

**Manuscript Number: HEP4-22-0432**

**Population based meta-analysis and gene set enrichment identifies FXR/RXR pathway as common to fatty liver disease and lipids**

Samuel K. Handelman<sup>#1</sup>, Yindra M. Puentes<sup>#1,2</sup>, Annapurna Kuppa<sup>#1</sup>, Yanhua Chen<sup>1</sup>, Xiaomeng Du<sup>1</sup>, Mary F. Feitosa<sup>3</sup>, Nicholette D. Palmer<sup>4^</sup>, Elizabeth K. Speliotes<sup>1,2\*^</sup>.

#Equally contributed

^ Equally contributed to supervision

1. Division of Gastroenterology and Hepatology, University of Michigan Health System, Ann Arbor, MI, 48109, USA.
2. Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, 48109, USA.
3. Division of Statistical Genomics, Department of Genetics, Washington University, Saint Louis, Missouri 63108, USA.
4. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA.

**\* Corresponding authors contact information:**

Elizabeth K. Speliotes, MD, PhD, MPH  
Keith Henley MD Collegiate Professor of Gastroenterology  
Professor of Medicine  
Division of Gastroenterology and Hepatology  
University of Michigan Health System  
1150 West Medical Center Drive  
Ann Arbor, MI 48109  
Phone: 734-647-2964  
Fax: 734-763-2535  
E-mail: [espeliot@med.umich.edu](mailto:espeliot@med.umich.edu)

**Financial support:** The AGES-Reykjavik Study is funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The Old Order Amish Study was supported by NIH research grants K01 DK067207, R01 AG18728, R01HL088119, U01 HL72515, and U01 HL084756. Partial funding was also provided by the Diabetes Research and Training Center of Maryland

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/hep4.2066](https://doi.org/10.1002/hep4.2066)

(P60 DK079637) and the Nutrition and Obesity Research Center of Maryland (P30DK072488). Support for FamHS was provided by the National Heart, Lung and Blood Institute (NHLBI) grant R01 HL08770003 and R01 HL117078 and National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant R01 DK089256. FHS is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract No. N01-HC-25195, HHSN268201500001I and 75N92019D00031). Funding for SHARe Affymetrix genotyping was provided by NHLBI Contract N02-HL64278. SHARe Illumina genotyping was provided under an agreement between Illumina and Boston University. Support for the GENOA study was provided by the NIH, grant numbers HL085571, HL087660 and HL100245 from NHLBI. Support for the IRASFS was provided by the NHLBI grants R01 HL060944, R01 HL061019, R01 HL060919, R01 HL060894 and R01 HL061210. Genotyping and analysis were supported by NIDDK grants DK085175, DK118062 and R01 DK118062. JHS is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the NHLBI and the National Institute on Minority Health and Health Disparities (NIMHD). Analyses in the UKBB were done under approved project 18120 (EKS). The authors also wish to thank the staffs and participants of the JHS. EKS, SKH, YP, AK, YC, XD, LFB, MFF, PAP, and NDP were supported by R01 DK106621 (to EKS), R01 DK107904 (to EKS), R01 DK128871 (to EKS and NDP), and University of Michigan. NDP was supported by R01 DK128871. SKH is presently an employee of Eli Lilly, but this work was done when we was working at the University of Michigan.

**Conflict of interest: None**

**Author Contributions:**

Study concept and design: EKS

Study supervision: EKS, NDP

Analysis and interpretation of data, statistical analysis: SKH, YP

Drafting of the manuscript: SKH, AK, NDP, EKS

Critical revision of the manuscript for important intellectual content: EKS, SKH, YP, AK, YC, XD, MFF, NDP

**ABSTRACT**

**Background:** Nonalcoholic fatty liver disease (NAFLD) is prevalent worldwide. NAFLD is associated with elevated serum triglycerides (TG) and low-density lipoprotein cholesterol (LDL) and reduced high-density lipoprotein cholesterol (HDL). Both NAFLD and blood lipid levels are genetically influenced and may share a common genetic etiology. **Methods:** We used genome-

wide association studies (GWAS) ranked genes and gene set enrichment analysis to identify pathways that affect serum lipids and NAFLD. We identified credible genes in these pathways and characterized missense variants in these for effects on serum traits. **Results:** We used MAGENTA to identify 58 enriched pathways from publicly available TG, LDL and HDL GWAS (N=99,000). Three of these pathways were also enriched for associations with European-ancestry NAFLD GWAS (N=7,176). One pathway, FXR/RXR activation was replicated for association in an African-ancestry NAFLD GWAS (N=3,214) and plays a role in serum lipids and NAFLD. Credible genes (proteins) in FXR/RXR activation include those associated with cholesterol/bile/bilirubin transport/absorption (*ABCC2*(MRP2), *ABCG5*(ABCG5), *ABCG8*(ABCG8), *APOB*(APOB), *FABP6*(ILBP), *MTTP*(MTP), *SLC4A2*(AE2)) and nuclear hormone mediated control of metabolism (*NR0B2*(SHP), *NR1H4*(FXR), *PPARA*(PPAR), *FOXO1*(FOXO1A)) or other pathways *FETUB*(FETUB). Missense variants in *ABCC2*(MRP2), *ABCG5*(ABCG5), *ABCG8*(ABCG8), *APOB*(APOB), *MTTP*(MTP), *NR0B2*(SHP), *NR1H4*(FXR), *PPARA*(PPAR) that associate with serum LDL levels also associate with serum liver function tests in UK Biobank. **Discussion:** Genetic variants in *NR1H4*(FXR) that protect against liver steatosis increase serum LDL cholesterol while variants in other members of the family have congruent effects on these traits. Human genetic pathway enrichment analysis can help guide therapeutic development by identifying effective targets for NAFLD/serum lipid manipulation while minimizing side effects. In addition, new missense variants could be used in companion diagnostics to determine their influence on drug effectiveness.

**Keywords:** NAFLD, Dyslipidemia, Genetic Epidemiology, Metabolic Syndrome, Pathway Analysis

## INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) is the hepatic manifestation of metabolic syndrome and has risen to a global prevalence ranging from 23-25% (1). The hallmark of NAFLD is increased intrahepatic triglyceride accumulation, i.e., steatosis, in the absence of excessive alcohol consumption. NAFLD may progress to more severe liver disease including nonalcoholic steatohepatitis (NASH), fibrosis, cirrhosis, and hepatocellular carcinoma. There are few treatments for NAFLD in part due to a poor understanding of its etiology.

At least 50% of patients diagnosed with NAFLD also have dyslipidemia (2); most often this is elevated serum concentrations of triglycerides (TG), increased low-density lipoprotein cholesterol (LDL) and reduced serum concentrations of high-density lipoprotein cholesterol

(HDL). Both serum lipids and NAFLD are heritable and genome-wide association studies (GWAS) have identified specific variants implicating genes that associate with these processes. A meta-analysis of GWAS in 7,176 individuals of European ancestry identified variants in or near five loci that associate with NAFLD (3). Variants near *GCKR* and *TM6SF2* also associate with serum lipid traits but in opposite directions suggesting a common genetic etiology to both serum and liver lipid accumulation that has not been extensively explored.

Here we systematically test for pathway enriched for genetic associations between NAFLD and serum lipids. We used Meta-Analysis Gene-set Enrichment of variant Associations (MAGENTA) (4) to identify pathways that have specific genetic associations between NAFLD and lipid traits. We identify common enriched genes and identify exonic variants in some genes and characterize their effects on multiple serum measures in UK Biobank. These results show how this type of analysis can identify effective targets for therapeutic intervention of NAFLD and serum LDL levels, identify their associations with other organ systems and identify coding mutations that may affect drug response in clinical trials.

## EXPERIMENTAL PROCEDURES

### Ethics statement

All research in this study was approved by the Institutional Review Board of the University of Michigan (Ann Arbor, MI). IRB approval was not required to use publicly accessible GWAS summary data. For institutional review board approvals and consent documentation, refer to the cited studies.

