Asian Journal of Psychiatry

RNA-sequencing of peripheral whole blood of individuals at ultra-high-risk for psychosis – a longitudinal perspective

--Manuscript Draft--

Manuscript Number:	ASJP-D-23-01667R1
Article Type:	Full Length Article
Keywords:	Schizophrenia; Ultra-high-risk; Psychosis; RNA-sequencing; Youth and Adolescent
Corresponding Author:	Wilson Wen Bin Goh, PhD Nanyang Technological University SINGAPORE
First Author:	Samuel Ming Xuan Tan, BSc
Order of Authors:	Samuel Ming Xuan Tan, BSc
	Jie Yin Yee, BSc
	Sugam Budhraja, BSc
	Balkaran Singh, BSc
	Zohreh Doborjeh, PhD
	Maryam Doborjeh, PhD
	Nikola Kasabov, PhD
	Edmund Lai, PhD
	Alexander Sumich, PhD
	Jimmy Lee, MD
	Wilson Wen Bin Goh, PhD
Abstract:	Background The peripheral blood is an attractive source of prognostic biomarkers for psychosis conversion. There is limited research on the transcriptomic changes associated with psychosis conversion in the peripheral whole blood. Study Design We performed RNA-sequencing of peripheral whole blood from 65 ultra-high-risk (UHR) participants and 70 healthy control participants recruited in the Longitudinal Youth-at-Risk Study (LYRIKS) cohort. 13 UHR participants converted in the study duration. Samples were collected at 3 timepoints, at 12-months interval across a 2-year period. We examined whether the genes differential with psychosis conversion contain schizophrenia risk loci. We then examined the functional ontologies and GWAS associations of the differential genes. We also identified the overlap between differentially expressed genes across different comparisons. Study results Genes containing schizophrenia risk loci were not differentially expressed in the peripheral whole blood in psychosis conversion. The differentially expressed genes in psychosis conversion are enriched for ontologies associated with cellular replication. The differentially expressed genes in psychosis conversion are associated with non-neurological GWAS phenotypes reported to be perturbed in schizophrenia and psychosis but not schizophrenia and psychosis phenotypes themselves. We found minimal overlap between the genes differential with psychosis conversion and the genes that are differential between pre-conversion and non-conversion samples. Conclusion The associations between psychosis conversion and peripheral blood-based biomarkers are likely to be indirect. Further studies to elucidate the mechanism behind potential indirect associations are needed.
Suggested Reviewers:	Valeria Mondelli Professor, King's College London valeria.mondelli@kcl.ac.uk

	Prof Mondelli specializes in psychoimmunology, which is highly relevant in this manuscript.
Opposed Reviewers:	
Response to Reviewers:	

Dear Editors,

Thank you for the opportunity to revise and resubmit our manuscript titled “RNA-sequencing of peripheral whole blood of individuals at ultra-high-risk for psychosis – a longitudinal perspective” by Tan et al. to be considered for publication as an Article in the Asian Journal of Psychiatry. Together with the revised manuscript, we have attached a document with our responses to the reviewer’s comments as well an annotated version of the revised manuscript to where all changes have been marked. We hope the annotated version will help to facilitate evaluations of the revised manuscript.

The novelty of our study has been shifted slightly to reflect a change in the framing of the study as requested by the reviewer. We will describe our study and the revised novelty in this cover letter:

This study is part of the Longitudinal Youth at-Risk Study (LYRIKS) which is a Singapore-based longitudinal study for psychosis conversion in adolescents and early adults designated as ultra-high risk of psychosis conversion (Lee et al, Schizophr Res, 2013). Whole blood samples were collected over 2 years at 12 months interval including baseline and when the participant converted to psychosis. We were able to obtain blood samples of participants before and after they converted to psychosis respectively as well as samples from participants who were diagnosed as high-risk but did not convert to psychosis, and matched controls over the study duration.

Particular novelty of study are: 1) At the time of writing and to the best of our knowledge, we are the first study to conduct a gene ontology enrichment analysis on the genes differential in psychosis conversion using a longitudinal cohort. 2) Due to cohort inclusion criteria and strict drug laws in Singapore (the country of origin for this study), our findings are minimally confounded by cannabis use. This is significant as cannabis use has been shown to confer additional risk to psychosis conversion and over 50% of the individuals who converted to psychosis in comparable studies reported a use of cannabis which (Chaumette et al, Schizophr Bull, 2019) and (Mongan et al, Jama Psych, 2020). 3) By using a linear mixed-effect model-based method in our analysis, we were able to account for intra-individual correlations inherent to longitudinal data. This additional correlation was not accounted for in existing similar studies (Chaumette et al, Schizophr Bull, 2019) and (Mongan et al, Jama Psych, 2020).

Using power analysis methods described in (Bi and Liu, BMC Bioinfo, 2016), we found that our study is sufficiently powered at 0.851. We performed RNA-sequencing of the whole blood samples. Key results include 1) The genes correlated with psychosis conversion are predominately immune related and while the genes are not associated with psychosis and schizophrenia phenotypes, they are associated many phenotypes reported to be perturbed in schizophrenia, including diseases and disorders such as “insomnia” and “coronary artery

disease“ and physiological parameters such as “fatty acid measurement” and “systolic blood pressure”. Based on these findings, we hypothesize that that the relation between peripheral blood-based biomarkers and psychosis conversion is likely to be indirect. 2) We showed that there is minimal overlap between genes that are correlated with psychosis conversion and genes that correlate with labels useful for identifying participants likely to convert (i.e. pre-conversion versus non-conversion). Results indicating this distinction is important for any future biomarker studies.

We believe that the findings of this paper will be of interest to the readers of your journal and valuable in informing future clinical studies into potential biomarkers for psychosis conversion.

We would like to thank the Editor and Reviewers for their advice and comments. It has been very helpful in improving our paper. Along with this letter and the revised manuscript, we also have included an annotated version of the revised manuscript where the locations of changes have been marked (strikethrough for deleted text and red for inserted text) together with comments justifying each change. Please find our responses to your comments and an overview of the changes made below.

Reviewer #1: This is an interesting study assessing RNA-sequencing in 65 UHR and 70 HC subjects. Though the findings are interesting, and authors appear to have a valuable set of data, some amendments need to be made.

Introduction is poorly written and should be reframed. It lacks a comprehensive review of the current literature in the topic of the proposed study. Also, some sentences are strangely inserted in the text, without a strong connection with the remaining of the text (e.g., the one about recreational drug use).

Response: Thank you for this suggestion. Key changes to the introduction include:

We described the current state of study in peripheral blood-based markers of psychosis conversion in UHR individuals, the limitations of existing studies, and the value of our study (page 4-6 in the annotated version). The sentence about recreational drug use has been modified to be more specific: "recreational drug" has been specified as "cannabis". The relevance of cannabis use has been more precisely defined: "over 50% of the individual who converted to psychosis in comparable studies Mongan et al. and Chaumette et al. reported a use of cannabis, representing a potential confounding factor."

Key points of the revised introduction include:

1. While there had been many studies that examined the transcriptomic changes associated with psychosis conversion, the majority are either cross-sectional or only involved a small set of candidate biomarkers.
2. Transition into psychosis is an evolving process and employing cross-sectional designs tend to introduce significant bias (Maxwell and Cole 2007)
3. Using a small set of candidate biomarkers precludes the discovery of novel candidate markers and the ability to perform gene set enrichment analysis.
4. We could only identify 2 studies that utilised high-throughput profiling on a longitudinal cohort study data, namely Chaumette et al. 2019 and Mongan et al. 2020.
5. Neither Chaumette et al nor Mongan et al examined the gene ontologies of the gene differentially expressed with psychosis conversion. The methods utilised in the 2 studies (DESEQ and SVM respectively) were not designed to model intra-individual correlations inherent in longitudinal data.
6. over 50% of the individual who converted to psychosis in Mongan et al. and Chaumette et al. reported a use of cannabis. As cannabis has been reported to confer additional risk to psychosis transition, it represents a confounding factor.
7. Our study provides the following value:

a. Utilised linear mixed-effect model-based method for identifying differentially expressed genes – thereby accounting for intra-individual correlations in the data.
b. We examined the gene ontology of the genes differentially expressed in psychosis conversion.
c. We examined whether the genes that are differential with psychosis conversion are necessarily differential between pre-conversion and non-conversion participants.
d. The drug free nature of our cohort allows for an analysis without confounding by cannabis and other illicit substances.
e. As an additional opportunity, we were also able to examine whether the genes containing loci likely to confer significant risk of schizophrenia identified in a recent Psychiatric Genomic Consortium meta-analysis (Trubetskoy et al. 2022) were differentially expressed in the peripheral blood with psychosis conversion.

These changes have been worked into the introduction/background. We hope these are satisfactory, and now provides an improved framing of our research contribution.

In Table 1, the different comparisons should be ordered by increased adjusted p-value.

Response: Thank you for this comment. The rows in Table 1 and Table 2 have been reordered by ascending adjusted p-value

The second paragraph of the Discussion section should describe with much more detail the findings of the study. Which genes were correlated with mitosis? What are the roles of the genes found? Authors should also better explain why this pattern is indicative of immune upregulation, and other assumptions like "hematological phenotypes" found in their sample.

Please provide rationale for the arguments hypothesized, so that the findings make physiological sense to the reader.

Response:

Thank you for this comment, this comment brought to our attention that the 2nd paragraph of the discussion was lacking in elaboration. We have made the following changes:

1. The sections regarding the GO enrichment analysis has been reworked to explain why we believe the enriched ontologies is indicative of cellular replication/mitosis and consequently an immune upregulation.
    a. GO terms were grouped by commonalities (e.g: mitotic spindle assembly, chromosomal segregation) in the results. Example: regulation of chromosome segregation (GO:0051983) and mitotic sister chromatid segregation (GO:0000070) were grouped into chromosomal segregation
    b. The various commonalities in the terms namely: mitotic spindle assembly and organization, chromosomal segregation, and DNA replication and cell cycle regulation are all processes of cell replication. This pattern is highly suggestive of the source cells undergoing cellular replication.
    c. Since the source of the RNA are leukocytes, a pattern of immune cell proliferation and thus immune upregulation.

Regarding the assumption of “haematological phenotypes”, the term was intended to broadly describe phenotypes typically expected to be associated with SNPs located within genes expressed by cells in the peripheral blood (e.g. mean corpuscular volume and basophile count). We now realised that using this term may appear presumptuous. Hence, the 3rd paragraph of the discussion has been reworked to avoid is over generalising the enriched phenotypes. The changes made to the 3rd paragraph is described in greater detail below in response to the comment on the aptness of the term “co-morbidities”.

English should be improved. E.g., the use of the word "terms" in line 239; line 82, "In total, 392 RNA libraries from the were collect from the…"

Response:

1. Thank you for the comment. Upon scrutinizing the text, we realised that in addition to the language, the term “RNA libraries” is not appropriate as at this stage, the RNA has not been extracted and the samples exist in venous whole blood. Line 82 has been rewritten to reflect this. It now reads as “In total, 392 venous whole blood samples were collected” (page 7 in annotated version). We greatly appreciate the opportunity to rectify this mistake.
2. The word “term” in line 239 has been replaced by the more appropriate “phenotype” (page 16 in annotated version).

We have also made language tweaks across the rest of the manuscript to improve readability and flow.

In the Discussion, use of term comorbidity might be inadequate, as authors did not investigate other illnesses in their sample. Nor were their findings related to comorbidities. Genes might be related to other physical entities but stating that they are related to comorbidities in schizophrenia is not quite right. As a matter of fact, immune response is possibly an integral component of disease pathophysiology, and not a comorbidity.

Response:

Thank you for this very astute comment. We understand that immune response is possibly an integral component of disease pathophysiology and not a comorbidity. Our original intention for the usage of the term “co-morbidity” was to group disorders/diseases such as insomnia and coronary artery diseases from the non-disorder/disease phenotypes such as systolic blood pressure.

Upon further consideration, we agree with the reviewer that the term “co-morbidities” is inappropriate. We now describe insomnia and coronary artery disease phenotypes as “disorders and diseases frequently observed in individuals with schizophrenia”.

The interpretation we wish to convey was that while the differentially expressed genes were not enriched in SNPs associated with schizophrenia and psychosis GWAS phenotypes, many of the enriched phenotypes were reported to be perturbed in individuals with schizophrenia and psychosis.

Based on this pattern, we hypothesized that the connection between the peripheral blood and psychosis conversion is likely to be indirect.

Upon scrutiny, we understood that the writing is unsatisfactory and may cause confusion. Furthermore, while implied by its absence in Table 3, the result that the differentially expressed genes were not enriched in SNPs associated with GWAS phenotypes was not explicitly stated.

We apologize for these oversights.

We have made the following amendments to our text:

1. We have included a segment to explicitly state that the differentially expressed genes were not enriched in SNPs associated with GWAS phenotypes relating to psychosis and schizophrenia (annotated version page 13). We have also added a new supplementary table listing the number of overlap and adjusted p-value of the various GWAS phenotypes associated with psychosis and schizophrenia (Supplementary Table S6).
2. We discussed the lack of enrichment for SNP associated in schizophrenia and psychosis in the discussion. This result is consistent with the recent meta-analysis by the Psychiatric Genomic Consortium (Trubetskoy et al. 2022) that revealed that most of the SNP conferring significant risk to schizophrenia were expressed neuronal cells in the CNS (page 16 in the annotated version)
3. Insomnia and coronary artery disease is now described as “disorders and diseases frequently observed in individuals with schizophrenia“ (page 17 in the annotated version)
4. We rephrased our original hypothesis
    a. Original: “neurological changes associated with psychosis conversion are accompanied by a range of non-neurological concomitant changes, many of which have been reported to be correlated with and or co-morbidities of schizophrenia”
    b. New: “We hypothesize that the association between psychosis conversion and biomarkers readily observable in the peripheral whole blood is likely to be indirect” (page 17 in annotated version)
5. We elaborated upon the implications of our hypothesis to include important future research questions.
    a. Original: “Consequently, using blood-based markers to detect co-morbidities and concomitant changes of psychosis and schizophrenia could represent a meaningful direction in predicting clinical outcomes in UHR individuals.”

    b. New: “Whether a differential expression of genes enriched for SNPs associated with these phenotypes necessarily indicate a perturbation of the phenotype, and how are these non-neurological phenotypes implicated in the pathophysiology of psychosis will be important research directions to evaluate this hypothesis. If positive evidence of both questions can be found, using blood-based markers to detect concomitant changes of psychosis and schizophrenia could represent a viable avenue for predicting clinical outcomes in UHR individuals.” (page 17 in annotated version)

Please refrain from using the term "schizophrenic" as it is stigmatizing (line 240).

