Quantitative genome-scale analysis of human liver reveals dysregulation of glycosphingolipid pathways in progressive nonalcoholic fatty liver disease

Running title: Genome-scale metabolic model of progressive NAFLD

Partho Sen[1,2,*], Olivier Govaere[3], Tim Sinioja[4], Aidan McGlinchey[1], Dawei Geng[4], Vlad Ratziu[5], Elisabetta Bugianesi[6], Jörn M. Schattenberg[7], Antonio Vidal-Puig[8], Michael Allison[9], Simon Cockell[3], Ann K. Daly[3], Tuulia Hyötyläinen[4], Quentin M. Anstee[3,10,*§], Matej Orešič[1,2,* §]

[§]Joint Senior Authors

[1]School of Medical Sciences, Örebro University, 70281 Örebro, Sweden

[2]Turku Bioscience, University of Turku and Åbo Akademi University, FI-20520 Turku, Finland

[3]Translational & Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK

[4]Department of Chemistry, Örebro University, 70281 Örebro, Sweden

[5]Assistance Publique-Hôpitaux de Paris, hôpital Beaujon, University Paris-Diderot, Paris, France

[6]Department of Medical Sciences, Division of Gastro-Hepatology, A.O. Città della Salute e della Scienza di Torino, University of Turin, Turin, Italy

[7]Metabolic Liver Research Programm, Department of Medicine, University Hospital Mainz, Mainz, Germany

[8]University of Cambridge Metabolic Research Laboratories, Wellcome-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK

[9]Liver Unit, Department of Medicine, Cambridge Biomedical Research Centre, Cambridge University NHS Foundation Trust, United Kingdom

[10]Newcastle NIHR Biomedical Research Center, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

**Correspondence*: quentin.anstee@newcastle.ac.uk, matej.oresic@oru.se

ABSTRACT

Nonalcoholic fatty liver disease (NAFLD) is a well defined chronic liver diseases closely related with metabolic disorders. The prevalence of NAFLD is rapidly increasing worldwide, while the pathology and the underlying mechanisms driving NAFLD are not fully understood. In NAFLD, a series of metabolic changes takes place in the liver. However, the alteration of the metabolic pathways in the human liver along the progression of NAFLD, *i.e.*, the transition from nonalcoholic steatosis (NAFL) to steatohepatitis (NASH) through cirrhosis remains to be discovered. Here, we sought to examine the metabolic pathways of the human liver across the full histological spectrum of NAFLD. We analyzed the whole liver tissue transcriptomic (RNA-Seq) and serum metabolomics data obtained from a large, prospectively enrolled cohort of histologically characterized patients derived from the European NAFLD Registry (n=206), and developed genome-scale metabolic models (GEMs) of human hepatocytes at different stages of NAFLD. The integrative approach employed in this study has enabled us to understand the regulation of the metabolic pathways of human liver in NAFL, and with progressive NASH-associated fibrosis (F0–F4). Our study identified several metabolic signatures in the liver and blood of these patients, specifically highlighting the alteration of vitamins (A, E) and glycosphingolipids (GSLs), and their link with complex glycosaminoglycans (GAGs) in advanced fibrosis. The study provides insights into the underlying pathways of the progressive fibrosing steatohepatitis. Furthermore, by applying genome-scale metabolic modeling (GSMM), we were able to identify the metabolic differences among carriers of widely validated genetic variants associated with NAFLD / NASH disease severity in three genes *(PNPLA3*, *TM6SF2* and *HSD17B13)*.

KEYWORDS: Nonalcoholic fatty liver disease; NAFLD; NASH; steatohepatitis; metabolic models; pathways; transcriptomics; lipidomics; metabolomics; gene variants; genotypes.

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) represents a histological spectrum characterized by increased lipid accumulation in the hepatocytes, that encompasses non-alcoholic fatty liver (steatosis, NAFL) and an inflammatory form termed as nonalcoholic steatohepatitis (NASH) in which progressive hepatic fibrosis occurs, ultimately leading to cirrhosis and hepatocellular carcinoma (HCC) in some patients[1]. NAFLD is associated with features of the metabolic syndrome including obesity, type 2 diabetes mellitus (T2DM), dyslipidemia, and hypertension[2,3]. The global prevalence of NAFLD in the general population has been estimated to be 25%, whilst the prevalence of NASH has been estimated to range from 3 – 5%[4]. The prevalence of NAFLD has been increasing proportionately with the epidemics of obesity and is predicted to continue to rise[5].

The natural history of NAFLD is highly variable, with substantial interpatient variation in disease severity and outcome predicted by the degree of fibrosis[6,7]. NAFLD is generally considered to be a complex disease trait driven by an obesogenic environment acting on a background of genetic susceptibility[8]. However, knowledge about the processes that contribute to the observed variation in severity remains incomplete. There is a pressing need to elucidate the pathophysiological processes that occur as NAFL progresses to NASH and fibrosis stage increases towards cirrhosis. Such knowledge will define novel biomarkers that better risk stratify patients, helping to individualize their care, and aid the development of novel therapeutic strategies and drug targets.

Progression of NAFLD is characterized by distinct metabolic changes in the liver[9]. Previously, we showed that there is an excess accumulation of liver triacylglycerols (TGs) in NAFLD, particularly those with low carbon number and double bond content[10], which reflects increased *de novo* lipogenesis in NAFLD[11,12]. Over the past decade, GSMM has emerged as a powerful tool to study metabolism in human cells[13-15], including in modeling metabolism of human liver under healthy and disease conditions[16-18]. In a recent study, we charted metabolic activities associated with degree of steatosis in human liver from 16 NAFLD cases by integrating genome-wide transcriptomics data from human liver biopsies, and metabolic flux data measured across the human splanchnic vascular bed using genome-scale metabolic modeling (GSMM)[18]. A similar GSMM study in 32 patients with NAFLD plus 20 controls identified serine deficiency and suggested that the serum concentration of glycosaminoglycans (GAGs) (e.g., chondroitin and heparan sulphates) may aid the staging of NAFLD[17]. Both studies, whilst informative, only used transcriptomic data from relatively small patient cohorts with crude histological characterization of disease severity and so neither was able to adequately address grade of steatohepatitis or stage of fibrosis with sufficient granularity.

In this study, we aim to examine the changes in hepatic metabolic processes and pathways that occur across the full histological spectrum of NAFLD, exploring the evolving changes during the progression of NAFL to NASH with increasing fibrosis stages (F0 up to F4) including cirrhosis. Whole liver tissue RNA-Seq transcriptomic data from a large cohort of histologically characterized patients derived from the European NAFLD Registry (n=206)[19,20] was integrated into genome-scale metabolic models (GEMs) (**Table 1** and **Figure 1**). The integrative approach employed in this study has informed our understanding of how liver metabolism is altered in NAFL and with progressive NASH-associated fibrosis. Furthermore, GSMM predicted the metabolic differences among carriers of widely validated genetic variants associated with NAFLD / NASH disease severity in three genes

(PNPLA3, *TM6SF2* and *HSD17B13)*[8,21-23].

RESULTS

Genome-scale metabolic models of human hepatocytes in NAFLD

A GEM of human hepatocyte (*iHepatocytes2322*) was first developed by Mardinoglu *et al.*, by combining the *iHepatocyte1154* GEM[24] with previously published liver models. *iHepatocytes2322* comprises 2,322 genes, 5,686 metabolites and 7,930 reactions found in the human liver[17]. Here, by using NAFLD stage-specific transcriptomics data generated from the liver biopsies of 206 patients with NAFLD (**Table 1**)[20], we contextualized the *iHepatocytes2322* model (**Methods**), and developed GEMs of hepatocytes across the full spectrum of NAFLD severity (NAFL through NASH F0-F4) (**Figure 1**). A stepwise strategy that combines *GIMME*[25] and *E*-flux[26] algorithm was developed for the contextualization of GEMs, *i.e.,* selection of stage-specific reactions and their associated genes and metabolites from the *iHepatocytes2322* (**Methods**). The number of genes, metabolites and reactions included in these models are given in (**Figure S1**).

The GEMs were used to study the metabolic differences across the NAFLD spectrum. To increase the statistical power of differential expression analysis of the transcriptomic data required for GSMM, cases were grouped by disease severity and clinical implications for outcome[7]: 'minimal' disease – NAFL + NASH (F0-1) (n= 85); 'mild' disease – NAFL + NASH (F0-2) (n= 138); 'clinically significant' non-cirrhotic fibrosis – (NASH F2-3) (n= 107); and 'advanced' fibrosis – (NASH F3-4) (n= 68).

Regulation of metabolic subsystems of human hepatocytes in steatosis and progressive stages of fibrosis

Flux variability analysis (FVA) is a computational method used to estimate the minimum and maximum flux carried by each metabolic reaction of a GEM[27], in this case, the stage-specifc hepatocyte GEMs developed. A high or low flux span (max – min reaction flux) denotes the extent of variability (metabolic alteration) exhibited by each reaction within a metabolic pathway or subsystem of a GEM (**Methods**).

FVA has shown remarkable alterations in the metabolic subsystems across various stages of NAFLD (**Figure 2**). Average flux spans of the reaction fluxes in the metabolic subsystems were markedly higher at NASH F4, whilst lower flux spans were evident at NAFL and NASH F01 stages (**Figure 2a**), suggesting that alteration of metabolic pathways were higher in NASH F4 as compared to the other stages of NASH-associated fibrosis. In contrast, lower flux spans at NASH F01 and NAFL suggests, at these stages, the metabolic pathways of hepatocytes are tightly regulated[28,29] (**Figure 2a-b**). Intriguingly, the results of FVA showed that the alterations of metabolic pathways in NASH F01 were similar to NAFL (**Figure 2a**).

FVA identified 55 metabolic subsystems altered across one or more stages of NAFLD (**Figure 2b**). In NAFL, the % of relative average flux span (RFS) of the metabolic subsystems associated with leukotriene, glycerophospholipid (GPL), ascorbate and aldarate, retinol (vitamin A), riboflavin (vitamin B1), nicotinamide metabolism (vitamin B3), and fatty acid activation were increased (>20%) (**Figure 2b**). Higher %RFS values across the ether lipid metabolism, pantothenate and CoA

biosynthesis were observed at NASH F01 as compared to NAFL (**Figure 2a-b**). In NASH F2, %RFS values across the bile acid, steroid and alanine, aspartate glutamate metabolism was increased. Intriguingly, NASH F3 and F4 stages showed a proportional change of % RFS values across several metabolic subsystems, except for the aromatic amino acid, fatty acid, vitamin (A, B1) metabolism which showed higher %RFS values at NASH F4 (**Figure 2b**).

Next, we performed an unbiased non-uniform random sampling (RS)[30] and estimated the flux differences between the metabolic subsystems of the human hepatocytes at various stages of NAFLD. Our results suggest that, fatty acid (FA) metabolism was elevated whilst the β-oxidation of FA was decreased in NASH (F0-4) as compared to NAFL (**Figure 3**). These changes were also apparent in 'clinically significant' non-cirrhotic and / or 'advanced' (*vs.* 'minimal') fibrosis. At these stages, vitamin (A, B1, D, E) metabolism was differentially regulated; average fluxes of vitamin E and D were decreased whilst an increase in retinol metabolism (vitamin A) was observed (**Figure 3**). Intriguingly, an increase in the sphingolipid (SL) metabolism and a decrease in the glycosphingolipid (GSL) (cerebrosides / globo-, lacto- and neolacto series) metabolism was marked with the progression of NASH-associated fibrosis (**Figure 3**). Moreover, the average fluxes of GAGs, particularly chondroitin, heparin and keratin sulphate, cholesterol, aromatic amino acid, ether lipid, eicosanoid metabolism, bile acid biosynthesis and glycolysis / gluconeogenesis were altered at one or more stages of NASH-associated fibrosis *vs.* NAFL (**Figure 3**).

Reporter metabolite analysis of human hepatocytes unveiled the intrinsic regulation of GSLs, GAGs and cholesterols in advance fibrosis

Reporter metabolite (RM) analysis is a metabolite-centric differential approach that aids in identifying metabolites in a network around which significant transcriptional changes occur[31,32]. RMs can predict hot spots in a metabolic network that are significantly altered between two different conditions, in this case, different stages of NAFLD.

RM analysis revealed that several species of eicosanoids, estrone, steroids and retinoic acids were increased ($p<0.05$) in NASH F2 stage as compared to NAFL (**Figure S2**). RMs of NASH F3 (*vs.* NAFL) showed a remarkable pattern with several species of GSLs, particularly, cerebrosides (glucosylceramides (GlcCers), lactosylceramides (LacCers), digalactosylceramides), globosides and gangliosides (GM1-, GM2-alpha) were decreased. At this stage, RMs of GAGs (heparin, keratin sulphates) and cholesterols were also decreased. Intriguingly, these GAGs were linked to GSLs by 3'-Phosphoadenosine-5'-phosphosulfate (PAPs) and globotriaosylceramides which were also decreased (**Figure 4**). In addition, RMs of multiple eicosanoids, prostaglandins (PGs), leukotrienes (LTs) and arachidonic and retinoic acid derivates were increased (**Figure 4**).

Interestingly, RMs of GSLs, GAGs, and cholesterols remain decreased in cirrhosis (NASH F4 vs. NAFL) (**Figure S3**). A similar directional change in the RMs of GSLs, GAGs, cholesterols, gamma-tocopherols (Vitamin E) was observed in 'advanced' fibrosis – (NASH F3-4) vs. 'minimal disease – NAFL + NASH (F0-1) (**Figure S4**) and 'mild disease' – NAFL + NASH (F0-2), respectively (**Figure S5**). In addition, RMs of aromatic and sulphur containing amino acids, *i.e.*, phenylalanine, tyrosine, tryptophan, histidine and methionine were increased in the 'advanced' fibrosis vs. 'mild disease'. These amino acids were associated with apolipoproteins (apo), primarily involved in the formation

of low-density lipoprotein particles (LDL and VLDL). Sensitivity analysis of the hepatocyte models showed that, with the progression of NASH-associated fibrosis the intracellular concentration of apolipoprotein, fatty acyl carrier protein (ACP) and cholesterols tends to be limiting towards the liver biomass (Shadow price[33], SP > 0) (**Figure S6**). However, no significant change in the GSLs, GAGs, cholesterols was observed in the 'clinically significant' non-cirrhotic fibrosis – (NASH F2-3) vs. 'minimal disease' (**Figures S7** and **S8**).

Regulation of sphingolipid and ceramide pathways in advanced fibrosis

GSLs and Cers are essential intermediates of sphingolipid (SL) pathways (**Figure 5a**). They play a significant role in maintaining the integrity of the plasma membrane. Our results suggest that, several species of GSLs and related pathways were altered in 'advanced' fibrosis (**Figures 3-4** and **S8**). At this stage, the GSLs were associated with GAGs (heparin, keratin sulphates) and cholesterols (**Figures 4** and **S8**). Therefore, the intracellular regulation of SL pathways might play a critical role in the progression of fibrosis.

Next, we identified the metabolic genes (MGs) of the sphingolipid (SL) pathways that were differentially expressed (p.adjusted < 0.05) with the progression of NASH-associated fibrosis. MGs and fluxes of *denovo* SL pathway (R2, R3, R4) were downregulated in 'advanced' fibrosis *vs.* 'minimal' and 'mild' disease stages (**Figure 5a-c**). Moreover, the expression of sphingomyelinase (*SMase)* and the flux through the (R5: *SMase*) pathway was increased, which suggests production of ceramides (Cers) from sphingomyelin (SM) *via SMase* (R5) is facilitated in 'advanced' fibrosis as compared to 'minimal' and 'mild' disease stages respectively (**Figure 5a-c**). The majority of MGs and fluxes of the GSL pathways, notably (R7, R8, R9a) were downregulated along the fibrosis progression (**Figure 5a-c**). A marked increase in the activity of the sphingosine recycling pathway (R10) was also observed. Taken together, in 'advanced' fibrosis the hepatocytes tend to enhance the Cer production from SMs. It also decreases the synthesis of GSLs, particularly GlcCers, LacCers and GalCers from Cers (**Figure 5a-c**). Excessive accumulation of Cers in the hepatocyte may be required to exert cellular signaling, modulate inflammatory response or generate apoptotic stress signals[34,35].

Marked alterations of Cers and GSLs levels in the serum of the patients with advanced fibrosis

By applying lipidomics to serum samples obtained from NAFLD patients in the present study, we measured the levels of Cers and GSLs. We found that the serum levels of several Cer species were increased (Tukey's HSD, $p<0.05$) (**Figure 6a-d**), whilst GSLs (HexCers, GlcCers) were decreased (Tukey's HSD, $p<0.05$) in the 'advanced' fibrosis (NASH F3) as compared to 'minimal' and / or 'mild' disease stages (**Figure 6e-f**). Overall, the directional changes (up or downregulated) of the predicted RMs of GSLs in the liver were comparable with the differences (increase or decrease) in the serum GSL levels of the patients at different stages of NASH-associated fibrosis. Intriguingly, there was an abrupt decrease in the levels of Cers and HexCers at NASH F4 (cirrhosis) stage, which may be attributed to the alterations in the SL pathway(s) in the liver.

Impact of genetic variants associated with NAFLD on liver metabolic pathways

Genetic susceptibility plays a vital role in the development of NAFLD[8,36]. Here, we studied the most widely validated genetic modifier and strongest genetic risk factors for NAFLD, *i.e.*, PNPLA3 (rs738409), TM6SF2 (rs58542926) and HSD17B13 (rs72613567)[8,21,36-38]. Some of these gene variants are known to alter TAG hydrolysis[39] and retinol (Vitamin A) metabolism[40,41] and increase the severity of NAFLD[37,39,42]. However, little is known about the metabolic differences conferred by these variants.

