated. Then, 5,453 CDSs (Supplementary File 3) were identified from 5,453 genes of HU-11/CBS4732. Furthermore, only 13 single-exon genes of HU-11/CBS4732 or NCYC495 were not detected to be expressed using the RNA-seq data SRP124832, while 87 genes (data not shown) of DL-1 have been reported to be not expressed in the previous study (Ravin et al., 2013).

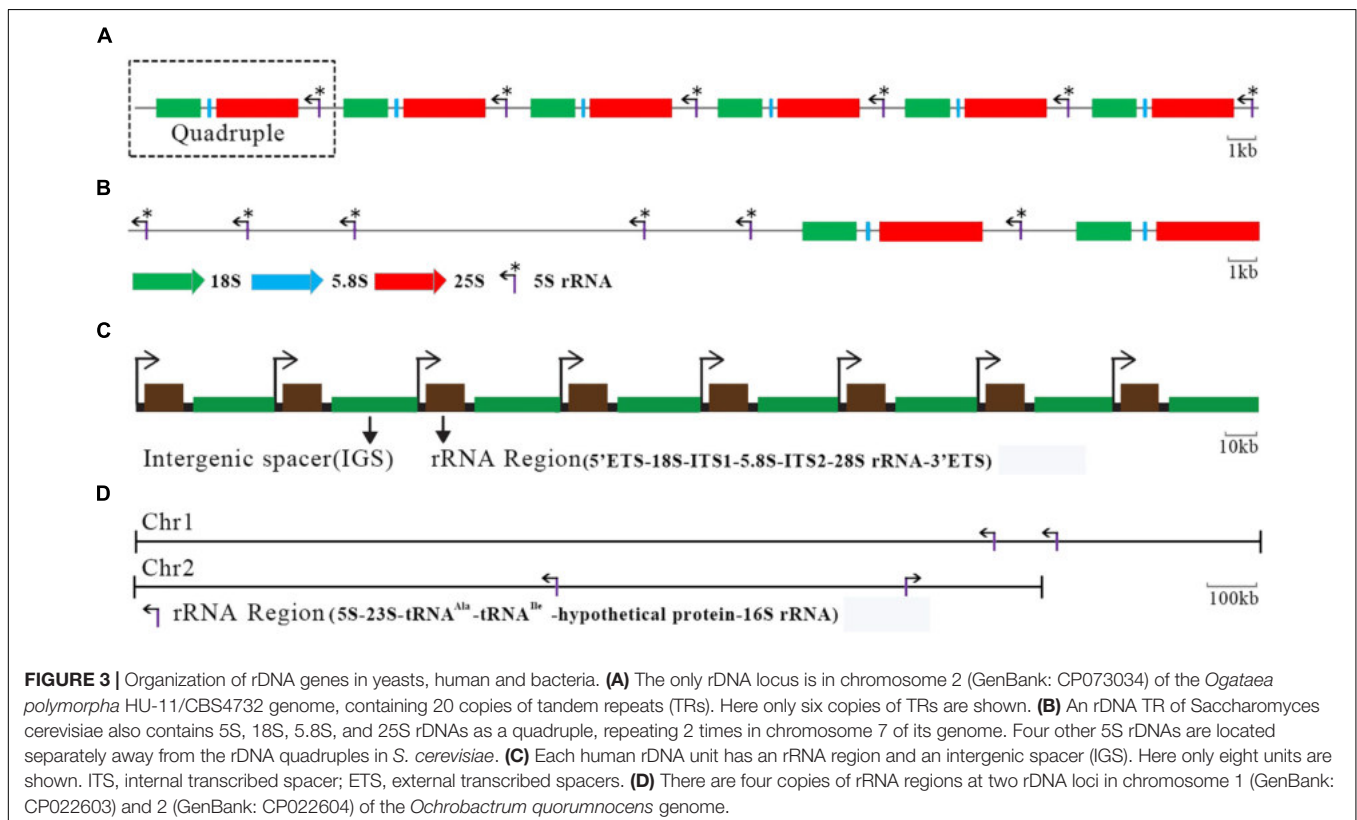
## Organization of rDNA Genes

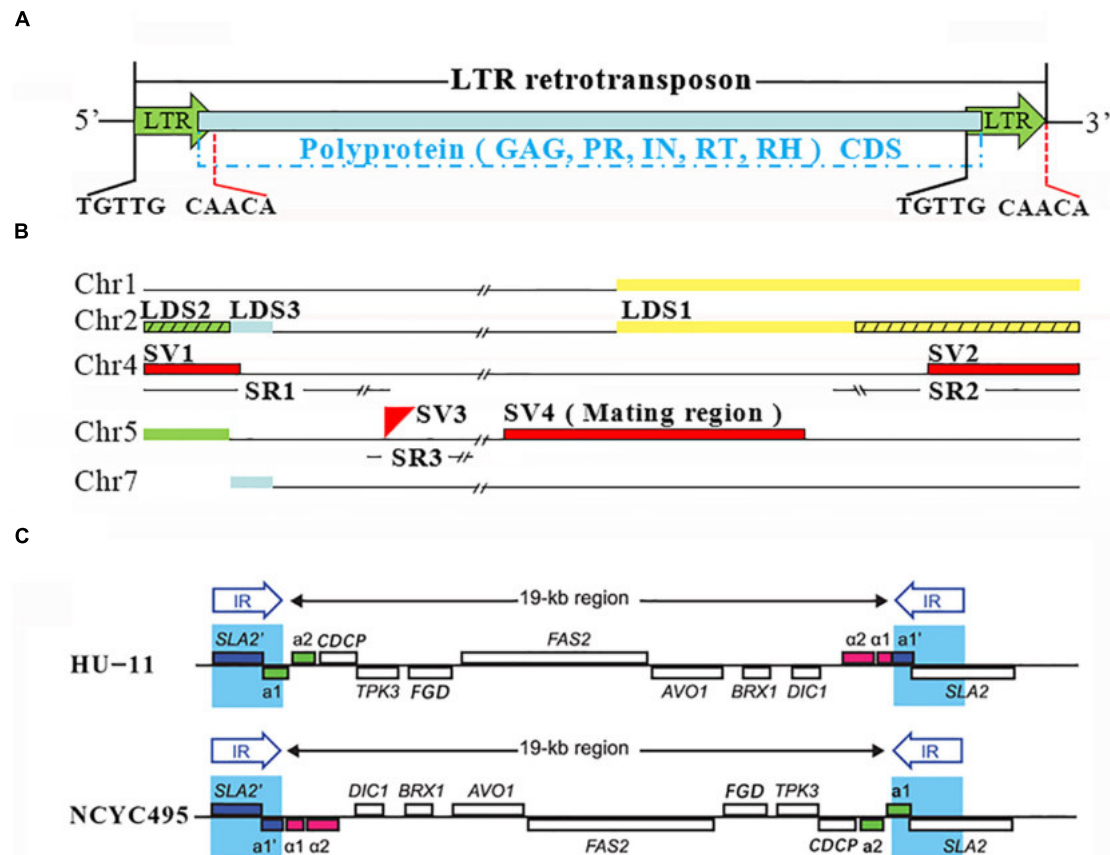
An rDNA TR of HU-11/CBS4732, NCYC495, or DL-1 encodes 5S, 18S, 5.8S, and 25S rRNAs (named as quadruple in the present study), with a length of 8,145 bp (Supplementary File 1). As the largest TR region (~162 Kbp) in the HU-11/CBS4732 genome, the only rDNA locus is in chromosome 2. The organization of rDNA TRs is conserved in the *Ogataea* genus. TRs of HU-11/CBS4732 and NCYC495 rDNAs share a very high nucleotide (nt) sequence identity of 99.5% (8,115/8,152), while TRs of HU-11/CBS4732 and DL-1 rDNAs share a comparatively low nt sequence identity of 97% (7,530/7,765). The major difference of rDNAs among *Ogataea* spp. lie in their copy numbers. The copy number of rDNA TRs was

estimated as 20 in the HU-11/CBS4732 genome (Figure 3A), while that was estimated as 6 and 25 in NCYC495 and DL-1, respectively (Ravin et al., 2013). An rDNA TR of *Saccharomyces cerevisiae* also contains 5S, 18S, 5.8S, and 25S rDNAs as a quadruple, repeating two times in chromosome 7 of its genome (Figure 3B), however, four other 5S rDNAs are located separately away from the rDNA quadruples in *S. cerevisiae*. Different from *O. polymorpha* or *S. cerevisiae* with only one rDNA locus, *Pichia pastoris* GS115 carries several rDNA loci, which are interspersed in three of its four chromosomes. Since the genome of *P. pastoris* GS115 (GenBank assembly: GCA\_001708105) is incomplete and poorly annotated, we estimated the copy number of its rDNAs as three or more. In animals, rDNAs encoding 18S, 5.8S, and 28S rRNAs are also organized in TRs and transcribed into a single RNA precursor by RNA polymerase I. As a typical example, human has approximately 200–600 rDNA copies (Figure 3C) distributed in short arms of the five acrocentric chromosomes (chromosomes 13, 14, 15, 21, and 22) (Agrawal and Ganley, 2018). In prokaryotic cells, 5S, 23S, and 16S rRNA genes are typically organized as a co-transcribed operon. There may be one or more copies of the operon dispersed in the genome and the copy numbers typically range from 1 to 15 in bacteria. As a typical example, *Ochrobactrum quorumnocens* has four copies of co-transcribed operons at two rDNA loci (Figure 3D) in chromosome 1 (GenBank: CP022603) and 2 (GenBank: CP022604). Compared to *S. cerevisiae*, human, and *O. quorumnocens* rDNAs (Figures 3B–D), the rDNAs of

*O. polymorpha* are more closely organized. By the organization in TR, 20 copies of *O. polymorpha* rDNA quadruples are transcribed in a large (>162 Kbp) co-transcribed operon, suggesting that their transcription can be regulated with higher efficiency. This genomic feature may contribute to the high yield characteristics of *O. polymorpha*.