Response:

Thank you for pointing this out. We did not mean to offend any groups. This was an oversight on our part. Consistent with the remainder of the text, the line has been amended to read “individual with schizophrenia” (line 326 in the annotated version)

- Peripheral whole blood samples were collected from UHR individuals at 12 months intervals.
- Profiling by RNA-sequencing indicates that psychosis conversion in UHR individuals is associated with immune dysregulation.
- Genes differential in psychosis conversion are not associated with schizophrenia and psychosis phenotypes but are associated with many phenotypes reported to be perturbed in schizophrenia. We hypothesize an indirect connection between the peripheral blood and psychosis conversion.
- Uncovering mechanistic links between psychosis conversion and the perturbed phenotypes along with the connection between the perturbed phenotypes and the genes differentially expressed with psychosis conversion will be important research directions.
- Minimal overlap between gene differential for psychosis conversion and genes differential between pre-conversion sample from non-conversion samples.

Funding and Acknowledgement

This research/project is supported by the National Research Foundation, Singapore under its Industry Alignment Fund – Prepositioning (IAF-PP) Funding Initiative (SDSC-2020-005). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore. National Research Foundation, Singapore was not involved in the study design; data collection, analysis and interpretation of data; the writing of the report; and in the decision to submit the article.

The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg).

Figure 1

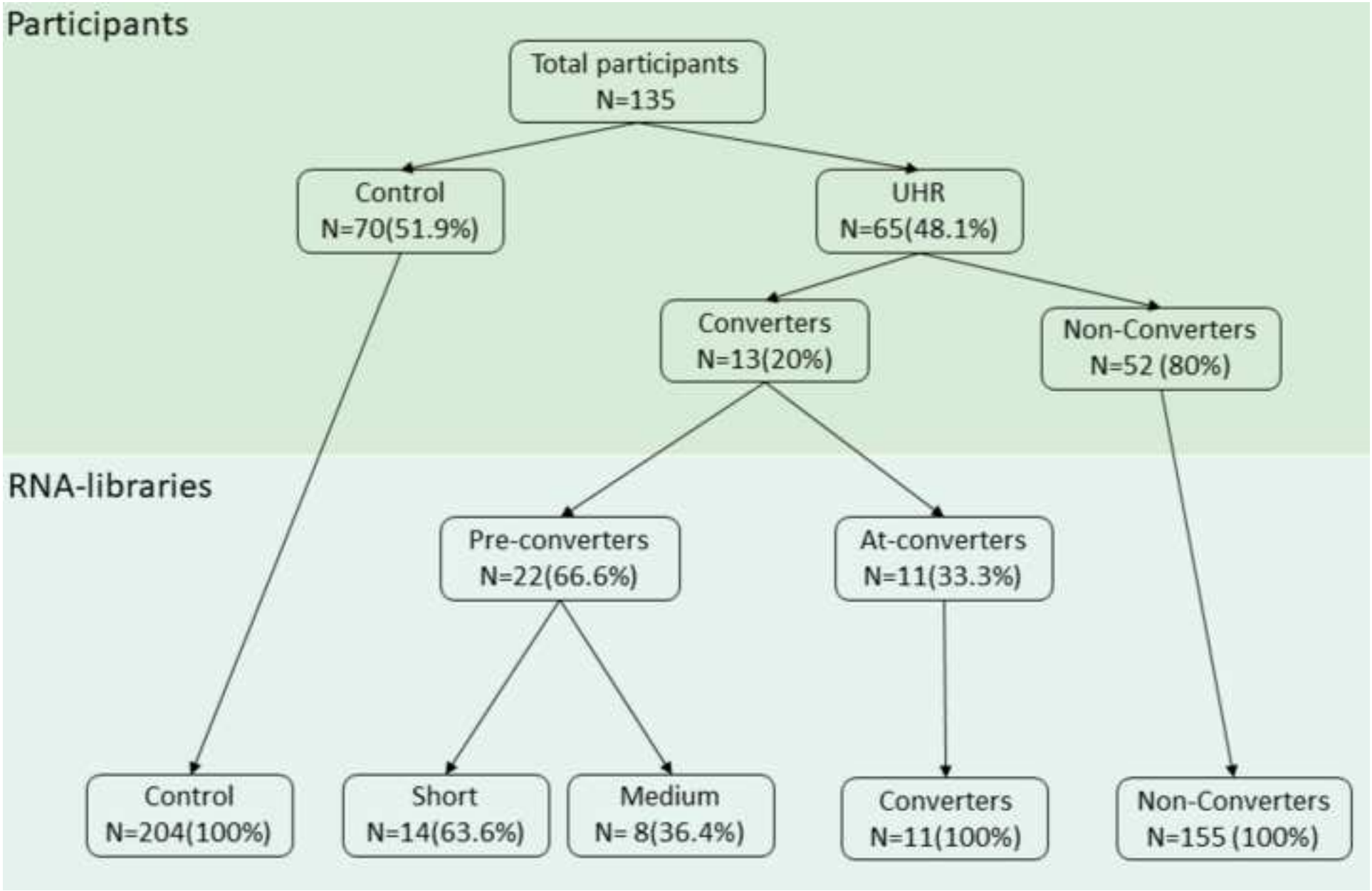

Table 3: Enriched GWAS phenotypes in SHORT vs CVT and MEDIUM vs CVT

	GWAS Phenotype	SHORT vs CVT adj.p-val	MEDIUM vs CVT adj.p-val
Both SHORT vs CVT and MEDIUM vs CVT	body height	2.65E-71	4.90E-07
	systolic blood pressure	0.004808	2.74E-06
SHORT vs CVT only	acute myeloid leukemia	0	-
	neuroimaging measurement	2.31E-75	-
	insomnia	2.28E-45	-
	educational attainment	5.67E-25	-
	vaginal microbiome measurement	7.79E-16	-
	smoking initiation	6.80E-12	-
	rheumatic heart disease	3.92E-10	-
	lipid measurement	9.13E-07	-
	phospholipid measurement	1.95E-05	-
	response to tenofovir	4.47E-05	-
	mean corpuscular hemoglobin	0.000135	-
	brain measurement	0.000139	-
	gut microbiome measurement	0.000331	-
	basophil count	0.000412	-
	total cholesterol measurement	0.000431	-
	brain volume measurement	0.001301	-
	mean corpuscular volume	0.003454	-
	erythrocyte count	0.004056	-
	attention deficit hyperactivity disorder	0.005285	-
	coronary artery disease	0.007824	-
	mean reticulocyte volume	0.009638	-
	high density lipoprotein cholesterol measurement	0.028402	-
	basophil percentage of leukocytes	0.029636	-
MEDIUM vs CVT only	triglyceride measurement	-	2.03E-07
	body mass index	-	5.64E-07

Table 2: Top 10 enriched gene ontologies in SHORT vs CVT and MEDIUM vs CVT

	Term	Overlap	Adjusted P-value	Combined Score	Genes
SHORT vs CVT	mitotic spindle organization (GO:0007052)	10/157	4.47E-05	187.6127	CDC20;TPX2;CENPF;KIF4A;BUB1B;BIRC5;KIF2C;BUB1;SKA1;DLGAP5
	microtubule cytoskeleton organization involved in mitosis (GO:1902850)	9/128	4.47E-05	198.4685	CDC20;CENPF;KIF4A;BUB1B;BIRC5;KIF2C;BUB1;SKA1;DLGAP5
	regulation of mitotic cell cycle phase transition (GO:1901990)	9/188	5.52E-04	104.6555	CDC20;TPX2;ANLN;CENPF;CDK1;KIF14;BUB1B;HMMR;CDC6
	mitotic sister chromatid segregation (GO:0000070)	7/102	5.52E-04	149.2213	KIF18B;CDK1;KIF14;NCAPG;KIF2C;CEP55;DLGAP5
	regulation of chromosome segregation (GO:0051983)	4/18	5.52E-04	562.6262	KIF2C;CDC6;MKI67;BUB1
	hepatocyte differentiation (GO:0070365)	3/6	5.52E-04	1938.618	PROX1;E2F7;E2F8
	DNA replication (GO:0006260)	7/108	5.70E-04	136.0071	EXO1;PCLAF;CDK1;MCM10;CLSPN;CDC6;TICRR
	spindle assembly checkpoint signaling (GO:0071173)	4/21	5.78E-04	438.5553	CDC20;CENPF;BUB1B;BUB1
	mitotic spindle assembly checkpoint signaling (GO:0007094)	4/21	5.78E-04	438.5553	CDC20;CENPF;BUB1B;BUB1
	mitotic spindle checkpoint signaling (GO:0071174)	4/21	5.78E-04	438.5553	CDC20;CENPF;BUB1B;BUB1
MEDIUM vs CVT	monovalent inorganic anion homeostasis (GO:0055083)	1/10	0.045367	1307.487	CA12
	mitotic DNA replication checkpoint signaling (GO:0033314)	1/10	0.045367	1307.487	ORC1
	chloride ion homeostasis (GO:0055064)	1/10	0.045367	1307.487	CA12
	mitotic nuclear membrane disassembly (GO:0007077)	1/12	0.045367	1032.951	CCNB2
	mitotic DNA integrity checkpoint signaling (GO:0044774)	1/13	0.045367	932.0597	ORC1
	nuclear membrane disassembly (GO:0051081)	1/14	0.045367	847.7047	CCNB2

	DNA replication checkpoint signaling (GO:0000076)	1/17	0.047186	661.819	ORC1
	positive regulation of smoothened signaling pathway (GO:0045880)	1/21	0.050957	506.0053	TEDC2
	regulation of transcription involved in G1/S transition of mitotic cell cycle (GO:0000083)	1/32	0.061207	296.3194	ORC1
	bicarbonate transport (GO:0015701)	1/35	0.061207	264.3328	CA12

Table 1: Top 10 differentially Expressed Genes Across Different Comparisons

	Gene symbol	log2FC	Avg Expr	adj.P.Val	Log-odds DE
SHORT vs CVT	IGKV1-6	1.749106	2.151455	0.00223	6.270656
	PCOLCE2	2.446494	-3.358	0.004089	4.545012
	CYP51A1P1	1.8136	-2.89244	0.011334	0.954209
	ITGA8	1.922005	-3.78934	0.017899	-0.271
	AC005220.1	-1.80242	-4.31701	0.022967	-0.7087
	EEF1A1P30	-1.91238	-3.86173	0.024979	-0.92225
	AP003419.3	-2.01198	-3.92194	0.032333	-1.41701
	AC253536.3	1.709154	-2.73394	0.032544	-1.46671
	AC079331.2	-1.92406	-3.66073	0.034124	-1.51914
	AC136618.1	-1.94269	-3.90204	0.034201	-1.51579
SHORT vs non-CVT	AC106872.3	1.063223	-3.23106	0.005703	4.106383
	UBE3AP2	-1.10486	-1.9454	0.019576	1.907311
	FAM27E4	1.414184	-4.00346	0.02079	1.97393
	ENSG00000286996	-1.1608	-2.15429	0.023742	1.057269
	CYP51A1P1	-1.50222	-2.89244	0.026721	0.658721
	ENSG00000286065	-1.04637	-2.17584	0.048531	-0.92696
MEDIUM vs CVT	AC145423.3	-1.55434	-2.22482	0.02214	3.62668
	LINC02289	-1.19849	2.632654	0.029279	3.027541
	IGKV2D-24	2.204792	-0.45198	0.034103	2.084801
	TEDC2	1.551655	-1.87729	0.034103	2.405233
	AC027607.1	-1.21044	-0.97772	0.034103	2.478262
	ORC1	1.1773	0.716462	0.034103	2.660014
	ENSG00000285966	-1.48894	-0.74495	0.042874	1.717976
	ENSG00000285534	-1.58104	-1.40456	0.042963	1.531203
	CCNB2	1.436071	0.766532	0.042963	1.507651
	CA12	-1.95492	-3.34769	0.048307	1.306614
MEDIUM vs non-CVT	CA12	2.050455	-3.34769	2.25E-03	5.587628
	LINC01629	2.100376	-3.24698	5.55E-03	4.679412
	AC145423.3	1.351728	-2.22482	5.77E-03	3.728576
	DAAM2	2.204425	0.572607	1.63E-02	2.135852
	ADGRB3	1.788875	-3.28153	1.80E-02	1.908368
	ATP2C2-AS1	1.427606	-3.21542	1.80E-02	2.111974
	TMEM119	1.419754	-1.03113	1.98E-02	1.569783
	AC020703.1	1.3693	-3.22812	2.20E-02	1.530564
	MAOA	2.38041	-3.00218	3.23E-02	0.886514
	ENSG00000286996	-1.39376	-2.15429	4.99E-02	-0.09548

Financial Disclosure

The authors declare no financial conflict of interest.

Conflict of Interest

The authors declare no conflict of interest.

Abstract

Background

The peripheral blood is an attractive source of prognostic biomarkers for psychosis conversion. There is limited research on the transcriptomic changes associated with psychosis conversion in the peripheral whole blood.

Study Design

We performed RNA-sequencing of peripheral whole blood from 65 ultra-high-risk (UHR) participants and 70 healthy control participants recruited in the Longitudinal Youth-at-Risk Study (LYRIKS) cohort. 13 UHR participants converted in the study duration. Samples were collected at 3 timepoints, at 12-months interval across a 2-year period. We examined whether the genes differential with psychosis conversion contain schizophrenia risk loci. We then examined the functional ontologies and GWAS associations of the differential genes. We also identified the overlap between differentially expressed genes across different comparisons.

Study results

Genes containing schizophrenia risk loci were not differentially expressed in the peripheral whole blood in psychosis conversion. The differentially expressed genes in psychosis conversion are enriched for ontologies associated with cellular replication. The differentially expressed genes in psychosis conversion are associated with non-neurological GWAS phenotypes reported to be perturbed in schizophrenia and psychosis but not schizophrenia and psychosis phenotypes themselves. We found minimal overlap between the genes differential

with psychosis conversion and the genes that are differential between pre-conversion and non-conversion samples.

Conclusion

The associations between psychosis conversion and peripheral blood-based biomarkers are likely to be indirect. Further studies to elucidate the mechanism behind potential indirect associations are needed.