Here, we developed GEMs for hepatocytes in the individuals carrying one of these three common gene variants, *PNPLA3* (GC, GG) (n=69)*, TM6SF2* (CT, TT) (n=13)*, HSD17B13* (-T, TT) (n=21) compared with wild type (WTs) (n=36) *i.e.*, individuals homozygous for all the three gene variants: *PNPLA3 (CC) , TM6SF2 (CC) and HSD17B13 (--)* (**Tables 1** and **S1**)*,* using genome-wide transcriptomics data and the *iHepatocytes2322* model. The number of metabolites and reactions included in these models are given in (**Figure S1**). Selection of individuals exclusively carrying *PNPLA3* or *TM6SF2* or *HSD17B13* gene variants from the study cohort[43] is given in (**Table S1**).

FVA of the hepatocyte models of the gene variants showed that, the average flux span across the metabolic subsystems in *PNPLA3* variant carrier was markedly higher than *TM6SF2 and HSD17B13* variant carriers, whilst lowest flux variation was found in the WT model (**Figure 7a**). FVA identified 31 metabolic subsystems that were altered in at least one gene variant carrier (**Figure 7b**). Interestingly, *PNPLA3* variant carriers showed higher values of %RFS (>50%) across the ether lipid, vitamin E, leukotriene, starch and sucrose, pantothenate and CoA biosynthesis, and glycerophospholipid and tryptophan metabolism (**Figure 7b**). Furthermore, *PNPLA3* (heterozygous: GC, n=47) variant carriers showed larger %RFS values across these subsystems as compared to (homozygous: GG, n=22) variant carriers (**Figure S9**). On the otherhand, lower values of %RFS was marked across fatty acid activation and biosynthesis in the hepatocytes of the *PNPLA3* variant carrier as compared to *TM6SF2 and HSD17B13* variant carriers, suggesting that fatty acid metabolism might be tightly regulated and essential for the cellular function of the *PNPLA3* variant carriers (**Figure 7b**). Also, random sampling of the flux states of *PNPLA3* variant carriers *vs.* WT also showed an increase in the fluxes across the ether lipid and fatty acid metabolism (**Figure S10**).

DISCUSSION

GSMM of human hepatocytes in NASH-associated fibrosis (n=206) identified several metabolic signatures and pathways markedly regulated at different stages of fibrosis. The previous GSMM study of human NAFLD[44] also identified altered metabolic pathways in NAFLD, however, that study was confined to a relatively small patients group, not representing the full spectrum of NAFLD. Here, we used genome-wide transcriptomics (RNA-seq)[43] data that provided a dynamic range of genes, enzymes / reactions, metabolites and their interactions, to model stage-specific GEMs for human hepatocytes in NAFLD (**Figure 1**). In addition, the study cohort allowed us to stratify and model the metabolic differences in three major genetic variants (*PNPLA3, TM6SF2* and *HSD17B13*) associated with risk and severity of NAFLD.

Our results implicate changes in the levels of vitamins (A, E) in the NASH-associated fibrosis progression. In the hepatocytes, a significant increase in the RMs of retinoic acid derivatives were observed in NASH F2, F3 (*vs.* NAFL) and in 'clinically significant' non-cirrhotic fibrosis (*vs.* 'minimal

disease')(**Figure S2**, **Figures 4** and **Figure S7**). A small increase in the serum levels of Vitamin A (retinol) and retinyl palmitate was observed at NASH F01 stage, whilst retinyl palmitate was markedly decreased at NASH F4 stage (**Figure 6g-h**). At this stage, the activated hepatic stellate cells tend to lose the retinyl esters stored in the liver, ultimately leading to vitamin A deficiency[42,45]. Vitamin A is a regulator of glucose and lipid metabolism in the liver and adipose tissue and may attenuate in the development of NAFLD[42,45]. The PNPLA3 gene product is known to have retinyl ester hydrolase activity, and *PNPLA3-I148M* is associated with low serum retinol level with enhanced retinyl esters in the liver of NAFLD patients[40,41]. However, GSMM have not identified any differences in vitamin A metabolism related to *PNPLA3* variant carriage. Intriguingly, Vitamin E derivatives were subsequently decreased across the stages of fibrosis. High dose of Vitamin E supplementation has been shown to improve histological steatohepatitis over placebo in the PIVENS randomized controlled trial of pioglitazone or vitamin E in patients with NASH[46].

We observed a dynamic regulation of complex GSLs in the advanced fibrosis. At this stage, majority of cerebrosides (HexCers: glucosylceramides (GlcCers) and lactosylceramides (LacCers)), and globosides were decreased in the patient groups. Intriguingly, the serum concentrations of GSLs (HexCers, GlcCers) were abruptly decreased in the patients with advance fibrosis or cirrhosis (F4) (**Figure 6e-f**). Ceramides (Cers) are the key intermediates of sphingolipid metabolism that promote cellular proliferation, differentiation and cell death[47,48]. Cers interact with several pathways involved in insulin resistance, oxidative stress, inflammation, and apoptosis, that are linked to NAFLD[48,49]. Understanding the role of ceramides in the staging of NASH-associated fibrosis is of great diagnostic interest[49]. In normal physiological conditions, Cers converts to GSLs, which prevents its excessive accumulation in the cells[50]. Our data suggest that in patients with advanced fibrosis, the serum concentrations of Cers increase up to NASH F3, and decrease at NASH F4 stages. Differential expression of metabolic genes and flux analysis showed an increase of Cer production *via* SMase, whilst production of GSLs (GlcCer, GalCer and LacCers) decreases (**Figure 5**). Thus, the conversion of Cers to GSLs (*via* glucosylceramide synthase, GCS) might have been compromised in 'advanced' fibrosis. Intriguingly, we found that the GSLs (LacCers and globosides) were associated with GAGs (heparin, keratin sulphate), which was also altered in the progression of NASH-associated fibrosis. It is well demonstrated that inhibiting GSL synthesis in obese mice has improved glucose homeostasis and markedly reduced the development of NAFL[51].

Taken together, the current study has identified several novel metabolic signatures and pathways in staging NAFLD. It enabled us to understand, how different metabolites and their intermediates are regulated across various stages of NASH-associated fibrosis. While some of our key results corroborate with the earlier findings, others allowed us to elucidate the dys(regulation) of metabolic pathways in NAFLD. Moreover, GSMM has demonstrated several stage-specific metabolic changes, which might help discover biomarkers, identify drug targets, and ultimately lead to effective therapeutic strategies for NASH. To this end, our data indicate the significance of GSL pathways and targeting these pathways might ameliorate the liver pathology associated with NAFL and NASH-associated fibrosis. However, some of these findings remains to be validated by *in vivo* and/or *in vitro* experiments.

METHODS

Cohort description

A full description of the transcriptomic analysis of the whole liver RNA-Seq data used in this study has previously been reported[20]. Patient samples were derived from the European NAFLD Registry (NCT04442334)[19] and comprised snap-frozen biopsy samples and associated clinical data from 206 patients diagnosed with histologically characterized NAFLD in France, Germany, Italy, or the United Kingdom. All biopsies were centrally scored by two expert liver pathologists according to the semiquantitative NASH-Clinical Research Network 'NAFLD Activity Score' (NAS)[52]. Fibrosis was staged from F0 to F4 (cirrhosis). Patients with alternate diagnoses and etiologies were excluded, including excessive alcohol intake (30 g per day for males and 20 g per day for females), viral hepatitis, autoimmune liver diseases, and steatogenic medication use. As previously described[19], collection and use of samples and clinical data for this study were approved by the relevant local and/or national Ethical Review Committee covering each participating center, with all patients providing informed consent for participation. All participant recruitment and informed consent processes at recruitment centers were conducted in compliance with nationally accepted practice in the respective territory and in accordance with the World Medical Association Declaration of Helsinki 2018.

All samples obtained in this study were histologically scored by two expert liver pathologists according to the widely accepted, semiquantitative NASH-Clinical Research Network 'NASH Activity Score' (NAS) system[52] and grouped according to histopathological grade of NASH and stage of fibrosis from F0 up to F4 (cirrhosis)[20].

Transcriptomics

Transcriptomic (RNA-seq) data associated with this study was obtained from[20]; also available in the NCBI GEO repository (accession GSE135251).

Genotyping

Genotypes were obtained from the GWAS data described in detail in[8]. Some data was also obtained from RNA-seq as reported elsewhere[43].

Genome-scale metabolic modelling

Contextualization of GEMs

Functional GEMs of human hepatocytes were developed by step-wise combining Gene Inactivity Moderated by Metabolism and Expression *(GIMME)*[25] and *E-Flux*[26] algorithms, applied to *iHepatocytes2322* as a template model. By integrating NAFLD stage-specific whole tissue transcriptomics (RNASeq) data with *iHepatocytes2322,* we contextualized and developed hepatocyte-GEMs for different stages of NASH-associated fibrosis, *i.e.*, 'minimal' NAFL + NASH (F0-1) (n= 85); 'mild' NAFL + NASH (F0-2) (n= 138); 'clinically significant' non-cirrhotic fibrosis (NASH F2-3) (n= 107); and 'advanced' fibrosis (NASH F3-4) (n= 68). The stage-specific transcriptomics data was used to score the active reactions encoded by the metabolic genes (MGs) of *iHepatocytes2322*. Accordingly, the feasibility of a particular reaction(s) to be included or discarded in the hepatocyte-GEM was determined. Subsequently, a draft GEM for each stage, that

include the active reactions, and their associated genes and metabolites was generated. The draft models were subjected to gap filling by assigning metabolic tasks performed using '*FitTask*' function deployed in the RAVEN suite[53]. All the models were able to carry out 256 metabolic tasks (*e.g.*, gluconeogenesis, fatty acid biosynthesis, oxidation *etc*.)[17,54], which was tested using '*CheckTask*' function. The blocked reactions were rectified or removed.

A similar approach was adapted to contextualize GEMs (*iHepatocytes2322*) for hepatocytes in the individuals exhibiting carriage of gene variants, *PNPLA3* (GC, GG) (n=69)*, TM6SF2* (CT, TT)(n=13)*, HSD17B13* (-T, TT)(n=21) and wild type (WTs)(n=36).

Normalized flux differences

We deployed an unbiased and non-uniform random sampling (RS) method[30] that finds solutions among the feasible flux distributions of the metabolic network. By applying RS on the context / stage-specific GEMs, we estimated the flux distribution (sampled, n=1000). Only reactions having non-zero measurable fluxes (<1e-6) fluxes were considered in the pair-wise (case – control) comparision. The absolute values of the reaction fluxes were grouped by their corresponding subsystems. The fluxes per subsystem were averaged and subjected to quantile normalization. Thereby, the flux differences between the case *vs.* control was estimated as stated in[30,55]. RS was performed using RAVEN 2.0 suite using '*MOSEK 8*' solver[53].

Flux Variability Analysis

A loop-less version of fastFVA[56] implemented in COBRA Toolbox v3.0[57] was used to prevent the formation of futile cycles. By using FVA, we estimated the flux range of reactions included in each context / stage-specific GEM for NAFLD, and thereby, the flux span (max – min) was estimated. Only reactions having non-zero flux span were considered. All the flux spans per subsystem was averaged. Percentage of relative average flux span (% RFS) per metabolic subsystem for each stage of NAFLD was estimated.

Sensitivity Analysis

Shadow price (SP) of a metabolite determines its impact on the cellular biomass, when the intracellular concentration of the metabolite deviates by one-unit from the steady-state[33]. If a metabolite has a positive SP, an increase in the concentration of the metabolite will increase the cellular biomass, suggesting the metabolite is limiting for the biomass. A negative SP suggests metabolite is available in excess, and zero SP suggests that, the concentration of metabolite remains insensitive towards the cellular biomass. Of note, the sign of the SP depends on the solver used. SP was evaluated using 'optimizeCbModel' function implemented in the COBRA Toolbox v3.0[57].

Mixed integer linear programming (MILP) was performed using '*MOSEK 8*' solver (licensed for the academic user) integrated in the RAVEN 2.0 suite[53]. Linear programming (LP) and optimization was performed using '*ILOG-IBM CPLEX (version 128)*' solver. Quality control and sanity checks[58] were performed using COnstraint-Based Reconstruction and Analysis Toolbox (Cobra toolbox v3.0)[57]. Simulations were performed using Cobra toolbox (v3.0) and RAVEN suite. All the operations were performed in MATLAB 2017b (Mathworks, Inc., Natick, MA, USA).

Reporter metabolite analysis

RM analysis was performed by using '*reporterMetabolites*' function of the RAVEN 2.0 suite[53]. RMs that were significantly ($p < 0.05$) altered between two different NAFLD patient groups were listed.

Data visualization and graphics

Several libraries / packages of R v3.6.0, a statistical programming language such as *'Heatmap.2'*, *'boxplot'*, '*beanplot*', '*gplot*', and '*ggplot2*' and Cytoscape v3.8.0 software were used for data visualization and network analysis.

Analysis of lipids and polar metabolites

Lipidomics analysis

Serum samples were randomized and extracted using a modified version of the previously-published Folch procedure[59]. The maternal samples were analysed as one batch and the cord blood samples as a second batch. In short, 10 µL of 0.9% NaCl and, 120 µL of CHCl3: MeOH (2:1, v/v) containing the internal standards (c = 2.5 µg/mL) was added to 10 µL of each serum sample. The standard solution contained the following compounds: 1,2-diheptadecanoyl-sn-glycero-3-phosphoethanolamine (PE(17:0/17:0)), N-heptadecanoyl-D-erythro-sphingosylphosphorylcholine (SM(d18:1/17:0)), N-heptadecanoyl-D-erythro-sphingosine (Cer(d18:1/17:0)), 1,2-diheptadecanoyl-sn-glycero-3-phosphocholine (PC(17:0/17:0)), 1-heptadecanoyl-2-hydroxy-sn-glycero-3-phosphocholine (LPC(17:0)) and 1-palmitoyl-d31-2-oleoyl-sn-glycero-3-phosphocholine (PC(16:0/d31/18:1)), were purchased from Avanti Polar Lipids, Inc. (Alabaster, AL, USA), and, triheptadecanoylglycerol (TG(17:0/17:0/17:0)) was purchased from Larodan AB (Solna, Sweden). The samples were vortex mixed and incubated on ice for 30 min after which they were centrifuged (9400 × g, 3 min). 60 µL from the lower layer of each sample was then transferred to a glass vial with an insert and 60 µL of CHCl3: MeOH (2:1, v/v) was added to each sample. The samples were stored at -80 °C until analysis.

Calibration curves using 1-hexadecyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine (PC(16:0e/18:1(9Z))), 1-(1Z-octadecenyl)-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine (PC(18:0p/18:1(9Z))), 1-stearoyl-2-hydroxy-sn-glycero-3-phosphocholine (LPC(18:0)), 1-oleoyl-2-hydroxy-sn-glycero-3-phosphocholine (LPC(18:1)), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (PE(16:0/18:1)), 1-(1Z-octadecenyl)-2-docosahexaenoyl-sn-glycero-3-phosphocholine (PC(18:0p/22:6)) and 1-stearoyl-2-linoleoyl-sn-glycerol (DG(18:0/18:2)), 1-(9Z-octadecenoyl)-sn-glycero-3-phosphoethanolamine (LPE(18:1)), N-(9Z-octadecenoyl)-sphinganine (Cer(d18:0/18:1(9Z))), 1-hexadecyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoethanolamine (PE(16:0/18:1)) from Avanti Polar Lipids, 1-Palmitoyl-2-Hydroxy-sn-Glycero-3-Phosphatidylcholine (LPC(16:0)), 1,2,3 trihexadecanoalglycerol (TG(16:0/16:0/16:0)), 1,2,3-trioctadecanoylglycerol (TG(18:0/18:0/18:)) and 3β-hydroxy-5-cholestene-3-stearate (ChoE(18:0)), 3β-Hydroxy-5-cholestene-3-linoleate (ChoE(18:2)) from Larodan, were prepared to the following concentration levels: 100, 500, 1000, 1500, 2000 and 2500 ng/mL (in CHCl3:MeOH, 2:1, v/v) including 1250 ng/mL of each internal standard.

The samples were analyzed by ultra-high-performance liquid chromatography quadrupole time-of-

flight mass spectrometry (UHPLC-QTOFMS). Briefly, the UHPLC system used in this work was a 1290 Infinity II system from Agilent Technologies (Santa Clara, CA, USA). The system was equipped with a multi sampler (maintained at 10 °C), a quaternary solvent manager and a column thermostat (maintained at 50 °C). Injection volume was 1 µL and the separations were performed on an ACQUITY UPLC® BEH C18 column (2.1 mm × 100 mm, particle size 1.7 µm) by Waters (Milford, MA, USA). The mass spectrometer coupled to the UHPLC was a 6545 QTOF from Agilent Technologies interfaced with a dual jet stream electrospray (Ddual ESI) ion source. All analyses were performed in positive ion mode and MassHunter B.06.01 (Agilent Technologies) was used for all data acquisition. Quality control was performed throughout the dataset by including blanks, pure standard samples, extracted standard samples and control serum samples, including in-house serum and a pooled QC with an aliquot of each sample was collected and pooled and used as quality control sample.

Relative standard deviations (% RSDs) for identified lipids in the control serum samples (n = 13) and in the pooled serum samples (n = 54) were on average 22.4% and 17.5%, respectively.