### GWAS data and cohort descriptions

GWAS summary statistics from five large-scale meta-analysis formed the basis of this analysis (**ST1**). Lipid associations (TG, LDL, and HDL) for the pathway analysis were drawn from the Global Lipids Genetics Consortium (GLGC) (5). NAFLD associations were drawn from meta-analysis results in European-ancestry populations (3) and in African-ancestry populations from the Genetics of Obesity-related Liver Disease (GOLD) Consortium (6). Among the GOLD cohorts, computed tomography (CT) scanning with a standardized protocol was used to measure hepatic steatosis which is correlated with decreased liver attenuation ( $r^2 = 0.92$ ) (7). Associations for body mass index (BMI) (8) and waist-to-hip ratio adjusted for BMI (WHRadjBMI) (9) were drawn from the Genetic Investigation of Anthropometric Traits (GIANT) Consortium. Associations for systolic and diastolic blood pressure (SBP and DBP, respectively)

were drawn from the International Consortium of Blood Pressure GWAS (10). UK Biobank has been described extensively elsewhere (11) and was expanded to include laboratory measures, quantitative traits, and biomarkers.

### **MAGENTA analysis to identify pathways that affect serum lipids and NAFLD**

Functional enrichment analysis as implemented in MAGENTA v2.4 (4) was used to test gene sets for enrichment of genetic association signals. Briefly, MAGENTA uses linkage disequilibrium (LD) pattern information to identify independent variants which are assigned to a gene based on genomic coordinates, i.e., 110kb upstream and 40kb downstream of the open reading frame. It then assigns the p-value of the most significant variant to the gene, corrected for gene size, the independent variants per kilobase, the genetic distance and the LD using a stepwise multivariate regression approach. To evaluate pathway enrichment, the 75<sup>th</sup> percentile cutoff for the rank of the gene was used. MAGENTA compares the number of genes in the gene set with gene-wise p-values less than this cutoff to the expected number of genes, then computes a nominal gene set enrichment p-value as well as a false-discovery rate (FDR) based on 10,000 randomly chosen gene sets of identical size to the putatively enriched gene set.

MAGENTA was used to identify pathways from 10,994 gene sets from Ingenuity, Gene Ontology, REACTOME, BIOCARTA, KEGG and PANTHER that were enriched for genetic associations with serum lipid traits, i.e., TG, LDL, and HDL. Serum lipid genetic association enriched pathways were tested for genetic association enrichment pathways with NAFLD in European-ancestry individuals. Pathways that were enriched for serum lipids and European ancestry NAFLD were then tested for genetic association enrichment with NAFLD in African-ancestry individuals. A 75<sup>th</sup> percentile cutoff for the rank of corrected gene p-values to evaluate enrichment and significance at a MAGENTA permutation FDR of  $\alpha=0.1$  to account for multiple testing were specified. In addition, European ancestry NAFLD association enriched pathways were tested for genetic association enrichment with BMI, WHRadjBMI, DBP and SBP.

### **Relative credibility of individual NAFLD genes**

To identify credible genes in the FXR/RXR Activation pathway that were identified in the genetic enrichment for NAFLD, p-values for association with NAFLD from the European and African ancestries were meta-analyzed using Fisher's method. These meta-analysis p-values are converted to minimum Bayes Factors.

### **Pleiotropic effects of non-synonymous credible gene variants in the UK Biobank**

We carried out association analysis using publicly available UK Biobank summary statistics (<https://pan.ukbb.broadinstitute.org/>) for non-synonymous variants in credible genes and then prioritized them based on the lowest p-value across all traits considered. FDR adjusted p-values were calculated based on 81 variants tested for the most significant trait for the association analysis. SNP-specific annotation information was obtained using ANNOVAR (12). LD information between variants was obtained using NIH LDlink (<https://ldlink.nci.nih.gov>) for European-ancestry populations from 1000 Genomes. An  $r^2$  of  $< 0.5$  was used to determine independence. Laboratory measures and other quantitative traits considered from UK Biobank were: BMI, albumin, alkaline phosphatase (ALP), alanine aminotransferase (ALT), apolipoprotein A (APOA), apolipoprotein B (APOB), aspartate aminotransferase (AST), direct bilirubin (DBili), urea, calcium, cholesterol, creatinine, C-reactive protein (CRP), cystatin C, gamma-glutamyl transferase (GGT), glucose, glycated hemoglobin (HbA1c), HDL, insulin-like growth factor-1 (IGF-1), LDL, lipoprotein A, oestradiol, phosphate, rheumatoid factor, sex hormone binding globin (SHBG), total bilirubin (TBili), testosterone, total protein, TG, urate, vitamin D, (DBP, SBP and waist circumference (Waist.Cir). Co-heritability was also estimated between these traits using the LCV package (13) across all variants for which European ancestry LD scores (14) were available.

## RESULTS

### Setting and Participants

Publicly available GWAS summary statistics from European-ancestry studies were used for serum lipids (N=100,184 for TG, 95,454 for LDL, 99,900 for HDL and 96,598 for TG) (5), blood pressure (N=69,395 for SBP and DBP) (10) and anthropometric (N=249,796 for BMI and 77,167 for WHRadjBMI) (8) (9) traits. For NAFLD analysis, data from the GOLD Consortium were used (3) (6). This included 7,176 individuals of European-ancestry and 3,124 individuals of African-ancestry with computed tomography (CT) measured NAFLD.

### Pathways enriched for lipid trait associated genes are also enriched in NAFLD

To identify pathways common to serum lipids and NAFLD European populations, GWAS summary statistics for LDL, HDL and triglycerides were used to assign genetic association p-values to all genes in the human genome using MAGENTA v2.4 (**Figure 1**). Gene set enrichment analysis identified 58 gene sets that were enriched for genetic associations with lipid traits (FDR $<0.1$ ) (**ST2**). Summary statistics from the GOLD European-ancestry NAFLD GWAS

were then used to assign genetic association p-values to all genes in the genome and tested against the 58 lipid associated pathways to assess enrichment for genetic associations with European ancestry NAFLD results. Three pathways, hepatic cholestasis ( $P_{FDR}=0.054$ ), FXR/RXR activation ( $P_{FDR}=0.062$ ) and chylomicron-mediated lipid transport ( $P_{FDR}=0.076$ ) were found to be enriched for genetic associations with liver attenuation (**Table 1**). Of these, only FXR/RXR activation, as defined by Ingenuity, also showed significant enrichment ( $P_{FDR}=0.0089$ ) for liver attenuation in individuals of African ancestry (**Table 1**). To test whether the pathway enrichment was specific to lipids and NAFLD, we also tested metabolic syndrome traits for joint genetic association with NAFLD. We assigned genetic association values to genes based on GWAS summary statistics for BMI, WHRadjBMI, DBP or SBP. None of the lipid/NAFLD associated pathways were enriched for associations with any of these traits supporting that those pathways were specific for serum lipids and NAFLD (**ST3**).

### Credible gene identification

Top associated NAFLD signals in European and African-ancestry individuals were examined in the FXR/RXR Activation pathway and found to be driven by different variants in the different ancestries (**ST4**, **ST5**). The signals were distinct even when extended LD in European-ancestry populations was considered, except for the lead variants in *APOB* which were the same in both ancestries.