Besides the high similarity of genomic organization, the rDNAs of *O. polymorpha* HU-11/CBS4732 and *S. cerevisiae* share high nt sequence identities of 95.3% (1720/1805), 96.2% (152/158), 92% (3,111/3,381), and 96.7% (117/121) for 18S, 5.8S, 25S, and 5S rDNAs, respectively. These identities are little lower than those of *O. polymorpha* NCYC495 and DL-1, indicated that the rDNA genes are more conservative than the protein-coding genes. Therefore, rDNA is an important genomic feature for the detection, identification, classification and phylogenetic analysis of *Ogataea* spp. Unexpectedly, we found that the rDNAs (GenBank: FN392325) of *O. polymorpha* HU-11/CBS4732 and *P. pastoris* GS115 have nt sequence identities of 87.3% (1477/1691), 80% (84/105), and 80.5% (2,073/2,576) for 18S, 5.8S, and 25S rDNAs, respectively, which are much lower than those of *S. cerevisiae*. These results are not consistent with those of a previous study (Ravin et al., 2013), in which phylogenetic analysis using 153 protein-coding genes showed that *O. polymorpha* and *Pichia pastoris* GS115 are members of a clade that is distinct from the one that *S. cerevisiae* belongs to. Based on the previous study, HU-11/CBS4732 is phylogenetically closer to *P. pastoris* GS115 than *S. cerevisiae*. The nt sequence identities of rDNAs between HU-11/CBS4732 and





**FIGURE 4** | Long terminal repeat (LTR) retrotransposons, large duplicated segments and structural variations. **(A)** NCYC495 and HU-11/CBS4732 share identical 322-bp LTRs, which are flanked by TCTTG and CAACA at their 5' and 3' ends. Three of seven LTR-rt of HU-11/CBS4732 have no homologs in the NCYC495 genome due to misassembly. A LTR-rt consists of 5' LTR, 3' LTR and a single open reading frame (ORF) encoding a putative polyprotein. This polyprotein, if translated, can be processed into truncated gag (GAG), protease (PR), integrase (IN), reverse transcriptase (RT) and RNase H (RH). **(B)** Chr1, 2, 4, 5, and 7 represent the chromosomes (GenBank: CP073033, 34, 36, 37, and 39) of the HU-11/CBS4732 genome. Three large duplicated segments (LDSs) named LDS1 (in yellow color), 2 (in green color) and 3 (in blue color) are supposed to be included in both NCYC495 and HU-11/CBS4732 genomes. However, LDS2 and a 14,090 bp part of LDS1 (indicated by black slashes) were not assembled into chromosome 2 of the NCYC495 genome. The genomic differences between the HU-11/CBS4732 and NCYC495 only include three structural variations (SVs), named SV1, 2 and 3 (in red color). The three SVs are located in three large syntenic regions (SRs) of HU-11/CBS4732, NCYC495, and DL-1 genomes with very high nt sequence identities, named SR1, 2 and 3. SV4 is a 22.6-Kb DNA region which functions in the determination of the yeast mating-type (MAT). **(C)** The graphic elements used to represent the genomes and genes were originally used in the previous study (Hanson et al., 2017). The HU-11 genome (GenBank: CP073033-40) contains a 22.6-Kb MAT region where MAT $\alpha$  can be transcribed, while the NCYC495 genome (RefSeq: NW\_017264698-704) contains an identical 22.6-Kb MAT region where MAT $\alpha$  can be transcribed.

*P. pastoris* GS115 are supposed to be higher than those between HU-11/CBS4732 and *S. cerevisiae*.

## Long Terminal Repeat Retrotransposons

LTRs with lengths of 322 bp were discovered in all seven chromosomes of HU-11/CBS4732. These LTRs with the low GC content of 29% (94/322) are flanked by TCTTG and CAACA at their 5' and 3' ends (Figure 4A). In HU-11/CBS4732, a total of 14 LTRs are present in seven copies of intact LTR-rt (Supplementary File 1), which were identified as components of Tpa5 LTR-rt (GenBank: AJ439553) from *Pichia angusta* CBS4732 (a former name of *O. polymorpha* CBS4732) in a previous study (Neuveglise et al., 2002). A LTR-rt consists of 5' LTR, 3' LTR, and a single open reading frame (ORF) encoding a putative polyprotein (Figure 4A). This polyprotein, if translated,

can be processed into truncated Gag (GAG), protease (PR), integrase (IN), reverse transcriptase (RT), and RNase H (RH). Based on the gene order (PR, IN, RT, and RH), the LTR-rt of HU-11/CBS4732 were classified into the Ty5 type of the Ty1/copia group (Ty1, 2, 4, and 5 types) (Agrawal and Ganley, 2018).

With the length corrected from 4,883 bp to 4,882 bp, a sequence (GenBank: AJ439558) was used as the reference of Tpa5 LTR-rt to search for homologs. The results confirmed that HU-11/CBS4732 is phylogenetically closest to NCYC495 and they share identical LTRs. However, the 322-bp LTRs of HU-11/CBS4732 and NCYC495 are quite distinct from the 282-bp LTRs of DL-1, which were reported as 290-bp solo LTRs in the previous study (Ravin et al., 2013). In addition, the amino acid (aa) sequences of the polyprotein with the length of 1417 aa in HU-11/CBS4732 and NCYC495 LTR-rt

are distinct from those in DL-1. Based on the records in the UniProt Knowledgebase (UniProtKB), *O. polymorpha* strains DL-1, ATCC26012, BCRC20466, JCM22074, and NRRL Y-7560 have nearly the same aa sequences (UniProt: W1QI12) of the polyprotein. The above results suggest that the LTR-rt is another important genomic feature for the detection, identification, classification and phylogenetic analysis of *Ogataea* spp. Using RNA-seq data of NCYC495 (SRA: SRP124832), we discovered that the genes encoding the polyproteins in the LTR-rt of *O. polymorpha* are transcribed. If these putative polyproteins can be translated merits further studies.

In the previous study, 50,000 fragments of 13 *Hemiascomycetes* species were used to identify LTR-rt. However, the analysis was probably biased as it was based on only random sequences (approximately 1 kb on average) without the related genome information (Neueglise et al., 2002). In the present study, seven copies of intact LTR-rt (As described above) were precisely located in the HU-11/CBS4732 genome (**Figure 2A**), with five in the sense strands of chromosome 1, 2, 3, 6, and 7 (named LTR-rt1, 2, 3, 6, and 7) and two in the antisense strands of chromosome 5 and 7 (named LTR-rt5R and 7R). LTR-rt1, 3, and 6 share very high nt identities of 99.9% with each other. LTR-rt1 or 3 contains a single ORF encoding a polyprotein with the same aa sequence, while LTR-rt6 contains a single ORF with a 42-bp insertion (encoding RSSLFDVPCSTVD), compared to LTR-rt1 and 3. LTR-rt2, 5R, 7, and 7R contain several single nucleotide polymorphisms (SNPs), small insertions and deletions (InDels), which break the single ORFs into several ORFs. Genome comparison revealed that the homologs of LTR-rt2, 3, and 5R in HU-11/CBS4732 are present in the NCYC495 genome with very high nt identities of 99.9%, while the homologs of LTR-rt1, 7, and 7R, however, are absent in the NCYC495 genome. Further analysis determined that their absence in the NCYC495 genome resulted from misassembly.

## Large Duplicated Segments in Subtelomeric Regions

Syntenic comparison revealed that *O. polymorpha* HU-11/CBS4732 is so phylogenetically close to NCYC495 that the syntenic regions cover nearly 100% of their genomes, however, HU-11/CBS4732 and NCYC495 are significantly distinct from DL-1. HU-11/CBS4732 and NCYC495 share a nt identity of 99.5% through their whole genomes, including the rDNA regions and LTR-rt. In contrast, HU-11/CBS4732 and DL-1 share a comparatively low nt identity (<95%). Subsequently, the detection of structural variations (SVs) was performed between the HU-11/CBS4732 and NCYC495 genomes. Further analysis revealed that most of detected SVs are errors in the assembly of NCYC495 genome (**Figure 4B**), particularly including: (1) LTR-rt1, 7, and 7R (absent in NCYC495) which need be included in the NCYC495 genome; (2) two large deletions (absent in NCYC495) which need be added at 5' and 3' ends of chromosome 2 of NCYC495; and (3) an over-assembled large segment (absent in HU-11/CBS4732) at 3' end of chromosome 6 (NW\_017264698:1509870-1541475), which need be removed from chromosome 6 of NCYC495. Before the correction of

above errors, (1), (2), and (3) were confirmed by long PacBio subreads. Particularly, (3) was confirmed as the telomeric region at the 3' end of chromosome 6, which was wrongly assembled as the junction region at 5' end of the over-assembled segment in the previous study and the reason is that the copy number of TRs [GGCGGGT]<sub>n</sub> (NW\_017264698:1509840-1509869) in this telomeric region was under-estimated using short sequencing data.

Two large deletions [one type of SV (Zhang et al., 2016)] in the NCYC495 genome (As described above) are “false-positive” SVs caused by the misassembly of LDSs (**Supplementary File 1**) in the subtelomeric regions. In contrast, all LDSs were correctly assembled in the HU-11/CBS4732 genome. Using long (>30 Kb) PacBio subreads, human curation (see section “Materials and Methods”) was performed to verify the locations of the LDSs, particularly three LDSs named LDS1, 2 and 3 with lengths of ~27,850, ~5,100, and ~2,500 bp, respectively (**Figure 4B**): (1) LDS1 and its paralog are present at 3' ends of chromosome 2 and 1 in the HU-11/CBS4732 genome, respectively, while the paralog of LDS1 was correctly assembled into 3' end of chromosome 1 of NCYC495, but a 14,090 bp part of LDS1 was not assembled into 3' end of chromosome 2, which corresponds to a large deletion; and (2) LDS2 and its paralog are present at 5' ends of both chromosomes 2 and 5 in the HU-11/CBS4732 genome, while the paralog of LDS2 was correctly assembled into 5' end of chromosome 5, but LDS2 was not assembled into 5' end of chromosome 2 of NCYC495, which corresponds to the other large deletion. Different from LDS1 and LDS2, LDS3 and its paralog were correctly assembled in the NCYC495 genome. LDS3 is downstream of LDS2 in chromosome 2, and the paralog of LDS3 is present at 5' end of chromosome 7 (**Figure 4B**). LDS1 and 2 had not been discovered before the present study, mainly because they are nearly identical to their paralogs. Particularly, there are only four mismatches and one 1-bp gap between LDS1 and its paralog. As an important finding, telomeric TR-like sequences [ACCCGCC]<sub>n</sub> or [ACCCGCC]<sub>n</sub> (n > 2) were discovered at 3' ends of LDS2 and its paralog (located on both chromosomes 2 and 5), and at 3' end of LDS3's paralog (located in chromosome 7). The discovery of these sequences (**Supplementary File 1**) indicated that these LDSs were integrated at 5' ends of the old telomeric regions by their 3' ends.

