Longitudinal multi-omics analyses of the gut–liver axis reveals metabolic dysregulation in hepatitis C infection and cirrhosis

A list of authors and their affiliations appears at the end of the paper

The gut and liver are connected via the portal vein, and this relationship, which includes the gut microbiome, is described as the gut–liver axis. Hepatitis C virus (HCV) can infect the liver and cause fibrosis with chronic infection. HCV has been associated with an altered gut microbiome; however, how these changes impact metabolism across the gut–liver axis and how this varies with disease severity and time is unclear. Here we used multi-omics analysis of portal and peripheral blood, faeces and liver tissue to characterize the gut–liver axis of patients with HCV across a fibrosis severity gradient before ($n = 29$) and 6 months after ($n = 23$) sustained virologic response, that is, no detection of the virus. Fatty acids were the major metabolites perturbed across the liver, portal vein and gut microbiome in HCV, especially in patients with cirrhosis. Decreased fatty acid degradation by hepatic peroxisomes and mitochondria was coupled with increased free fatty acid (FFA) influx to the liver via the portal vein. Metatranscriptomics indicated that *Anaerostipes hadrus*-mediated fatty acid synthesis influences portal FFAs. Both microbial fatty acid synthesis and portal FFAs were associated with enhanced hepatic fibrosis. *Bacteroides vulgatus*-mediated intestinal glycan breakdown was linked to portal glycan products, which in turn correlated with enhanced portal inflammation in HCV. Paired comparison of patient samples at both timepoints showed that hepatic metabolism, especially in peroxisomes, is persistently dysregulated in cirrhosis independently of the virus. Sustained virologic response was associated with a potential beneficial role for *Methanobrevibacter smithii*, which correlated with liver disease severity markers. These results develop our understanding of the gut–liver axis in HCV and non-HCV liver disease aetiologies and provide a foundation for future therapies.

The gut microbiome is an outsourcing of genes by the host to maximize calorie and environmental exploitation that directly influences host energy regulation, metabolism and immunity[1,2]. The liver is the nexus between the gut and the remainder of the host and is itself vulnerable to perturbations in this complex biology[3,4]. Alterations in the gut–liver axis have been well demonstrated in non-alcoholic fatty liver disease (NAFLD) and chronic liver disease from non-metabolic aetiologies[5]. Viral hepatitis including chronic hepatitis C virus (HCV)

is a major cause of liver disease that results in persistent inflammation and subsequent fibrosis in the liver. This hepatic fibrosis ultimately progresses to end-stage liver disease known as cirrhosis. Early stage of cirrhosis, referred to as 'compensated cirrhosis', is associated with intact hepatic function. Over time, this can progress to hepatic dysfunction referred to as 'decompensated cirrhosis'. It is in such patients with decompensated cirrhosis that alteration in the normal composition of intestinal microbial species, also known as gut dysbiosis, has been described[5]. In the past decade, clearance of HCV infection called sustained virologic response (SVR) has become a reality with the advent of direct-acting anti-viral therapy. However, cirrhosis persists in the immediate post-SVR period with a gradual regression of fibrosis over years. Thus, HCV-associated liver disease can serve as a model to delineate alterations in the gut–liver axis due to fibrosis, both with and without the presence of the initial trigger, that is, HCV. Lastly, to complete studies investigating the gut–liver axis, the portal vein should be interrogated as it is the most direct signalling conduit between the gut microbiome and the liver. Core aberrancies in these three biological compartments, their integration across liver disease severity, and changes over time lack detailed characterization.

In this Article, to this end, we have added insight into the role of the gut–liver axis in chronic liver disease by exploring the portal vein in patients with HCV across fibrosis severity at two timepoints, that is, in chronic HCV-associated compensated liver disease (HCVi) followed by re-evaluation approximately 6 months after HCV elimination (Fig. 1). To achieve this, we simultaneously collected blood samples from peripheral and portal veins, liver biopsies and faeces at the two timepoints, that is, before and after SVR. Specifically, combination therapy with sofosbuvir and velpatasvir, oral inhibitors of viral replication, was used to achieve SVR. We explored the metabolic interplay in the gut–liver axis by performing metabolomics on serum, RNA transcriptomics on liver and faeces, and microbial 16S ribosomal RNA analysis on faeces. The severity of hepatic fibrosis was measured on histopathological samples at both HCVi and SVR timepoints utilizing a scoring system derived by Ishak[6]. By utilizing a multilayered integrative analysis maximizing high-throughput molecular techniques we have characterized the biological landscape in the gut microbiome, the portal vein, the liver and their interactions across distinct stages of liver disease severity. This was a hypothesis-generating study utilizing HCV as a model to elucidate the role of the gut–liver axis in fibrosis rather than HCV pathogenesis.

Our comprehensive multi-omics analysis has revealed dysregulated energy homeostasis in the gut–liver axis with a predominant disruption in fatty acid (FA) metabolism (Extended Data Fig. 1). In the liver, metabolic alterations were localized to peroxisomes and mitochondria in HCVi and persisted in advanced fibrosis after SVR. In the microbiome, there was enhanced transcriptional activity of *Anaerostipes hadrus*-mediated FA synthesis and functional predominance of mucin-degrading *Bacteroides vulgatus* with worsening HCVi disease severity. By integrating the three biological compartments, we have shown how these hepatic and microbial metabolic perturbations directly relate to host inflammation in HCVi via immune and metabolic signals circulating in the portal vein.

Results

In total, 29 patients completed initial evaluation (HCVi) and 23 patients completed re-evaluation approximately 6 months after SVR, that is, on average 0.99 years (range 0.73–1.25 years) from sofosbuvir/velpatasvir initiation (Fig. 1). All patients underwent sampling of portal and peripheral blood, faeces and liver tissue at both timepoints for paired analysis. To elucidate perturbations in the gut–liver axis related to fibrosis, patients were stratified utilizing Ishak fibrosis scores[6] on liver biopsies from HCVi and SVR timepoints. Patients with Ishak fibrosis score '0–4' were characterized as non-cirrhotics and '5–6' as cirrhotics (Supplementary Table 1). There was no difference in fibrosis or direct portal pressure between HCVi and SVR (Extended Data Fig. 2).

Host metabolism is downregulated in HCVi and fibrosis

Transcriptomics was performed on paired liver biopsies from HCVi and SVR that identified 7,866 differentially expressed genes (DEGs) in HCVi compared with SVR (DeSeq2, false discovery rate (FDR) *P* value <0.1). Over-representation analysis was performed on DEGs using Kyoto Encyclopedia of Genes and Genomes (KEGG) (FDR *P* value <0.1) (Fig. 2a). As expected, HCVi showed upregulated hepatic inflammatory and anti-viral pathways enriched in interferon genes as well as enhanced hepatic expression of *IFNG* and *IFNL1* compared with SVR (Source Data Fig. 2). Intriguingly, every pathway downregulated in HCVi was metabolic, including metabolism of FAs, branched-chain amino acids, aromatic amino acids, peroxisomes and peroxisome proliferator-activated receptor (PPAR) signalling. When analysed using Gene Ontology (GO) by cellular component, genes downregulated in HCVi localized to peroxisomes and mitochondria, organelles fundamental for energy and redox balance[7] (over-representation analysis, Fisher's exact test; FDR *P* value <0.1) (Fig. 2b). Mitochondrial and peroxisomal dysfunction in FA oxidation has been shown across liver disease aetiologies, including HCV[8,9]. Reduced FA degradation (FoldEnrichment −5.23, FDR *P* value <0.0001), peroxisome FA oxidation (for example, *ACOX* FoldChange −0.67, FDR *P* value <0.0001), catalase (*CAT*) (FoldChange −0.66, FDR *P* value <0.0001) and retinol metabolism (*DHRS4* FoldChange −0.32, FDR *P* value 0.029) in HCVi exemplifies peroxisomes' role in redox imbalance in liver disease[10,11] (Fig. 2c and Extended Data Fig. 3).

We next explored if hepatic metabolic dysfunction translates into circulatory changes. Compared with SVR, HCVi showed elevated peripheral and portal levels of a wide range of metabolites, many of which belonged to categories with decreased hepatic degradation (FDR *P* value <0.1) (Fig. 2d,e). Increased levels of such metabolites alongside decreased hepatic processing could have negative consequences in HCVi from impaired energy extraction and/or excess of metabolically active compounds with implications for insulin resistance, onco-metabolism, hepatic steatosis and sarcopenia[12–15].

Lastly, we assessed whether hepatic aberrations in HCV are related to fibrosis and persist in SVR fibrosis. Hepatic genes and portal metabolites co-expressed in both HCVi and SVR and linked to liver disease severity independent of HCV were explored using consensus weighted gene correlation network analysis (WGCNA) on paired HCVi and SVR samples. Only one hepatic module 'MEred' inversely correlated with fibrosis (Pearson correlation co-efficient −0.47, unadjusted *P* value 0.02) and only one portal metabolite module 'MEyellow' positively correlated with direct portal pressure (Pearson correlation co-efficient 0.60, unadjusted *P* value 0.002) in both HCVi and SVR (Fig. 3a,b). Hepatic 'MEred' module was enriched in metabolism notably FA degradation (FoldEnrichment 8.87, FDR *P* value <0.0001), PPAR signalling (FoldEnrichment 4.78, FDR *P* value 0.0013) and peroxisomes (FoldEnrichment 3.84, FDR *P* value 0.010); and portal 'MEyellow' module was most enriched in free fatty acids (FFAs) (Fig. 3c,d). These findings expand on the current knowledge of serum metabolic alterations in HCV and SVR[16].

In summary, our results extend knowledge of downregulated hepatic mitochondrial and peroxisomal FA catabolism and PPAR signalling in HCV cirrhosis[9]. Many of these metabolic concepts, especially related to FA metabolism and peroxisome function, remain perturbed in advanced fibrosis after SVR.

Portal metabolites and gut microbiome in HCVi liver disease

Our findings of hepatic metabolic dysregulation accentuated in fibrosis were an opportunity to explore the gut–liver axis in fibrosis. As the portal vein is the major conduit for metabolic signals between the gut microbiome and the liver, we explored portal vein metabolomics and the gut microbiome across liver disease severity within HCVi. First, serum metabolic signature in HCVi was analysed in relation to fibrosis using similarity network fusion (SNF) and spectral clustering analysis on 1,541 metabolites measured in portal and peripheral serum[17] (Fig. 4a). There was a significant difference in mean Ishak fibrosis scores of the
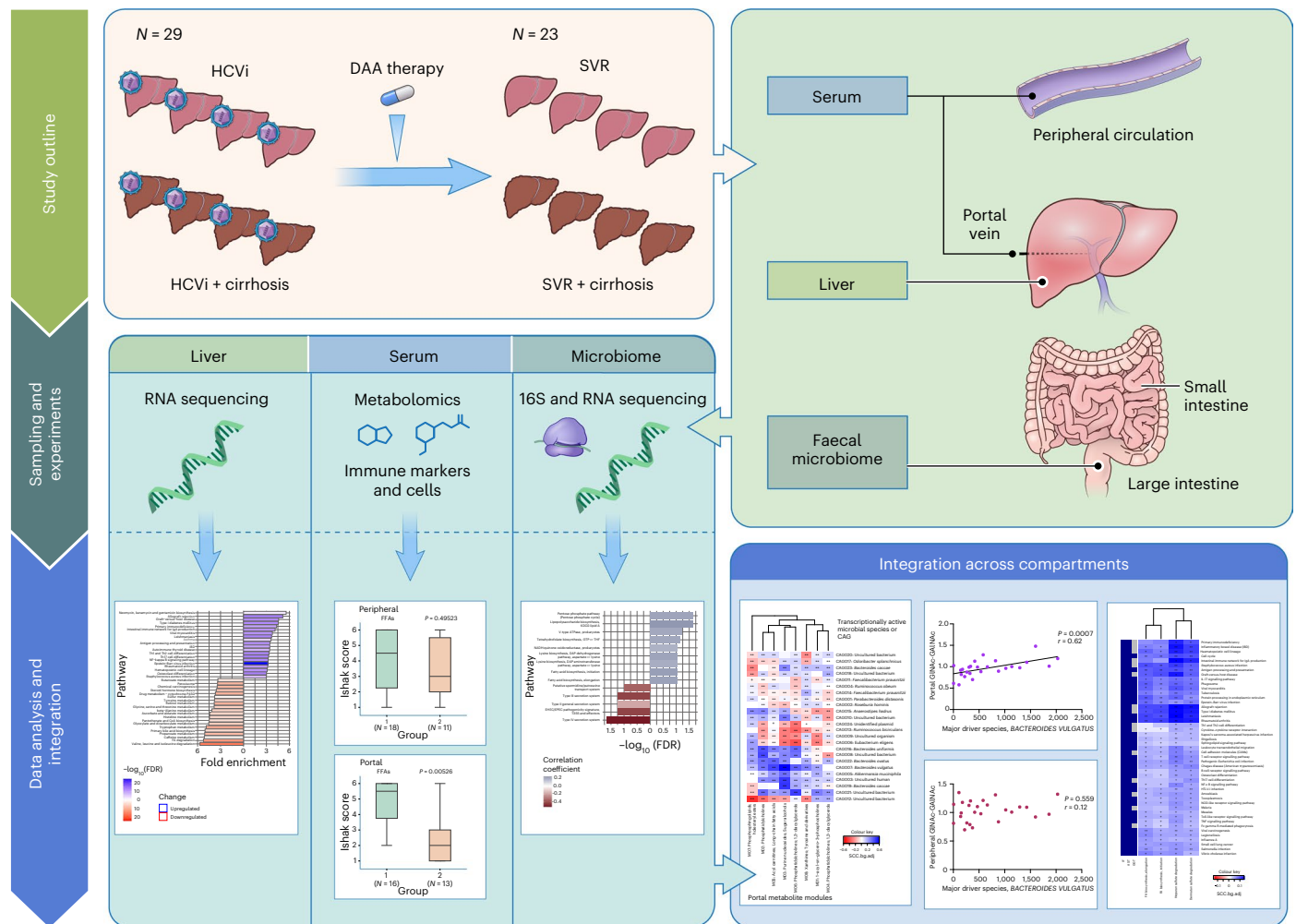
Fig. 1 | Study design. Overview of the clinical protocol with two-timepoint analysis, that is, HCVi and 6 months following SVR. At each timepoint, samples were obtained from peripheral serum, portal serum, liver biopsy and faeces. Experimental procedures, analysis and integration were carried out as described in Methods. DAA, direct-acting anti-viral therapy.

patient groups derived from portal metabolomics SNF ($P = 0.0066$), but with no difference based on peripheral metabolomics SNF ($P = 0.23$). This was validated using logistic regression modelling (Supplementary Table 6). Spectral clustering was then performed on 25 HCVi patient networks generated from individual super-pathways (Supplementary Table 7). FFA was the only super-pathway where HCVi patient groups showed a significant difference in mean Ishak fibrosis scores based on spectral clustering of portal ($P = 0.005$), but not peripheral FFA ($P = 0.49$) (Fig. 4b). This was validated using logistic regression (Supplementary Table 6).

Unbiased clustering of patients with HCVi into early and advanced fibrosis based on only portal metabolomic profile encouraged investigation of gut microbial composition and transcriptional activity in HCVi. Faecal 16S rRNA analysis showed no significant relationship between microbial phyla or genera and fibrosis in HCVi (Extended Data Fig. 4a,b and Supplementary Fig. 1a,b). Next, we explored changes in transcriptional activity that may precede compositional alterations by examining HCVi faecal metatranscriptomics. Microbial RNA sequences were aligned to MetaHIT Consortium, 889,668 nucleotide sequences were captured, 4,718 microbial KEGG Orthology (KO) genes were annotated and KO genes were grouped into KEGG functional modules correlated with liver disease severity markers using background-adjusted median Spearman correlation ($SCC_{bg.adj.}$)[18]. This unbiased exploration of microbial function revealed a direct association of fibrosis and hepatic necroinflammation with microbial energy metabolism including FAs,

amino acids, and glycans in HCVi (Fig. 4c, Extended Data Fig. 5a, b) (FDR P value <0.1). Specifically, fibrosis positively correlated with microbial FA biosynthesis, initiation ($SCC_{bg.adj.}$ 0.22, FDR P value 0.092) and elongation ($SCC_{bg.adj.}$ 0.21, FDR P value 0.092), but not with β-oxidation ($SCC_{bg.adj.}$ −0.11, FDR P value 0.45). Hepatic necroinflammation positively correlated with microbial glycan degradation; heparan sulfate (aspartate aminotransferase (AST), $SCC_{bg.adj.}$ 0.25, FDR P value 0.040; γ-glutamyl transferase (GGT), $SCC_{bg.adj.}$ 0.21, FDR P value 0.059) and dermatan sulfate (GGT, $SCC_{bg.adj.}$ 0.28, FDR P value 0.094) degradation.

