

# Genome-wide Analysis Identifies Novel Gallstone-susceptibility Loci Including Genes Regulating Gastrointestinal Motility

Cameron J Fairfield<sup>1</sup>, Thomas M Drake<sup>1</sup>, Riinu Pius<sup>1</sup>, Andrew D Bretherick<sup>2</sup>, Archie Campbell<sup>13,4</sup>, David W Clark<sup>5</sup>, Jonathan A Fallowfield<sup>6</sup>, Caroline Hayward<sup>2</sup>, Neil C Henderson<sup>6</sup>, Andrii Iakovliev<sup>7</sup>, Peter K Joshi<sup>5</sup>, Nicholas L Mills<sup>8</sup>, David J Porteous<sup>3</sup>, Prakash Ramachandran<sup>6</sup>, Robert K Semple<sup>8</sup>, Catherine A Shaw<sup>1</sup>, Cathie LM Sudlow<sup>1</sup>, Paul RHJ Timmers<sup>2,5</sup>, James F Wilson<sup>2,5</sup>, Stephen J Wigmore<sup>9</sup>, Athina Spiliopoulou<sup>7\*</sup>, Ewen M Harrison<sup>1\*</sup>

---

<sup>1</sup> Centre for Medical Informatics, Usher Institute, University of Edinburgh

<sup>2</sup> MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh

<sup>3</sup> Centre for Genomic and Experimental Medicine, Institute of Genetics & Cancer, University of Edinburgh

<sup>4</sup> Health Data Research UK, University of Edinburgh

<sup>5</sup> Centre for Global Health Research, Usher Institute, University of Edinburgh

<sup>6</sup> Centre for Inflammation Research, Queen's Medical Research Institute, University of Edinburgh

<sup>7</sup> Centre for Population Health Sciences, Usher Institute, University of Edinburgh

<sup>8</sup> Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh

<sup>9</sup> Department of Clinical Surgery, Division of Health Sciences, University of Edinburgh

\*EMH and AS are joint senior authors. EMH is the corresponding author

## Abstract

### Objective

Genome-wide association studies (GWAS) have identified several risk loci for gallstone disease. As with most polygenic traits, it is likely many genetic determinants are undiscovered. The aim of this study was to identify novel genetic variants that represent new targets for gallstone research and treatment.

### Design

We performed a GWAS of 28,627 gallstone cases and 348,373 controls in the UK Biobank and a GWA meta-analysis (43,639 cases and 506,798 controls) with the FinnGen cohort. We assessed pathway enrichment using gene-based then gene-set analysis and tissue expression of implicated genes in Genotype-Tissue Expression project data. We constructed a polygenic risk score (PRS) and evaluated phenotypic traits associated with the score.

### Results

Seventy-five risk loci were identified ( $P < 5 \times 10^{-8}$ ) of which forty-six were novel. Pathway enrichment revealed associations with lipid homeostasis, glucuronidation, phospholipid metabolism and gastrointestinal motility. *ANO1* and *TMEM147*, both in novel loci, are strongly expressed in the gallbladder and gastrointestinal tract. Both regulate gastrointestinal motility. The gallstone risk allele rs7599-A leads to suppression of hepatic *TMEM147* expression suggesting the protein protects against gallstone formation. Individuals in the highest decile of the PRS demonstrated a 6-fold increased risk of gallstones compared to the lowest risk category. The PRS was strongly associated with increased body mass index, serum liver enzyme and C-reactive protein concentrations and decreased lipoprotein cholesterol concentrations.

### Conclusion

This GWAS demonstrates the polygenic nature of gallstone risk and identifies 46 novel susceptibility loci. For the first time, we implicate genes influencing gastrointestinal motility in the pathogenesis of gallstones.

## Summary Box

### What is already known on this subject?

- Genome-wide association studies (GWAS) have identified 29 genetic variants within independent loci which increase the risk of gallstone disease.
- Most of these variants lie within or near to genes that regulate lipid or bile acid metabolism. Gallstones are known to have a significant genetic component with 25-50% of gallstone disease due to genetic risk factors.
- Much of this risk is not accounted for by the known gallstone-susceptibility loci.

### What are the new findings?

- We performed a GWAS in the UK Biobank (28,627 gallstone cases, 348,373 controls) and a GWA meta-analysis (43,639 cases and 506,798 controls) with the FinnGen cohort.
- We identified a total of 75 gallstone-susceptibility loci with 46 of these being new and the remaining 29 being those already identified.
- We annotated the variants based on their position within or near to genes and assessed pathway enrichment through gene-set analysis.
- We identify two novel gallstone-susceptibility loci in which the lead variants lie within genes governing gastrointestinal motility which highly expressed in gallbladder (*ANO1* and *TMEM147*).
- We demonstrate further loci involved in primary cilia function.
- We report significant association of a polygenic risk score with gallstones using independent subsets of the study population.