Introduction

The ultra-high-risk (UHR) criteria was introduced to identify individuals likely in the prodromal phase of psychosis (A. R. Yung et al. 1996). Accurate and early identification is essential for targeted interventions to improve clinical outcomes (Fusar-Poli, McGorry, and Kane 2017). Psychometric assessments such as the Comprehensive Assessment for At-Risk Mental States (CAARMS)(Alison R. Yung et al. 2005) and the Structured Interview for Prodromal Syndromes (SIPS)(Miller et al. 2003) have been shown to have high sensitivity but poor specificity(Oliver et al. 2022). There is thus motivation to supplement the UHR designation with other measurements.

The peripheral blood is an attractive source for biomarker discovery (Liew et al. 2006). Many studies have been conducted in attempts to identify blood-based biomarkers of psychosis conversion (Ota et al. 2019; Demars et al. 2020; Chaumette et al. 2019; Mongan et al. 2020; English et al. 2018; Perkins et al. 2015; Laskaris et al. 2019). However, the majority of existing studies are either cross-sectional (Ota et al. 2019; Laskaris et al. 2019) or only assessed the expression changes of a selected list of candidate genes (English et al. 2018; Perkins et al. 2015;

Demars et al. 2020). Transition into psychosis is an evolving process and employing cross-sectional designs tend to introduce significant bias (Maxwell and Cole 2007). Additionally, assessing a relatively small set of candidate genes precludes the possibility of discovering novel biomarkers and the characterization of transcriptomic changes using methods such as Gene Set Enrichment Analysis (GSEA) (Subramanian et al. 2005). To date, only 2 studies profiled the transcriptomic changes of psychosis conversion in a longitudinal study cohort using high-throughput methods (Chaumette et al. 2019; Mongan et al. 2020).

Existing studies are limited in a few aspects: First, at the time of writing, no studies have examined the gene ontologies of the genes differentially expressed in psychosis conversion. Second, it is unknown if the genes differentially expressed in psychosis conversion are necessarily differential between pre-conversion and non-conversion participants – a property that reflects the suitability and reliability of the genes as prognostic biomarker. Third, the intra-individual correlation inherent to longitudinal data were not accounted for in existing studies (Chaumette et al. 2019; Mongan et al. 2020). Fourth, over 50% of converters in existing studies reported a history of cannabis use (Chaumette et al. 2019; Mongan et al. 2020) – this is a potential confounding factor as cannabis use has been strongly evidenced to confers additional risk in transition to psychosis (Gage, Hickman, and Zammit 2016).

We performed RNA-sequencing on whole blood samples from a subsample of the LYRIKS cohort (Lee et al. 2013). Peripheral whole blood samples were collected from the LYRIKS cohort at 12 months interval and individuals with a history of substance use (including cannabis) were excluded from the study (Lee et al. 2013).

Using a linear-mixed effect model-based method (Hoffman and Roussos 2021), we examined the differential expression of gene in psychosis conversion and the gene ontology enrichment of the differentially expressed genes. The linear-mixed effect model based method allows us to model the intra-individual correlations within the data (Diggle 2013). Along with the gene ontology enrichment analysis, the differential expression analysis also provides us an opportunity to examine whether a set of 120 genes containing high-risk loci identified by a recent large scale meta-analysis are differentially expressed with psychosis conversion in the peripheral blood (Trubetskoy et al. 2022). We then examined whether the genes that are differential with psychosis conversion are necessarily differential between pre-conversion and non-conversion participants. The drug-free nature of our cohort also allows us to provide a perspective into the transcriptomic changes of psychosis conversion with minimal confounding by cannabis and other illicit substances use.

This study aims to expand upon the existing understanding of blood-based biomarkers of psychosis conversion and provide directions for future studies in this domain.

Methods

Participants

The Longitudinal Youth-at-Risk study (LYRIKS) was a prospective, observational study conducted in Singapore on youths aged 14 to 29 to assess risk factors for psychosis (Lee et al., 2013). Briefly, participants were followed up over 2 years and were recruited from psychiatric clinics, various community agencies including educational institutes and social services, or were

self-referred. Participants were excluded if they (i) had a past or current history of psychosis or intellectual disability, (ii) were currently using illicit substances, (iii) were taking mood stabilizers, (iv) had previous antipsychotic exposure of more than 5 mg haloperidol per day for 3 weeks (or equivalent) or were on an antipsychotic at the point of recruitment, or (iv) had medical causes associated with their attenuated psychotic symptoms.

Ethics approval for this study was provided by the National Healthcare Group's Domain Specific Review Board (approval number 2009/00167). After complete description of the study to the participants, written informed consent was obtained. The present study utilised whole blood samples from 135 participants (70 control (CTRL) and 65 UHR) from the original LYRIKS cohort (Lee et al. 2013). All samples from participants who converted were included. Only samples from participants with at least 2 follow-ups were included for control and participants who did not convert. UHR designation and psychosis threshold were assessed using CAARMS. Venous whole blood samples were collected from the participants at 12-month intervals. Participants who converted at any point in the study were removed from the study following a final sample collection. Of the 65 UHR participants, 13 converted to psychosis (CVT) and 52 did not (non-CVT). In total, 392 venous whole blood samples were collected (**Supplementary Table S1**).

Demographic

The control group (CTRL) consist of 25 [35.7%] females with mean [SD] age of 21.1 [3.07] years and 45 [64.3%] males with mean [SD] age of 22.8 [3.51] years. The UHR group consist of 46 [70.8%] males with mean [SD] age of 21.4 [3.08] years and 19 [29.2%] females with mean

[SD] age of 21.9 [3.05] years. Of the 65 UHR participants, 52 participants did not develop psychosis over the 2-year study period (non-CVT). Of the 52 non-CVT participants, 17 [32.7%] are female with a mean [SD] age of 22.2 [3.3] years and 35 [67.3%] are male with mean [SD] age of 21.4 [3.08] years. The remaining 13 [20%] UHR participants with mean [SD] age of 20.3 [3.13] years developed psychosis over the study period (CVT); 2 [15.4%] are females. The median time to conversion from baseline was 11 months [IQR 19 - 5 = 14].

No statistically significant difference in sex proportion was found between CVT participants and CTRL participants (Chi-squared = 2.06, p = 0.205) or with non-CVT participants (Chi-squared = 3.06; P= 0.080). No statistically significant difference in age was found between CVT participants and CTRL participants (t = 1.05, 95% CI, [-1.05, 3.11]; P = 0.310) or non-CVT participants (t = -0.48142; 95% CI, [-2.58, 1.62]; P = 0.636).

Sequencing and pre-processing

A sample of venous whole blood was collected from all participants into Tempus™ Blood RNA tube (Thermo Fisher Scientific) and stored at −80◦C. Total RNA was extracted with the Tempus™ Spin RNA Isolation Kit (Thermo Fisher Scientific) according to manufacturer's protocol. Quality and concentration of extracted RNA was measured with 2100 Bioanalyzer System (Agilent). The libraries were prepped and sequenced using 8 Illumina NovaSeq (2x151bp) sequencing lanes. Sequencing quality was assessed using FastQC v0.11.9 (Schmieder and Edwards 2011). Sequences are aligned to Ensembl GRCh38 using STAR 2.7.9a (Dobin et al. 2013). Aligned reads are quantified using featureCount from the Subread 2.0.3 R Package (Liao, Smyth, and Shi 2019). Genes with zero variance and mean count < 1 were discarded yielding a

total of 21810 genes. All processing and analysis were performed in the R Programming language Version 4.0.3 (R Core Team 2013).

Power analysis

We performed power analysis using ssizeRNA R package (Bi and Liu 2016). We defined our criteria at absolute log2 fold-change > 1 and false discover rate = 0.05. Following the low number of differentially expressed genes in similar studies (Mongan et al. 2020; Chaumette et al. 2019), we set our expected proportion of differentially expressed genes at 0.005. We found that a sample size of 11 is sufficient to achieve a statistical power of 0.81. At 13 converter samples, our study has a power of 0.851.

Differential gene analysis

We divided the samples into 5 groups (**Figure 1**):

(1) Samples from participants who did not develop psychosis (non-CVT)
(2) Samples from participants post-conversion (CVT)
(3) Samples collected less than, or equal to 12 months before conversion (SHORT)
(4) Samples collected more than 12 months before conversion (MEDIUM)
(5) Samples from participants healthy controls (CTRL)

Samples in SHORT and MEDIUM will be referred to as pre-conversion samples (pre-CVT). *Sex* and *time* were included as model covariates where *time* is the month of sample collection. Longitudinal data involves repeated measurements from the same participant across time

which introduces intra-individual correlation. Repeated measurements result in intra-individual correlations that needs to be accounted for in the model. One method capable of modelling such correlation is the linear mixed effect model (Diggle 2013). We modelled known co-variates in the data via a linear mixed effect model using the DREAM model (Hoffman and Roussos 2021) from the voom-limma pipeline (Law et al. 2014). We modelled the slope of each sample across time and sample intercept as random effects:

$$y = group + sex + time + (time|sample) + (0|sample)$$

p-value adjustment (adjusted p-value) for multiple testing was addressed via Benjamini-Hochberg correction (Benjamini and Hochberg 1995). Threshold for differential expression was defined at an absolute log fold-change of > 1 and an adjusted p-value of < 0.05. List of contrasts used can be found in **Supplementary Notes.**

<Insert Figure 1>

Functional profiling via Gene Ontology analysis

Enrichment analysis for Gene Ontology (GO) functional terms was performed using Enrichr (Chen et al. 2013) webtool (which employs a Fisher's exact test). We examined the top 10 enriched ontologies ranked by adjusted p-value for analysis and interpretation.

Phenotype association analysis

GWAS phenotype associations were queried from the Genecards database (Fishilevich et al. 2017; Stelzer et al. 2016). GWAS phenotypes with less than 5 occurrences amongst the

differential genes were dropped. We perform Fisher's exact tests with Bonferroni correction to determine if the differential genes are enriched for SNP associated with a particular phenotype found in the European Bioinformatics Institute (EBI) GWAS catalog (Buniello et al. 2019). The same EBI catalogue is used in Genecards (Stelzer et al. 2016). Further details on data processing procedure are detailed in **Supplementary Notes**.

Results

Differential gene expression analysis

We identified 12, and 130 genes differentially expressed in MEDIUM vs CVT, and SHORT vs CVT respectively. The 10 differential genes with largest fold changes are listed in **Table 1,** the full list of differential genes can be found in **Supplementary Table S2-S3.** No genes were found to be differential in MEDIUM vs SHORT. We also identified 6, and 22 genes differentially expressed in MEDIUM vs non-CVT, and SHORT vs non-CVT respectively (**Table 1**) (full list in **Supplementary Table S4-S5**). No differential genes were identified when comparing CTRL vs CVT. We then examined whether the genes reported to contain high-risk loci in a recent large-scale meta-analysis were differentially expressed in our analysis (Trubetskoy et al. 2022). None of the genes associated with those loci were differentially expressed in our analyses.

< Insert Table 1 >

Gene ontology enrichment analysis

We performed enrichment analysis of the differential genes found in SHORT vs CVT and MEDIUM vs CVT to identify enriched gene ontologies. The enriched ontologies in SHORT vs CVT contain commonalities namely: mitotic spindle assembly and organization (GO:0007052; GO:0071173; GO:0007094; GO:0071174), chromosomal segregation (GO:0051983; GO:0000070), and DNA replication and cell cycle regulation (GO:1901990; GO:0006260) **(Table 2)**. Commonalities in the enriched ontologies in MEDIUM vs CVT include: cell cycle checkpoint (GO:0044774; GO:0033314; GO:0000076; GO:0000083), and nuclear membrane disassembly (GO:0007077; GO:0051081) **(Table 2)**. However, the overlap is minimal between the GO gene sets and the genes differential in MEDIUM vs CVT, likely due to the small number of differential genes in this comparison.

< Insert Table 2 >

GWAS phenotypes enrichment analysis

We examined the GWAS phenotypes associated with SNPs within the differentially expressed genes for SHORT vs CVT and MEDIUM vs CVT. We found 25 and 4 GWAS phenotypes that are enriched amongst the GWAS phenotypes of genes differential in SHORT vs CVT and MEDIUM vs CVT respectively **(Table 3)**. The differentially expressed genes were not enriched for any schizophrenia and psychosis GWAS phenotypes **(Supplementary Table S6)**.

< Insert Table 3 >

Comparison between genes differential in psychosis conversion and genes differential in pre-conversion and non-conversion

We examined whether the genes differential in psychosis conversion are also differential between pre-conversion and non-conversion participants. We identified the intersection (denoted by the symbol ∩) between the set of genes differential in psychosis conversion and the set of genes differential between pre-conversion and non-conversion. We found 2 genes in SHORT vs CVT ∩ SHORT vs non-CVT and 5 gene in both MEDIUM vs CVT ∩ MEDIUM vs non-CVT. CYP51A1P1 and UBE3AP2 were differential in both SHORT vs CVT and SHORT vs non-CVT. CA12, COL9A2 and pseudogenes ENSG00000285966, ENSG00000248936, ENSG00000278112 were differential in both MEDIUM vs CVT and MEDIUM vs non-CVT.

Discussion

Our study possesses several properties that allow for unique perspectives into the nuances of peripheral blood-based transcriptomic changes in UHR and psychosis conversion.

First, at the time of writing, we are the first to conduct a gene ontology enrichment analysis on the genes differential in psychosis conversion using a longitudinal cohort. Second, we employed a mixed-effect model to account for the intra-individual correlation within a longitudinal data – a factor unaccounted for in previous studies (Chaumette et al. 2019; Mongan et al. 2020). Third, owing to the recruitment criteria, our study cohort is minimally confounded by substance use. Fourth, at the time of writing, this study is the first to examine whether the gene differential in psychosis conversion is also differential between pre-conversion and non-conversion participants.

The genes identified in recent whole exome (Singh et al. 2022) and large scale GWAS (Trubetskoy et al. 2022) analyses to confer high risk of schizophrenia were not differentially expressed in our analyses. This is consistent with the finding that the genes were mostly expressed in synaptic junctions rather than the peripheral blood (Trubetskoy et al. 2022). This finding highlights the difficulty of employing schizophrenia genetic risk loci as a basis of identifying biomarkers in the peripheral blood.

The commonalities in the enriched gene ontologies include 1) mitotic spindle assembly and organization, 2) chromosomal segregation, 3) DNA replication and cell cycle regulation cell cycle checkpoint, and 4) nuclear membrane disassembly – all processes of cellular replication. This pattern suggests that the source cells are proliferating. Considering that nucleated cells in

241 the whole blood are generally leukocytes, the source cells are predominately leukocytes and
242 this observation is consistent with a pattern of immune upregulation widely observed in
243 schizophrenia and psychosis conversion (Gardiner et al. 2013; Tomasik et al. 2016; Xu et al.
244 2012). Disentangling biomarkers specific to psychosis conversion from markers of general
245 immune responses will be a crucial step in the identification of peripheral blood-based
246 biomarkers of psychosis.