## Structural Variations Between HU-11/CBS4732 and NCYC495

The major errors in the assembly of NCYC495 genome (RefSeq: NW\_017264698-704) include incomplete rDNAs, misassembly of LTR-rt, under-assembly of LDSs and over-assembly of a large segment. After we corrected these errors using the syntenic regions, only four SVs between HU-11/CBS4732 and NCYC495 remained and were named as SV1, SV2, SV3, and SV4 (**Figure 4B**). Further analysis revealed that SV4 is a very special “false-positive” SV. Actually, SV4 is a 22.6-Kb DNA region (**Figure 4C**) which functions in the determination of the yeast mating-type (MAT). According to previous studies (Hanson et al., 2017), yeast mating generally occurs between two haploid cells with opposite genotypes (MAT $\alpha$  and MAT $\alpha$ )

at this locus, to form a diploid zygote (MAT $\alpha$ / $\alpha$ ). *Ogataea* spp. contain both a MAT $\alpha$  locus and a MAT $\alpha$  locus in chromosome 5, approximately 19 Kb apart (**Figure 4C**). The two MAT loci are beside two copies of an identical 2-Kb DNA sequence, which form two inverted repeats (IRs). During MAT switching, the two copies of the IR recombine, inverting the orientation of the 19-Kb region relative to the rest of the chromosome. The MAT locus proximal to the centromere is not transcribed, probably due to silencing by centromeric heterochromatin, whereas the distal MAT locus is transcribed. The HU-11 genome contains a 22.6-Kb MAT region (MAT-HU11) where MAT $\alpha$  can be transcribed, while the NCYC495 genome contains a 22.6-Kb MAT region (MAT-NCYC495) where MAT $\alpha$  can be transcribed. There is only one 1-bp gap between the large segments MAT-HU11 and the reverse-complementary sequence of MAT-NCYC495 (**Supplementary File 1**). Therefore, the HU-11 (GenBank: CP073033-40) and NCYC495 (RefSeq: NW\_017264698-704) genomes represent genomes of *O. polymorpha* MAT $\alpha$  and MAT $\alpha$  cells, respectively. MAT regions can't be used as a genomic marker to characterize different *O. polymorpha* strains, as MAT switching can be induced without environmental signals. For example, we found that MAT switching of HU-11 even occur under optimal growth

conditions (pH = 5.5,  $T = 32^{\circ}\text{C}$ ), although the frequency is extremely low (1/264).

Only three SVs (SV1, SV2, and SV3) are true-positive. SV1 and SV2 are present at 5' and 3' ends of chromosome 4, respectively, while the location of SV3 is close to 5' ends of chromosome 5 (**Figure 4C**). Five sequences involved in these three SVs are SV1-CBS4732 and SV2-CBS4732 in the HU-11/CBS4732 genome and SV1-NCYC495, SV2-NCYC495, and SV3-NCYC495 in the NCYC495 genome (**Supplementary File 1**). These five sequences can be used to identify *O. polymorpha* strains, particularly CBS4732, NCYC495, and their derivative strains. Blasting the five sequences to the NCBI NT database, we found that SV1-CBS4732 and SV2-NCYC495 are nearly identical (>98%) to their orthologs at 5' and 3' ends of chromosome 4 in the DL-1 genome (GenBank: CP080319), respectively, while SV1-NCYC495 and SV2-CBS4732 have no homologs in chromosome 4 of DL-1. As an insertion into the NCYC495 genome, SV3-NCYC495 has a very high nt sequence identity (>91%) to its homolog in the DL-1 genome. Further analysis showed that the three SVs are located in three large syntenic regions (SRs) of HU-11/CBS4732, NCYC495, and DL-1 genomes with very high nt sequence identities (>95%). Three SRs are: (1) SR1 with a length of 161,844 bp at 5' ends of chromosome 4; (2) SR2 with a length

**TABLE 2** | Twenty five different genes between HU-11/CBS4732 and NCYC495.

Gene	Locus	Orthologs (CBS4732/NCYC495/DL-1)	Function
OGAPO_03767	SV1- CBS4732	OGAPO_03767/-/HPODL_03767	12-oxophytodienoate reductase 3 (OPR3)
OGAPO_03766	SV1- CBS4732	OGAPO_03766/-/HPODL_03766	Aminotriazole resistance protein
OGAPODRAFT_24127	SV1-NCYC495	-/OGAPODRAFT_24127/-	Myo-inositol transporter 1
OGAPODRAFT_16381	SV1-NCYC495	-/OGAPODRAFT_16381/-	Aldo keto reductase (ARK)
OGAPODRAFT_24129	SV1-NCYC495	-/OGAPODRAFT_24129/-	Amidase
OGAPODRAFT_12876	SV1-NCYC495	-/OGAPODRAFT_12876/-	MFS transporter
OGAPODRAFT_16382	SV1-NCYC495	-/OGAPODRAFT_16382/-	NADP-dependent alcohol dehydrogenase 6
OGAPO_00001	SV2- CBS4732	OGAPO_00001/-/-	Aminotriazole resistance protein
OGAPO_00002	SV2- CBS4732	OGAPO_00002/-/-	Aryl-alcohol dehydrogenase
OGAPO_00003	SV2- CBS4732	OGAPO_00003/-/-	Sterol regulatory element-binding protein ECM22
OGAPO_00004	SV2- CBS4732	OGAPO_00004/-/-	Agmatine ureohydrolase
OGAPO_00005	SV2- CBS4732	OGAPO_00005/-/-	P-loop containing nucleoside triphosphate hydrolase protein
OGAPO_00006	SV2- CBS4732	OGAPO_00006/-/-	MFS general substrate transporter
OGAPO_00007	SV2- CBS4732	OGAPO_00010/-/-	Acetylornithine aminotransferase, mitochondrial (ARG8)
OGAPO_00008	SV2- CBS4732	OGAPO_00012/-/-	Aldo keto reductase (ARK)
OGAPODRAFT_13497*	SV2-NCYC495	OGAPO_13497*/HPODL_00892	Basic amino-acid permease
OGAPODRAFT_76936	SV2-NCYC495	-/OGAPODRAFT_76936/HPODL_00891	Transcriptional activator protein DAL81
OGAPODRAFT_16706	SV2-NCYC495	-/OGAPODRAFT_16706/HPODL_00890	DUF1479-domain-containing protein
OGAPODRAFT_37951	SV2-NCYC495	-/OGAPODRAFT_37951/HPODL_02394	MFS domain-containing protein
OGAPODRAFT_93168	SV3-NCYC495	-/OGAPODRAFT_93168/HPODL_04518	MFS domain-containing protein
OGAPODRAFT_15973*	SV3-NCYC495	OGAPO_15973*/HPODL_04520	MFS sugar transporter
OGAPODRAFT_75778	SV3-NCYC495	-/OGAPODRAFT_75778/HPODL_04517	Adenosine deaminase
OGAPODRAFT_75779	SV3-NCYC495	-/OGAPODRAFT_75779/HPODL_04516	Zn(2)-C6 fungal-type domain-containing protein

The genomic differences between *Ogataea polymorpha* HU-11/CBS4732 and NCYC495 include SNPs, small InDels, and only three SVs (**Figure 4B**). Five sequences (SV1-CBS4732, SV2-CBS4732, SV1-NCYC495, SV2-NCYC495, and SV3-NCYC495) were involved in these three SVs. Only 25 genes (**Supplementary File 1**) were involved in the three SVs between HU-11/CBS4732 and NCYC495. 10 genes (OGAPO\_03766-67 and OGAPO\_00003-08) of HU-11/CBS4732 and 11 genes (OGAPODRAFT\_24127, 16381, 24129, 12876, 16382, 76936, 16706, 37951, 93168, 75778, and 75779) of NCYC495 have no orthologs in NCYC495 and HU-11/CBS4732, respectively. \*Two genes (OGAPODRAFT\_13497 and 15973) in NCYC495 were significantly changed into two other ones (OGAPO\_13497 and 15973) in HU-11/CBS4732. Six genes (OGAPO\_00003-08) in a major part of SV2-CBS4732 have no homologs in NCYC495 or DL-1. OGAPO\_, OGAPODRAFT\_, and HPODL\_ are prefix of gene IDs of HU-11/CBS4732, NCYC495, and DL-1, respectively.

of 81,748 bp at 3' ends of chromosome 4; and (3) SR3 with a length of 11,087 bp close to 5' ends of chromosome 5. The above results revealed that many recombination events occurred in chromosome 4 and 5 of CBS4732 and NCYC495' ancestors after their divergence, particularly: (1) recombination events occurred at 5' end of chromosome 4 of the NCYC495' ancestor, resulting in the acquisition of SV1-NCYC495; (2) recombination events occurred at 3' end of chromosome 4 of the CBS4732' ancestor, resulting in the acquisition of SV2-CBS4732; (3) recombination events occurred close to 5' end of chromosome 5 of the CBS4732' ancestor, resulting in the loss of SV3-CBS4732 (the hypothetical homolog of SV3-NCYC495).

Only a few genes (predicted as 25) were involved in the three SVs between HU-11/CBS4732 and NCYC495 (Table 2). Among the 25 genes (Table 2), 10 genes (OGAPO\_03766-67 and OGAPO\_00003-08) of HU-11/CBS4732 and 11 genes (OGAPODRAFT\_24127, 16381, 24129, 12876, 16382, 76936, 16706, 37951, 93168, 75778, and 75779) of NCYC495 have no orthologs in NCYC495 and HU-11/CBS4732, respectively and two genes (OGAPODRAFT\_13497 and 15973) in NCYC495 were significantly changed into two other ones (OGAPO\_13497 and 15973) in HU-11/CBS4732, resulting in different aa sequences. Blasting the proteins encoded by these 25 genes (Supplementary File 1) to the UniProt database, we found that the proteins encoded by five genes (OGAPODRAFT\_24127, 16381, 24129, 12876, and 16382) in SV1-NCYC495 have the highest sequence similarities to their homologs encoded by six genes (HPODL\_02401, 02402, 02403, 02404, 02405, and 02398) at 3' end of chromosome 6 (RefSeq: NC\_027860) in the DL-1 genome. The above results suggest that SV1-NCYC495 from chromosome 6 of NCYC495' ancestor was acquired by chromosome 4 of NCYC495 via translocation. The proteins encoded by six genes (OGAPO\_00003-08) in a major part (more than 80%) of SV2-CBS4732 have no homologs in *Ogataea polymorpha* NCYC495 or DL-1, but have the highest sequence similarities to their homologs in *O. thermophila*, followed by *O. philodendri* and *O. hagerorum*. These six proteins also have homologs in the *Cyberlindnera jadinii* strain NRRL Y-1542. Furthermore, we found that the proteins encoded by two genes (OGAPO\_00001-02) in the minor part of SV2-CBS4732 (from chromosome 4) have the highest sequence similarities to their homologs encoded by genes in other chromosomes. These findings revealed more combination events occurred between chromosome 4 and other chromosomes within the genome of NCYC495' ancestor or CBS4732' ancestor.

## CONCLUSION

The *O. polymorpha* strain CBS4732 *ura3* $\Delta$  (named HU-11) is a nutritionally deficient mutant derived from CBS4732 by a 5-bp insertion of "GAAGT" into the 32nd position of the *URA3* CDS; this insertion causes a frame-shift mutation of the *URA3* CDS, resulting in the loss of the *URA3* functions. Since the difference between the genomes of CBS4732 and HU-11 is only five nts, HU-11 has the same reference genome

as CBS4732 (noted as HU-11/CBS4732). In the present study, we have assembled the full-length genome of *O. polymorpha* HU-11/CBS4732 using high-depth PacBio and Illumina data. 5' and 3' telomeric, subtelomeric, rDNA, LTR-rts, low complexity, and other repeat regions were curated to improve the genome quality. In brief, the main findings include complete rDNAs, complete LTR-rts, three LDSs in subtelomeric regions and three SVs between the HU-11/CBS4732 and NCYC495 genomes. SV1, LDS1, LDS2, and LDS3 were validated using the strand-specific RNA-seq data of NCYC495 (SRA: SRP124832). These findings are very important for the assembly of full-length genomes of yeast and the correction of assembly errors in the published genomes of *Ogataea* spp.