### How might it impact on clinical practice in the foreseeable future?

- The genes and pathways identified represent novel targets for development of medication targeting primary or secondary prevention of gallstones. This may be of particular benefit to those unable to undergo cholecystectomy.
- The individual variants or polygenic risk score identified in this GWAS could form the basis for identification of individuals with high risk of gallstones to support screening or treatment of gallstone disease. This screening may be of particular benefit in populations with elevated risk of gallstones such as haemolytic disease or bariatric surgery.

## Introduction

Gallstone disease is one of the most common reasons for acute presentation to hospitals worldwide<sup>1</sup> and generates the greatest healthcare expenditure of any gastrointestinal condition in the United States<sup>2</sup>. Prevalence of gallstones reaches 20-40% in high-risk populations<sup>3,4</sup> and gallstones are highly heritable<sup>5</sup>. Identification of novel pathways influencing gallstone development may improve risk stratification in high risk populations and support development of medications targeting gallstone formation and other cholestatic disease for use in primary and secondary prevention.

A total of seven genome-wide association studies (GWAS) have been published revealing 29 gallstone-promoting loci<sup>6-12</sup>. Most gallstones are composed of cholesterol which crystallises in bile at high concentration and/or in bile lacking protective bile acids<sup>13</sup>. Many of the identified variants relate to cholesterol and bile acid metabolism. Despite evidence of impaired gallbladder and gastrointestinal motility contributing to the formation of gallstones, there have been no gallstone-susceptibility variants identified that influence these characteristics<sup>14</sup>. Using larger and more richly phenotyped cohorts may identify such variants and contribute to the knowledge of gallstone formation.

The most recent GWAS and GWAS meta-analysis used publicly-available summary GWAS data from the Global Biobank Engine (GBE)<sup>15</sup> based on UK Biobank (UKB) data<sup>16</sup>. GBE is based on three-level ICD-10 and ICD-9 codes and considered gallstones to be code “K80” (“cholelithiasis”) or self-reported gallstones at study enrolment. The UKB captures several other informative variables including OPCS codes, other ICD-9 and 10 codes, self-reported procedures relating to gallstones and primary care data. By incorporating these additional data, a greater proportion of patients with gallstone disease can be classified as cases rather than misclassified as controls.

The aim of this study was to perform an updated GWAS using more richly phenotyped participants to identify novel pathways contributing to gallstone formation. The genetic associations identified were replicated in the Generation Scotland: Scottish Family Health Study (GS-SFHS) cohort<sup>17</sup> and meta-analysed with FinnGen summary association statistics<sup>18</sup>. A polygenic risk score was derived and analysis for gene-environment interaction undertaken.

## Materials and Methods

Methods for identification of participants, determination of gallstone status and genotyping for the UKB, GS-SFHS and the FinnGen cohort are provided in Supplementary Materials 1. Diagnostic codes used for the identification of gallstones are shown in Supplementary Tables 1-3. The UKB received ethical approval (Research Ethics Committee [REC] reference number: 11/NW/0382). UKB data access was approved under projects 30439 (phenotype data) and 19655 (genotype data). Ethical approval for GS-SFHS was granted by NHS Tayside REC (reference number: 05/S1401/89). The FinnGen study protocol was approved by the Ethical Review Board of the Hospital District of Helsinki and Uusimaa (Nr HUS/990/2017).

## Genome-wide Association Study

### Initial Analysis

Genotyped and imputed SNPs were analysed. Participants of self-reported European ancestry were included and outliers for heterozygosity and unexpected runs of homozygosity were excluded. One randomly-selected participant from each pair of related individuals in the UKB (Kinship > 0.0884<sup>19</sup>) was excluded.

Association with gallstones was analysed using logistic regression adjusted for age, sex, the first 20 genetic principal components and batch, with batch included as a random effect. Imputation dosage was used for imputed SNPs.

$$gallstones \sim age + sex + (1|batch) + pc1 + pc2 + \dots + pc20 + genotype$$

SNPs were excluded based on minor allele frequency (MAF < 0.001), imputation quality (INFO < 0.3) and departure from Hardy-Weinberg equilibrium (HWE:  $P < 5 * 10^{-6}$ ). The MAF was selected on the basis of ensuring 25 or more minor alleles in the cases ( $MAF = 25 / (2 * n_{cases})$ ) which corresponded to a MAF of 0.0004. This was rounded up to 0.001.