To determine transcriptionally active microbial species, de novo assembly was performed on the microbial metatranscriptome by co-abundance clustering using MGS canopy algorithm[19]. Taxonomically annotated co-abundance gene groups (CAGs) were analysed using leave-one-out analysis[18]. Fibrosis-associated microbial 'FA biosynthesis initiation' was driven by *CAG003: uncultured bacterium* followed by *CAG015: A. hadrus*, a FA-synthesizing gut commensal[20,21](Supplementary Table 2). Similarly, GGT-associated microbial 'heparan sulfate degradation' was predominantly driven by *CAG007: B. vulgatus*, a glycan-metabolizing gut commensal[22].

Microbial function linked to portal metabolites in HCVi

To elucidate an influence of microbial metabolism on the host via portal circulation, portal metabolite WGCNA modules were correlated with transcriptionally active microbial CAGs ($SCC_{bg.adj.}$, FDR P value <0.1) (Fig. 5a). The major transcriptionally active microbial species for FA
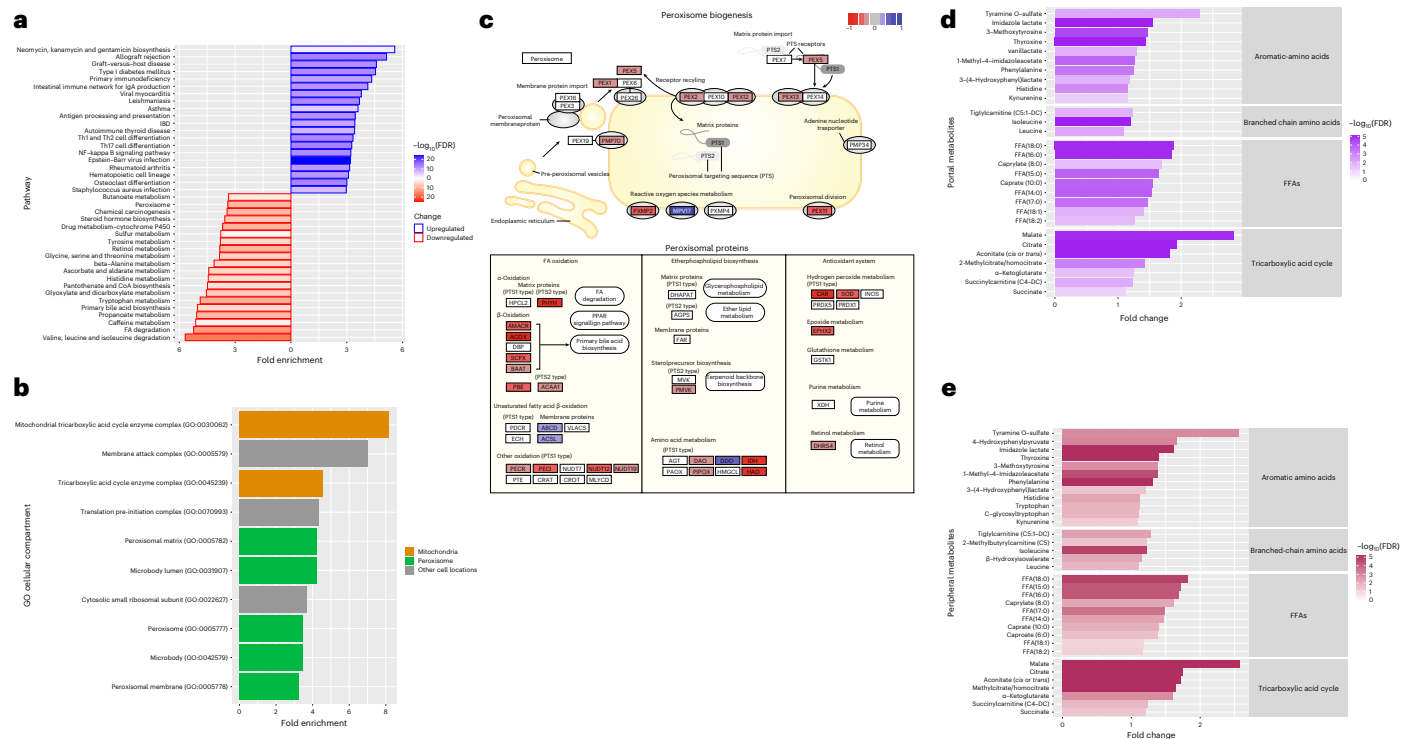
Fig. 2 | Longitudinal evaluation revealed reduced hepatic metabolism in peroxisomes and mitochondria alongside higher circulatory levels of corresponding metabolites in HCVi compared with SVR. a, Over-representation analysis on the hepatic DEGs with FDR *P* value <0.1 in HCVi compared with SVR mapped to the KEGG pathway database using NetworkAnalyst software. Visualized are the 20 most enriched hepatic KEGG pathways, FDR *P* value <0.1. Bars shaded blue represent hepatic pathways enriched in DEGs upregulated in HCVi, and bars shaded orange represent pathways enriched in DEGs downregulated in HCVi compared with SVR (*n* = 22). **b**, Predominant cellular location of the 2,380 DEGs downregulated in HCVi with FDR *P* value <0.1, identified using GO Cellular Component Database. Visualized are the ten most enriched cellular locations for hepatic DEGs decreased in HCVi compared with SVR, FDR *P* value <0.1 (*n* = 22). **c**, Detailed illustration using GAGE R of hepatic DEGs in the KEGG functional pathway 'Peroxisome' downregulated in HCVi compared with SVR, FDR *P* value <0.1. DEGs highlighted blue for fold change >0 (that is, increased in HCVi) and red for fold change <0 (that is, decreased in HCVi) compared with SVR (*n* = 22). **d,e**, Paired comparison of alterations in serum metabolite levels between HCVi and SVR (two-sided Wilcoxon matched-pairs signed-rank test, and FDR *P* value <0.1). Visualization of representative portal (**d**) and peripheral (**e**) metabolites elevated in HCVi compared with SVR belonging to categories with downregulated hepatic metabolism (*n* = 23).

biosynthesis *CAG015: A. hadrus* showed strongest association with portal modules M02 (SCC$_{bg.adj.}$ 0.28, FDR *P* value 0) and M07 (SCC$_{bg.adj.}$ 0.27, FDR *P* value 0) that predominantly contained complex lipids and FFAs. Microbial contribution to portal FFA was reinforced from the direct associations of portal FFA with *CAG015: A. hadrus* (for example, portal palmitate, Spearman correlation co-efficient 0.50, unadjusted *P* value 0.0084) and with KO genes for microbial FA biosynthesis (Supplementary Table 3). Of note, FFA associations with *CAG015: A. hadrus* were limited to long-chain FFA, not short-chain FFA. This was interesting as *A. hadrus* is known to influence host health through short-chain FA synthesis, and to our knowledge there is no pre-existing literature on its role in long-chain FA synthesis[20,21]. Our findings support a role for *A. hadrus*-mediated FA biosynthesis on the availability of long-chain FFAs in portal circulation. Furthermore, strong correlations of long-chain FFAs with liver enzymes, pro-inflammatory cytokines and macrophage activation only in portal vein (for example, portal FFA16:0 with AST, Spearman correlation co-efficient 0.61, unadjusted *P* value 0.0003; IL-8, Spearman correlation co-efficient 0.66, unadjusted *P* value ≤0.0001) suggests pathological implications of microbially derived long-chain FFAs in HCVi liver disease progression (Fig. 5b).

In HCVi, the transcriptionally active species for mucosal glycan degradation *CAG007: B. vulgatus* was the major driver for most disease associated microbial functions (Fig. 4d). This was important as microbial glycan metabolism influences intestinal homeostasis and inflammation[23,24]. During calorie imbalance, *B. vulgatus* shifts energy extraction to mucosal glycans, and by degrading mucin

it becomes a pathobiont causing barrier dysfunction and inflammation, best studied in inflammatory bowel disease (IBD)[22,25]. Our findings highlight this concept in HCVi liver disease with hepatic metabolic dysfunction akin to 'calorie imbalance'. Inferring translocation from peripheral blood limits accurate appraisal of gut-derived signals before hepatic processing. Direct evaluation of portal signals uncovered a possible contribution of *B. vulgatus* to portal glycan products and a pro-inflammatory role of portal glycan products, concepts not fully explored in liver disease. Transcriptionally active *CAG007: B. vulgatus* showed the strongest correlation with portal module M03 (SCC$_{bg.adj.}$ 0.60, FDR *P* value 0) containing diverse glycans including *N*-acetylglucosamine/*N*-acetylgalactosamine (GlNAc-GalNAc) and *N*-acetylneuraminic acid/sialic acid (Fig. 5a). *CAG007: B. vulgatus* directly correlated with GlNAc-GalNAc only in portal (Spearman correlation co-efficient 0.62, unadjusted *P* value 0.0007), but not peripheral serum (Spearman correlation co-efficient 0.12, unadjusted *P* value 0.559) (Fig. 5c). Only in portal serum was GlNAc-GalNAc elevated in HCVi-Cir compared with HCVi-NC (Fig. 5d). Both GlNAc-GalNAc and sialic acid strongly correlated with markers of inflammation, including AST, alkaline phosphatase (ALP), GGT, sCD14, sCD163, IL-6, IL-8, TNFα, CXCL9 and HMGB1 (for example, portal GlNAc-GalNAc with TNFα, Spearman correlation co-efficient 0.63, unadjusted *P* value 0.0003); CD4 and CD8 T cells, intestinal α4β7 T cells; and an unfavourable lipid profile (Supplementary Table 4). Negative consequences of *B. vulgatus*-mediated glycan degradation were supported by elevated
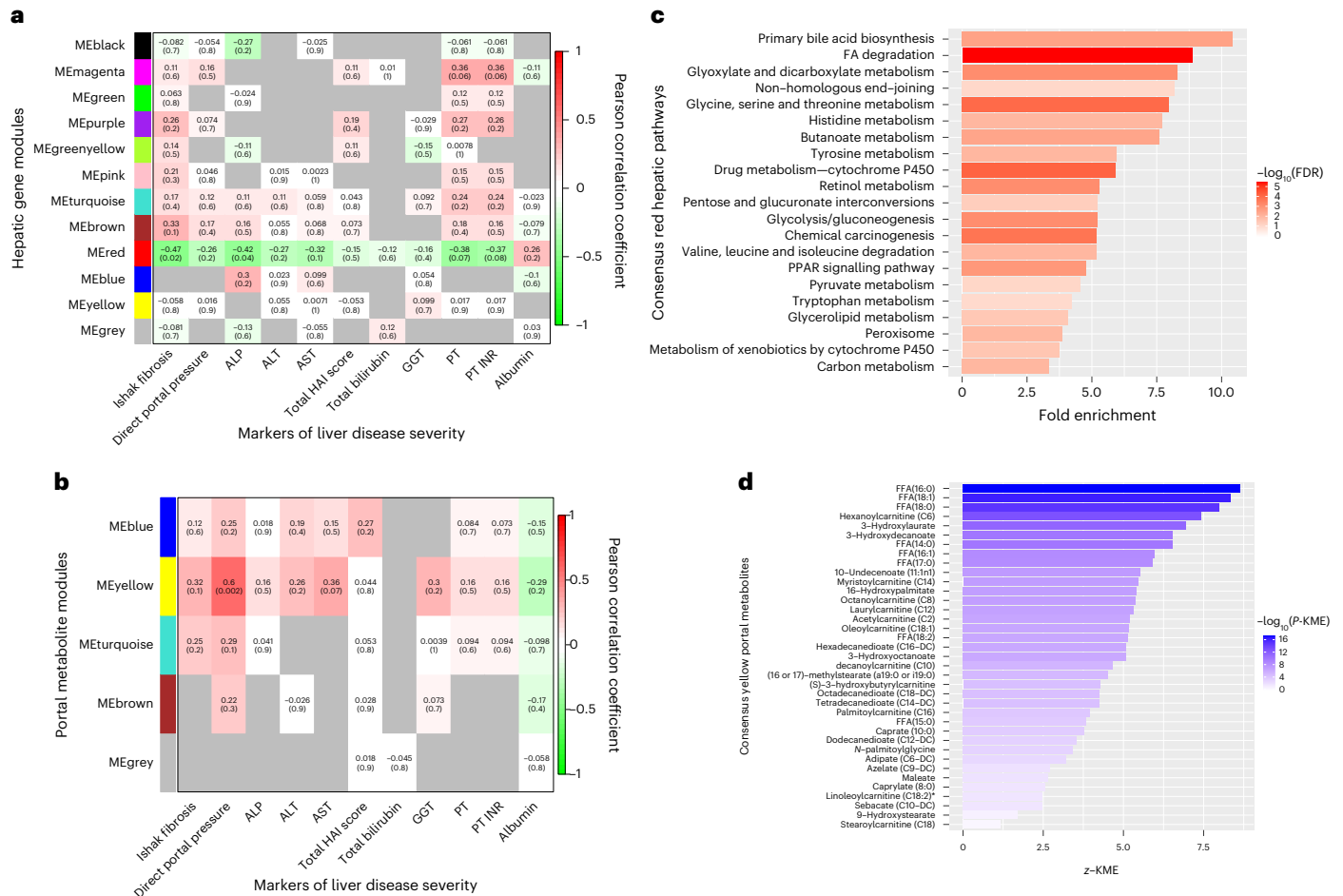
Fig. 3 | Fibrosis was linked to decreased hepatic FA metabolism, and direct portal pressure to increased portal FFAs independent of HCV, n = 22.
a, Correlation heat map of liver disease severity markers (*x* axis) and consensus WGCNA hepatic gene modules preserved at both timepoints, that is, hepatic genes co-expressed in both HCVi and SVR (*y* axis). 'ME' is a module or cluster of genes that are co-expressed, and each ME module is randomly assigned a colour. Within each cell, the upper row indicates the Pearson correlation co-efficient and the lower row with parenthesis the two-sided Fisher's asymptotic unadjusted *P* value. Cells greyed out indicate no consensus formed owing to opposite directions for Pearson correlation co-efficient in HCVi and SVR. Only one hepatic module 'MEred' inversely correlated with fibrosis in both HCVi

and SVR. **b**, Correlation heat map of liver disease severity markers (*x* axis) and consensus WGCNA portal metabolite modules preserved in both HCVi and SVR (*y* axis) as in Fig. 3a. 'ME' is a module or cluster of metabolites that are most strongly correlated with each other, and each ME module is randomly assigned a colour. Only one portal module 'MEyellow' positively correlated with direct portal pressure in both HCVi and SVR. **c**, Over-representation analysis of hepatic genes in consensus WGCNA module 'MEred' (NetworkAnalyst, FDR *P* value <0.1). **d**, Individual FFA contained in the consensus WGCNA module 'MEyellow.' *z*-KME indicates *z* score for KME (eigengene-based connectivity or metabolite module membership), and *P*-KME indicates two-sided unadjusted *P* value calculated for each KME value.

markers of gut dysbiosis (IL-18) and intestinal dysfunction (zonulin) in HCVi compared with SVR (Extended Data Fig. 6). Although cor-relative, exploring metabolic and immune markers in portal vein has revealed how *B. vulgatus*-mediated glycan degradation may impair intestinal homeostasis in HCVi liver disease[22]. Portal glycan products and their correlates of gut permeability, dysbiosis, inflammatory signals and α4β7 T cells are emerging therapeutic targets in IBD; our findings suggest therapeutic themes in hepatic disorders[26,27].

We explored a possible relationship between disease-associated microbial functions and hepatic immune homeostasis on a transcrip-tional level. This inter-omics integration showed direct associations of microbial FA biosynthesis and intestinal glycan degradation with multiple hepatic pathways for inflammation, immunity and dysbio-sis (SCC$_{bg.adj.}$, FDR *P* value <0.1) (Fig. 5e). Extra-hepatic FFAs without hepatic lipid oxidation can potentiate oxidative stress and inflam-mation in liver disease, a concept most studied in NAFLD but also in HCV[8,28]. We speculate that, in patients with HCVi with advanced fibro-sis, higher portal long-chain FFA derived from *A. hadrus* potentiates

hepatic injury due to impaired hepatic mitochondrial and peroxisomal function. Such a pathological role of microbial long-chain FFA is not well elucidated[29–32].