To determine whether particular genes were enriched for associations in both European- and African-ancestry individuals, genes above the 75<sup>th</sup> percentile were examined. Twenty-three genes of the 54 effective gene sets were above the 75<sup>th</sup> percentile across European- and African-ancestry individuals. Eleven genes were above the 75<sup>th</sup> percentile cutoff for NAFLD in both European- and African-ancestry populations; this was not a statistically significant overlap ( $p=0.58$ , Fisher's exact test). To identify credible genes that were responsible for the genetic enrichment for NAFLD, we meta-analyzed the p-values for association with NAFLD from the European and African ancestries and converted the meta-analysis p-values to minimum Bayes Factors. Genes that had a minimum Bayes Factor at least  $\sqrt{10}$ -fold versus the median gene in the pathway were defined as credible. **Figure 2** provides a graphical representation of the statistical evidence of the gene (protein) associations in the different ancestries. *APOB* (*APOB*), *ABCC2* (*MRP2*), *ABCG8* (*ABCG8*), *ABCG5* (*ABCG5*), *NRIH4* (*FXR*), *FOXO1* (*FOXO1A*), *PPARA* (*PPAR*), *FETUB* (*FETUB*), and *FABP6* (*ILBP*) had highly statistically significant associations (Bayes Factor  $> \sqrt{10}$ -fold) in both ancestries. *SLC4A2* (*AE2*), *NR0B2* (*SHP*) and *MTTP* (*MTP*) had strong associations in European-ancestry populations and less significant

Author Manuscript

associations in African-ancestry populations. Lead variants in these credible genes are summarized in **Table 2**. All these variants had a Z-score of 3 or greater for promoting fatty liver in European-ancestry populations in GOLD. The functions of the corresponding genes/proteins are summarized in **Table 2** and graphically in **Figure 3**. They play a role in cholesterol (*ABCG8* (*ABCG8*), *ABCG5* (*ABCG5*), *APOB* (*APOB*), *MTTP* (*MTP*)), bile acid (*ABCC2* (*MRP2*), *FABP6* (*ILBP*), *NR1H4* (*FXR*), *PPARA* (*PPAR*)) and glucose (*FOXO1* (*FOXO1A*), *NR0B2* (*SHP*)) biology as well as solute secretion (*SLC4A2* (*AE2*)) and tumor biology (*FETUB* (*FETUB*)).

### Variants in credible genes associate with a range of laboratory measures in the UK Biobank

To better define the potential effects of credible genes on physiology, we identified and evaluated non-synonymous variants in credible genes for statistically significant associations with continuous serum and anthropometry measures in UK Biobank (**ST6, ST7**). Variants in high LD were collapsed into single entries when  $r^2 > 0.9$ . *ABCG5* R50C and *ABCG8* D19H share a common haplotype (rs11887534 and rs6756629,  $r^2 = 0.95$  in European-ancestry populations), as do *ABCC2* (*MRP2*) V1188E and C1515Y (rs17222723 and rs8187710,  $r^2 = 0.98$  in European-ancestry populations), and *APOB* (*APOB*) L2739P and S4338N (rs676210 and rs1042034,  $r^2 = 1.00$  in European-ancestry populations); all other variants were independent. We show the association for up to two variants per gene (three in *APOB* (*APOB*) to show the range of effect) with an FDR adjusted p-value less than 0.1 for associations with serum lipids or liver function tests (**Figure 4, ST7**) and with all traits (**ST6**). **ST6** contains all non-synonymous variants in the credible genes and association p-value with all quantitative traits reported in the UK Biobank. These non-synonymous variants were generally not in LD with the lead variants from **Table 2** (see **ST8**).

For results presented in **Figure 4**, effect sizes are oriented towards the LDL-increasing allele owing to LDL having sufficient sample size for accurate effect estimate, unless otherwise noted. As seen in Figure 4, variants in *ABCC2* (*MRP2*) (V1188E, C1515Y), *ABCG5* (*ABCG5*) (G27A), *ABCG5/8* (*ABCG5/8*) (C50R, H19D), and *NR1H4* (*FXR*) (T183M) were associated with increased serum LDL but decreased serum ALT, the most sensitive and specific serum marker of hepatocyte injury. Variants in *MTTP* (*MTP*) (S688G), *PPARA* (*PPAR*) (A286V) and *NR0B2* (*SHP*) (G171A) were associated with increased serum LDL and increased serum ALT. Variants in *APOB* (*APOB*) (T98I, R1923H) and *FABP6* (*ILBP*) (M124I) had statistically significant associations with LDL but did not have statistically significant associations with ALT. A variant in *SLC4A2* (*AE2*) (Q311R) had a strong association with bilirubin. Variants in *APOB* (*APOB*) (T98I,



R1923H, L2739P/S4338N) associated with increased LDL, increased serum TG and decreased serum HDL. Variants in *APOB* (*APOB*) that increased LDL also associated with decreased ALP (T98I, L2739P/S4338N) and decreased direct bilirubin (T98I, R1923H). Variants in *FABP6* (*ILBP*) (M124I) associated with increased serum LDL were also associated with increased serum HDL. A variant in *PPARA* (*PPAR*) (A268V) associated with increased serum LDL also associated with increased serum TG. A variant in *NR1H4* (*FXR*) (T183M) that was associated with serum LDL was associated with serum HDL. A variant in *NR0B2* (*SHP*) (G171A) that was associated with LDL was also associated with serum TG and serum HDL. Other notable novel associations include *NR1H4* (*FXR*) T183M with CRP and sex hormone binding globulin and *NR0B2* (*SHP*) (G171A) with CRP (**ST6**).

For the credible genes, except for *FETUB* (*FETUB*) and *FOXO1* (*FOXO1A*), at least one non-synonymous variant had a significant association on one or more of the 35 serum quantitative traits (**ST6**). In contrast, many of these variants did not associate with overall obesity as measured using BMI or abdominal obesity as measured using Waist Cir (**ST6**). These results support the initial observation that gene associations with the FXR/RXR pathway are specific to serum lipids and liver function and not to the broader set of serum quantitative traits.

### **Specific genetic variants dissociate genetically co-heritable traits.**

We also examined the effects of variants on traits (**ST6**) relative to the co-heritability of traits (**ST9**) in the European population in the UK Biobank. *SLC4A2* (*AE2*) Q311R decreases bilirubin but does not affect other liver function tests consistent with the genetic co-heritability estimates between these traits. In contrast, *ABCC2* (*MRP2*) V1188E/C1515Y and G921S and *ABCG5/ABCG8* (*ABCG5/ABCG8*) C50R H19D decrease bilirubin and ALT, AST, and ALP. The pronounced increase of sex hormone binding globulin (*SHBG*) in *NR1H4* (*FXR*) T183M is expected with decreased ALT (15). ALT and HDL are negatively correlated across the genome, so *NR0B2* (*SHP*) G171A shows this pattern, while *NR1H4* (*FXR*) T183M shows associations in the same direction for these traits. Finally, LDL is generally, although not strongly, positively correlated with ALT and AST across the genome. Therefore, the effects of *ABCC2* (*MRP2*) (V1188E/C1515Y), *ABCG5/ABCG8* (*ABCG5/ABCG8*) (G27A or C50R/H19D) and *NR1H4* (*FXR*) (T183M) all increase LDL while decreasing ALT and AST.

## **DISCUSSION**

We have performed a systematic evaluation of pathways enriched for genetic associations between NAFLD and serum lipids with results supporting a causal role for genetic variation. For some members of the pathway, effects on serum lipids and NAFLD were congruent whereas others were opposing. This is important for therapeutic development as toxicity to the liver is one of the most common reasons for drug failure. Knowledge of potential toxicities will inform drug development targeted towards altering serum lipids.

The novel approach implemented in this study focuses on pathway level reproducibility across populations and is not subject to the limitations of requiring variant reproducibility across populations. Although there is overlapping signal in the FXR/RXR pathway for both European and African ancestry populations, the individual variants are largely different. Even if the associations of individual variants mostly generalize between populations (16), differences in allele frequencies could result in a differential set of above-threshold genes between groups, especially if these genes are subject to selection (17). These results also show that highly significant p-values were not an inevitable consequence of follow-up studies in a large sample like the UK Biobank, which may be overpowered. However, these data support a pathway- or systems-level interpretation of the findings owing to the genetic evidence in two independent populations as studied herein. Therefore, the interpretation of these findings is dependent on the molecular biology and biochemistry of the individual genes.