Genome-wide significance was determined as  $P < 5 * 10^{-8}$  in the UKB and Bonferroni-corrected  $P_C = 0.05$  in the replication cohorts (0.05 divided by the number of loci tested). Replication was undertaken by analysing the lead SNP within each locus or a linkage disequilibrium (LD) proxy.

P values smaller than  $4.94 * 10^{-324}$  (the limit of numerical precision on a Linux 64-bit operating system) were rounded to  $4.94 * 10^{-324}$  (Supplementary Materials 1).

### Genome-Wide Association Meta-analysis

A genome-wide association meta-analysis was undertaken using UKB data and FinnGen data in an inverse variance-weighted fixed effects model. SNPs shared between the datasets (10,044,759) were analysed. RsID was converted to the format “rs1234:A:T” to avoid ambiguity at multi-allelic SNPs.

## Linkage Disequilibrium Clumping and Conditional Analyses

LD clumping was performed using the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) web application version 1.3.5d<sup>20</sup>. Loci were established for each lead SNP with a minimum distance of 250Kb between loci and using an  $r^2 < 0.25$  to indicate separate independent SNPs within the same locus. Details of the FUMA parameters are available in Supplementary Materials 2A. Five loci with rare lead SNPs not present in the 1000 Genomes Phase 3 reference panel ( $MAF_{UKB} \leq 0.01-0.001$ ) were generated manually by creating a 500 KB window centred on the lead SNP.

Each locus was re-analysed whilst conditioning on the lead SNP and further signals with genome-wide significance were identified. This process was repeated with additional lead SNPs until no remaining SNPs reached genome-wide significance.

## Population Stratification

The genomic inflation factor  $\lambda_{gc}$  was calculated using the summary GWAS statistics. Evidence of test statistic inflation in  $\lambda_{gc}$  was investigated with linkage disequilibrium score regression<sup>21</sup>.

## Sensitivity Analysis

As the definition of gallstones was based on clinical detection of gallstones it is possible that detection bias may influence the results. Gallstones detected during investigations for other conditions may be recorded despite causing no symptoms. Genetic variants causing other hepatic pathology may therefore influence the likelihood of detection, rather than formation, of gallstones. Individuals requiring surgical management of gallstones are likely to be symptomatic, minimising detection bias. A sensitivity analysis in which cases underwent a cholecystectomy, or any gallstone-specific procedure as defined in Supplementary Table 2, and controls were free from gallstone disease was undertaken at each of the identified lead SNPs. Those with gallstones who did not undergo surgery were excluded.

## Association of Lithogenic Loci with Phenotypic Traits

Associations with serum biochemistry values for each identified variant were assessed in an age- and sex-adjusted linear regression model. Serum bilirubin (total, conjugated and unconjugated), alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transferase (GGT), alkaline phosphatase (ALP), total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides, apolipoprotein A and B (ApoA and ApoB), lipoprotein A (LipoA), haemoglobin (Hb), red blood cells (RBC), reticulocyte percentage, C-reactive protein (CRP) and Cystatin C were assessed. Each trait was assessed visually via histogram and log-transformed in the case of skewed distribution.

## Gene-Environment Interaction

The *UGT1A1* polymorphism rs4148325 associates with bilirubin and causes Gilbert's syndrome but is a comparatively modest driver of gallstones in GWAS<sup>12,22</sup>. In patients with haemolytic disease or cystic fibrosis, the variant is a significant risk factor for gallstones<sup>23-25</sup>. This raises the possibility of a gene-environment interaction in which defects of bilirubin metabolism only lead to gallstone formation in the context of increased bilirubin production or decreased bilirubin excretion. Each identified locus was re-analysed for a gene-environment interaction with serum unconjugated bilirubin as a covariate and as an interaction term with genotype. A Wald test was performed to determine the presence of a gene-environment interaction and significance was determined using a Bonferroni correction. This was not conducted at the known *UGT1A1* locus as unconjugated bilirubin is likely to be a mediator of the relationship between the variant and gallstones.

## Polygenic Risk Score

A polygenic risk score (PRS) was derived using a random 80%/20% split of the UKB cohort. A separate GWAS was conducted in the 80% and the summary associations statistics used to derive the PRS which was validated in the 20% split using PRSice<sup>26</sup>. This was conducted as an additive model with default PRSice parameters on directly genotyped SNPs. The PRS was evaluated for association with the same phenotypic traits as individual risk loci in a simple, unadjusted linear or logistic regression.