The present study preliminarily revealed the relationship between *O. polymorpha* CBS4732, NCYC495, and DL-1. HU-11/CBS4732 is so phylogenetically close to NCYC495 that the syntenic regions cover nearly 100% of their genomes. Moreover, HU-11/CBS4732 and NCYC495 share a nucleotide identity of 99.5% through their whole genomes, including the rDNA regions and LTR-rts. The genomic differences between HU-11/CBS4732 and NCYC495 include SNPs, small InDels, and only three SVs. CBS4732 and NCYC495 can be regarded as the same strain in basic research and industrial applications. The HU-11 (GenBank: CP073033-40) and NCYC495 (RefSeq: NW\_017264698-704) genomes represent genomes of *O. polymorpha* MAT $\alpha$  and MAT $\alpha$  cells, respectively. Large segments SV1-CBS4732, SV2-CBS4732, SV1-NCYC495, SV2-NCYC495, and SV3-NCYC495 involved in the three SVs can be used to identify *O. polymorpha* strains, particularly CBS4732, NCYC495, and their derivative strains. Among these five large segments, SV2-CBS4732 merits further investigation. The proteins encoded by six genes in a major part of SV2-CBS4732 have no homologs in *Ogataea polymorpha* NCYC495 or DL-1, but have the highest sequence similarities to their homologs in *O. thermophila*, followed by *O. philodendri* and *O. hagerorum*. These six proteins also have homologs in the *Cyberlindnera jadinii* strain NRRL Y-1542. As most genome sequences do not contain complete subtelomeric regions where the six genes locate, the origin of these genes are still not determined.

Only with the exact sequences of subtelomeric regions, can we discover the SV1 and SV2. Furthermore, we reported for the first time LDSs in the subtelomeric regions of *Ogataea* genomes. LDS1 and LDS2 had not been discovered before the present study, mainly because they are nearly identical to their paralogs. A computational study (Brown, 2010) showed that subtelomeric gene families are evolving and expanding much faster than gene families which do not contain subtelomeric genes in yeasts. This previous study also concluded that the extraordinary instability of eukaryotic subtelomeres supports rapid adaptation to novel niches by promoting gene recombination and duplication followed by the functional divergence of the alleles. Our results indicated that large segment duplication in subtelomeric regions occurs in a size to the extent of ~27,850 bp and suggests that the genome expansion in methylotrophic yeasts is mainly driven by large segment duplication in subtelomeric regions. The discovery of telomeric TR-like sequences at 3'

ends of the LDSs indicated that they were integrated at 5' ends of the old telomeric regions by their 3' ends. The exact LDS and telomeric TR-like sequences are very important for the investigation of the molecular mechanism (if *via* recombination or not) that underlies large segment duplication in subtelomeric regions.

## MATERIALS AND METHODS

The *Ogataea polymorpha* strain HU-11 (CGMCC No. 1218) was preserved in the China General Microbiological Culture Collection Center (CGMCC). DNA extraction and quality control were performed as described in our previous study (Wang Y. et al., 2016). A 500 bp DNA library was constructed as described in our previous study (Wang Y. et al., 2016) and sequenced on the Illumina HiSeq X Ten platform. A 10 Kb DNA library was constructed and sequenced on the PacBio Sequel platforms, according to the manufacturer's instruction. The cleaning and quality control of PacBio data was performed with the software SMRTlink v5.0 ( $-\text{minLength} = 50, -\text{minReadScore} = 0.8$ ), while the cleaning and quality control of Illumina data was performed with the software Fastq\_clean (Zhang et al., 2014) v2.0. PacBio data was used to assemble the HU-11/CBS4732 draft genome with the software MECAT (Xiao et al., 2016) v1.2. To polish the draft genome, the software BWA was used to align Illumina data to the HU-11/CBS4732 draft genome. Then, the software samtools was used to obtain the BAM and pileup files from the alignment results. Perl scripts were used to extract the consensus sequence from the pileup file. This polishing procedure was repeatedly performed until human curation started. The reference genomes of *O. polymorpha* HU-11/CBS4732, NCYC495 and DL-1 are available at the NCBI GenBank or RefSeq database under the accession numbers CP073033-40, NW\_017264698-704 and NC\_027860-66. Another genome of *O. polymorpha* DL-1 (GenBank: CP080316-22) was also used for syntenic comparison and SV detection, as this genome is more complete than the genome (RefSeq: NC\_027860-66).

The genome sequences of 34 species in the *Ogataea* or *Candida* genus were downloaded from the Genome-NCBI datasets and their accession numbers were included in **Supplementary File 1**. Syntenic comparison of genomes was performed using the CoGe website,<sup>1</sup> only for visualization. The detailed syntenic comparison and SV detection were performed locally using the software blast v2.9.0 and Perl scripts. Using the software IGV (Thorvaldsdóttir et al., 2013) v2.0.34, human curation of the poly(GC) regions, 5' and 3' telomeric, subtelomeric, rDNA, LTR-rt, low complexity, and other repeat regions was performed with 103,345 long (>20 Kbp) PacBio subreads. The curation criteria is: (1) the junctions between large segments (e.g., LTR-rt, LDSs, or SVs) must be spanned by a long PacBio subread; and (2) the corrected nucleotides must be confirmed by more than 5 long PacBio subreads. To estimate

the frequency of MAT switching, each long PacBio subread was counted with human curation. Statistical computation and plotting were performed using the software R v2.15.3 with the Bioconductor packages (Gao et al., 2014). Prediction of protein-coding genes (>150 bp) was performed using the software AUGUSTUS (Hoff and Stanke, 2013) v2.7.0. Strand-specific RNA-seq data (SRA: SRP124832) was used to curate gene annotations of HU-11/CBS4732, NCYC495 and DL-1. As the reads in the data SRP124832 correspond to the reverse-complementary counterpart of transcripts, they were transformed into their reverse-complementary sequences for all the analyses in the present study.

## DATA AVAILABILITY STATEMENT

The complete genome sequence of the *O. polymorpha* HU-11/CBS4732 is available at the NCBI GenBank database under the accession number CP073033-40, in the project PRJNA687834.

## AUTHOR CONTRIBUTIONS

SG conceived the project and drafted the manuscript. SG and DW supervised the present study. JC assembled the HU-11/CBS4732 genome, prepared the figures, tables, and **Supplementary Material**. JB and HF executed the experiments. SG, QS, and TY analyzed the data. SG, HW, WB, and JR revised the manuscript. All authors have read and approved the manuscript.

## FUNDING

This work was supported by the Natural Science Foundation of China (31872388) to HF, Natural Science Foundation of Guangdong Province of China (2021A1515011072) to JB, and Tianjin Key Research and Development Program of China (19YFZCSY00500) to SG. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## ACKNOWLEDGMENTS

We appreciate the help equally from the people listed below. They are Dawei Huang, Huaijun Xue, Yanqiang Liu, Bingjun He, Qiang Zhao, and Zhen Ye from College of Life Sciences, Nankai University.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.855666/full#supplementary-material>

<sup>1</sup><https://genomeevolution.org/CoGe>



## REFERENCES

- Agrawal, S., and Ganley, A. (2018). The conservation landscape of the human ribosomal RNA gene repeats. *PLoS One* 13:e0207531. doi: 10.1371/journal.pone.0207531
- Brown, C. A. (2010). Rapid expansion and functional divergence of subtelomeric gene families in yeast. *Curr. Biol.* 20, 895–903. doi: 10.1016/j.cub.2010.04.027
- Gao, S., Ou, J., and Xiao, K. (2014). *R Language and Bioconductor in Bioinformatics Applications (Chinese Edition)*. Tianjin: Tianjin Science and Technology Translation Publishing Ltd.
- Hanson, S. J., Byrne, K. P., Wolfe, K. H., and Joseph, H. (2017). Flip/flop mating-type switching in the methylotrophic yeast *Ogataea polymorpha* is regulated by an Efg1-Rme1-Ste12 pathway. *PLoS Genet.* 13:e1007092. doi: 10.1371/journal.pgen.1007092
- Hoff, K. J., and Stanke, M. (2013). WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Res.* 41, 123–128. doi: 10.1093/nar/gkt418
- Levine, D. W., and Cooney, C. L. (1973). Isolation and characterization of a thermotolerant methanol-utilizing yeast. *Appl. Environ. Microbiol.* 26, 982–990. doi: 10.1128/am.26.6.982-990.1973
- Massoud, R. R., Hollenberg, C. P., Juergen, L., Holger, W., Eike, G., Christian, W., et al. (2003). The *Hansenula polymorpha* (strain CBS4732) genome sequencing and analysis. *FEMS Yeast Res.* 4, 207–215. doi: 10.1016/S1567-1356(03)00125-9
- Morais, J. O. F., and Maia, M. H. D. (1959). Estudos de microorganismos encontrados em leitos de despejos de caldas de destilarias de Pernambuco. II. Uma nova especie de *Hansenula: H. polymorpha*. *An. Esc. Super. Quim. Univ. Recife* 1, 15–20.
- Neueglise, C., Feldmann, H., Bon, E., Gaillardin, C., and Casaregola, S. (2002). Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. *Genome Res.* 12, 930–943. doi: 10.1101/gr.219202
- Ravin, N. V., Eldarov, M. A., Kadnikov, V. V., Beletsky, A. V., and Skryabin, K. G. (2013). Genome sequence and analysis of methylotrophic yeast *Hansenula polymorpha* DL1. *BMC Genomics* 14:837. doi: 10.1186/1471-2164-14-837
- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 14, 178–192. doi: 10.1093/bib/bbs017
- Wang, H., He, X., and Zhang, B. (2007). *A Method to Construct Ogataea polymorpha Strains and Its Application*. CN: 200410080517.2.
- Wang, H., Wang, C., Wang, Y., and Yang, J. (2011). *A Recombinant Hirudin Gene and Its Application*. CN: 200810103154.8.
- Wang, H., Wang, C., and Yang, J. (2016). *A High-Dose Recombinant B Hepatitis Vaccine Expressed in Ogataea spp.* CN: 201610178526.8.
- Wang, Y., Wang, Z., Chen, X., Zhang, H., Guo, F., Zhang, K., et al. (2016). The complete genome of *Brucella suis* 019 provides insights on cross-species infection. *Genes* 7, 1–12. doi: 10.3390/genes7020007
- Wickerham, L. J. (1951). Taxonomy of yeasts. *Tech. Bull. U. S. Dep. Agric.* 6, 781–782.
- Xiao, C. L., Chen, Y., Xie, S. Q., Chen, K. N., and Xie, Z. (2016). MECAT: an ultra-fast mapping, error correction and de novo assembly tool for single-molecule sequencing reads. *bioRxiv* [Preprint] doi: 10.1101/089250
- Xu, X., Bei, J., Xuan, Y., Chen, J., Chen, D., Barker, S. C., et al. (2020). Full-length genome sequence of segmented RNA virus from ticks was obtained using small RNA sequencing data. *BMC Genomics* 21:641. doi: 10.1186/s12864-020-07060-5
- Zhang, F., Xu, T., Mao, L., Yan, S., Chen, X., Wu, Z., et al. (2016). Genome-wide analysis of Dongxiang wild rice (*Oryza rufipogon* Griff.) to investigate lost/acquired genes during rice domestication. *BMC Plant Biol.* 16:103. doi: 10.1186/s12870-016-0788-2
- Zhang, M., Zhan, F., Sun, H., Gong, X., Fei, Z., and Gao, S. (2014). “Fastq\_clean: an optimized pipeline to clean the Illumina sequencing data with quality control,” in *Proceedings of the 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, (Belfast: IEEE).