Among genes within the FXR/RXR pathway, results from genetic variants highlight consistency with the known biology, i.e. their role in cholesterol homeostasis or metabolic syndrome (18) is well-characterized. *APOB* (*APOB*) is the primary apolipoprotein in chylomicrons and LDL with increased expression observed in liver tissue of NAFLD patients (19). *PPARA* (peroxisome proliferator-activated receptor alpha) (*PPAR*) is part of the nuclear receptor superfamily which is highly expressed in liver (20). *PPARA* (*PPAR*) is a master regulator of lipid metabolism, modulating hepatic fatty acid transport and beta-oxidation (21). In animal studies, a protective effect preventing steatosis has been observed. This was postulated to be via the inhibition of hepatic lipid and lipoperoxide accumulation (22) although these findings may not extend to humans (23). ATP-binding cassette transporters G5 and G8 (*ABCG5*/*ABCG5*)/*ABCG8*(*ABCG8*)) are required for efficient cholesterol trafficking. In liver cells they mediate hepatic cholesterol and plant sterol excretion into bile (24). Upregulation of transgenic murine *ABCG5/ABCG8* leads to reduced absorption of dietary cholesterol and increased cholesterol efflux into bile. *NR1H4* (*FXR*) is reported to be a positive regulator of the reverse cholesterol transport pathway, known to promote increased biliary disposal of peripheral hepatic cholesterol into fecal matter via (*ABCG5*(*ABCG5*)/*ABCG8*(*ABCG8*)) (25). Downregulation of

*NR1H4* (FXR) in NAFLD can lead to lower (*ABCG5* (*ABCG5*)/*ABCG8* (*ABCG8*)) expression, possibly resulting in cholesterol accumulation as seen in *ABCG5/ABCG8* deficient mice (26). As a novel result, *NR1H4* (FXR) T183M has a prominent effect on SHBG, which both contributes to adverse fatty liver effects (27) and may be a key mediator of the link between fatty liver and insulin resistance (28). The carrier protein fetuin B *FETUB* (*FETUB*) is synthesized in the liver and in human hepatocytes. Treatment with FXR-agonists has been shown to increase *FETUB* (*FETUB*) expression (29). Taken together, our results indicated that liver fat and serum lipids share at least some genetic causes; however, the underlying biological roles are not entirely understood.

In this work we show that the FXR/RXR activation pathway is enriched for genetic associations with both serum lipids and liver attenuation. Increased LDL with a decrease in liver fat has been observed with the FXR agonist obeticholic acid (30). There is currently substantial debate surrounding whether this is an on target or off target effect or ascribed to the drug at all. Our data shows that a missense variant in *NR1H4* (FXR; T183M) reduces liver enzyme concentrations and increases serum LDL supporting an on-target effect; although the functional impact, i.e., gain or loss of function, was not directly assessed. Further support is seen with oral administration of chenodeoxycholic acid, a bile acid and natural FXR/RXR ligand, which also increases plasma LDL (31) further supporting that FXR agonists may have an LDL increasing effect. In animal models, chenodeoxycholic acid (31) and obeticholic acid (32) reduce hepatic triglyceride content but show worsening serum lipid profiles as a side-effect (33). This effect is opposite of what is seen epidemiologic studies where NAFLD correlates with higher serum LDL, triglycerides, and liver function tests (ALT/AST) and lower HDL. Several variants in the genes in the FXR/RXR pathway have similar atypical associations with reduced levels of liver enzymes and increased LDL levels. This includes *ABCC2* (*MRP2*) (V1188E, C1515Y), *ABCC5* (G27A) and *ABCG5/8* (*ABCG5/8*) (C50R/H19D). This suggests that global targeting of these genes to reduce liver fat may cause undesirable and inadvertent increases in LDL.

In contrast, variants in several other genes including *PPARA* (*PPAR*) (A286V) and *NR0B2* (*SHP*) (G171A) simultaneously decrease liver function tests and serum LDL. Treatment with PPAR-alpha agonists, including fibrates or selective agonists, reduce LDL as well as serum triglycerides without increasing liver fat in mice (34) and humans (35) further supporting this congruent effect on serum lipids and NAFLD. *NR0B2* (*SHP*) interacts with *FOXO1* (*FOXO1A*) to cause FOXO1-mediated Glucose-6-Phosphatase transcription (36). Drugs targeting SHP do not yet exist; however, genetic analysis suggest SHP may be a promising target for simultaneous reductions in LDL and liver fat. Although we found one mutation in *MTTP* (*MTP*) (S688G), its

associations were not as severe as other reported mutations, e.g., reduced serum cholesterol and increased risk for NAFLD (37). Not surprisingly, targeting *MTTP* (MTP) with lomitapide, an inhibitor of *MTTP* (MTP), decreases serum LDL in familial hypercholesterolemia but increases liver fat as a side-effect (38). These data validate the use of genetics to predict both on target effects such as lowering cholesterol as well as on target side effects such as increasing liver fat.

Biologically there is support for the role of implicated credible genes and their effect on liver fat accumulation and serum lipids. *APOB* and *MTTP* play a role in the formation of chylomicrons in the intestine and VLDL in liver. Mutations that interfere with their function can lead to decreases in LDL and TG in serum and an increase in fat deposition in intestine and liver (39). *ABCG5/ABCG8* are present on the canalicular side of hepatocytes and can serve to excrete sterols including cholesterol into bile. Mutation of these proteins in mice (26) or humans leads to higher absorption of cholesterol from the intestine and higher serum LDL levels and cardiovascular disease risk (40). Overexpression of these proteins leads to increased cholesterol efflux into bile and cholelithiasis (41). *SLC4A2* (AE2) regulates the intracellular pH by exchanging intracellular bicarbonate for extracellular chloride ion in hepatocytes and cholangiocytes (42), which are involved in bile production. *MRP2* is also a canalicular transporter that transports organic ions, glutathione, and bile. Mutations of this transporter would also lead to accumulation of bile in hepatocytes and increase LDL in serum. *ILBP* regulates bile acid trafficking in enterohepatic circulation. Overall, the decreased activity of *APOB*, *MTTP*, *ABCG5/ABCG8*, and *MRP2* prevent normal efflux of lipids, cholesterol, or bile from liver, and for some intestine, and can lead to damage of these organs.

In summary, we identified the FXR/RXR pathway as playing a role in serum lipids and NAFLD. We identified missense variants in *APOB* (*APOB*), *MTTP* (MTP), *ABCG5/ABCG8* (*ABCG5/ABCG8*), *ABCC2* (*MRP2*), *FABP6* (*ILBP*), *SLC4A2* (AE2), *NR0B2* (SHP), *NR1H4* (FXR), and *PPARA* (PPAR) that affect serum lipids or liver ALT, consistent with the role of these genes in lipid and liver processes. We showed that a missense variant in *NR1H4* (FXR) that associated with liver ALT and serum LDL cholesterol. While the exact molecular mechanism remains to be determined, this effect is supported by additional biomarkers, e.g., AST, GGT and HDL. We have also shown that missense variants in *PPARA* (PPAR) and *NR0B2* (SHP) decrease ALT and serum LDL. In mice *PPARA* knock-out mice have increased liver steatosis and liver inflammation on various diets (reviewed in (43)). A *PPARA* agonist Wy-14643 was able to prevent hepatic steatosis and inflammation in mice on a methionine and choline deficient (MCD) diet but not in *PPARA* knock-out mice suggesting an on-target effect of increasing fatty acid oxidation. (44) Studies in humans have not been done but our data would suggest that

altering *PPARA* (PPAR) function with the missense variant, e.g., V268A, would decrease liver fat, serum LDL and serum TG. In support of this, fibrates in humans are weak PPAR agonists and have been shown to decrease both serum LDL, serum TG and liver fat (45), thus supporting this as a target for NAFLD and combined serum hyperlipidemia. Our data also shows that *NR0B2* (SHP) A171G associates with liver fat, serum LDL and serum TG. The mechanism of how affecting *NR0B2* (SHP) molecularly results in these lipid effects is not known. One intriguing possibility is that SHP has been found to negatively regulate PPAR in cardiomyocytes (46); therefore, its loss would result in increased PPAR activity which, as noted above, results in decreased liver fat and serum LDL cholesterol.

