# Molecular Phenomics and Metagenomics of Hepatic Steatosis in Non-Diabetic Obese Women

Lesley Hoyles[1,10], José-Manuel Fernández-Real[2,10,11], Massimo Federici[3,10,11], Matteo Serino[4,5], James Abbott[1], Julie Charpentier[4,5], Christophe Heymes[4,5], Jèssica Latorre Luque[2], Elodie Anthony[6], Richard H. Barton[1], Julien Chilloux[1], Antonis Myridakis[1], Laura Martinez-Gili[1], José Maria Moreno-Navarrete[2], Fadila Rayah[6], Vincent Azalbert[4,5], Vincent Blasco-Baque[4,5], Josep Puig[2], Gemma Xifra[2], Wifredo Ricart[2], Christopher Tomlinson[1], Mark Woodbridge[1], Marina Cardellini[3], Francesca Davato[3], Iris Cardolini[3], Ottavia Porzio[7,8], Paolo Gentilieschi[7], Frédéric Lopez[4,5], Fabienne Foufelle[9], Sarah A. Butcher[1], Elaine Holmes[1], Jeremy K. Nicholson[1], Catherine Postic[6], Rémy Burcelin[4,5,11], Marc-Emmanuel Dumas[1,11]

[1] Division of Integrative Systems Medicine and Digestive Diseases, Department of Surgery and Cancer, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom

[2] Department of Endocrinology, Diabetes and Nutrition, Hospital of Girona “Dr Josep Trueta”, Universitat of Girona and CIBERobn Pathophysiology of Obesity and Nutrition, Instituto de Salud Carlos III, Madrid, Spain

[3] Department of Systems Medicine, University of Rome Tor Vergata, Via Montpellier 1 00133, Rome, Italy

[4] Institut National de la Santé et de la Recherche Médicale (INSERM), Toulouse, France

[5] Université Paul Sabatier (UPS), Unité Mixte de Recherche (UMR) 1048, Institut des Maladies Métaboliques et Cardiovasculaires (I2MC), Team 2: 'Intestinal Risk Factors, Diabetes, Dyslipidemia, and Heart Failure' F-31432 Toulouse Cedex 4, France

[6] Institut Cochin Inserm U1016 CNRS UMR 8104, Université Paris Descartes, 24 rue du Faubourg Saint Jacques, 75014 Paris

[7] Department of Experimental Medicine and Surgery, University of Rome Tor Vergata

[8] Department of Laboratory Medicine, Bambino Gesù Children's Hospital, Piazza S.Onofrio 4, 00165 Roma, Italy

[9] Sorbonne Universités, UPMC Univ Paris 06, UMRS 1138, Centre de Recherche des Cordeliers, F-75006, Paris, France.

[10] These authors contributed equally to this work.

[11] Correspondence should be sent to should be addressed to J.M.F.-R.(jmfreal@idibgi.org), M.F. (federicm@uniroma2.it), R.B. (remy.burcelin@inserm.fr), or M.-E.D. (m.dumas@imperial.ac.uk)

**ABSTRACT**

**The role of molecular signals from the microbiome and their coordinated interactions with those from the host in hepatic steatosis – notably in obese patients and as risk factors for insulin resistance and atherosclerosis – needs to be understood. We reveal molecular networks linking gut microbiome and host phenome to hepatic steatosis in a cohort of non-diabetic obese women. Steatotic patients had low microbial gene richness and increased genetic potential for processing of dietary lipids and endotoxin biosynthesis (notably from Proteobacteria), hepatic inflammation and dysregulation of aromatic and branched-chain amino acid (AAA and BCAA) metabolism. We demonstrated that faecal microbiota transplants and chronic treatment with phenylacetic acid (PAA), a microbial product of AAA metabolism, successfully trigger steatosis and BCAA metabolism. Molecular phenomic signatures were predictive (AUC = 87%) and consistent with the gut microbiome making an impact on the steatosis phenome (>75% shared variation) and, therefore, actionable via microbiome-based therapies.**

**Main**

Hepatic steatosis is a multi-factorial phenotype common to several chronic conditions such as insulin resistance, atherosclerosis and fatty liver disease, with increasing worldwide prevalence related to the obesity epidemic[1-5]. The gut microbiota recently emerged as a pivotal transducer of environmental influences (i.e., dietary components, drug treatments) to exert protective or detrimental effects on several host tissues and systems, including regulation of intermediary metabolism, liver function and cardiovascular disorders, either directly via translocation or indirectly through microbial metabolism or function in metabolic disorders[6-8,9-11]. Rodent studies demonstrated the role of the gut microbiome in liver disease and to the stratification of Type 2 diabetes (T2D) and cardiovascular disorders (CVD). Microbiome-associated factors involve, for instance, bacterial lipopolysaccharides (LPS) or methylamines such as trimethylamine (TMA) and trimethylamine *N*-oxide (TMAO)[12,13] playing a role in the development of insulin resistance and atherosclerosis[6,14,15]. Hepatic steatosis is a shared mechanism for the development of T2D and CVD in humans in both non-alcoholic and virus-associated fatty liver disease[16] but the physiological mechanisms behind this interplay remain poorly understood[17,18].

Here, we take advantage of the advances in high-throughput sequencing and phenotyping technologies to characterize in humans physiological mechanisms responsible for the integrated interactions between signals from the gut metagenome and the host molecular phenome (a comprehensive set of molecular phenotypes useful to identify subgroups of patients[17]) of hepatic steatosis. We introduce a unique integrative multi-omics and precision medicine approach combining shotgun metagenomics, liver transcriptomics, metabolomics in plasma and urine and clinical phenotyping to reveal the molecular mechanisms and multi-scalar interactions involved in the physiology of steatosis in a new cohort of non-diabetic obese women we recruited as part of the FLORINASH consortium.

In-depth analyses of faecal metagenomics and phenomics reveal a robust signature highlighting a tight crosstalk between the microbiome, host gene expression and metabolism in hepatic steatosis involving low microbial gene richness[19] (MGR) and imbalances in aromatic amino acid (AAA) and branched-chain amino acid (BCAA) metabolism[20,21]. Based on the results obtained in our clinical

study, we then demonstrated a causal role of the microbiota-produced metabolite most strongly associated with steatosis, namely phenylacetic acid (PAA), in the triggering of the hepatic steatosis phenome by faecal microbiota transplants (FMT) and by testing PAA on primary cultures of human hepatocytes and in mice.

**Cohort design and identification of clinical confounders**

To characterize the hepatic steatosis phenome, we established two unique and independent cohorts of women – negative for viral hepatitis – from Italy and Spain who elected for bariatric surgery. We focused on morbid obesity in non-diabetic women to examine liver steatosis variability. In particular, we excluded patients with known T2D to avoid the confounding influence of long-term hyperglycemia[22] or medications such as metformin[23,24] on the microbiome (see *Methods* for full inclusion and exclusion criteria). The degree of hepatic fat was defined according to the joint guidelines from the European Associations for the study of Liver, Obesity and Diabetes (EASL, EASO, EASD, see *Methods*)[25,26,27]. Given the impact of the microbiome on insulin resistance[6,28], we also performed oral glucose tolerance tests (OGTT) and euglycemic hyperinsulinemic clamps (EHC) (**Supplementary Table 1**). Clinical phenotypes were complemented by faecal metagenomics and molecular phenomics (plasma and urine metabolomes and liver transcriptomes) for association studies[29,30]. We then devised a data-driven hypothesis generation and validation strategy (**Fig. 1**). We first identified age, cohort and BMI as confounders, while all other clinical variables were mediators or had no effect on the generalized linear models (**Supplementary Fig. 1**, **Supplementary Table 2**). These three confounders were taken into account in all subsequent partial Spearman's rank-based correlation (pSRC) patterns across clinical variables.

**Metagenomic signatures of hepatic steatosis**

To obtain detailed taxonomic and functional information in hepatic steatosis for the faecal microbiome, we sequenced the patients' faecal metagenome and data were processed using our in-house pipeline, performing QC checks, filtering, and binning of reads into taxonomic kingdoms (**Supplementary Table 3**, **Supplementary Fig. 2**); metagenome assembly, gene prediction and clustering, functional annotation of gene clusters and comparison with the HMP Integrated Gene Catalog (IGC)[31] were performed. A total of 19,140,155 predicted genes were identified, which formed 3,902,787 gene

clusters. Of these genes, 2,320,286 mapped to the IGC, while 1,582,501 were novel (90 % query length, 95 % identity; **Supplementary Table 4**). We derived gene counts, *i.e.*, a measure of MGR, based on average values obtained from 30 samplings of 7 million randomly sampled reads that mapped to genes[8], resulting in a mean of 558,246 ± 154,249 genes across the samples (**Supplementary Table 5**), which is in the same order of magnitude as previous reports[19]. Remarkably MGR was significantly anti-correlated with hepatic steatosis (**Fig. 2a**; liver steatosis 0 665,063 ± 126,062 vs liver steatosis 3 517,989 ± 126,062 genes, $n$ = 10 patients both groups, $P$ = 0.03 Wilcoxon rank sum test) and with a number of markers of liver function, including γ-glutamyltransferase, alanine aminotransferase and inflammation (C-reactive protein) as well as with echography-assessed liver steatosis (**Fig. 2b**). Our data demonstrate for the first time the association of MGR with liver steatosis in a BMI-adjusted context, reinforcing previous observations for body weight and liver cirrhosis[19].

To determine whether specific microbes were responsible for this correlation, we assessed the abundance of prokaryotes within the metagenomes. Several taxa were significantly associated with liver steatosis and other related clinical parameters (**Fig. 2c–e**, **Supplementary Fig. 3a**, **Supplementary Fig. 4**, **Supplementary Table 6**): at the phylum level Proteobacteria, Actinobacteria and Verrucomicrobia were significantly correlated with liver steatosis, while Firmicutes and Euryarchaeota were significantly anti-correlated, whereas species diversity (calculated using the Chao1 estimator) was not correlated with liver steatosis (**Supplementary Fig. 3b**).

We next investigated associations between microbial function, by mapping our microbial gene catalog onto KEGG modules, and clinical phenotypes, thus revealing positive associations of hepatic steatosis with microbial carbohydrate, lipid and amino acid metabolism (**Supplementary Fig. 5**). These data suggest a change in microbial metabolism may contribute to liver health in morbidly obese women. Of particular relevance, LPS and peptidoglycan biosynthesis was significantly correlated with liver steatosis (**Supplementary Fig. 5**); this increase in LPS biosynthetic potential being consistent with an increased representation of Gram-negative Proteobacteria in steatosis, as observed in rodents[32]. These pathway-level analyses also highlight an increase in bacterial biosynthetic potential for fatty acids and sugars and various amino acids including BCAAs (Val, Leu and Ile) and AAAs (Trp,

Tyr and Phe) associated with steatosis and insulin resistance; this increase in BCAA biosynthesis further confirming previous reports in obesity and insulin resistance contexts[28,33].

**Impact of the microbiome on the hepatic steatosis phenome**

To reveal metabolic phenotypes possibly involved in steatosis related to the gut microbiome and liver steatosis, we performed metabolic profiling of urine and plasma by $^{1}$H-NMR spectroscopy. A metabolome-wide association study (MWAS)[30] resulted in 124 metabolite signals in urine and 80 in plasma correlated with hepatic steatosis and associated clinical traits (**Supplementary Fig. 6, 7**). Strikingly, the majority of liver steatosis-associated metabolites in plasma and urine were also associated with low MGR (**Fig. 3a, b**, **Supplementary Fig. 8**, **Supplementary Table 7**). Among the top liver steatosis metabolites (also associated with low MGR), we observed a significant correlation with BCAAs in plasma (leucine p-FDR = $4.69\times10^{-5}$; valine p-FDR = $1.72\times10^{-4}$; isoleucine p-FDR = $9.72\times10^{-5}$, **Fig. 3a**) and a significant increase in urine (leucine p-FDR = $6.1\times10^{-4}$; valine p-FDR = $1.73\times10^{-3}$; isoleucine p-FDR = 0.024, **Fig. 3b**) consistent with reports in obese patients[20,28]. Plasma choline and phosphocholine were not anti-correlated with liver steatosis for the 56 patients (but were anti-correlated in the larger cohort, *n*=102, **Supplementary Fig. 9**), whereas increased choline excretion was observed in liver steatosis (**Fig. 3b**), which is consistent with previous reports regarding choline bioavailability[6,34]. Remarkably, urinary hippurate was associated with MGR, echoing similar associations recently observed with Shannon diversity index obtained from 16S rRNA gene sequence profiling[35]. Among the microbial–mammalian co-metabolites significantly associated with steatosis and low MGR, plasma PAA (p-FDR = $4.69\times10^{-5}$) showed the strongest positive association (**Fig. 3a**). High MGR observed in non-steatotic patients was significantly correlated with a number of gut-derived microbial metabolites, such as urinary phenylacetylglutamine (p-FDR = $3.10\times10^{-9}$), plasma acetate (p-FDR = 0.009) and TMAO (p-FDR = 0.006) (**Supplementary Table 7**), a microbial-host co-metabolite playing a role in insulin resistance and atherosclerosis[6,14,15]. We further confirmed that TMAO, but not TMA, was marginally anti-correlated with steatosis by UPLC-MS/MS using isotopically-labelled standards[36,37], which is consistent with recent reports on the role of TMAO in metabolic homeostasis[15,38] (**Supplementary Table 8**). Altogether, these results suggest for the first time the existence of a metabolic phenotype associated with hepatic steatosis and low MGR, pinpointing

169 elevated BCAAs, AAAs and microbial metabolites coupled to a potential imbalance in hepatic
170 oxidation and conjugation of those microbial substrates.
171
172 To identify hepatic molecular mechanisms associated with the gut microbiome, we complemented our
173 phenome coverage by profiling hepatic (liver biopsy) transcriptomes from the same set of patients.
174 We identified 3,386 and 3,201 genes significantly positively and negatively correlated, respectively,
175 with liver steatosis (pSRC p-FDR < 0.05) (**Supplementary Table 9**). Furthermore, 3,581 human
176 genes significantly correlated (p-FDR < 0.05) with MGR: the pathways associated with the 1,776
177 genes significantly positively correlated with MGR largely matched those significantly anti-correlated
178 with steatosis (**Supplementary Fig. 10**). However, the 1,805 anti-correlated with MGR matched those
179 positively associated with hepatic steatosis (**Supplementary Fig. 10**), consistent with an anti-
180 correlation between liver steatosis and MGR (**Supplementary Table 9**, **Supplementary Table 10**).
181
182 To generate molecular hypotheses that could be useful for microbiota-related next-generation
183 therapeutic strategies we performed a hepatic signalling pathway impact analysis (SPIA, see
184 *Methods*) including the 2,277 genes intersecting the liver steatosis and MGR-associated genes. In
185 particular hepatic genes associate with non-specific pathways involved in the core immune response
186 to clearance of viral and bacterial (Proteobacteria, Gram-negative) infections (*i.e.*, viral
187 carcinogenesis; pathogenic *Escherichia coli* infection, shigellosis), alcoholism and insulin resistance
188 (**Fig. 3c**). Enrichment analyses (see *Methods*) of the hepatic genes significantly associated with MGR
189 further highlighted a significant (p-FDR < 0.2) over-representation of KEGG pathways associated with
190 the proteasome, phagosome, insulin resistance, glucagon signalling and non-specific responses to
191 microbial (Gram-negative, viral) infections (**Fig. 3d**). Among the overlapping genes co-associated with
192 hepatic steatosis and low MGR, *LPL* (lipoprotein lipase) was among the most correlated with hepatic
193 steatosis, while *ACADSB* (short/branched chain acyl-CoA dehydrogenase) and *INSR* (insulin
194 receptor) were the most anti-correlated (**Fig. 3e**), suggesting a molecular basis for the observation
195 that individuals with low MGR have a reduced capacity to respond to insulin exemplified by decreased
196 glucose disposal rate (during the EHC) and increased HOMA-IR (as shown in **Fig. 2b** and previously
197 reported in ref 28).
198

We further complemented our analyses of the hepatic transcriptome by assessing the topology of a directional network made of 2,277 genes significantly associated with liver steatosis and low MGR mapped onto KEGG pathways involved in liver disease, by aggregating all the KEGG networks with at latest one gene in common with the genes included in the NAFLD pathway. To analyse the topology of this resulting network, we computed shortest paths between the significant genes and derived the betweenness centrality metric[39,40], *i.e.*, the number of shortest paths passing through a particular gene product, to evaluate how central these genes are in the network. Betweenness centrality further highlights clusters of central genes channelling a high proportion of the shortest paths involving cAMP-related genes (*CREB3L4*, *PRKACA*, *CRTC2)*, innate immunity (Nuclear Factor Kappa B subunit 1, *NFKB1)* and *INSR* amongst others (**Fig. 3f**). Overall, hepatic gene expression is concordant with the metabolic signature obtained in plasma and urine showing elevated BCAAs concomitantly associated with low MGR, liver steatosis and insulin resistance, highlighting the interconnection among these three parameters. Genetic manipulation of *INSR* in the hepatocyte displayed a NAFLD phenotype[41,42] and the gut microbiome has recently been shown in rodents to interfere with *INSR* activation in the liver[43]. These results provide in humans a validation of numerous rodent-based hypotheses.

**Steatosis-associated microbiota and microbial metabolites modulate the steatosis phenome**

Our results document a strong contribution of the gut microbiome to the hepatic steatosis phenome. The increased microbial capacity for metabolism of BCAAs and metabolism of AAAs such as phenylalanine, tyrosine and tryptophan in liver steatosis (**Supplementary Fig. 5**) – phenylalanine metabolism resulting in PAA production – is supported by circulating metabolic markers (**Fig. 3a-b**), suggesting potentially causal mechanisms involving the microbiome in the steatosis phenome. In particular, our results strengthen the contribution of the gut microbiome to increased levels of circulating BCAAs in the host[28,33] – a metabolic phenotype gaining a central role in metabolic disorders[20]. This disruption of the gut–liver axis is further exemplified by the increase in inflammatory response, ER stress and phagosome pathways associated with a decrease in insulin signalling and small-molecule catabolic processes, conceivably altogether contributing to impaired BCAA and AAA metabolism as well as detoxification of liver steatosis-associated microbial compounds.

We then tested whether faecal microbial communities from donors with hepatic steatosis (steatosis grade 3) could trigger steatosis molecular mechanisms to recipient mice when compared with samples from donors with no hepatic steatosis (grade 0) (**Fig. 4a**). Donors with hepatic steatosis (n=3, steatosis grade 3) were randomly selected. Among subjects without hepatic steatosis (n=3, grade 0), we chose those that were similar in age, BMI, and fasting glucose to those with steatosis. For instance, fasting glucose was 87.3 ± 16.7 mg/dL in subjects without steatosis and 97.3 ± 6.4 mg/dL in the steatosis group ($P$ = 0.39). After a short antibiotic treatment and wash-out period and four consecutive daily faecal microbiota transplantations (FMTs), the recipient mice were fed a chow diet for 2 weeks. In the former group, this procedure resulted in a moderate but rapid accumulation of hepatic triglycerides (**Fig. 4b**). We also observed an increased *Fabp4* expression and plasma valine concentration compared with mice that received samples from patients without liver steatosis (**Supplementary Fig. 11a,b**), showing the general impact of the steatosis-associated microbiota from human donors on mouse liver lipid accumulation. By permutation testing seven-fold cross-validated O-PLS models using the donor human microbiome composition, we could successfully predict recipient mouse phenome responses, especially for steatosis, hepatic triglyceride content, *Fabp4* and plasma valine levels (**Fig. 4c** and **Supplementary Fig. 11c-e**, 1,000 random permutations, see *Methods*), highlighting the statistical robustness of the prediction between human donor microbiomes and recipient mouse phenome. We then derived significant associations between the donor microbiota composition and the mouse phenome, showing that the steatosis-associated microbiota influences multiple patterns of association with hepatic triglycerides, circulating BCAAs and TMAO (**Fig. 4d**). Similar, yet weaker associations were also observed between the mouse phenome and recipient mouse microbiota evaluated by faecal 16S rRNA gene amplicon analysis (**Supplementary Fig. 12**). The rapid hepatic lipid accumulation suggested a causal role of the human faecal microbiota in the triggering of hepatic steatosis which over a long-term period could lead to a highly significant liver lipid depot further contributing to hepatic insulin resistance.