**Conflict of Interest:** HW was employees by Tianjin Hemu Health Biotechnological Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chang, Bei, Shao, Wang, Fan, Yau, Bu, Ruan, Wei and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## **APPENDIX 2**

### **CONFERENCE ABSTRACTS REVERENT TO THE THESIS TOPIC**

Order	Title	Authors	Journal Name	Access domains / Identifier
<b>Noncoding RNA and Colorectal Cancer</b>				
1	A lncRNA-histone acetyltransferase complex induces colorectal oncogenesis through regulation of a metabolic kinase network	Christos Polytarchou, Maria Hatzia Apostolou, Marina Koutsioumpa, Niki Christodoulou, <b>Tung On Yau</b> , Eleni Birli, Odette Pomenya, Cristina Montiel-Duarte, Swapna Mahurkar Joshi, Daniel W. Hommes, Hein W. Verspaget, Jun Yu and Dimitrios Iliopoulos	<i>Cancer Research</i> . <b>2020</b> ;80(16 Suppl): Abstract nr LB-184.	Journal URL: <a href="https://cancerres.aacrjournals.org/content/80/16_Supplement/LB-184">cancerres.aacrjournals.org/content/80/16_Supplement/LB-184</a> DOI: <a href="https://doi.org/10.1158/1538-7445.AM2020-LB-184">10.1158/1538-7445.AM2020-LB-184</a>
2	A Novel MicroRNA Panel for Non-Invasive Diagnosis and Prognosis of Colorectal Cancer	Jessie Qiaoyi Liang, <b>Tung On Yau</b> , Tsz Ching Dorothy Yau, Chun Ho Szeto, Flora Sha Zhao, Francis K. Chan, Joseph J. Sung, Jun Yu	<i>Gastroenterology</i> . <b>2019</b> ;156 (6), S-496.	Journal URL: <a href="https://gastrojournal.org/article/S0016-5085(19)38106-5/fulltext">gastrojournal.org/article/S0016-5085(19)38106-5/fulltext</a> ScienceDirect: <a href="https://sciedirect.com/science/article/abs/pii/S0016508519381065">sciedirect.com/science/article/abs/pii/S0016508519381065</a> DOI: <a href="https://doi.org/10.1016/S0016-5085(19)38106-5">10.1016/S0016-5085(19)38106-5</a>
3	Faecal-based noninvasive biomarkers microRNA-221 and microRNA-18a for colorectal cancer screening	<b>Tung On Yau</b> , Chung Wah Wu, Joseph Jao Yiu Sung, Jun Yu, Ceen-Ming Tang	<i>Anticancer Research</i> ; <b>2015</b> ;35(2),1140-1141.	URL: <a href="https://ar.iiarjournals.org/content/35/2/1139">ar.iiarjournals.org/content/35/2/1139</a>
4	Establishment of Taqman Probe-Based Quantitative PCR Assays for Evaluation of Bacterial Markers in Human Fecal Samples	Qiaoyi Liang, <b>Tung On Yau</b> , Francis KL Chan, Joseph JY Sung, Jun Yu	<i>Gastroenterology</i> . <b>2014</b> ;146(5), S-179-S-180.	Journal URL: <a href="https://gastrojournal.org/article/S0016-5085(14)60639-9/fulltext">gastrojournal.org/article/S0016-5085(14)60639-9/fulltext</a> ScienceDirect: <a href="https://sciedirect.com/science/article/abs/pii/S0016508514606399">sciedirect.com/science/article/abs/pii/S0016508514606399</a> DOI: <a href="https://doi.org/10.1016/S0016-5085(14)60639-9">10.1016/S0016-5085(14)60639-9</a>
<b>Clostridioides difficile infection</b>				
5	Circulating microRNAs linked to immunometabolic traits associate with faecal microbiota transplantation for <i>clostridioides difficile</i> infection	Tanya Monaghan, Tahseen Jilani, Marcin Frankowski, Odette Pomenya, <b>Tung On Yau</b> , Niki Christodoulou, Maria Hatzia Apostolou, Iwona Wojcik, Maja Pucic-Bakovic, Frano Vuckovic, Thomas Louie, Gordan Lauc, Dina Kao, Christos Polytarchou	<i>Gut</i> . <b>2019</b> ;68, A185.	Journal URL: <a href="https://gut.bmj.com/content/68/Suppl_2/A185.2">gut.bmj.com/content/68/Suppl_2/A185.2</a> DOI: <a href="https://doi.org/10.1136/gutjnl-2019-BSGAbstracts.353">10.1136/gutjnl-2019-BSGAbstracts.353</a>
6	Fecal Microbiota Transplantation Regulates Circulating microRNA Expression Levels in Patients with Recurrent <i>Clostridioides</i> Infection	Tanya Monaghan, Odette Pomenya, <b>Tung On Yau</b> , Niki Christodoulou, Maria Hatzia Apostolou, Tahseen Jilani, Dina H. Kao, Christos Polytarchou	<i>Gastroenterology</i> . <b>2019</b> ;156(6), S-85.	Journal URL: <a href="https://gastrojournal.org/article/S0016-5085(19)37000-3/fulltext">gastrojournal.org/article/S0016-5085(19)37000-3/fulltext</a> ScienceDirect: <a href="https://sciedirect.com/science/article/abs/pii/S0016508519370003">sciedirect.com/science/article/abs/pii/S0016508519370003</a> DOI: <a href="https://doi.org/10.1016/S0016-5085(19)37000-3">10.1016/S0016-5085(19)37000-3</a>

Order	Title	Authors	Journal Name	Access domains / Identifier
<b>Epstein-Barr virus-associated Gastric Cancer</b>				
7	Integrative Identification of EBV-Associated Variations at Genomic, Epigenomic and Transcriptomic Levels in Gastric Cancer	Qiaoyi Liang, Xiaotian Yao, Senwei Tang, <b>Tung On Yau</b> , Junhong Zhao, Joseph J.Y.Sung, Jun Yu	<i>Gastroenterology</i> . <b>2013</b> ;155 (5), S-525.	Journal URL: <a href="http://gastrojournal.org/article/S0016-5085(13)61949-6/fulltext">gastrojournal.org/article/S0016-5085(13)61949-6/fulltext</a> ScienceDirect: <a href="http://sciencedirect.com/science/article/abs/pii/S0016508513619526">sciencedirect.com/science/article/abs/pii/S0016508513619526</a> DOI: <a href="https://doi.org/10.1016/S0016-5085(13)61952-6">1016/S0016-5085(13)61952-6</a>
<b>Non-Alcoholic Steatohepatitis</b>				
8	Hepatic CXCR3 Promotes Non-Alcoholic Steatohepatitis Through Inflammation, Lipid Accumulation and Autophagy Deficiency	Xiang Zhang, Eagle SH Chu, <b>Tung On Yau</b> , Xiaoxing Li, Joseph J.Y. Sung, Jun Yu	<i>Gastroenterology</i> . <b>2014</b> ;146(5), S-992.	Journal URL: <a href="http://gastrojournal.org/article/S0016-5085(14)63350-3/fulltext">gastrojournal.org/article/S0016-5085(14)63350-3/fulltext</a> ScienceDirect: <a href="http://sciencedirect.com/science/article/abs/pii/S0016508514633503">sciencedirect.com/science/article/abs/pii/S0016508514633503</a> DOI: <a href="https://doi.org/10.1016/S0016-5085(14)63350-3">10.1016/S0016-5085(14)63350-3</a>
9	The role of CXCR3 in the pathogenesis of non-alcoholic steatohepatitis.	Xiang Zhang, Eagle S.H. Chu, <b>Tung On Yau</b> , Xiaoxing Li, Joseph JY Sung, Jun Yu	<i>Hepatology International</i> . <b>2014</b> ;8(S1), S339-340.	Journal URL: <a href="http://link.springer.com/article/10.1007/s12072-014-9519-7">link.springer.com/article/10.1007/s12072-014-9519-7</a> DOI: <a href="https://doi.org/10.1007/s12072-014-9519-7">10.1007/s12072-014-9519-7</a>
10	Role of Interferon $\gamma$ -Inducible Protein 10 in the Pathogenesis of Non-Alcoholic Steatohepatitis	Xiang Zhang, Eagle SH Chu, <b>Tung On Yau</b> , Xiaoxing Li, Joseph J.Y. Sung, Jun Yu	<i>Gastroenterology</i> ; <b>2013</b> ;144(5), S-948.	Journal URL: <a href="http://gastrojournal.org/article/S0016-5085(13)63525-8/fulltext">gastrojournal.org/article/S0016-5085(13)63525-8/fulltext</a> ScienceDirect: <a href="http://sciencedirect.com/science/article/abs/pii/S0016508513635258">sciencedirect.com/science/article/abs/pii/S0016508513635258</a> DOI: <a href="https://doi.org/10.1016/S0016-5085(13)63525-8">10.1016/S0016-5085(13)63525-8</a>
11	CXCL10 Induces Hepatocyte Apoptosis and Autophagy in Experimental Non-Alcoholic Steatohepatitis	Xiang Zhang, Xiaojuan Wang, Eagle SH Chu, <b>Tung On Yau</b> , Minnie Y. Go, Joseph J.Y. Sung, Jun Yu	<i>Gastroenterology</i> ; <b>2013</b> ;144(5), S-949.	Journal URL: <a href="http://gastrojournal.org/article/S0016-5085(13)63529-5/fulltext">gastrojournal.org/article/S0016-5085(13)63529-5/fulltext</a> ScienceDirect: <a href="http://sciencedirect.com/science/article/abs/pii/S0016508513635295">sciencedirect.com/science/article/abs/pii/S0016508513635295</a> DOI: <a href="https://doi.org/10.1016/S0016-5085(13)63529-5">10.1016/S0016-5085(13)63529-5</a>

## A NOVEL MICRORNA PANEL FOR NON-INVASIVE DIAGNOSIS AND PROGNOSIS OF COLORECTAL CANCER

Jessie Qiaoyi Liang, **Tung On Yau**, Tsz Ching Dorothy Yau, Chun Ho Szeto, Flora Sha Zhao, Francis K. Chan, Joseph J. Sung, Jun Yu