## Pathway Enrichment

We conducted a gene-based analysis in MAGMA to identify genes and gene-sets showing significant association with gallstones. Methods for gene-based and gene-set analysis have been described previously<sup>27</sup>. In brief, summary association statistics, following FUMA-based annotation, were grouped by gene and the average  $\chi^2$  statistic across all SNPs in each gene computed and converted to gene-based P values. Genes demonstrating significant association (after Bonferroni correction) were taken forward for a competitive gene-set analysis in which gene-sets were tested for association with the phenotype. This tests whether the mean association of the phenotype with genes in the gene set is greater than that of genes not in the gene set. The mean association was adjusted for gene size, gene density (based on genetic distance using between-SNP LD), the inverse of the mean minor allele frequency in the gene, and the log value of these three variables. This adjustment is necessary to prevent spurious associations driven by the number and frequency of possible mutations within a given gene rather than a true effect. 4,761 curated gene sets and 5,917 gene ontology (GO) terms were available for analysis from Molecular Signatures Database v6.2. Gene tissue expression analysis was undertaken in MAGMA using the identified genes. Loci without clear relation to gallstones were assessed by literature search based on the FUMA gene annotations for genes nearest to the lead SNP to identify further pathways not present or identified within the curated gene-sets. Lead SNPs were analysed in the Genotype-Tissue expression project (GTEx) to assess their effect on expression of putative causal genes.

## **Plotting and Statistical Analysis of Results**

Genomic analyses were performed using SNPtest (version 2.5.4 for gene-environment interaction model, version 2.5.2 for all other analyses) and QCTOOLS (version 2.0.6)<sup>28</sup> within a Linux high-performance compute cluster. GWA meta-analysis was undertaken in METAL<sup>29</sup>. Post-GWAS analysis, regression analyses and plotting were performed using R version 3.6.3<sup>30</sup>. Methods and results are reported in accordance with the STREGA guidelines<sup>31</sup> (Supplementary Materials 3).

## **Patient Involvement**

During the design stage of the project, the NHS Lothian Public and Patient Involvement and Engagement was involved in drafting lay summaries of the proposed research. Published results will be returned to the UK Biobank team who inform participants of research outputs.



## Results

### Gallstones Cohort

A total of 502,616 participants entered the UKB study of which 377,000 were taken forward for GWAS. A history of gallstones was present in 28,627 with 348,373 controls. The median follow-up available was 9.0 years (range 7.3- 11.9). Baseline demographics of the UKB cohort are shown in Supplementary Table 4.

A total of 24,096 participants entered GS-SFHS of which 6,317 unrelated individuals were taken forward for GWAS. A history of gallstones was present in 1,089 with 5,228 controls. The mean follow-up available was 10.7 years (range 9.6-14.7) (Figure 1).

In FinnGen 173,437 participants were analysed for gallstones (15,012 cases, 158,425 controls).

### Genome-wide Association Study

#### Initial UKB Analysis

A total of 4,065 SNPs were identified with genome-wide significant P values after removal of low-quality SNPs (see Supplementary Materials 4). An additional 11,539 SNPs demonstrated borderline significance ( $P < 5 * 10^{-5}$ ). LD-clumping revealed 50 loci with at least one significant signal (see Supplementary Materials 2B). Five additional loci contained variants with a significant association with gallstones for which the lead SNP was not present in the 1000 Genomes reference panel (MAF 0.01 to 0.001) and were therefore not annotated by the FUMA application. LD-clumping for those loci was performed using boundaries of 250KB from the lead SNP.

Of the 55 loci identified in the UKB, 27 loci are newly identified variants, 7 are known from previous GWAS using other datasets and 21 variants were previously identified only in studies using GBE data<sup>10,12</sup> (Table 1). Four variants were replicated in GS-SFHS (rs11887534, rs1993453, rs1260326 and rs62129966;  $P < 0.00091$ , corrected for 55 variants). 33 variants from the UKB GWAS reached the replication threshold in the FinnGen cohort ( $P < 0.00012$ , corrected for 42 variants available) and 38 reached nominal significance ( $P < 0.05$ ). 10 of the 27 novel loci identified in the UKB GWAS reached the replication threshold in FinnGen. Results for replication in GS-SFHS and FinnGen at each of the 55 loci identified in the UKB are shown in Supplementary Table 5. Loci with replication in either GS-SFHS or FinnGen are highlighted in bold in Table 1.