To highlight the potential of novel microbial compounds to directly affect the hepatic steatosis phenome, we selected PAA due to the convergence of metagenomic and metabolomic observations: *i*) there is increased abundance of microbial gene pathways associated with its production in metagenomic sequences (**Supplementary Fig. 5**), and *ii*) it is the strongest microbial metabolite

associated with steatosis in our MWAS models (**Fig. 3a**). We compared its effects with the effects of palmitic acid, a free fatty acid known to trigger hepatic steatosis in human primary hepatocytes[44], using a full factorial design. We assessed lipid accumulation, expression of genes involved in steatosis as well as BCAA metabolism and consumption. PAA initiates molecular mechanisms leading to triglyceride accumulation in human primary hepatocytes in synergy with palmitic acid (**Fig. 5a-b**) and induces expression of lipid metabolism genes (*LPL* and *FASN*, **Fig. 5c-d**). PAA induced *INSR* expression contrary to palmitic acid and participated in the reduction of *GLUT2* expression (**Fig. 5e-f**). We next investigated AKT phosphorylation, which was significantly lowered by PAA, suggesting PAA reduces the response to insulin (**Fig. 5g**). PAA increased *ACADSB* expression (**Fig. 5h**) and resulted in an increased utilization of BCAA from the cell medium (**Fig. 5i-k**). We then treated mice with PAA for 2 weeks to confirm the increase in hepatic triglycerides and excreted isoleucine (**Fig. 5l-m**). These results suggest that PAA, as one of our top hepatic steatosis-associated microbial metabolites, significantly increases hepatic BCAA utilization and hepatic lipid accumulation.

**Integrative data crosstalk and steatosis signatures**

We finally quantified the crosstalk among gut microbiome, clinical phenotypes, liver transcriptome, urine and plasma metabolomes by estimating the proportion of shared variation amongst the different tables through *Rv* coefficients (**Fig. 6**, see *Methods*). A high proportion of information (79–97%) was shared between matching datasets (**Fig. 6a**, **Supplementary Table 11**), suggesting a strong similarity between metagenomic and phenomic data; the weakest (79.44%) being between urinary metabolome and clinical parameters. The metagenomic data shared 92–93% similarity with clinical parameters, liver transcriptome and plasma metabolome, while they only shared 74.68% with the urinary metabolome. This statistical crosstalk analysis suggests that, although metagenomic and phenomic data have strong similarity, there is still information attached to each original dataset which, if pooled together, could result in a robust signature.

We then built a multivariate model integrating metagenomic, transcriptomic and metabolomic information by fitting an orthogonal partial least squares discriminant analysis (O-PLS-DA) and tested its ability to correctly predict new samples during a seven-fold cross-validation through random permutation testing (**Fig. 6b**, 10,000 random permutations, $P = 0.0029$). We derived a bootstrapped

Receiver-Operator Characteristic (ROC) curve for the cross-validated models illustrating the ability of the model to correctly predict new samples (AUC=87%, **Fig. 6c, Supplementary Table 12**) of the binary prediction of steatosis (*i.e.*, steatosis vs. no steatosis) using cross-validated scores derived from seven-fold cross-validation of the O-PLS-DA model (see *Methods*), thereby confirming the joint predictive power of molecular phenomics and metagenomics. The predictive power of the phenome model is driven by the hepatic transcriptome (AUC 85%) that directly relates to the affected organ, but the excreted phenome and plasma metabolome both reach 73% and 79%, respectively. This AUC is particularly relevant as the non-invasive basal clinical data yielded 58%, which only increases through addition of more invasive metabolic challenges (OGTT and EHC, AUC 69%). Altogether, these predictive models based on molecular phenomics and metagenomics further support the idea that these molecular signatures used to generate hypotheses are robust and ultimately suggesting that the link tethering the microbiome to hepatic steatosis is robust too.

## DISCUSSION

In this study, we performed an in-depth clinical characterization of well-phenotyped non-diabetic obese women from Spain and Italy. We then reveal molecular networks between the gut microbiome and the hepatic steatosis phenome in this population of morbidly obese women, through computational integration of individual metagenomes, metabolomes and hepatic transcriptomes with histological steatosis scores. The robustness of our phenome signatures and the experimental follow-ups show that hepatic steatosis is negatively associated with MGR and the microbiome contributes to the steatosis phenome. The striking association between low MGR and hepatic steatosis is consistent with clinical and preclinical results confirming the role of the microbiome in rodent models[7] of non-alcoholic fatty liver disease and the role of MGR in metabolic disease[19,45].

We then functionally characterized an increased gut microbial amino-acid metabolism in steatotic subjects that has a profound impact on their liver transcriptome, biofluid metabolomes and liver fat accumulation, leading eventually to fatty liver. We found an anti-correlation pattern between steatosis and MGR was valid for the most significant steatosis-associated genes and metabolites, thereby suggesting that the reduction in MGR is a key factor that imbalances microbiome metabolic pathways leading to a steatosis-associated phenome, as observed for obesity[19,45]. From this tight crosstalk, we

319 further depict a coordinated disruption of the gut–liver axis in hepatic steatosis that manifests itself
320 across the faecal metagenome, hepatic transcriptome and biofluid metabolome. For instance, the
321 increased Proteobacteria frequency in hepatic steatosis is mirrored by an increase in microbial-
322 associated functional pathways related to endotoxin production and immune response in steatotic
323 patients – both at the hepatic and circulating levels. Our study further confirms the impact of LPS and
324 putatively other microbial products on liver lipid accumulation in humans[46], as previously proposed in
325 rodent models[32].
326
327 By integrating numerous biological measurements, our data analysis strategy implemented a detailed
328 functional analysis of the patient faecal metagenomes and molecular phenomes, offering novel
329 insights for the integrative physiology of hepatic steatosis. For instance, the increased microbial
330 potential for BCAA production, a phenomenon already reported for insulin resistance and obesity[28,33],
331 is mirrored by an increase in the BCAA pool in biofluids. Also, our bioinformatic analysis of
332 metagenomic sequences combined with metabolomic data suggested a direct role for microbial
333 degradation of AAAs into PAA in patients with steatosis. Our preclinical studies in rodents and primary
334 culture of human hepatocytes corroborated the role of this metabolite, amongst others, as an example
335 of a microbially-related metabolite involved in hepatic steatosis. By subsequently focussing on a
336 unique microbiome-associated feature such as PAA, which was selected through converging patterns
337 observed in microbial gene functions and biofluid metabolomes, we identified a novel mechanism by
338 which the microbiome facilitates steatosis, *via* increased BCAA utilisation and AAA metabolism.
339 Whilst acknowledging the complexity of the microbiome–host interplay, it should be noted that
340 although PAA is an exemplar metabolite highlighted in our human dataset, its effects are here limited
341 to triggering steatosis-associated molecular mechanisms and it is unlikely to be the sole player in
342 steatosis. The PAA effects are most likely part of a much broader, multifactorial process orchestrated
343 by the microbiome and involving many factors that warrant further studies.
344
345 The demonstration that the faecal microbiota obtained from patients with steatosis (grade 3, >66%)
346 initiated hepatic lipid accumulation and affected the phenome of recipient mice through FMTs
347 reinforces the causal role of the microbiota in steatosis. Not only did the human donor microbiota from
348 patients with steatosis trigger hepatic triglyceride accumulation in recipient mice, but it also affected

their circulating metabolome and hepatic transcriptome, through an increase in circulating valine levels and an increased expression of genes involved in lipid metabolism. Moreover, the characteristics of the donor microbiota predicts the extent of the phenomic response in the donor mice, which echoes recent reports on the prodromal role of the microbiota for metabolic response to diet in animal models[15,47] and humans[48]. The successful replication of the steatotic phenotype using human donor material for FMT in mice represents a key translational link between metagenomic studies in patients with NAFLD, hepatic fibrosis or cirrhosis[8,49,50] and previous FMT studies that had only been established for NAFLD with mouse donors[7].

Altogether, we propose a model in which the microbiome orchestrates three possibly complementary contributions to hepatic steatosis in obesity: i) reduced MGR – indicative of deleterious changes in microbiome functions – can trigger steatosis and increase the BCAA pool; ii) microbially-produced PAA and possibly other related metabolites facilitate hepatic lipid accumulation *via* a synergetic increase in BCAA utilization in the TCA cycle; and iii) microbially-associated factors such as LPS induce inflammation in hepatocytes.

Similar to Qin *et al.* (2014) who studied the faecal metagenome of liver cirrhosis patients[8], our data indicate a slight shift of the faecal microbiome in patients with steatosis to one more similar to that found in the human small intestine and oral cavity. For example, patients with steatosis had fewer *Lachnospiraceae* and *Ruminococcaceae* responsible for butyrate production and were enriched in *Acidaminococcus* and *Escherichia* spp. *Bacteroides* spp. were associated with insulin resistance, concordant with observations from Pedersen *et al.*[28], who showed *Bacteroides vulgatus* was one of the main species contributing to insulin resistance, and circulating levels of BCAAs in humans[28].

In conclusion, this work offers a unique clinical resource and integrated analysis of metagenomics with molecular phenomics of hepatic steatosis in non-diabetic obese women coupled with experimental validations in cellular and animal models. Not only does our work further validate previous studies in humans[49], but it also confirms hypotheses formulated in rodent models, such as the role of LPS, in which the gut microbiome was shown to influence gene pathways involved in the immune system and metabolic disorders (*i.e.*, inflammation impacting host metabolism[7,32,51]).

Ultimately, this integrated database and modelling approach also suggests new potentially causal mechanisms in hepatic steatosis involving BCAA- and AAA-derived metabolites. Our investigations further support the view that the molecular crosstalk between the microbiome and its human host is of utmost importance for patient health and highlights the need for integrative analyses of metagenomes and broad-sense phenomes[52,53]. Our study establishes a comprehensive understanding of the microbial factors affecting human metabolic disease states for precision medicine, thereby laying the groundwork for targeted FMT therapies and pharmacotherapies to promote hepatic metabolic homeostasis.

**Online Content**

Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Acknowledgements**

We thank Professor Nigel J. Gooderham for critical reading of the manuscript. This work was supported by EU-FP7 FLORINASH (Health-F2-2009-241913) to RB, MF, JMFR, FF, CP, EH and JKN. This work used the computing resources of the UK MEDical BIOinformatics partnership – aggregation, integration, visualisation and analysis of large, complex data (UK MED-BIO), which is supported by the Medical Research Council (grant number MR/L01632X/1). LH is in receipt of an MRC Intermediate Research Fellowship in Data Science (MR/L01632X/1, UK MED-BIO). This work was also supported by funding to RB (Region 2009-2014; Agence Nationale de la Recherche ANR Bactimmunodia & GAD), to MF (MIUR PRIN 2015MPESJS_004, RF-2011-02349921 and Fondazione Roma Non-Communicable Diseases 2014), to MED (EU METACARDIS under agreement HEALTH-F4-2012-305312, Neuron II under agreement 291840, and the MRC MR/ M501797/1) and to JMFR (FIS project 15/01934, CIBERobn Pathophysiology of Obesity and Nutrition and FEDER funds).

**Author Contributions.** RB, JMFR, MF, FL, FF, CP, EH and JKN designed the study and supervised all parts of the project. RB is the project leader and chaired the consortium. MED led data integration and elaborated the primary interpretation of analytical outcomes with LH, in close collaboration with

MF, JMFR and RB. LH implemented the microarray data analysis workflow, CT and MW developed the data repository. JA developed the metagenomic data analysis pipeline in collaboration with LH, and SAB supervised the development of the data repository and the pipeline. LH, JA, RHB, and MED performed data analyses. JMFR and MF designed the clinical protocol and oversaw the clinical activities. MC, FD, IC, OP, PG, JP, GX and WR recruited and phenotyped patients, collected biological samples and physiological data. MS, VA and VB performed RNA and DNA extractions, RB and MS supervised DNA sequencing and gene profiling. JLL and JMMN performed cell culture experiments, FR, EA, JC and CH performed animal work. RHB, JC and LMG performed metabolic profiling of plasma and urine by $^{1}$H-NMR, EH and JKN supervised metabolic profiling. AM performed methylamine quantifications. LH and MED drafted the first versions of the paper with critical and substantial contributions from MF, JMFR and RB. All authors provided support and constructive criticism throughout the project and approved the final version.

**FIGURE LEGENDS**

**Figure 1. Flowchart showing approach used for the integration of clinical, molecular phenomics and metagenomic information and biological validations. *a*,** Confounder and modifier analysis performed on the FLORINASH clinical markers identified three confounders: age, BMI and country (*n* = 105). Subsequent analyses were performed using partial Spearman rank-based correlation (pSRC) coefficients adjusted for age, BMI and country. ***b*,** Metagenome-wide and phenome-wide association of taxonomic abundance data with clinical markers (*n* = 56 patients). ***c*,** Network analysis of hepatic transcriptome (*n* = 56 patients). ***d*,** Metabonome-Wide Association Study based on plasma (*n* = 56) and urine (*n* = 56) $^{1}$H-NMR spectra. ***e*,** Integrative comparison analysis using *Rv* coefficients (*n* = 56). ***f*,** Predictive performance of an O-PLS-DA model integrating all metagenomic and phenomic modalities for prediction of non-alcoholic fatty liver (no hepatic steatosis, score = 0, *n* = 10 vs. steatosis, score > 0, *n* = 46) in ROC curves.

**Figure 2. Association between liver steatosis, microbial gene richness (MGR) and metagenomic data in obese women. *a***, MGR was significantly anti-correlated with liver steatosis. ***b***, Correlation of MGR with clinical data (p-FDR values shown). ***c***, Association of genus-level abundance data with clinical data. +, p-FDR < 0.05. ***d,*** Prokaryotic taxa significantly (p-FDR < 0.05) anti-correlated with liver steatosis at the phylum and genus levels. ***e*,** Prokaryotic taxa significantly (p-FDR < 0.05) correlated with liver steatosis at the phylum and genus levels. (No liver steatosis = 10; liver steatosis 1 = 22; liver steatosis 2 = 14; liver steatosis 3 = 10 for all panels.)

**Figure 3**. **Association of metabolomic and transcriptomic data with liver steatosis and microbial gene richness (MGR). *a***, Plasma metabolites most significantly (p-FDR < 0.05) partially correlated with liver steatosis. ***b***, Urinary metabolites most significantly (p-FDR < 0.05) partially correlated with liver steatosis. ***c*,** SPIA evidence plot for the intersection of the 2,277 genes significantly associated with liver steatosis and MGR. Each signaling pathway is represented by one dot. The pathways at the right of the red oblique line are significant (< 0.2) after Bonferroni correction of the global *P* values, pG, obtained by combining the pPERT and pNDE using the normal inversion method. The pathways at the right of the blue oblique line are significant (< 0.2) after a FDR correction of the global *P* values, pG. The yellow node represents the KEGG pathway 'Non-alcoholic

fatty liver disease (NAFLD) – *Homo sapiens* (human)'; 05222, Small cell lung cancer; 4914, Progesterone-mediated oocyte maturation. ***d***, Enrichr (KEGG pathway) analysis of the hepatic genes significantly (p-FDR < 0.05) correlated and anti-correlated with MGR. ***e***, The ten hepatic genes most significantly (p-FDR < 0.05) correlated and anti-correlated with liver steatosis. ***f***, Network analysis of the 2,277 hepatic steatosis – MGR intersecting genes. The correlation values for liver steatosis were used to generate the network: the bluer a node, the more significantly anti-correlated liver steatosis is with the hepatic gene; the redder a node, the more significantly correlated liver steatosis is with the hepatic gene. Analysis of betweenness centrality[39,40] showed *CREB3L4*, *PRKACA*, *CRTC2*, *OGT*, *INSR*, *NFKB1*, *PPP1CA*, *IKBKG*, *MAP3K7*, *MAPK9*, *ITGAV*, *RRAS2*, *RPS6KA2*, *PHKA1*, *PHKB*, *BRAF*, *ALDOC*, *PFKL*, *EFNA1*, *FGF12*, *ANGPT4*, *PDGFB*, *VEGFB*, *FGFR4*, *MAP2K2*, *TAPBP*, *ALDH3A2*, *ALDH7A1*, *GPI and GNAI3* to be (from highest betweenness centrality to lowest) the 30 genes having most control over the network. Genes with no network connections have been removed for clarity. (No liver steatosis = 10; liver steatosis 1 = 22; liver steatosis 2 = 14; liver steatosis 3 = 10 for all panels.)

**Figure 4. Transfer of steatotic and metabolic phenotypes to mice through FMT of material from patients with liver steatosis grade 3.** ***a***, FMT protocol. ***b***, Hepatic triglycerides in recipient mice. ***c***, Permutation tests for goodness of fit ($R^2$) and prediction ($Q^2$) parameters obtained from a seven-fold cross-validated O-PLS regression model quantitatively predicting recipient mouse hepatic lipid accumulation from human donor microbiome composition. **d**, Association between recipient mouse phenome and human donor microbiota (n = 44). The quality of a given O-PLS model is usually assessed by goodness-of-fit ($R^2$) and goodness-of-prediction ($Q^2$). The $R^2$ parameter corresponds to the explained variance of the model whilst the $Q^2$ parameter corresponds to the predicted variance, as assessed by seven-fold cross-validation of the given model. The significance of the $R^2$ and $Q^2$ parameters is then assessed by using 10,000 random permutations of the class membership variable. The horizontal axis corresponds to the correlation between the original class membership (on the right) and the permuted class membership (10,000 permutations on the left of the plot). The vertical axis corresponds to the $R^2$ (green dots) and $Q^2$ (blue dots) coefficients. The green ($R^2$) and blue ($Q^2$) lines are both increasing form left to right suggesting that the original $R^2$ and $Q^2$ parameters on the right are significantly different from both populations of $R^2$ and $Q^2$ parameters obtained from models

fitted with random permutations of the class membership on the left. Data obtained from FMT protocols performed independently with faecal material from three patients with liver steatosis (grade 3, >66% steatosis) and three control patients (grade 0, <5% steatosis), *n* = 8 recipient mice per donor. Data are mean ± s.e.m., * $p < 0.05$.

**Figure 5. Microbial PAA induces liver steatosis and BCAA use in primary human hepatocytes and mice.** ***a***, Micrographs of primary human hepatocytes stained with Oil Red O. ***b***, Quantification of lipid accumulation. ***c***, *LPL* expression in hepatocytes. ***d***, *FASN* expression in hepatocytes. ***e***, *INSR* expression in hepatocytes. ***f***, *GLUT2* expression in hepatocytes. ***g***, AKT phosphorylation in hepatocytes. ***h***, *ACADSB* expression in hepatocytes. ***i***, Valine in hepatocyte cell medium. ***j***, Leucine in hepatocyte cell medium. ***k***, Isoleucine in hepatocyte cell medium. ***l***, Hepatic triglycerides in PAA-treated mice. ***m***, Isoleucine in urine from PAA-treated mice (*n* = 8-10 per group). Data obtained with *n* = 4 observations per group unless stated otherwise. Data are mean ± s.e.m., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Abbreviations: CTRL, control group; PAA, phenylacetic acid treatment group; PA, palmitic acid treatment group; PA+PAA, palmitic acid and phenylacetic acid treatment group.

**Figure 6. Phenome-wide crosstalk and predictive modelling.** ***a***, Metagenome–phenome matrix correlation network computed for the patients with matching metagenomic and phenomic profiles (*n* = 56) using the modified *Rv* correlation matrix coefficient. Each phenomic table corresponds to a node and edges represent the relationships between tables, *i.e.*, the per cent of shared information, derived from the $Rv^2$ matrix correlation coefficient corresponding to the proportion of variance shared by the two tables – which like a squared Pearson's correlation coefficient ($r^2$) – corresponds to the proportion of explained variance between two variables. ***b***, Discriminative power of a supervised multivariate model (OPLS-DA) fitted with patients with matching metagenomic and phenomic profiles (*n* = 56) to predict new samples, using random permutation testing (10,000 iterations). ***c***, Performance of classification of liver steatosis status (*n* = 10, vs. others, *n* = 46) based on matching molecular phenomic and gut metagenomic profiles. A ROC curve was obtained for the cross-validated model predictions derived from the O-PLS-DA model, reaching an AUC of 87.07%, corresponding to the successful prediction rate. Groups for all panels are: no steatosis (grade 0), *n* = 10; steatosis (grades 1-3), *n* = 46.