Mass spectrometry data processing was performed using the open source software package MZmine 2.18[60]. The following steps were applied in this processing: (i) Crop filtering with a m/z range of 350 – 1200 m/z and an RT range of 2.0 to 12 minutes, (ii) Mass detection with a noise level of 750, (iii) Chromatogram builder with a minimum time span of 0.08 min, minimum height of 1000 and a m/z tolerance of 0.006 m/z or 10.0 ppm, (iv) Chromatogram deconvolution using the local minimum search algorithm with a 70% chromatographic threshold, 0.05 min minimum RT range, 5% minimum relative height, 1200 minimum absolute height, a minimum ration of peak top/edge of 1.2 and a peak duration range of 0.08 - 5.0, (v), Isotopic peak grouper with a m/z tolerance of 5.0 ppm, RT tolerance of 0.05 min, maximum charge of 2 and with the most intense isotope set as the representative isotope, (vi) Peak filter with minimum 12 data points, a FWHM between 0.0 and 0.2, tailing factor between 0.45 and 2.22 and asymmetry factor between 0.40 and 2.50, (vii) Join aligner with a m/z tolerance of 0.009 or 10.0 ppm and a weight for of 2, a RT tolerance of 0.1 min and a weight of 1 and with no requirement of charge state or ID and no comparison of isotope pattern, (viii) Peak list row filter with a minimum of 10% of the samples (ix) Gap filling using the same RT and m/z range gap filler algorithm with an m/z tolerance of 0.009 m/z or 11.0 ppm, (x) Identification of lipids using a custom database search with an m/z tolerance of 0.009 m/z or 10.0 ppm and a RT tolerance of 0.1 min, and (xi) Normalization using internal standards PE(17:0/17:0), SM(d18:1/17:0), Cer(d18:1/17:0), LPC(17:0), TG(17:0/17:0/17:0) and PC(16:0/d30/18:1)) for identified lipids and closest ISTD for the unknown lipids followed by calculation of the concentrations based on lipid-class concentration curves.

Analysis of polar metabolites

Serum samples were randomized and sample preparation was carried out as described previously[61]. The maternal samples were analysed as one batch and the cord blood samples as a second batch. In summary, 400 µL of MeOH containing ISTDs (heptadecanoic acid, deuterium-labeled DL-valine, deuterium-labeled succinic acid, and deuterium-labeled glutamic acid, c = 1 µg/mL) was added to 30 µl of the serum samples which were vortex mixed and incubated on ice for 30 min after which they were centrifuged (9400 × g, 3 min) and 350 µL of the supernatant was

collected after centrifugation. The solvent was evaporated to dryness and 25 µL of MOX reagent was added and the sample was incubated for 60 min at 45 °C. 25 µL of MSTFA was added and after 60 min incubation at 45 °C 25 µL of the retention index standard mixture (n-alkanes, c=10 µg/mL) was added.

The analyses were carried out on an Agilent 7890B GC coupled to 7200 QTOF MS. Injection volume was 1 µL with 100:1 cold solvent split on PTV at 70 °C, heating to 300 °C at 120 °C/minute. Column: Zebron ZB-SemiVolatiles. Length: 20m, I.D. 0.18mm, film thickness: 0.18 µm. With initial Helium flow 1.2 mL/min, increasing to 2.4 mL/min after 16 mins. Oven temperature program: 50 °C (5 min), then to 270°C at 20 °C/min and then to 300 °C at 40 °C/min (5 min). EI source: 250 °C, 70 eV electron energy, 35µA emission, solvent delay 3 min. Mass range 55 to 650 amu, acquisition rate 5 spectra/s, acquisition time 200 ms/spectrum. Quad at 150 °C, 1.5 mL/min N2 collision flow, aux-2 temperature: 280 °C.

Calibration curves were constructed using alanine, citric acid, fumaric acid, glutamic acid, glycine, lactic acid, malic acid, 2-hydroxybutyric acid, 3-hydroxybutyric acid, linoleic acid, oleic acid, palmitic acid, stearic acid, cholesterol, fructose, glutamine, indole-3-propionic acid, isoleucine, leucine, proline, succinic acid, valine, asparagine, aspartic acid, arachidonic acid, glycerol-3-phosphate, lysine, methionine, ornithine, phenylalanine, serine and threonine purchased from Sigma-Aldrich (St. Louis, MO, USA) at concentration range of 0.1 to 80 µg/mL. An aliquot of each sample was collected and pooled and used as quality control samples, together with a NIST SRM 1950 serum sample and an in-house pooled serum sample. Relative standard deviations (% RSDs) of the metabolite concentrations in control serum samples showed % RSDs within accepted analytical limits at averages of 27.2% and 29.2% for in-house QC abd pooled QC samples.

SUPPLEMENTAL INFORMATION

Supplemental Information includes fourteen figures (**Figures S1-10**) and one table (**Table S1**).

DATA AND SOFTWARE AVAILABILITY

The GEMs for human hepatocytes contextualized for various stages of NAFLD were deposited in BioModels[62] (www.ebi.ac.uk/biomodels/), and assigned an identifier MODEL2102060002. The RNA-Seq data are available in the NCBI GEO repository (accession GSE135251). The lipidomic / metabolomic datasets generated in this study are available at the NIH Common Fund's National Metabolomics Data Repository (NMDR) website, the Metabolomics Workbench (https://www.metabolomicsworkbench.org).

ACKNOWLEDGMENTS

We would like to thank the EPoS (Elucidating Pathways of Steatohepatitis) (http://epos-nafld.eu) investigators.

FUNDING SOURCE

This study has been supported by the EPoS (Elucidating Pathways of Steatohepatitis) consortium funded by the Horizon 2020 Framework Program of the European Union under Grant Agreement 634413, the LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) consortium

funded by the Innovative Medicines Initiative (IMI2) Program of the European Union under Grant Agreement 777377. Q.M.A. is a Newcastle NIHR Biomedical Research Centre investigator. T.H. and M.O. received funding from the Novo Nordisk Foundation (Grant no. NNF20OC0063971). T.V.P. laboratory was funded by the MRC MDU programme grant.

AUTHOR CONTRIBUTIONS

M.O. proposed and design the study. Q.M.A., A.K.D and M.O. assisted with the formulation of the study design. M.O., Q.M.A., A.K.D., and T.H. supervised the study. P.S. performed GSMM and statistical analysis. T.S. and D.G. performed metabolomics analysis, supervised by T.H. A.M. performed metabolomics data analysis. O.G. and S.C. performed the transcriptomic analysis, supervised by A.K.D. V.R. E.B. J.M.S., A.V.-P., M.A., Q.M.A. contributed to clinical data acquisition in the EPoS cohort, coordinated by Q.M.A.. P.S. and M.O. wrote the manuscript. All authors critically revised the manuscript for intellectual content and approved the final manuscript.

M.O. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

DECLARATION OF INTERESTS

J.M.S declares Consultancy: Boehringer Ingelheim, BMS, Genfit, Gilead Sciences, Intercept Pharmaceuticals, Madrigal, Novartis, Novo Nordisk, Nordic Bioscience, Pfizer, Roche, Sanofi, Siemens Healthcare GmbH. Research Funding: Gilead Sciences, Boehringer Ingelheim. Speakers Bureau: Falk Foundation MSD Sharp & Dohme GmbH. M.A. reports consultancy/advisory with MedImmune/Astra Zeneca and E3Bio, as well as honoraria from Intercept and grant support from GSK and Takeda. E.B. reports advisory/consulting for BMS, Genfit SA, Gilead, Intercept, and Novartis. Q.M.A. reports grants, personal fees and other from Allergan/Tobira, other from E3Bio, other from Eli Lilly & Company Ltd, other from Galmed, grants, personal fees and other from Genfit SA, personal fees and other from Gilead, other from Grunthal, other from Imperial Innovations, grants and other from Intercept Pharma Europe Ltd, other from Inventiva, other from Janssen, personal fees from Kenes, other from MedImmune, other from NewGene, grants, personal fees and other from Pfizer Ltd, other from Raptor Pharma, grants and other from Novartis Pharma AG, grants from Abbvie, personal fees and other from BMS, grants from GSK, other from NGMBio, other from Madrigal, other from Servier, other from EcoR1, other from 89Bio, other from Altimmune, grants and other from AstraZeneca, other from Axcella, other from Blade, other from BNN Cardio, other from Celgene, other from Cirius, other from CymaBay, other from Genentech, other from HistoIndex, other from Indalo, other from IQVIA, other from Metacrine, other from North Sea Therapeutics, personal fees and other from Novo Nordisk, other from Poxel, other from Terns, other from Viking Therapeutics, grants from Glympse Bio, other from PathAI, outside the submitted work.

All other authors declare that they have no competing interests.

REFERENCES

1 Anstee, Q. M., Reeves, H. L., Kotsiliti, E., Govaere, O. & Heikenwalder, M. From NASH to HCC: current concepts and future challenges. *Nat Rev Gastroenterol Hepatol* **16**, 411-428, doi:10.1038/s41575-019-0145-7 (2019).

2 Anstee, Q. M., Targher, G. & Day, C. P. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. *Nat Rev Gastroenterol Hepatol* **10**, 330-344, doi:10.1038/nrgastro.2013.41 (2013).

3 Labenz, C. *et al.* Predictors of advanced fibrosis in non-cirrhotic non-alcoholic fatty liver disease in Germany. *Aliment Pharmacol Ther* **48**, 1109-1116, doi:10.1111/apt.14976 (2018).

4 Younossi, Z. *et al.* Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* **15**, 11-20, doi:10.1038/nrgastro.2017.109 (2018).

5 Estes, C. *et al.* Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016-2030. *J Hepatol* **69**, 896-904, doi:10.1016/j.jhep.2018.05.036 (2018).

6 McPherson, S. *et al.* Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: Implications for prognosis and clinical management. *J Hepatol* **62**, 1148-1155, doi:10.1016/j.jhep.2014.11.034 (2015).

7 Taylor, R. S. *et al.* Association Between Fibrosis Stage and Outcomes of Patients With Nonalcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis. *Gastroenterology* **158**, 1611-1625 e1612, doi:10.1053/j.gastro.2020.01.043 (2020).

8 Anstee, Q. M. *et al.* Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort(). *J Hepatol* **73**, 505-515, doi:10.1016/j.jhep.2020.04.003 (2020).

9 Masoodi, M. *et al.* Metabolomics and Lipidomics in NASH: From identifying biomarkers to the development of non-invasive diagnostic tests. *Nat Rev Gastroenterol Hepatol* (2021, in press).

10 Oresic, M. *et al.* Prediction of non-alcoholic fatty-liver disease and liver fat content by serum molecular lipids. *Diabetologia* **56**, 2266-2274, doi:10.1007/s00125-013-2981-2 (2013).

11 Westerbacka, J. *et al.* Splanchnic Balance of Free Fatty Acids, Endocannabinoids, and Lipids in Subjects With Nonalcoholic Fatty Liver Disease. *Gastroenterology* **139**, 1961-U1232, doi:10.1053/j.gastro.2010.06.064 (2010).

12 Kotronen, A. *et al.* Hepatic Stearoyl-CoA Desaturase (SCD)-1 Activity and Diacylglycerol but Not Ceramide Concentrations Are Increased in the Nonalcoholic Human Fatty Liver. *Diabetes* **58**, 203-208, doi:10.2337/db08-1074 (2009).

13 Bordbar, A. *et al.* Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation. *Mol Syst Biol* **8**, 558, doi:10.1038/msb.2012.21 (2012).

14 Sen, P. *et al.* Metabolic Alterations in Human Peripheral Blood Mononuclear Cells Associate with Progression to Islet Autoimmunity and Type 1 Diabetes. *bioRxiv*, 658500, doi:10.1101/658500 (2019).

15 Sen, P., Kemppainen, E. & Orešič, M. Perspectives on Systems Modelling of Human Peripheral Blood Mononuclear Cells. *Frontiers in molecular biosciences* **4**, 96 (2017).

16 Jerby, L., Shlomi, T. & Ruppin, E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol Syst Biol* **6**, 401, doi:10.1038/msb.2010.56 (2010).

17 Mardinoglu, A. *et al.* Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat Commun* **5**, 3083, doi:10.1038/ncomms4083 (2014).

18 Hyotylainen, T. *et al.* Genome-scale study reveals reduced metabolic adaptability in patients with non-alcoholic fatty liver disease. *Nature Communications* **7**, doi:10.1038/ncomms9994 (2016).

19 Hardy, T. *et al.* The European NAFLD Registry: A real-world longitudinal cohort study of nonalcoholic fatty liver disease. *Contemp Clin Trials* **98**, 106175, doi:10.1016/j.cct.2020.106175 (2020).

20 Govaere, O. *et al.* Transcriptomic Profiling Across the Non-Alcoholic Fatty Liver Disease Spectrum Reveals Gene Signatures for Steatohepatitis and Fibrosis. *Science Translational Medicine* **12**, eaba4448 (2020).

21 Liu, Y. L. *et al.* TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. *Nat Commun* **5**, 4309, doi:10.1038/ncomms5309 (2014).

22 Abul-Husn, N. S. *et al.* A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. *N Engl J Med* **378**, 1096-1106, doi:10.1056/NEJMoa1712191 (2018).

23 Romeo, S. *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* **40**, 1461-1465, doi:10.1038/ng.257 (2008).

24 Agren, R. *et al.* Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol* **8**, e1002518, doi:10.1371/journal.pcbi.1002518 (2012).

25 Becker, S. A. & Palsson, B. O. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol* **4**, e1000082, doi:10.1371/journal.pcbi.1000082 (2008).

26 Colijn, C. *et al.* Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. *PLoS computational biology* **5**, e1000489 (2009).

27 Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC bioinformatics* **11**, 489 (2010).

28 Jensen, P. A. & Papin, J. A. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* **27**, 541-547, doi:10.1093/bioinformatics/btq702 (2011).

29 Nielsen, J. It is all about metabolic fluxes. *J Bacteriol* **185**, 7031-7035, doi:10.1128/jb.185.24.7031-7035.2003 (2003).

30 Bordel, S., Agren, R. & Nielsen, J. Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput Biol* **6**, e1000859, doi:10.1371/journal.pcbi.1000859 (2010).

31 Cakir, T. *et al.* Integration of metabolome data with metabolic networks reveals reporter reactions. *Mol Syst Biol* **2**, 50, doi:10.1038/msb4100085 (2006).

32 Patil, K. R. & Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2685-2689, doi:10.1073/pnas.0406811102 (2005).

33 Palsson, B. Ø. *Systems biology: properties of reconstructed networks*. (Cambridge university press, 2006).

34 Zhang, T., de Waard, A. A., Wuhrer, M. & Spaapen, R. M. The Role of Glycosphingolipids in Immune Cell Functions. *Front Immunol* **10**, 90, doi:10.3389/fimmu.2019.00090 (2019).

35 Apostolidis, S. A. *et al.* Phosphatase PP2A is requisite for the function of regulatory T cells. *Nature immunology* **17**, 556 (2016).

36 Trepo, E. & Valenti, L. Update on NAFLD genetics: From new variants to the clinic. *J Hepatol* **72**, 1196-1209, doi:10.1016/j.jhep.2020.02.020 (2020).

37 Luukkonen, P. K. *et al.* Impaired hepatic lipid synthesis from polyunsaturated fatty acids in TM6SF2 E167K variant carriers with NAFLD. *J Hepatol* **67**, 128-136, doi:10.1016/j.jhep.2017.02.014 (2017).

38 Ma, Y. *et al.* 17-Beta Hydroxysteroid Dehydrogenase 13 Is a Hepatic Retinol Dehydrogenase Associated With Histological Features of Nonalcoholic Fatty Liver Disease. *Hepatology* **69**, 1504-1519, doi:10.1002/hep.30350 (2019).

39 He, S. *et al.* A sequence variation (I148M) in PNPLA3 associated with nonalcoholic fatty liver disease disrupts triglyceride hydrolysis. *J Biol Chem* **285**, 6706-6715, doi:10.1074/jbc.M109.064501 (2010).

40 Kovarova, M. *et al.* The Genetic Variant I148M in PNPLA3 Is Associated With Increased Hepatic Retinyl-Palmitate Storage in Humans. *J Clin Endocrinol Metab* **100**, E1568-1574, doi:10.1210/jc.2015-2978 (2015).

41 Mondul, A. *et al.* PNPLA3 I148M Variant Influences Circulating Retinol in Adults with Nonalcoholic Fatty Liver Disease or Obesity. *J Nutr* **145**, 1687-1691, doi:10.3945/jn.115.210633 (2015).

42 Pettinelli, P. *et al.* Altered hepatic genes related to retinol metabolism and plasma retinol in patients with non-alcoholic fatty liver disease. *PLoS One* **13**, e0205747, doi:10.1371/journal.pone.0205747 (2018).

43 Govaere, O. *et al.* Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci Transl Med* **12**, doi:10.1126/scitranslmed.aba4448 (2020).

44 Mardinoglu, A. *et al.* Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature communications* **5** (2014).

45 Saeed, A., Dullaart, R. P. F., Schreuder, T., Blokzijl, H. & Faber, K. N. Disturbed Vitamin A Metabolism in Non-Alcoholic Fatty Liver Disease (NAFLD). *Nutrients* **10**, doi:10.3390/nu10010029 (2017).

46 Sanyal, A. J. *et al.* Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. *N Engl J Med* **362**, 1675-1685, doi:10.1056/NEJMoa0907929 (2010).

47 Turpin-Nolan, S. M. & Bruning, J. C. The role of ceramides in metabolic disorders: when size and localization matters. *Nat Rev Endocrinol* **16**, 224-233, doi:10.1038/s41574-020-0320-5 (2020).

48 Gault, C. R., Obeid, L. M. & Hannun, Y. A. An overview of sphingolipid metabolism: from synthesis to breakdown. *Adv Exp Med Biol* **688**, 1-23, doi:10.1007/978-1-4419-6741-1_1 (2010).