Insights from re-evaluating the gut microbiome after SVR
Re-evaluating the gut microbiome after SVR uncovered a putative role of *Methanobrevibacter smithii*, dominant archaea for methane metab-olism, in cirrhosis without HCV. Co-expressed microbial functional KO gene modules were correlated with liver disease severity markers after SVR (WGNCA) (Fig. 6a). Only 'MEGreen' inversely correlated with liver disease severity (fibrosis, Pearson correlation co-efficient −0.49, unadjusted *P* value 0.02; direct portal pressure, Pearson cor-relation co-efficient −0.59, unadjusted *P* value 0.003). This module was most enriched in methane metabolism (over-representation analysis, FoldEnrichment 7.79, FDR *P* value <0.0001) (Fig. 6b). Of 18 KO genes in 'methane metabolism', 16 KO genes taxonomically anno-tated to *M. smithii* using BLASTN >99% identity and >95% coverage (Supplementary Table 8).
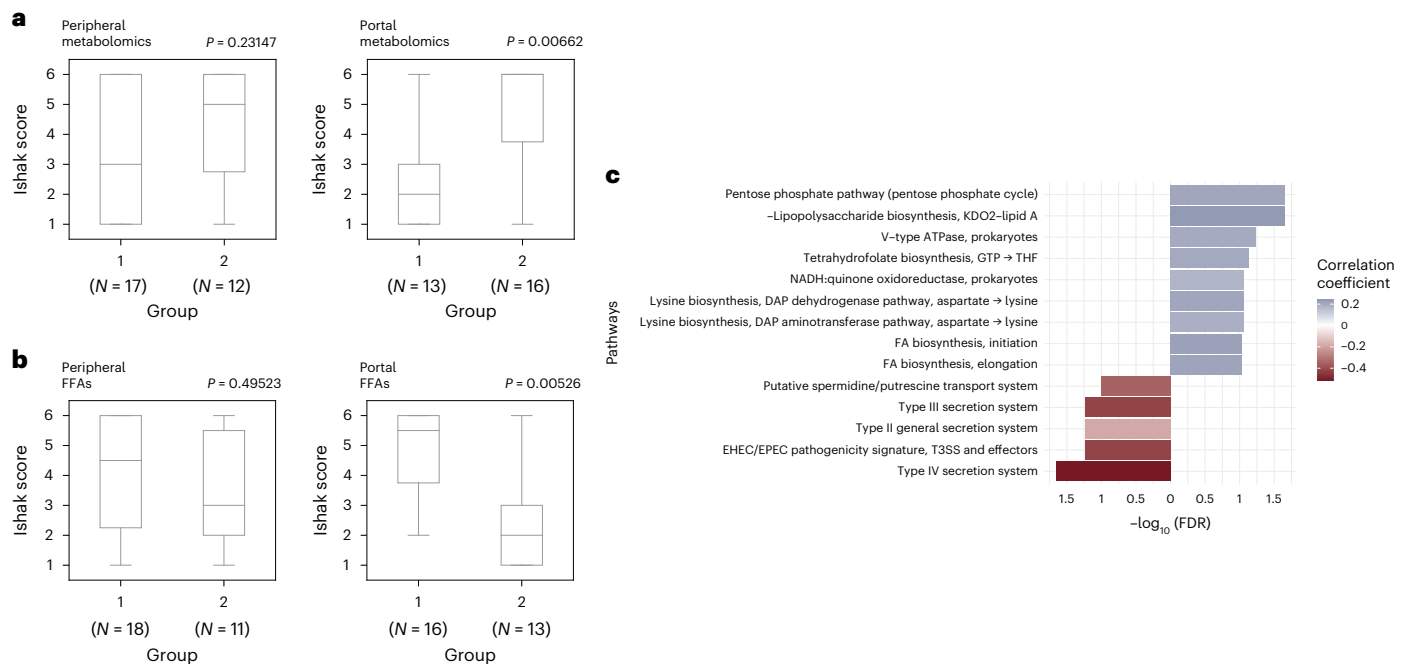
Fig. 4 | Only portal metabolomics clustered patients with HCVi into early and advanced fibrosis. HCVi disease severity was linked to increased microbial FA synthesis and glycan degradation driven by transcriptionally active *A. hadrus* and *B. vulgatus*. **a**, SNF and spectral clustering performed on 1,541 metabolites in HCVi. SNF from portal (not peripheral) metabolomics clustered patients with HCVi into groups with significantly different mean Ishak fibrosis scores (SNF, Welch's *t*-test, unadjusted two-sided *P* value) (*n* = 29). Peripheral metabolomics group 1: minimum and lower bound (25th percentile) 1, median (50th percentile) 3, maximum and upper bound (75th percentile) 6, no outliers; peripheral metabolomics group 2: minimum 1, lower bound (25th percentile) 2.75, median (50th percentile) 5, maximum and upper bound (75th percentile) 6, no outliers; portal metabolomics group 1: minimum and lower bound (25th percentile) 1, median (50th percentile) 2, maximum 6, upper bound (75th percentile) 3, no outliers; portal metabolites group 2: minimum 1, lower bound (25th percentile) 3.75, median (50th percentile) 6, maximum and upper bound (75th percentile) 6, no outliers. **b**, Within major metabolic categories, the HCVi patient network derived from portal (not peripheral) FFAs could cluster patients with HCVi into groups with significantly different mean Ishak fibrosis scores (spectral clustering, Welch's *t*-test, unadjusted two-sided *P* value) (*n* = 29). Peripheral FFA group 1: minimum 1, lower bound (25th percentile) 2.25, median (50th percentile) 4.5, maximum and upper bound (75th percentile) 6, no outliers; peripheral FFA group 2: minimum 1, lower bound (25th percentile) 2, median (50th percentile) 3, maximum 6, upper bound (75th percentile) 5.5, no outliers; portal FFA group 1: minimum 2, lower bound (25th percentile) 3.75, median (50th percentile) 5.5, maximum and upper bound (75th percentile) 6, no outliers; portal FFA group 2: minimum and lower bound (25th percentile) 1, median (50th percentile) 2, maximum 6, upper bound (75th percentile) and 3, no outliers. **c**, Correlation of microbial KEGG functional modules with hepatic fibrosis in HCVi using SCC$_{bg.adj.}$, two-sided FDR *P* value <0.1 (*n* = 26).

M. smithii is critical for intestinal energy homeostasis. By eliminating hydrogen it ensures glycan fermentation by saccharolytic bacteria maximizing energy harvest[33,34]. Relevance of inter-microbial linkages of *Methanobrevibacter* in SVR was exemplified by gut ecology analysis on 16S rRNA using Sparse Correlations for Compositional data (SparCC)[35] (two-sided pseudo *P* value <0.05) (Fig. 6c). Most significant taxonomic linkages of *Methanobrevibacter* genus were with saccharolytic *Clostridiales*. A beneficial role for *Methanobrevibacter* was also suggested by compelling anti-inflammatory associations of *Methanobrevibacter* 16S rRNA genus abundance and *M. smithii* transcriptional activity with liver enzymes, cytokines and an unfavourable lipid profile in SVR (Supplementary Table 5).

We speculate that decreased *M. smithii* methane metabolism in advanced fibrosis after SVR alters the luminal pH, saccharolytic gut commensals, short-chain FFA availability and ultimately intestinal homeostasis; concepts with profound biological implications[21].

Discussion

A comprehensive multi-omics integration including the major conduit for host–microbiome crosstalk, the portal vein, revealed energy metabolism in particular FA metabolism as the fundamental disturbance in the gut-liver axis in HCVi. Temporal study design (Fig. 1) showed that cirrhosis is a persistent state of metabolic dysregulation centred in peroxisomes even after SVR. We have laid out an atlas for liver disease alterations and uncovered authentic human pathophysiology.

Chronic inflammation is an energy-demanding state with a profound impact on metabolism when occurring in the central metabolic organ, the liver[36,37]. Hepatic metabolic dysregulation, particularly mitochondrial FA metabolism, has been investigated across liver disease aetiologies[38]. Mitochondrial dysfunction impedes energy extraction from FAs and promotes intra-hepatic lipid accumulation, redox imbalance and inflammasome activation[8]. Our findings echo this with decreased hepatic mitochondrial FA catabolism in HCVi accentuated in cirrhosis. However, as important is the identification of peroxisomes as a major site of disrupted FA metabolism with worsening fibrosis in HCVi and SVR. Peroxisomes and mitochondria are co-dependent, and despite peroxisomes' role in energy, lipid metabolism and redox balance, they remain under-appreciated in liver disease[7,39,40]. Peroxisome dysfunction in fibrosis may lead to FA-induced hepatic oxidative stress as PPAR signalling is crucial for hepatic FA utilization[41]. Therapeutic relevance of this concept is supported by use of PPAR agonists across liver disease aetiologies, including NAFLD and primary biliary cholangitis[42,43]. Thus, even though HCV can now be cured, our study adds biological insight into metabolically targeted therapies, highlights the importance of treating HCVi at earlier stages and offers a provocative therapeutic approach of peroxisome rescue in SVR cirrhosis.

To our knowledge, portal metabolites have been studied only in decompensated cirrhosis[44–46]. Characterization of the portal metabolomic landscape in compensated liver disease was a critical element of our study. Unlike peripheral serum, portal metabolomics has
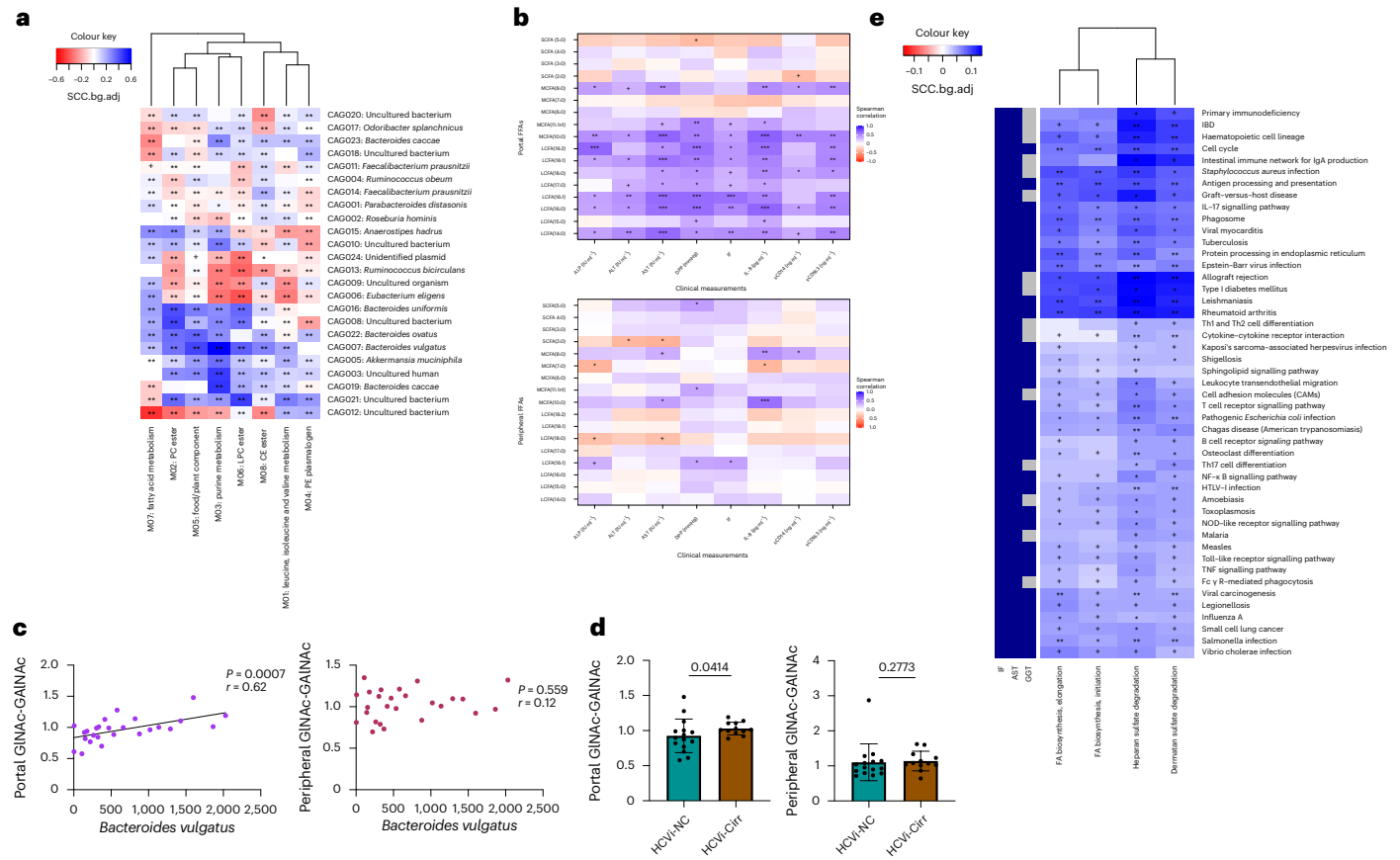
Fig. 5 | In HCVi, transcriptionally active *A. hadrus* and *B. vulgatus* were directly linked to portal FFA and glycan products, respectively. Microbially derived portal signals and disease-associated microbial functions correlated with enhanced circulatory and hepatic pathways of inflammation.
a, Correlation heat map of inter-omics associations between transcriptionally active microbial species (CAGs) and portal metabolite WGCNA modules in HCVi. $SCC_{bg.adj}$, Mann–Whitney U test, two-sided, FDR P value +, 0.05–0.1, *0.05–0.01, **0.01–0.001, ***<0.0001) ($n = 26$). Portal modules are numbered M01 through M08, annotated for most prevalent metabolic subpathway. PC, phosphatidylcholine; LPC, lysophosphatidylcholine; CE, ceramide; PE, phophatidylethanolamine. **b**, Correlation heat map between portal (upper) and peripheral (lower) serum FFAs and markers of liver disease severity and inflammation in HCVi (Spearman R, two-sided FDR P value +, 0.05–0.1,

*0.05–0.01, **0.01–0.001, ***<0.0001) ($n = 29$). SCFA, short-chain FFA; MCFA, medium-chain FFA; LCFA, long-chain FFA; DPP, direct portal pressure; IF, Ishak fibrosis score. **c**, Transcriptionally active *CAG007: B. vulgatus* correlated with portal (left) and peripheral (right) N-acetylglucosamine-galactosamine (GlNAc-GalNAc) (Spearman correlation co-efficient, two-sided) ($n = 29$). **d**, Compared with HCVi-NC, GlNAc-GalNAc was higher in HCVi-Cirr but only in portal serum (two-sided Mann–Whitney, HCVi-Cirr $n = 13$ versus HCVi-NC $n = 16$). Scatter plots with bars, data are presented as median ± IQR. **e**, Correlation heat map of inter-omics associations between microbial KEGG functional modules (x axis) and hepatic KEGG pathways (y axis). Displayed are only hepatic pathways and microbial functional modules that positively correlated with Ishak fibrosis score (IF), AST and GGT (FDR P value <0.1) as indicated by blue column on the left y axis ($SCC_{bg.adj}$, Mann–Whitney U test, two-sided FDR P value as per **a**) ($n = 26$).

highlighted that in HCVi fibrosis higher gut-derived portal FFA may in fact accelerate liver injury, compounded with decreased hepatic mitochondrial and peroxisomal FA metabolism.

Studies on the gut microbiome in liver disease have combined aetiologies and inferred function from composition, and only few re-evaluated after eliminating the disease trigger[47,48]. Our study addressed these deficits by focusing on one aetiology, exploring microbial metatranscriptome and re-evaluating after SVR. Furthermore, associations between transcriptionally active microbial species and portal vein signals were mapped. Despite lack of association with microbial composition, distinct microbial functions were associated with HCVi fibrosis. Microbial FA synthesis was increased with fibrosis in HCVi as there was reduced hepatic FA metabolism. Transcriptionally active *A. hadrus* was responsible for this paradoxical rise in microbial FA synthesis. *A. hadrus* impacts human health via short-chain FA synthesis; however, our inter-omics analysis suggests a direct contribution of *A. hadrus* to host long-chain FFA[20,21]. This underscores the yet undiscovered aspects of gut microbiome's influence on human pathophysiology.

Contrary to our findings of a positive association between fibrosis and microbial FA synthesis, previous studies suggest a beneficial role of microbial long-chain FA in liver and intestinal injury[31,32]. This contrast may be due to difference in species, acuity and nature of injury, and sampling luminal, not portal, FAs. Another reason may be the context-specific nature of microbiome's influence on host. For example, *Lachnospiraceae*-mediated lipid metabolism is beneficial for colonic health at a 'local level' but paradoxically harmful in hepatic dysfunction, that is, in IBD with and without primary sclerosing cholangitis[49]. Given the pathological role of extra-hepatic FFA in liver disease, we suggest an opportunity for therapeutic manipulation of *A. hadrus* and microbial FA synthesis to reduce hepatic oxidative injury and ultimately slow fibrosis progression[8,28].