## REFERENCES


1. Saltiel, A. R. & Kahn, C. R. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **414,** 799–806 (2001).
2. Kahn, S. E., Hull, R. L. & Utzschneider, K. M. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* **444,** 840–846 (2006).
3. Meex, R. C. R. & Watt, M. J. Hepatokines: linking nonalcoholic fatty liver disease and insulin resistance. *Nat Rev Endocrinol* **13,** 509–520 (2017).
4. Petersen, M. C., Vatner, D. F. & Shulman, G. I. Regulation of hepatic glucose metabolism in health and disease. *Nat Rev Endocrinol* **13,** 572–587 (2017).
5. Adams, L. A., Anstee, Q. M., Tilg, H. & Targher, G. Non-alcoholic fatty liver disease and its relationship with cardiovascular disease and other extrahepatic diseases. *Gut* **66,** 1138–1153 (2017).
6. Dumas, M.-E. *et al.* Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 12511–12516 (2006).
7. Le Roy, T. *et al.* Intestinal microbiota determines development of non-alcoholic fatty liver disease in mice. *Gut* (2012). doi:10.1136/gutjnl-2012-303816
8. Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513,** 59–64 (2014).
9. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
10. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).
11. Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341,** 1241214 (2013).
12. Craciun, S. & Balskus, E. P. Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 21307–21312 (2012).
13. Hoyles, L. *et al.* Metabolic retroconversion of trimethylamine N-oxide and the gut microbiota. *bioRxiv* 225581 (2017). doi:10.1101/225581
14. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472,** 57–63 (2011).
15. Dumas, M.-E. *et al.* Microbial-Host Co-metabolites Are Prodromal Markers Predicting Phenotypic Heterogeneity in Behavior, Obesity, and Impaired Glucose Tolerance. *Cell Rep* **20,** 136–148 (2017).
16. Lonardo, A. *et al.* Fatty liver is associated with an increased risk of diabetes and cardiovascular disease - Evidence from three different disease models: NAFLD, HCV and HIV. *World J. Gastroenterol.* **22,** 9674–9693 (2016).
17. Houle, D., Govindaraju, D. R. & Omholt, S. Phenomics: the next challenge. *Nat. Rev. Genet.* **11,** 855–866 (2010).
18. Dumas, M.-E., Kinross, J. & Nicholson, J. K. Metabolic phenotyping and systems biology approaches to understanding metabolic syndrome and Fatty liver disease. *Gastroenterology* **146,** 46–62 (2014).
19. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500,** 541–546 (2013).
20. Newgard, C. B. *et al.* A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab.* **9,** 311–326 (2009).
21. Jang, C. *et al.* A branched-chain amino acid metabolite drives vascular fatty acid transport and causes insulin resistance. *Nat. Med.* **22,** 421–426 (2016).
22. Karlsson, F. H. *et al.* Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498,** 99–103 (2013).
23. Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528,** 262–266 (2015).
24. Wu, H. *et al.* Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug. *Nat. Med.* **23,** 850–858 (2017).
25. European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD)European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease.

*Diabetologia* **59,** 1121–1140 (2016).
26. European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD)European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J. Hepatol.* **64,** 1388–1402 (2016).
27. Kleiner, D. E. *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41,** 1313–1321 (2005).
28. Pedersen, H. K. *et al.* Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **535,** 376–381 (2016).
29. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490,** 55–60 (2012).
30. Holmes, E. *et al.* Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **453,** 396–400 (2008).
31. Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32,** 834–841 (2014).
32. Cani, P. D. *et al.* Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* **56,** 1761–1772 (2007).
33. Shoaie, S. *et al.* Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome. *Cell Metab.* **22,** 320–331 (2015).
34. Spencer, M. D. *et al.* Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. *Gastroenterology* **140,** 976–986 (2011).
35. Pallister, T. *et al.* Hippurate as a metabolomic marker of gut microbiome diversity: Modulation by diet and relationship to metabolic syndrome. *Scientific Reports* **7,** 13670 (2017).
36. Kadar, H. *et al.* A multiplexed targeted assay for high-throughput quantitative analysis of serum methylamines by ultra performance liquid chromatography coupled to high resolution mass spectrometry. *Arch. Biochem. Biophys.* **597,** 12–20 (2016).
37. Plovier, H. *et al.* A purified membrane protein from Akkermansia muciniphila or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat. Med.* **23,** 107–113 (2017).
38. Schugar, R. C. *et al.* The TMAO-Producing Enzyme Flavin-Containing Monooxygenase 3 Regulates Obesity and the Beiging of White Adipose Tissue. *Cell Rep* **19,** 2451–2461 (2017).
39. Davidovic, L. *et al.* A metabolomic and systems biology perspective on the brain of the fragile X syndrome mouse model. *Genome Res.* **21,** 2190–2202 (2011).
40. Rodriguez-Martinez, A. *et al.* MetaboSignal: a network-based approach for topological analysis of metabotype regulation via metabolic and signaling pathways. *Bioinformatics* **33,** 773–775 (2017).
41. Biddinger, S. B. *et al.* Hepatic insulin resistance directly promotes formation of cholesterol gallstones. *Nat. Med.* **14,** 778–782 (2008).
42. Michael, M. D. *et al.* Loss of insulin signaling in hepatocytes leads to severe insulin resistance and progressive hepatic dysfunction. *Mol. Cell* **6,** 87–97 (2000).
43. Fujisaka, S. *et al.* Antibiotic effects on gut microbiota and metabolism are host dependent. *J. Clin. Invest.* **126,** (2016).
44. Latorre, J. *et al.* Decreased lipid metabolism but increased FA biosynthesis are coupled with changes in liver microRNAs in obese subjects with NAFLD. *Int J Obes (Lond)* **41,** 620–630 (2017).
45. Cotillard, A. *et al.* Dietary intervention impact on gut microbial gene richness. *Nature* **500,** 585–588 (2013).
46. Sharifnia, T. *et al.* Hepatic TLR4 signaling in obese NAFLD. *Am. J. Physiol. Gastrointest. Liver Physiol.* **309,** G270–8 (2015).
47. Thaiss, C. A. *et al.* Persistent microbiome alterations modulate the rate of post-dieting weight regain. *Nature* **540,** 540–551 (2016).
48. Zeevi, D. *et al.* Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **163,** 1079–1094 (2015).
49. Boursier, J. *et al.* The severity of nonalcoholic fatty liver disease is associated with gut dysbiosis and shift in the metabolic function of the gut microbiota. *Hepatology* **63,** 764–775 (2016).
50. Loomba, R. *et al.* Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection of Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease. *Cell Metab.* **25,** 1054–1062.e5 (2017).
51. Serino, M. *et al.* Metabolic adaptation to a high-fat diet is associated with a change in the gut

microbiota. *Gut* **61,** 543–553 (2012).
52. Claesson, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488,** 178–184 (2012).
53. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352,** 560–564 (2016).

## SUPPLEMENTARY METHODS

**Primary human hepatocyte culture and treatments.** Cryopreserved primary human hepatocytes (HH) were obtained commercially (Innoprot, Bizkaia, Spain) and cultured with hepatocytes medium (Innoprot) supplemented with 5% fetal bovine serum, 1% hepatocytes growth supplement (mixture of growth factors, hormones and proteins necessary for culture of primary hepatocytes), and 100 U/mL penicillin and streptomycin. HH were grown on poly-L-lysine pre-coated cell dishes at 37 °C and 5% $CO_2$ atmosphere following manufacturer's recommendations. Cells were treated 24 h after seeding with phenylacetic acid (PAA), palmitic acid (PA) or a combination of both. Compounds were prepared as follows: 136.16 mg of PAA (P16621, Sigma, San Luis, MO) were dissolved in 10 mL of phosphate-buffered saline (PBS) and 27.84 mg of PA (Sigma, San Luis, MO) in 1 mL sterile water to obtain both components at 100 mM stock solutions. Bovine serum albumin (BSA, 5 %) was prepared in serum-free DMEM and then mixed with PA stock solution for at least 1 h at 40 ºC to obtain a 5 mM solution. HH were treated with PAA 10 mM, PA 200 µM or a combination of both for 24 h. BSA and PBS were used as vehicles. All experimental conditions were performed in four biological replicates.

After treatment, cells were washed with PBS and collected with Qiazol for RNA purification. Total RNA was extracted and purified using RNeasy Mini Kit (QIAgen, Gaithersburg, MD) following the manufacturer's protocol. Gene expression procedures were assessed using LightCycler 480 Real-Time PCR System (Roche Diagnostics SL, Barcelona, Spain), using TaqMan technology suitable for relative genetic expression quantification. Fatty acid (FA) accumulation was tested with Oil Red O staining. Briefly, after treatment cells were washed twice with PBS, fixed with paraformaldehyde 7% for 1 h and dipped in isopropanol 60% before staining with Oil Red O (Sigma, Lyon, France) for 10 min at room temperature. Pictures were taken with an inverted microscope. For quantification, 100% isopropanol was added to elute Oil Red O and optical density was monitored spectrophotometrically at 500 nm (Cytation5, Biotek). Finally, insulin resistance analysis was performed as follows: HH were maintained in starvation for 1 h after treatment. Insulin (100 nM in serum-free DMEM medium) was used for stimulation of insulin pathway for 10 min. Then, cells were collected and homogenized in 50 $\mu$L of lysis buffer (Cell Signaling Technology, Barcelona, Spain) and cell debris was discarded by centrifugation (10 min, 15,000 r.p.m. at 4 ºC). Protein amount was determined using the Lowry assay

(Biorad, Madrid, Spain). Protein extracts were separated by SDS-PAGE and transferred to nitrocellulose membrane by conventional procedures. Membranes were immunoblotted with phosphorylated (Ser473) Akt serine/threonine kinase (pAkt) and total Akt (Cell Signaling Technology). Anti-rabbit IgG coupled to horseradish peroxidise was used as secondary antibody. Horseradish peroxidase activity was detected by chemiluminescence and quantification of protein expression was performed using Scion image software. Cell-based assays were not performed in a blind manner.

**PAA treatment in mice.** Procedures were carried out according to the French guidelines for the care and use of experimental animals (Animal authorization agreement n° CEEA34.AFB/CP.082.12, validated by the University Paris Descartes Ethical Committee). 10-week old C57BL/6J mice (Janvier labs) were used for *in vivo* PAA treatments. Mice were maintained in a 12 h light/dark cycle with water and standard diet (65% carbohydrate, 11% fat, and 24% protein) supplemented or not with 0.8% of PAA (Sigma) for 2 weeks. The mouse experiments were not performed in a blind manner. Experimental groups (*n* = 8-9 per group) were randomly allocated.

**Faecal microbiota transplantation in mice.** All animal experimental procedures were approved by the local ethical committee (approval number 31-278) of Rangueil University Hospital (Toulouse, France). Faecal microbiota and faecal water transplantation were performed as previously described[54,55] so that 20 mg per day per mouse (C57BL6 male, 8 week old, Charles River) were administered for four consecutive days. Briefly, six faecal samples from subjects with (n=3) and without (n=3) hepatic steatosis matched for age and BMI (*see* **Supplementary Table 1**) were suspended separately in sterile reduced PBS ($N_2$ gas and thioglycolic acid, Sigma Aldrich, St. Louis, MO). The faecal matter was used to treat 8-week-old mice. First, eight mice per patient were treated for 7 days with an antibiotic mixture (neomycin, ampicillin, metronidazole as described[54,55]). A 4-day wash-out period ensured elimination of the antibiotics. The mice were then gavaged once-a-day for four consecutive days with the faecal matter suspended in the buffer. Two weeks later the mice were sacrificed; livers and plasma were collected and frozen before assay. The mouse experiments were not performed in a blind manner. Experimental groups were randomly allocated.

**Western blot analysis (mouse and primary hepatocytes).** Proteins from liver tissue were extracted from whole cell lysates. Proteins were subjected to 10% SDS-PAGE gels and electroblotted to nitrocellulose membranes. Rabbit polyclonal antibodies of Akt (Cell Signaling, 9272), pAkt (ser473, Cell Signaling, 4060) and pAkt (thr308, Cell Signaling, 3038) were used at 1:1000.

**Recruitment of patients and processing of samples.** All subjects gave written informed consent, validated and approved by the ethical committee of the Hospital Universitari Dr Josep Trueta (Comitè d'Ètica d'Investigació Clínica, approval number 2009 046) and Policlinico Tor Vergata University of Rome (Comitato Etico Indipendente, approval number 28-05-2009). The human subject cohort comprised 105 morbidly obese women at the Endocrinology Service of the Hospital Universitari de Girona Dr Josep Trueta (Girona, Spain, *n* = 44) and at the Center for Atherosclerosis of Policlinico Tor Vergata University of Rome (Rome, Italy, *n* = 61). Sample size was not determined by statistical methods.

Inclusion criteria: Pre-established inclusion criteria were: all subjects were of Caucasian origin; the subjects reported a stable body weight 3 months preceding the study, and were not given a liquid diet before surgery, were free of any infections, including use of antibiotics, 1 month before surgery and had no systemic disease.

Exclusion criteria: Pre-established exclusion criteria were: subjects with known medical history of diabetes or self-reported use of hypoglycemic agents, presence of liver disease, specifically HCV infection and tumor disease, and subjects with thyroid dysfunction were excluded by biochemical work-up. Alcohol consumption >20 g/day was an exclusion criterion. Hepatitis B was routinely excluded before the surgical procedure (anti-HB virus antibodies), iron overload: serum ferritin was below 200 ng/mL in all subjects, autoimmune hepatitis was excluded by histology and exclusion of viral hepatitis, alpha-1 antitrypsin deficiency was excluded by anamnestic data and clinical evidence, drug-induced liver injury was excluded using a drug questionnaire.

722 Stool and biofluid samples from all of the subjects were obtained during the week before elective
723 gastric bypass surgery, during which the liver biopsy was sampled. Liver samples were collected in
724 RNA*later*, fragmented and immediately flash-frozen in liquid nitrogen before storage at −80 °C.
725
726 **Histology on liver biopsies (human).** The investigators were blind to group allocations. A
727 pathologist and a radiologist in each center assigned groups independently. Liver biopsies were
728 analysed by a single expert pathologist. The liver samples were stained with hematoxylin and eosin,
729 Masson's trichrome and reticulin. Excessive hepatic fat accumulation, associated with insulin
730 resistance, is defined by the presence of liver steatosis in >5% of hepatocytes according to
731 histological analysis[21,22]
732
733 **Clinical biochemistry (human).** Plasma glucose concentrations were measured in duplicate by the
734 glucose oxidase method using a Beckman glucose analyser II (Beckman Instruments, Brea,
735 California). Duplicate samples were used for plasma insulin determination by the immunoradiometric
736 assay (Medgenix Diagnostics, Fleunes, Belgium). The coefficients of variation (intra-assay) were 5.2
737 % at a concentration of 10 mU/Land 3.4 % at 130 mU/L. The coefficients of variation (inter-assay)
738 were 6.9 % and 4.5 % at 14 and 89 mU/L, respectively. Total plasma cholesterol was measured by an
739 enzymatic, colorimetric method through the cholesterol esterase/cholesterol oxidase/peroxidase
740 reaction (Cobas CHOL2). HDL cholesterol was quantified by a homogeneous enzymatic colorimetric
741 assay through the cholesterol esterase/cholesterol oxidase/peroxidase reaction (Cobas HDLC3).
742 Total plasma triglycerides were measured by an enzymatic, colorimetric method with glycerol
743 phosphate oxidase and peroxidase (Cobas TRIGL). LDL cholesterol was calculated using the
744 Friedewald formula. Cortisol was determined by routine laboratory test[56].
745
746 **Euglycemic hyperinsulinemic clamp (human).** Insulin action was determined by the euglycemic
747 hyperinsulinemic clamp (EHC). After an overnight fast, two catheters were inserted into an antecubital
748 vein, one for each arm, used to administer constant infusions of glucose and insulin, and to obtain
749 arterialized venous blood samples. A 2-h EHC was initiated by a two-step primed infusion of insulin
750 (80 mU/m$^2$/min for 5 min, 60 mU/m$^2$/min for 5 min) immediately followed by a continuous infusion of
751 insulin at a rate of 40 mU/m$^2$/min (regular insulin; Actrapid, Novo Nordisk, NJ). Glucose infusion

752 began at minute 4 at an initial perfusion rate of 2 mg/kg/min, then was adjusted to maintain plasma
753 glucose concentration at 4.9–5.5 mmol/L. Blood samples were collected every 5 min for determination
754 of plasma glucose and insulin. Insulin sensitivity was assessed as the mean glucose infusion rate
755 during the last 40 min. In the stationary equilibrium, the amount of glucose administered (M) equals
756 the glucose taken by the body tissues and is a measure of overall insulin sensitivity. A 75-g oral
757 glucose-tolerance test (OGTT) in accordance with American Diabetes Association criteria was also
758 performed[57].
759
760 **Metabolic profiling and phenotyping by $^{1}$H-NMR spectroscopy.** All $^{1}$H-NMR spectra were
761 acquired using Bruker DRX600 spectrometers (Rheinstetten, Germany) running under TopSpin, with
762 either a 5 mm TXI probe operating at 600.13 MHz or a 5mm BBI probe operating at 600.44 MHz. All
763 runs were carried out using Bruker BACS60 sample handling automation; prior to each run the 90°
764 pulse length was determined and set for the run. The field frequency was locked on $D_2O$ as solvent.
765 In all experiments, water suppression was carried out by noise irradiation during the 2 s recycle delay
766 (RD). For all experiments, 128 scans were recorded into 32K data points with a spectral width of 20
767 ppm, and an exponential function was applied to the FID prior to the Fourier transformation, which
768 resulted in a line broadening of 0.3 Hz. All urine and plasma NMR spectra were automatically phased,
769 baseline-corrected and referenced either to trimethylsilylpropionate TSP (δ 0.0) for urine, or the center
770 of the α-glucose anomeric doublet (δ 5.23) for plasma, using in-house MatLab (The MathWorks,
771 Natick, Massachusetts) scripts. Baseline and peak alignment quality control was done by individual
772 verification for each spectrum and occasionally a spectrum was manually adjusted. Spectral line-
773 shape quality was also individually assessed, and occasionally spectra were re-acquired during the
774 same sample run. The spectra were all then imported to Matlab and the region around the water
775 resonance (δ 4.7–4.9 ppm for urine and δ 4.5–5.0 ppm for plasma) was zeroed. The NMR data arrays
776 then underwent spectral median fold-change normalization[58] using a probabilistic quotient
777 normalization (PQN) algorithm, performed with in-house scripts.
778
779 Urine samples. Urines were thawed at room temperature from frozen storage at -80 °C and briefly
780 centrifuged to allow clean supernatant aliquoting into a 5 mm NMR tube. A high $D_2O$ (80:20) buffer
781 was operationally prepared by weighing 5.77 g of $Na_2HPO_4$, 1.05 g $NaH_2PO_4$, 33.65 mg TSP and 80

782 mg $NaN_3$ into a flask, with the addition of 180 mL of $D_2O$ and 20 mL $H_2O$ to make approximately 200
783 mL of buffer. Urine samples were prepared by adding 150 µL of phosphate buffer to 350 µL of urine in
784 5 mm NMR tubes, and the mixture was then briefly vortexed. The primary data acquisitions were
785 made using the standard 1-D pulse program *noesypr1d*, [Recycle delay (RD)-90°-$t_1$-90°-$t_m$-90°-
786 acquire free induction decay (FID)][6].
787
788 Plasma samples. Plasma samples were primarily stored at -80 °C in heparinized lithium tubes, though
789 a few early samples were in EDTA tubes. A 0.9 % (w/v) NaCl solution was prepared with 80 %:20 %
790 (v/v) $H_2O$:$D_2O$, with 200 mg/L added $NaN_3$ to inhibit microbial activity. After thawing plasma at room
791 temperature, 350 µL aliquots were carefully extracted by micropipette to avoid any coagulates and
792 placed in a 5 mm NMR tube, with 150 µL of isotonic 0.9 % saline solution "extender" then being
793 added and gently vortexed briefly to make a final volume of 500 µL. $^1$H-NMR spectra of the plasma
794 samples were acquired employing two 1-D NMR experiments. Acquisitions were made using a
795 standard 1-D pulse program, *noesypr1d,* [Recycle delay (RD)-90°-$t_1$-90°-$t_m$-90°- acquire free induction
796 decay (FID)], and also a Carr-Purcell-Meiboom-Gill (CPMG) [RD-90°-(τ-180°-τ) n-acquire FID] using
797 the pulse program *cpmgpr*, where *n* = 100, the number of spin echoes and t= 400 µs, the CPMG
798 delay time), yielding a 2 nτ spin-echo cycle for a total of 80 ns. The CPMG data were those used for
799 all subsequent metabolic modeling of plasma, due to the useful partial suppression by CPMG of
800 intensity from the ultra-broad lipoprotein signals present[6].
801
802 **Plasma methylamine quantification by UPLC-MS/MS.** Methylamines were quantified as previously
803 described[36,37]. Plasma samples (10 uL) were spiked with 10 µL isotopically labelled Internal
804 Standards (IS) ($^{13}C_3$/$^{15}$N-TMA, $d_9$-TMAO and $d_9$-choline in water; 1 mg/L, Sigma-Aldrich). TMA was
805 derivatized to its ethoxy- analogue with the addition of 45 µL of derivatization solution (15g/L ethyl 2-
806 bromoacetate, 1% $NH_4OH$ in 1:1 acetonitrile/water). The reaction was completed after 30 min at room
807 temperature. Protein/lipid precipitation solution (935 µL; 94% acetonitrile/5% water/1% formic acid)
808 was added; samples were centrifuged for 15 min (4 °C, 20,000g) and were transferred to UPLC-
809 autosampler vials. 2 uL were injected to a Waters Acquity UPLC-Xevo TQ-S UPLC-MS/MS system
810 equipped with an Acquity BEH HILIC (2.1 × 100 mm, 1.7 µm) chromatographic column. An isocratic
811 elution was applied with 10 mM ammonium formate in 95:5 (v/v) acetronitrile:water for 6.3 min at 750

μL/min and 50 °C. Positive electrospray (ESI+) was used as ionization source. The monitored transitions were the following: for derivatized-TMA, +146->+118/59 *m*/*z* (23/27 V); for derivatised-$^{13}C3/^{15}N$-TMA, +150->+63 *m*/*z* (27 V); for TMAO, +76->+59/58 *m*/*z* (12/13 V); for $d_9$-TMAO, +85->+68 *m*/*z* (18 V); for choline, +104->+45/60 *m*/*z* (22/20 V) and for $d_4$-choline, +108->+60 *m*/*z* (20 V).