While the current study was well-powered using the largest set of existing studies for the traits of interest, it is not without limitations. Among these, there remains uncertainty surrounding the causal genes and mechanisms as they relate to NAFLD. This is partially attributed to the reliance of MAGENTA on proximity, providing the necessary *enrichment* of causal genes needed for the analysis. Further, this study relies on a range of liver function traits, including liver attenuation and liver enzymes. These measures, which represent noninvasive proxies, are imperfect indicators of underlying liver disease, especially in the setting of excess liver fat. Moreover, there is always concordance for measures of liver fat risk and changes in liver function tests. Finally, these pathways should not be considered in isolation. Results from animal models (47) of bile acids suggest NAFLD resolution are contradictory and implicate additional processes (48, 49) not addressed here. Continued deep phenotyping (50) and characterization of the epidemiological relationships between liver function tests and liver fat accumulation are needed. Given the growing population burden of fatty liver disease, further mechanistic characterization of the bile acid pathway in fatty liver and dyslipidemia could reap tremendous rewards in advancing health.

Pathway-level associations provide insight into disease etiology and contributory processes. We used a gene set enrichment approach to highlight pathways that affect related traits, i.e., serum lipids and liver attenuation, with some distinct features. We identified genes involved in FXR/RXR activation with effects on both serum lipids and NAFLD in humans. In many cases, non-synonymous variants in the pathway were found to have strong associations with both serum lipids, especially LDL, and liver function tests. The associations of pathway members on serum lipids and liver function tests are diverse suggesting that not all members of the pathway, if affected, will impact the same traits. This work highlights both promises and pitfalls in the use of genetic data to explore potential side effects from therapies targeting individual genes. Our

observations are consistent with the variation seen in humans and further support disease treatment and prevention considering individual variability, i.e., precision medicine.

## REFERENCES

1. Lazarus JV, Mark HE, Anstee QM, Arab JP, Batterham RL, Castera L, Cortez-Pinto H, et al. Advancing the global public health agenda for NAFLD: a consensus statement. *Nat Rev Gastroenterol Hepatol* 2022;19:60-78.
2. Wilson PWF, D'Agostino RB, Parise H, Sullivan L, Meigs JB. Metabolic Syndrome as a Precursor of Cardiovascular Disease and Type 2 Diabetes Mellitus. *Circulation* 2005;112:3066-3072.
3. Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, Gudnason V, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet* 2011;7:e1001324.
4. Segre AV, Consortium D, investigators M, Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* 2010;6.
5. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;466:707-713.
6. Palmer ND, Musani SK, Yerges-Armstrong LM, Feitosa MF, Bielak LF, Hernaez R, Kahali B, et al. Characterization of European ancestry nonalcoholic fatty liver disease-associated variants in individuals of African and Hispanic descent. *Hepatology* 2013;58:966-975.
7. Iwasaki M, Takada Y, Hayashi M, Minamiguchi S, Haga H, Maetani Y, Fujii K, et al. Noninvasive evaluation of graft steatosis in living donor liver transplantation. *Transplantation* 2004;78:1501-1505.
8. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Lango Allen H, et al. Association analysis of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010;42:937-948.
9. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, Thorleifsson G, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* 2010;42:949-960.
10. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011;478:103-109.
11. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779.
12. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
13. O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* 2018;50:1728-1734.
14. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, Hemani G, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 2017;33:272-279.
15. Gluud C. Testosterone and alcoholic cirrhosis. *Epidemiologic, pathophysiologic and therapeutic studies in men. Dan Med Bull* 1988;35:564-575.

16. Carlson CS, Matise TC, North KE, Haiman CA, Fesinmeyer MD, Buyske S, Schumacher FR, et al. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol* 2013;11:e1001661.
17. Ding K, Kullo IJ. Geographic differences in allele frequencies of susceptibility SNPs for cardiovascular disease. *BMC Med Genet* 2011;12:55.
18. Priest C, Tontonoz P. Inter-organ cross-talk in metabolic syndrome. *Nat Metab* 2019;1:1177-1188.
19. Higuchi N, Kato M, Tanaka M, Miyazaki M, Takao S, Kohjima M, Kotoh K, et al. Effects of insulin resistance and hepatic lipid accumulation on hepatic mRNA expression levels of apoB, MTP and L-FABP in non-alcoholic fatty liver disease. *Exp Ther Med* 2011;2:1077-1081.
20. Yamakawa-Kobayashi K, Ishiguro H, Arinami T, Miyazaki R, Hamaguchi H. A Val227Ala polymorphism in the peroxisome proliferator activated receptor alpha (PPARalpha) gene is associated with variations in serum lipid levels. *J Med Genet* 2002;39:189-191.
21. Lefebvre P, Chinetti G, Fruchart JC, Staels B. Sorting out the roles of PPAR alpha in energy metabolism and vascular homeostasis. *J Clin Invest* 2006;116:571-580.
22. Ip E, Farrell GC, Robertson G, Hall P, Kirsch R, Leclercq I. Central role of PPARalpha-dependent hepatic lipid turnover in dietary steatohepatitis in mice. *Hepatology* 2003;38:123-132.
23. Fernandez-Miranda C, Perez-Carreras M, Colina F, Lopez-Alonso G, Vargas C, Solis-Herruzo JA. A pilot trial of fenofibrate for the treatment of non-alcoholic fatty liver disease. *Dig Liver Dis* 2008;40:200-205.
24. Graf GA, Yu L, Li WP, Gerard R, Tuma PL, Cohen JC, Hobbs HH. ABCG5 and ABCG8 are obligate heterodimers for protein trafficking and biliary cholesterol excretion. *J Biol Chem* 2003;278:48275-48282.
25. Escola-Gil JC, Llaverias G, Julve J, Jauhiainen M, Mendez-Gonzalez J, Blanco-Vaca F. The cholesterol content of Western diets plays a major role in the paradoxical increase in high-density lipoprotein cholesterol and upregulates the macrophage reverse cholesterol transport pathway. *Arterioscler Thromb Vasc Biol* 2011;31:2493-2499.
26. Yu L, Hammer RE, Li-Hawkins J, Von Bergmann K, Lutjohann D, Cohen JC, Hobbs HH. Disruption of *Abcg5* and *Abcg8* in mice reveals their crucial role in biliary cholesterol secretion. *Proc Natl Acad Sci U S A* 2002;99:16237-16242.
27. Gluud C, Bennett P, Svenstrup B, Micic S. Effect of oral testosterone treatment on serum concentrations of sex steroids gonadotrophins and prolactin in alcoholic cirrhotic men. Copenhagen Study Group for Liver Diseases. *Aliment Pharmacol Ther* 1988;2:119-128.
28. Hua X, Li M, Pan F, Xiao Y, Cui W, Hu Y. Non-alcoholic fatty liver disease is an influencing factor for the association of SHBG with metabolic syndrome in diabetes patients. *Sci Rep* 2017;7:14532.
29. Murakami T, Walczak R, Caron S, Duhem C, Vidal V, Darteil R, Staels B. The farnesoid X receptor induces fetuin-B gene expression in human hepatocytes. *Biochem J* 2007;407:461-469.
30. Siddiqui MS, Van Natta ML, Connelly MA, Vuppalanchi R, Neuschwander-Tetri BA, Tonascia J, Guy C, et al. Impact of obeticholic acid on the lipoprotein profile in patients with non-alcoholic steatohepatitis. *J Hepatol* 2020;72:25-33.
31. Ghosh Laskar M, Eriksson M, Rudling M, Angelin B. Treatment with the natural FXR agonist chenodeoxycholic acid reduces clearance of plasma LDL whilst decreasing circulating PCSK9, lipoprotein(a) and apolipoprotein C-III. *J Intern Med* 2017;281:575-585.
32. Briand F, Brousseau E, Quinsat M, Burcelin R, Sulpice T. Obeticholic acid raises LDL-cholesterol and reduces HDL-cholesterol in the Diet-Induced NASH (DIN) hamster model. *Eur J Pharmacol* 2018;818:449-456.
33. Pencek R, Marmon T, Roth JD, Liberman A, Hooshmand-Rad R, Young MA. Effects of obeticholic acid on lipoprotein metabolism in healthy volunteers. *Diabetes Obes Metab* 2016;18:936-940.