**Background and Aim:** MicroRNAs play important roles in the development of colorectal cancer (CRC). Multiple miRNAs have shown to be of diagnostic and/or prognostic value for CRC, but their clinical application is limited due to application of non-targeted methods or single target detection. In this study, we identified and evaluated the utility of a new panel of miRNAs in the stool-based non-invasive diagnosis and mucosa-based prognosis of CRC. **Experimental Design:** Stool samples from 381 subjects (184 CRC, 60 advanced adenoma and 137 control subjects) and primary CRC tissues from 123 patients were collected. A panel of miRNAs was selected by analyzing genome-wide miRNA expression profiles in CRC. A multiplex RT-qPCR assay and scoring algorithms for diagnosis and prognosis were devised. **Results:** By integrative analysis of TCGA small RNA sequencing data and in-house miRNA array data, a panel of 5 miRNAs differentially expressed in CRC tissues compared to normal colon tissues (miR-92a, miR-21, miR-135b, miR-145 and miR-133a) was selected. Then a stem-loop and probe based multiplex RT-qPCR assay was established for the convenient quantification of the five miRNAs. A scoring algorithm to combine all five miRNAs for CRC diagnosis (C-index) was trained by logistic regression on qPCR data from a training cohort of 60 CRC and 60 control fecal samples. Results from the validation cohort showed that, among the individual miRNAs, fecal miR-92a performed best in distinguishing CRC patients from controls, with an area under receiver operating curve (AUROC) of 0.782 (sensitivity=72% and specificity=72% by Youden's index method). C-index showed significantly improved diagnostic performance compared to individual miRNAs, with an AUROC of 0.849 ( $P=0.001$  vs miR-92a by pairwise comparison of ROCs). At 80% specificity, C-index showed a sensitivity of 81% for CRC diagnosis, which was further improved by fecal immunochemical test (FIT) to 90% ( $P=0.010$ ). For detection of advanced adenoma (specificity=80%), sensitivity of C-index (33%) was significantly higher than FIT (17%,  $P=0.035$ ) and was improved to 43% by combining with FIT. Moreover, another scoring algorithm for prognosis (P-index) was developed to combine mucosal miR-21, miR-92a, miR-145 and miR-133a by proportional-hazards regression models. Kaplan-Meier survival analysis showed that a high P-index was significantly associated with shortened survival in CRC patients (HR=3.74 (95% CI: 1.93 to 7.24),  $P=9.5e-05$ ). Multivariate analysis showed that P-index was an independent risk factor for poor survival of CRC patients (HR=2.53 (95%CI: 1.18 to 5.42),  $P=0.017$ ). **Conclusions:** This study identified a panel of 5 CRC-related miRNAs and developed a multiplex RT-qPCR and scoring platform that could be conveniently applied in clinical settings for stool-based non-invasive diagnosis and mucosa-based prognosis of CRC.

## FAECAL-BASED NONINVASIVE BIOMARKERS MICRORNA-221 AND MICRORNA-18A FOR COLORECTAL CANCER SCREENING

**Tung On Yau**, Chung Wah Wu, Joseph Jao Yiu Sung, Jun Yu, Ceen-Ming Tang

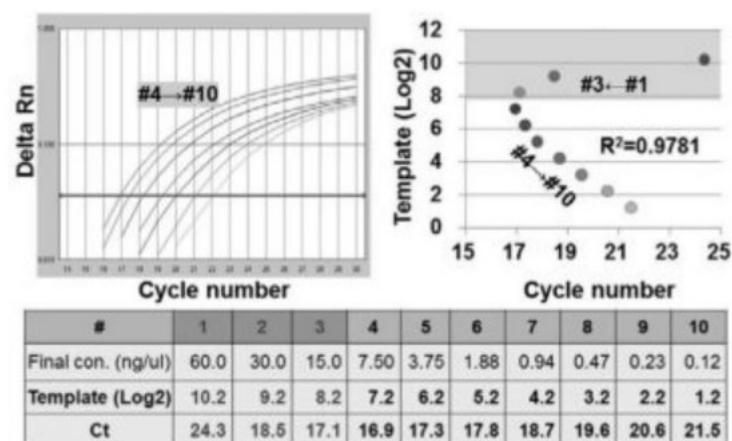
Institute of Digestive Disease, The Chinese University of Hong Kong, Hong Kong; Department of Pharmacology, University of Oxford, UK.

**Objective:** microRNA (miRNA) detection in faeces is a new approach for colorectal cancer (CRC) screening. The study objective was to classify non-invasive biomarkers miR-221 and miR-18a in faecal samples for the screening of CRC. **Methods:** A miRNA microarray containing 667 miRNAs was performed to classify miRNA expression in primary CRC tissues. We focused on two significantly upregulated miRNAs; miR-221 and miR-18a. They were subsequently verified in 40 pairs of CRC tissues and 595 faecal samples (198 healthy controls, 199 polyps, and 198 CRCs). **Results:** The levels of miR-221 and miR-18a were significantly upregulated in the miRNA expression microarray. The levels of miR-221 and miR-18a were also raised in 40 CRC tumours compared to their respective adjacent normal tissues. In faecal samples, miR-221 and miR-18a showed a significant increasing trend from healthy controls to late stages of CRC ( $P < 0.0001$ ). miR-221 and miR-18a levels were both significantly higher in subjects with CRC stages I / II (miR-221:  $P < 0.0001$ , miR-18a:  $P < 0.0001$ ) and CRC stages III / IV (miR-221:  $P = 0.0004$ , miR-18a:  $P < 0.0001$ ) compared to healthy controls. The AUROC of miR-221 and miR-18a were 0.73 and 0.67 for CRC patients compared with healthy controls, respectively. No significant differences in the levels of faecal-based miR-221 and miR-18a were found between patients with proximal, distal and rectal CRC lesions. Antibiotic intake in CRC patients did not alter the levels of faecal-based miRNA-221 and miRNA-18a. **Conclusion:** Faecal-based miR-221 can be utilised as a non-invasive biomarker for CRC screening.

## Establishment of Taqman Probe-Based Quantitative PCR Assays for Evaluation of Bacterial Markers in Human Fecal Samples

Qiaoyi Liang, **Tung On Yau**, Francis K. L. Chan, Joseph J. Y. Sung, Jun Yu

**Background and aim:** With the widespread application of pyrosequencing and metagenome sequencing in the investigation of human intestinal microbiota, disease-associated bacterial markers are emerging. There is an urgent need to develop a reliable, rapid, and cost-effective method to evaluate these markers for potential clinical applications. However, the "absolute" quantities of target markers commonly obtained using the standard curve method, where DNA concentration serves as an internal control for quantitative PCR (qPCR), is only as good as the DNA quantification method and does not eliminate the possibility of human DNA contamination. This study aimed to establish a TaqMan probe-based qPCR assay for quantification of total bacteria to serve as an internal control for future target marker abundance assessment. This internal control assay was further tested with a gene marker from the well-known colorectal cancer (CRC)-associated *Fusobacterium* in human fecal samples. **Methods and Results:** We designed a degenerate primer-probe (VIC-labeled) set with amplicon size suitable for qPCR quantification targeting a 146-bp region of the 16S rRNA genes, covering >90% of the eubacterial population within the Ribosomal Database Project Release version 10.8. Tests using different samples indicated that this internal control assay was capable of evaluating total bacteria with DNA template of 0.5 ng/ul to avoid false-negative results for samples with low *Fusobacterium*. Therefore, reliable quantification of the target marker can be obtained relatively by duplex qPCR in a single reaction with appropriate DNA template concentration, but independently on the basis of absolute DNA quantity. Evaluation in 112 CRC patients and 162 healthy individuals using the duplex qPCR assay showed that the *Fusobacterium* marker abundance was significantly higher in the CRC patients than in the healthy individuals, as expected ( $0.0363 \pm 0.0118$  vs.  $0.0009 \pm 0.0003$ ;  $P < 0.0001$ , Mann Whitney test). **Conclusions:** This study established the first internal control TaqMan probe-based qPCR assay suitable for eubacterial marker quantification. This internal control assay paves the way for easy and reliable evaluation of eubacterial markers using duplex or even multiplex TaqMan probe-based qPCR in the future.



**Figure 1.** An example of qPCR evaluation on serially diluted samples from one fecal DNA sample. qPCR results correlated well with template quantity when final DNA concentrations were 10 ng/ul inhibited PCR amplification.

## CIRCULATING MICRORNAS LINKED TO IMMUNOMETABOLIC TRAITS ASSOCIATE WITH FAECAL MICROBIOTA TRANSPLANTATION FOR CLOSTRIDIODES DIFFICILE INFECTION

1 Tanya Monaghan\*, 2 Tahseen Jilani, 3 Marcin Frankowski, 4 Odette Pomenya, 4 **Tung On Yau**, 4 Niki Christodoulou, 4 Maria Hatzia Apostolou, 5 Iwona Wojcik, 5 Maja Pucic-Bakovic, 5 Frano Vuckovic, 5 Thomas Louie, 6 Gordan Lauc, 7 Dina Kao, 4 Christos Polyarchou.

1 NIHR Nottingham Digestive Diseases Biomedical Research Centre, University Of Nottingham, Nottingham, UK; 2 School of Computer Science, University of Nottingham, Nottingham, UK; 3 Faculty of Chemistry, Adam Mickiewicz University, Poznan, Poland; 4 Department of Biosciences, John van Geest Cancer Centre, School of Science and Technology, Nottingham Trent University, Nottingham, UK; 5 Genos Glycoscience Research Laboratory, Zagreb, Croatia; 6 University of Calgary, Calgary, Canada; 7 University of Alberta, Edmonton, Canada

10.1136/gutjnl-2019-BSGAbstracts.353

**Introduction** The molecular mechanisms underlying successful faecal microbiota transplantation (FMT) for recurrent *C. difficile* infection (rCDI) remain poorly understood. The aim of this study was to characterise alterations in circulating microRNAs and immunometabolic traits following FMT for rCDI. **Methods** We analysed a subset of serum samples previously acquired from a prospective multicentre, randomized trial of FMT delivered by capsule vs colonoscopy in the management of rCDI [NCT02254811]. 126 sera from 42 patients at screening, 4 and 12 weeks post FMT [12M, median age 68.5 yrs (2–1); 30 F, 53 yrs (2–1)] were included. MicroRNA panel v3 and the nCounter platform (Nanostring Tech) were used for the analysis of 800 microRNAs. Quantitative PCR and 3'UTR reporter assays were employed to verify microRNA inflammatory protein targets in colonic epithelial and peripheral blood mononuclear cells. Biometal levels were assessed using inductively coupled plasma mass spectrometry (ICP-MS). Hydrophilic interaction ultra-performance liquid chromatography (HILIC-UPLC) and nano-liquid chromatography coupled with electrospray mass spectrometry (nanoLCESI-MS) were utilised to profile the total serum and IgG Fc N-glycome. Pathway analysis was performed using Metacore software. All statistical analyses including non-parametric longitudinal method (nparLD) followed by Wilcoxon signed-rank test for pairwise comparisons and linear mixed modelling were performed in SPSS v.24 and R 3.5.1. **Results** MicroRNA profiling revealed an upregulation in the levels of 64 circulating microRNAs 4 and 12 weeks following successful FMT for rCDI compared to screening time point. MicroRNA signatures coincide with a reduction in circulating selenium and copper, and regulate levels of interleukin-12B, (IL-12B), IL-18 and fibroblast growth factor-21 (FGF21) as well as serum protein N-glycosylation traits. MicroRNA alterations reveal commonalities with several types of cancer and multiple sclerosis, and link metabolic traits to immune cell survival and differentiation. **Conclusions** These findings contribute to a greater understanding of the molecular mechanisms underlying FMT and identify new potential targets for therapeutic intervention.