**Transcriptomics.** Vials containing snap-frozen liver biopsy samples (one per patient) were sent on dry ice to MiltenyiBiotec (Germany), where RNA was extracted from samples using standard extraction protocols (Trizol). RNA was quality-checked [electropherograms, gel images and RNA integrity number (RIN)] using an Agilent 2100 Bioanalyzer platform (Agilent Technologies); RNA with a RIN of greater than six was of sufficient quality for gene expression profiling experiments[59]. For linear T7-based amplification of RNA, 100 ng of each total RNA sample was used. To produce Cy3-labelled cRNA, the RNA was amplified and labeled using the Agilent Low Input Quick Amp Labeling Kit according to the manufacturer's instructions. Amounts of cRNA and dye incorporated were measured using a spectrophotometer (ND-1000; NanoDrop Technologies). Hybridization of the Agilent Whole Human Genome Oligo Microarrays, 4×44K was done according to the Agilent 60-mer oligo microarray processing protocol using the Agilent Gene Expression Hybridization Kit. After two washes with Agilent Gene Expression Wash Buffer and one with acetonitrile, the fluorescence signals of the hybridized Agilent microarrays were detected using Agilent's Microarray Scanner System. The image files were read using Agilent Feature Extraction Software to determine feature intensities (i.e. to produce the raw data).

Microarray data were processed and normalized using R and the BioConductor package LIMMA (Linear Models for Microarray Data), with the modifications for single channel data implemented[60]. Quality of data was assessed using pseudo MA plots and box plots on raw data. Background correction was done (method = 'normexp', offset = 16, normexp.method='rma'). Normalization of the green channel between arrays was done using 'cyclicloess' between pairs of arrays. Control and low-expressed probes were filtered out of the data. Probes that were at least 10 % brighter than the negative controls on at least one array were kept. The batch effect among samples was removed using removeBatchEffect based on 'Batch'[60]. Probes with which no genes (i.e. no Entrez ID) were associated were removed from the batch-corrected data. Probe data were averaged based on

association to a particular gene. The processed data submitted to ArrayExpress (accession E-MTAB-4856) represent the normalized, batch-corrected data with average values for genes. Human KEGG pathways (KGML format) were downloaded from the KEGG PATHWAY database (http://www.genome.jp/kegg/pathway.html) on 29 April 2016 and used in SPIA[61] and network (KEGGgraph, RBGL)[62] analyses. Network analysis was performed using the genes significantly correlated with NAFLD and a set of 20 KEGG pathways involving at least one gene belonging to KEGG liver disease pathway: hsa04151 PI3K-Akt signaling pathway, hsa04145 Phagosome, hsa04010 MAPK signaling pathway, hsa04024 cAMP signaling pathway, hsa04141 Protein processing in endoplasmic reticulum, hsa03010 Ribosome, hsa04060 Cytokine-cytokine receptor interaction, hsa04120 Ubiquitin mediated proteolysis, hsa05206 MicroRNAs in cancer, hsa03050 Proteasome, hsa04931 Insulin resistance, hsa04910 Insulin signaling pathway, hsa04932 Non-alcoholic fatty liver disease (NAFLD), hsa04612 Antigen processing and presentation, hsa04620 Toll-like receptor signaling pathway, hsa04621 NOD-like receptor signaling pathway, hsa05100 Bacterial invasion of epithelial cells, hsa00280 Valine, leucine and isoleucine degradation, hsa00010 Glycolysis/Gluconeogenesis and hsa04923 Regulation of lipolysis in adipocytes.

**16S rRNA gene sequencing (mouse)**

Fecal and ileal content were extracted and sequenced by Vaiomer (Vaiomer SAS, Labège, France) as previously described[63]. Briefly, total DNA was extracted from fecal and ileal content using the using the QIAamp DNA Stool Mini Kit (QIAgen, Hilden, Germany) after two mechanical lysis steps in a bead beater (TissueLyser; Qiagen,); first 3 min at 30 Hz with 5 mm stain steel bead (Qiagen) then two times for 30 sec at 20 Hz with Mobio 0.1 mm glass beads (Qiagen).

The quality and quantity of extracted nucleic acids were evaluated by gel electrophoresis (1% [w/w] agarose in Tris/borate/ethylenediaminetetraacetic acid 0.5×) and NanoDrop 2000 UV spectrophotometer (Thermo Scientific, Waltham, MA, USA). The V3-V4 hypervariable regions of the 16S rDNA were amplified by two steps PCR using Vaiomer V2 primers and sequenced using MiSeq Reagent Kit v3 (2x300 bp Paired-End Reads, Illumina, San Diego, CA, USA) as previously described[63]. The MiSeq sequences were then analysed using the bioinformatics pipeline established by Vaiomer using FROGS v1.4.0 [64]. Briefly, after demultiplexing barcoded Illumina paired reads,

872 single read sequences are cleaned and paired for each sample independently into longer fragments.
873 Operational taxonomic units (OTU) are produced with via single-linkage clustering and taxonomic
874 assignment is performed in order to determine community profiles (generated by Blast+ v2.2.30+
875 against the Silva v128 Parc databank restricted to the bacterial kingdom)[64].
876
877 **Metagenomics (human).** Shotgun sequencing data were generated for 56 patients. Faecal total DNA
878 was extracted from frozen feces using the QIAamp DNA mini stool kit (Qiagen, Courtaboeuf, France),
879 slightly modified by adding a bead- (≤106 µm diameter) beating step (6500 rpm, 3 x 30 s) as
880 previously described[51]. Full details of the pipeline (SCAMP) used to process and analyse
881 metagenomic data are available[65]. Pipeline scripts and instructions for obtaining the independently
882 distributed programs and databases are available from http://www.imperial.ac.uk/bioinformatics-data-
883 science-group/resources/software. Briefly, raw sequence data were assessed for presence of adapter
884 sequences and trimmed using Trim Galore!
885 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/trim_galore_User_Guide_v0.4.1.pdf)
886 to remove low-quality bases (Q < 20) from the 3' end of reads and discarding trimmed reads shorter
887 than 50 nt. Reads were binned to higher taxa (human, parasites, fungi, protozoa/helminths, plants
888 and prokaryotes, **Supplementary Fig. 2**), by alignment to reference databases (**Supplementary**
889 **Table 13**) using the BWA MEM algorithm (https://arxiv.org/abs/1303.3997). Reads that did not map to
890 any reference dataset were assumed to be prokaryotic in origin and subjected to further analysis.
891 MetaPhlAn2.0[66,67] was used to identify the taxonomic composition of each sample and assess the
892 abundance of prokaryotes within the metagenomes. *Bacteroides ovatus*, *Bac. uniformis*, *Bac.*
893 *vulgatus*, *Blautia obeum*, [*Ruminococcus*] *torques*, *Faecalibacterium prausnitzii* and *Subdoligranulum*
894 spp. were detected in all 56 samples. Partial correlations adjusted for age, BMI and country were
895 done on taxa meeting a previously published criterion (median relative abundance of >0.01 % in one
896 or more steatosis groups)[8]. Metagenome assembly was carried out in two rounds using IDBA-UD[68],
897 with an initial independent assembly carried out for each sample. Unassembled reads were then
898 pooled and subjected to a second round of assembly in assembly to improve the representation of
899 low-abundance sequences. *Ab-initio* gene prediction was carried out using MetaGeneMark[69,70]. The
900 resulting predictions were translated, and the protein sequences clustered using the cluster-fast
901 method of UCLUST[71], with a 95% identity cut-off. Centroid sequences from each cluster were used to

form a non-redundant gene catalogue used for downstream analysis. Gene abundance in each sample was determined by alignment of the reads using BWA MEM against the gene catalogue, determining the number of reads mapped to each gene sequence and normalizing as described[19]. Functional annotation was carried out by mapping to the KEGG protein database (version 73.1, downloaded on 10 February 2015) using USEARCH[71] with an e-value cutoff of $1x10^{-9}$.

**Determination of microbial gene richness (MGR).** MGR was derived essentially as described previously[8,19]. Briefly, data were downsized to adjust for sequencing depth and technical variability by randomly selecting 7 million reads mapped to the merged gene catalogue (of 3,902,787 genes) for each sample and then computing the mean number of genes over 30 random drawings. This was done for all 56 patients for whom metagenomic data were available. Results are shown in **Supplementary Table 5**.

**Statistical analyses.** Linear modeling was used to identifier confounders and modifiers within the clinical data, with missing values were replaced by group medians. Metagenomic, transcriptomic and metabolic profiling data were not normally distributed. On the basis of these analyses, partial Spearman rank-based correlations (pSRC) were used to assess associations among the various datasets, with BMI, age and country included as confounders in all analyses. All results were adjusted for multiple testing using the Benjamin and Hochberg procedure (p-FDR) unless otherwise stated. Data are presented as median ± sd. Multivariate matrix correlations were performed using to compare the information between tables as previously described[72] using the modified *Rv* coefficient due to high collinearity in the data[73]. Predictive multivariate models were built using orthogonal partial least squares discriminant analysis (O-PLS-DA) as previously described[74]. The predictive power of O-PLS-DA models was initially assessed using seven-fold cross-validation[74], to derive $Q^2_{Yhat}$ goodness-of-prediction parameters. The significance of the $Q^2_{Yhat}$ parameter was then derived by $H_0$ permutation testing (10,000 iterations)[75] and the predictive ability of the cross-validated O-PLS-DA models was evaluated using bootstrapped Receiver Operator Characteristic (ROC) curves.

**Accession numbers.** The raw metagenomic sequence data (with human-associated reads removed) have been deposited under study accession number PRJEB14215 (secondary accession number

ERP015847). The raw 16S rRNA gene sequence data associated with the mouse FMT work have been deposited under study accession number PRJEB24891.

## EXTENDED DATA TABLES

**Supplementary Table 1.** Clinical information for 105 female bariatric-surgery patients whose liver biopsies were assessed for hepatic steatosis.

**Supplementary Table 2.** Use of linear models to examine effects of confounders/modifiers on analyses (*n* = 105).

**Supplementary Table 3.** Read count and binning data for metagenomic data for each patient (*n* = 56).

**Supplementary Table 4.** Number of metagenome genes found in IGC at different cut-off values (*n* = 56).

**Supplementary Table 5.** Gene counts (MGR) determined for each patient for whom metagenomic data were available (*n* = 56).

**Supplementary Table 6.** Spearman's ranked based partial correlations of liver steatosis with taxonomic abundance data (*n* = 56).

**Supplementary Table 7.** Spearman's ranked based partial correlation (taking into account age, BMI, cohort) of MGR with metabolic phenotyping data (*n* = 56).

**Supplementary Table 8.** Methylamine quantifications by UPLC-MS/MS and Spearman's rank based partial correlation with steatosis (taking into account age, BMI and country; *n* = 60).

**Supplementary Table 9.** Spearman's ranked based partial correlations (taking into account BMI, age, cohort) of liver steatosis with hepatic transcriptome data for the patients for whom metagenomic data were available (*n* = 56).

**Supplementary Table 10.** Spearman's ranked based partial correlations (taking into account age,

BMI, country) of MGR with hepatic transcriptome data (*n* = 56).

**Supplementary Table 11.** Evaluation of shared variance between metagenome and phenome through *Rv* matrix correlation coefficients (*n* = 56).

**Supplementary Table 12.** Areas under the curve (AUC) for bootstrapped Receiver Operator Characteristic curves obtained from 7-fold cross-validated O-PLS-DA models for binary classification between no steatosis (grade 0), *n* = 10; steatosis (grades 1-3), *n* = 46.

**Supplementary Table 13.** Source and composition of reference datasets used in processing of metagenomic data.

978 **EXTENDED DATA FIGURES**

979

980 **Supplementary Figure 1.** Determination of distinction between confounders and modifiers, for
981 inclusion of confounders in partial correlations (*n* = 105). **a**, Effect of country of recruitment on clinical
982 data. Red, Spain; blue, Italy. **b**, Based on linear modeling BMI, country and age were found to be
983 confounders. Significant differences between the data for the Spanish and Italian cohorts were
984 determined using Student's *t* test. **c**, Example of identification of modifiers rather than confounders,
985 using glucose disposal rate (M) (mg/(kg/min)). Any change in slope of the line between both models
986 indicates that M (mg/(kg/min)) is a mediator not a confounder and can, therefore, not be corrected for
987 in partial correlations. **d**, Heatmap of partial Spearman rank-based correlations between clinical
988 parameters adjusted for age, BMI and country.

989

990 **Supplementary Figure 2.** Breakdown of binning of metagenomic DNA to different kingdoms (*n* = 56).
991 **a**, Total DNA (reads). The majority of faecal DNA belonged to prokaryotes (archaea and bacteria). **b**,
992 Plant DNA. Plant-associated DNA was predominated by dietary sources of plant material. **c**,
993 Parasite/helminth DNA. Among the parasite/helminth DNA in samples, reads from *Trichuristrichiura*
994 (human whipworm) predominated.

995

996 **Supplementary Figure 3.** Additional analyses of taxonomic data generated using MetaPhlAn2.0 and
997 the metagenomic sequence data (*n* = 56). **a**, Upper two rows: prokaryotic species significantly (p-FDR
998 < 0.05) anti-correlated with liver steatosis; lower two rows, prokaryotic species significantly (p-FDR <
999 0.05) correlated with liver steatosis. **b**, Species richness, measured using Chao1, was not significantly
1000 correlated with liver steatosis ($p = 0.0750$).

1001

1002 **Supplementary Figure 4.** Heatmaps showing partial Spearman rank-based correlation of abundance
1003 data at different taxonomic ranks with clinical data for the 56 patients whose metagenomes were
1004 analyzed. +, p-FDR < 0.05.

1005

1006 **Supplementary Figure 5.** Heatmaps showing partial Spearman rank-based correlation of
1007 metagenome-derived KEGG pathway data with clinical data for the 56 patients whose metagenomes

were analyzed. +, p-FDR < 0.05.

**Supplementary Figure 6.** Heatmap showing partial Spearman rank-based correlation of urinary metabolites with clinical data for the 56 patients included in the metagenomic study. +, p-FDR < 0.05. Only significant annotated urinary metabolites are shown.

**Supplementary Figure 7.** Heatmap showing partial Spearman rank-based correlation of plasma metabolites with clinical data for the 56 patients included in the metagenomic study. +, p-FDR < 0.05. Only significant annotated plasma metabolites are shown.

**Supplementary Figure 8.** $^{1}$H-NMR-based Metabolome-Wide Association Study in urine and plasma for MGR and steatosis. Red dots, significantly (p-FDR < 0.05) correlated with MGR or steatosis; blue dots, significantly (p-FDR < 0.05) anti-correlated with MGR or steatosis; grey dots, not significantly correlated with MGR or steatosis.

**Supplementary Figure 9.** Heatmap showing partial Spearman rank-based correlation of plasma metabolites with clinical data for the 102 patients within the FLORINASH cohort for whom plasma metabolomes were available. +, p-FDR < 0.05. Only annotated plasma metabolites are shown.

**Supplementary Figure 10.** Enrichr[76,77] was used to identify KEGG pathways related to genes significantly correlated (pSRC) with hepatic steatosis for 56 patients. Additional significant (p-FDR <0.05) results are shown for the KEGG pathways associated with genes positively correlated with steatosis.

**Supplementary Figure 11. Additional recipient mouse phenotypes predicted from donor microbiota composition.** ***a***, *Fabp4* gene expression in liver of recipient mice. ***b*** Plasma valine measured by $^{1}$H-NMR. **c-d**, Permutation tests (n = 10,000) for goodness of fit ($R^2$) and prediction ($Q^2$) parameters obtained from a seven-fold crossvalidated O-PLS regression model quantitatively predicting recipient mouse phenomes from human donor microbiome composition: **c**, hepatic *Fabp4*, **d**, plasma valine. Data obtained from FMT protocols performed with independent 3 patients with liver

steatosis (grade 3, >66% steatosis) and 3 control patients (grade 0, <5% steatosis), *n* = 8 recipient mice per donor. Data are mean ± s.e.m., * $p < 0.05$.

**Supplementary Figure 12. Link between mouse phenotypes and their microbiota composition.** Heatmap obtained Spearman's ranked based correlations between mouse phenotypes and Family taxonomical level derived from 16S rRNA gene amplicon analysis (p-FDR<0.05, n = 43).

## SUPPLEMENTARY REFERENCES


54. Nicolas, S. *et al.* Transfer of dysbiotic gut microbiota has beneficial effects on host liver metabolism. *Mol. Syst. Biol.* **13,** 921 (2017).
55. Grasset, E. *et al.* A Specific Gut Microbiota Dysbiosis of Type 2 Diabetic Mice Induces GLP-1 Resistance through an Enteric NO-Dependent and Gut-Brain Axis Mechanism. *Cell Metab.* **25,** 1075–1090.e5 (2017).
56. Grasa, M. D. M. *et al.* Modulation of SHBG binding to testosterone and estradiol by sex and morbid obesity. *Eur. J. Endocrinol.* **176,** 393–404 (2017).
57. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2018. *Diabetes Care* **41,** S13–S27 (2018).
58. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in $^{1}$H NMR metabonomics. *Anal Chem* **78,** 4281–4290 (2006).
59. Fleige, S. & Pfaffl, M. W. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine* **27,** 126–139 (2006).
60. Smyth, G. K. in *… and computational biology solutions using R and …* 397–420 (Springer-Verlag, 2005). doi:10.1007/0-387-29362-0_23
61. Tarca, A. L. *et al.* A novel signaling pathway impact analysis. *Bioinformatics* **25,** 75–82 (2009).
62. Zhang, J. D. & Wiemann, S. KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. *Bioinformatics* **25,** 1470–1471 (2009).
63. Lluch, J. *et al.* The Characterization of Novel Tissue Microbiota Using an Optimized 16S Metagenomic Sequencing Pipeline. *PLoS ONE* **10,** e0142334 (2015).
64. Escudié, F. *et al.* FROGS: Find, Rapidly, OTUs with Galaxy Solution. *Bioinformatics* **40,** 299 (2017).
65. Abbott, J. C. *et al.* ScAMP: Scalable Automated Metagenomics Pipeline. *in prep*
66. Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9,** 811–814 (2012).
67. Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12,** 902–903 (2015).
68. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28,** 1420–1428 (2012).
69. Besemer, J. Heuristic approach to deriving models for gene finding. *Nucleic Acids Res.* **27,** 3911–3920 (1999).
70. Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38,** e132–e132 (2010).
71. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).
72. Dumas, M.-E. *et al.* Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. *Anal Chem* **78,** 2199–2208 (2006).
73. Smilde, A. K., Kiers, H. A. L., Bijlsma, S., Rubingh, C. M. & van Erk, M. J. Matrix correlations for high-dimensional data: the modified RV-coefficient. *Bioinformatics* **25,** 401–405 (2009).
74. Cloarec, O. *et al.* Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies. *Anal Chem* **77,** 517–526 (2005).
75. Blaise, B. J. *et al.* Metabotyping of Caenorhabditis elegans reveals latent phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 19808–19812 (2007).
76. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14,** 128 (2013).
77. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44,** W90–7 (2016).