34. Takei K, Han SI, Murayama Y, Satoh A, Oikawa F, Ohno H, Osaki Y, et al. Selective peroxisome proliferator-activated receptor-alpha modulator K-877 efficiently activates the peroxisome proliferator-activated receptor-alpha pathway and improves lipid metabolism in mice. *J Diabetes Investig* 2017;8:446-452.
35. Kostapanos MS, Kei A, Elisaf MS. Current role of fenofibrate in the prevention and management of non-alcoholic fatty liver disease. *World J Hepatol* 2013;5:470-478.
36. Zhang Y, Hagedorn CH, Wang L. Role of nuclear receptor SHP in metabolism and cancer. *Biochim Biophys Acta* 2011;1812:893-908.
37. Hsiao PJ, Lee MY, Wang YT, Jiang HJ, Lin PC, Yang YH, Kuo KK. MTP-297H polymorphism reduced serum cholesterol but increased risk of non-alcoholic fatty liver disease—a cross-sectional study. *BMC Med Genet* 2015;16:93.
38. Blom DJ, Aversa MR, Meagher EA, du Toit Theron H, Sirtori CR, Hegele RA, Shah PK, et al. Long-Term Efficacy and Safety of the Microsomal Triglyceride Transfer Protein Inhibitor Lomitapide in Patients With Homozygous Familial Hypercholesterolemia. *Circulation* 2017;136:332-335.
39. Feingold KR. *Introduction to Lipids and Lipoproteins*. South Dartmouth, MA: Endotext, 2021.
40. Yu XH, Qian K, Jiang N, Zheng XL, Cayabyab FS, Tang CK. ABCG5/ABCG8 in cholesterol excretion and atherosclerosis. *Clin Chim Acta* 2014;428:82-88.
41. Yu L, Li-Hawkins J, Hammer RE, Berge KE, Horton JD, Cohen JC, Hobbs HH. Overexpression of ABCG5 and ABCG8 promotes biliary cholesterol secretion and reduces fractional absorption of dietary cholesterol. *J Clin Invest* 2002;110:671-680.
42. Banales JM, Prieto J, Medina JF. Cholangiocyte anion exchange and biliary bicarbonate excretion. *World J Gastroenterol* 2006;12:3496-3511.
43. Liss KH, Finck BN. PPARs and nonalcoholic fatty liver disease. *Biochimie* 2017;136:65-74.
44. Ip E, Farrell G, Hall P, Robertson G, Leclercq I. Administration of the potent PPARalpha agonist, Wy-14,643, reverses nutritional fibrosis and steatohepatitis in mice. *Hepatology* 2004;39:1286-1296.
45. Shipman KE, Strange RC, Ramachandran S. Use of fibrates in the metabolic syndrome: A review. *World J Diabetes* 2016;7:74-88.
46. Rodriguez-Calvo R, Chanda D, Oligschlaeger Y, Miglianico M, Coumans WA, Barroso E, Tajés M, et al. Small heterodimer partner (SHP) contributes to insulin resistance in cardiomyocytes. *Biochim Biophys Acta Mol Cell Biol Lipids* 2017;1862:541-551.
47. Postic C, Girard J. The role of the lipogenic pathway in the development of hepatic steatosis. *Diabetes Metab* 2008;34:643-648.
48. Jiao N, Baker SS, Chapa-Rodriguez A, Liu W, Nugent CA, Tsompana M, Mastrandrea L, et al. Suppressed hepatic bile acid signalling despite elevated production of primary and secondary bile acids in NAFLD. *Gut* 2018;67:1881-1891.
49. Distrutti E, Santucci L, Cipriani S, Renga B, Schiaroli E, Ricci P, Donini A, et al. Bile acid activated receptors are targets for regulation of integrity of gastrointestinal mucosa. *J Gastroenterol* 2015;50:707-719.
50. Kettunen J, Demirkan A, Wurtz P, Draisma HH, Haller T, Rawal R, Vaarhorst A, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016;7:11122.

## FIGURE LEGENDS

**Figure 1.** Study Design of the two-stage pathway analysis.



Abbreviations: GWAS - Genome-wide association study; FXR/RXR - farnesoid X receptor/retinoid X receptor; MAGENTA - Meta-Analysis Gene-set Enrichment of variant Associations; HDL - high-density lipoprotein; LDL - low-density lipoprotein; TG – triglycerides; NAFLD - Non-alcoholic fatty liver disease.

**Figure 2. Summary of gene-wise associations with liver attenuation.** Each point corresponds to a single gene, with a gene-wise p-value on each axis (log scale) in either the African or European subgroups of GOLD. Filled circles represent credible genes, which are labelled with gene names. The bars represent the MAGENTA enrichment cutoff.

Abbreviations: GOLD - Genetics of Liver Disease; MAGENTA - Meta-Analysis Gene-set Enrichment of variant Associations; *ABCC2* (ATP Binding Cassette Subfamily C Member 2; MRP2 (Multidrug resistance-associated protein 2); *ABCG5* and *ABCG8* (ATP-binding cassette (ABC) transporters G5 and G8); *APOB* (Apolipoprotein B); *FABP6* - Fatty Acid Binding Protein 6; ILBP - Ileal lipid-binding protein; *MTTP* (MTP) - microsomal triglyceride transfer protein; *NR0B2* - Nuclear Receptor Subfamily 0 Group B Member 2; SHP - Small heterodimer partner; *NR1H4* - Nuclear Receptor Subfamily 1 Group H Member 4; FXR - farnesoid X receptor; *PPARA* (PPAR) - Peroxisome Proliferator Activated Receptor Alpha; *SLC4A2* - solute carrier family 4 member 2; AE2- Anion exchange protein 2; *FETUB* (FETUB) - Fetuin B; *FOXO1* (FOXO1A) - Forkhead Box O1.

**Figure 3. Summary of physiological role of credible genes.** This graphic summarizes the well characterized role of each gene/protein (see text for additional detail.) An intestinal enterocyte is shown in beige on the left with the lumen at the top of the figure; hepatocytes in blue are shown on the right. In both cells, cytosolic and nuclear proteins are shown in purple symbols while membrane-associated proteins are shown in brown symbols.

Abbreviations: ALP - Alkaline phosphatase; ALT - alanine transaminase; AST - aspartate aminotransferase; GGT - gamma-glutamyl transferase; DBili - direct bilirubin; HDL - high-density lipoprotein; LDL - low-density lipoprotein; TG - triglycerides; MRP2 - Multidrug resistance-associated protein 2; *ABCG5* and *ABCG8* (ATP-binding cassette (ABC) transporters G5 and G8); *APOB* - Apolipoprotein B; ILBP - ileal lipid-binding protein; MTP - microsomal triglyceride transfer protein; SHP - small heterodimer partner; FXR - farnesoid X receptor; PPAR - Peroxisome Proliferator Activated Receptor Alpha; AE2- Anion exchange protein 2; VLDL - very low density lipoprotein; FETUB - Fetuin B; FOXO1A - Forkhead Box O1.

**Figure 4: UK Biobank laboratory associations for non-synonymous variants in credible genes.** Each row corresponds to a gene and non-synonymous variant (or a pair of variants in the case of high linkage disequilibrium ( $r^2 > 0.5$ ); one haplotype includes amino acid substitutions in both *ABCG5* and *ABCG8*. The additional columns provide an inverse-normal-transformed effect size (expressed as upward red (increased % of a standard deviation) and downward blue

(decreased % of a standard deviation) arrows) for individual traits. All the effect sizes are oriented towards the LDL-increasing allele (see ST7 for additional details).