We have identified *B. vulgatus* as the major transcriptionally active species responsible for most microbial functions linked to HCVi liver disease severity. *B. vulgatus* can become a mucin degrader during calorie imbalance and mediate gut inflammation and dysbiosis, best studied in IBD[22–26]. In HCVi, *B. vulgatus* was not only the major functionally active species for mucosal glycan degradation but directly linked
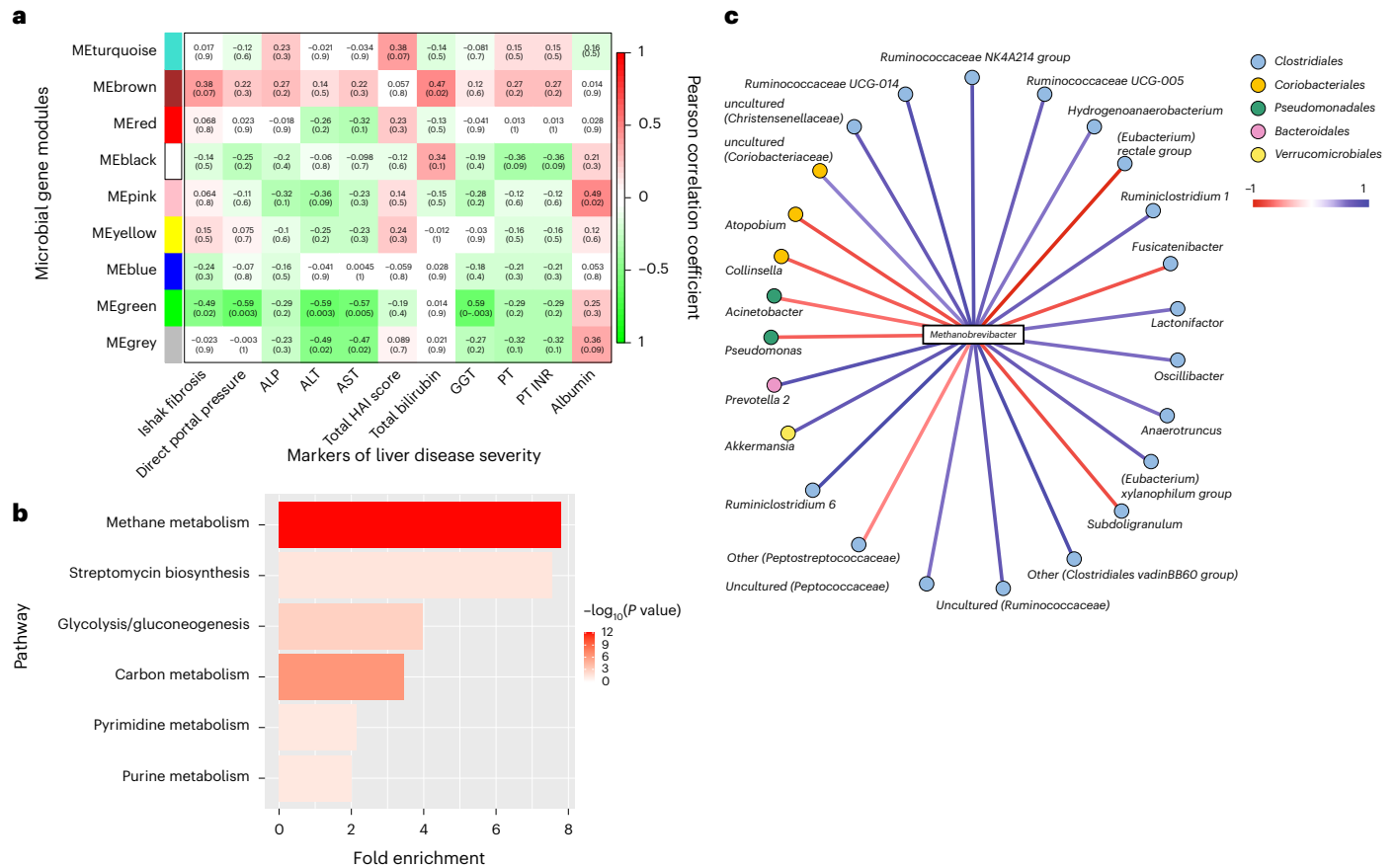
Fig. 6 | A beneficial role of *Methanobrevibacter* and methane metabolism reduced in fibrosis after SVR, *n* = 23. a, Correlation heat map of liver disease severity markers (*x* axis) and co-expressed microbial functional KO gene modules derived (*y* axis) in SVR (WGCNA). 'ME' is a module or cluster of genes that are co-expressed, and each ME module is randomly assigned a colour. Within each cell, upper row indicates Pearson correlation co-efficient and lower row with parenthesis two-sided Fisher's asymptotic unadjusted *P* value. Only one microbial KO gene module 'MEgreen' inversely correlated with multiple markers of liver disease severity. **b**, Pathway enrichment on the microbial KO genes in the

'MEGreen' module using MicrobiomeAnalyst R package, Shotgun Data Profiling, unadjusted *P* < 0.05. **c**, Inter-microbial linkages of *Methanobrevibacter* in SVR using SparCC on relative abundance of genera derived from faecal 16S rRNA. Displayed are 25 microbial genera that reached statistical significance (two-sided pseudo *P* values <0.05) for interactions with *Methanobrevibacter* genus. Nodes represent microbial genera coloured by their constituent taxonomic order. Lines are coloured blue or orange per scale on the side of the figure for SparCC correlation co-efficient.

to enhanced hepatic transcription of inflammatory pathways. Thus, we have demonstrated that functional predominance of *B. vulgatus* has negative effects on energy and immune homeostasis in HCVi, as studied in other liver disease aetiologies[46,50]. As there was no difference in relative abundance of microbial phyla, we propose that a functional change in *B. vulgatus* precedes compositional changes mostly described in decompensated cirrhosis[47,48]. Inhibition of *B. vulgatus*-mediated glycan degradation through manipulation of diet or intestinal immunity offers an earlier therapeutic intervention to preserve intestinal homeostasis in chronic liver disease.

Inferring directionality from longitudinal analysis, we hypothesize that hepatic metabolic dysfunction from HCV-induced inflammation leads to functional over-representation of 'less favourable' microbial metabolic pathways. This speculation is supported by studies linking *B. vulgatus* with altered hepatic lipids, bile acids and vitamin A[51–53]. Of note, this analysis is associative and portal long-chain FA and glycans may not be directly derived from the gut microbiome but share disease associations, a concept worthy of further exploration.

We acknowledge that, owing to the intrinsic nature of human studies, data presented here are observational and correlative. By virtue of this being a hypothesis-generating study, multiple test corrections were employed to minimize associations due to chance. Validity of the results was supported by the consistency in findings across biological

compartments and patient subsets. A major strength of our study was the longitudinal design with paired data from the same patients before and after HCV elimination that allows inference of directionality. Exploring authentic human biology in this manner is the first step in revealing processes that require mechanistic validation through future work. Specifically, concomitant alterations in hepatic, portal and microbial FA metabolism should be validated in an animal model of chronic liver disease. Microbial functions of FA synthesis and glycan degradation as well as the major transcriptionally active microbial species for these functions, that is, *A. hadrus* and *B. vulgatus* could be manipulated to assess for changes in liver disease severity. Similarly, hepatic peroxisome and mitochondrial rescue, for example, with PPAR agonists, could be undertaken to explore changes in disease and disease-associated microbial functions. Animal-based experimental models would also address two limitations of using a human cohort. First, although a temporal analysis allowed us to compare HCV with a presumably improved state (SVR), we cannot confirm complete resolution owing to the inability to obtain portal blood and/or liver tissue samples from healthy controls. Second, gut–liver axis components that would add further insight but were not sampled include intestinal lymphatics, bile and luminal microenvironment.

We have highlighted the central role of mitochondria and peroxisomes in hepatic metabolic dysregulation and identified enhanced

microbial FA biosynthesis and glycan metabolism driven by transcriptionally active *A. hadrus* and *B. vulgatus* in HCVi liver disease. Only by evaluating the portal vein were these shared metabolic signals between the gut microbiome and liver shown to have a role in enhanced portal and hepatic inflammation, thus potentially accelerating liver disease. This global energy dysregulation in the gut–liver axis has clinical implications in liver disease manifestations such as hepatic steatosis, sarcopenia, malnutrition and encephalopathy[54]. By performing temporal analysis, we have shown persistent hepatic metabolic disruption in advanced fibrosis even after elimination of the initial trigger (HCV). This study highlights the centrality of energy metabolism in cirrhosis and suggests its relevance in non-HCV aetiologies of cirrhosis. Concepts of microbial glycan metabolism and a therapeutic role for peroxisome rescue have been described in viral infections, including severe acute respiratory syndrome coronavirus 2 and metabolic syndrome, suggesting universal mechanisms with therapeutic implications in chronic inflammation as it relates to the host and microbiome[1,55–57].

Methods

Patient selection and study design

Before enrolment, we calculated the sample size as follows: sample size of seven patients per group will provide the study with a statistical power of 80% at a 95% confidence level, to detect a difference of 60% in portal vein microbial product detection rate between the two groups. The sample size was increased to 10 patients in the minimal fibrosis group and 20 patients in the advanced fibrosis group to allow for withdrawals and technical failures that might occur at a higher rate in cirrhotic patients. Thirty-six patients with chronic HCV infection were assessed at the National Institutes of Health Clinical Center, of whom 30 subjects (the accrual ceiling) were found eligible and agreed to participate. Major exclusion criteria for enrolment included other aetiologies of chronic liver disease, decompensated liver disease, carcinoma including hepatocellular carcinoma, and alcohol use of more than seven drinks per week. One patient was excluded for incidental finding of hepatocellular carcinoma after signing consent. All patients signed informed consent for participation into The National Institute of Diabetes, Digestive and Kidney Diseases, and the National Institute of Arthritis and Musculoskeletal Diseases, Institutional Review Board at the National Institutes of Health approved protocol (NCT02400216). Patients were compensated for study participation. Of the initial 29 patients in HCVi, a total of 23 patients completed every component of post-treatment evaluation as one patient died and four patients declined re-enrolment. One patient re-consented but did not complete sample collection after SVR and thus was excluded from all analysis at the SVR timepoint. All data were collected and processed at the National Institutes of Health Clinical Center. The recruitment began on 29 May 2015 and ended on 11 March 2016. Data collection began on 1 June 2015 and ended on 14 February 2017.

This was not a randomized control trial. The study was a proof of concept. As such, the intent was not to perform a randomized controlled study. Rather, the design of the study was aimed at comparing paired samples before and after an intervention (HCV treatment).

All patients had HCV RNA >5,000 IU ml⁻¹ with genotype distributions recorded (Supplementary Table 1). Patients with cirrhosis had Child–Pugh score A (compensated liver disease). Medication use and pre-existing medical conditions are reported in Supplementary Table 9. No patient had used an anti-microbial agent within 3 months before enrolment, and only 3 of the 29 patients had been on a proton pump inhibitor at the time of initial enrolment. One patient had a previous diagnosis of diabetes mellitus. Between the HCVi and SVR time interval there was no significant change in body mass index or haemoglobin A1c (Supplementary Table 1).

Sample collection

Percutaneous ultrasound-guided puncture of the left or right hepatic lobe in proximity to a distal portal vein branch was performed with a 17 G needle. Two hepatic core biopsy specimens were obtained with a coaxially introduced 18 G Temno needle (Temno Evolution, MeritMedical). The 17 G needle was then repositioned to puncture the proximate portal vein branch under ultrasound guidance. Catheterization of the portal vein under fluoroscopic guidance was accomplished with a Grebset (Teleflex) over a 0.018 inch guidewire. Portal pressures and venous blood samples were obtained through the 5 F braided sheath. The sheath was then withdrawn into the hepatic parenchyma and the puncture track embolized with Gelfoam pledgets (Pfizer Medical) for haemostasis (Supplementary Video).

Peripheral blood samples were obtained from an antecubital vein drawn into a 3.5 ml Z Serum Sep. Clot Activator (Ref 454067P, Greiner Bio-One). For each patient, this was performed at the same time as the portal vein sampling. Serum samples were processed by centrifugation at 2,000*g* for 10 min within 4 h of being drawn and subsequently stored at −80 °C until analysis.

Stool samples were collected in a sealable container within 2 days of serum sample collection, and stool container was stored at +4 °C immediately after collection. Samples were aliquoted into Eppendorf tubes and stored at −80 °C for further analysis within 8 h of initial collection.

Of note, blood, faecal and liver tissue samples were collected and analysed blinded in the HCVi cohort. As the same patients returned for SVR, the collection of the blood, faecal and liver tissue samples in SVR was not blinded. However, after collection, all SVR samples were coded and analysed blinded.

Dietary survey

Dietitians conducted a computer-assisted (Nutrition Data System for Research, Nutrition Coordinating Center) 24 h recall using the multiple pass method to determine dietary intake in the day preceding the faecal samples. To assess chronic dietary habits, subjects were asked to electronically complete the Diet History Questionnaire II past year with portion size version, which is a 150-question food frequency questionnaire to determine the relative mass of 161 macro- and micronutrients consumed. In total, 28 patients with HCVi and 23 patients with SVR completed this analysis.

Sample size for each analysis

For all analysis on serum, plasma, whole blood and microbial 16S data at individual timepoints, 29 patients with HCVi and 23 with SVR were included. Patients were stratified using Ishak fibrosis score from liver biopsies corresponding to HCVi and SVR timepoints. Ishak fibrosis scores 0–4 were characterized as non-cirrhotic, and Ishak fibrosis scores 5 and 6 as cirrhotics. One patient had a change in Ishak score from '6' to '0' after SVR but with an inadequate sample size (9 mm) and an unchanged elevated direct portal pressure at the SVR timepoint. Thus, he was stratified as cirrhotic in SVR cohort on the basis of his original HCVi Ishak fibrosis score. This resulted in an overall distribution of patients as HCVi-Cirr $n = 13$ and HCVi-NC $n = 16$; SVR-Cirr $n = 9$ and SVR-NC $n = 14$. Owing to failure of patient samples to meet quality-control parameters, analysis on liver transcriptome had HCVi $n = 27$ (HCVi-Cirr $n = 12$ and HCVi-NC $n = 15$) and SVR $n = 23$ (SVR-Cirr $n = 9$ and SVR-NC $n = 14$). Owing to data filtering, analysis on microbial transcriptome had HCVi $n = 26$ and SVR $n = 23$. Lastly, for paired analysis on serum, plasma and microbial 16S rRNA data $n = 23$, and for liver transcriptomics and microbial metatranscriptomics $n = 22$ owing to data filtering as above.

Histological and RNA-sequencing analysis on liver biopsies

Liver biopsy samples were scored in a blinded manner by a hepatopathologist Dr David Kleiner. Fibrosis was scored on liver biopsy samples using Ishak fibrosis score, and inflammation was scored using the Hepatic Activity Index (HAI)[6,58]. For RNA-sequencing analysis, liver tissue sample (10 mg) was snap frozen and stored at −80 °C. Total

RNA extraction was performed simultaneously on the liver biopsies collected at HCVi and SVR timepoints. To minimize batch effect, samples were coded without regard to their respective timepoints and processed in a blinded manner. This was done using TRIzol (catalogue number 15596026) and Qiagen RNA Extraction Kit (catalogue number 74104). A poly-A selection was performed on the total RNA samples using NEBNext Poly (A) Selection kit (catalogue number E7490S). The RNA library was prepared using the poly-A selected RNA and ScriptSeq RNA library Prep Kit (catalogue number SSV21106). All complementary DNA libraries were quantified using KAPA Biosystems Illumina qPCR Kit (Roche, catalogue number 07960140001), normalized and submitted to the NIDDK Genomics Core for Illumina HiSeq 4000 sequencing. Raw sequencing files were aligned to the *Homo sapiens* hg38 reference genome using STAR in Partek Flow (Version 10.0) (Computer software, Partek Inc. 2020). Sequences were filtered using a cut-off threshold of <40 million reads with Phred scores <30 for all 50 bp sequenced. Of the 24,380 genes identified, batch effect removal, pre-analysis data filtering and differential gene expression were performed using DESeq2 R package. Two patients (from HCVi cohort) failed quality control and were excluded from analysis involving liver transcriptome data from HCVi, resulting in $n = 27$ for HCVi and $n = 23$ for SVR.

Clinical markers of liver disease

Biochemical assays were performed on a Cobas C 501 system to measure alanine aminotransferase (ALT), AST, albumin, and total and direct bilirubin. Complete blood counts were drawn in 3 ml K2 EDTA tubes (Ref 367856, Becton, Dickinson and Company), measured on Sysmex system.