**Fig. 1** **a**

## Patient recruitment and clinical markers

### Clinical recruitment

Spain (n = 44)

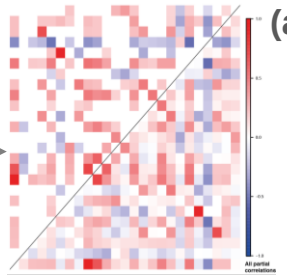
Italy (n = 61)

### Primary phenotype steatosis (liver histology)

Inclusion criteria:
Obese women of Caucasian origin, body weight stable > 3 months, no antibiotics < 1 month

Exclusion criteria:
Alcohol consumption >20 g/day, iron overload, hepatitis B&C, auto-immune hepatitis, thyroid dysfunction, type 2 diabetes, cancer

### Confounder & modifier analyses (age, BMI, country)

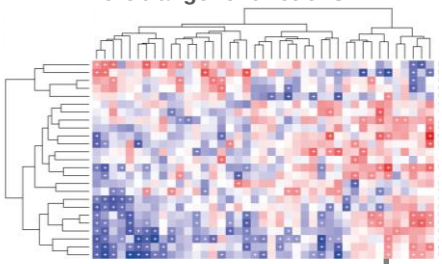


Spearman's partial correlations for hepatic steatosis, adjusted for age, BMI & country with Benjamini and Hochberg multiple testing corrections

hepatic steatosis

## b Metagenome-wide association

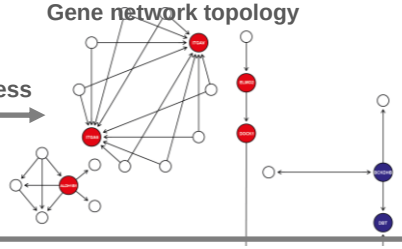
Taxonomical abundance (394 species)
Microbial gene richness (10 million genes)
Microbial gene functions



gene richness

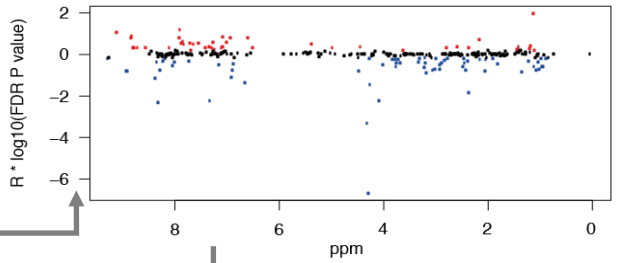
## c Transcriptome-wide association

Genes associated with steatosis and MGR
Gene enrichment
Gene network topology



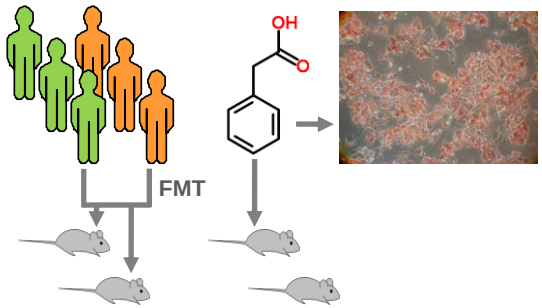
## d Metabolome-wide association

Urine and plasma $^{1}$H NMR
Plasma UPLC-MS/MS
Metabolites associated with steatosis and MGR



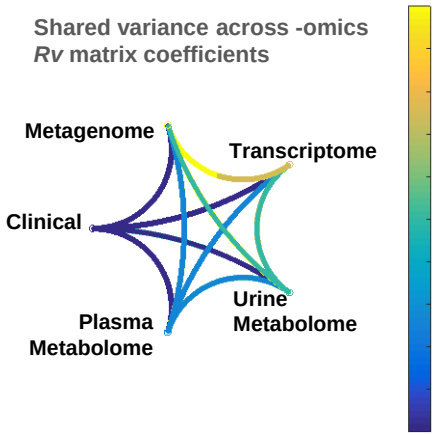

## e Biological validations

Fecal microbiome transplants in mice
Oral gavage in mice
Primary human hepatocytes



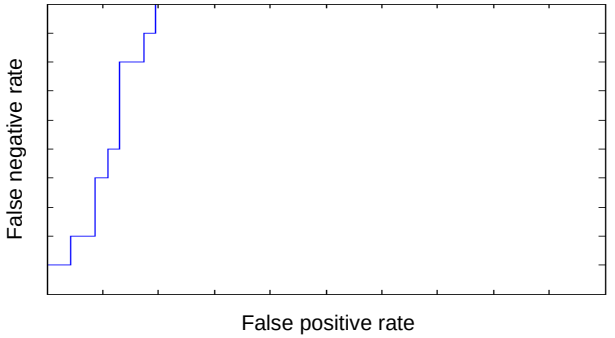

## f Cross-matrix information

Shared variance across -omics
*Rv* matrix coefficients




## g Integrative Modelling

Multi-omics predictive O-PLS-DA models
7-fold crossvalidation and 10,000 $H_0$ permutations
Bootstrapped ROC analysis



# Fig. 2

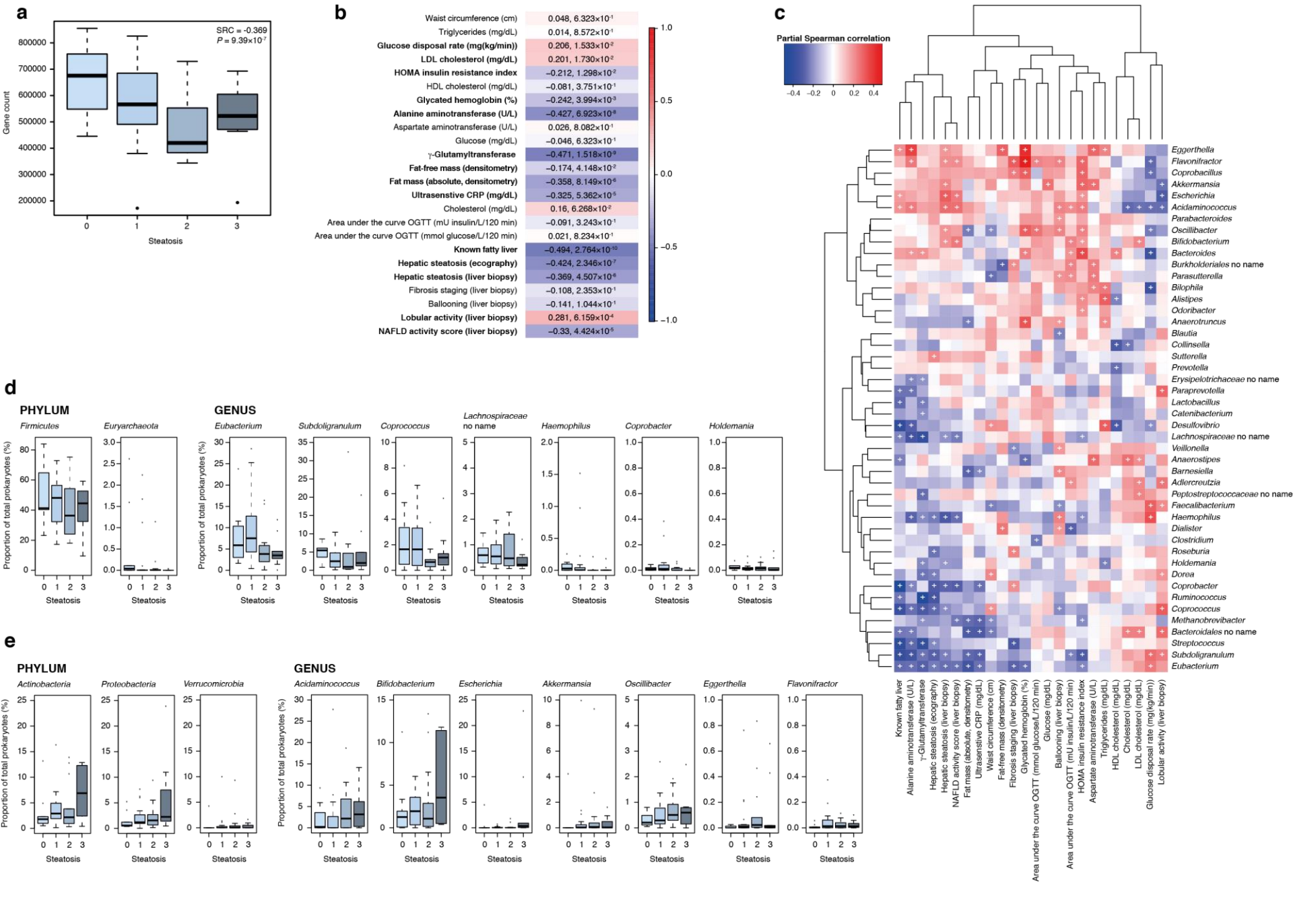
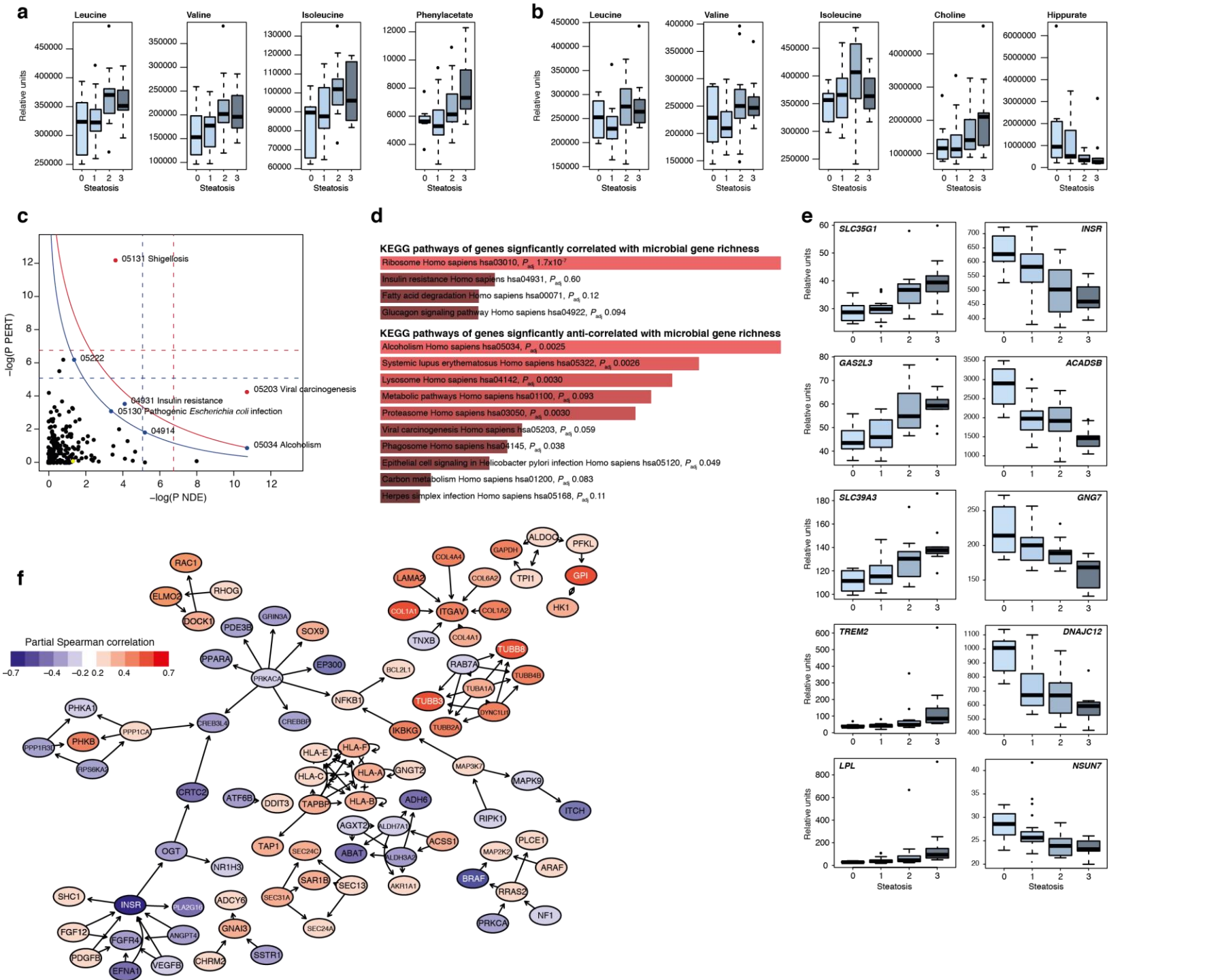
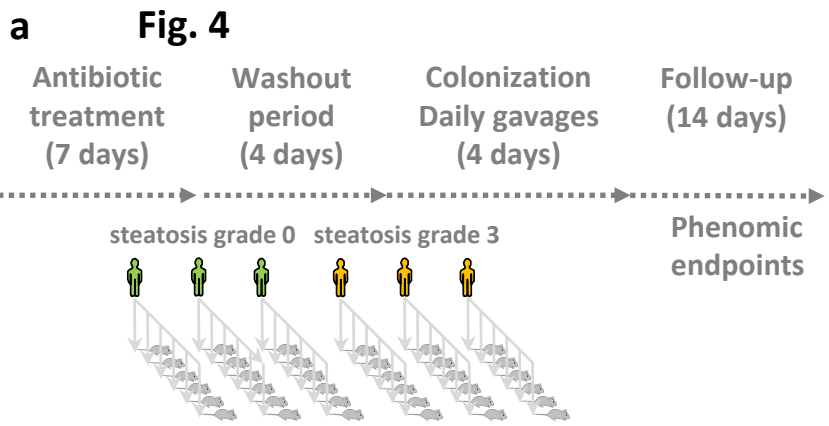
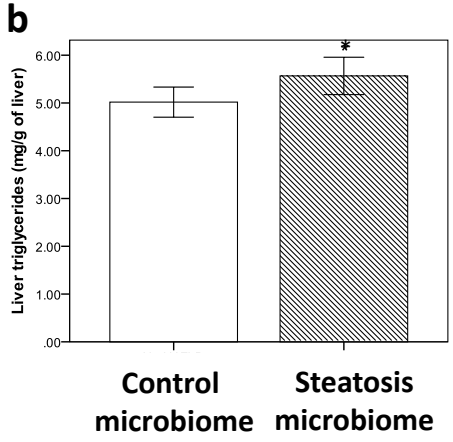

a
SRC = -0.369
P = 9.39×10-7
Gene count
Steatosis
b
Waist circumference (cm) 0.048, 6.323×10-1
Triglycerides (mg/dL) 0.014, 8.572×10-1
Glucose disposal rate (mg(kg/min)) 0.206, 1.533×10-2
LDL cholesterol (mg/dL) 0.201, 1.730×10-2
HOMA insulin resistance index −0.212, 1.298×10-2
HDL cholesterol (mg/dL) −0.081, 3.751×10-1
Glycated hemoglobin (%) −0.242, 3.994×10-3
Alanine aminotransferase (U/L) −0.427, 6.923×10-8
Aspartate aminotransferase (U/L) 0.026, 8.082×10-1
Glucose (mg/dL) −0.046, 6.323×10-1
γ-Glutamyltransferase −0.471, 1.518×10-9
Fat-free mass (densitometry) −0.174, 4.148×10-2
Fat mass (absolute, densitometry) −0.358, 8.149×10-6
Ultrasenstive CRP (mg/dL) −0.325, 5.362×10-5
Cholesterol (mg/dL) 0.16, 6.268×10-2
Area under the curve OGTT (mU insulin/L/120 min) −0.091, 3.243×10-1
Area under the curve OGTT (mmol glucose/L/120 min) 0.021, 8.234×10-1
Known fatty liver −0.494, 2.764×10-10
Hepatic steatosis (ecography) −0.424, 2.346×10-7
Hepatic steatosis (liver biopsy) −0.369, 4.507×10-6
Fibrosis staging (liver biopsy) −0.108, 2.353×10-1
Ballooning (liver biopsy) −0.141, 1.044×10-1
Lobular activity (liver biopsy) 0.281, 6.159×10-4
NAFLD activity score (liver biopsy) −0.33, 4.424×10-5
c
Partial Spearman correlation
Eggerthella
Flavonifractor
Coprobacillus
Akkermansia
Escherichia
Acidaminococcus
Parabacteroides
Oscillibacter
Bifidobacterium
Bacteroides
Burkholderiales no name
Parasutterella
Bilophila
Alistipes
Odoribacter
Anaerotruncus
Blautia
Collinsella
Sutterella
Prevotella
Erysipelotrichaceae no name
Paraprevotella
Lactobacillus
Catenibacterium
Desulfovibrio
Lachnospiraceae no name
Veillonella
Anaerostipes
Barnesiella
Adlercreutzia
Peptostreptococcaceae no name
Faecalibacterium
Haemophilus
Dialister
Clostridium
Roseburia
Holdemania
Dorea
Coprobacter
Ruminococcus
Coprococcus
Methanobrevibacter
Bacteroidales no name
Streptococcus
Subdoligranulum
Eubacterium
Known fatty liver
Alanine aminotransferase (U/L)
γ-Glutamyltransferase
Hepatic steatosis (ecography)
Hepatic steatosis (liver biopsy)
NAFLD activity score (liver biopsy)
Fat mass (absolute, densitometry)
Ultrasenstive CRP (mg/dL)
Waist circumference (cm)
Fat-free mass (densitometry)
Fibrosis staging (liver biopsy)
Glycated hemoglobin (%)
Area under the curve OGTT (mmol glucose/L/120 min)
Glucose (mg/dL)
Ballooning (liver biopsy)
Area under the curve OGTT (mU insulin/L/120 min)
HOMA insulin resistance index
Aspartate aminotransferase (U/L)
Triglycerides (mg/dL)
HDL cholesterol (mg/dL)
Cholesterol (mg/dL)
LDL cholesterol (mg/dL)
Glucose disposal rate (mg/kg/min)
Lobular activity (liver biopsy)
d
PHYLUM
Firmicutes
Euryarchaeota
GENUS
Eubacterium
Subdoligranulum
Coprococcus
Lachnospiraceae no name
Haemophilus
Coprobacter
Holdemania
Proportion of total prokaryotes (%)
Steatosis
e
PHYLUM
Actinobacteria
Proteobacteria
Verrucomicrobia
GENUS
Acidaminococcus
Bifidobacterium
Escherichia
Akkermansia
Oscillibacter
Eggerthella
Flavonifractor
Proportion of total prokaryotes (%)
Steatosis