The D19H variant of *ABCG8* (rs11887534) demonstrated a highly significant result ( $P < 4.94 * 10^{-324}$ ). 23 adjacent variants demonstrated identical P values and almost identical  $\beta$ -coefficients. As rs11887534 is a known protein-altering variant (D19H) and previously identified as the lead SNP by other GWAS, this was considered as the lead SNP.

### GWA Meta-analysis

GWA meta-analysis with FinnGen data revealed a total of 63 loci associated with gallstone disease ( $P < 5 * 10^{-8}$ ). Of these, 43 were identified in the UKB, one was previously identified with GBE data<sup>10</sup> and a further 19 were novel loci bringing the total number of novel gallstone-susceptibility loci to 46 and the overall total of gallstone-susceptibility loci to 75. Of the 29 previously reported loci, 28 were identified in the UKB GWAS and all 29 were identified in the meta-analysis. Twelve loci identified in the UKB GWAS did not retain significance on meta-analysis. The 63 lead SNPs from the meta-analysis are shown in Table 2.

A combined circular Manhattan plot for association of variants with gallstones in the meta-analysis as well as the 12 loci only identified in the UKB is shown in Figure 2. Rectangular Manhattan plots for the meta-analysis and separately for the initial UKB analysis are available in Supplementary Figures 1 to 4.

### Conditional Analyses

Conditional analyses, adjusting for the lead SNP, were undertaken for the 55 loci identified in the UKB GWAS. After conditional analyses, a total of 6 loci were found to have additional independent signal suggesting that more than one variant within these loci has a causal relationship with gallstones. The remaining 49 loci did not have any additional SNPs reaching genome-wide significance ( $P < 5 * 10^{-8}$ ).

The *ABCG8* locus (rs11887534) contained 10 variants with independent and significant association with gallstones within *ABCG8*, *ABCG5* and *DYNC2LI1*. The *TM4SF4* locus (rs4681515) contained 2 variants, the *ABCB4* locus (rs7802555) contained 3 variants, the *SLC10A2* locus (rs144846334) contained 3 variants, the *FOXA3* locus (rs34255979) contained 2 variants and the *TNRC6B* locus (rs11089985) contained 2 variants with independent and significant association with gallstones. Full results of the conditional analyses are provided in Supplementary Materials 4. This study did not detect a second independent signal previously identified at the locus containing *TTC39B*<sup>9</sup> although a more stringent genome-wide significance threshold was used in this study ( $P = 5 * 10^{-8}$  versus  $P = 6.49 * 10^{-4}$ ).

### Population Stratification

After adjusting for the first 20 genetic principal components, there was evidence of test statistic inflation ( $\lambda_{gc} = 1.103$ ). Test statistic inflation is seen in traits caused by multiple genes (polygenicity) but can also be a feature of unmeasured population substructure in which genetically-similar individuals also share similar environmental exposures<sup>32</sup>. Linkage disequilibrium score regression determines the extent to which polygenicity drives the test statistic inflation. The linkage disequilibrium score regression intercept was

1.042 and the proportion of test statistic inflation ascribed to causes other than polygenicity was estimated to be 10.9-17.2% confirming that polygenicity is the main driver of test statistic inflation.  $\lambda_{gc}$  was similar to the GBE GWAS<sup>12</sup>. The quantile-quantile plot for the UKB is shown in Supplementary Figure 5.

## Sensitivity Analysis

A sensitivity analysis was conducted in each of the 55 lead SNPs using 23,422 individuals undergoing surgery for gallstones with 348,373 controls with no record of gallstones. 11 loci did not retain significance at the genome-wide threshold. Ten retained a suggestive association ( $P < 5 * 10^{-5}$ ; *ABO*, *ADAMTS20*, *FOXP1*, *GMDS-DT*, *GOT1*, chr18:rs542917327, *ITCH*, *NPAS2*, *SH3BP4*, *TBC1D32*) whilst one did not ( $P = 2.75 * 10^{-4}$ ; *LMNB1*). Direction and size of effect was very similar for all 55 loci suggesting detection bias had minimal influence on results. Full results at each locus are available in Supplementary Materials 4. A forest plot comparing effect sizes is shown in Supplementary Figure 6.