49 Pagadala, M., Kasumov, T., McCullough, A. J., Zein, N. N. & Kirwan, J. P. Role of ceramides in nonalcoholic fatty liver disease. *Trends Endocrinol Metab* **23**, 365-371, doi:10.1016/j.tem.2012.04.005 (2012).

50 Ishay, Y., Nachman, D., Khoury, T. & Ilan, Y. The role of the sphingolipid pathway in liver fibrosis: an emerging new potential target for novel therapies. *Am J Physiol Cell Physiol* **318**, C1055-C1064, doi:10.1152/ajpcell.00003.2020 (2020).

51 Zhao, H. *et al.* Inhibiting glycosphingolipid synthesis ameliorates hepatic steatosis in obese mice. *Hepatology* **50**, 85-93, doi:10.1002/hep.22970 (2009).

52 Kleiner, D. E. *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313-1321, doi:10.1002/hep.20701 [doi] (2005).

53 Wang, H. *et al.* RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. *PLoS Comput Biol* **14**, e1006541, doi:10.1371/journal.pcbi.1006541 (2018).

54 Mardinoglu, A. *et al.* Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol Syst Biol* **9**, 649, doi:10.1038/msb.2013.5 (2013).

55 Osterlund, T., Nookaew, I., Bordel, S. & Nielsen, J. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst Biol* **7**, 36, doi:10.1186/1752-0509-7-36 (2013).

56 Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC Bioinformatics* **11**, 489, doi:10.1186/1471-2105-11-489 (2010).

57 Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3. 0. *arXiv preprint arXiv:1710.04038* (2017).

58 Thiele, I. & Palsson, B. O. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* **5**, 93-121, doi:10.1038/nprot.2009.203 (2010).

59 Nygren, H., Seppanen-Laakso, T., Castillo, S., Hyotylainen, T. & Oresic, M. *Liquid Chromatography-Mass Spectrometry (LC-MS)-Based Lipidomics for Studies of Body Fluids and Tissues*. Vol. 708 (2011).

60 Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *Bmc Bioinformatics* **11**, doi:10.1186/1471-2105-11-395 (2010).

61 Castillo, S., Mattila, I., Miettinen, J., Oresic, M. & Hyotylainen, T. Data Analysis Tool for Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry. *Analytical Chemistry* **83**, 3058-3067, doi:10.1021/ac103308x (2011).

62 Malik-Sheriff, R. S. *et al.* BioModels-15 years of sharing computational models in life science. *Nucleic Acids Res* **48**, D407-D415, doi:10.1093/nar/gkz1055 (2020).

Figure Legends

Figure 1. Study design and schematic illustration. Key steps in the reconstruction of stage-specific genome-scale metabolic models (GEMs) of human hepatocytes in NAFLD. Labels F(0-4) represents different stages of fibrosis.

Figure 2. Flux variability across the metabolic subsystems of human hepatocytes in NAFLD. a) Showing cumulative distribution / proportion of average flux span (max - min) of all the reactions within a metabolic subsystem at different stages of NAFLD. **b)** Stacked bar plot showing the % of relative average flux span of each metabolic subsystem at different stages (color coded) of NAFLD.

Figure 3. Metabolic flux states of human hepatocytes in NAFLD. Normalized average flux differences across the metabolic subsystems of the hepatocytes at different stages of NAFLD. Red and blue color denotes increase and decrease of average reaction fluxes within a metabolic subsystem at two different stages of NAFLD.

Figure 4. Reporter metabolites of NASH F3 vs. NAFL. A metabolic-centric view of reporter metabolite (RM) modules that were significantly altered ($p<0.05$) between NASH F3 *vs.* NAFL. Orange and blue color denotes up and downregulated respectively. Each node represents a 'RM' and single or double lines represent reversible or irreversible metabolic reactions respectively. RMs that belong to a particular chemical class are color coded.

Figure 5. Regulation of sphingolipid metabolism in NAFLD. a) A schematic illustration of sphingolipid (SL) metabolism in human hepatocytes. *SPT: serine palmitoyl-CoA transferase, CerS: ceramide synthase, and DES: dihydroceramide desaturase, SMase: sphingomyelinase pathway S1P: sphingosine-1-phosphate, CDase: ceramidases, SK: sphingosine kinase, S1PPase: sphingosine phosphate phosphatases GCDase: glucosidase, GalCDase: galactosidase, C1P: ceramide 1-phosphate.* The modeled reactions are labeled with 'R#'. **b)** Log_2 fold change (FC) of the differentially expressed (p.adjusted < 0.05) metabolic genes (MGs) of SL pathways at two different stages of NAFLD. **c)** Normalized flux differences across the SL reactions of the hepatocytes at different stages of NAFLD. Red and blue color denotes increase and decrease of reaction fluxes within at two different stages of NAFLD.

Figure 6. Serum metabolic levels of ceramides, glycosphingolipids and retinoids in the patients at different stages of NAFLD. a-i) Beanplots showing the log2 intensities of the metabolites across different stages of NAFLD. Significant differences (Tukey's HSD, $p < 0.05$) are shown by '*'. The dotted line denotes the mean of the population, and the black dashed lines in the bean plots represent the group mean.

Figure 7. Flux variability across the metabolic subsystems of human hepatocytes in the genetic modifiers of NAFLD. a) Showing cumulative distribution / proportion of average flux span of all the reactions within a metabolic subsystem in the hepatocytes of the carriers of three major genetic variants (*PNPLA3, TM6SF2* and *HSD17B13*) associated with risk and severity of NAFLD and the wild types (WT)(**Table S1**) models. **b)** Stacked bar plot showing the % of relative average flux span of each metabolic subsystem in the hepatocyte of the carriers of different genetic variants.

FIGURE 1

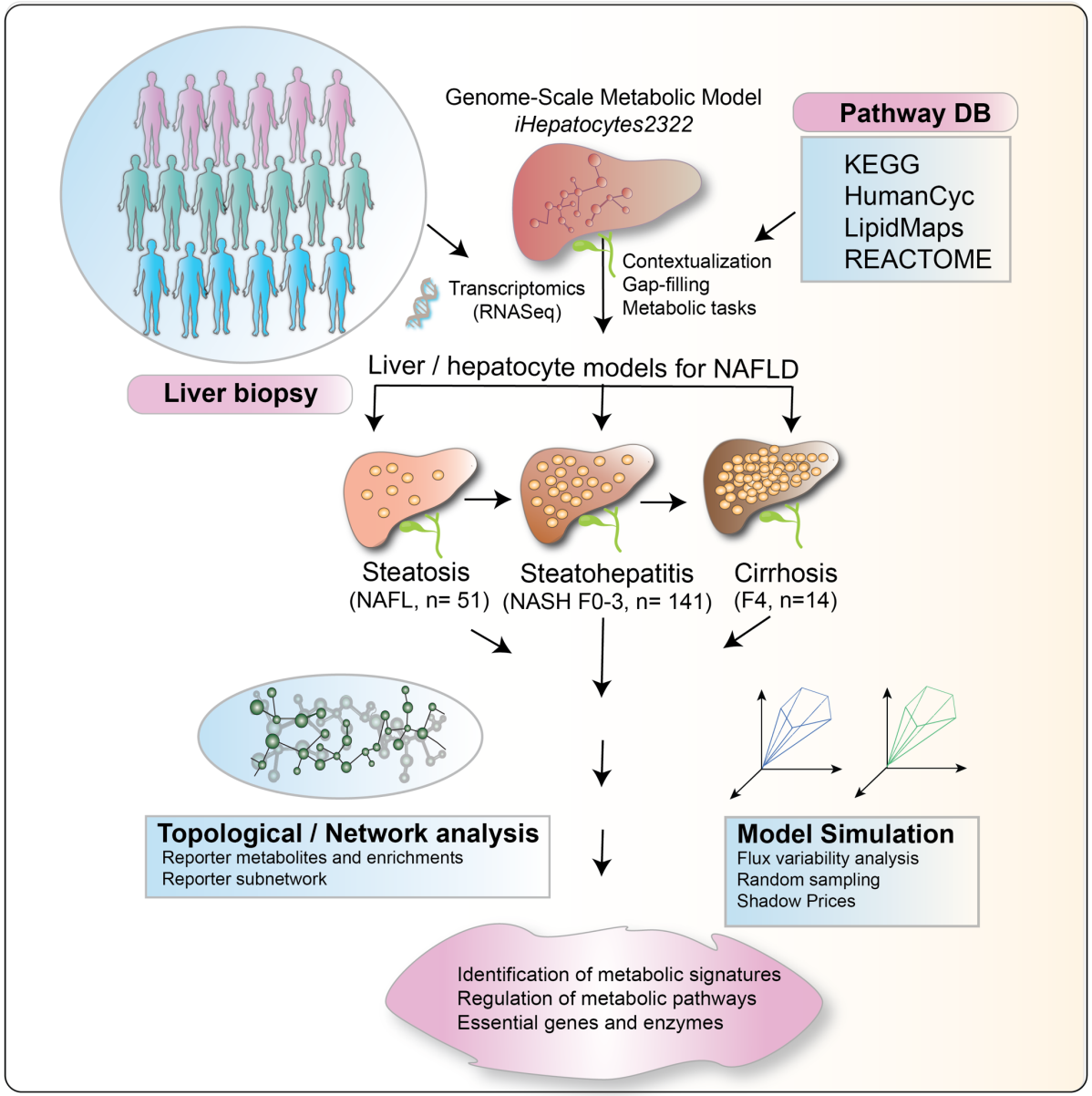

Genome-Scale Metabolic Model
iHepatocytes2322
Pathway DB
KEGG
HumanCyc
LipidMaps
REACTOME
Transcriptomics
(RNASeq)
Contextualization
Gap-filling
Metabolic tasks
Liver biopsy
Liver / hepatocyte models for NAFLD
Steatosis
(NAFL, n= 51)
Steatohepatitis
(NASH F0-3, n= 141)
Cirrhosis
(F4, n=14)
Topological / Network analysis
Reporter metabolites and enrichments
Reporter subnetwork
Model Simulation
Flux variability analysis
Random sampling
Shadow Prices
Identification of metabolic signatures
Regulation of metabolic pathways
Essential genes and enzymes

FIGURE 2

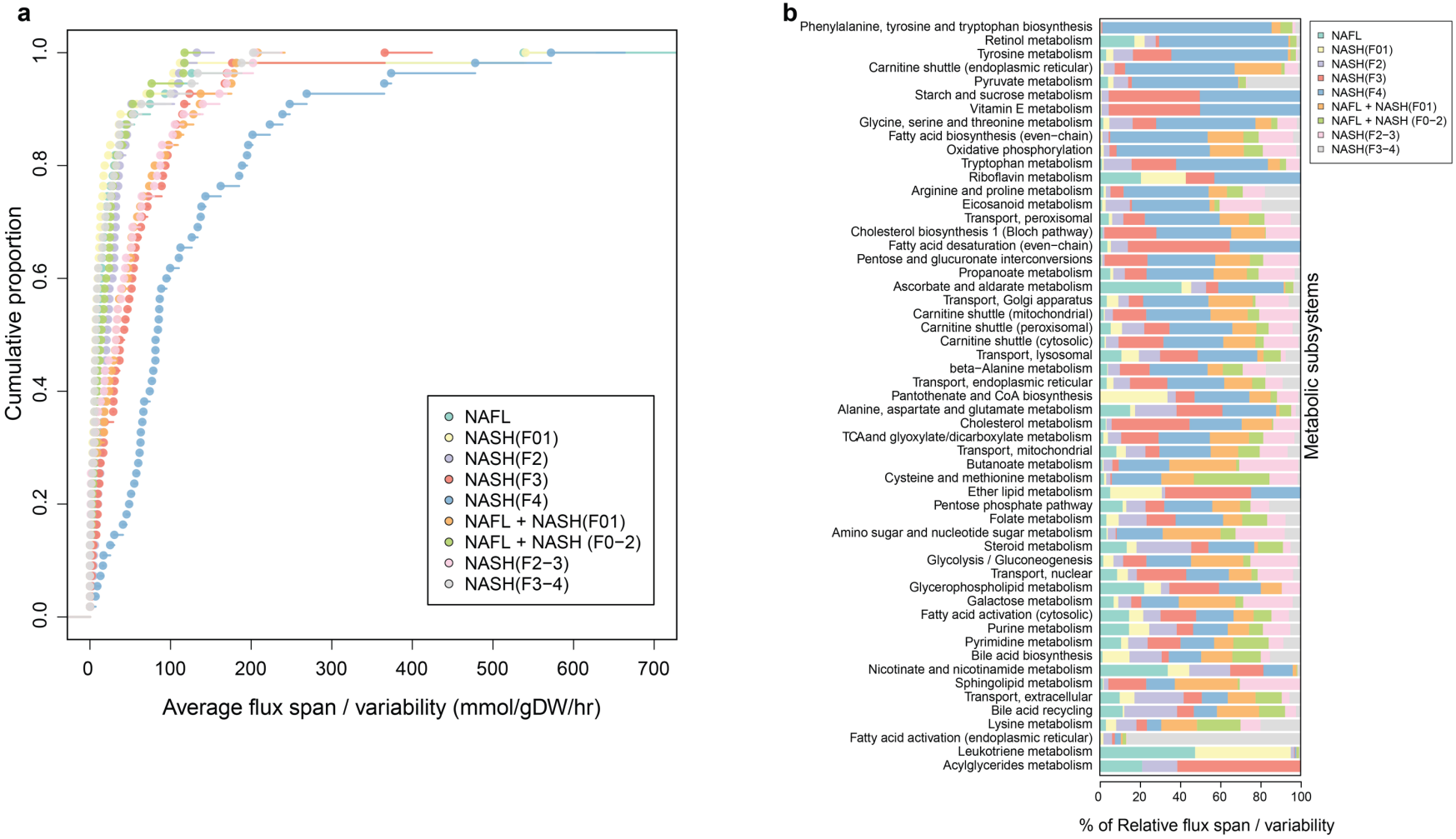

a
Cumulative proportion
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500
600
700
Average flux span / variability (mmol/gDW/hr)
NAFL
NASH(F01)
NASH(F2)
NASH(F3)
NASH(F4)
NAFL + NASH(F01)
NAFL + NASH (F0−2)
NASH(F2−3)
NASH(F3−4)
b
Phenylalanine, tyrosine and tryptophan biosynthesis
Retinol metabolism
Tyrosine metabolism
Carnitine shuttle (endoplasmic reticular)
Pyruvate metabolism
Starch and sucrose metabolism
Vitamin E metabolism
Glycine, serine and threonine metabolism
Fatty acid biosynthesis (even−chain)
Oxidative phosphorylation
Tryptophan metabolism
Riboflavin metabolism
Arginine and proline metabolism
Eicosanoid metabolism
Transport, peroxisomal
Cholesterol biosynthesis 1 (Bloch pathway)
Fatty acid desaturation (even−chain)
Pentose and glucuronate interconversions
Propanoate metabolism
Ascorbate and aldarate metabolism
Transport, Golgi apparatus
Carnitine shuttle (mitochondrial)
Carnitine shuttle (peroxisomal)
Carnitine shuttle (cytosolic)
Transport, lysosomal
beta−Alanine metabolism
Transport, endoplasmic reticular
Pantothenate and CoA biosynthesis
Alanine, aspartate and glutamate metabolism
Cholesterol metabolism
TCAand glyoxylate/dicarboxylate metabolism
Transport, mitochondrial
Butanoate metabolism
Cysteine and methionine metabolism
Ether lipid metabolism
Pentose phosphate pathway
Folate metabolism
Amino sugar and nucleotide sugar metabolism
Steroid metabolism
Glycolysis / Gluconeogenesis
Transport, nuclear
Glycerophospholipid metabolism
Galactose metabolism
Fatty acid activation (cytosolic)
Purine metabolism
Pyrimidine metabolism
Bile acid biosynthesis
Nicotinate and nicotinamide metabolism
Sphingolipid metabolism
Transport, extracellular
Bile acid recycling
Lysine metabolism
Fatty acid activation (endoplasmic reticular)
Leukotriene metabolism
Acylglycerides metabolism
Metabolic subsystems
0
20
40
60
80
100
% of Relative flux span / variability
NAFL
NASH(F01)
NASH(F2)
NASH(F3)
NASH(F4)
NAFL + NASH(F01)
NAFL + NASH (F0−2)
NASH(F2−3)
NASH(F3−4)