## FECAL MICROBIOTA TRANSPLANTATION REGULATES CIRCULATING MICRORNA EXPRESSION LEVELS IN PATIENTS WITH RECURRENT CLOSTRIDIODES INFECTION

Tanya Monaghan, Odette Pomenya, **Tung On Yau**, Niki Christodoulou, Maria Hatzia Apostolou, Tahseen Jilani, Dina H. Kao, Christos Polytarchou

**Background:** MicroRNAs may play a crucial role in bridging communication between the gut microbiome and host to maintain intestinal homeostasis and prevent disease. MicroRNA signatures may harbour diagnostic and prognostic properties. The aim of this study was to characterise alterations in circulating microRNAs following fecal microbiota transplantation (FMT) for recurrent *C. difficile* infection (rCDI). **Methods:** We analysed a subset of sera from a prospective multicentre, randomised trial of FMT delivered by capsule vs colonoscopy in the management of rCDI [NCT02254811]. 126 sera from 42 patients at screening, 4 and 12 weeks post FMT [12M, median age 68.5 yrs (28-81); 30F, 53 yrs (20-91)] were included. MicroRNA panel v3 and the nCounter platform (Nanostring Tech) were used for the analysis of 800 microRNAs. Circulating microRNAs were isolated using the miRNeasy Serum/Plasma Kit (Qiagen), purified and concentrated using Amicon Ultra YM-3 columns (Millipore). Colonic epithelial and peripheral blood mononuclear cells were transfected with microRNAs using Lipofectamine RNAiMax (Life Technologies). Total cell RNA was isolated with RNeasy Plus Mini Kit (Qiagen) and subjected to RT and qPCR using iScript Reverse Transcription and iTaq Universal SYBR Green Supermix (Bio-Rad). TargetScan was employed for microRNA target prediction. The 3'UTRs of microRNA targets were cloned in the psi-Check2 reporter vector (Promega) and target sequences were mutated using QuikChange II site-directed mutagenesis kit (Agilent). 3'UTR activities were assessed using the Dual-Luciferase Reporter Assay kit (Promega). MicroRNA changes were examined using a non-parametric longitudinal method (nparLD in R) followed by Wilcoxon signed-rank test for pairwise comparisons. Statistical significance was set at an alpha of 0.01. Pathway analysis was performed using the Metacore (Thomson Reuters) software. **Results:** MicroRNA profiling identified significant changes in the levels of 71 circulating microRNAs 4 and 12 weeks following FMT. Correlation analyses showed that the levels of miR-23a, miR-150 and miR-26b, inversely correlate with the serum protein levels of IL12B, IL18 and FGF21, in the same patients. qPCR and 3'UTR reporter assays verified that miR-23a, miR-150 and miR-26b, target directly the 3'UTR of IL12B, IL18 and FGF21 mRNAs, respectively. Pathway analysis of the microRNA signatures linked inflammatory signalling with metabolic traits, revealing commonalities with autoimmune diseases and cancer. **Conclusions:** To our knowledge, this is the first study reporting differential microRNA expression in the circulation of patients following FMT, and identifies specific microRNA signatures which may help predict the response to FMT. These findings support the need to decipher the specific molecular pathways by which gut microbes influence host microRNA expression following FMT.

## **Integrative Identification of EBV-Associated Variations At Genomic, Epigenomic and Transcriptomic Levels in Gastric Cancer**

Qiaoyi Liang, Xiaotian Yao, Senwei Tang, **Tung On Yau**, Junhong Zhao, Joseph J. Y. Sung, Jun Yu

**Background and Aim:** Epstein-Barr virus (EBV)-associated gastric cancer (GC) represents a distinct subtype of GC with unique clinicopathological features. However, the molecular basis discriminating EBV-associated GC from EBV-negative GC remains largely unknown. We aimed to analyze EBV-associated host variations at genomic, epigenomic and transcriptomic levels in GC. **Methods:** The EBV-infected AGS (AGS-EBV) cells and EBV-negative AGS cells were subjected to whole genome and transcriptome sequencing on an Illumina HiSeq 2000 platform. Genome-wide DNA methylation profiles were analyzed by Methyl-DNA immunoprecipitation-chip assay. Expression of candidate EBV genes was validated in 13 EBV(+)GC samples by RT-PCR. EBV-associated candidate mutations were verified in 20 EBV(+)GCs as compared with 100 EBV(-)GCs and 30 non-cancerous stomach samples by Sanger sequencing. Novel EBV-associated methylated genes were validated in EBV(+) and EBV(-) GC samples using bisulfite genomic sequencing (BGS). The molecular networks dysregulated by EBV infection were determined by KEGG pathway enrichment analysis using WebGestalt. **Results:** Transcriptome analysis of AGS-EBV revealed expression of 9 well-known EBV genes in GC (BARF0, BHRF1, LMP2A, etc) and 68 genes not reported in GC (BNLF2a, BBRF2, BFRF1A, etc). Further examination confirmed expression of BNLF2a, BFRF2 and BFRF1A in 54%, 85% and 77% of primary EBV(+)GC samples respectively. Whole genome sequencing revealed 45 novel non-synonymous single nucleotide variations (SNVs), 11 exonic indels, and 24 genes disrupted by structural variations in AGS-EBV. Novel mutated genes were corroborated to be significantly associated with EBV(+)GCs compared to EBV(-)GCs, including AKT2 ( $P<0.0001$ ), CCNA1 ( $P=0.0013$ ) and TGFBR1 ( $P=0.0301$ ). Integrative epigenome/transcriptome analysis revealed 216 genes downregulated by promoter hypermethylation and 46 genes upregulated by hypomethylation. Transcriptional down-regulations of several novel genes (ACSS1, FAM3B, IHH, NEK9, SLC7A8 and TRABD) mediated by promoter hypermethylation were validated in AGS-EBV compared with AGS by BGS and demethylation re-expression assays. Among them, IHH and FAM3B were further confirmed to be EBV-associated methylated genes in primary EBV(+)GCs. KEGG pathway enrichment analysis revealed two novel pathways (Axon guidance and Focal adhesion) being significantly affected by both genomic alterations and epigenomic hypermethylation caused by EBV infection. **Conclusions:** Expression of novel EBV genes is evident in EBV-associated GC. EBV infection causes host genomic alterations and mainly hypermethylation to dysregulate key pathways such as Axon guidance and Focal adhesion during a subtype of gastric carcinogenesis. This first integrative analysis of EBV-associated variations provides intriguing insight into the pathogenesis of EBV-associated GC.

## Hepatic CXCR3 Promotes Non-Alcoholic Steatohepatitis Through Inflammation, Lipid Accumulation and Autophagy Deficiency

Xiang Zhang, Eagle SH Chu, **Tung On Yau**, Xiaoxing Li, Joseph J. Y. Sung, Jun Yu

**Background and aims:** Cytokines/chemokines and their receptors play crucial roles in inflammatory conditions and might be the lynchpin in the transition of simple steatosis to nonalcoholic steatohepatitis (NASH). Chemokine receptor CXCR3 correlates with chronic liver inflammation. However, the specific contribution of CXCR3 to the development of NASH is largely unknown. We aimed to elucidate the role of CXCR3 in the development of NASH and to investigate the potential effect of its antagonist in suppressing NASH development. **Methods:** CXCR3 knockout (CXCR3 KO), wildtype (WT) littermates and db/db mice were fed with control diet or methionine and choline-deficient (MCD) diet to induce NASH for 4 weeks. Mouse AML-12 hepatocytes were cultured in MCD medium in the presence or absence of a CXCR3 specific inhibitor (NIBR2130) for 24 hours. A series of assays including cytokine profiling, protein-DNA binding activity, co-immunoprecipitation of CXCR3 and its ligand were performed. **Results:** WT mice fed with MCD diet developed steatohepatitis and showed 2.5-fold higher hepatic CXCR3 mRNA expression compared to the control diet fed WT mice with normal liver histology ( $P < 0.05$ ). Similarly, CXCR3 mRNA expression increased 4.0-fold in db/db mice fed with MCD than those fed with control diet. Compared with WT littermates, CXCR3 KO mice were more resistant to MCD-induced steatohepatitis as evidenced by the significantly ameliorated histological grading of hepatic inflammation ( $P < 0.0001$ ) and steatosis ( $P < 0.0001$ ); reduced hepatic triglyceride ( $P < 0.0001$ ), lipid peroxide ( $P < 0.05$ ) and serum ALT levels ( $P < 0.0001$ ). Induction of CXCR3 in WT mice fed with MCD was associated with the increased expression of hepatic pro-inflammatory chemokines and cytokines including TNF- $\alpha$ , monocyte chemoattractant protein 1 (MCP-1) and interleukin-5. CXCR3 was also associated with the activation of NF- $\kappa$ B signaling pathway as evidenced by increased NF- $\kappa$ B DNA binding activity and enhanced phospho-NF- $\kappa$ B p65, p50 expression. Moreover, CXCR3 induced steatosis was mediated by increased hepatic lipogenesis through inducing liver X receptors (LXR $\alpha$  and LXR $\beta$ ), sterol regulatory element binding protein-1c (SREBP-1c), stearoyl-CoA desaturase-1 (SCD-1) and FAS. In particular, autophagy deficiency was involved in CXCR3 induced steatohepatitis as indicated by p62/SQSTM1 accumulation. We further revealed that CXCR3 is a functional CXCL10 receptor in hepatocytes responsible for signaling onset of hepatic inflammation in steatohepatitis. Inhibition of CXCR3 using a highly selective CXCR3 antagonist NIBR2130 suppressed MCD-induced hepatocytes injury in AML-12 hepatocytes. **Conclusions:** CXCR3 plays a pivotal role in NASH development by promoting inflammation, fatty acid accumulation and autophagy deficiency. Pharmacologic blockade of CXCR3 is a potential novel approach for NASH treatment.