## Association of Lithogenic Loci with Phenotypic Traits

Assessment of serum lipids was undertaken for each of the 55 lithogenic variants from the UKB in all 377,000 individuals taken forward for GWAS. The lithogenic alleles were heterogeneous in their influence on serum lipids with some lithogenic alleles promoting hyperlipidaemia whilst others demonstrated a protective effect (see Figure 3). The lithogenic variants of *ABCG8* and *HNF4A* (rs11887534-C and rs1800961-T respectively) both resulted in reductions of cholesterol of 0.10mmol/L (95% CI 0.09-0.12). The lithogenic variants of *GCKR* and *ABO* demonstrated highly significant increases in serum cholesterol (rs1260326-C: 0.05mmol/L [95% CI 0.05-0.06]; and rs115478735-T: 0.07 [0.06-0.07]). Figures for the other serum biochemistry markers are available in Supplementary Figures 7-12. Numeric results are shown in Supplementary Materials 5.

## Gene-Environment Interaction

Gene-environment interaction of SNPs with serum unconjugated bilirubin was undertaken at 54 of the lead SNPs from the main GWAS (excluding *UGT1A1*) in the 377,000 individuals. After Bonferroni correction for 54 analyses, a significance level of 0.00093 was determined. No variants reached statistical significance although nominal significance was seen in *LIN28B* ( $P=0.012$ ), *ABCB4* ( $P=0.023$ ), *JMJD1C* ( $P=0.008$ ), *ANPEP* ( $P=0.041$ ), *CLDN7* ( $P=0.011$ ), *ATP8B1* ( $P=0.019$ ) and *HNF4A* ( $P=0.002$ ).

## Polygenic Risk Score

The PRS was significantly associated with gallstone disease in the validation cohort (maximum variance explained on the liability scale  $R^2 = 2.5\%$ ,  $P = 2.62 * 10^{-181}$ , when using a GWAS P value threshold of 0.0001). There was a 6-fold increase of risk of gallstones in those with the highest decile of genetic risk compared to those in the lowest with an approximately linear increase in risk with each increasing PRS decile (Figure 4). The PRS was strongly associated with increased BMI, waist circumference, hip circumference, liver

enzymes (other than ALP), cystatin C and CRP (greater genetically-predicted risk of gallstones resulted in higher values for those traits). The PRS was associated with reductions in cholesterol, HDL, LDL, ApoB and ALP. Generally, for all traits with which the PRS was significantly associated there was an approximately linear increase (or decrease) with each increase in PRS decile and no apparent threshold effect. There were no obvious differences in levels of trigs, ApoA, LipoA, blood count markers, HbA1c or glucose (Supplementary Figures 13-15).

## Pathway Enrichment

172 genes showed significant association with gallstones ( $P < 2.59 * 10^{-6}$ ; Supplementary Materials 4) and were highly expressed in liver ( $P = 4.09 * 10^{-6}$ ; gallbladder expression was not available in MAGMA). 13 gene-sets were identified as significantly associated with gallstones (Table 3). Many of these gene-sets relate to glucuronidation, lipid metabolism and phospholipid transport. Additionally, Waldenström's macroglobulinaemia and curated gene-sets of cancer drug resistance were identified which both contained *ABCB1* and *ABCB4*, both regulate biliary composition but are highly pleiotropic.

Literature search identified two genes governing gastrointestinal motility (*ANO1* and *TMEM147*). Rs7599 lies within the 3'UTR region and the lithogenic allele (A) is associated with significantly reduced expression of *TMEM147* in the liver ( $P = 3.0 * 10^{-47}$ ). Data on effects of rs56363382 and all other significant SNPs in the *ANO1* loci were not available from GTEx. Both *ANO1* and *TMEM147* were found to be highly expressed in gallbladder mucosa in the Human Protein Atlas. 5 identified genes were involved in primary cilia function (*DYNC2LI1*, *TBC1D32*, *ADAMTS20*, *PKD2L1* and *POC1B*).

## Discussion

We report the largest and most detailed GWA of gallstones conducted worldwide with functional annotation of variants and gene-set analysis. We analysed 377,000 individuals (28,627 cases and 348,373 controls) from the UKB and replicated findings in a Scottish cohort (GS-SFHS) and a Finnish Cohort (FinnGen). We performed a genome-wide meta-analysis with UKB and FinnGen data. We identified 46 new lithogenic loci and replicated all 29 previously identified loci bringing the total number of identified gallstone-susceptibility loci to 75. Of the 75 loci, 12 were identified only in the UKB GWAS and not replicated or significant in the GWA meta-analysis. Through pathway enrichment we identified 13 gene-sets strongly associated with gallstones. Through FUMA annotation we identified two novel pathways that may influence gallstone development, gastrointestinal motility and ciliogenesis. Our results highlight important pathways that may be targeted by novel therapies and form the basis for further research into gallstones and other cholestatic disease.