FIGURE 3

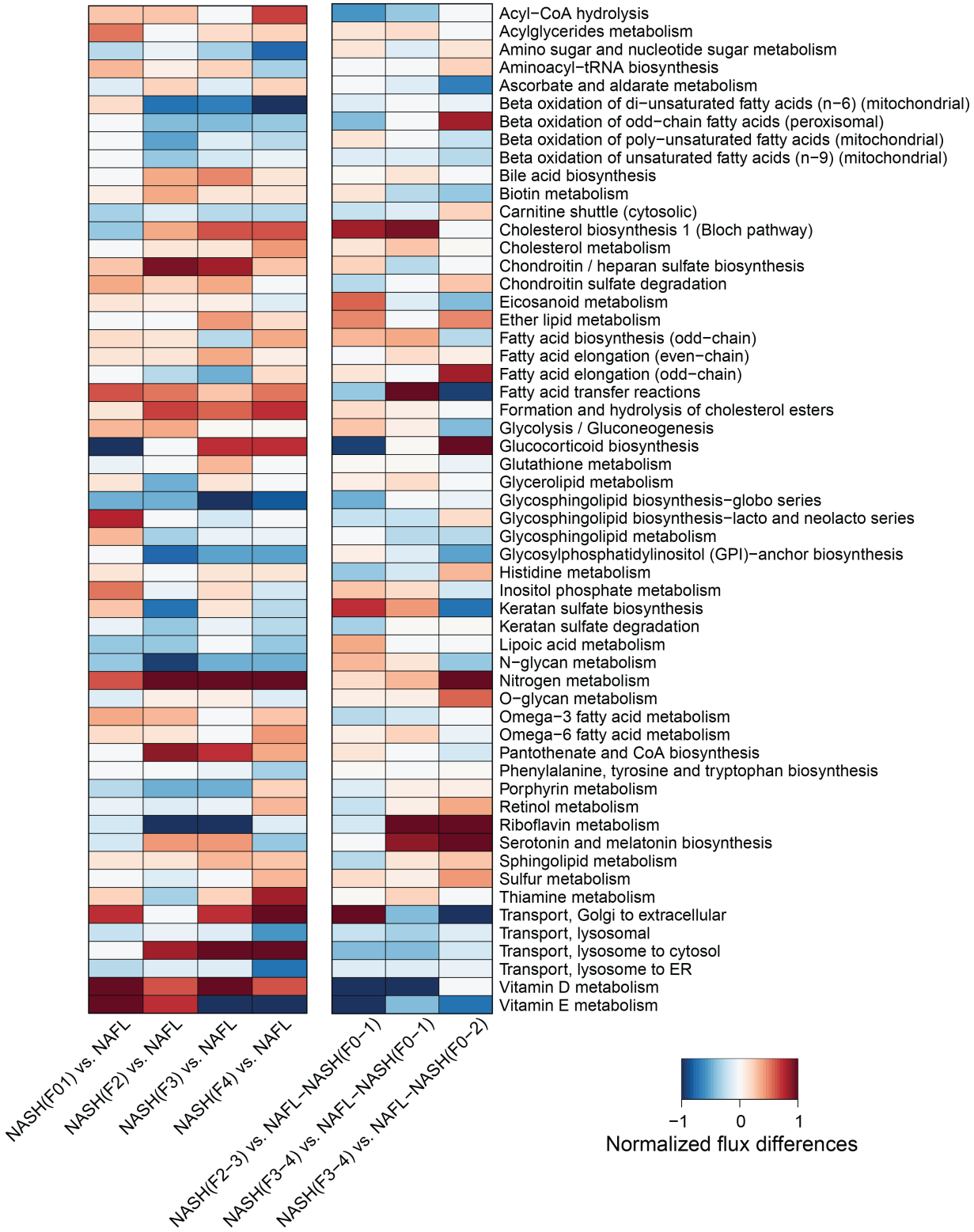

Acyl−CoA hydrolysis
Acylglycerides metabolism
Amino sugar and nucleotide sugar metabolism
Aminoacyl−tRNA biosynthesis
Ascorbate and aldarate metabolism
Beta oxidation of di−unsaturated fatty acids (n−6) (mitochondrial)
Beta oxidation of odd−chain fatty acids (peroxisomal)
Beta oxidation of poly−unsaturated fatty acids (mitochondrial)
Beta oxidation of unsaturated fatty acids (n−9) (mitochondrial)
Bile acid biosynthesis
Biotin metabolism
Carnitine shuttle (cytosolic)
Cholesterol biosynthesis 1 (Bloch pathway)
Cholesterol metabolism
Chondroitin / heparan sulfate biosynthesis
Chondroitin sulfate degradation
Eicosanoid metabolism
Ether lipid metabolism
Fatty acid biosynthesis (odd−chain)
Fatty acid elongation (even−chain)
Fatty acid elongation (odd−chain)
Fatty acid transfer reactions
Formation and hydrolysis of cholesterol esters
Glycolysis / Gluconeogenesis
Glucocorticoid biosynthesis
Glutathione metabolism
Glycerolipid metabolism
Glycosphingolipid biosynthesis−globo series
Glycosphingolipid biosynthesis−lacto and neolacto series
Glycosphingolipid metabolism
Glycosylphosphatidylinositol (GPI)−anchor biosynthesis
Histidine metabolism
Inositol phosphate metabolism
Keratan sulfate biosynthesis
Keratan sulfate degradation
Lipoic acid metabolism
N−glycan metabolism
Nitrogen metabolism
O−glycan metabolism
Omega−3 fatty acid metabolism
Omega−6 fatty acid metabolism
Pantothenate and CoA biosynthesis
Phenylalanine, tyrosine and tryptophan biosynthesis
Porphyrin metabolism
Retinol metabolism
Riboflavin metabolism
Serotonin and melatonin biosynthesis
Sphingolipid metabolism
Sulfur metabolism
Thiamine metabolism
Transport, Golgi to extracellular
Transport, lysosomal
Transport, lysosome to cytosol
Transport, lysosome to ER
Vitamin D metabolism
Vitamin E metabolism
NASH(F01) vs. NAFL
NASH(F2) vs. NAFL
NASH(F3) vs. NAFL
NASH(F4) vs. NAFL
NASH(F2−3) vs. NAFL−NASH(F0−1)
NASH(F3−4) vs. NAFL−NASH(F0−1)
NASH(F3−4) vs. NAFL−NASH(F0−2)
−1
0
1
Normalized flux differences

FIGURE 4

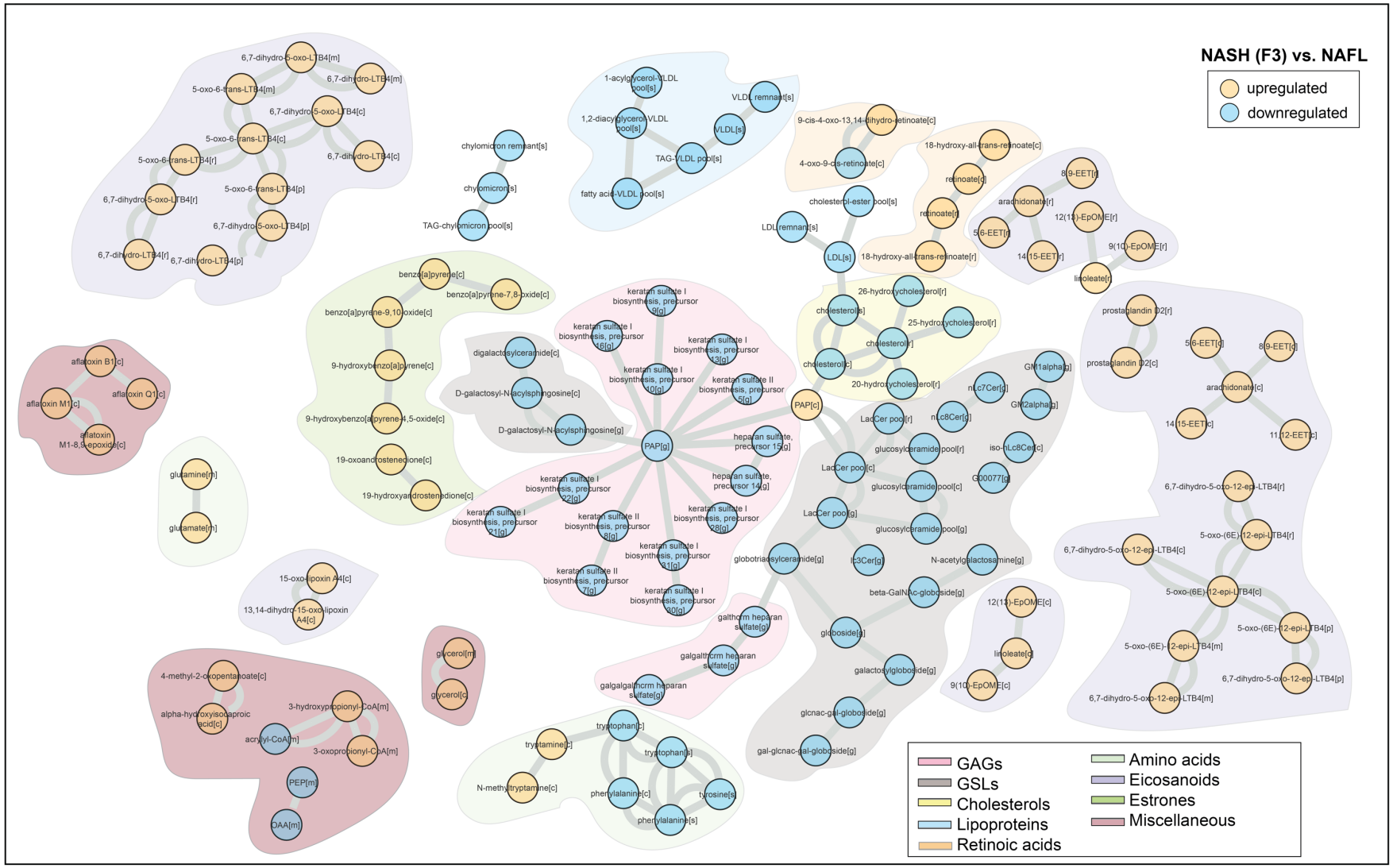

NASH (F3) vs. NAFL
upregulated
downregulated
GAGs
GSLs
Cholesterols
Lipoproteins
Retinoic acids
Amino acids
Eicosanoids
Estrones
Miscellaneous
6,7-dihydro-5-oxo-LTB4[m]
5-oxo-6-trans-LTB4[m]
6,7-dihydro-LTB4[m]
6,7-dihydro-5-oxo-LTB4[c]
5-oxo-6-trans-LTB4[c]
5-oxo-6-trans-LTB4[r]
6,7-dihydro-LTB4[c]
5-oxo-6-trans-LTB4[p]
6,7-dihydro-5-oxo-LTB4[r]
6,7-dihydro-5-oxo-LTB4[p]
6,7-dihydro-LTB4[r]
6,7-dihydro-LTB4[p]
chylomicron remnant[s]
chylomicron[s]
TAG-chylomicron pool[s]
1-acylglycerol-VLDL pool[s]
VLDL remnant[s]
1,2-diacylglycerol-VLDL pool[s]
VLDL[s]
TAG-VLDL pool[s]
fatty acid-VLDL pool[s]
9-cis-4-oxo-13,14-dihydro-retinoate[c]
4-oxo-9-cis-retinoate[c]
18-hydroxy-all-trans-retinoate[c]
retinoate[c]
cholesterol-ester pool[s]
retinoate[r]
LDL remnant[s]
LDL[s]
18-hydroxy-all-trans-retinoate[r]
8(9-EET[r]
arachidonate[r]
12(13)-EpOME[r]
5(6-EET[r]
14(15-EET[r]
9(10)-EpOME[r]
linoleate[r]
benzo[a]pyrene[c]
benzo[a]pyrene-7,8-oxide[c]
benzo[a]pyrene-9,10-oxide[c]
9-hydroxybenzo[a]pyrene[c]
9-hydroxybenzo[a]pyrene-4,5-oxide[c]
19-oxoandrostenedione[c]
19-hydroxyandrostenedione[c]
keratan sulfate I biosynthesis, precursor 9[g]
keratan sulfate I biosynthesis, precursor 16[g]
keratan sulfate I biosynthesis, precursor 13[g]
keratan sulfate I biosynthesis, precursor 10[g]
keratan sulfate II biosynthesis, precursor 15[g]
heparan sulfate, precursor 15[g]
heparan sulfate, precursor 14[g]
keratan sulfate I biosynthesis, precursor 22[g]
keratan sulfate I biosynthesis, precursor 21[g]
keratan sulfate II biosynthesis, precursor 8[g]
keratan sulfate I biosynthesis, precursor 28[g]
keratan sulfate I biosynthesis, precursor 31[g]
keratan sulfate II biosynthesis, precursor 7[g]
keratan sulfate I biosynthesis, precursor 30[g]
PAP[g]
PAP[c]
digalactosylceramide[c]
D-galactosyl-N-acylsphingosine[c]
D-galactosyl-N-acylsphingosine[g]
26-hydroxycholesterol[r]
cholesterol[s]
25-hydroxycholesterol[r]
cholesterol[r]
cholesterol[c]
20-hydroxycholesterol[r]
GM1alpha[g]
nLc7Cer[g]
GM2alpha[g]
LacCer pool[r]
nLc8Cer[g]
glucosylceramide pool[r]
iso-nLc8Cer[c]
LacCer pool[c]
glucosylceramide pool[c]
G00077[g]
LacCer pool[g]
glucosylceramide pool[g]
globotriaosylceramide[g]
lc3Cer[g]
N-acetylgalactosamine[g]
beta-GalNAc-globoside[g]
galthorm heparan sulfate[g]
galgalthorm heparan sulfate[g]
galgalgalthorm heparan sulfate[g]
globoside[g]
galactosylgloboside[g]
glcnac-gal-globoside[g]
gal-glcnac-gal-globoside[g]
12(13)-EpOME[c]
linoleate[c]
9(10)-EpOME[c]
prostaglandin D2[r]
prostaglandin D2[c]
5(6-EET[c]
8(9-EET[c]
arachidonate[c]
14(15-EET[c]
11(12-EET[c]
6,7-dihydro-5-oxo-12-epi-LTB4[r]
5-oxo-(6E)-12-epi-LTB4[r]
6,7-dihydro-5-oxo-12-epi-LTB4[c]
5-oxo-(6E)-12-epi-LTB4[c]
5-oxo-(6E)-12-epi-LTB4[p]
5-oxo-(6E)-12-epi-LTB4[m]
6,7-dihydro-5-oxo-12-epi-LTB4[p]
6,7-dihydro-5-oxo-12-epi-LTB4[m]
aflatoxin B1[c]
aflatoxin Q1[c]
aflatoxin M1[c]
aflatoxin M1-8,9-epoxide[c]
glutamine[m]
glutamate[m]
15-oxo-lipoxin A4[c]
13,14-dihydro-15-oxo-lipoxin A4[c]
glycerol[m]
glycerol[c]
4-methyl-2-oxopentanoate[c]
alpha-hydroxyisocaproic acid[c]
3-hydroxypropionyl-CoA[m]
acryl-CoA[m]
3-oxopropionyl-CoA[m]
PEP[m]
OAA[m]
tryptophan[c]
tryptamine[c]
tryptophan[s]
N-methyltryptamine[c]
phenylalanine[c]
tyrosine[s]
phenylalanine[s]

FIGURE 5

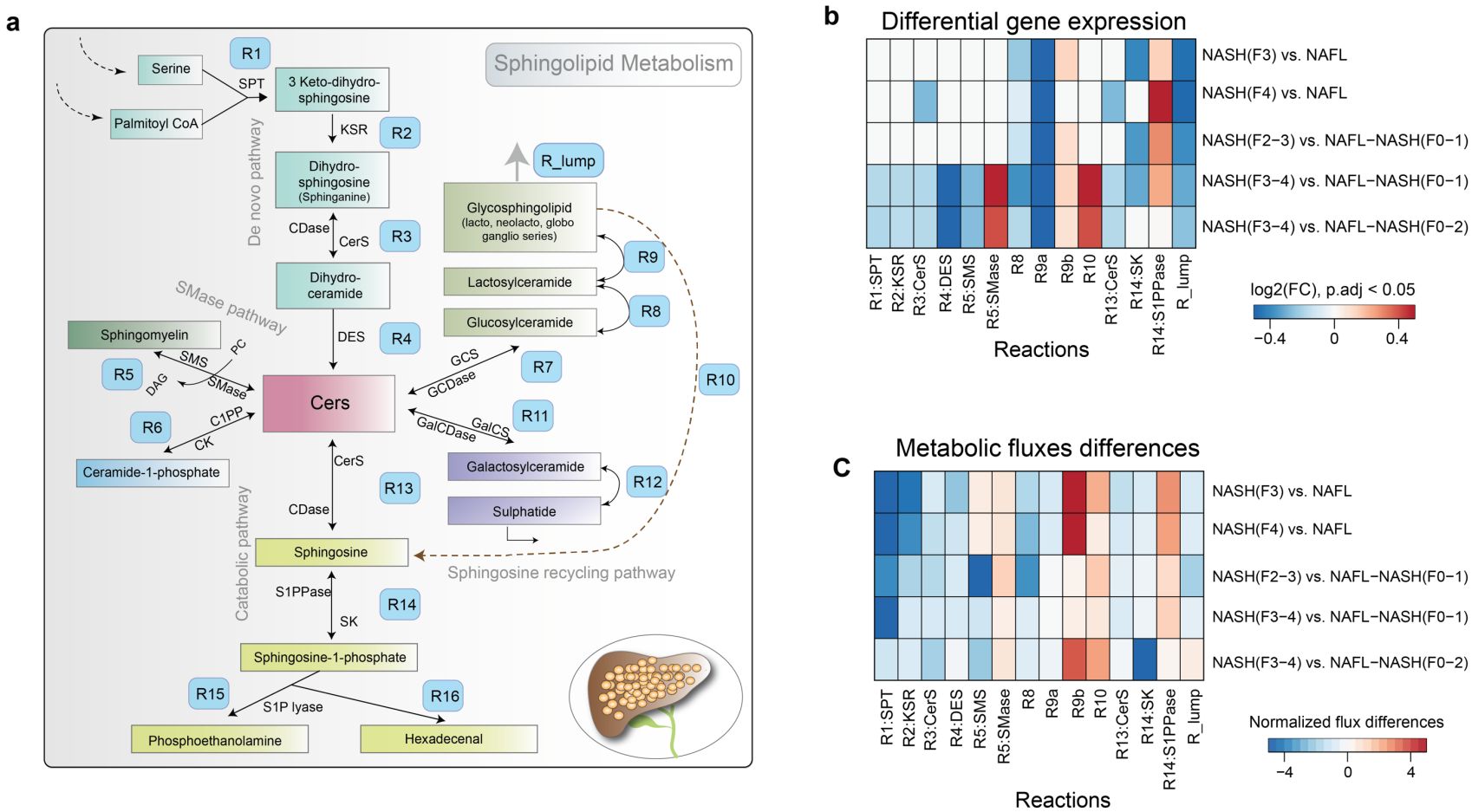

a
Sphingolipid Metabolism
Serine
Palmitoyl CoA
R1
SPT
3 Keto-dihydro-sphingosine
KSR
R2
Dihydro-sphingosine (Sphinganine)
CDase
CerS
R3
Dihydro-ceramide
DES
R4
De novo pathway
SMase pathway
Sphingomyelin
SMS
PC
DAG
SMase
R5
Cers
R6
C1PP
CK
Ceramide-1-phosphate
CerS
CDase
R13
Catabolic pathway
Sphingosine
S1PPase
SK
R14
Sphingosine-1-phosphate
R15
S1P lyase
R16
Phosphoethanolamine
Hexadecenal
R_lump
Glycosphingolipid (lacto, neolacto, globo ganglio series)
R9
Lactosylceramide
R8
Glucosylceramide
GCS
GCDase
R7
R10
GalCDase
GalCS
R11
Galactosylceramide
R12
Sulphatide
Sphingosine recycling pathway
b
Differential gene expression
NASH(F3) vs. NAFL
NASH(F4) vs. NAFL
NASH(F2−3) vs. NAFL−NASH(F0−1)
NASH(F3−4) vs. NAFL−NASH(F0−1)
NASH(F3−4) vs. NAFL−NASH(F0−2)
R1:SPT
R2:KSR
R3:CerS
R4:DES
R5:SMS
R5:SMase
R8
R9a
R9b
R10
R13:CerS
R14:SK
R14:S1PPase
R_lump
Reactions
log2(FC), p.adj < 0.05
−0.4
0
0.4
c
Metabolic fluxes differences
NASH(F3) vs. NAFL
NASH(F4) vs. NAFL
NASH(F2−3) vs. NAFL−NASH(F0−1)
NASH(F3−4) vs. NAFL−NASH(F0−1)
NASH(F3−4) vs. NAFL−NASH(F0−2)
R1:SPT
R2:KSR
R3:CerS
R4:DES
R5:SMS
R5:SMase
R8
R9a
R9b
R10
R13:CerS
R14:SK
R14:S1PPase
R_lump
Reactions
Normalized flux differences
−4
0
4