Serum immune and microbial markers

In total, 65 serum markers were measured in HCVi cohort and 61 of the 65 serum markers were measured again in SVR. IL-2, IL-4, IL-6, IL-8, IL-10, IL-12p70, IL-13 and TNFα were measured with the V-PLEX Proinflammatory Panel 1 Human Kit (Meso Scale Diagnostics, catalogue number K15049D-1); GM-CSF, IL-1α, IL-5, IL-7, IL-12/IL-23p40, IL-15, IL-16, IL-17 and TNFβ were assessed with the V-PLEX Cytokine Panel 1 Human Kit (Meso Scale Diagnostics, catalogue number K15050D-1); IL-18 was measured with Human IL-18 Kit (Meso Scale Diagnostics, catalogue number K151MCD-2); interferons (IFNα, IFNβ, IFNγ and IFNλ) were measured using U-PLEX Interferon Combo Human (Meso Scale Diagnostics, catalogue number K094K-1); VEGFR1, bFGF, PlGF, Tie2, VEGF-A, VEGF-C and VEGF-D were measured with the V-PLEX Angiogenesis Panel 1 Human Kit (Meso Scale Diagnostics, catalogue number K15190D-1). E-selectin, P-selectin, sICAM3 and thrombomodulin were measured with the Human Vascular Injury Panel 1 Human Kit (Meso Scale Diagnostics, catalogue number K15135C-1). SAA, CRP, sVCAM1 and sICAM1 were measured with the V-PLEX Vascular Injury Panel 2 Human Kit (Meso Scale Diagnostics, catalogue number K15198D-1). TGFβ was measured with Human TGF-β 1 Kit (Meso Scale Diagnostics, catalogue number K151IUC-1). Eotaxin, MDC, CCL26, MIP1α, MIP1β, TARC, MCP1, MCP4 and CXCL10 were measured with the V-PLEX Chemokine Panel 1 Human Kit (R&D Systems, catalogue number K15047D-1); soluble CD163 (sCD163) and soluble CD14 (sCD14) were measured with a Quantikine ELISA kit for human sCD163 and human sCD14 (R&D Systems, catalogue numbers DC1630, and DC140 respectively); PDGF-AA and PDGF-BB were measured with Human/Mouse PDGF-AA Quantikine Elisa Kit and Human PDGF-BB Quantikine Elisa Kit (R&D Systems, catalogue numbers DAA00B and DBB00, respectively). CCL5, CXCL4 and CXCL9 were measured with Human CCL5/RANTES Quantikine ELISA Kit, Human PF4/CXCL4 Quantikine ELISA Kit and Human CXCL9/MIG Quantikine ELISA Kit (R&D Systems, catalogue numbers DRN00B, DPF40 and DCX900, respectively); FGF19 with Human FGF-19 Quantikine ELISA Kit (R&D Systems, catalogue number DF1900). Zonulin was measured using Mybiosource Human Zonulin ELISA Kit (catalogue number MBS706368). Lipopolysaccharide was measured with

Lonza QCL-1000 120 Test Kit (catalogue number 50-647U); endotoxin with Lonza Kinetic-QCL 192 Test Kit (catalogue number 50-650U), lipoteichoic acid with General LTA ELISA Kit (catalogue number MBS288308), peptidoglycan with Mybiosource Human Peptidoglycan (PG) ELISA Kit (catalogue number MBS751887) and 1,3-β-D-glucan with Fungitell 1,3-β-D-Glucan ELISA Kit (Associates of Cape Cod Incorporated) (catalogue number FT001). HMGB1 was measured with ELISA HMGB1, 96DET Reagent, HMGB1 ELISA kit (Tecan, catalogue number ST51011). PDGFRa was measured with PDGFRa Human ELISA Kit (Cedarlane Labs, catalogue number SEC060HU), and iC3b was measured with Microvue iC3b (Quidel, catalogue number A006). Each assay was conducted following the respective manufacturers' protocols. All assays were performed in serum and in duplicate.

Flow cytometry

EDTA anti-coagulated peripheral and portal blood samples were processed for flow cytometry using a whole blood lysis method, stained with fluorescent antibodies, collected with a FACS Canto II (Becton Dickinson) and analysed using FCS Express software (De Novo). Lymphocytes were identified on the basis of a gate (Supplementary Fig. 2) established by forward and side angle scatter and confirmed using anti-CD45 and anti-CD14. B cells were identified by directly conjugated monoclonal antibodies: anti-CD20, anti-CD19, anti-CD5, anti-CD10, anti-IgM, anti-CD38 and anti-CD27. Irrelevant, directly conjugated, murine IgG1 was used to ascertain background staining. All monoclonal antibodies were obtained from Becton Dickinson, except for anti-Vβ-11 and anti-CD45RA (Beckman-Coulter), anti-IgM and anti-Vα-24 (BioLegend), anti-CD4, anti-CD45, anti-CD14, anti-CD19, anti-CD10 and anti-CD27 (Life Technologies) and anti- α4β7 (NIH AIDS Reagent Program, National Institute of Allergy and Infectious Diseases (NIAID), NIH). For information on antibody dilutions/amounts, company names and catalogue numbers for antibodies used, refer to Supplementary Table 10.

Non-targeted global metabolite profiling

Metabolon conducted the global metabolomics assays in peripheral and portal serum at both timepoints as detailed below. Sample handling, quality control and data extraction along with biochemical identification, data curation, quantification and data normalizations were performed as detailed below[59].

The original HCVi metabolomic data from 29 patients contained a total of 1,541 metabolites and were used for metabolomic analysis within HCVi (SourceData_Metabolites_IndividualCohorts). Similarly, the SVR metabolomic data from 23 patients contained a total of 1,786 metabolites and were used for metabolomic analysis within SVR (SourceData_Metabolites_IndividualCohorts). To perform paired analysis on the 23 patients in both HCVi and SVR cohorts, 12 anchor samples or technical replicates from the HCVi cohort were re-submitted and analysed simultaneously with the SVR serum samples to facilitate merging of the two datasets. A 50% fill value was required for each metabolite for the purposes of merging, and consequently, metabolites that did not meet this criterion were excluded from the final merged dataset. This anchored analysis generated 1,256 metabolites to be used for paired analysis between HCVi and SVR cohorts (SourceData_Metabolites_PairedCohorts). All samples were analysed on Metabolon's global metabolic profiling (HD4) and complex lipid panel (CLP) platforms. The experimental samples were loaded in a balanced and equivalent manner across the analytical platforms and analysed without any further normalization.

Metabolites were extracted in methanol with vigorous shaking for 2 min (Glen Mills Genogrinder 2000) followed by centrifugation. The resulting extract was divided into aliquots, dried and then reconstituted in acidic or basis liquid chromatography (LC)-compatible solvents for analysis by four ultraperformance liquid chromatography–tandem mass spectrometry (UPLC–MS/MS) methods[60].

Two aliquots were analysed using acidic, positive ion conditions, chromatographically optimized for either hydrophilic or hydrophobic compounds, respectively. For detection of the hydrophilic compounds, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1 × 100 mm, 1.7 μm) using water and methanol, containing 0.05% perfluoropentanoic acid and 0.1% formic acid. For the hydrophobic compounds, the extract was gradient eluted from the C18 column (Waters UPLC BEH C18-2.1 × 100 mm, 1.7 μm) using methanol, acetonitrile, water, 0.05% perfluoropentanoic acid and 0.01% formic acid and was operated at an overall higher organic content. A third aliquot was analysed using basic negative ion optimized conditions and gradient eluted using a separate dedicated C18 column (Waters UPLC BEH C18-2.1 × 100 mm, 1.7 μm) using water and methanol containing 6.5 mM ammonium bicarbonate. A final aliquot was analysed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1 × 150 mm, 1.7 μm) using a gradient consisting of water and acetonitrile with 10 mM ammonium formate. The mass spectrometry (MS) analysis alternated between MS and data-dependent MS^2 scans using dynamic exclusion, and the scan ranged from 80 to 1,000 m/z. Metabolites were identified by automated comparison of the ion features in the experimental samples to a reference library of chemical standard entries that included retention time, molecular weight (m/z), preferred adducts and in-source fragments as well as associated MS spectra, and were curated by visual inspection for quality control using software developed at Metabolon[61].

Measurement of circulating short-chain FAs

Detection and quantification of peripheral and portal plasma short-chain FAs were analysed and characterized by UPLC–MS/MS, utilizing a Thermo Scientific Vanquish UPLC and a Thermo Scientific Altis triple quadrupole mass spectrometer with heated electrospray ionization (ESI; HESI-II, Thermo Scientific) in negative ion mode (3,500 V). The internal standard (IS) solution was prepared in MeOH containing 2-ethylbutyric acid. The short-chain FFA standards were mixed with acetic acid (C2), propionic acid (C3), butyric acid (C4), valeric acid (C5) and caproic acid (C6) at different ranges of concentrations. Fifty microlitres of mixed short-chain FFA standards was added to 300 μl of IS solution and vortexed for 5 min, then 250 μl of mixture was transferred into an LC–MS vial. Fifty microlitres of plasma sample was mixed with 300 μl of IS solution, vortexed vigorously for 5 min, and centrifuged at 4 °C, then 250 μl of clear supernatant was transferred into an LC–MS vial. The derivatization was carried out for both standards and samples by adding 20 μl of 200 mM 3-nitrophenylhydrazine, 20 μl of 320 mM N-(3-dimethylaminopropyl)-N′-ethylcarbodiimide, HCl (EDAC) in 75% methanol and 20 μl of 16% pyridine in methanol. The derivatization reaction was incubated at 4 °C for 24 h. Finally, the derivatization solutions (2 μl) were analysed by means of UPLC–ESI–MS/MS. A reverse-phase analysis was performed via an Acquity UPLC BEH C18 column (1.7 μm, 2.1 × 100 mm) at 40 °C, and the samples were maintained in the autosampler at 4 °C. The mobile phase consisting of solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in Acetonitrile) was delivered at a flow rate of 0.35 ml min⁻¹, 12 min for each injection. The gradient elution was as follows: B% = 15, 15, 55, 100 and 15 (0, 0.25, 7.25, 8.75 and 11.25 min). Quantitation of the short-chain FFA were based on MS/MS transitions. Standards were calibrated with $R^2 > 0.99$.

Measurement of circulating lipoprotein particles

Portal and peripheral plasma lipids (total cholesterol, triglycerides and high-density lipoprotein (HDL)-C) were measured by the Cobas6000 analyser (Roche Diagnostics). Plasma lipoprotein particles numbers were measured by nuclear magnetic resonance (NMR) on the Vantera Analyzer (LabCorp), which uses Lp4 deconvolution algorithm (LP4) to quantify lipoprotein subspecies. NMR has inherently high resolving power to discriminate HDL particles differing only slightly in size, but previous algorithms did not fully exploit this capability. The LP4

algorithm measures seven different HDL subspecies with improved precision and corrects a prior systematic overestimation of the absolute concentrations of HDL particles (responsible for apoA-1/HDL-P ratios that were improbably low)[62]. New HDL subclass signal to particle conversion factors were determined by regression of the NMR subclass signal areas against plasma apoA-1 concentrations, producing 'calibrated' HDL particle concentrations (cHDLP) that are ~30% lower than previous HDLP values. This platform quantified six HDL subspecies, H1P through H7P numbered from lowest HDL particle concentration (H1P) to highest HDL particle concentration (H7P). Also corrected was an aspect of prior deconvolution models that led to systematic underestimation of LDL particle concentrations (owing to imperfect modelling of the plasma protein background signal). As a result, calibrated LP4 LDL particle concentrations (cLDLP) are now higher by about 350 nmol l⁻¹ while remaining highly correlated ($r = -0.95$) with previous LDL-P values and retaining equivalently strong associations with cardiovascular outcomes. Finally, linear regressions of subclass signal areas against independent chemical measures of cholesterol, triglycerides and apolipoproteins from a large population sample have produced conversion factors enabling the reporting of NMR-derived lipid and apolipoprotein concentrations. The current software version reports 45 parameters, while simultaneously measuring a novel NMR inflammation biomarker (GlycA)[62] plus five multimarkers that combine selected NMR parameters into 'scores' for the assessment of diabetes risk or prediction of longevity. These include Lipoprotein Insulin Resistance Index (LP-IR), Insulin Resistance Diabetes Risk Factor Index (IRDRF), Short-term Diabetes Risk Factor Index (SDRF), 5-Year Diabetes Risk Factor Index (DRF5) and NMR Longevity Index (LGVX). We have reported these scores calculated from lipoprotein particles measured in both portal and peripheral plasma.

Bioinformatics and integrative analyses on omics data

SNF. We utilized SNF, developed by Wang et al.[17], a data integration method that systematically captures both shared and complementary information from different data sources[17]. All computations were carried out on R 4.0.2 (R Core Team 2020) with the R packages SNFtool and bnstruct (https://www.r-project.org).

HCVi metabolites. We first confirmed that the peripheral and portal datasets contain the same set of analytes. Within each dataset, 1,541 metabolites spanned 25 pathway classes, for example, lipid, nucleotide, amino acid and so on. We performed the following steps on each dataset, independent of each other. Each metabolite class was normalized using the standardNormalization function, which was then used to compute the distances between the patients ($n = 29$) using the dist2 function. The resulting 29-by-29 distance matrix was transformed to an affinity matrix through a scaled exponential similarity kernel. We used the affinityMatrix ($K = 10$, $\sigma = 0.5$) function for the transformation. K represents the number of nearest neighbours, whereas σ is a hyperparameter that measures the variance of local model. The affinity matrix, W, describes a patient similarity network graph where the edge between patient i and j has an edge weight of $W(i,j)$, the degree of similarity between the patients. We repeated this process for all 25 pathway classes to obtain 25 affinity matrices. The matrices were fused using the SNF ($K = 10$, $t = 25$) function. K represents the number of neighbours in the k-nearest neighbours' part of the SNF analysis algorithm, and t is the number of iterations in the fusion process. The fused matrix contains the comprehensive information of the 25 patient networks. Finally, we applied spectralClustering(C=2) function on the fused affinity matrix to cluster the patients into two groups. Using Welch's t-test, we compared the mean Ishak scores of the groups.

HCVi cytokines. We first verified that the peripheral and portal datasets share the same set of analytes. There were 65 common cytokines across the datasets. Data pre-processing for the cytokine data involved

missing data omission and imputation. Following Wang et al., the cytokines with more than 20% missing data were omitted from the analyses (none met this cut-off). The remaining missing data were imputed using the *k*-nearest neighbours' algorithm from the R package bnstruct. Unlike the metabolite datasets, where we were able to group the analytes on the basis of their pathway classes, there was no obvious way to group the cytokines. Thus, we produced only one affinity matrix for each dataset, and hence the fusion algorithm was not necessary in this case. We followed the steps described earlier to construct two affinity matrices, one for the peripheral and the other for the portal dataset, that convey the patient similarity information. Using spectral clustering, we obtained two groups. We used Welch's *t*-test to compare the means of Ishak scores of the groups.

Logistic regression. For logistic regression modelling, we utilized the Python package Scikit-Learn (https://scikit-learn.org)[63]. All computations were performed with Python version 3.9.5. The patients' Ishak scores were transformed to a binary variable. To evenly distribute the patients, Ishak scores less than or equal to 3 were labelled '0', and the scores greater than 3 were labelled '1'. Following the algorithm of spectral clustering, we obtained *L*, the normalized Laplacian of the affinity matrix, and constructed a 29-by-2 matrix whose columns are the eigenvectors associated with two smallest eigenvalues of *L*. We trained and tested logistic regression models on this 29-by-2 matrix. The models were trained and tested with 20 different combinations of train/test sets with 23 training samples and 6 testing samples. The model accuracy was evaluated with the mean precision and the mean F1 score of the test sets.

16S rRNA analysis on faecal samples
HCVi and SVR faecal samples were collected and flash frozen with storage at −80 °C. PowerSoil DNA Isolation Kit (MO BIO, catalogue number 12888-100) was used to extract total DNA following a modified protocol for DNA extraction from faeces[64]. Magnetic bead purification was used in place of column binding purification. The 16S region was amplified from total DNA using primers targeting the V4 region of 16S SSU rRNA (515f-806r). Paired-end sequencing of the 16S amplicons was conducted by the NIDDK Genomics Core on the Illumina MiSeq (2 × 150 bp). Paired-end FASTQ files were processed and analysed with QIIME v1.9.1 on the Nephele platform from the NIAID Office of Cyber Infrastructure and Computational Biology (OCICB) in Bethesda, MD https://nephele.niaid.nih.gov. For pre-processing, the minimum Phred quality score was 19 and the Phred offset was 33. Reads were joined using the following parameters: max bad run length of 3, minimum overlap length of 10, and 25% difference within overlap. For alignment, operational taxonomic units were matched to known bacterial sequences using the reference database Greengenes, used at 99% sequence similarity.

Gut ecology analysis on faecal 16S rRNA using SparCC.
To infer the taxon–taxon correlations from the faecal 16S rRNA, we utilized SparCC[35], a technique for estimating correlation values from compositional data. The computations were carried out on Python 3.9.5 with the Python module SparCC (https://github.com/JCSzamosi/SparCC3). Firstly, the correlation between the relative abundance of microbial genera derived from faecal 16S rRNA were computed (342 genera in total). Then we generated 100 shuffled datasets, as described by Friedman et al., and computed correlations for each of the shuffled dataset. Finally, for each component pair, pseudo *P* values were calculated to determine statistical significance of the correlations with two-sided comparison. Pseudo *P* value for each component pair is defined as proportion of shuffled datasets for which the corresponding correlation value at least as extreme as the original data.

Microbial metatranscriptomics analysis on faecal samples
Using the same faecal samples as noted in the 16S rRNA analysis section, total RNA extraction was performed using TRIzol (catalogue number 15596026) and Qiagen RNA Extraction Kit (catalogue number 74104). A DNAase (Ambion Turbo DNA Free Kit; Invitrogen, catalogue number AM1907) was added to digest DNA, leaving RNA to form a cDNA library. All cDNA libraries were quantified using KAPA Biosystems Illumina qPCR kit (Roche, catalogue number 07960140001). The cDNA libraries were normalized and submitted to the NIDDK Genomics Core for Illumina HiSeq 4000 sequencing. Every faecal sample had a minimum of 40 million single-end reads trimmed to have minimum Phred quality score of 30 for each base pair.