Two newly identified loci are highly expressed in the gallbladder and govern gastrointestinal motility. *ANO1* (also known as *TMEM16A*) is a calcium-activated chloride channel widely expressed in the gastrointestinal tract with roles in automaticity and contraction of the smooth muscle cells via the interstitial cells of Cajal<sup>33</sup>. Administration of *ANO1* inhibitors (TMinh-23) in murine models impedes gastric emptying with improved oral but not parenteral glucose tolerance supporting the role of *ANO1* as a gastric motility promoter<sup>33,34</sup>. *ANO1* is also expressed in the biliary epithelium where it has another role in secretion of fluid from the apical cholangiocyte membrane in response to bile acids suggesting it may promote increased bile flow<sup>35</sup>. *ANO1* may influence gallstone formation via gallbladder contractility and biliary flow. Given the possible impact on gallbladder motility and biliary composition, we suggest that any trials of TMinh-23 assess for development of gallstones as an important safety consideration. *TMEM147* is known to inhibit transfer of M3 muscarinic receptors to the cell surface<sup>36</sup>. M3 receptors are highly expressed in smooth muscle cells of the gallbladder and may influence gallbladder contractility and biliary stasis<sup>37</sup>. The variants may also influence gallstone formation through increased colonic transit time which is also associated with gallstones and believed to alter enterohepatic bile acid circulation<sup>14</sup>.

Five loci relate to function of primary cilia. *TBC1D32* controls ciliogenesis within the neural tube and interacts with the sonic hedgehog protein which has wider roles in organogenesis and axonal guidance<sup>38</sup>. The metalloproteinase encoded by *ADAMTS20* also plays a pleiotropic role in ciliogenesis and hedgehog signaling with mutant variants causing ciliopathies<sup>39</sup>. *POC1B* is essential for formation of the centriole and ciliogenesis during cell division<sup>40</sup>. *PKD2L1* is involved in regulating calcium concentration within the primary cilia<sup>41</sup> and disorders of calcium signaling of the primary cilia are thought to result in cholestasis<sup>42</sup>. Three independent signals within the *ABCG8* locus were intronic or 3'-UTR variants of *DYNC2LI1*. *DYNC2LI1* has a role in ciliogenesis with mutations causing severe ciliopathies<sup>43</sup>. Variants within the sonic hedgehog gene and *DYNC2LI1* have previously been associated with gallbladder cancer and gallstone formation<sup>44</sup> whilst *TBC1D32* and *ADAMTS20* (novel loci, not replicated) and *PKD2L1* and *POC1B* (identified in the meta-

analysis) are novel loci. We suggest that abnormalities in cholangiocyte ciliogenesis and ciliary function may contribute to biliary stasis with gallstone formation. Abnormal expression or function of cholangiocyte cilia causes other cholestatic disease and malignancies<sup>45</sup>. At present, therapies targeting primary cilia are limited. Ricolinostat, a selective inhibitor of histone deacetylase 6, is under investigation in several malignancies but its effect on gallstones remains unknown<sup>45</sup>.

MAGMA analysis identified thirteen gene-sets. Most were related to lipid homeostasis and glucuronidation. Gene-sets for Waldenström's macroglobulinaemia and multidrug-resistance pathways were enriched with enrichment driven by *ABCB4* and *ABCB1*. *ABCB4* encodes a phosphatidylcholine flippase which effluxes of phosphatidylcholine into bile<sup>46</sup>. Phosphatidylcholine acts as a solvent for biliary cholesterol. *ABCB4* mutations are a cause of early-onset gallstones and low-phospholipid-associated cholelithiasis<sup>47</sup>. *ABCB1* is a neighbouring gene involved in phospholipid and drug excretion with broad specificity<sup>46</sup>. The pathway enrichments are most likely a feature of pleiotropy as these genes have several distinct functions. However, serum IgG and IgM extracted from patients with Waldenström's macroglobulinaemia results in nucleation of biliary cholesterol, possibly representing an alternate mechanism directly linking gallstones with Waldenström's macroglobulinaemia<sup>48</sup>.