FIGURE 6

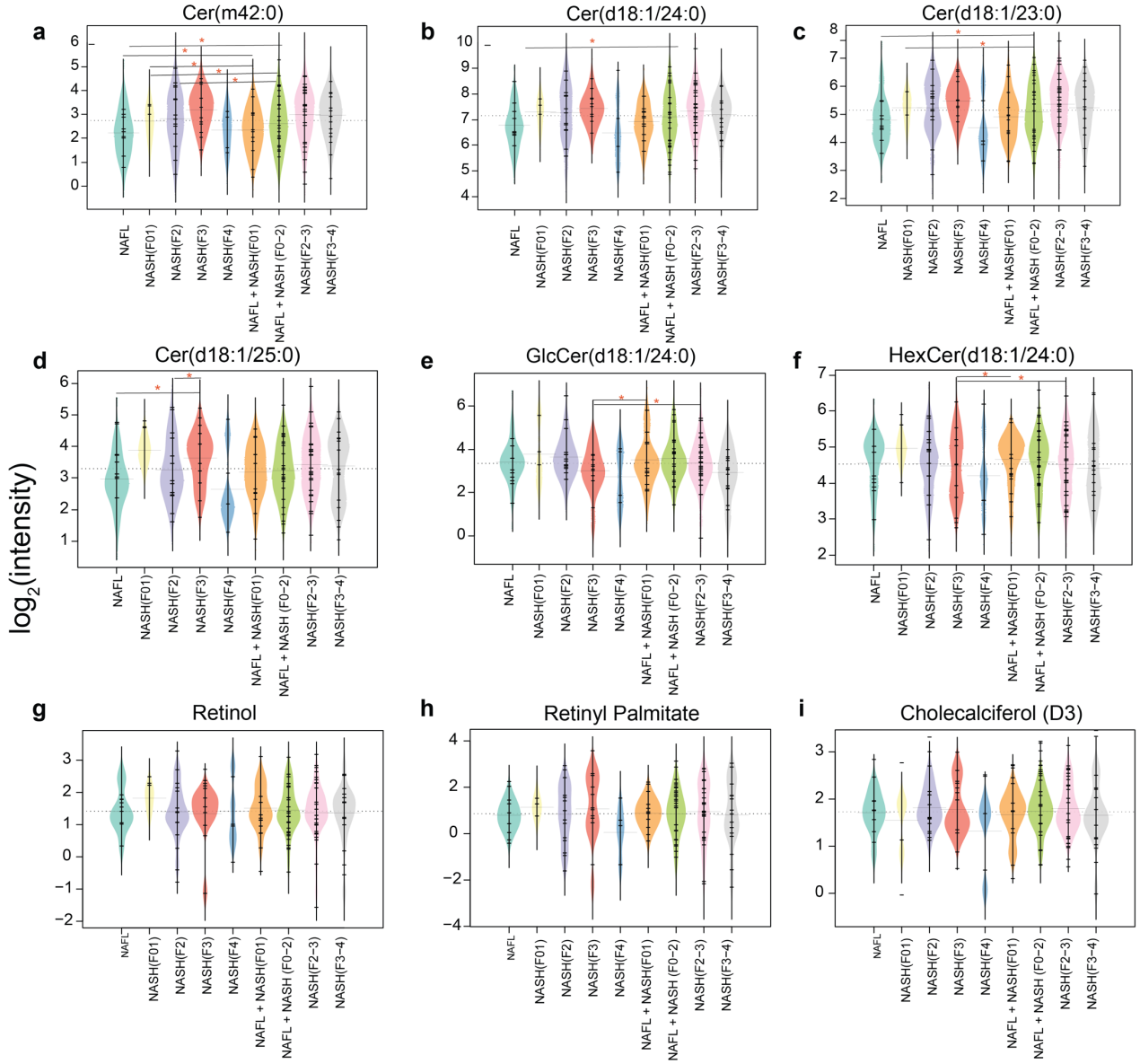

a
Cer(m42:0)
b
Cer(d18:1/24:0)
c
Cer(d18:1/23:0)
d
Cer(d18:1/25:0)
e
GlcCer(d18:1/24:0)
f
HexCer(d18:1/24:0)
g
Retinol
h
Retinyl Palmitate
i
Cholecalciferol (D3)
log2(intensity)
NAFL
NASH(F01)
NASH(F2)
NASH(F3)
NASH(F4)
NAFL + NASH(F01)
NAFL + NASH (F0−2)
NASH(F2−3)
NASH(F3−4)

FIGURE 7

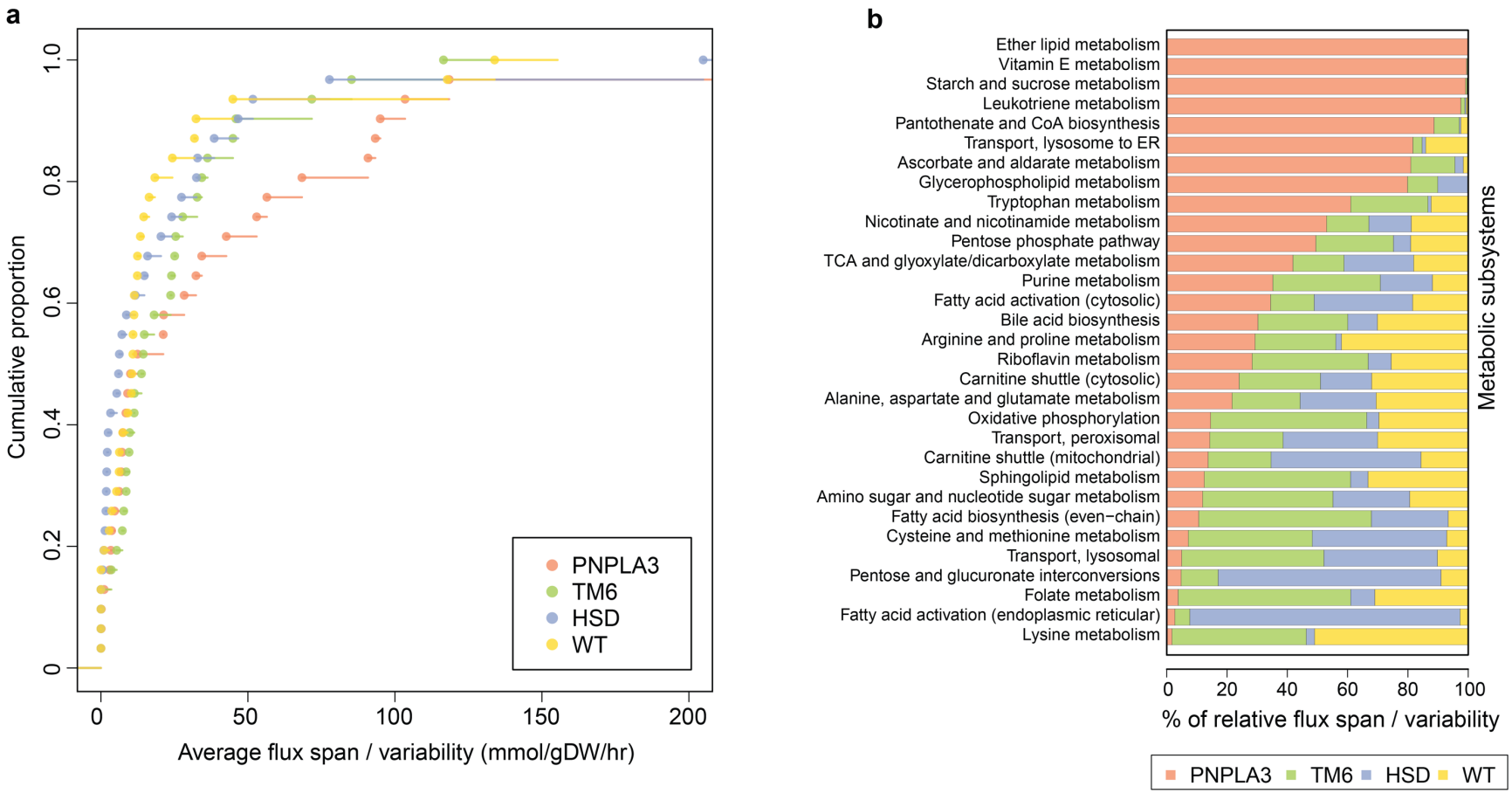

a
Cumulative proportion
0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
Average flux span / variability (mmol/gDW/hr)
PNPLA3
TM6
HSD
WT
b
Ether lipid metabolism
Vitamin E metabolism
Starch and sucrose metabolism
Leukotriene metabolism
Pantothenate and CoA biosynthesis
Transport, lysosome to ER
Ascorbate and aldarate metabolism
Glycerophospholipid metabolism
Tryptophan metabolism
Nicotinate and nicotinamide metabolism
Pentose phosphate pathway
TCA and glyoxylate/dicarboxylate metabolism
Purine metabolism
Fatty acid activation (cytosolic)
Bile acid biosynthesis
Arginine and proline metabolism
Riboflavin metabolism
Carnitine shuttle (cytosolic)
Alanine, aspartate and glutamate metabolism
Oxidative phosphorylation
Transport, peroxisomal
Carnitine shuttle (mitochondrial)
Sphingolipid metabolism
Amino sugar and nucleotide sugar metabolism
Fatty acid biosynthesis (even−chain)
Cysteine and methionine metabolism
Transport, lysosomal
Pentose and glucuronate interconversions
Folate metabolism
Fatty acid activation (endoplasmic reticular)
Lysine metabolism
Metabolic subsystems
0
20
40
60
80
100
% of relative flux span / variability
PNPLA3
TM6
HSD
WT

Table 1. Demographics of the EPoS RNA-Seq Cohort

Clinical features	Patients (n)	Values (mean ± SD and / or patients 'n')
Age (mean ± SD)	206	54 (± 11.87)
Sex		
male	206	123
female		83
BMI (mean ± SD)	204	31.34 (± 5.04)
T2DM	206	
no		96
yes		110
HBA1C (mmol/mol ± SD)	135	48.06 (± 14.54)
ALT (mean± SD)	204	67.12 (± 41.61)
AST (mean± SD)	201	44.67 (± 23.08)
Steatosis grade	206	
0		0
1		60
2		73
3		73
Ballooning	206	
0		52
1		98
2		56
Kleiner lobular Inflammation	206	
0		16
1		95
2		80
3		15
SAF lobular	206	

inflammation		
0		16
1		140
2		50
Brunt fibrosis stage	206	
0		38
1		47
2		53
3		54
4		14
NASH	206	
no		53
yes		153
NAS score ≥ 4	206	
no		58
yes		148
SAF activity score ≥ 2	206	
no		53
yes		153
HSD17B13 rs72613567 (--/-T/TT) [unknown]	188	120/61/7 [18]
PNPLA3 rs738409 (CC/GC/GG)	206	75/89/42
TM6SF2 rs58542926 (CC/CT/TT)	206	156/48/2

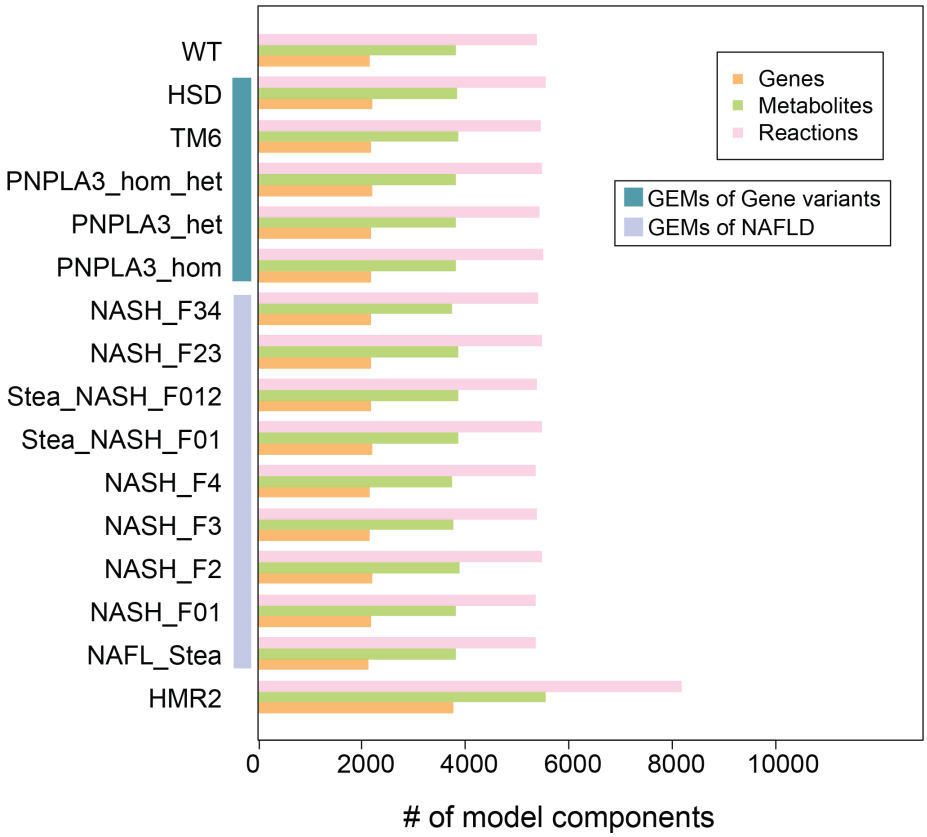
Quantitative genome-scale analysis of human liver reveals dysregulation of glycosphingolipid pathways in progressive nonalcoholic fatty liver disease

Supplementary Notes

Sen *et al.,*

Supplementary Figure 1

Genome-scale metabolic models of human hepatocytes in NAFLD. A bar plot comparing the number of genes, metabolites and reactions included in the genome-scale metabolic models (GEMs) of human hepatocytes contextualized for various stages of NAFLD. GEMs contextualized for three major genetic variants (*PNPLA3, TM6SF2* and *HSD17B13*) associated with risk and severity of NAFLD are shown as PNPLA3, TM6 and HSD respectively. 'het' and 'hom' represents homozygous and heterozygous respectively. 'HMR2' is Human Metabolic Reaction (GEM) version2.0 [1] and 'WT' is wild type.