247 Gene differential with psychosis conversion were not enriched with SNPs associated
248 with schizophrenia or psychosis phenotypes. This is consistent with a recent finding that found
249 that the genes containing the SNPs conferring substantial risk to schizophrenia were
250 predominately expressed by neuronal cells in the central nervous system (Trubetskoy et al.
251 2022). Several of the enriched phenotypes have been reported to be perturbed in individuals
252 with schizophrenia, these include cell population ratios such as neutrophil to lymphocyte ratio
253 (Karageorgiou, Milas, and Michopoulos 2019), physiological measurements such as fatty acid
254 measurement and systolic blood pressure (Andreassen et al. 2013), disorders and diseases
255 frequently observed in individuals with schizophrenia such as insomnia and coronary artery
256 disease (Palmese et al. 2011; Ifteni et al. 2014), and variations in the gut microbiome (Nguyen
257 et al. 2019; Zhu et al. 2020). We hypothesize that the association between psychosis conversion
258 and biomarkers readily observable in the peripheral whole blood is likely to be indirect.
259 Whether a differential expression of genes enriched for SNPs associated with these phenotypes
260 necessarily indicate a perturbation of the phenotype, and how are these non-neurological
261 phenotypes implicated in the pathophysiology of psychosis will be important research
262 directions to evaluate this hypothesis. If positive evidence of both questions can be found,

using blood-based markers to detect concomitant changes of psychosis and schizophrenia could represent a viable avenue for predicting clinical outcomes in UHR individuals.

Genes differential with psychosis conversion are not necessarily differential between pre-conversion and non-conversion individuals. Of the genes that exist in the overlaps, CYP51A1P1 and UBE3AP2 were differential in both SHORT vs CVT and SHORT vs non-CVT. CYP51A1P1 is a pseudogene for cytochrome P450 family 51 subfamily A member 1 (CYP51A1). The cytochrome P450 family proteins are predominantly involved in lipids and cholesterol metabolism (Stelzer et al. 2016). UBE3AP2 is a pseudogene for ubiquitin protein ligase 3A (UBE3A) and was reported to associate with hypertension (Stelzer et al. 2016). While certain rare variants in UBE3A has been proposed to confer risks to schizophrenia (Singh et al. 2022; Trubetskoy et al. 2022), it is unclear if this correlation might extend to pseudogenes of UBE3A and to psychosis conversion. Genes CA12, COL9A2 were differential in both MEDIUM vs CVT and MEDIUM vs non-CVT. CA12 is a carbonic anhydrase which catalyzes the reversible hydration of carbon dioxide. GWAS associated studies reported that CA12 is associated with a large range of phenotypes including schizophrenia symptom severity (Stelzer et al. 2016). COL9A2 codes for an alpha chain subunit in type IX collagen. Associations between COL9A2 and schizophrenia are indirect, via phenotypes such as cognitive measurements (Lam et al. 2018), testosterone measurements (Moore et al. 2013), and waist-hip-ratio (Andreassen et al. 2013). At the time of writing, the functions of ENSG00000285966, ENSG00000248936, and ENSG00000278112 are unknown and will not be discussed. Based on our study design, these genes and pseudogenes are theoretical candidate biomarkers for psychosis conversion. However, the lack of robust links to pathways linked to the pathology of schizophrenia

dampens our confidence in the ability of these genes and pseudogenes to serve as biomarkers of psychosis conversion.

Limitation

Our study contains several limitations. Firstly, the sample size of 392 used in this study is relatively small especially since samples are distributed into multiple time-points across various groups. Hence, the inferential statistical analyses performed in this study may be underpowered. However, we opine that a reliable (and useful) biomarker should be robust and detectable, even against the backdrop of small sample sizes. Secondly, we did not distinguish non-CVT participants who remitted from the UHR status from non-CVT participants who maintained UHR status to avoid further reducing sample sizes across groups. Thirdly, the etiology of psychosis conversion is inherently heterogenous (Zwicker, Denovan-Wright, and Uher 2018) and without a basis for pre-defining all medically relevant subtypes, our study design was unable to account for the heterogenous nature of psychosis conversion. Finally, the overlaps in our gene ontology enrichment analysis are small especially in MEDIUM vs CVT, we advise the readers to exercise discretion when considering the results.

Conclusion

The identification of peripheral blood-based biomarkers for psychosis is an area of great clinical value. Our study expanded upon existing studies by illustrating various important nuances in the identification of peripheral blood-based biomarkers of psychosis conversion and establishes directions for future investigations. We identified differentially expressed genes in

psychosis conversion, which were consistent with a pattern of immune upregulation. The SNPs in the genes differential in psychosis conversion were associated with a range of phenotypes reported to be perturbed in schizophrenia and psychosis but are not associated with schizophrenia and psychosis phenotypes themselves. We also showed that the genes correlated with psychosis conversion are not necessarily differential between pre-conversion and non-conversion individuals. The genes containing risk loci for schizophrenia were not differentially expressed in the peripheral whole blood in psychosis conversion. From these findings, we hypothesize that the associations between psychosis conversion and peripheral blood biomarkers are indirect. Mechanistic studies to examine such indirect associations will be a valuable research direction.

Data availability

The results of each contrast in the differential genes analysis can be found in the Supplementary section. Supplementary information is available at the journal website. Raw transcripts counts are not available due to patient confidentiality requirements. Analysis code is available on request from the authors.

Author contribution

S.M.X.T., Y.J.Y., J.L., and W.W.B.G conceptualized and designed the research; S.M.X.T., Y.J.Y., performed the research; S.M.X.T. performed that statistical analysis; S.M.X.T. and Y.J.Y. wrote the manuscript; Y.J.Y. and J.L. were responsible for the acquisition of the data; S.M.X.T., B.S., S.B., Z.D., M.D., N.K., E.L., A.S, Y.J.Y., J.L., and W.W.B.G interpretated the results; B.S., S.B., Z.D.,

M.D., N.K., E.L., A.S, Y.J.Y., J.L., and W.W.B.G reviewed the work. All authors approve of the submitted version of the manuscript.

Transparency declaration

We affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

References

Addington, Jean, Kristin S. Cadenhead, Barbara A. Cornblatt, Daniel H. Mathalon, Thomas H. McGlashan, Diana O. Perkins, Larry J. Seidman, et al. 2012. "North American Prodrome Longitudinal Study (NAPLS 2): Overview and Recruitment." *Schizophrenia Research* 142 (1–3): 77–82. https://doi.org/10.1016/j.schres.2012.09.012.

Andreassen, Ole A., Srdjan Djurovic, Wesley K. Thompson, Andrew J. Schork, Kenneth S. Kendler, Michael C. O'Donovan, Dan Rujescu, et al. 2013. "Improved Detection of Common Variants Associated with Schizophrenia by Leveraging Pleiotropy with Cardiovascular-Disease Risk Factors." *The American Journal of Human Genetics* 92 (2): 197–209. https://doi.org/10.1016/j.ajhg.2013.01.001.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bi, Ran, and Peng Liu. 2016. "Sample Size Calculation While Controlling False Discovery Rate for Differential Expression Analysis with RNA-Sequencing Experiments." *BMC Bioinformatics* 17 (1): 146. https://doi.org/10.1186/s12859-016-0994-9.

Buniello, Annalisa, Jacqueline A L MacArthur, Maria Cerezo, Laura W Harris, James Hayhurst, Cinzia Malangone, Aoife McMahon, et al. 2019. "The NHGRI-EBI GWAS Catalog of Published Genome-Wide Association Studies, Targeted Arrays and Summary Statistics 2019." *Nucleic Acids Research* 47 (D1): D1005–12. https://doi.org/10.1093/nar/gky1120.

Chaumette, Boris, Oussama Kebir, Juliette Pouch, Bertrand Ducos, Fekrije Selimi, ICAAR study group, Raphael Gaillard, and Marie-Odile Krebs. 2019. "Longitudinal Analyses of Blood Transcriptome During Conversion to Psychosis." *Schizophrenia Bulletin* 45 (1): 247–55. https://doi.org/10.1093/schbul/sby009.

Chen, Edward Y, Christopher M Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Vaz Meirelles, Neil R Clark, and Avi Ma'ayan. 2013. "Enrichr: Interactive and Collaborative HTML5 Gene List Enrichment Analysis Tool." *BMC Bioinformatics* 14 (1): 128. https://doi.org/10.1186/1471-2105-14-128.

Demars, Fanny, Oussama Kebir, Aude Marzo, Anton Iftimovici, Catherine Schramm, ICAAR Study Group, Isabelle Amado, et al. 2020. "Dysregulation of Peripheral Expression of the YWHA Genes during Conversion to Psychosis." *Scientific Reports* 10 (1): 9863. https://doi.org/10.1038/s41598-020-66901-1.

Diggle, Peter, ed. 2013. *Analysis of Longitudinal Data*. Second Paperback Edition. Oxford Statistical Science Series 25. Oxford: Oxford University Press.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics* 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

English, Jane A, Lorna M Lopez, Aoife O'Gorman, Melanie Föcking, Magdalena Hryniewiecka, Caitriona Scaife, Sophie Sabherwal, et al. 2018. "Blood-Based Protein Changes in Childhood Are Associated With Increased Risk for Later Psychotic Disorder: Evidence From a Nested Case–Control Study of the ALSPAC Longitudinal Birth Cohort." *Schizophrenia Bulletin* 44 (2): 297–306. https://doi.org/10.1093/schbul/sbx075.

Fishilevich, Simon, Ron Nudel, Noa Rappaport, Rotem Hadar, Inbar Plaschkes, Tsippi Iny Stein, Naomi Rosen, et al. 2017. “GeneHancer: Genome-Wide Integration of Enhancers and Target Genes in GeneCards.” *Database* 2017 (January). https://doi.org/10.1093/database/bax028.

Fusar-Poli, Paolo, Patrick D. McGorry, and John M. Kane. 2017. “Improving Outcomes of First-Episode Psychosis: An Overview.” *World Psychiatry* 16 (3): 251–65. https://doi.org/10.1002/wps.20446.

Gage, Suzanne H., Matthew Hickman, and Stanley Zammit. 2016. “Association Between Cannabis and Psychosis: Epidemiologic Evidence.” *Biological Psychiatry* 79 (7): 549–56. https://doi.org/10.1016/j.biopsych.2015.08.001.

Gardiner, Erin J., Murray J. Cairns, Bing Liu, Natalie J. Beveridge, Vaughan Carr, Brian Kelly, Rodney J. Scott, and Paul A. Tooney. 2013. “Gene Expression Analysis Reveals Schizophrenia-Associated Dysregulation of Immune Pathways in Peripheral Blood Mononuclear Cells.” *Journal of Psychiatric Research* 47 (4): 425–37. https://doi.org/10.1016/j.jpsychires.2012.11.007.

Hoffman, Gabriel E, and Panos Roussos. 2021. “Dream: Powerful Differential Expression Analysis for Repeated Measures Designs.” Edited by Jan Gorodkin. *Bioinformatics* 37 (2): 192–201. https://doi.org/10.1093/bioinformatics/btaa687.

Ifteni, Petru, Christoph U. Correll, Victoria Burtea, John M. Kane, and Peter Manu. 2014. “Sudden Unexpected Death in Schizophrenia: Autopsy Findings in Psychiatric Inpatients.” *Schizophrenia Research* 155 (1–3): 72–76. https://doi.org/10.1016/j.schres.2014.03.011.

Karageorgiou, Vasilios, Gerasimos P. Milas, and Ioannis Michopoulos. 2019. “Neutrophil-to-Lymphocyte Ratio in Schizophrenia: A Systematic Review and Meta-Analysis.” *Schizophrenia Research* 206 (April): 4–12. https://doi.org/10.1016/j.schres.2018.12.017.

Lam, Max, Jimmy Lee, Attilio Rapisarda, Yuen Mei See, Zixu Yang, Sara-Ann Lee, Nur Amirah Abdul-Rashid, et al. 2018. “Longitudinal Cognitive Changes in Young Individuals at Ultrahigh Risk for Psychosis.” *JAMA Psychiatry* 75 (9): 929. https://doi.org/10.1001/jamapsychiatry.2018.1668.

Laskaris, Liliana, Andrew Zalesky, Cynthia Shannon Weickert, Maria A. Di Biase, Gursharan Chana, Bernhard T. Baune, Chad Bousman, et al. 2019. “Investigation of Peripheral Complement Factors across Stages of Psychosis.” *Schizophrenia Research* 204 (February): 30–37. https://doi.org/10.1016/j.schres.2018.11.035.

Law, Charity W, Yunshun Chen, Wei Shi, and Gordon K Smyth. 2014. “Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts.” *Genome Biology* 15 (2): R29. https://doi.org/10.1186/gb-2014-15-2-r29.

Lee, Jimmy, Gurpreet Rekhi, Natasha Mitter, Yioe Ling Bong, Michael S. Kraus, Max Lam, Attilio Rapisarda, et al. 2013. “The Longitudinal Youth at Risk Study (LYRIKS) — An Asian UHR Perspective.” *Schizophrenia Research* 151 (1–3): 279–83. https://doi.org/10.1016/j.schres.2013.09.025.

Liao, Yang, Gordon K Smyth, and Wei Shi. 2019. “The R Package Rsubread Is Easier, Faster, Cheaper and Better for Alignment and Quantification of RNA Sequencing Reads.” *Nucleic Acids Research* 47 (8): e47–e47. https://doi.org/10.1093/nar/gkz114.

Liew, Choong-Chin, Jun Ma, Hong-Chang Tang, Run Zheng, and Adam A. Dempsey. 2006. "The Peripheral Blood Transcriptome Dynamically Reflects System Wide Biology: A Potential Diagnostic Tool." *Journal of Laboratory and Clinical Medicine* 147 (3): 126–32. https://doi.org/10.1016/j.lab.2005.10.005.

Maxwell, Scott E., and David A. Cole. 2007. "Bias in Cross-Sectional Analyses of Longitudinal Mediation." *Psychological Methods* 12 (1): 23–44. https://doi.org/10.1037/1082-989X.12.1.23.

Miller, T. J., T. H. McGlashan, J. L. Rosen, K. Cadenhead, J. Ventura, W. McFarlane, D. O. Perkins, G. D. Pearlson, and S. W. Woods. 2003. "Prodromal Assessment With the Structured Interview for Prodromal Syndromes and the Scale of Prodromal Symptoms: Predictive Validity, Interrater Reliability, and Training to Reliability." *Schizophrenia Bulletin* 29 (4): 703–15. https://doi.org/10.1093/oxfordjournals.schbul.a007040.