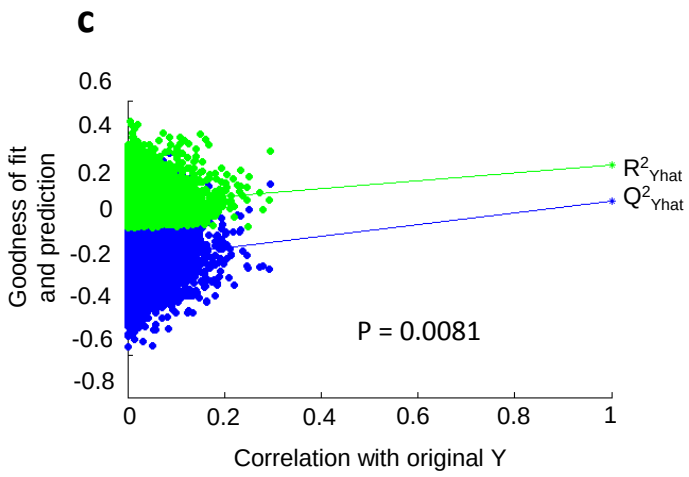
Fig. 3


a
Leucine
Valine
Isoleucine
Phenylacetate
Relative units
Steatosis
b
Leucine
Valine
Isoleucine
Choline
Hippurate
Relative units
Steatosis
c
05131 Shigellosis
05222
05203 Viral carcinogenesis
04931 Insulin resistance
05130 Pathogenic Escherichia coli infection
04914
05034 Alcoholism
−log(P PERT)
−log(P NDE)
d
KEGG pathways of genes signficantly correlated with microbial gene richness
Ribosome Homo sapiens hsa03010, P_adj 1.7x10-7
Insulin resistance Homo sapiens hsa04931, P_adj 0.60
Fatty acid degradation Homo sapiens hsa00071, P_adj 0.12
Glucagon signaling pathway Homo sapiens hsa04922, P_adj 0.094
KEGG pathways of genes signficantly anti-correlated with microbial gene richness
Alcoholism Homo sapiens hsa05034, P_adj 0.0025
Systemic lupus erythematosus Homo sapiens hsa05322, P_adj 0.0026
Lysosome Homo sapiens hsa04142, P_adj 0.0030
Metabolic pathways Homo sapiens hsa01100, P_adj 0.093
Proteasome Homo sapiens hsa03050, P_adj 0.0030
Viral carcinogenesis Homo sapiens hsa05203, P_adj 0.059
Phagosome Homo sapiens hsa04145, P_adj 0.038
Epithelial cell signaling in Helicobacter pylori infection Homo sapiens hsa05120, P_adj 0.049
Carbon metabolism Homo sapiens hsa01200, P_adj 0.083
Herpes simplex infection Homo sapiens hsa05168, P_adj 0.11
e
SLC35G1
INSR
GAS2L3
ACADSB
SLC39A3
GNG7
TREM2
DNAJC12
LPL
NSUN7
Relative units
Steatosis
f
Partial Spearman correlation
−0.7 −0.4 −0.2 0.2 0.4 0.7
RAC1
ELMO2
RHOG
DOCK1
GRIN3A
PDE3B
SOX9
PPARA
EP300
PRKACA
BCL2L1
NFKB1
CREB3L4
CREBBP
PHKA1
PPP1CA
PHKB
PPP1R3D
RPS6KA2
IKBKG
HLA-E
HLA-F
HLA-C
HLA-A
GNGT2
HLA-B
TAPBP
ATF6B
DDIT3
CRTC2
OGT
NR1H3
TAP1
SEC24C
SAR1B
SEC13
SEC31A
SEC24A
ABAT
ALDH7A1
ALDH3A2
AKR1A1
ACSS1
ADH6
MAP3K7
MAPK9
ITCH
RIPK1
MAP2K2
PLCE1
ARAF
BRAF
RRAS2
NF1
PRKCA
SHC1
INSR
PLA2G16
ADCY6
GNAI3
SSTR1
CHRM2
FGF12
FGFR4
ANGPT4
PDGFB
VEGFB
EFNA1
ALDOC
PFKL
GAPDH
TPI1
GPI
HK1
COL4A4
COL6A2
LAMA2
COL1A1
ITGAV
COL1A2
TNXB
COL4A1
TUBB8
RAB7A
TUBB4B
TUBA1A
TUBB3
DYNC1LI1
TUBB2A

# Fig. 4


a
Antibiotic treatment (7 days)
Washout period (4 days)
Colonization Daily gavages (4 days)
Follow-up (14 days)
steatosis grade 0
steatosis grade 3
Phenomic endpoints


b
Liver triglycerides (mg/g of liver)
6.00
5.00
4.00
3.00
2.00
1.00
.00
*
Control microbiome
Steatosis microbiome


c
Goodness of fit and prediction
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
R2Yhat
Q2Yhat
P = 0.0081
0
0.2
0.4
0.6
0.8
1
Correlation with original Y

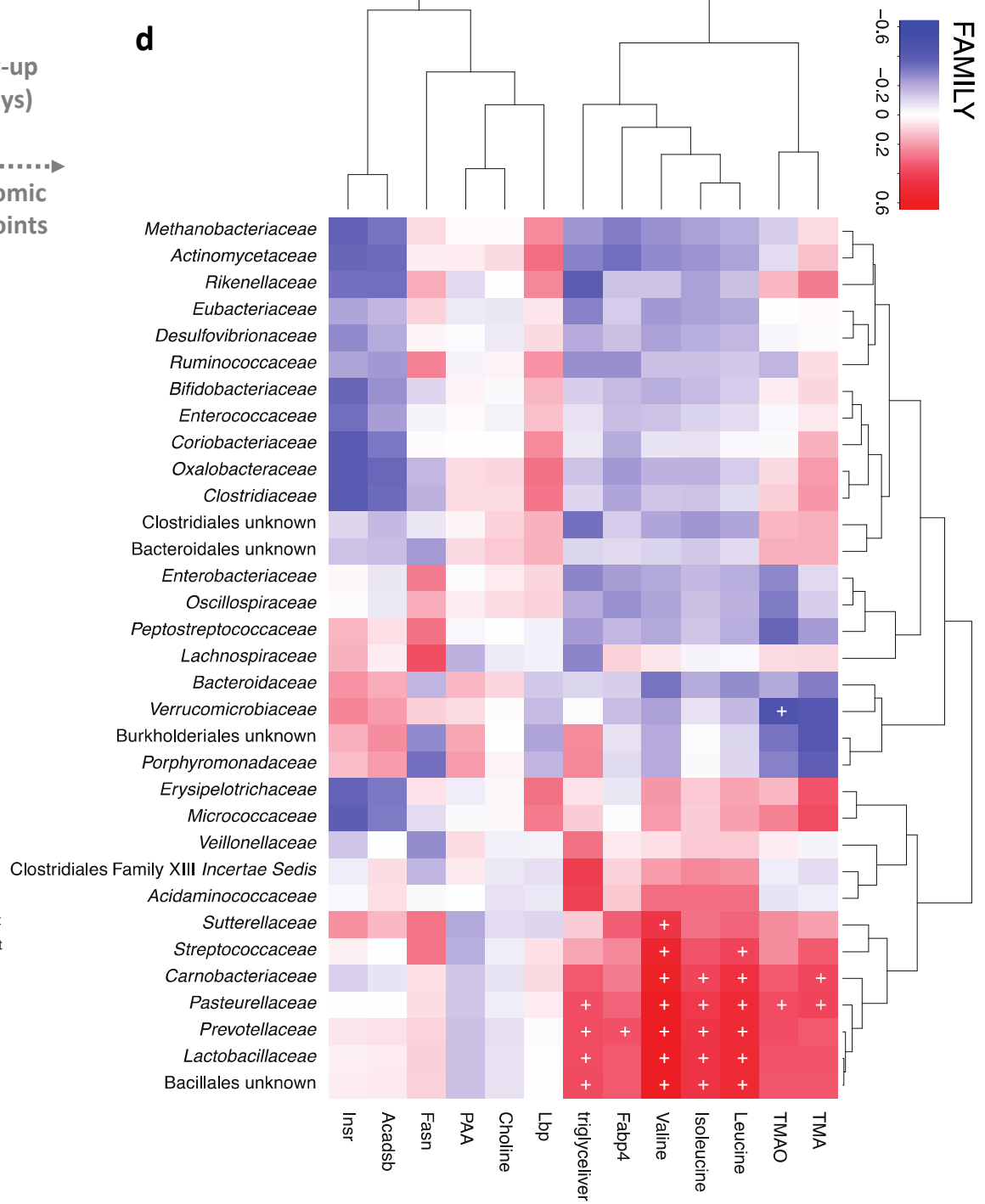

d
FAMILY
0.6
0.2
0
-0.2
-0.6
Methanobacteriaceae
Actinomycetaceae
Rikenellaceae
Eubacteriaceae
Desulfovibrionaceae
Ruminococcaceae
Bifidobacteriaceae
Enterococcaceae
Coriobacteriaceae
Oxalobacteraceae
Clostridiaceae
Clostridiales unknown
Bacteroidales unknown
Enterobacteriaceae
Oscillospiraceae
Peptostreptococcaceae
Lachnospiraceae
Bacteroidaceae
Verrucomicrobiaceae
Burkholderiales unknown
Porphyromonadaceae
Erysipelotrichaceae
Micrococcaceae
Veillonellaceae
Clostridiales Family XIII Incertae Sedis
Acidaminococcaceae
Sutterellaceae
Streptococcaceae
Carnobacteriaceae
Pasteurellaceae
Prevotellaceae
Lactobacillaceae
Bacillales unknown
Insr
Acadsb
Fasn
PAA
Choline
Lbp
triglyceliver
Fabp4
Valine
Isoleucine
Leucine
TMAO
TMA

Fig.5

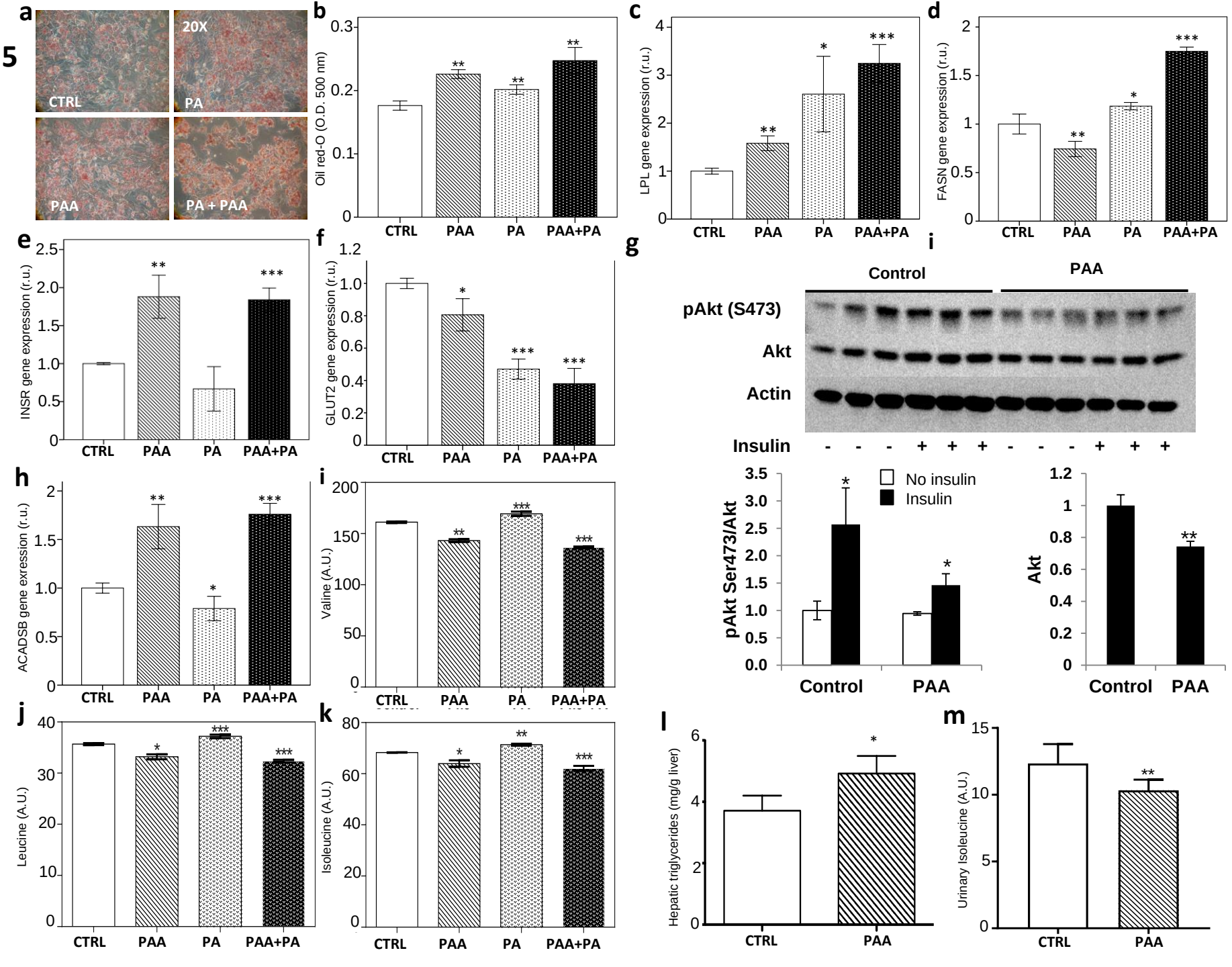
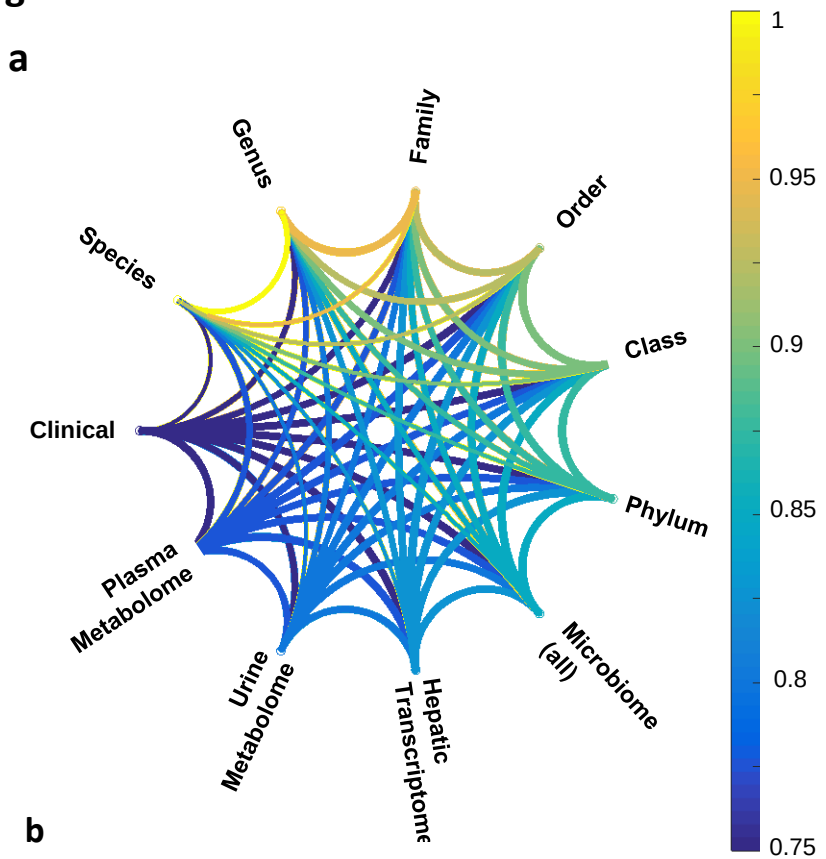
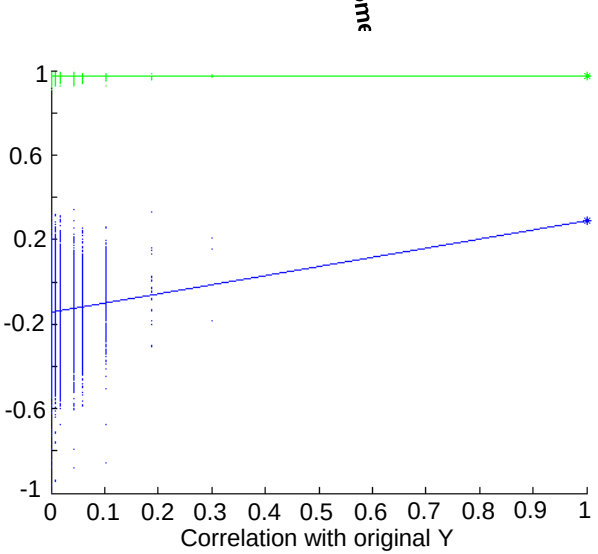
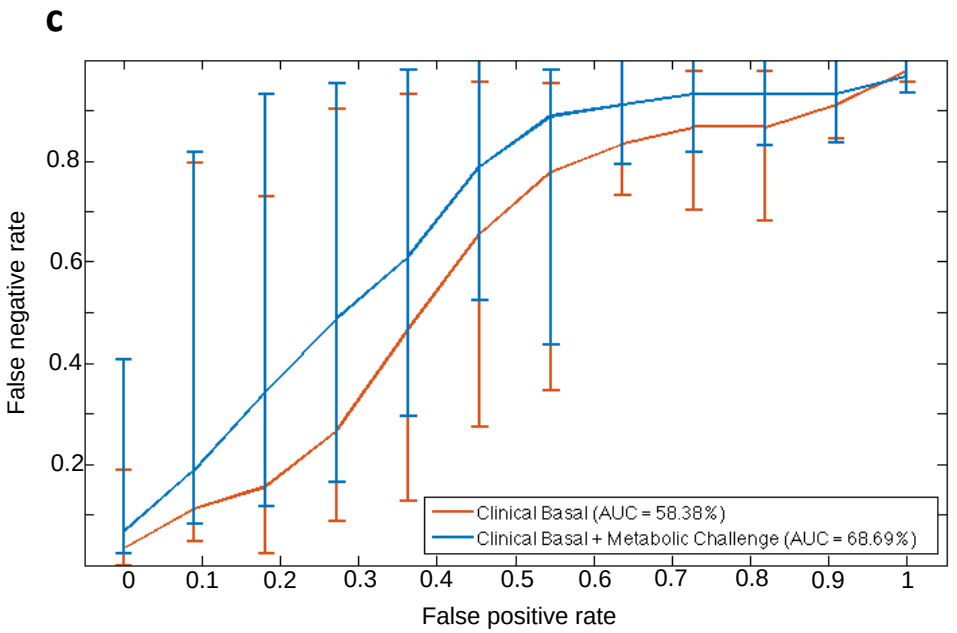
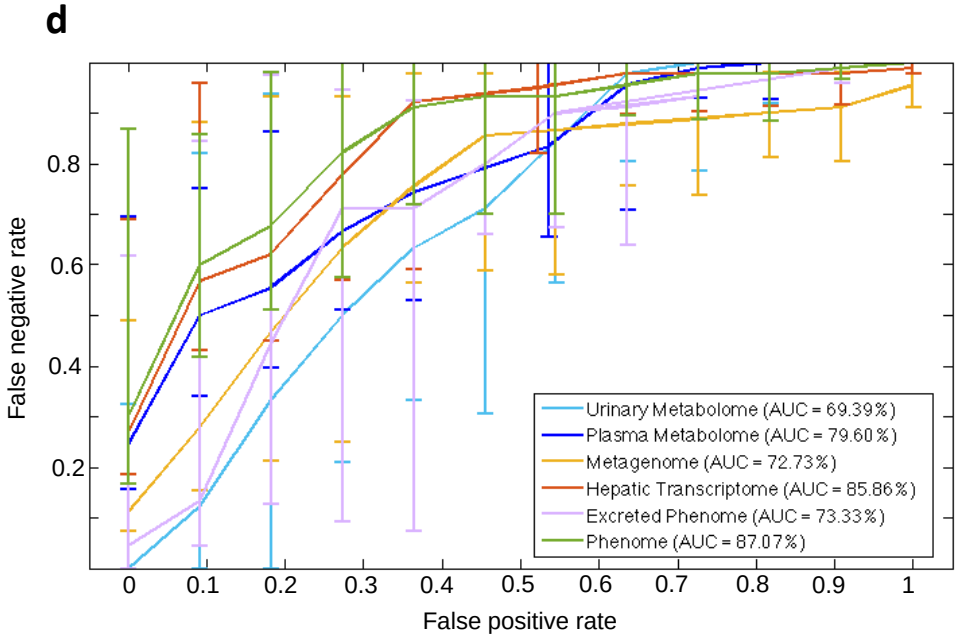

a
CTRL
20X
PA
PAA
PA + PAA
b
Oil red-O (O.D. 500 nm)
CTRL
PAA
PA
PAA+PA
c
LPL gene expression (r.u.)
d
FASN gene expression (r.u.)
e
INSR gene expression (r.u.)
f
GLUT2 gene expression (r.u.)
g
Control
PAA
pAkt (S473)
Akt
Actin
Insulin
i
pAkt Ser473/Akt
No insulin
Insulin
Akt
h
ACADSB gene exression (r.u.)
i
Valine (A.U.)
j
Leucine (A.U.)
k
Isoleucine (A.U.)
l
Hepatic triglycerides (mg/g liver)
m
Urinary Isoleucine (A.U.)