Supplementary Figure 2

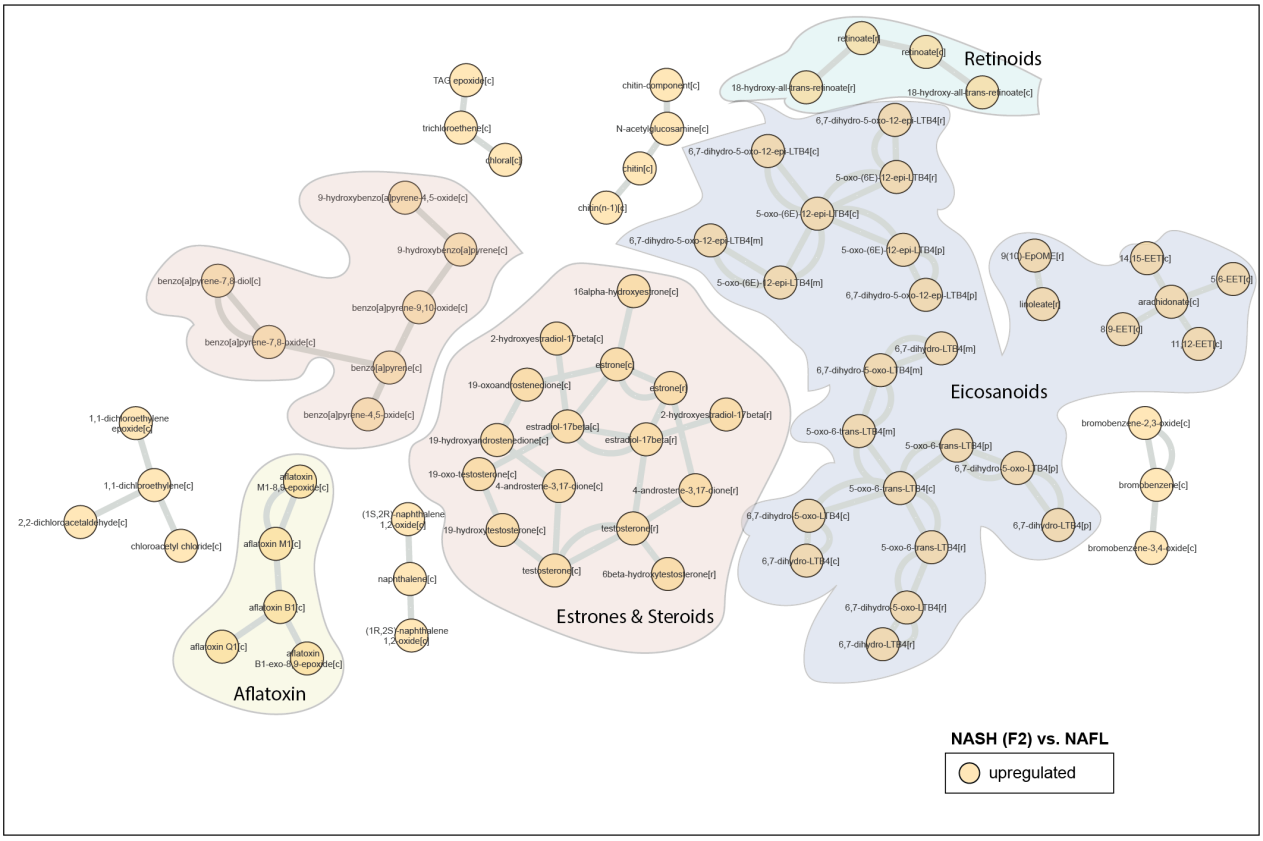


Figure S2. A metabolic-centric view of reporter metabolite (RM) modules that were significantly altered ($p<0.05$) between NASH F2 *vs.* NAFL. Orange color denotes upregulated. Each node represents a 'RM' and single or double lines represent reversible or irreversible metabolic reactions respectively. RMs that belong to a particular chemical class are color coded.

Supplementary Figure 3

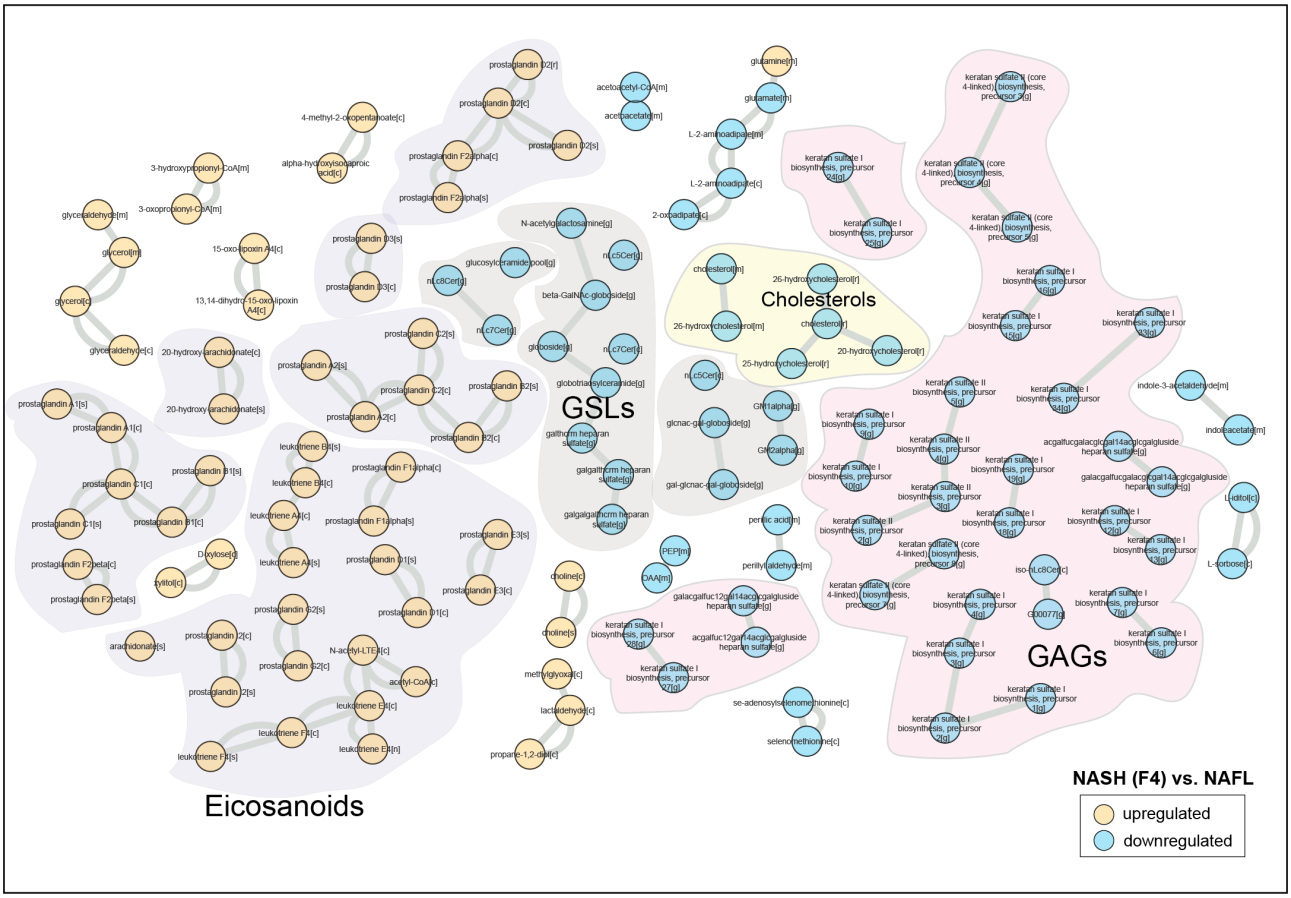


Figure S3. A metabolic-centric view of reporter metabolite (RM) modules that were significantly altered ($p<0.05$) between NASH F4 *vs.* NAFL. Orange and blue color denotes up and downregulated respectively. Each node represents a 'RM' and single or double lines represent reversible or irreversible metabolic reactions respectively. RMs that belong to a particular chemical class are color coded.

Supplementary Figure 4

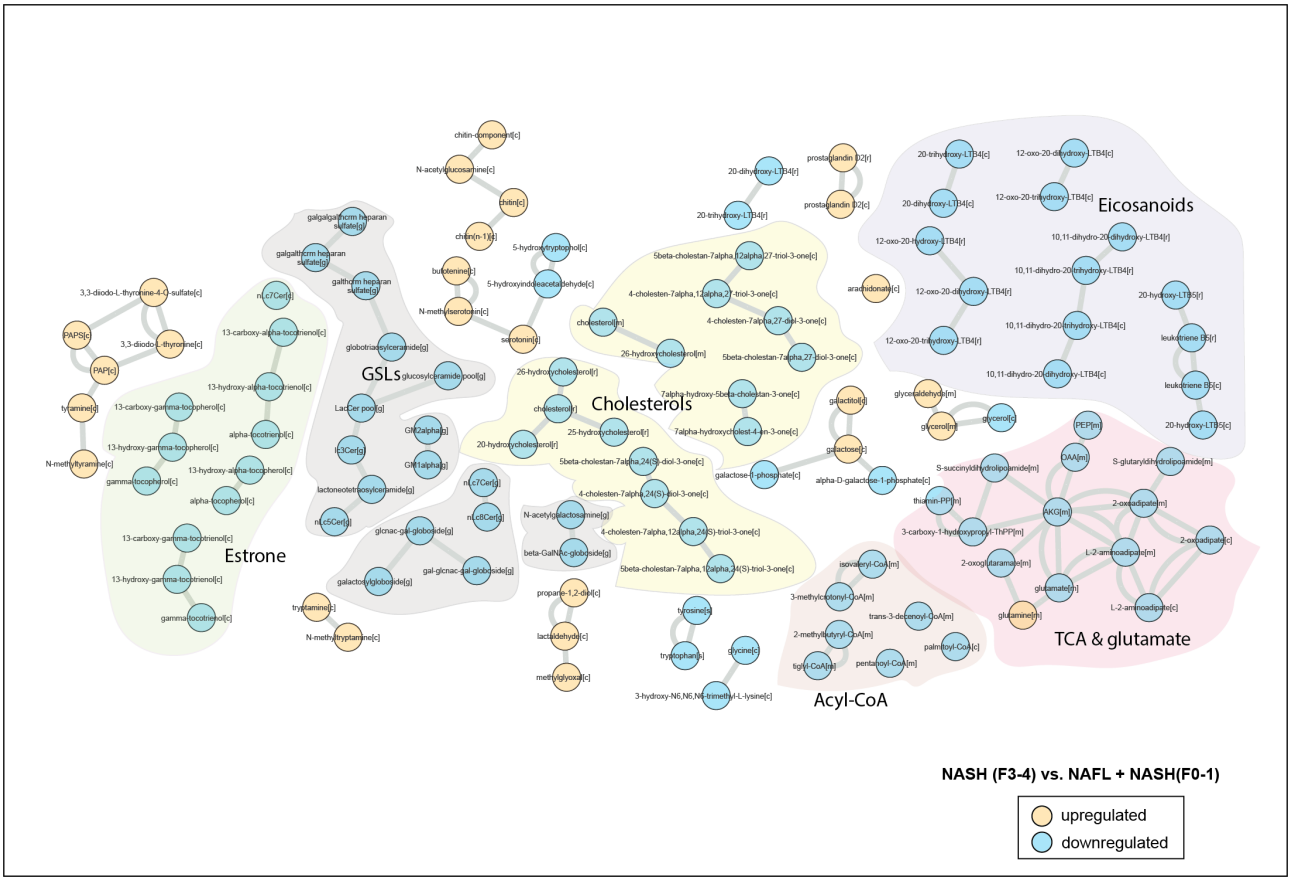


Figure S4. A metabolic-centric view of reporter metabolite (RM) modules that were significantly altered ($p<0.05$) between 'advanced' fibrosis *vs.* 'minimum' disease. Orange and blue color denotes up and downregulated respectively. Each node represents a 'RM' and single or double lines represent reversible or irreversible metabolic reactions respectively. RMs that belong to a particular chemical class are color coded.

Supplementary Figure 5

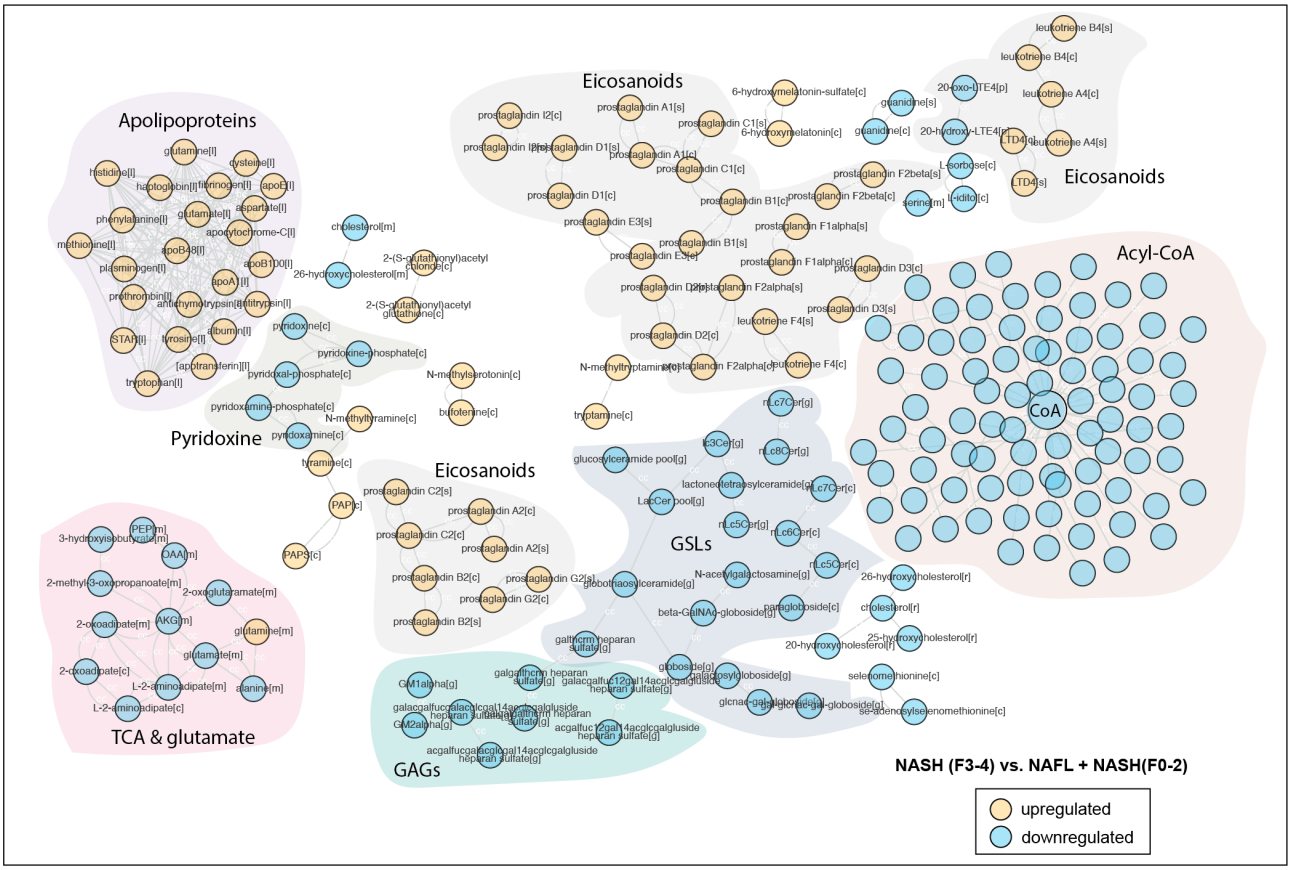


Figure S5. A metabolic-centric view of reporter metabolite (RM) modules that were significantly altered ($p<0.05$) between 'advanced' fibrosis *vs.* 'mild disease. Orange and blue color denotes up and downregulated respectively. Each node represents a 'RM' and single or double lines represent reversible or irreversible metabolic reactions respectively. RMs that belong to a particular chemical class are color coded.

Supplementary Figure 6

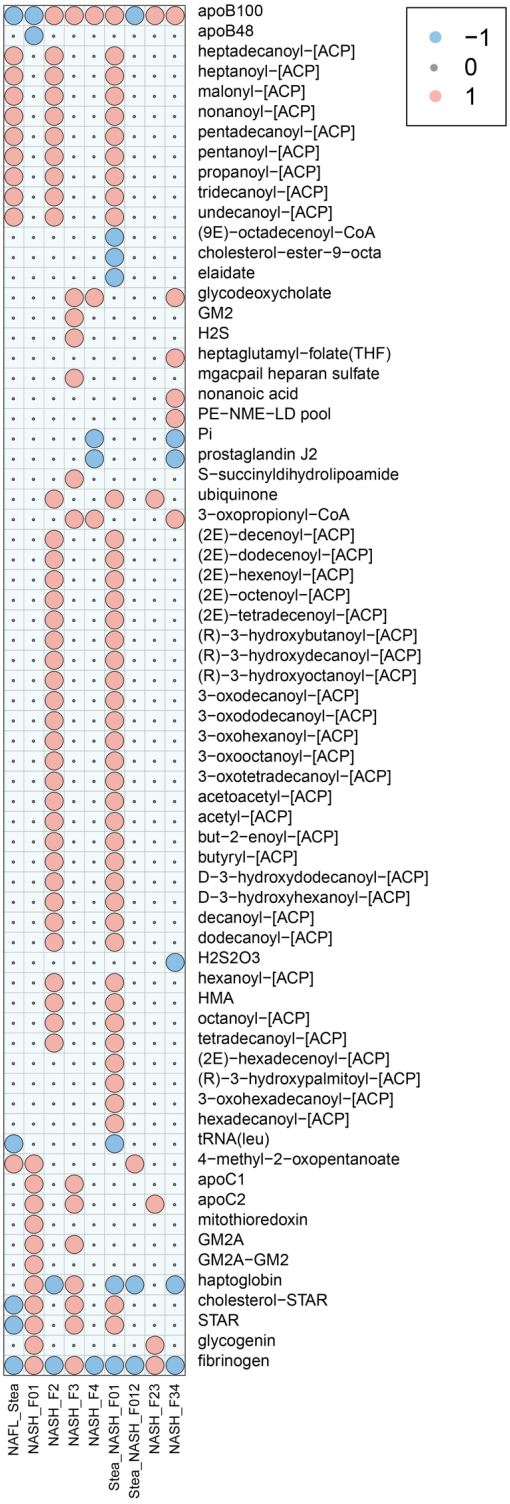


Figure S6. Showing shadow price (SP) of the metabolites across various stages of NAFLD estimated by flux imbalance analysis (FIA). Blue and orange color denotes negative and positive SPs respectively.