Mongan, David, Melanie Föcking, Colm Healy, Subash Raj Susai, Meike Heurich, Kieran Wynne, Barnaby Nelson, et al. 2020. "Development of Proteomic Prediction Models for Transition to Psychotic Disorder in the Clinical High-Risk State and Psychotic Experiences in Adolescence." *JAMA Psychiatry* 78 (1): 77–90. https://doi.org/10.1001/jamapsychiatry.2020.2459.

Moore, Loretta, Merribel Kyaw, Ans Vercammen, Rhoshel Lenroot, Jayashri Kulkarni, Jackie Curtis, Maryanne O'Donnell, Vaughan J Carr, C Shannon Weickert, and Thomas W Weickert. 2013. "Serum Testosterone Levels Are Related to Cognitive Function in Men with Schizophrenia." *Psychoneuroendocrinology* 38 (9): 1717–28.

Nguyen, Tanya T., Tomasz Kosciolek, Yadira Maldonado, Rebecca E. Daly, Averria Sirkin Martin, Daniel McDonald, Rob Knight, and Dilip V. Jeste. 2019. "Differences in Gut Microbiome Composition between Persons with Chronic Schizophrenia and Healthy Comparison Subjects." *Schizophrenia Research* 204 (February): 23–29. https://doi.org/10.1016/j.schres.2018.09.014.

Oliver, Dominic, Maite Arribas, Joaquim Radua, Gonzalo Salazar de Pablo, Andrea De Micheli, Giulia Spada, Martina Maria Mensi, et al. 2022. "Prognostic Accuracy and Clinical Utility of Psychometric Instruments for Individuals at Clinical High-Risk of Psychosis: A Systematic Review and Meta-Analysis." *Molecular Psychiatry*, June. https://doi.org/10.1038/s41380-022-01611-w.

Ota, Vanessa Kiyomi, Patricia Natalia Moretti, Marcos Leite Santoro, Fernanda Talarico, Leticia Maria Spindola, Gabriela Xavier, Carolina Muniz Carvalho, et al. 2019. "Gene Expression over the Course of Schizophrenia: From Clinical High-Risk for Psychosis to Chronic Stages." *Npj Schizophrenia* 5 (1): 5. https://doi.org/10.1038/s41537-019-0073-0.

Palmese, Laura B., Pamela C. DeGeorge, Joseph C. Ratliff, Vinod H. Srihari, Bruce E. Wexler, Andrew D. Krystal, and Cenk Tek. 2011. "Insomnia Is Frequent in Schizophrenia and Associated with Night Eating and Obesity." *Schizophrenia Research* 133 (1–3): 238–43. https://doi.org/10.1016/j.schres.2011.07.030.

Perkins, Diana O., Clark D. Jeffries, Jean Addington, Carrie E. Bearden, Kristin S. Cadenhead, Tyrone D. Cannon, Barbara A. Cornblatt, et al. 2015. "Towards a Psychosis Risk Blood Diagnostic for Persons Experiencing High-Risk Symptoms: Preliminary Results From the NAPLS Project." *Schizophrenia Bulletin* 41 (2): 419–28. https://doi.org/10.1093/schbul/sbu099.

Quattrone, Diego, Laura Ferraro, Giada Tripoli, Caterina La Cascia, Harriet Quigley, Andrea Quattrone, Hannah E. Jongsma, et al. 2021. “Daily Use of High-Potency Cannabis Is Associated with More Positive Symptoms in First-Episode Psychosis Patients: The EU-GEI Case–Control Study.” *Psychological Medicine* 51 (8): 1329–37. https://doi.org/10.1017/S0033291720000082.

R Core Team. 2013. “R: A Language and Environment for Statistical Computing.”

Schmieder, R., and R. Edwards. 2011. “Quality Control and Preprocessing of Metagenomic Datasets.” *Bioinformatics* 27 (6): 863–64. https://doi.org/10.1093/bioinformatics/btr026.

Singh, Tarjinder, Timothy Poterba, David Curtis, Huda Akil, Mariam Al Eissa, Jack D. Barchas, Nicholas Bass, et al. 2022. “Rare Coding Variants in Ten Genes Confer Substantial Risk for Schizophrenia.” *Nature* 604 (7906): 509–16. https://doi.org/10.1038/s41586-022-04556-w.

Stelzer, Gil, Naomi Rosen, Inbar Plaschkes, Shahar Zimmerman, Michal Twik, Simon Fishilevich, Tsippi Iny Stein, et al. 2016. “The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses.” *Current Protocols in Bioinformatics* 54 (1). https://doi.org/10.1002/cpbi.5.

Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, et al. 2005. “Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (43): 15545–50. https://doi.org/10.1073/pnas.0506580102.

Tomasik, Jakub, Hassan Rahmoune, Paul C. Guest, and Sabine Bahn. 2016. “Neuroimmune Biomarkers in Schizophrenia.” *Schizophrenia Research*. Elsevier B.V. https://doi.org/10.1016/j.schres.2014.07.025.

Trubetskoy, Vassily, Antonio F. Pardiñas, Ting Qi, Georgia Panagiotaropoulou, Swapnil Awasthi, Tim B. Bigdeli, Julien Bryois, et al. 2022. “Mapping Genomic Loci Implicates Genes and Synaptic Biology in Schizophrenia.” *Nature* 604 (7906): 502–8. https://doi.org/10.1038/s41586-022-04434-5.

United Nations Office on Drugs and Crime. 2006. “World Drug Report 2006.” United Nations, Office on Drugs and Crime. https://www.unodc.org/unodc/en/data-and-analysis/WDR-2006.html.

Xu, Junzhe, Jingchun Sun, Jingchun Chen, Lily Wang, Anna Li, Matthew Helm, Steven L Dubovsky, Silviu-Alin Bacanu, Zhongming Zhao, and Xiangning Chen. 2012. “RNA-Seq Analysis Implicates Dysregulation of the Immune System in Schizophrenia.” *BMC Genomics* 13 (8): 1–10. https://doi.org/10.1186/1471-2164-13-S8-S2.

Yung, A. R., P. D. McGorry, C. A. McFarlane, H. J. Jackson, G. C. Patton, and A. Rakkar. 1996. “Monitoring and Care of Young People at Incipient Risk of Psychosis.” *Schizophrenia Bulletin* 22 (2): 283–303. https://doi.org/10.1093/schbul/22.2.283.

Yung, Alison R., Alison R. Yung, Hok Pan Yuen, Patrick D. Mcgorry, Lisa J. Phillips, Daniel Kelly, Margaret Dell’olio, et al. 2005. “Mapping the Onset of Psychosis: The Comprehensive Assessment of At-Risk Mental States.” *Australian & New Zealand Journal of Psychiatry* 39 (11–12): 964–71. https://doi.org/10.1080/j.1440-1614.2005.01714.x.

Zhu, Feng, Yanmei Ju, Wei Wang, Qi Wang, Ruijin Guo, Qingyan Ma, Qiang Sun, et al. 2020. "Metagenome-Wide Association of Gut Microbiome Features for Schizophrenia." *Nature Communications* 11 (1): 1612. https://doi.org/10.1038/s41467-020-15457-9.
Zwicker, Alyson, Eileen M. Denovan-Wright, and Rudolf Uher. 2018. "Gene–Environment Interplay in the Etiology of Psychosis." *Psychological Medicine* 48 (12): 1925–36. https://doi.org/10.1017/S003329171700383X.

Abstract

Background

The peripheral blood is an attractive source of prognostic biomarkers for psychosis conversion. ~~It is also unclear whether genes correlated with psychosis conversion can reliably identify patients who will eventually experience psychosis and schizophrenia in the future.~~ There is limited research on the transcriptomic changes associated with psychosis conversion in the peripheral whole blood.

Study Design

We performed RNA-sequencing of peripheral whole blood from 65 ultra-high-risk (UHR) participants and 70 healthy control participants recruited in the Longitudinal Youth-at-Risk Study (LYRIKS) cohort. 13 UHR participants converted in the study duration. Samples were collected at 3 timepoints, at 12-months interval across a 2-year period. ~~We examined the functional ontologies and genetic associations of the differential genes correlated with psychosis conversion. We then compared the sets of differentially expressed genes across different comparisons.~~ We examined whether the genes differential with psychosis conversion contain schizophrenia risk loci. We then examined the functional ontologies and GWAS associations of the differential genes. We also identified the overlap between differentially expressed genes across different comparisons.

Study results

Commented [ST1]: This change has been made to better reflect the revised framing of the research; the lack of high-throughput longitudinal studies into the transcriptomic changes in psychosis conversion

Commented [ST2]: Changes made to reflect the addition of whether "genes differential with psychosis conversion contain schizophrenia risk loci"

~~The differentially expressed genes are enriched for ontologies associated with cellular replication. GWAS phenotypes associated with co-morbidities of schizophrenia such as "insomnia" and "coronary artery disease" as well as physiological parameters associated with schizophrenia such as "fatty acid measurements", "body height", "systolic blood pressure", and "gut microbiome measurements" were also observed.~~ Genes containing schizophrenia risk loci were not differentially expressed in the peripheral whole blood in psychosis conversion. The differentially expressed genes in psychosis conversion are enriched for ontologies associated with cellular replication. The differentially expressed genes in psychosis conversion are associated with non-neurological GWAS phenotypes reported to be perturbed in schizophrenia and psychosis but not schizophrenia and psychosis phenotypes themselves. We found minimal overlap between the genes differential with psychosis conversion and the genes that are differential between pre-conversion and non-conversion samples.

Commented [ST3]: This change reflects the changes made in the discussion section

Conclusion

~~Our findings are consistent with theories of immune dysregulation in psychosis. Using blood-based markers to detect co-morbidities and concomitant changes of psychosis might represent a viable direction in prognosis in UHR individuals. Our results also illustrate the distinction between genes differential with psychosis conversion with genes useful as biomarkers of psychosis conversion.~~

The associations between psychosis conversion and peripheral blood-based biomarkers are likely to be indirect. Further studies to elucidate the mechanism behind potential indirect associations are needed.

Introduction

The ultra-high-risk (UHR) criteria was introduced to ~~define patients in the prodrome of psychosis~~ identify individuals likely in the prodromal phase of psychosis (A. R. Yung et al. 1996). Accurate and early identification is essential for ~~potential~~ targeted interventions to improve clinical outcomes (Fusar-Poli, McGorry, and Kane 2017). Psychometric assessments such as the Comprehensive Assessment for At-Risk Mental States (CAARMS)(Alison R. Yung et al. 2005) and the Structured Interview for Prodromal Syndromes (SIPS)(Miller et al. 2003) have been shown to have high sensitivity but poor specificity(Oliver et al. 2022). There is thus motivation to supplement the UHR designation with other measurements.

~~Attempts have been made to identify candidate blood-based biomarkers of psychosis conversion in UHR using high-throughput methods such as high-throughput proteomic(Mongan et al. 2020) and RNA-sequencing(Chaumette et al. 2019). However, large proportions of participants in comparable cohorts studies such as North American Prodrome Longitudinal Study (NAPLS) and the EUropean network of national schizophrenia networks studying Gene-Environment Interactions (EU-GEI) reported recreational drug use which is highly correlated with positive symptom severity(Addington et al. 2012; Quattrone et al. 2021). Furthermore, existing studies only included blood samples collected at baseline and at conversion. This design likely to result in large variations in the latency to conversion between samples collected before conversion and could mask changes only observable in the few months before conversion.~~

Commented [ST4]: Changes made to improve precision of wording

The peripheral blood is an attractive source for biomarker discovery (Liew et al. 2006). Many studies have been conducted in attempts to identify blood-based biomarkers of psychosis conversion (Ota et al. 2019; Demars et al. 2020; Chaumette et al. 2019; Mongan et al. 2020; English et al. 2018; Perkins et al. 2015; Laskaris et al. 2019). However, the majority of existing studies are either cross-sectional (Ota et al. 2019; Laskaris et al. 2019) or only assessed the expression changes of a selected list of candidate genes (English et al. 2018; Perkins et al. 2015; Demars et al. 2020). Transition into psychosis is an evolving process and employing cross-sectional designs tend to introduce significant bias (Maxwell and Cole 2007). Additionally, assessing a relatively small set of candidate genes precludes the possibility of discovering novel biomarkers and the characterization of transcriptomic changes using methods such as Gene Set Enrichment Analysis (GSEA) (Subramanian et al. 2005). To date, only 2 studies profiled the transcriptomic changes of psychosis conversion in a longitudinal study cohort using high-throughput methods (Chaumette et al. 2019; Mongan et al. 2020).

Existing studies are limited in a few aspects: First, at the time of writing, no studies have examined the gene ontologies of the genes differentially expressed in psychosis conversion. Second, it is unknown if the genes differentially expressed in psychosis conversion are necessarily differential between pre-conversion and non-conversion participants – a property that reflects the suitability and reliability of the genes as prognostic biomarker. Third, the intra-individual correlation inherent to longitudinal data were not accounted for in existing studies (Chaumette et al. 2019; Mongan et al. 2020). Fourth, over 50% of converters in existing studies reported a history of cannabis use (Chaumette et al. 2019; Mongan et al. 2020) – this is a

potential confounding factor as cannabis use has been strongly evidenced to confers additional risk in transition to psychosis (Gage, Hickman, and Zammit 2016).

~~In this study, we performed RNA-sequencing on whole blood samples from the Longitudinal Youth-at-Risk Study (LYRIKS) cohort(Lee et al. 2013). The LYRIKS cohort was recruited in Singapore, which have one of the lowest substance abuse rate in the world(United Nations Office on Drugs and Crime 2006) and participants with history of substance or anti-psychotic use were excluded(Lee et al. 2013). Blood samples were collected from the LYRIKS cohort at 12 months interval: allowing for additional stratification of samples collected before conversion. We performed differential gene expression analysis to identify gene ontology enrichment analysis to examine the transcriptomic changes in the peripheral blood as a participant convert to psychosis. We also examined whether the gene that are differential with psychosis conversion are necessarily useful in identifying samples collected before conversion from samples from patients who do not convert to psychosis.~~

We performed RNA-sequencing on whole blood samples from a subsample of the LYRIKS cohort (Lee et al. 2013). Peripheral whole blood samples were collected from the LYRIKS cohort at 12 months interval and individuals with a history of substance use (including cannabis) were excluded from the study (Lee et al. 2013).