Fig. 6

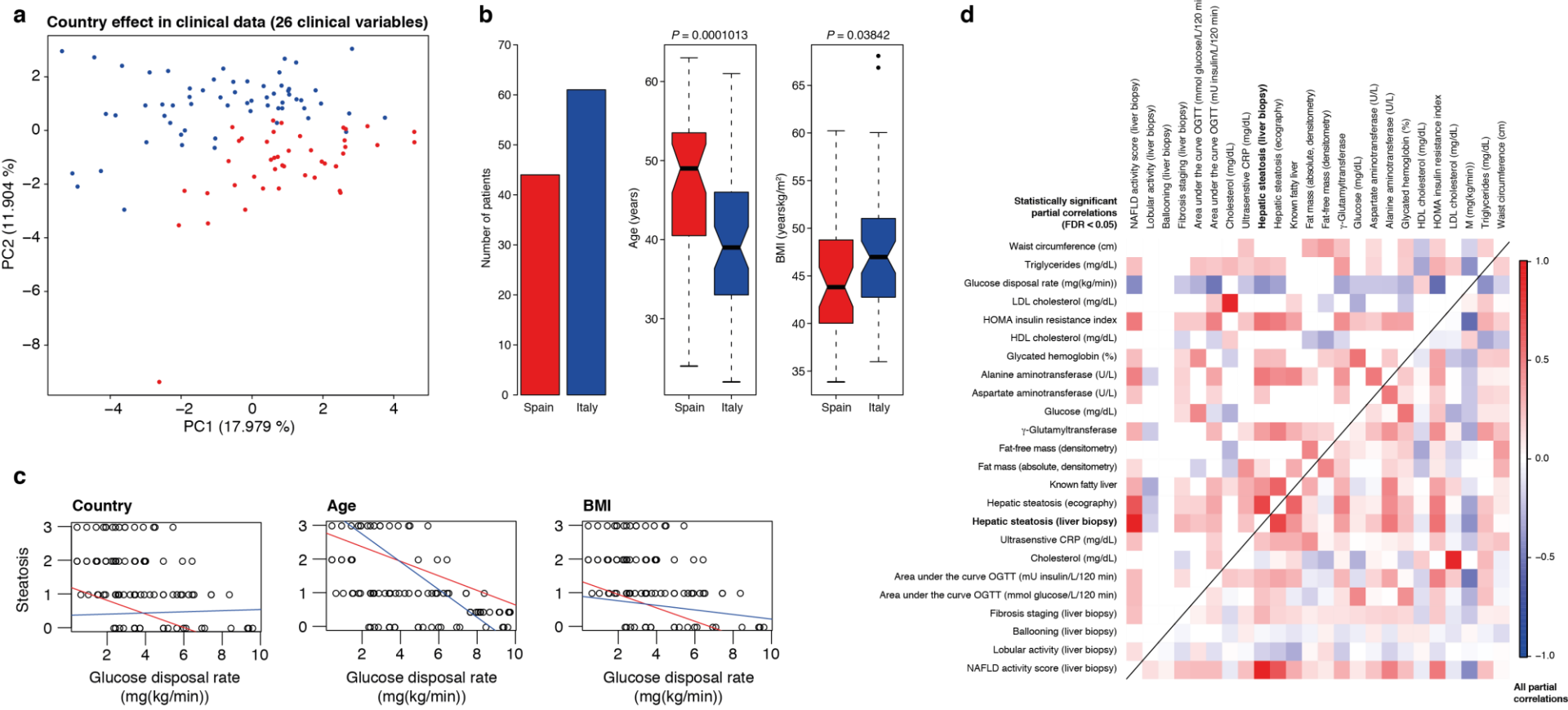

a
Genus
Family
Order
Species
Class
Clinical
Phylum
Plasma Metabolome
Urine Metabolome
Hepatic Transcriptome
Microbiome (all)
1
0.95
0.9
0.85
0.8
0.75


b
1
0.6
0.2
-0.2
-0.6
-1
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Correlation with original Y

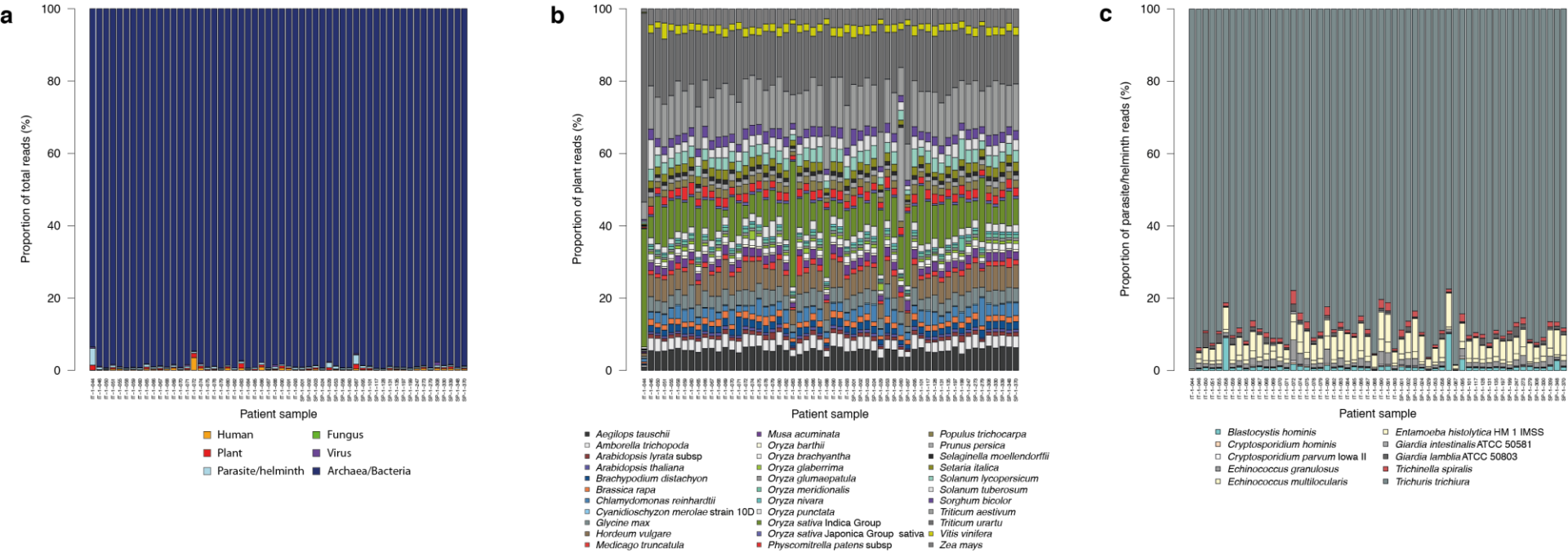

c
False negative rate
0.8
0.6
0.4
0.2
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
False positive rate
Clinical Basal (AUC = 58.38%)
Clinical Basal + Metabolic Challenge (AUC = 68.69%)


d
False negative rate
0.8
0.6
0.4
0.2
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
False positive rate
Urinary Metabolome (AUC = 69.39%)
Plasma Metabolome (AUC = 79.60%)
Metagenome (AUC = 72.73%)
Hepatic Transcriptome (AUC = 85.86%)
Excreted Phenome (AUC = 73.33%)
Phenome (AUC = 87.07%)