Abbreviations: ALP - Alkaline phosphatase; ALT - alanine transaminase; AST - aspartate aminotransferase; GGT - gamma-glutamyl transferase; DBili - direct bilirubin; HDL - high-density lipoprotein; LDL - low-density lipoprotein; TG - triglycerides; *ABCC2* (ATP Binding Cassette Subfamily C Member 2; *MRP2* (Multidrug resistance-associated protein 2) ; *ABCG5* and *ABCG8* (*ABCG5* and *ABCG8*, ATP-binding cassette (ABC) transporters G5 and G8); *APOB* (*APOB*) - Apolipoprotein B; *FABP6* - Fatty Acid Binding Protein 6; ILBP - ileal lipid-binding protein; *MTTP* (*MTP*) - microsomal triglyceride transfer protein; *NR0B2* - Nuclear Receptor Subfamily 0 Group B Member 2; SHP - Small heterodimer partner; *NR1H4* - Nuclear Receptor Subfamily 1 Group H Member 4; FXR - farnesoid X receptor; *PPARA* (*PPAR*) - Peroxisome Proliferator Activated Receptor Alpha; *SLC4A2* - solute carrier family 4 member 2; AE2- Anion exchange protein 2; MAC - median minor allele count; FDR - False Discovery Rate.

## TABLE LEGENDS

**Table 1. Lipid-related gene sets enriched for NAFLD.** For each gene set the database source, effective number of genes in each gene set, the expected number of genes above the threshold (roughly 25% of the effective size) and the actual number of genes above the threshold within the pathway is provided. Statistical significance includes the nominal enrichment p-value and the FDR-adjusted p-value.

Abbreviations: NAFLD - Non-alcoholic fatty liver disease; FXR/RXR - farnesoid X receptor/retinoid X receptor; FDR - False Discovery Rate.

**Table 2. Lead variants in credible genes that significantly increase liver steatosis in the GOLD cohort.** For each credible gene, the gene symbol (protein symbol), terminal process in the Ingenuity subpathway and relative credibility defined as the Bayes Factor vs. *APOE* (the median gene in the FXR/RXR pathway) is provided. The lead variant reported by MAGENTA includes the effect allele (lower liver density, higher liver fat). Effect sizes were significant at FDR < 0.1.

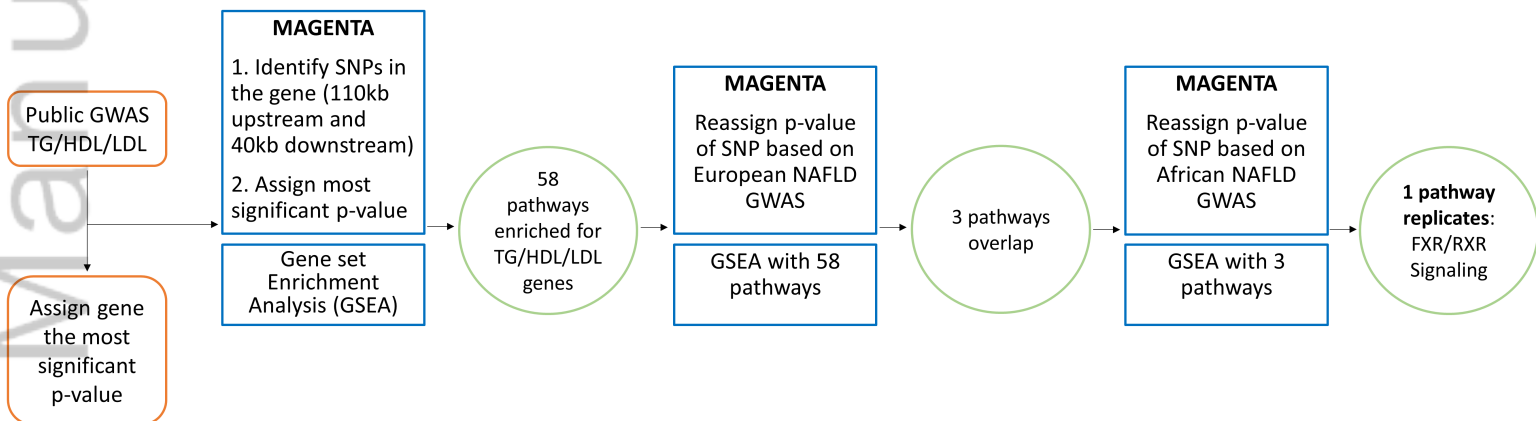
Abbreviations: GOLD - Genetics of Liver Disease; MAGENTA - Meta-Analysis Gene-set Enrichment of variant Associations; *APOE* - Apolipoprotein E; *ABCC2* (ATP Binding Cassette Subfamily C Member 2; *MRP2* (Multidrug resistance-associated protein 2) ; *ABCG5* and *ABCG8* (*ABCG5* and *ABCG8*, ATP-binding cassette (ABC) transporters G5 and G8); *APOB* (*APOB*) - Apolipoprotein B; *FABP6* - Fatty Acid Binding Protein 6; ILBP - ileal lipid-binding protein; *MTTP* (*MTP*) - microsomal triglyceride transfer protein; *NR0B2* - Nuclear Receptor Subfamily 0 Group B Member 2; SHP - Small heterodimer partner; *NR1H4* - Nuclear Receptor Subfamily 1 Group H Member 4; FXR - farnesoid X receptor; *PPARA* (*PPAR*) - Peroxisome Proliferator Activated Receptor Alpha; *SLC4A2* - solute carrier family 4 member 2; AE2- Anion exchange protein 2; *FETUB* (*FETUB*) - Fetuin B; *FOXO1* (*FOXO1A*) - Forkhead Box O1.

Table 1. Lipid-related gene sets enriched for NAFLD.

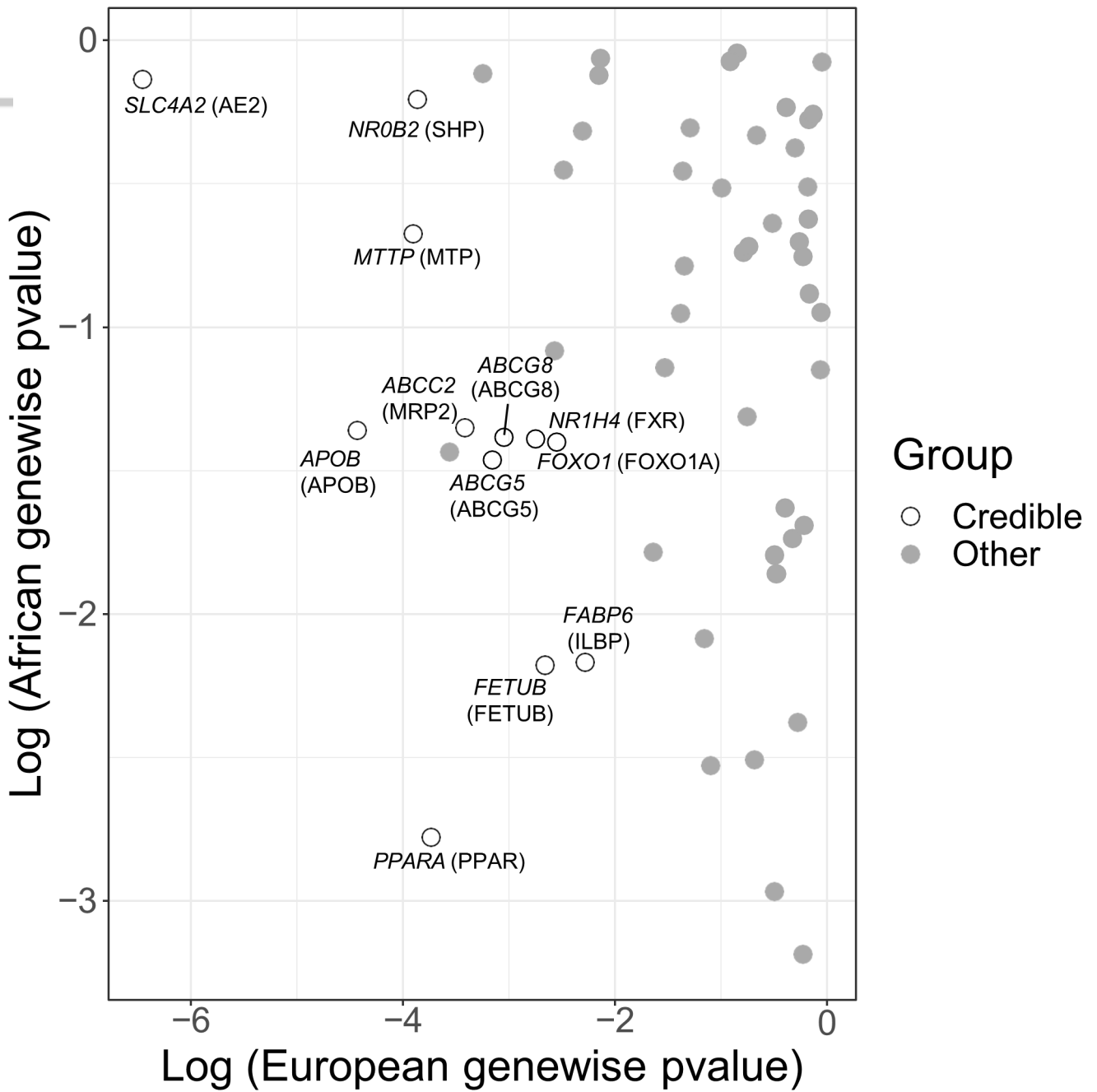
Database	Gene Set	Effective Gene Set Size	Expected (# Genes >75 <sup>th</sup> %)	# Genes >75 <sup>th</sup> %	Nominal Enrichment P-value	FDR-adjusted Enrichment P-value
<b><i>European Ancestry (N=7,176)</i></b>						
Ingenuity	Hepatic Cholestasis	58	15	24	0.0046	0.054
Ingenuity	FXR/RXR Activation	54	14	23	0.0036	0.062
Reactome	Chylomicron Mediated Lipid Transport	13	3	8	0.0057	0.076
<b><i>African Ancestry (N=3,124)</i></b>						
Ingenuity	FXR/RXR Activation	56	14	23	0.0058	0.0089