Using a linear-mixed effect model-based method (Hoffman and Roussos 2021), we examined the differential expression of gene in psychosis conversion and the gene ontology enrichment of the differentially expressed genes. The linear-mixed effect model based method allows us to model the intra-individual correlations within the data (Diggle 2013). Along with the gene ontology enrichment analysis, the differential expression analysis also provides us an

opportunity to examine whether a set of 120 genes containing high-risk loci identified by a recent large scale meta-analysis are differentially expressed with psychosis conversion in the peripheral blood (Trubetskoy et al. 2022). We then examined whether the genes that are differential with psychosis conversion are necessarily differential between pre-conversion and non-conversion participants. The drug-free nature of our cohort also allows us to provide a perspective into the transcriptomic changes of psychosis conversion with minimal confounding by cannabis and other illicit substances use.

This study aims to expand upon the existing understanding of blood-based biomarkers of psychosis conversion and provide directions for future studies in this domain.

Commented [ST5]: Revision of the introduction to better frame the study

Methods

Participants

The Longitudinal Youth-at-Risk study (LYRIKS) was a prospective, observational study conducted in Singapore on youths aged 14 to 29 to assess risk factors for psychosis (Lee et al., 2013). Briefly, participants were followed up over 2 years and were recruited from psychiatric clinics, various community agencies including educational institutes and social services, or were self-referred. Participants were excluded if they (i) had a past or current history of psychosis or intellectual disability, (ii) were currently using illicit substances, (iii) were taking mood stabilizers, (iv) had previous antipsychotic exposure of more than 5 mg haloperidol per day for 3 weeks (or equivalent) or were on an antipsychotic at the point of recruitment, or (iv) had medical causes associated with their attenuated psychotic symptoms.

Ethics approval for this study was provided by the National Healthcare Group's Domain Specific Review Board (approval number 2009/00167). After complete description of the study to the participants, written informed consent was obtained. The present study utilised whole blood samples from 135 participants (70 control (CTRL) and 65 UHR) from the original LYRIKS cohort (Lee et al. 2013). All samples from participants who converted were included. Only samples from participants with at least 2 follow-ups were included for control and participants who did not convert. UHR designation and psychosis threshold were assessed using CAARMS. Venous whole blood samples were collected from the participants at 12-month intervals. Participants who converted at any point in the study were removed from the study following a final sample collection. Of the 65 UHR participants, 13 converted to psychosis (CVT) and 52 did not (non-CVT). In total, 392 venous whole blood samples were collected ~~RNA libraries were collected from the 135 participants across 5 timepoints namely: Baseline, 6 months, 12 months, 18 months, and 24 months (or conversion in CVT)~~ (**Supplementary Table S1**).

Commented [ST6]: This change was made in response to the comment "English should be improved. E.g., the use of the word "terms" in line 239; line 82, "In total, 392 RNA libraries from the were collect from the…""

As explained in the accompanying letter, "RNA libraries" is replaced by "venous whole blood samples"

The portion relating to the various timepoint has been deleted as it is redundant to the information in Supplementary Table S1 and disrupts the flow of the sentence.

Demographic

The control group (CTRL) consist of 25 [35.7%] females with mean [SD] age of 21.1 [3.07] years and 45 [64.3%] males with mean [SD] age of 22.8 [3.51] years. The UHR group consist of 46 [70.8%] males with mean [SD] age of 21.4 [3.08] years and 19 [29.2%] females with mean [SD] age of 21.9 [3.05] years. Of the 65 UHR participants, 52 participants did not develop psychosis over the 2-year study period (non-CVT). Of the 52 non-CVT participants, 17 [32.7%] are female with a mean [SD] age of 22.2 [3.3] years and 35 [67.3%] are male with mean [SD] age of 21.4 [3.08] years. The remaining 13 [20%] UHR participants with mean [SD] age of 20.3

[3.13] years developed psychosis over the study period (CVT); 2 [15.4%] are females. The median time to conversion from baseline was 11 months [IQR 19 - 5 = 14].

No statistically significant difference in sex proportion was found between CVT participants and CTRL participants (Chi-squared = 2.06, p = 0.205) or with non-CVT participants (Chi-squared = 3.06; P= 0.080). No statistically significant difference in age was found between CVT participants and CTRL participants (t = 1.05, 95% CI, [-1.05, 3.11]; P = 0.310) or non-CVT participants (t = -0.48142; 95% CI, [-2.58, 1.62]; P = 0.636).

Sequencing and pre-processing

A sample of venous whole blood was collected from all participants into Tempus™ Blood RNA tube (Thermo Fisher Scientific) and stored at −80◦C. Total RNA was extracted with the Tempus™ Spin RNA Isolation Kit (Thermo Fisher Scientific) according to manufacturer's protocol. Quality and concentration of extracted RNA was measured with 2100 Bioanalyzer System (Agilent). The libraries were prepped and sequenced using 8 Illumina NovaSeq (2x151bp) sequencing lanes. Sequencing quality was assessed using FastQC v0.11.9 (Schmieder and Edwards 2011). Sequences are aligned to Ensembl GRCh38 using STAR 2.7.9a (Dobin et al. 2013). Aligned reads are quantified using featureCount from the Subread 2.0.3 R Package (Liao, Smyth, and Shi 2019). Genes with zero variance and mean count < 1 were discarded yielding a total of 21810 genes. All processing and analysis were performed in the R Programming language Version 4.0.3 (R Core Team 2013).

Power analysis

We performed power analysis using ssizeRNA R package (Bi and Liu 2016). We defined our criteria at absolute log2 fold-change > 1 and false discover rate = 0.05. Following the low number of differentially expressed genes in similar studies (Mongan et al. 2020; Chaumette et al. 2019), we set our expected proportion of differentially expressed genes at 0.005. We found that a sample size of 11 is sufficient to achieve a statistical power of 0.81. At 13 converter samples, our study has a power of 0.851.

Differential gene analysis

We divided the samples into 5 groups (**Figure 1**):

(1) Samples from participants who did not develop psychosis (non-CVT)

(2) Samples from participants post-conversion (CVT)

(3) Samples collected less than, or equal to 12 months before conversion (SHORT)

(4) Samples collected more than 12 months before conversion (MEDIUM)

(5) Samples from participants healthy controls (CTRL)

Samples in SHORT and MEDIUM will be referred to as pre-conversion samples (pre-CVT). *Sex* and *time* were included as model covariates where ~~*time* ∈ {0,6,12,18,24} is the set of number of months for follow-up since the sample was collected~~ *time* is the month of sample collection. Longitudinal data involves repeated measurements from the same participant across time which introduces intra-individual correlation. Repeated measurements result in intra-individual correlations that needs to be accounted for in the model. ~~The approach we chose is to deal with longitudinal data~~ One method capable of modelling such correlation is the linear

Commented [ST7]: This change was made to improve sentence flow

Commented [ST8]: This change was made to improve sentence flow

mixed effect model (Diggle 2013). We modelled known co-variates in the data via a linear mixed effect model using the DREAM model (Hoffman and Roussos 2021) from the voom-limma pipeline (Law et al. 2014). We modelled the slope of each sample across time and sample intercept as random effects:

$$y = group + sex + time + (time|sample) + (0|sample)$$

p-value adjustment (adjusted p-value) for multiple testing was addressed via Benjamini-Hochberg correction (Benjamini and Hochberg 1995). Threshold for differential expression was defined at an absolute log fold-change of > 1 and an adjusted p-value of < 0.05. List of contrasts used can be found in **Supplementary Notes.**

<Insert Figure 1>

Functional profiling via Gene Ontology analysis

Enrichment analysis for Gene Ontology (GO) functional terms was performed using Enrichr (Chen et al. 2013) webtool (which employs a Fisher's exact test). We examined the top 10 enriched ontologies ranked by adjusted p-value for analysis and interpretation.

Phenotype association analysis

GWAS phenotype associations were queried from the Genecards database (Fishilevich et al. 2017; Stelzer et al. 2016). GWAS phenotypes with less than 5 occurrences amongst the differential genes were dropped. We perform Fisher's exact tests with Bonferroni correction to determine if the differential genes are enriched for SNP associated with a particular phenotype

found in the European Bioinformatics Institute (EBI) GWAS catalog (Buniello et al. 2019). The same EBI catalogue is used in Genecards (Stelzer et al. 2016). Further details on data processing procedure are detailed in **Supplementary Notes**.

Results

Differential gene expression analysis

We identified 12, and 130 genes differentially expressed in MEDIUM vs CVT, and SHORT vs CVT respectively. The 10 differential genes with largest fold changes are listed in **Table 1,** the full list of differential genes can be found in **Supplementary Table S2-S3.** No genes were found to be differential in MEDIUM vs SHORT. We also identified 6, and 22 genes differentially expressed in MEDIUM vs non-CVT, and SHORT vs non-CVT respectively (**Table 1**) (full list in **Supplementary Table S4-S5**). No differential genes were identified when comparing CTRL vs CVT. We then examined whether the genes reported to contain high-risk loci in a recent large-scale meta-analysis were differentially expressed in our analysis (Trubetskoy et al. 2022). None of the genes associated with those loci were differentially expressed in our analyses.

Commented [ST9]: This segment was originally intended as a supplementary analysis. With the reframing of the study, we believe it is more appropriate to merge it with this analysis

< Insert Table 1 >

Gene ontology enrichment analysis

We performed enrichment analysis of the differential genes found in SHORT vs CVT and MEDIUM vs CVT to identify enriched gene ~~ontology terms~~ ontologies. The enriched ontologies in SHORT vs CVT ~~The genes differential in SHORT vs CVT are mostly enriched for GO terms associated with mitosis **(Table 2)**. The enriched GO terms in MEDIUM vs CVT are also consistent with the theme of mitosis observed in SHORT vs CVT **(Table 2)**.~~ contain commonalities namely: mitotic spindle assembly and organization (GO:0007052; GO:0071173; GO:0007094; GO:0071174), chromosomal segregation (GO:0051983; GO:0000070), and DNA replication and cell cycle regulation (GO:1901990; GO:0006260) **(Table 2)**. Commonalities in the enriched ontologies in MEDIUM vs CVT include: cell cycle checkpoint (GO:0044774; GO:0033314; GO:0000076; GO:0000083), and nuclear membrane disassembly (GO:0007077; GO:0051081) **(Table 2)**. However, the overlap is minimal between the GO gene sets and the genes differential in MEDIUM vs CVT, likely due to the small number of differential genes in this comparison.

< Insert Table 2 >

GWAS phenotypes enrichment analysis

We examined the GWAS phenotypes associated with SNPs within the differentially expressed genes for SHORT vs CVT and MEDIUM vs CVT. We found 25 and 4 GWAS phenotypes that are enriched amongst the GWAS phenotypes of genes differential in SHORT vs CVT and MEDIUM vs CVT respectively **(Table 3)**. ~~The enriched GWAS phenotypes are largely associated with relative abundance of leukocyte populations and related cellular measurements. Our~~

Commented [ST10]: This change was made in response to the comment:

"The second paragraph of the Discussion section should describe with much more detail the findings of the study. Which genes were correlated with mitosis? What are the roles of the genes found? Authors should also better explain why this pattern is indicative of immune upregulation, and other assumptions like "hematological phenotypes" found in their sample. Please provide rationale for the arguments hypothesized, so that the findings make physiological sense to the reader."

The ontologies are grouped based on their commonalities

~~observation is consistent with prior observations that psychosis conversion is associate with immune response (Tomasik et al. 2016). We also noted well known co-morbidities of schizophrenia such as "insomnia" and "coronary artery disease" (Palmese et al. 2011; Ifteni et al. 2014). Enrichment in GWAS phenotypes relating to physiological measurements such as "fatty acid measurements", "body height", "systolic blood pressure", and "gut microbiome measurements" were also observed.~~ The differentially expressed genes were not enriched for any schizophrenia and psychosis GWAS phenotypes **(Supplementary Table S6)**.

< Insert Table 3 >

Comparison between genes differential in psychosis conversion and genes differential in pre-conversion and non-conversion

We examined whether the genes differential in psychosis conversion are also differential between pre-conversion and non-conversion participants. We identified the intersection (denoted by the symbol ∩) between the set of genes differential in psychosis conversion and the set of genes differential between pre-conversion and non-conversion. We found 2 genes in SHORT vs CVT ∩ SHORT vs non-CVT and 5 gene in both MEDIUM vs CVT ∩ MEDIUM vs non-CVT. CYP51A1P1 and UBE3AP2 were differential in both SHORT vs CVT and SHORT vs non-CVT. CA12, COL9A2 and pseudogenes ENSG00000285966, ENSG00000248936, ENSG00000278112 were differential in both MEDIUM vs CVT and MEDIUM vs non-CVT.

Commented [ST11]: This change was made in response to the comment:

"In the Discussion, use of term comorbidity might be inadequate, as authors did not investigate other illnesses in their sample. Nor were their findings related to comorbidities. Genes might be related to other physical entities but stating that they are related to comorbidities in schizophrenia is not quite right. As a matter of fact, immune response is possibly an integral component of disease pathophysiology, and not a comorbidity."

The result that schizophrenia and psychosis GWAS phenotypes were not enrich was explicitly stated, a new supplementary table (S6) was added.

Discussion

Our study possesses several properties that allow for unique perspectives into the nuances of peripheral blood-based transcriptomic changes in UHR and psychosis conversion.

~~First, to our knowledge, our study is the first to partition pre-conversion samples based on their time-to-conversion. Second, owing to strict anti-drug laws in Singapore and the exclusion of participants with a history substance or anti-psychotic use, the LYRIKS cohort is minimally contaminated by substance use(Lee et al. 2013). We showed that the genes correlated with psychosis conversion are predominantly enriched for immune related biological functions and that many of the GWAS phenotypes enriched those genes related to both immunological phenotypes and known co-morbidities of schizophrenia. We also showed that genes correlated with psychosis conversion are not necessarily the same as those useful for prognosis.~~ First, at the time of writing, we are the first to conduct a gene ontology enrichment analysis on the genes differential in psychosis conversion using a longitudinal cohort. Second, we employed a mixed-effect model to account for the intra-individual correlation within a longitudinal data – a factor unaccounted for in previous studies (Chaumette et al. 2019; Mongan et al. 2020). Third, owing to the recruitment criteria, our study cohort is minimally confounded by substance use. Fourth, at the time of writing, this study is the first to examine whether the gene differential in psychosis conversion is also differential between pre-conversion and non-conversion participants.