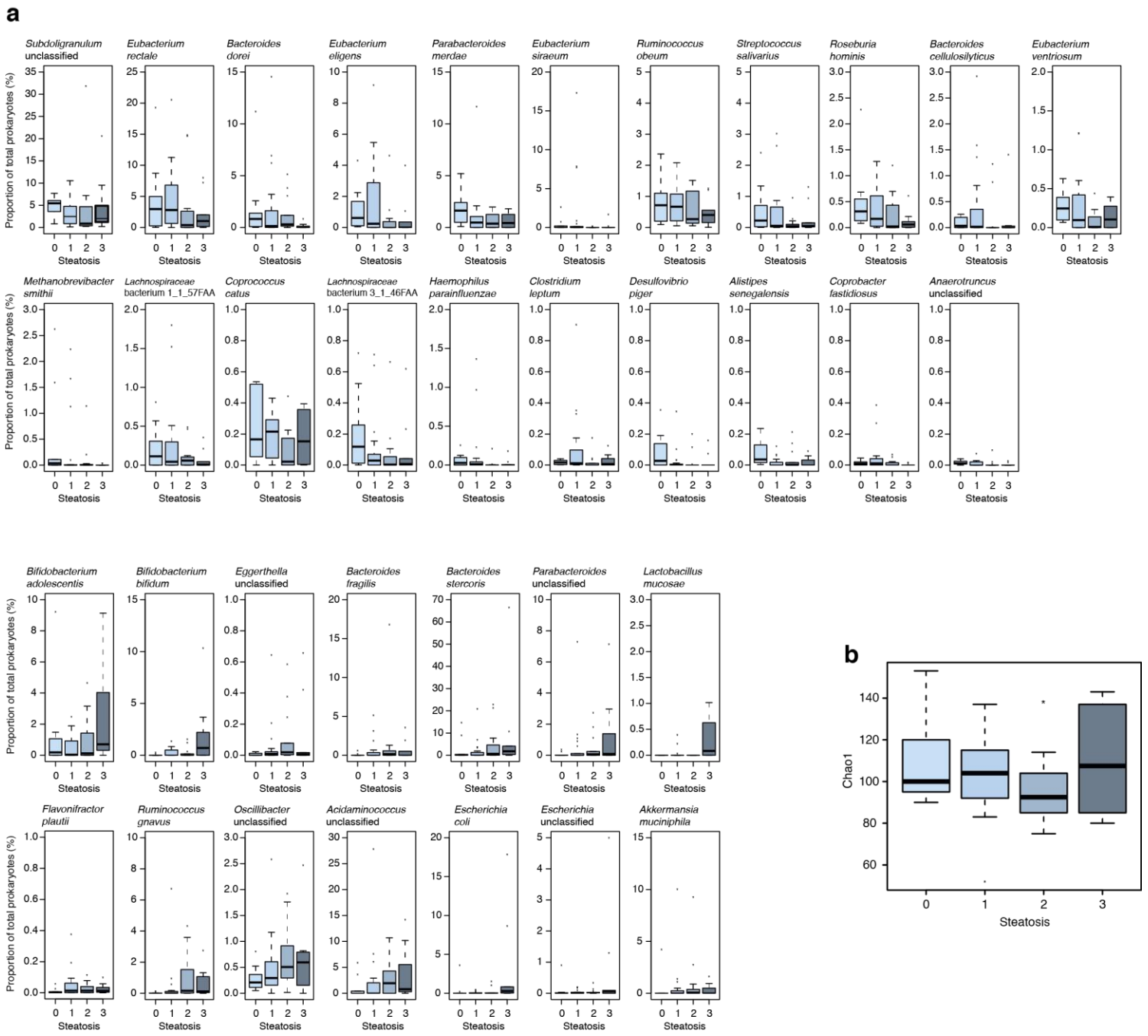
# Supplementary Fig. 1



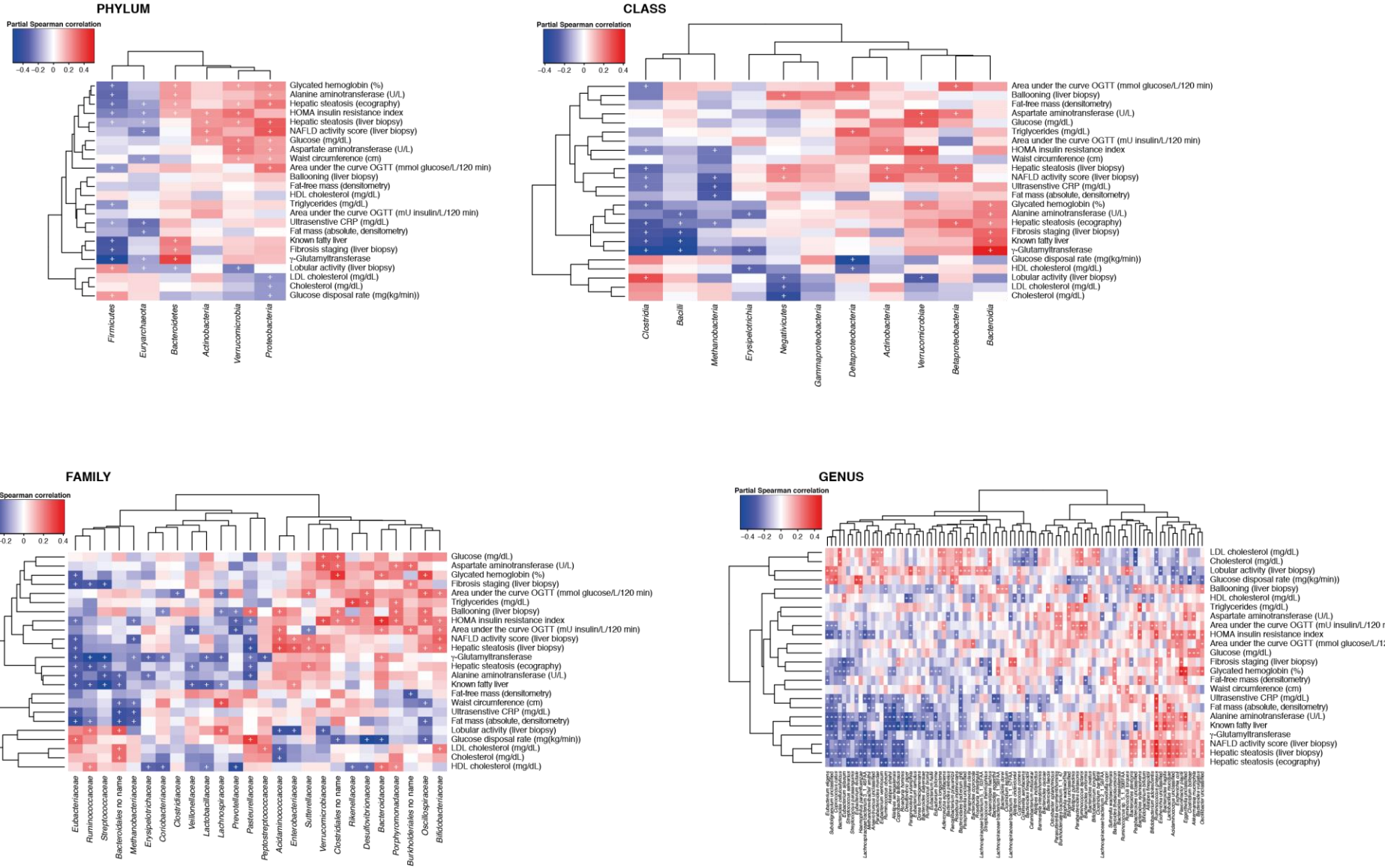
# Supplementary Fig. 2



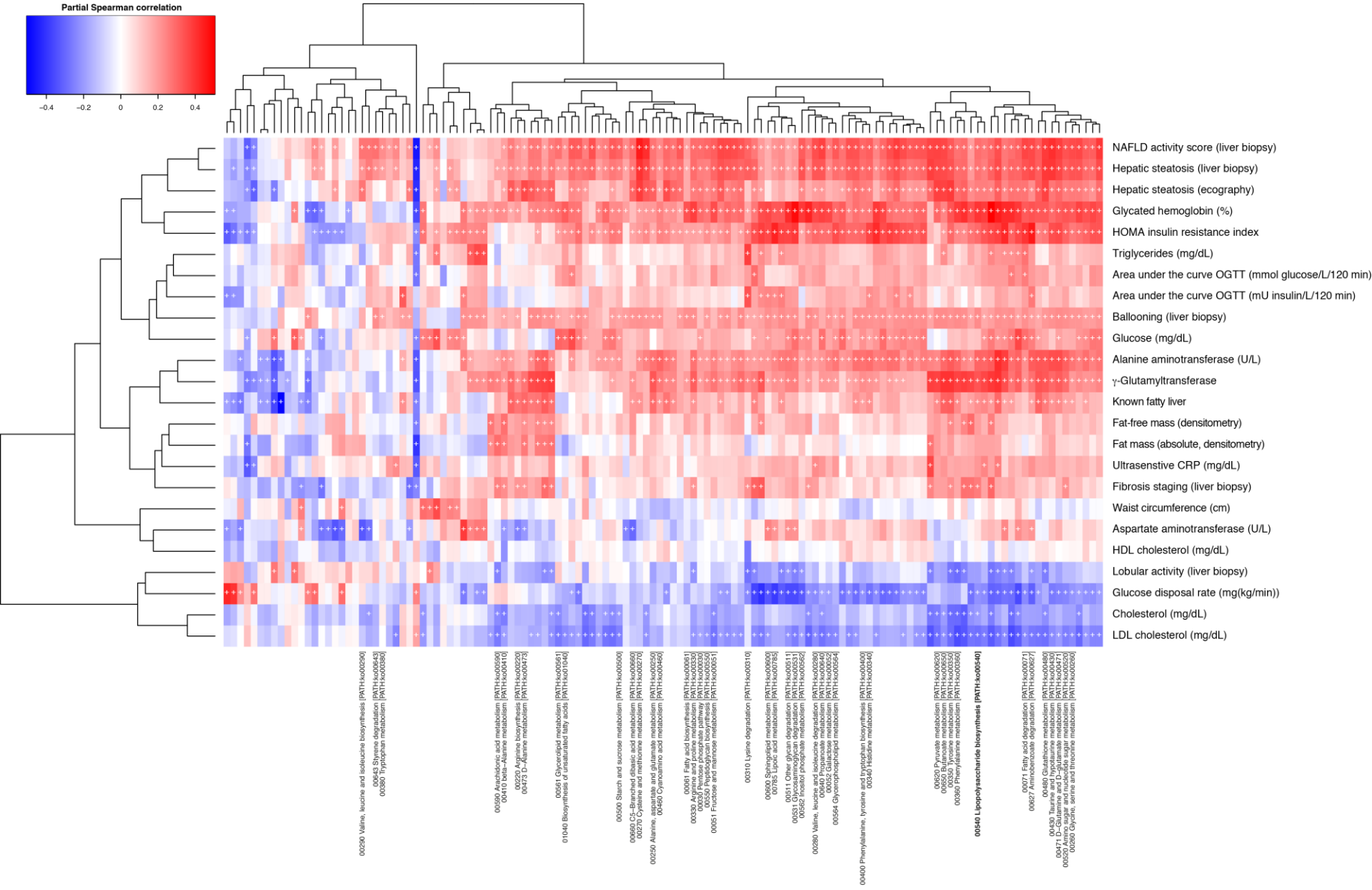
# Supplementary Fig. 3



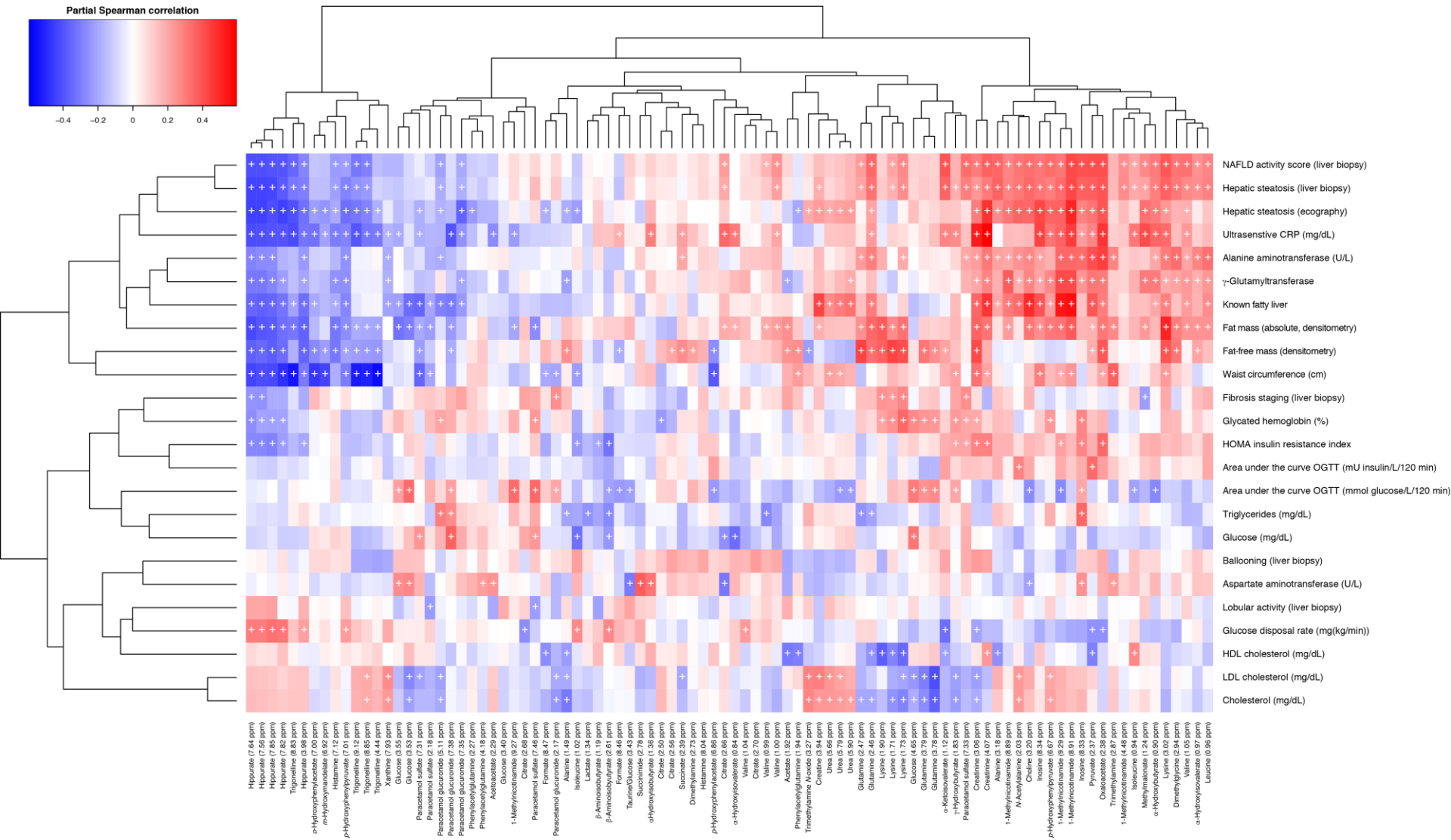
# Supplementary Fig. 4



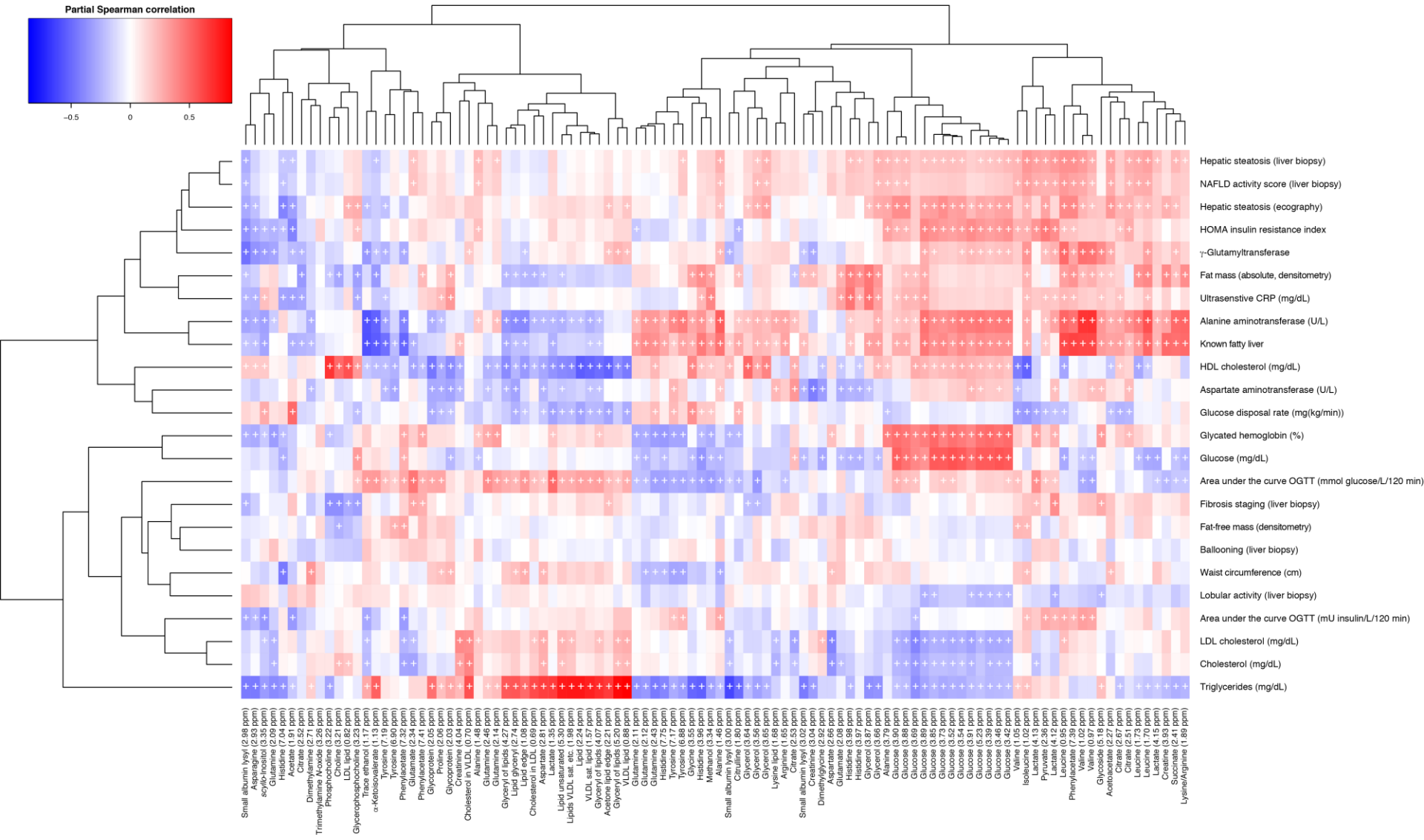
# Supplementary Fig. 5



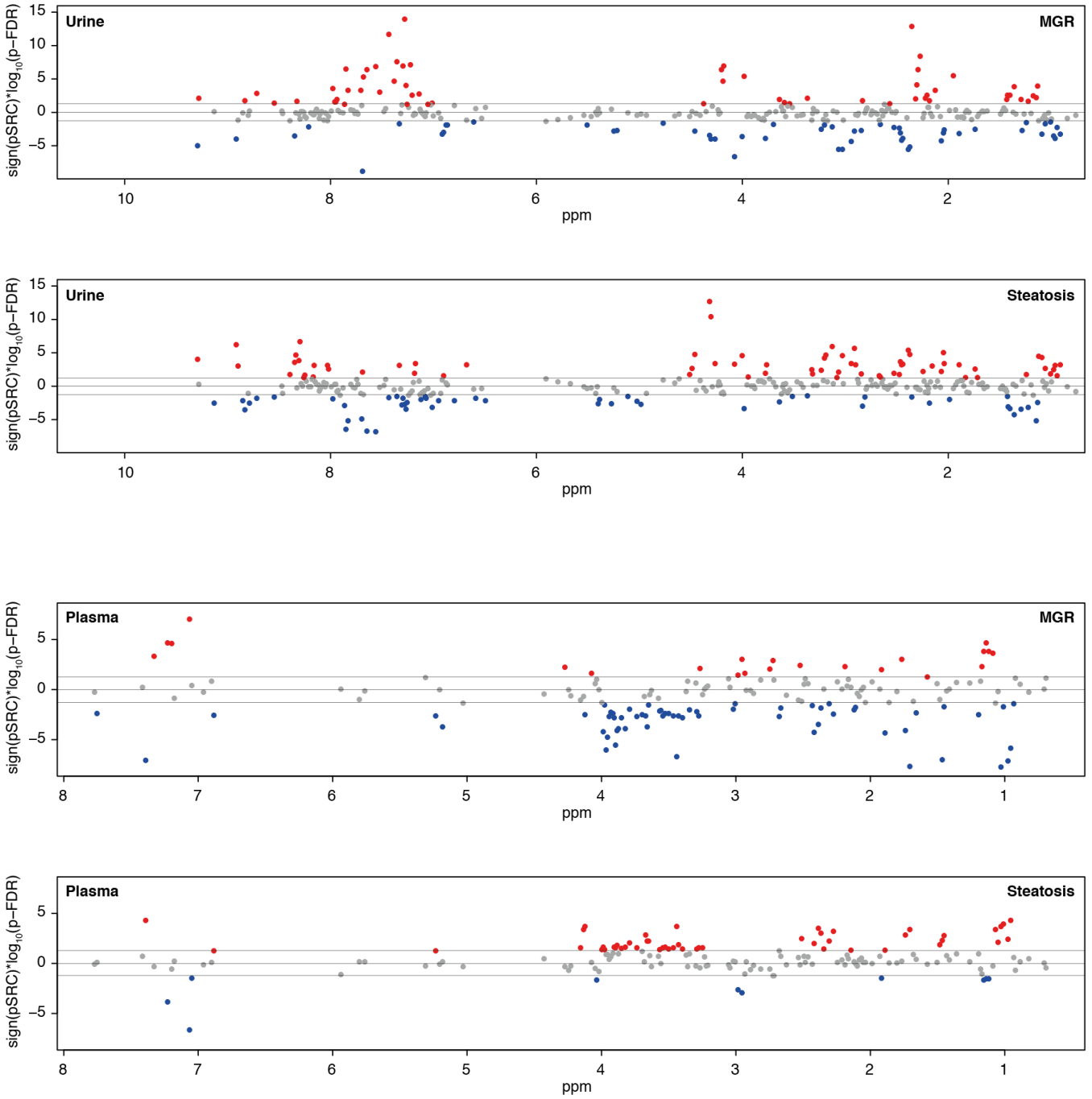
# Supplementary Fig. 6



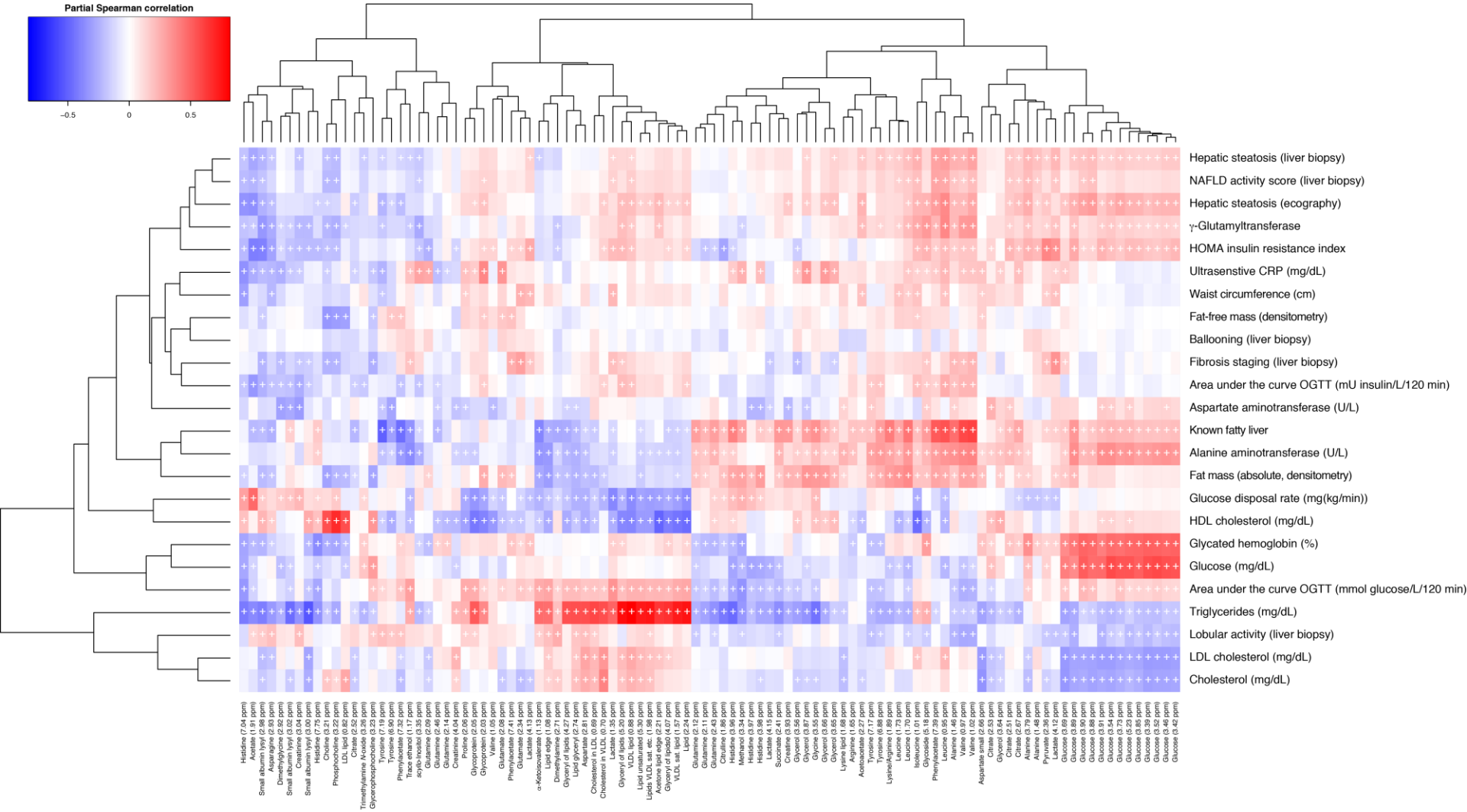
# Supplementary Fig. 7



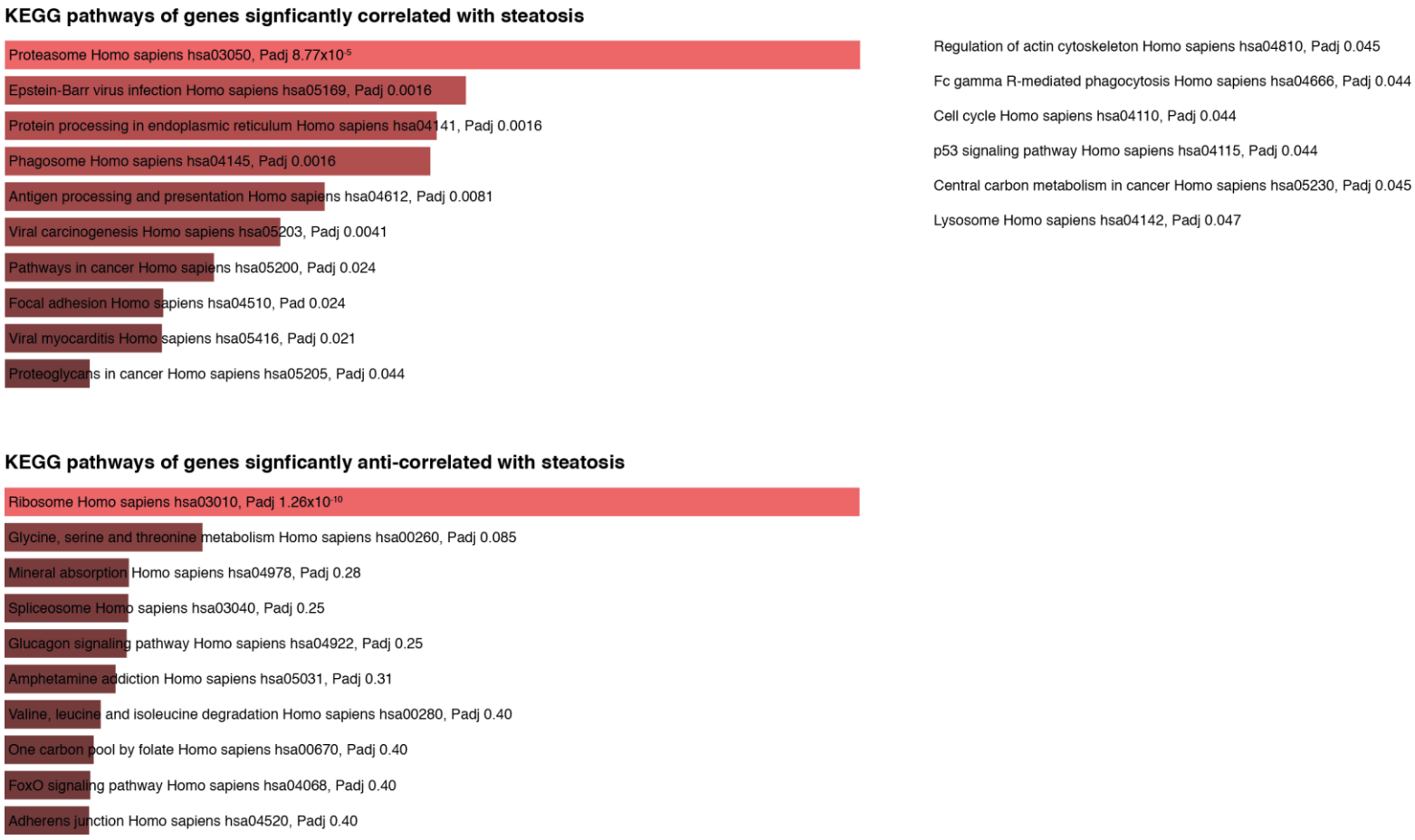
# Supplementary Fig. 8



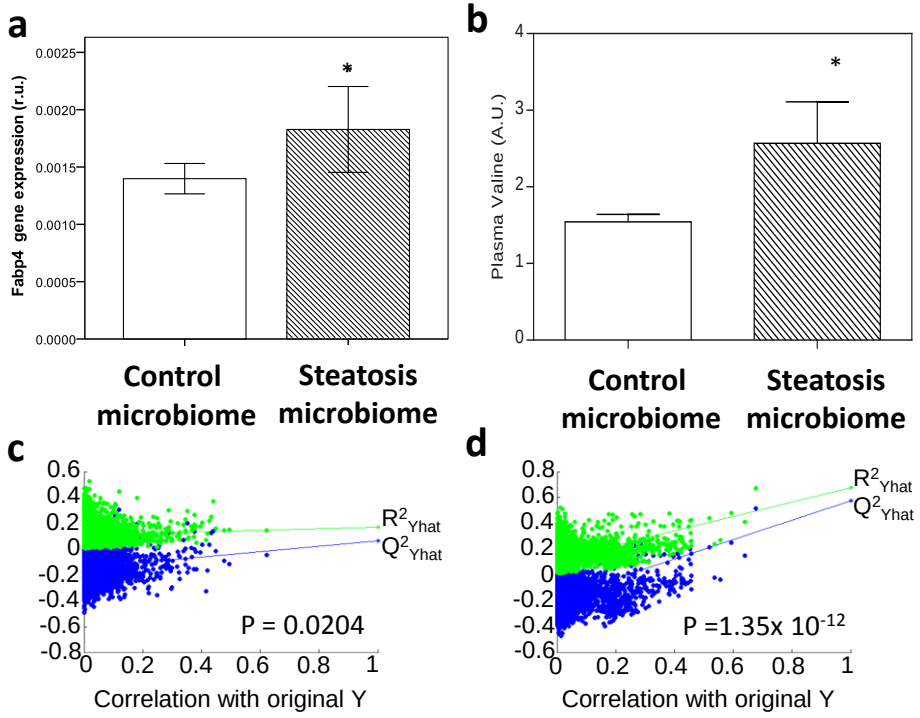
# Supplementary Fig. 9



# Supplementary Fig. 10



**Supplementary Fig. 11**



# Supplementary Fig. 12