**Table 2. Lead variants in credible genes that significantly increase liver steatosis in the GOLD cohort.**

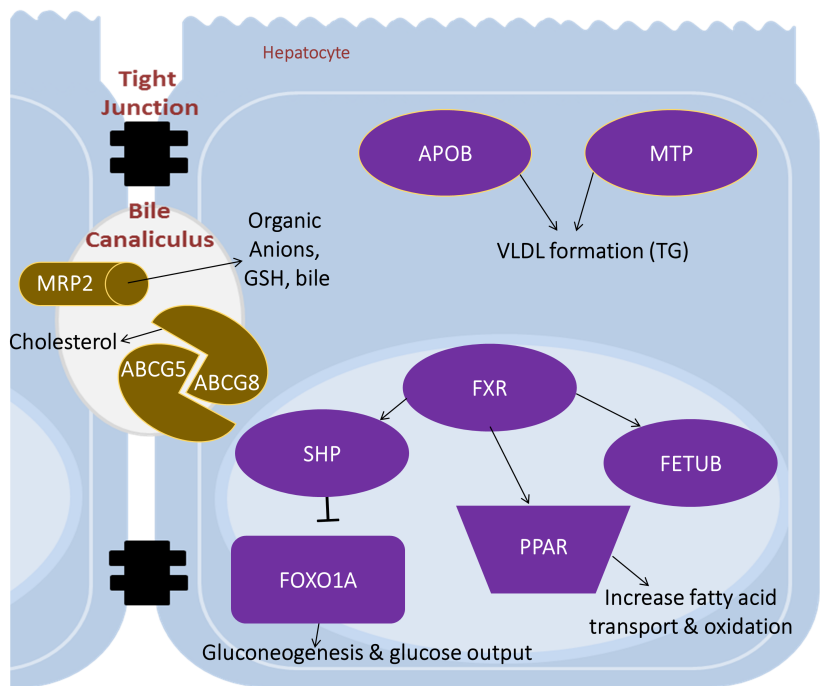
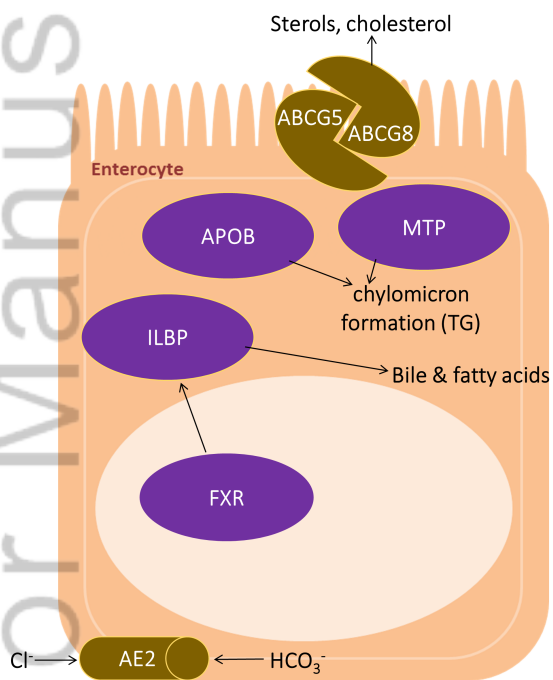
<b>Gene (Protein)</b>	<b>Ingenuity Subpathway</b>	<b>Relative Credibility</b>	<b>Variant-effect allele</b>
<i>ABCC2</i> (MRP2)	Bile Acid Secretion	8	rs10883407-A
<i>ABCG5, ABCG8</i> (ABCG5, ABCG8)	Cholesterol Secretion	7,6	rs17031754-T
<i>APOB</i> (APOB)	Cholesterol Secretion	19	rs478588-A
<i>FABP6</i> (ILBP)	Bile Acid Absorption	6	rs2546379-G
<i>FETUB</i> (FETUB)	Tumor suppression	8	rs2889755-T
<i>FOXO1</i> (FOXO1A)	Gluconeogenesis	4	rs2755212-C
<i>MTTP</i> (MTP)	Cholesterol secretion	7	rs12645746-C
<i>NR0B2</i> (SHP)	Gluconeogenesis; hepatocytic uptake	4	rs6658653-C
<i>NR1H4</i> (FXR)	Bile Acid Absorption, Synthesis	5	rs12297245-A
<i>PPARA</i> (PPAR)	Bile Acid synthesis; Fatty acid transport and oxidation	36	rs5768754-A
<i>SLC4A2</i> (AE2)	Bicarbonate secretion	39	rs2373929-G



HEP4\_2066\_Figure1\_071522.tif



HEP4\_2066\_Figure2\_071522.tif



HEP4\_2066\_Figure3\_071522.tif

Gene (Protein)	Variant	ALP	ALT	AST	GGT	DBili	HDL	LDL	TG
<b>ABCC2 (MRP2)</b>	V1188E		↓	↓	↓	↓		↑	
	C1515Y		↓	↓	↓	↓			
	S921G		↓	↓	↓	↓			
<b>ABCG5 (ABCG5)</b>	G27A		↓	↓				↑	
<b>ABCG5/ABCG8 (ABCG5/8)</b>	C50R H19D	↓	↓	↓		↓		↑	↑
<b>ABCG8 (ABCG8)</b>	K400T					↓		↑	
<b>APOB (APOB)</b>	T98I	↓				↓	↓	↑	↑
	R1923H					↓	↓	↑	↑
	L2739P S4338N	↓		↑			↓	↑	↑
<b>FABP6 (ILBP)</b>	M124I						↑	↑	
<b>MTTP (MTP)</b>	I155T		↑	↑				↑	↑
	S688G		↑					↑	
<b>NROB2 (SHP)</b>	G171A	↓	↑		↓		↓	↑	↑
<b>NR1H4 (FXR)</b>	T183M	↑	↓	↓	↓		↓	↑	
<b>PPARA (PPAR)</b>	A268V	↑	↑					↑	↑
<b>SLC4A2 (AE2)</b>	G26E			↑	↑	↑	↓		↑
	R311Q					↑			

HEP4\_2066\_Figure4\_071522.tif