Generation of a non-redundant metatranscriptomics reference gene catalogue and quantification of metatranscriptomics samples. Illumina raw metatranscriptomics data from 26 faecal samples in HCVi and 23 faecal samples in SVR were processed using MOCAT2 pipeline[65]. In brief, raw sequence reads were trimmed and quality filtered (MOCAT.pl rtf) with a length of 30 bp and quality of 20 bp cut-offs using FastX program (http://hannonlab.cshl.edu/fastx_toolkit/). High-quality reads were screened (MOCAT.pl -s hg19) for human contamination against provided human genome database (hg19, Genome Reference Consortium Human Reference[37]) using SOAPAligner (version 2.21). The screened reads were assembled into contiguous sequences (contigs) (MOCAT.pl -a -r hg19) with a minimum length of 500 bp using SOAPDenovo software (version 2.04). The construction of the non-redundant gene catalogue (MOCAT.pl -make_gene_catalog -assembly_type assembly -r hg19) was achieved by, first, predicting genes from long contigs (MOCAT.pl -sf samples -gp assembly -r hg19) by MetaGeneMark software (version 3.38), and then clustering them into non-redundant gene sets using CD-HIT.

High-quality reads were mapped to the annotated gene catalogue with 95% identity cut-off (MOCAT.pl -s samples.padded -r hg19 -identity 95) using SOAPAligner (version 2.21), where all uniquely mapped sequences were used to quantify microbial transcripts[65].

De novo assembly of metatranscriptome by co-abundance clustering; taxonomical annotation. The gene catalogue was clustered by co-abundance[19]. Briefly, canopy clustering algorithm performs a perpetual iteration of points (predicted genes) in multi-dimensional Pearson correlation space until the data converges, https://github.com/fplaza/mgs-canopy-algorithm. This method allows to perform de novo assembly of the metatranscriptome. The co-abundance clustering resulted in 696 CAGs in HCVi and 642 CAGs in SVR. For the downstream analysis only the largest CAGs with more than 700 genes in each, further referred to as transcriptionally active microbial species were selected: 24 CAGs in HCVi and 14 CAGs in SVR. To taxonomically annotate the transcriptionally active microbial species, catalogue genes from each CAG were mapped to known reference genomes using BLASTN (version 2.10.0+, NCBI nt database, March 2020 release https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome) at a threshold of 95% identity and filtered for genes with longer than 100 bp alignments. The transcriptionally active microbial species were assigned to a given genome with the most abundant species. Major driver species count table was constructed using the median gene transcription expression throughout the samples. For *M. smithii* correlations in SVR, microbial mRNA sequence reads were mapped to NCBI database, and the transcriptional activity was quantified based on NCBI taxonomic levels.

Profiling and taxonomic annotation of functional orthologues. For creating microbial functional profiles from KO genes, we used MetaHIT Consortium catalogue of over 3 million distinct nucleic acid sequences as a reference[66]. In total, 889,668 individual nucleotide sequences from MetaHit were aligned to the samples' trimmed reads using Bowtie in Partek Flow (Version 10.0) (Computer software, Partek 2020). The distinct nucleotide sequences were then summarized into 4,718 microbial KO genes. These KO genes were then summarized into microbiome

functional KEGG modules[67] based on annotations downloaded on 14 January 2014 to form metatranscriptomics functional potentials for the downstream analysis. Individual functional orthologues were taxonomically annotated using BLASTN (version 2.10.0+, NCBI nt database. March 2020 release) at a threshold of 99% identity and 95% coverage.

Driver-species analysis or leave-one-out analysis on the transcriptionally active microbial species. To identify the transcriptionally active microbial species that have the most contribution to the association between KEGG functional modules and the clinical phenotype, leave-one-out analysis was performed[18,68]. In each iteration, the phenotype–KEGG functional module association was calculated after excluding the genes from a given transcriptionally active microbial species. The importance of a given species was then defined as the highest change in median Spearman correlation coefficient between KOs and the clinical phenotype resulting from removing the genes from a given transcriptionally active microbial species.

Data dimensionality reduction
Knowledge-driven dimensionality reduction. Up-to-date human KEGG pathway gene sets were generated using kegg.gsets function from GAGE R package (version 2.201) (ref. [69]). A total of 19,960 hepatic genes were grouped into KEGG pathways[67] with a total number of 319 pathways. Microbial KO genes were mapped into KEGG functional modules as noted above.

Data-driven clustering. To evaluate which hepatic genes or portal metabolites are co-expressed in similar clusters both at HCVi and SVR timepoints, consensus clustering was performed on hepatic transcriptome and 600 Human Metabolome Database mapped portal metabolites using WGCNA (version 1.69) framework[70] available as an R package. Consensus clusters were assigned using modules constructed from paired samples at the two timepoints HCVi and SVR following WGCNA standards as described[70,71]. The parameter selection for consensus WGCNA construction on liver transcriptome included soft threshold β/power 16, minModuleSize 30, deepSplit 2; and for portal metabolites soft threshold β/power 12, minModuleSize 30, deepSplit 2. The consensus WGCNA modules were labelled by colours[70,71].

WGCNA was also utilized for clustering portal metabolomics and microbial KO genes for independent analyses in HCVi and SVR. A total of 600 Human Metabolome Database mapped portal metabolites in HCVi and microbial metatranscriptome KOs genes generated from alignment to the MetaHIT Consortium database in SVR were used to generate six portal metabolite modules and nine microbial KO modules, respectively. Co-expression correlations were calculated using bicor function, which performs biweight midcorrelations (a median-based correlation measure that is more robust to the presence of the outliers in the data). A signed, weighted metabolite network was constructed by applying the scale-free topology criterion to choose the soft threshold $\beta = 6$. Modules or clusters of densely inter-connected portal metabolites and microbial KO genes were determined by implementing a dynamic branch cutting method[71], using deepSplit of 4 for portal metabolites and deepSplit of 2 for microbial KO; and minModuleSize of 30 for both datasets. The metabolite and microbial KO profiles constituting a given cluster are summarized by the cluster eigenvectors (the first principal component of the metabolite abundances). To distinguish from consensus WGCNA modules, the single-timepoint WGCNA portal metabolite modules were labelled as numbers from M01 through M08 (Source Data Fig. 5).

Modules were then tested for association with markers of liver disease at HCVi and SVR timepoints, including Ishak fibrosis score, direct portal pressure, HAI, ALT, AST, ALP, GGT, total bilirubin, prothrombin time (PT), prothrombin time international normalized ratio (PT INR) and albumin using Pearson correlation, Fisher's asymptotic two-sided unadjusted P value <0.05.

Pathway enrichment and data visualization
Pathway enrichment on liver transcriptome comparing HCVi and SVR. To identify pathways altered in HCVi compared with SVR, we used the DEGs with FDR P value <0.1 obtained from DeSeq2 paired analysis comparing HCVi and SVR (2,743 upregulated and 2,380 downregulated DEGs). NetworkAnalyst https://www.networkanalyst.ca was used to perform over-representation analysis on the DEGs, and pathways were annotated to KEGG database using hypergeometric tests to compute FDR-corrected P values for enrichment. Fold enrichment for each pathway was calculated as the ratio of actual/expected gene hits. We also performed gene set enrichment on the above DEGs using GAGE R package (version 2.201) (ref. [69]) using curated gene sets from the KEGG database[67]. Pathways significantly altered (FDR P value <0.1) were then visualized using PathView R package (version 1.20.1).

Identification of predominant cellular location for hepatic pathways comparing HCVi and SVR. To identify the predominant cellular location for the hepatic DEGs downregulated in HCVi compared with SVR, we used GO by cellular component database. Specifically, over-representation analysis was performed with *Homo sapiens* reference database in PANTHER, Fisher's exact test with FDR P value <0.1 (ref. [72]). This cellular location enrichment was also utilized to group the 42 metabolic pathways inversely correlating with fibrosis in HCVi. Any pathway that had at least one hit in mitochondria and/or peroxisomes with FDR <0.1 in GO cellular component pathway analysis were classified as mitochondria and/or peroxisome. Pathways with no mitochondria and/or peroxisome hits were classified as 'Other'.

GSEA to determine hepatic pathways altered on the basis of fibrosis. To determine hepatic functional alterations in cirrhosis, we also performed gene set enrichment analysis (GSEA) on the HCVi liver transcriptome ($n = 27$) using enrichment methods previously published[73]. Pre-ranked analysis with cirrhosis as disease phenotype was performed on the DeSeq2 output comparing HCVi-Cirr and HCVi-NC subgroups (HCVi-Cirr $n = 12$ and HCVi-NC $n = 15$). Formula for pre-ranking was per GSEA protocol ($-\log_{10}$ unadjusted P value × sign (log fold change)]. The reference gene sets were curated from Baderlabs AllPathways_Go_noiea_keggappended (updated April 2019). Enrichment parameters to generate enrichment networks in Cytoscape included weighted analysis with 1,000 permutations, enrichment networks filtered for gene set size of 15–200, enrichment-unadjusted $P < 0.005$ and FDR P value <0.05, Jaccard overlap combined co-efficient of 0.375 and constant of 0.5. Owing to the large number of gene sets upregulated in HCVi-Cir, we further organized the 4,434 genes in the 379 upregulated pathways in HCVi-Cirr using ClusterMaker, WordCloud annotation, and then manually grouping the gene sets on the basis of fundamental biological processes to obtain a final summary network[73].

GSEA was also utilized to perform pathway enrichment analysis between HCVi and SVR within subgroups stratified by cirrhosis, that is, non-cirrhotic and cirrhotic subgroups. In this case, pre-ranked analysis was performed on the DeSeq2 output for the paired analysis between HCVi-NC versus SVR-NC and HCVi-Cirr versus SVR-Cirr. Of note, among the 22 patients who completed both timepoint evaluations, 2 patients were excluded from this analysis as they had a change in fibrosis category between the HCVi and SVR timepoints, that is, HCVi-Cirr and SVR-Cirr based on Ishak fibrosis scores 5–6, $n = 7$ and HCVi-NC versus SVR-NC based on Ishak fibrosis scores 0–4, $n = 13$. Enrichment network visualization of the pathways downregulated in HCVi compared with SVR subgroups was illustrated using Cytoscape with the same parameters as noted above.

Pathway enrichment on microbial metatranscriptome in SVR. Pathway enrichment on microbial functional KO gene module 'MEgreen' constructed using WGCNA in SVR was performed using a publicly available software, MicrobiomeAnalyst (MicrobiomeAnalystR package).

The list of KO genes in MEgreen was used as input in the Shotgun Data Profiling pipeline that revealed statistically significant KEGG functional pathways enriched in MEgreen (https://www.microbiomeanalyst.ca).

Statistical analysis for subgroup comparisons

Statistical analysis was performed using GraphPad Prism 8.0 (GraphPad Software) and R software (versions 3.5.0 and 4.0.2). Non-parametric tests were exclusively used for correlations and comparisons. All P values were two-sided and adjusted using FDR when applicable. Descriptive characteristics were summarized using median and inter-quartile range (IQR). Correlations between parameters of interest were assessed using Spearman correlation. Paired comparisons were performed using two-sided Wilcoxon matched-pairs signed-rank test and unpaired comparisons based on disease severity using two-sided Mann–Whitney U test. For all statistical analysis, the data distribution was assumed to be normal, but this was not formally tested.

Statistical analysis for data integration

All statistical analyses were carried out using R (version 3.5.0). Transcriptionally active driver species (CAGs), hepatic genes and microbial KOs present in fewer than three individuals were excluded from the analysis. The integration of different omics data types was achieved by using the pipeline[18,68]. In summary, after constructing the modules for hepatic KEGG pathways and microbiome KEGG functional modules, the dataset was filtered by choosing only features significantly (Benjamini–Hochberg FDR <0.1) associated with the clinical phenotype, that is, Ishak fibrosis score, and/or liver enzymes (ALT, AST, ALP and GGT). The phenotype correlation analyses were conducted with a Spearman rank correlation test corrected for background distribution in hepatic pathways and microbial modules using the $SCC_{bg.adj.}$ reported by Pedersen et al.[18]. Cross-omics associations on a cluster level were calculated using Mann–Whitney U- test, where the ranks of hepatic genes and microbial KOs within a given KEGG pathway or KEGG functional module were compared with the ranks of the rest of hepatic genes or KOs.

Additional resources

Clinical trial registry number for the protocol is NCT02400216.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Please note that the microbial and liver transcriptome sequence and microbial 16S rRNA sequence dataset has been made available in the BioProject repository. The accession number for this repository is PRJNA727609. The serum metabolomics data have been uploaded as source data (SourceData_Metabolites_IndividualCohorts and SourceData_Metabolites_PairedCohorts). Additional minimum input data necessary to interpret the figures and findings have been provided as source data where appropriate. *Homo sapiens* hg38 reference genome was sourced from https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/. Source data are provided with this paper.

Code availability

There was no custom code or mathematical algorithm utilized in this study.

References

1. Agus, A., Clément, K. & Sokol, H. Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut* https://doi.org/10.1136/gutjnl-2020-323071 (2020).
2. Zheng, D., Liwinski, T. & Elinav, E. Interaction between microbiota and immunity in health and disease. *Cell Res.* **30**, 492–506 (2020).
3. Tranah, T. H., Edwards, L. A., Schnabl, B. & Shawcross, D. L. Targeting the gut–liver–immune axis to treat cirrhosis. *Gut* https://doi.org/10.1136/gutjnl-2020-320786 (2020).
4. Macpherson, A. J., Heikenwalder, M. & Ganal-Vonarburg, S. C. The liver at the nexus of host–microbial interactions. *Cell Host Microbe* **20**, 561–571 (2016).
5. Tripathi, A. et al. The gut–liver axis and the intersection with the microbiome. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 397–411 (2018).
6. Ishak, K. et al. Histological grading and staging of chronic hepatitis. *J. Hepatol.* **22**, 696–699 (1995).
7. Fransen, M., Lismont, C. & Walton, P. The peroxisome-mitochondria connection: how and why?. *Int. J. Mol. Sci.* **18**, 6 (2017).
8. Mansouri, A., Gattolliat, C.-H. & Asselah, T. Mitochondrial dysfunction and signaling in chronic liver diseases. *Gastroenterology* **155**, 629–647 (2018).
9. Lupberger, J. et al. Combined analysis of metabolomes, proteomes, and transcriptomes of hepatitis C virus-infected cells and liver to identify pathways associated with disease development. *Gastroenterology* **157**, 537–551.e9 (2019).
10. Walker, C. L., Pomatto, L. C. D., Tripathi, D. N. & Davies, K. J. A. Redox regulation of homeostasis and proteostasis in peroxisomes. *Physiol. Rev.* **98**, 89–115 (2017).
11. Keller, H. et al. Fatty acids and retinoids control lipid metabolism through activation of peroxisome proliferator-activated receptor-retinoid X receptor heterodimers. *Proc. Natl Acad. Sci. USA* **90**, 2160–2164 (1993).
12. 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).
13. Ryan, D. G. et al. Coupling Krebs cycle metabolites to signalling in immunity and cancer. *Nat. Metab.* **1**, 16–33 (2019).
14. Yamakado, M. et al. Plasma amino acid profile associated with fatty liver disease and co-occurrence of metabolic risk factors. *Sci Rep.* **7**, 14485 (2017).
15. Dasarathy, S. & Merli, M. Sarcopenia from mechanism to diagnosis and treatment in liver disease. *J. Hepatol.* **65**, 1232–1244 (2016).
16. Meoni, G. et al. The metabolic fingerprints of HCV and HBV infections studied by nuclear magnetic resonance spectroscopy. *Sci Rep.* **9**, 4128 (2019).
17. Wang, B. et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* **11**, 333–337 (2014).
18. Pedersen, H. K. et al. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **535**, 376–381 (2016).
19. Nielsen, H. B. et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).
20. Koh, A., Vadder, F. D., Kovatcheva-Datchary, P. & Bäckhed, F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell* **165**, 1332–1345 (2016).
21. Oliphant, K. & Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome* **7**, 91 (2019).
22. Ramanan, D., Tang, M. S., Bowcutt, R., Loke, P. & Cadwell, K. Bacterial sensor Nod2 prevents inflammation of the small intestine by restricting the expansion of the commensal *Bacteroides vulgatus*. *Immunity* **41**, 311–324 (2014).
23. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–335 (2012).
24. Dias, A. M. et al. Glycans as critical regulators of gut immunity in homeostasis and disease. *Cell. Immunol.* **333**, 9–18 (2018).

25. Giron, L. B. et al. Sialylation and fucosylation modulate inflammasome-activating eIF2 signaling and microbial translocation during HIV infection. *Mucosal Immunol.* **13**, 753–766 (2020).

26. Huang, Y.-L., Chassard, C., Hausmann, M., von Itzstein, M. & Hennet, T. Sialic acid catabolism drives intestinal inflammation and microbial dysbiosis in mice. *Nat. Commun.* **6**, 8141 (2015).

27. Feagan, B. G. et al. Treatment of ulcerative colitis with a humanized antibody to the α4beta7 integrin. *N. Engl. J. Med.* **352**, 2499–2507 (2005).