Supplementary Figure 7

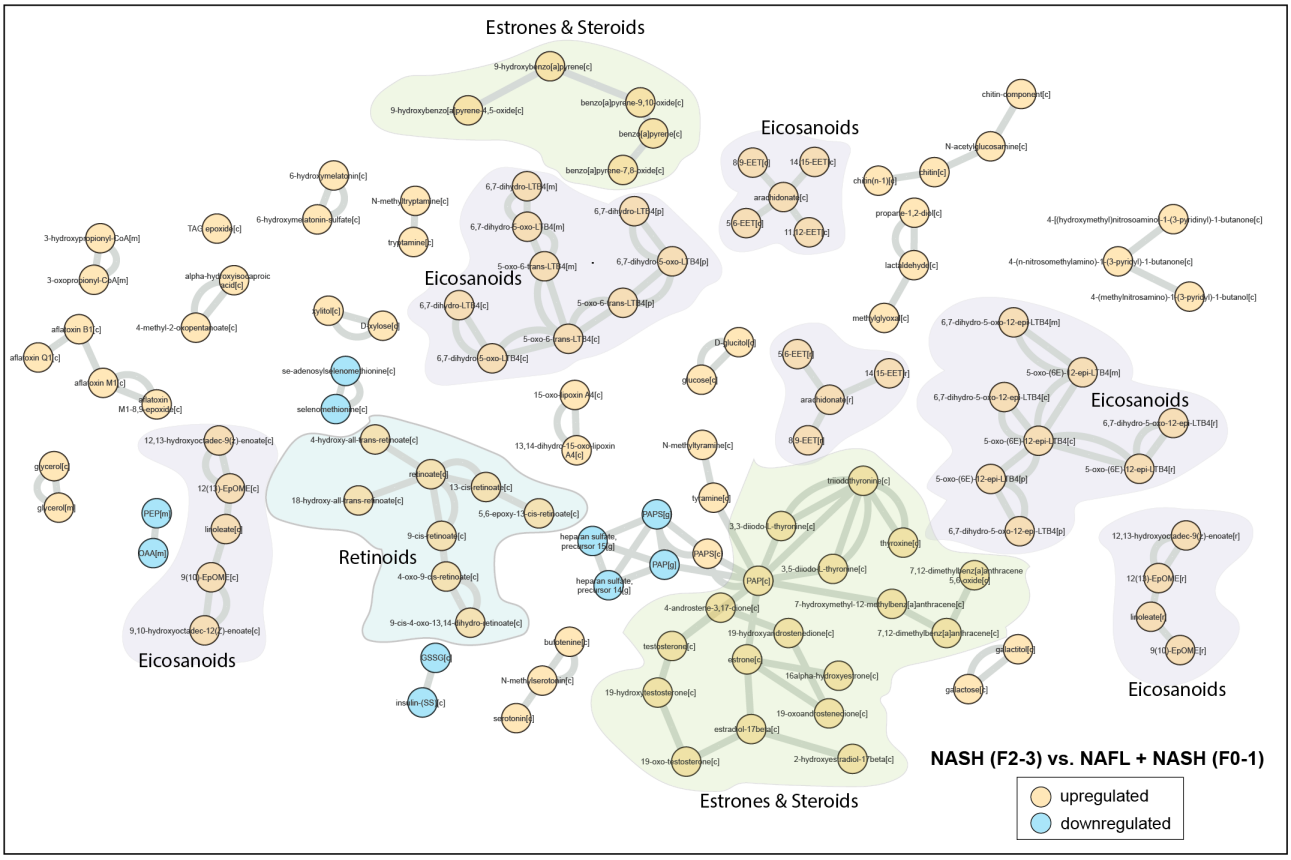


Figure S7. A metabolic-centric view of reporter metabolite (RM) modules that were significantly altered ($p<0.05$) between 'clinically significant' non-cirrhotic fibrosis *vs.* 'minimum' disease. Orange and blue color denotes up and downregulated respectively. Each node represents a 'RM' and single or double lines represent reversible or irreversible metabolic reactions respectively. RMs that belong to a particular chemical class are color coded.

Supplementary Figure 8

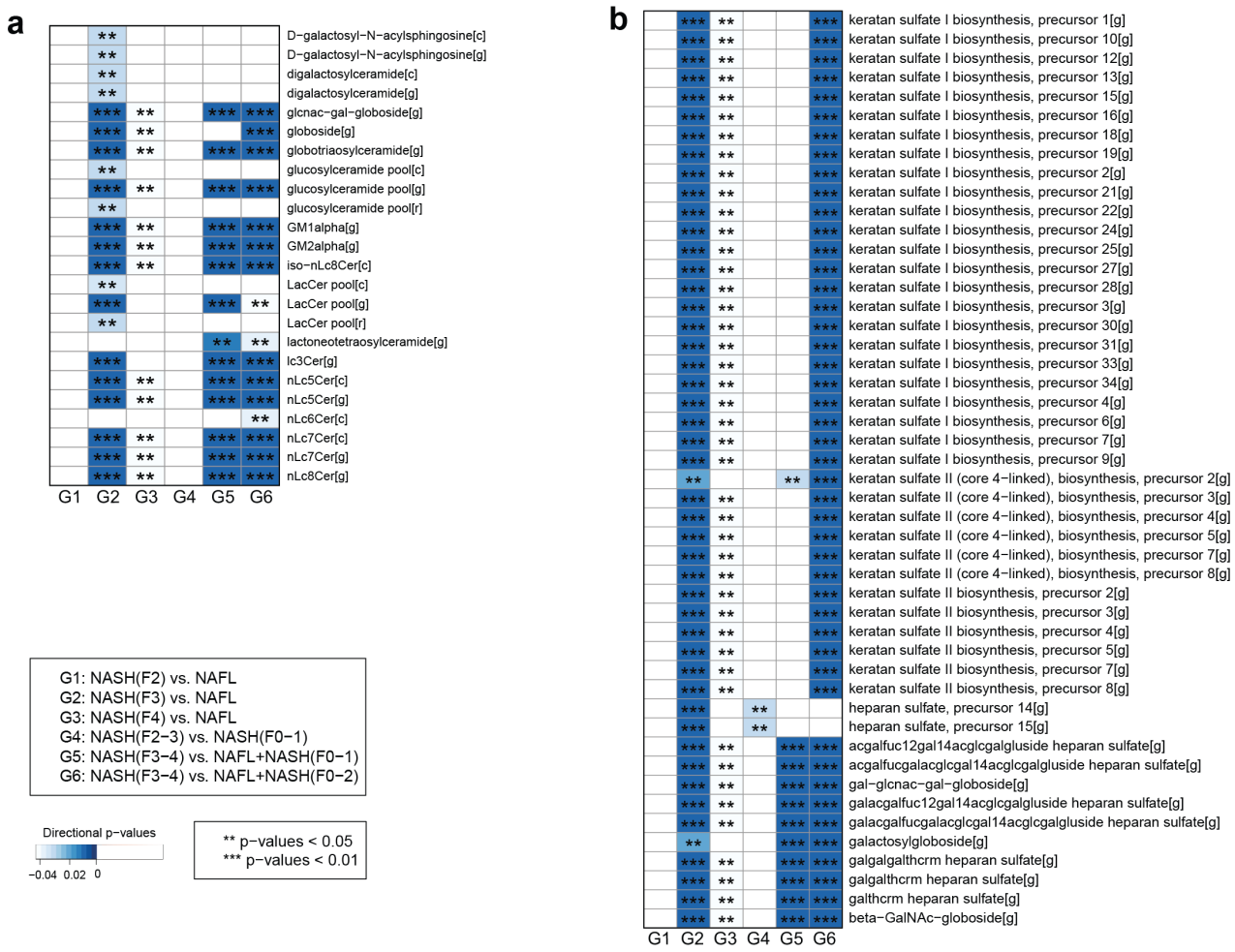


Figure S8. a-b) A heatmap showing RM clusters of glycosphingolipids (GSLs) and glycosaminoglycans (GAGs) that are significantly (***p<0.01, **p<0.05) downregulated (blue color) along the different stages NAFLD.

Supplementary Figure 9

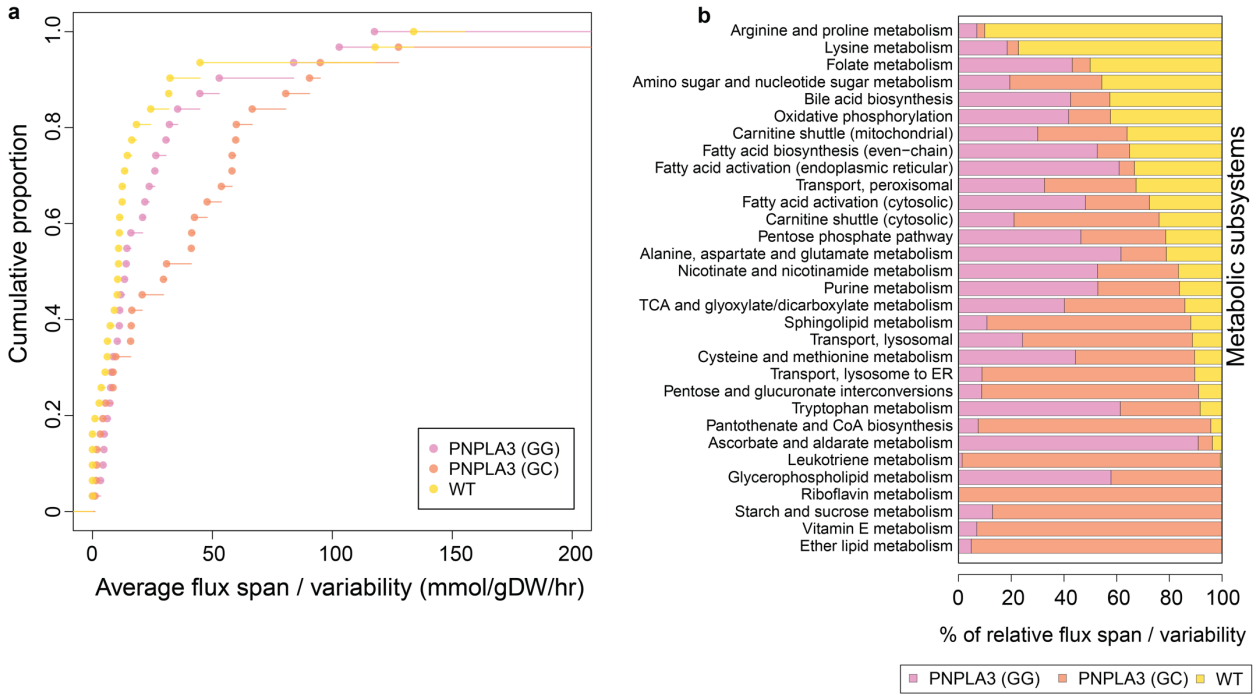


Figure S9. Flux variability across the metabolic subsystems of human hepatocytes carrying variants of PNPLA3 gene. a) Showing cumulative distribution / proportion of average flux span of all the reactions within a metabolic subsystem in the hepatocytes of the carriers of genetic variants of *PNPLA3* (homozygous: GG, n=47 and heterozygous: GC, n=22*)* and wild type (WT, n=36); i.e. individuals homozygous for all the three gene variants: *PNPLA3 (CC) , TM6SF2 (CC) and HSD17B13 (--)*(**Table S1**). b) Stacked bar plot showing the % of relative average flux span of each metabolic subsystem in the hepatocyte of the carriers of different genetic variants.

Supplementary Figure 10

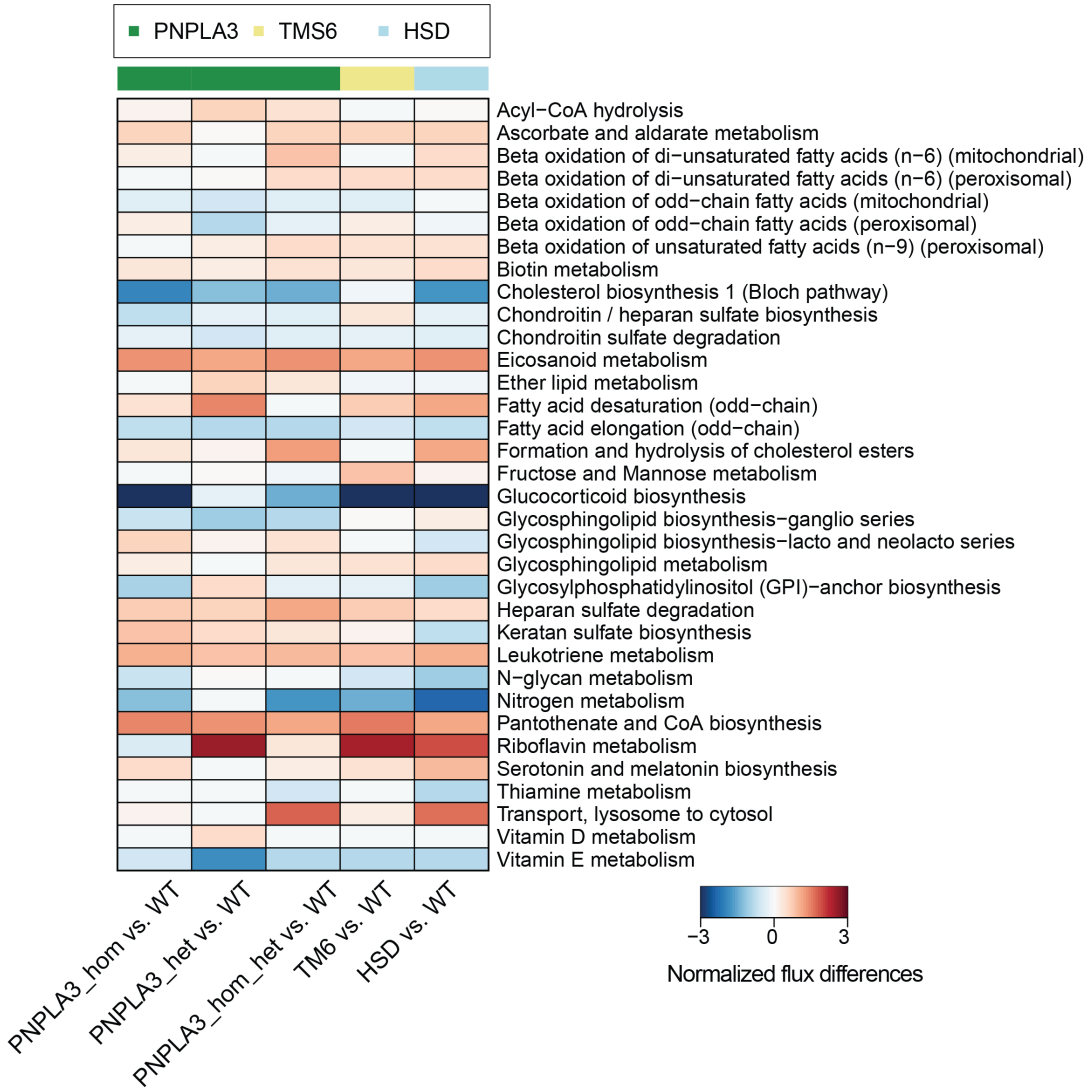


Figure S9. Normalized average flux differences in the hepatocytes of the carriers of three major genetic variants (*PNPLA3* (n=69), *TM6SF2* (n=13) and *HSD17B13* (n=21)) associated with risk and severity of NAFLD. Red and blue color denotes increase and decrease of fluxes in each subsystem between the gene variants and wild type (WT, n=36) models; i.e. individuals homozygous for all the three gene variants: *PNPLA3 (CC) , TM6SF2 (CC) and HSD17B13 (--)*. *'hom'* denotes homozygous (GG, n=47) , *'het'* denotes heterozygous (GC, n=22) and *'hom_het'* denotes combination of homo- and heterozygous (GG and GC, n=69)(**Table S1**).

Supplementary Table 1

Gene variants	RNA-seq cohort[2] (Individuals)	No. of participants
PNPLA3 rs738409 (CC/GC/GG)	206	
PNPLA3 rs738409 (CC)	75	**69** individuals exclusively exhibiting carriers of *PNPLA3 rs738409* (GG, GC) and homozygous / wild types for *TM6SF2 rs58542926* (CC) and *HSD17B13 rs72613567* (--).
PNPLA3 rs738409 (GG)	42	
PNPLA3 rs738409 (GC)	89	
TM6SF2 rs58542926 (CC/CT/TT)	206	
TM6SF2 rs58542926 (CC)	156	
TM6SF2 rs58542926 (CT)	2	**13** individuals exclusively exhibiting carriers of *TM6SF2 rs58542926* (CT, TT) and homozygous / wild types for *PNPLA3 rs73840* (CC) and *HSD17B13 rs72613567* (--).
TM6SF2 rs58542926 (TT)	48	
HSD17B13 rs72613567 (--/-T/TT)	188	
HSD17B13 rs72613567 (--)	120	**21** individuals exclusively exhibiting carriers of *HSD17B13 rs72613567* (-T, TT) and homozygous / wild types for *TM6SF2 rs58542926* (CC) and *PNPLA3 rs73840* (CC).
HSD17B13 rs72613567 (TT)	7	
HSD17B13 rs738409 (-T)	61	
Wild types, WTs (*PNPLA3 rs738409 (CC) + TM6SF2 rs58542926 (CC) + HSD17B13 rs72613567 (--))*	--	**36** individual wild types for *PNPLA3 rs738409, TM6SF2 rs58542926 and HSD17B13 rs72613567*)

Table S1. Selection of individuals carrying variants for *PNPLA3*, *TM6SF2* and *HSD17B13* genes associated with risk and severity of NAFLD.

References

1 Mardinoglu, A. *et al.* Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature communications* **5** (2014).

2 Govaere, O. *et al.* Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Science translational medicine* **12**, doi:10.1126/scitranslmed.aba4448 (2020).