Commented [ST12]: This section described the novelty of our study. It has been amended to reflect changes made to the framing of the study in the introduction

The genes identified in recent whole exome (Singh et al. 2022) and large scale GWAS (Trubetskoy et al. 2022) analyses to confer high risk of schizophrenia were not differentially

expressed in our analyses. This is consistent with the finding that the genes were mostly expressed in synaptic junctions rather than the peripheral blood (Trubetskoy et al. 2022). This finding highlights the difficulty of employing schizophrenia genetic risk loci as a basis of identifying biomarkers in the peripheral blood.

~~Our GO enrichment analysis showed that genes correlated with psychosis conversion are largely enriched for ontologies relating for mitosis. Considering that nucleated cells in the whole blood are generally leukocytes, this pattern is indicative of a general immune upregulation widely observed in schizophrenia and psychosis conversion (Gardiner et al. 2013; Tomasik et al. 2016; Xu et al. 2012). However, the low number of differential genes in MEDIUM vs CVT also meant that the overlaps with GO terms are small, we advise discretion when considering the GO analysis results of MEDIUM vs CVT.~~ The commonalities in the enriched gene ontologies include 1) mitotic spindle assembly and organization, 2) chromosomal segregation, 3) DNA replication and cell cycle regulation cell cycle checkpoint, and 4) nuclear membrane disassembly – all processes of cellular replication. This pattern suggests that the source cells are proliferating. Considering that nucleated cells in the whole blood are generally leukocytes, the source cells are predominately leukocytes and this observation is consistent with a pattern of immune upregulation widely observed in schizophrenia and psychosis conversion (Gardiner et al. 2013; Tomasik et al. 2016; Xu et al. 2012). Disentangling biomarkers specific to psychosis conversion from markers of general immune responses will be a crucial step in the identification of peripheral blood-based biomarkers of psychosis.

Commented [ST13]: This is a discussion of a result added in page 11 regarding whether the genes containing schizophrenia risk loci were differentially expressed in psychosis conversion

Commented [ST14]: This change was made in response to the comment

"The second paragraph of the Discussion section should describe with much more detail the findings of the study. Which genes were correlated with mitosis? What are the roles of the genes found? Authors should also better explain why this pattern is indicative of immune upregulation, and other assumptions like "hematological phenotypes" found in their sample. Please provide rationale for the arguments hypothesized, so that the findings make physiological sense to the reader."

~~Our GWAS phenotype enrichment analysis found that the genes correlated with psychosis conversion are generally associated with haematological GWAS phenotypes rather than neurological GWAS phenotypes. GWAS phenotypes relating to leukocyte percentages are of interest as cell population ratios such as neutrophil to lymphocyte ratio has been shown to be increased in chronic schizophrenia and psychosis conversion(Karageorgiou, Milas, and Michopoulos 2019). Many terms describing physiological measurements such as fatty acid measurement and systolic blood pressure are also reported to be different schizophrenic patients(Andreassen et al. 2013). We also observed some GWAS phenotypes describing known comorbidities of schizophrenia, such as insomnia and coronary artery disease (Palmese et al. 2011; Ifteni et al. 2014). And lastly, there is increasing attention into the role of the gut microbiome in mental disorders including schizophrenia and psychosis (Nguyen et al. 2019; Zhu et al. 2020). These observations suggest that the neurological changes associated with psychosis conversion are accompanied by a range of non-neurological concomitant changes, many of which have been reported to be correlated with and or co-morbidities of schizophrenia. How these non-neurological concomitant changes can be used to determine psychosis conversion risk should be a valuable research direction. Consequently, using blood-based markers to detect co-morbidities and concomitant changes of psychosis and schizophrenia could represent a meaningful direction in predicting clinical outcomes in UHR individuals.~~ Gene differential with psychosis conversion were not enriched with SNPs associated with schizophrenia or psychosis phenotypes. This is consistent with a recent finding that found that the genes containing the SNPs conferring substantial risk to schizophrenia were predominately expressed by neuronal cells in the central nervous system (Trubetskoy et al.

2022). Several of the enriched phenotypes have been reported to be perturbed in individuals with schizophrenia, these include cell population ratios such as neutrophil to lymphocyte ratio (Karageorgiou, Milas, and Michopoulos 2019), physiological measurements such as fatty acid measurement and systolic blood pressure (Andreassen et al. 2013), disorders and diseases frequently observed in individuals with schizophrenia such as insomnia and coronary artery disease (Palmese et al. 2011; Ifteni et al. 2014), and variations in the gut microbiome (Nguyen et al. 2019; Zhu et al. 2020). We hypothesize that the association between psychosis conversion and biomarkers readily observable in the peripheral whole blood is likely to be indirect. Whether a differential expression of genes enriched for SNPs associated with these phenotypes necessarily indicate a perturbation of the phenotype, and how are these non-neurological phenotypes implicated in the pathophysiology of psychosis will be important research directions to evaluate this hypothesis. If positive evidence of both questions can be found, using blood-based markers to detect concomitant changes of psychosis and schizophrenia could represent a viable avenue for predicting clinical outcomes in UHR individuals.

Genes differential with psychosis conversion are not necessarily differential between pre-conversion and non-conversion individuals. Of the genes that exist in the overlaps, CYP51A1P1 and UBE3AP2 were differential in both SHORT vs CVT and SHORT vs non-CVT. CYP51A1P1 is a pseudogene for cytochrome P450 family 51 subfamily A member 1 (CYP51A1). The cytochrome P450 family proteins are predominantly involved in lipids and cholesterol metabolism (Stelzer et al. 2016). UBE3AP2 is a pseudogene for ubiquitin protein ligase 3A (UBE3A) and was reported to associate with hypertension (Stelzer et al. 2016). While certain rare variants in UBE3A has been proposed to confer risks to schizophrenia (Singh et al. 2022;

Commented [ST15]: This change was made in response to the comment

"In the Discussion, use of term comorbidity might be inadequate, as authors did not investigate other illnesses in their sample. Nor were their findings related to comorbidities. Genes might be related to other physical entities but stating that they are related to comorbidities in schizophrenia is not quite right. As a matter of fact, immune response is possibly an integral component of disease pathophysiology, and not a comorbidity."

The discussion of the GWAS phenotypes has been reframed. We begin by discussing the result that we did not observe enrichment in genes containing schizophrenia risk loci. We then discuss that many of the enriched phenotypes have been reported to be perturbed in schizophrenia. Based on these findings, we hypothesize that this pattern suggest that the link between psychosis transition and blood-based biomarkers is indirect.

Commented [ST16]: This segment has been added to improve sentence flow

Trubetskoy et al. 2022), it is unclear if this correlation might extend to pseudogenes of UBE3A and to psychosis conversion. Genes CA12, COL9A2 were differential in both MEDIUM vs CVT and MEDIUM vs non-CVT. CA12 is a carbonic anhydrase which catalyzes the reversible hydration of carbon dioxide. GWAS associated studies reported that CA12 is associated with a large range of phenotypes including schizophrenia symptom severity (Stelzer et al. 2016). COL9A2 codes for an alpha chain subunit in type IX collagen. Associations between COL9A2 and schizophrenia are indirect, via phenotypes such as cognitive measurements (Lam et al. 2018), testosterone measurements (Moore et al. 2013), and waist-hip-ratio (Andreassen et al. 2013). At the time of writing, the functions of ENSG00000285966, ENSG00000248936, and ENSG00000278112 are unknown and will not be discussed. Based on our study design, these genes and pseudogenes are theoretical candidate biomarkers for psychosis conversion. However, the lack of robust links to pathways linked to the pathology of schizophrenia dampens our confidence in the ability of these genes and pseudogenes to serve as biomarkers of psychosis conversion.

~~Lastly, a recent large-scale meta-analysis identified a set of loci likely to confer substantial risk of schizophrenia (Singh et al. 2022; Trubetskoy et al. 2022). None of gene associated with those loci were differentially expressed in any of our analysis. This is unsurprising as most of those genes were reported to be expressed in synaptic junctions rather than the peripheral blood(Trubetskoy et al. 2022). This discrepancy further highlights the difficulty of generalizing findings based on schizophrenia cohorts to UHR cohorts and from schizophrenia to psychosis.~~

Commented [ST17]: This segment originally intended as a supplementary analysis had been merged with the differential gene expression analysis

Limitation

Our study contains several limitations. Firstly, the sample size of 392 used in this study is relatively small especially since samples are distributed into multiple time-points across various groups. Hence, the inferential statistical analyses performed in this study may be underpowered. However, we opine that a reliable (and useful) biomarker should be robust and detectable, even against the backdrop of small sample sizes. Secondly, we did not distinguish non-CVT participants who remitted from the UHR status from non-CVT participants who maintained UHR status to avoid further reducing sample sizes across groups. Thirdly, the etiology of psychosis conversion is inherently heterogenous (Zwicker, Denovan-Wright, and Uher 2018) and without a basis for pre-defining all medically relevant subtypes, our study design was unable to account for the heterogenous nature of psychosis conversion. Finally, the overlaps in our gene ontology enrichment analysis are small especially in MEDIUM vs CVT, we advise the readers to exercise discretion when considering the results.

Conclusion

The identification of peripheral blood-based biomarkers for psychosis is an area of great clinical value. Our study expanded upon existing studies by illustrating various important nuances in the identification of peripheral blood-based biomarkers of psychosis conversion and establishes directions for future investigations. We identified differentially expressed genes in psychosis conversion, which were consistent with a pattern of immune upregulation. The SNPs in the genes differential in psychosis conversion were associated with a range of phenotypes reported to be perturbed in schizophrenia and psychosis but are not associated with

schizophrenia and psychosis phenotypes themselves. We also showed that the genes correlated with psychosis conversion are not necessarily differential between pre-conversion and non-conversion individuals. The genes containing risk loci for schizophrenia were not differentially expressed in the peripheral whole blood in psychosis conversion. From these findings, we hypothesize that the associations between psychosis conversion and peripheral blood biomarkers are indirect. Mechanistic studies to examine such indirect associations will be a valuable research direction.

Commented [ST18]: We have added conclusion to provide a summary to the study

Data availability

The results of each contrast in the differential genes analysis can be found in the Supplementary section. Supplementary information is available at the journal website. Raw transcripts counts are not available due to patient confidentiality requirements. Analysis code is available on request from the authors.

Author contribution

S.M.X.T., Y.J.Y., J.L., and W.W.B.G conceptualized and designed the research; S.M.X.T., Y.J.Y., performed the research; S.M.X.T. performed that statistical analysis; S.M.X.T. and Y.J.Y. wrote the manuscript; Y.J.Y. and J.L. were responsible for the acquisition of the data; S.M.X.T., B.S., S.B., Z.D., M.D., N.K., E.L., A.S, Y.J.Y., J.L., and W.W.B.G interpretated the results; B.S., S.B., Z.D., M.D., N.K., E.L., A.S, Y.J.Y., J.L., and W.W.B.G reviewed the work. All authors approve of the submitted version of the manuscript.

Transparency declaration

We affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

References

Addington, Jean, Kristin S. Cadenhead, Barbara A. Cornblatt, Daniel H. Mathalon, Thomas H. McGlashan, Diana O. Perkins, Larry J. Seidman, et al. 2012. "North American Prodrome Longitudinal Study (NAPLS 2): Overview and Recruitment." *Schizophrenia Research* 142 (1–3): 77–82. https://doi.org/10.1016/j.schres.2012.09.012.

Andreassen, Ole A., Srdjan Djurovic, Wesley K. Thompson, Andrew J. Schork, Kenneth S. Kendler, Michael C. O'Donovan, Dan Rujescu, et al. 2013. "Improved Detection of Common Variants Associated with Schizophrenia by Leveraging Pleiotropy with Cardiovascular-Disease Risk Factors." *The American Journal of Human Genetics* 92 (2): 197–209. https://doi.org/10.1016/j.ajhg.2013.01.001.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bi, Ran, and Peng Liu. 2016. "Sample Size Calculation While Controlling False Discovery Rate for Differential Expression Analysis with RNA-Sequencing Experiments." *BMC Bioinformatics* 17 (1): 146. https://doi.org/10.1186/s12859-016-0994-9.

Buniello, Annalisa, Jacqueline A L MacArthur, Maria Cerezo, Laura W Harris, James Hayhurst, Cinzia Malangone, Aoife McMahon, et al. 2019. "The NHGRI-EBI GWAS Catalog of

441 Published Genome-Wide Association Studies, Targeted Arrays and Summary Statistics
442 2019." *Nucleic Acids Research* 47 (D1): D1005–12. https://doi.org/10.1093/nar/gky1120.
443 Chaumette, Boris, Oussama Kebir, Juliette Pouch, Bertrand Ducos, Fekrije Selimi, ICAAR study
444 group, Raphael Gaillard, and Marie-Odile Krebs. 2019. "Longitudinal Analyses of Blood
445 Transcriptome During Conversion to Psychosis." *Schizophrenia Bulletin* 45 (1): 247–55.
446 https://doi.org/10.1093/schbul/sby009.
447 Chen, Edward Y, Christopher M Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Vaz
448 Meirelles, Neil R Clark, and Avi Ma'ayan. 2013. "Enrichr: Interactive and Collaborative
449 HTML5 Gene List Enrichment Analysis Tool." *BMC Bioinformatics* 14 (1): 128.
450 https://doi.org/10.1186/1471-2105-14-128.
451 Demars, Fanny, Oussama Kebir, Aude Marzo, Anton Iftimovici, Catherine Schramm, ICAAR Study
452 Group, Isabelle Amado, et al. 2020. "Dysregulation of Peripheral Expression of the
453 YWHA Genes during Conversion to Psychosis." *Scientific Reports* 10 (1): 9863.
454 https://doi.org/10.1038/s41598-020-66901-1.
455 Diggle, Peter, ed. 2013. *Analysis of Longitudinal Data*. Second Paperback Edition. Oxford
456 Statistical Science Series 25. Oxford: Oxford University Press.
457 Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha,
458 Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal
459 RNA-Seq Aligner." *Bioinformatics* 29 (1): 15–21.
460 https://doi.org/10.1093/bioinformatics/bts635.
461 English, Jane A, Lorna M Lopez, Aoife O'Gorman, Melanie Föcking, Magdalena Hryniewiecka,
462 Caitriona Scaife, Sophie Sabherwal, et al. 2018. "Blood-Based Protein Changes in
463 Childhood Are Associated With Increased Risk for Later Psychotic Disorder: Evidence
464 From a Nested Case–Control Study of the ALSPAC Longitudinal Birth Cohort."
465 *Schizophrenia Bulletin* 44 (2): 297–306. https://doi.org/10.1093/schbul/sbx075.
466 Fishilevich, Simon, Ron Nudel, Noa Rappaport, Rotem Hadar, Inbar Plaschkes, Tsippi Iny Stein,
467 Naomi Rosen, et al. 2017. "GeneHancer: Genome-Wide Integration of Enhancers and
468 Target Genes in GeneCards." *Database* 2017 (January).
469 https://doi.org/10.1093/database/bax028.
470 Fusar-Poli, Paolo, Patrick D. McGorry, and John M. Kane. 2017. "Improving Outcomes of First-
471 Episode Psychosis: An Overview." *World Psychiatry* 16 (3): 251–65.
472 https://doi.org/10.1002/wps.20446.
473 Gage, Suzanne H., Matthew Hickman, and Stanley Zammit. 2016. "Association Between
474 Cannabis and Psychosis: Epidemiologic Evidence." *Biological Psychiatry* 79 (7): 549–56.
475 https://doi.org/10.1016/j.biopsych.2015.08.001.
476 Gardiner, Erin J., Murray J. Cairns, Bing Liu, Natalie J. Beveridge, Vaughan Carr, Brian Kelly,
477 Rodney J. Scott, and Paul A. Tooney. 2013. "Gene Expression Analysis Reveals
478 Schizophrenia-Associated Dysregulation of Immune Pathways in Peripheral Blood
479 Mononuclear Cells." *Journal of Psychiatric Research* 47 (4): 425–37.
480 https://doi.org/10.1016/j.jpsychires.2012.11.007.
481 Hoffman, Gabriel E, and Panos Roussos. 2021. "Dream: Powerful Differential Expression
482 Analysis for Repeated Measures Designs." Edited by Jan Gorodkin. *Bioinformatics* 37 (2):
483 192–201. https://doi.org/10.1093/bioinformatics/btaa687.