28. Negro, F. Abnormalities of lipid metabolism in hepatitis C virus infection. *Gut* **59**, 1279–1287 (2010).

29. Cronan, J. E. & Thomas, J. in *Methods in Enzymology* (ed. Hopwood, D.) Vol. 459 Ch. 17 395–433 (Academic Press, 2009).

30. Zhao, L. et al. Saturated long-chain fatty acid-producing bacteria contribute to enhanced colonic motility in rats. *Microbiome* **6**, 107 (2018).

31. Pujo, J. et al. Bacteria-derived long chain fatty acid exhibits anti-inflammatory properties in colitis. *Gut* https://doi.org/10.1136/gutjnl-2020-321173 (2020).

32. Chen, P. et al. Supplementation of saturated long-chain fatty acids maintains intestinal eubiosis and reduces ethanol-induced liver injury in mice. *Gastroenterology* **148**, 203–214.e16 (2015).

33. Samuel, B. S. & Gordon, J. I. A humanized gnotobiotic mouse model of host–archaeal–bacterial mutualism. *Proc. Natl Acad. Sci. USA* **103**, 10011–10016 (2006).

34. Borrel, G., Brugère, J.-F., Gribaldo, S., Schmitz, R. A. & Moissl-Eichinger, C. The host-associated archaeome. *Nat. Rev. Microbiol.* **18**, 622–636 (2020).

35. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).

36. Hotamisligil, G. S. Inflammation, metaflammation and immunometabolic disorders. *Nature* **542**, 177–185 (2017).

37. Gehrke, N. & Schattenberg, J. M. Metabolic inflammation—a role for hepatic inflammatory pathways as drivers of comorbidities in nonalcoholic fatty liver disease? *Gastroenterology* **158**, 1929–1947.e6 (2020).

38. Arvind, A., et al.) (MDText.com, Inc., 2000).

39. Lodhi, I. J. & Semenkovich, C. F. Peroxisomes: a nexus for lipid metabolism and cellular signaling. *Cell Metab.* **19**, 380–392 (2014).

40. Islam, S. M. T., Won, J., Khan, M., Chavin, K. D. & Singh, I. Peroxisomal footprint in the pathogenesis of nonalcoholic steatohepatitis. *Ann. Hepatol.* **19**, 466–471 (2020).

41. Kersten, S., Desvergne, B. & Wahli, W. Roles of PPARs in health and disease. *Nature* **405**, 421–424 (2000).

42. Pawlak, M., Lefebvre, P. & Staels, B. Molecular mechanism of PPARα action and its impact on lipid metabolism, inflammation and fibrosis in non-alcoholic fatty liver disease. *J. Hepatol.* **62**, 720–733 (2015).

43. Corpechot, C. et al. A placebo-controlled trial of bezafibrate in primary biliary cholangitis. *N. Engl. J. Med.* **378**, 2171–2181 (2018).

44. Tietz-Bogert, P. S. et al. Metabolomic profiling of portal blood and bile reveals metabolic signatures of primary sclerosing cholangitis. *Int. J. Mol. Sci.* **19**, 10 (2018).

45. Schierwagen, R. et al. Circulating microbiome in blood of different circulatory compartments. *Gut* **68**, 578–580 (2019).

46. Iebba, V. et al. Combining amplicon sequencing and metabolomics in cirrhotic patients highlights distinctive microbiota features involved in bacterial translocation, systemic inflammation and hepatic encephalopathy. *Sci. Rep.* **8**, 8210 (2018).

47. Albhaisi, S. A. M., Bajaj, J. S. & Sanyal, A. J. Role of gut microbiota in liver disease. *Am. J. Physiol. Gastrointest. Liver Physiol.* **318**, G84–G98 (2019).

48. Trebicka, J., Bork, P., Krag, A. & Arumugam, M. Utilizing the gut microbiome in decompensated cirrhosis and acute-on-chronic liver failure. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 167–180 (2021).

49. Vacca, M. et al. The controversial role of human gut lachnospiraceae. *Microorganisms* **8**, 573 (2020).

50. Schwimmer, J. B. et al. Microbiome signatures associated with steatohepatitis and moderate to severe fibrosis in children with nonalcoholic fatty liver disease. *Gastroenterology* **157**, 1109–1122 (2019).

51. Cai, W. et al. Ethanol extract of propolis prevents high-fat diet-induced insulin resistance and obesity in association with modulation of gut microbiota in mice. *Food Res. Int.* **130**, 108939 (2020).

52. Qi, X. et al. Gut microbiota–bile acid–interleukin-22 axis orchestrates polycystic ovary syndrome. *Nat. Med.* **25**, 1225–1233 (2019).

53. Hibberd, M. C. et al. The effects of micronutrient deficiencies on bacterial species from the human gut microbiota. *Sci. Transl. Med.* **9**, 390 (2017).

54. Tilg, H., Cani, P. D. & Mayer, E. A. Gut microbiome and liver diseases. *Gut* **65**, 2035–2044 (2016).

55. Zuo, T. et al. Alterations in gut microbiota of patients with COVID-19 during time of hospitalization. *Gastroenterology* **159**, 944–955.e8 (2020).

56. Chirinos, J. A. FEnofibRate as a Metabolic INtervention for Coronavirus Disease 2019. *NIH* https://clinicaltrials.gov/ct2/show/NCT04517396 (2021).

57. Kim, N. H., Han, K. H., Choi, J., Lee, J. & Kim, S. G. Use of fenofibrate on cardiovascular outcomes in statin users with metabolic syndrome: propensity matched cohort study. *Brit. Med. J.* **366**, l5125 (2019).

58. Knodell, R. G. et al. Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. *Hepatology* **1**, 431–435 (1981).

59. Evans, A. M., DeHaven, C. D., Barrett, T., Mitchell, M. & Milgram, E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal. Chem.* **81**, 6656–6667 (2009).

60. Bridgewater BR, E. A. High resolution mass spectrometry improves data quantity and quality as compared to unit mass resolution mass spectrometry in high-throughput profiling metabolomics. *J. Postgenomics Drug Biomark. Dev.* **4**, 2 (2014).

61. DeHaven, C. D., Evans, A. M., Dai, H. & Lawton, K. A. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *J. Cheminformatics* **2**, 9 (2010).

62. Connelly, M. A., Shalaurova, I. & Otvos, J. D. High-density lipoprotein and inflammation in cardiovascular disease. *Transl. Res. J. Lab. Clin. Med.* **173**, 7–18 (2016).

63. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *Mach. Learn. PYTHON* 6 https://scikit-learn.org

64. Ortiz, A. M. et al. Experimental microbial dysbiosis does not promote disease progression in SIV-infected macaques. *Nat. Med.* **24**, 1313–1316 (2018).

65. Kultima, J. R. et al. MOCAT2: a metagenomic assembly, annotation and profiling framework. *Bioinformatics* **32**, 2520–2523 (2016).

66. Ehrlich, S. D. in *Metagenomics of the Human Body* (ed. Nelson, K. E.) 307–316 (Springer, 2011).

67. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).

68. Pedersen, H. K. et al. A computational framework to integrate high-throughput '-omics' datasets for the identification of potential mechanistic links. *Nat. Protoc.* **13**, 2781–2800 (2018).

69. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 161 (2009).

70. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

71. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).

72. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).

73. Reimand, J. et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* **14**, 482–517 (2019).

Acknowledgements

Author contributions

All authors had a substantial contribution to this work. All authors provided approval for the final submitted version of the manuscript. Acquisition, analysis and interpretation of data were performed by all authors. G.M.Q., R.U., J.A.H., E.C.T., M.G. and O.E. substantively revised the work. O.E. contributed to the design of the work. R.O.A. and T.H. were responsible for the design of the work, the acquisition, analysis and interpretation of data, drafted the initial work and substantively revised it. Additionally, T.H. was responsible for conception of the work.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-022-01273-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-022-01273-y.

Correspondence and requests for materials should be addressed to Rabab O. Ali or Theo Heller.

Peer review information *Nature Microbiology* thanks Pieter Dorrestein, Nobuhiko Kamada, Eric Meissner and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Rabab O. Ali [1]✉, **Gabriella M. Quinn**[1], **Regina Umarova**[2], **James A. Haddad** [1], **Grace Y. Zhang**[1], **Elizabeth C. Townsend** [1], **Lisa Scheuing**[1], **Kareen L. Hill**[1], **Meital Gewirtz**[1], **Shakuntala Rampertaap**[3], **Sergio D. Rosenzweig**[3], **Alan T. Remaley**[4], **Jung Min Han**[5], **Vipul Periwal**[5], **Hongyi Cai**[6], **Peter J. Walter** [6], **Christopher Koh**[2], **Elliot B. Levy**[7], **David E. Kleiner** [8], **Ohad Etzion**[1] & **Theo Heller** [1]✉

[1]Translational Hepatology Section, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. [2]Liver Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. [3]Immunology Service, Department of Laboratory Medicine, Clinical Center, National Institutes of Health, Bethesda, MD, USA. [4]Cardiovascular and Pulmonary Branch of the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA. [5]Computational Medicine Section, Laboratory of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. [6]Clinical Mass Spectrometry Core, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. [7]Center for Interventional Oncology, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, MD, USA. [8]Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. ✉e-mail: rabab.ali226@gmail.com; theoh@intra.niddk.nih.gov
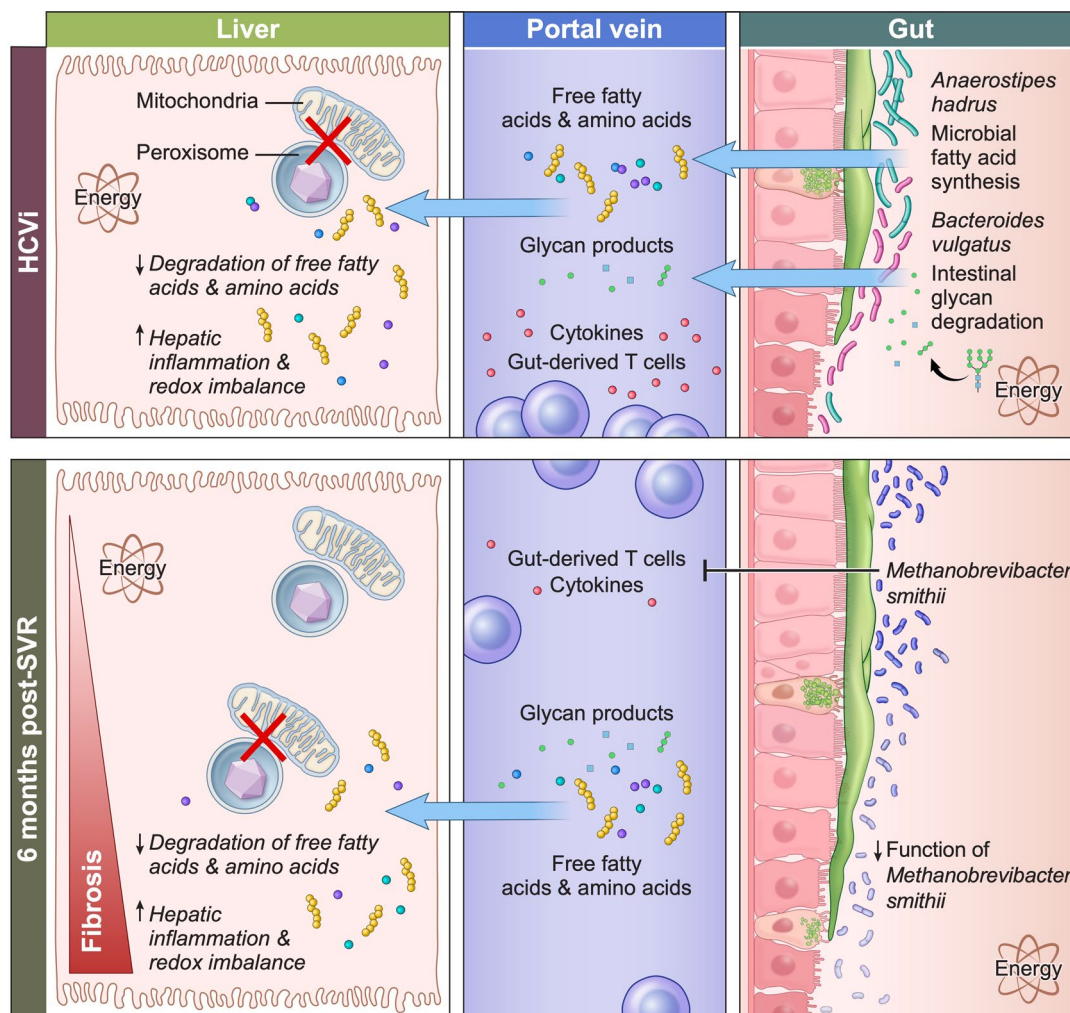
Extended Data Fig. 1 | Graphical Abstract. Hepatic metabolism in peroxisomes and mitochondria is decreased in chronic HCV infected patients (HCVi) compared to SVR. Fibrosis and necroinflammation in HCVi were linked to increased transcriptional activity of *Anaerostipes hadrus* mediated fatty acid synthesis and *Bacteroides vulgatus* mediated intestinal glycan degradation. Microbial-derived fatty acids and glycan products are elevated in portal circulation and linked to enhanced portal and hepatic inflammation in HCVi. Despite decreased hepatic and portal inflammation six months after SVR, hepatic metabolism and peroxisome function remains decreased in SVR with advanced fibrosis. *Methanobrevibacter smithii* showed decreased function in SVR fibrosis and may have anti-inflammatory properties.
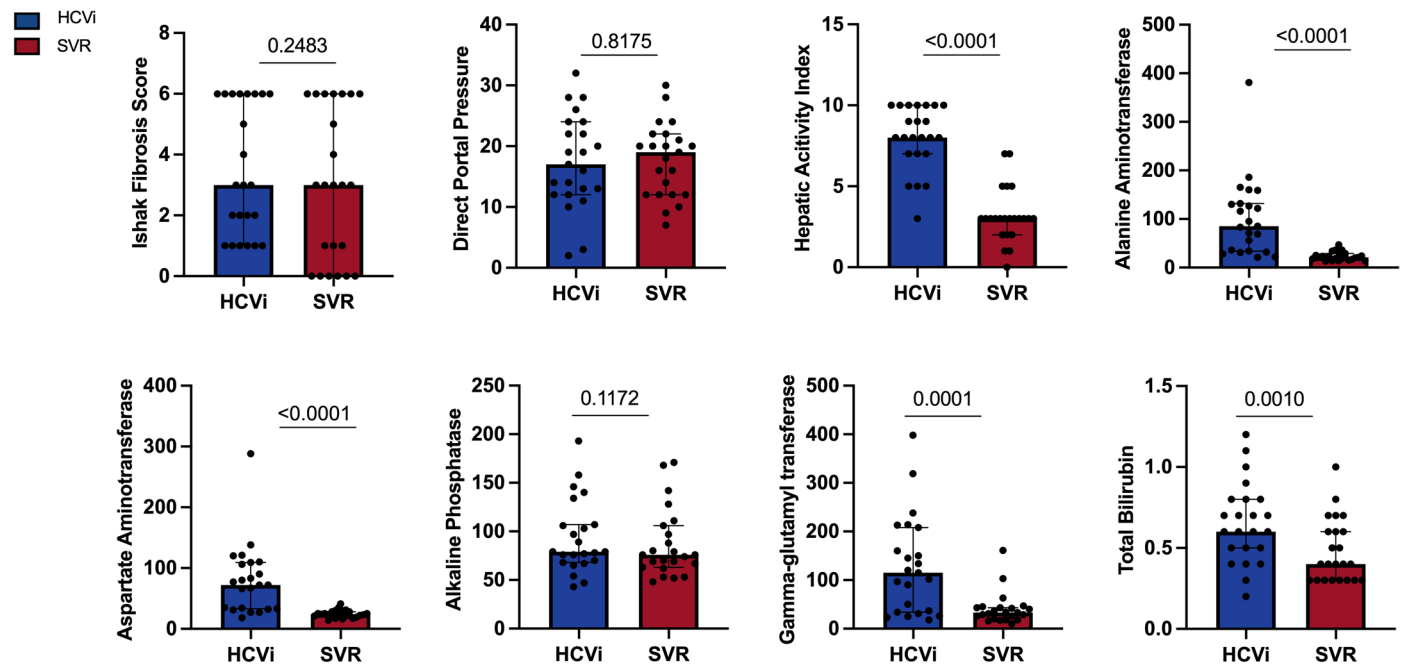
Extended Data Fig. 2 | Biochemical and histological markers of inflammation were elevated in HCVi compared to SVR. Wilcoxon matched pairs signed rank test, two-sided unadjusted p-value. Compared to SVR, HCVi patients had elevated serum markers of hepatocellular inflammation (ALT, alanine aminotransferase p < 0.0001; AST, aspartate aminotransferase p < 0.0001; GGT, gamma-glutamyl transferase p = 0.0001); histological marker of inflammation (HAI, Hepatic Activity index) p < 0.0001; and elevated serum total bilirubin p = 0.0010. HCVi showed no significant difference in fibrosis p = 0.2483, direct portal pressures p = 0.8175, or alkaline phosphatase p = 0.1172. Scatter plots with bars, data is presented as median values +/- IQR. *n* = 22.
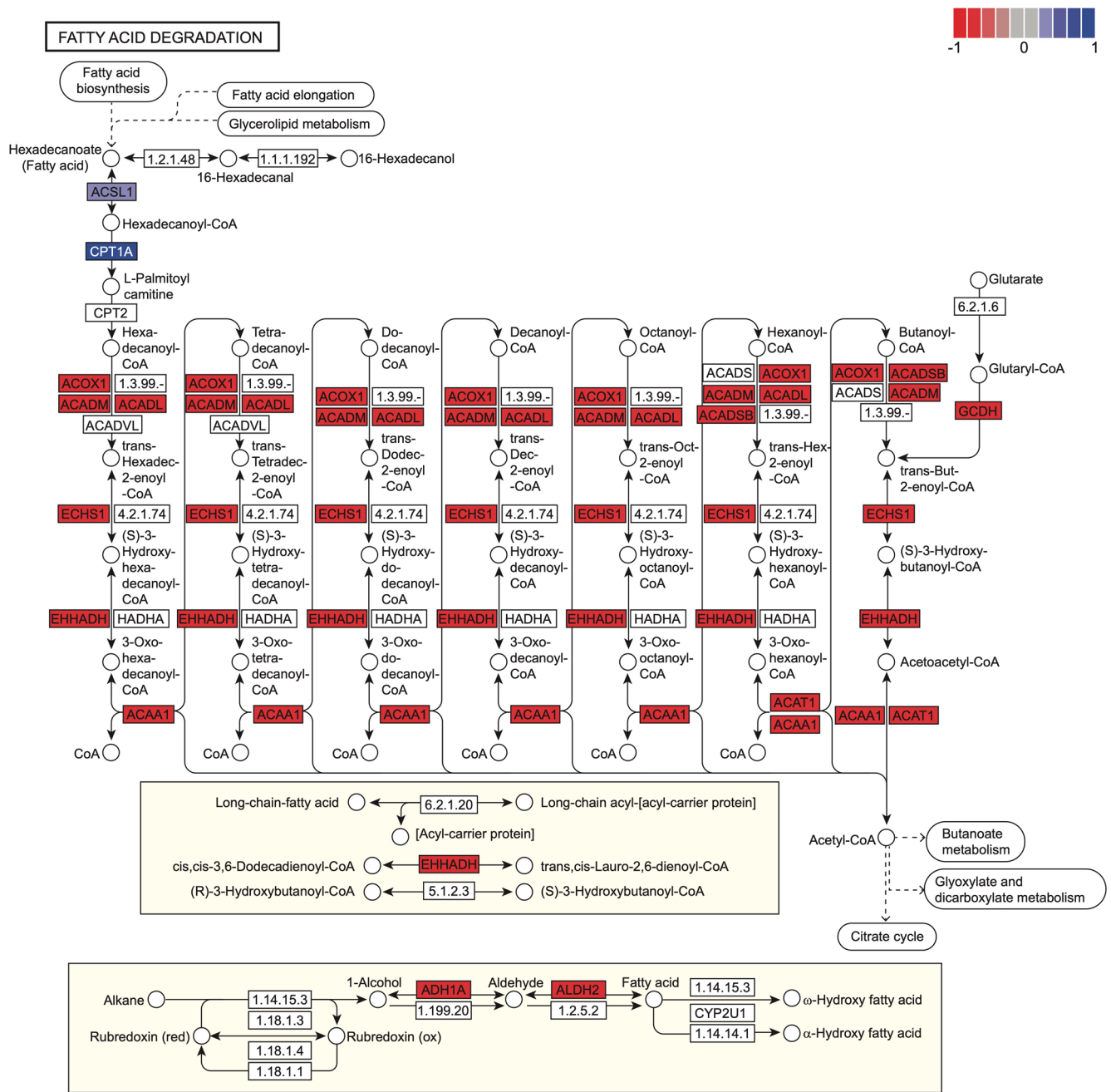
Extended Data Fig. 3 | Decreased hepatic fatty acid degradation in HCVi when compared to SVR. Graphical representation using GAGE R of the hepatic KEGG functional pathway 'Fatty Acid Degradation' significantly downregulated in HCVi compared to SVR (FDRp < 0.1). Within each pathway significant DEGs are highlighted blue for fold-change >0 and red for fold-change<0 in HCVi when compared with SVR (FDRp < 0.1). n = 22.