## The role of CXCR3 in the pathogenesis of non-alcoholic steatohepatitis

Xiang Zhang, Eagle S.H. Chu, **Tung On Yau**, Xiaoxing Li, Joseph JY Sung, Jun Yu

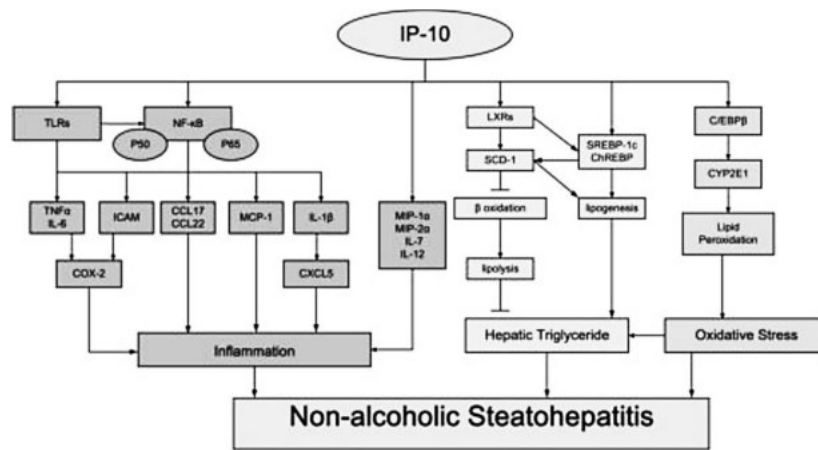
Institute of Digestive Disease and The Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, The Chinese University of Hong Kong, Hong Kong

**Background and aims:** Cytokines/chemokines and their receptors play crucial roles in inflammatory conditions and might be the lynchpin in the transition of simple steatosis to non-alcoholic steatohepatitis (NASH). Chemokine receptors have gained attention as potential target for novel therapeutics. Chemokine receptor CXCR3 correlates with chronic liver fibrosis and hepatitis C. However, the specific contribution of CXCR3 to the development of NASH is largely unknown. We aimed to elucidate the functional significance of CXCR3 in the development of NASH and to investigate the potential effect of its antagonist in suppressing NASH development. **Methods:** CXCR3 knockout (CXCR3<sup>-/-</sup>), wildtype (WT) littermates and db/db mice were fed with control diet or methionine and choline-deficient (MCD) diet to induce NASH for 4 weeks. Mouse AML-12 hepatocytes were cultured in MCD medium in the presence or absence of a CXCR3 specific inhibitor (NIBR2130) for 24 hours. A series of assays including cytokine profiling, protein-DNA binding activity, co-immunoprecipitation of CXCR3 and its ligand were performed. **Results:** WT mice fed with MCD diet developed steatohepatitis and showed 2.5-fold higher hepatic CXCR3 mRNA expression compared to the control diet-fed WT mice with normal liver histology ( $p < 0.05$ ). Similarly, CXCR3 mRNA expression increased 4.0-fold in db/db mice fed with MCD than those fed with control diet. Compared with WT littermates, CXCR3<sup>-/-</sup> mice were more resistant to MCD-induced steatohepatitis as evidenced by the significantly ameliorated histological grading of hepatic inflammation ( $p < 0.0001$ ) and steatosis ( $p < 0.0001$ ); reduced hepatic triglyceride ( $p < 0.0001$ ), lipid peroxide ( $p < 0.05$ ) and serum ALT levels ( $p < 0.0001$ ). Induction of CXCR3 in WT mice fed with MCD was associated with the increased expression of hepatic pro-inflammatory chemokines and cytokines including TNF $\alpha$ , monocyte chemoattractant protein 1 and interleukin-5. CXCR3 was also associated with the activation of nuclear factor- $\kappa$ B (NF- $\kappa$ B) signaling pathways as evidenced by increased NF- $\kappa$ B DNA binding activity, enhanced phospho-NF- $\kappa$ B p65, p50 expression and NF- $\kappa$ B downstream target intercellular adhesion molecule 1 expression. Moreover, CXCR3 induced steatosis was mediated by increased hepatic lipogenesis through inducing the expression of liver X receptors, sterol regulatory element binding protein-1c, stearoyl-CoA desaturase-1 and FAS. In particular, autophagy deficiency was involved in CXCR3 induced steatohepatitis. We further revealed that CXCR3 is a functional CXCL10 receptor in hepatocytes responsible for signaling onset of hepatic inflammation in steatohepatitis. Furthermore, inhibition of CXCR3 using a highly selective CXCR3 antagonist NIBR2130 suppressed MCD-induced hepatocytes injury in AML-12 hepatocytes. **Conclusions:** CXCR3 plays a pivotal role in NASH development by promoting inflammation, fatty acid accumulation and autophagy deficiency. Pharmacologic blockade of CXCR3 is a potential novel approach for NASH treatment.

## Role of Interferon $\gamma$ -Inducible Protein 10 in the Pathogenesis of Non-Alcoholic Steatohepatitis

Xiang Zhang, Eagle SH Chu, **Tung On Yau**, Xiaoxing Li, Joseph J. Y. Sung, Jun Yu

**Background & Aims:** Perpetuate liver inflammation is crucial in the pathogenesis of nonalcoholic steatohepatitis (NASH). Interferon  $\gamma$ -inducible protein 10 (IP-10) is a pro-inflammatory cytokine that correlates positively with the incidence of obesity and type 2 diabetes. However, the role of IP-10 in NASH remains elusive. We aimed to elucidate the functional significance of IP-10 in the development of NASH and to investigate the potential effect of IP-10 inhibitor in suppressing NASH development both in vitro and in vivo. **Methods:** IP-10 knockout (KO) and wildtype (WT) littermates were fed with control diet or methionine and choline-deficient (MCD) diet to induce NASH for 4 weeks. Neutralizing anti-IP-10 mAb or control mAb was injected into WT mice fed with MCD diet for 10 days. Mouse AML12 hepatocytes were cultured in MCD medium in the presence of anti-IP-10 or control mAb for 24 hours. **Results:** We first compared the IP-10 expression in 11 human NASH biopsies and 15 normal human liver tissues, and found that IP-10 mRNA expression was significantly higher in NASH than in the normal liver tissues ( $P < 0.001$ ). WT mice fed with MCD diet developed steatohepatitis and showed 5.8-fold higher IP-10 level than the control diet-fed WT mice with normal liver histology ( $P < 0.001$ ). Compared with WT littermates, IP-10 KO mice were more resistant to MCD-induced steatohepatitis as evidenced by the reduced level of hepatic inflammation ( $P < 0.01$ ), steatosis ( $P < 0.01$ ), triglyceride ( $P < 0.01$ ), lipidperoxide ( $P < 0.01$ ) and serum ALT ( $P < 0.001$ ). We revealed that the effect of IP-10 in promoting steatohepatitis was associated with the induction of pro-inflammatory cytokines TNF- $\alpha$ , IL6, IL-1 $\beta$ , monocyte chemoattractant protein 1 (MCP-1), macrophage inflammatory protein 1-alpha (MIP-1 $\alpha$ ) and MIP-2 $\alpha$ ; the upregulation of pro-inflammatory factors COX-2 and intercellular adhesion molecule 1 (ICAM); and the activation of nuclear factor- $\kappa$ B (NF- $\kappa$ B). IP-10 enhanced hepatic oxidative stress through upregulation of pro-oxidant cytochrome P450 2E1 (CYP2E1) and nuclear lipogenic CCAAT/enhancer binding protein  $\beta$  (C/EBP $\beta$ ). Moreover, the IP-10-induced steatosis was mediated by increased hepatic lipogenesis through inducing the expression of sterol regulatory element binding protein-1c (SREBP-1c), liver X receptors (LXRs) and stearoyl-CoA desaturase isoform-1 (SCD1). Furthermore, inhibition of IP-10 using a neutralizing anti-IP-10 antibody protected against the development of steatohepatitis in both AML-12 hepatocytes and mice. Molecular basis of the IP-10-induced NASH is shown in the figure. **Conclusions:** We demonstrate for the first time that IP-10 plays a pivotal role in the development of NASH by promoting inflammation, oxidative stress and fatty acid accumulation. IP-10 inhibition is a potential novel approach for the treatment of NASH.



## CXCL10 Induces Hepatocyte Apoptosis and Autophagy in Experimental Non-Alcoholic Steatohepatitis

Xiang Zhang, Xiaojuan Wang, Eagle SH Chu, **Tung On Yau**, Minnie Y. Go, Joseph J. Y. Sung, Jun Yu

**Background and Aims:** Aberrant expression of the pro-inflammatory chemokine CXCL10 is related to the severity of liver injury. Apoptosis and autophagy are two major forms of programmed cell death that regulates hepatocellular injury and may be crucial in nonalcoholic steatohepatitis (NASH) progression. Thus, we examined the association of CXCL10 with autophagy and apoptosis pathways in both in vitro and in vivo NASH models. **Methods:** Immortalized mouse hepatocytes (AML-12) were exposed to control or methionine and choline-deficient (MCD) medium in the presence of neutralizing anti-CXCL10 mAb or control mAb for 24 hours to investigate the role of CXCL10 in hepatocellular apoptosis and autophagy in vitro. CXCL10 knockout (KO) and C57BL/6 wildtype (WT) mice were fed with MCD diet for 4 weeks to induce NASH followed by systematic analysis of the CXCL10 effects on cell death in vivo. **Results:** AML-12 cells exposed to MCD medium developed significant steatosis with increased release of alanine transaminase (ALT) and oxidative injury. Analysis of AML-12 cells by Annexin V assay demonstrated that MCD medium induced a 2.9-fold increase of early apoptotic cells compared with control medium ( $P < 0.01$ ). However, CXCL10 inhibition using anti-CXCL10 mAb (1 $\mu$ g/mL) prevented MCD-induced apoptosis ( $P < 0.05$ ) in AML-12 hepatocytes, concomitant with the reduced mRNA expression of TNF $\alpha$ -related apoptosis-inducing ligand (TRAIL) receptors. Microtubule associated protein light chain 3 II (LC3-II), an autophagy marker, was also down-regulated by anti-CXCL10 mAb treatment in AML-12 cells cultured in MCD medium. The effects of CXCL10 on apoptosis and autophagy in NASH were further investigated in vivo. CXCL10 KO mice fed with MCD diet developed less severe NASH compared with WT littermates fed with the same diet. CXCL10 KO mice showed reduced cell apoptosis as indicated by less TUNEL-positive cells in the liver (0.94% versus 4.06%,  $P < 0.01$ ) and decreased protein expression of pro-apoptotic factors including TNF $\alpha$ , cleaved poly ADP-ribose polymerase and cleaved caspase-3. In parallel, MCD-fed CXCL10 KO mice showed decreased MCD-induced autophagy compared with WT mice fed with the same diet, as indicated by reduced expression of LC3-II and autophagy-related protein 7. Furthermore, autophagy and apoptosis modulators including toll-like receptor (TLR) 2 and TLR9 were ablated in the livers of CXCL10 KO mice. Concomitantly, JNK and p53, which are two apoptosis-regulatory factors involved in the modulation of autophagy, were observed to be down-regulated in the livers of MCD-fed CXCL10 KO mice compared with MCD-fed WT mice. **Conclusions:** CXCL10 is an

essential pro-apoptotic and pro-autophagic chemokine in NASH development through the regulation of JNK and p53. CXCL10 deletion inhibits hepatocyte apoptosis and autophagy induced by MCD both *in vitro* and *in vivo*.