Ifteni, Petru, Christoph U. Correll, Victoria Burtea, John M. Kane, and Peter Manu. 2014. "Sudden Unexpected Death in Schizophrenia: Autopsy Findings in Psychiatric Inpatients." *Schizophrenia Research* 155 (1–3): 72–76. https://doi.org/10.1016/j.schres.2014.03.011.

Karageorgiou, Vasilios, Gerasimos P. Milas, and Ioannis Michopoulos. 2019. "Neutrophil-to-Lymphocyte Ratio in Schizophrenia: A Systematic Review and Meta-Analysis." *Schizophrenia Research* 206 (April): 4–12. https://doi.org/10.1016/j.schres.2018.12.017.

Lam, Max, Jimmy Lee, Attilio Rapisarda, Yuen Mei See, Zixu Yang, Sara-Ann Lee, Nur Amirah Abdul-Rashid, et al. 2018. "Longitudinal Cognitive Changes in Young Individuals at Ultrahigh Risk for Psychosis." *JAMA Psychiatry* 75 (9): 929. https://doi.org/10.1001/jamapsychiatry.2018.1668.

Laskaris, Liliana, Andrew Zalesky, Cynthia Shannon Weickert, Maria A. Di Biase, Gursharan Chana, Bernhard T. Baune, Chad Bousman, et al. 2019. "Investigation of Peripheral Complement Factors across Stages of Psychosis." *Schizophrenia Research* 204 (February): 30–37. https://doi.org/10.1016/j.schres.2018.11.035.

Law, Charity W, Yunshun Chen, Wei Shi, and Gordon K Smyth. 2014. "Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts." *Genome Biology* 15 (2): R29. https://doi.org/10.1186/gb-2014-15-2-r29.

Lee, Jimmy, Gurpreet Rekhi, Natasha Mitter, Yioe Ling Bong, Michael S. Kraus, Max Lam, Attilio Rapisarda, et al. 2013. "The Longitudinal Youth at Risk Study (LYRIKS) — An Asian UHR Perspective." *Schizophrenia Research* 151 (1–3): 279–83. https://doi.org/10.1016/j.schres.2013.09.025.

Liao, Yang, Gordon K Smyth, and Wei Shi. 2019. "The R Package Rsubread Is Easier, Faster, Cheaper and Better for Alignment and Quantification of RNA Sequencing Reads." *Nucleic Acids Research* 47 (8): e47–e47. https://doi.org/10.1093/nar/gkz114.

Liew, Choong-Chin, Jun Ma, Hong-Chang Tang, Run Zheng, and Adam A. Dempsey. 2006. "The Peripheral Blood Transcriptome Dynamically Reflects System Wide Biology: A Potential Diagnostic Tool." *Journal of Laboratory and Clinical Medicine* 147 (3): 126–32. https://doi.org/10.1016/j.lab.2005.10.005.

Maxwell, Scott E., and David A. Cole. 2007. "Bias in Cross-Sectional Analyses of Longitudinal Mediation." *Psychological Methods* 12 (1): 23–44. https://doi.org/10.1037/1082-989X.12.1.23.

Miller, T. J., T. H. McGlashan, J. L. Rosen, K. Cadenhead, J. Ventura, W. McFarlane, D. O. Perkins, G. D. Pearlson, and S. W. Woods. 2003. "Prodromal Assessment With the Structured Interview for Prodromal Syndromes and the Scale of Prodromal Symptoms: Predictive Validity, Interrater Reliability, and Training to Reliability." *Schizophrenia Bulletin* 29 (4): 703–15. https://doi.org/10.1093/oxfordjournals.schbul.a007040.

Mongan, David, Melanie Föcking, Colm Healy, Subash Raj Susai, Meike Heurich, Kieran Wynne, Barnaby Nelson, et al. 2020. "Development of Proteomic Prediction Models for Transition to Psychotic Disorder in the Clinical High-Risk State and Psychotic Experiences in Adolescence." *JAMA Psychiatry* 78 (1): 77–90. https://doi.org/10.1001/jamapsychiatry.2020.2459.

Moore, Loretta, Merribel Kyaw, Ans Vercammen, Rhoshel Lenroot, Jayashri Kulkarni, Jackie Curtis, Maryanne O'Donnell, Vaughan J Carr, C Shannon Weickert, and Thomas W

528 Weickert. 2013. “Serum Testosterone Levels Are Related to Cognitive Function in Men
529 with Schizophrenia.” *Psychoneuroendocrinology* 38 (9): 1717–28.
530 Nguyen, Tanya T., Tomasz Kosciolek, Yadira Maldonado, Rebecca E. Daly, Averria Sirkin Martin,
531 Daniel McDonald, Rob Knight, and Dilip V. Jeste. 2019. “Differences in Gut Microbiome
532 Composition between Persons with Chronic Schizophrenia and Healthy Comparison
533 Subjects.” *Schizophrenia Research* 204 (February): 23–29.
534 https://doi.org/10.1016/j.schres.2018.09.014.
535 Oliver, Dominic, Maite Arribas, Joaquim Radua, Gonzalo Salazar de Pablo, Andrea De Micheli,
536 Giulia Spada, Martina Maria Mensi, et al. 2022. “Prognostic Accuracy and Clinical Utility
537 of Psychometric Instruments for Individuals at Clinical High-Risk of Psychosis: A
538 Systematic Review and Meta-Analysis.” *Molecular Psychiatry*, June.
539 https://doi.org/10.1038/s41380-022-01611-w.
540 Ota, Vanessa Kiyomi, Patricia Natalia Moretti, Marcos Leite Santoro, Fernanda Talarico, Leticia
541 Maria Spindola, Gabriela Xavier, Carolina Muniz Carvalho, et al. 2019. “Gene Expression
542 over the Course of Schizophrenia: From Clinical High-Risk for Psychosis to Chronic
543 Stages.” *Npj Schizophrenia* 5 (1): 5. https://doi.org/10.1038/s41537-019-0073-0.
544 Palmese, Laura B., Pamela C. DeGeorge, Joseph C. Ratliff, Vinod H. Srihari, Bruce E. Wexler,
545 Andrew D. Krystal, and Cenk Tek. 2011. “Insomnia Is Frequent in Schizophrenia and
546 Associated with Night Eating and Obesity.” *Schizophrenia Research* 133 (1–3): 238–43.
547 https://doi.org/10.1016/j.schres.2011.07.030.
548 Perkins, Diana O., Clark D. Jeffries, Jean Addington, Carrie E. Bearden, Kristin S. Cadenhead,
549 Tyrone D. Cannon, Barbara A. Cornblatt, et al. 2015. “Towards a Psychosis Risk Blood
550 Diagnostic for Persons Experiencing High-Risk Symptoms: Preliminary Results From the
551 NAPLS Project.” *Schizophrenia Bulletin* 41 (2): 419–28.
552 https://doi.org/10.1093/schbul/sbu099.
553 Quattrone, Diego, Laura Ferraro, Giada Tripoli, Caterina La Cascia, Harriet Quigley, Andrea
554 Quattrone, Hannah E. Jongsma, et al. 2021. “Daily Use of High-Potency Cannabis Is
555 Associated with More Positive Symptoms in First-Episode Psychosis Patients: The EU-GEI
556 Case–Control Study.” *Psychological Medicine* 51 (8): 1329–37.
557 https://doi.org/10.1017/S0033291720000082.
558 R Core Team. 2013. “R: A Language and Environment for Statistical Computing.”
559 Schmieder, R., and R. Edwards. 2011. “Quality Control and Preprocessing of Metagenomic
560 Datasets.” *Bioinformatics* 27 (6): 863–64.
561 https://doi.org/10.1093/bioinformatics/btr026.
562 Singh, Tarjinder, Timothy Poterba, David Curtis, Huda Akil, Mariam Al Eissa, Jack D. Barchas,
563 Nicholas Bass, et al. 2022. “Rare Coding Variants in Ten Genes Confer Substantial Risk
564 for Schizophrenia.” *Nature* 604 (7906): 509–16. https://doi.org/10.1038/s41586-022-
565 04556-w.
566 Stelzer, Gil, Naomi Rosen, Inbar Plaschkes, Shahar Zimmerman, Michal Twik, Simon Fishilevich,
567 Tsippi Iny Stein, et al. 2016. “The GeneCards Suite: From Gene Data Mining to Disease
568 Genome Sequence Analyses.” *Current Protocols in Bioinformatics* 54 (1).
569 https://doi.org/10.1002/cpbi.5.
570 Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert,
571 Michael A. Gillette, Amanda Paulovich, et al. 2005. “Gene Set Enrichment Analysis: A

572 Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles."
573 *Proceedings of the National Academy of Sciences of the United States of America* 102
574 (43): 15545–50. https://doi.org/10.1073/pnas.0506580102.
575 Tomasik, Jakub, Hassan Rahmoune, Paul C. Guest, and Sabine Bahn. 2016. "Neuroimmune
576 Biomarkers in Schizophrenia." *Schizophrenia Research*. Elsevier B.V.
577 https://doi.org/10.1016/j.schres.2014.07.025.
578 Trubetskoy, Vassily, Antonio F. Pardiñas, Ting Qi, Georgia Panagiotaropoulou, Swapnil Awasthi,
579 Tim B. Bigdeli, Julien Bryois, et al. 2022. "Mapping Genomic Loci Implicates Genes and
580 Synaptic Biology in Schizophrenia." *Nature* 604 (7906): 502–8.
581 https://doi.org/10.1038/s41586-022-04434-5.
582 United Nations Office on Drugs and Crime. 2006. "World Drug Report 2006." United Nations,
583 Office on Drugs and Crime. https://www.unodc.org/unodc/en/data-and-analysis/WDR-
584 2006.html.
585 Xu, Junzhe, Jingchun Sun, Jingchun Chen, Lily Wang, Anna Li, Matthew Helm, Steven L
586 Dubovsky, Silviu-Alin Bacanu, Zhongming Zhao, and Xiangning Chen. 2012. "RNA-Seq
587 Analysis Implicates Dysregulation of the Immune System in Schizophrenia." *BMC*
588 *Genomics* 13 (8): 1–10. https://doi.org/10.1186/1471-2164-13-S8-S2.
589 Yung, A. R., P. D. McGorry, C. A. McFarlane, H. J. Jackson, G. C. Patton, and A. Rakkar. 1996.
590 "Monitoring and Care of Young People at Incipient Risk of Psychosis." *Schizophrenia*
591 *Bulletin* 22 (2): 283–303. https://doi.org/10.1093/schbul/22.2.283.
592 Yung, Alison R., Alison R. Yung, Hok Pan Yuen, Patrick D. Mcgorry, Lisa J. Phillips, Daniel Kelly,
593 Margaret Dell'olio, et al. 2005. "Mapping the Onset of Psychosis: The Comprehensive
594 Assessment of At-Risk Mental States." *Australian & New Zealand Journal of Psychiatry*
595 39 (11–12): 964–71. https://doi.org/10.1080/j.1440-1614.2005.01714.x.
596 Zhu, Feng, Yanmei Ju, Wei Wang, Qi Wang, Ruijin Guo, Qingyan Ma, Qiang Sun, et al. 2020.
597 "Metagenome-Wide Association of Gut Microbiome Features for Schizophrenia." *Nature*
598 *Communications* 11 (1): 1612. https://doi.org/10.1038/s41467-020-15457-9.
599 Zwicker, Alyson, Eileen M. Denovan-Wright, and Rudolf Uher. 2018. "Gene–Environment
600 Interplay in the Etiology of Psychosis." *Psychological Medicine* 48 (12): 1925–36.
601 https://doi.org/10.1017/S003329171700383X.

602

603

RNA-sequencing of peripheral whole blood of individuals at ultra-high-risk for psychosis – a longitudinal perspective

Samuel Ming Xuan Tan (BSc)[1,2,3], Yee Jie Yin (BSc)[4], Sugam Budhraja (BSc)[5], Balkaran Singh (BSc)[5], Zohreh Doborjeh (PhD)[6], Maryam Doborjeh (PhD)[5], Nikola Kasabov (PhD)[5], Edmund Lai (PhD)[5], Alexander Sumich (PhD)[7], Jimmy Lee (MD)[2,4], Wilson Wen Bin Goh (PhD)[1,2,3]

[1] School of Biological Sciences, Nanyang Technological University, Singapore

[2] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

[3] Center for Biomedical Informatics, Nanyang Technological University, Singapore

[4] Research Division, Institute of Mental Health, Singapore

[5] Knowledge Engineering and Discovery Research Innovation, School of Engineering Computer and Mathematical Sciences, Auckland University of Technology, New Zealand

[6] School of Population Health, The University of Auckland, New Zealand

[7] School of Social Sciences, Nottingham Trent University, UK

Corresponding author:

Wilson Wen Bin Goh

wilsongoh@ntu.edu.sg

Correspondence address:

NTU Main Campus,

Experimental Medicine Building

59 Nanyang Drive

Singapore 636921

+65 65927871