a.

b.
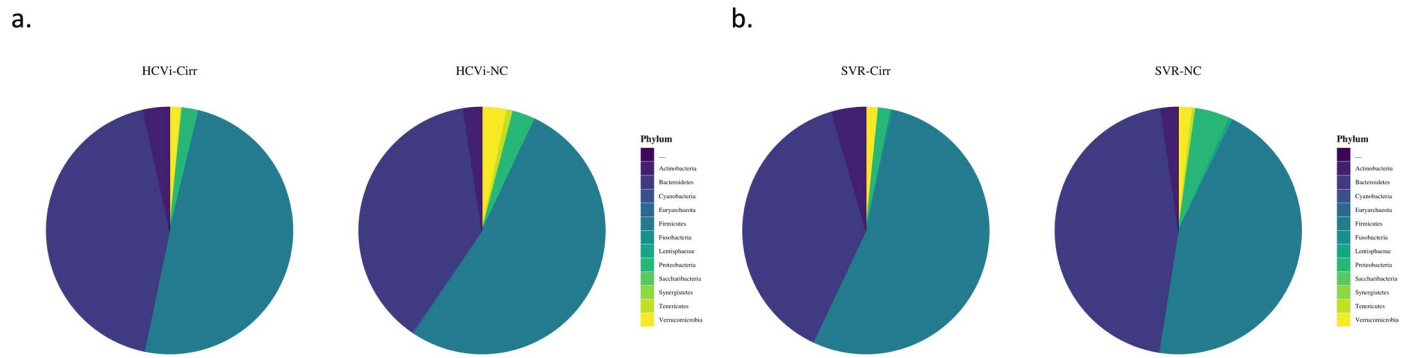
Extended Data Fig. 4 | No significant difference in gut microbial composition at phylum level based on presence of cirrhosis in HCVi or SVR. Fecal 16 S rRNA analysis was performed on fecal samples. Relative abundance of phyla plotted in pie-charts for patient subgroups based on cirrhosis. **a)** HCVi-Cirr ($n = 13$) and HCVi-NC ($n = 16$). **b)** SVR-Cirr ($n = 9$) and SVR-NC ($n = 14$). Analysis performed using QIIME.
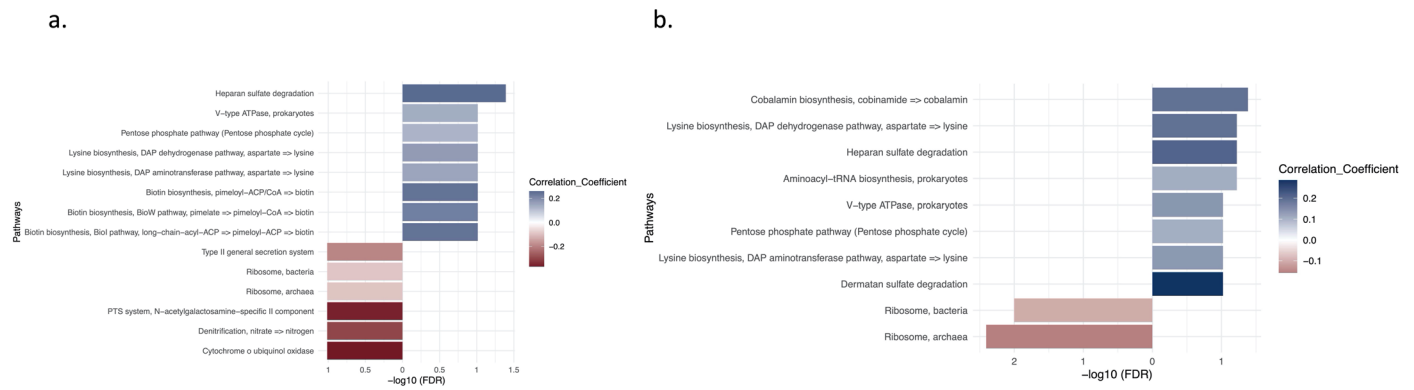
Extended Data Fig. 5 | Microbial metatranscriptome analysis revealed distinct microbial functions associated with hepatocellular injury in HCVi. Within HCVi cohort, AST **(a)** and GGT **(b)** significantly correlated with microbial KEGG functional modules for glycan degradation including heparan and dermatan sulfate degradation (SCC$_{bg.adj}$, two-sided, FDRp<0.1). $n = 26$.
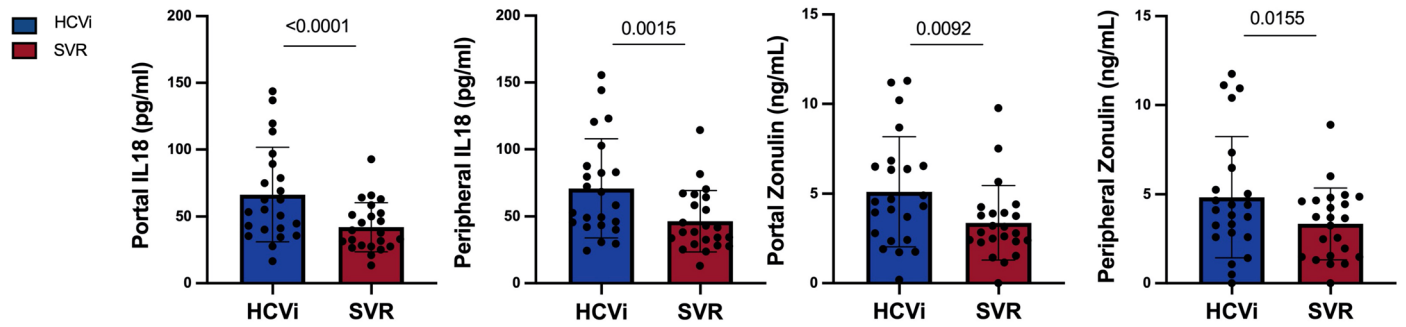
Extended Data Fig. 6 | Serum markers of gut dysbiosis and intestinal dysfunction elevated in HCVi when compared to SVR. Paired comparison of serum IL18 (marker of dysbiosis) and zonulin (marker of gut epithelial integrity) between HCVi and SVR cohorts (two-sided Wilcoxon matched pairs signed rank test, HCVi vs. SVR). Scatter plots with bars, data is presented as median values +/- IQR. Relative quantification of metabolites per Metabolon protocol. n = 23.

nature portfolio

Corresponding author(s):   Theo Heller, Rabab O. Ali

Last updated by author(s):   08/03/2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	There was no custom code or mathematical algorithm utilized in this study. See Methods for details of the commercial and publicly available software utilized in this study.
Data analysis	The following commercial and publicly available softwares were utilized in this study: R software (versions 3.5.0 and 4.0.2); Softwares Partek® Flow® (Version 10.0); DESeq2 R package (version 1.34.1); SNFtool and bnstruct R packages on R 4.0.2; Python package Scikit-Learn (version 3.9.5); Python module SparCC (Python 3.9.5); SOAPAligner (version 2.21); SOAPDenovo software (version 2.04); MetaGeneMark software (version 3.38); BLASTN (version 2.10.0+); GAGE R package (version 2.201); WGCNA (version 1.69); NetworkAnalyst (version 2019); PathView R package (version 1.20.1); and MicrobiomeAnalystR package.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Please note that the microbial and liver transcriptome sequence dataset has been made available in the BioProject repository. The DOI URL is https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA727609. The serum metabolomics data has been uploaded as Source Data

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Prior to enrollment, we calculated the sample size as follows: sample size of 7 patients per group will provide the study with a statistical power of 80% at a 95% confidence level, to detect a difference of 60% in portal vein microbial product detection rate between the two groups. The sample size was increased to 10 patients in the minimal fibrosis group and 20 patients in the advanced fibrosis group to allow for withdrawals and technical failures that might occur at a higher rate in cirrhotic patients.
Data exclusions	During the first evaluation of the 30 patients that signed informed consent, one patient was excluded for incidental finding of HCC after signing consent, n=29 in the HCVi cohort. During the second time-point evaluation, of the 24 patients that re-signed consent, one patient re-consented but did not complete sample collection after SVR, thus was excluded from all analysis at the SVR time-point, n=23 in SVR cohort. Due to failure of patient samples to meet quality control parameters (as described in the Methods), the hepatic transcriptome had n=27 in HCVi cohort, and n=23 for the SVR cohort. Similarly due to data filtering parameters, the microbial transcriptome had n=26 in HCVi cohort and n=23 in SVR cohort. For all paired analysis between HCVi and SVR, n=22. Please refer to the details in Methods Section.
Replication	Given the nature of the clinical trial and limited samples collected (esp. portal vein cannulation) replication of experiments was not possible.
Randomization	This was not a randomized control trial. The study was a proof-of-concept. As such, the intent was not to perform a randomized controlled study. Rather, the design of the study was aimed at comparing paired samples before and after an intervention (HCV treatment).
Blinding	The collection and final data analysis was performed unblinded as all samples were labeled to ensure accuracy and the final analysis required grouping patients by disease severity or HCV treatment status. However, all experimental techniques were performed blinded with the samples coded and randomized. The key for the codes was only broken after data acquisition to perform final analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	EDTA anticoagulated peripheral and portal blood samples were processed for flow cytometry using a whole blood lysis method, stained with fluorescent antibodies, collected with a FACS Canto II (Becton Dickinson [BD], San Jose, CA.) and analyzed using FCS Express software (De Novo, Glendale, CA). Lymphocytes were identified based on a gate established by forward and side angle scatter and confirmed using anti-CD45 and anti-CD14. B-cell were identified by directly conjugated monoclonal antibodies. B cells were identified with the following antibodies: anti-CD20, anti-CD19, anti-CD5, anti-CD10, anti-IgM, anti-CD38, and anti-CD27. Irrelevant, directly conjugated, murine IgG1 was used to ascertain background staining. All monoclonal antibodies were obtained from Becton Dickinson (San Jose, CA), except for anti-Vβ-11 and anti-CD45RA (Beckman-Coulter, Brea, CA), anti-IgM and anti-Vα-24 (BioLegend, San Diego, CA), anti-CD4, anti-CD45, anti-CD14, anti-CD19, anti-CD10, and anti-CD27 (Life Technologies, Carlsbad, CA), and anti- α4β7 (NIH AIDS Reagent Program, NIAID, NIH). Refer to SuppTable5 for information on antibody dilutions/amounts, company names, and catalog numbers for antibodies used.

Validation	All antibodies are validated running 40 healthy controls in the same format (cell population) as in patient protocol.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Please refer to Manuscript Table 1 for the covariate-relevant population characteristics and to Supplementary Table 2 for the relevant past and current diagnoses. Briefly, 29 patients with chronic hepatitis C were enrolled with a median age of 59, 18 of which were males, 19 Caucasians, 5 African American, 2 Asian, and 3 Hispanics. After HCV treatment, 23 patients with a median age of 58 completed re-evaluation after SVR, 14 of whom were males, 15 Caucasians, 5 African American, 1 Asian, and 2 Hispanics.
Recruitment	Patients were recruited to the study from the Hepatology Clinic at the National Institutes of Health Clinical Center. A total of 36 consecutive newly referred patients with hepatitis C were initially approached for enrollment. Of these, 2 did not meet eligibility criteria, and 4 declined enrollment. If patients were agreeable, the study was explained in detail and informed consent was obtained. Of the 30 that signed consent, 1 patient was excluded due to HCC detected only after signing consent. As patients were enrolled consecutively from a general Hepatology Clinic, self-selection bias was minimized. We did not recognize any other major biases that might have a significant impact on the results particularly as the design was a proof-of-concept non-randomized study.
Ethics oversight	The National Institute of Diabetes, Digestive and Kidney Diseases, and the National Institute of Arthritis and Musculoskeletal Diseases, Institutional Review Board at the National Institutes of Health.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT02400216
Study protocol	The full trial protocol is appended in the Supplementary Information.
Data collection	All data was collected and processed at the National Institutes of Health Clinical Center. The recruitment began on 5/29/2015 and ended on 3/11/2016. Data collection began on 6/1/2015 and ended on 2/14/2017.
Outcomes	As the design was a proof-of-concept non-randomized study, there were no relevant measures for clinical outcomes. Of note, all 29 patients from initial recruitment responded to direct acting antiviral therapy for HCV and achieved sustained virologic response.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	EDTA anti-coagulated peripheral and portal blood samples were processed for flow cytometry using a whole blood lysis method. See Online Methods for details.
Instrument	FACS Canto II (Becton Dickinson [BD]
Software	FCS Express software (De Novo, Glendale, CA)
Cell population abundance	Lymphocytes
Gating strategy	A back gating method is employed using CD45 vs CD14 to identify lymphocytes. Subsequently, a FSC vs SSC gate is used to identify the cell population and then gate on CD45 population.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.