Based on our results we suggest the following priorities for further research. 1. Direct genotyping at all 75 lead SNPs in individuals with corresponding cholecystectomy specimens for biochemical analysis of biliary content, gallbladder epithelial protein expression and cholangiocyte protein expression. 2. Assessment of gallbladder emptying with hepatobiliary scintigraphy in patients genotyped at *ANO1* and *TMEM147* to assess the influence of the lithogenic variants on gallbladder motility. 3. Assessment of gallstone risk in murine models with pharmacological modification of gallbladder motility (TMinh-23) or ciliotherapy (ricolinostat). 4. Validation of our findings in a large cohort in which radiological evaluation of gallstone disease has been undertaken.

The strengths of this study include a more complete definition of gallstones than in previous studies which reduces the rate of case misclassification. Notably, previous studies have identified up to 18,417 cases whilst we identify 28,627 from the UKB. We use three large cohorts with replication of the majority of novel findings. Many of the loci we identify have plausible relationships to gallstones which we explore in pathway analysis, tissue expression analyses and through assessing the impact of risk alleles on serum biomarkers. The limitations of this study include misclassification of individuals with asymptomatic gallstones that have not been detected clinically. Relying on clinical disease as the definition of exposure has been common to all earlier GWAS of gallstones although our selection of relevant codes has improved our sensitivity relative to other studies. The phenotypic characterisation of participants in the UKB is limited to less invasive procedures such as venepuncture and no data on biliary composition or pathological records of gallstone composition following cholecystectomy were available. This impedes the interpretation of cholesterol-influencing SNPs as SNPs may lead to reduced serum cholesterol through biliary excretion, increased serum cholesterol through intestinal absorption or a balanced combination of both with no net change in serum cholesterol. All three scenarios may lead to similar changes in biliary cholesterol yet demonstrate

opposing effects on serum cholesterol. Finally, the definition of gallstones in the FinnGen cohort was not identical to the UKB as we used the published summary association statistics from the FinnGen GWAS although the majority of UKB cases would have met the definition used in FinnGen.

In summary, we have performed a GWAS of gallstone disease using UKB data, performed a meta-analysis with the FinnGen cohort and identified 46 novel variants associated with gallstone disease. We identify the first known gallstone-susceptibility variants that influence gastrointestinal motility, and we identify key priorities for research into gallstone formation and preventative treatments.

## Funding

This work was funded by the Medical Research Council (MR/T008008/1).

## Acknowledgements

Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006] and is currently supported by the Wellcome Trust [216767/Z/19/Z]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award “Stratifying Resilience and Depression Longitudinally” (STRADL) Reference 104036/Z/14/Z).

This work was funded by a Medical Research Council Clinical Research Training Fellowship awarded to CJF (MR/T008008/1). ADB acknowledges funding from the Wellcome PhD training fellowship for clinicians (204979/Z/16/Z), the Edinburgh Clinical Academic Track (ECAT) programme. CH and JFW acknowledges support from the MRC Human Genetics Unit programme grant, “Quantitative traits in health and disease” (U. MC\_UU\_00007/10). NCH is supported by a Wellcome Trust Senior Research Fellowship in Clinical Science (ref. 219542/Z/19/Z). NLM is supported by the British Heart Foundation through a Senior Clinical Research Fellowship (FS/16/14/32023) and a Research Excellence Award (RE/18/5/34216). RKS is supported by the Wellcome Trust (grant 210752/Z/18/Z). AS acknowledges support from the Academy of Medical Sciences/the Wellcome Trust/ the Government Department of Business, Energy and Industrial Strategy/the British Heart Foundation/Diabetes UK Springboard Award [SBF006\1109].

## Competing Interests

No authors declare any competing interests.

## Ethical Approval

The UKB received ethical approval (Research Ethics Committee reference number: 11/NW/0382). Ethical approval was granted for Generation Scotland: Scottish Family Health Study by NHS Tayside Research Ethics Committee (reference number: 05/S1401/89).



## Author Contributions

CJF conducted the genome-wide association study and drafted the manuscript. CJF, ADB, PKJ, DWC, PRHJT and JFW developed the code required for data analysis. CJF, TMD and JFW assisted with data access requests. RP, EMH, AI and JFW assisted with data storage and server maintenance. DJP, CH and AC contributed data from Generation Scotland. CJF, EMH, AS, SJW, CAS, TMD, RKS, NLM, CLMS and RP were involved in study design and conception. JAF, NCH, PR, NLM, RKS, SJW and EMH were involved in critical appraisal of the manuscript and placing the research in the biological context. ADB, AC, CH, AI, JFW and AS were involved in critical appraisal of the manuscript in relation to GWAS methodology. CJF, EMH, RKS, AS and AI were involved in pathway enrichment analyses.

## References

1. Harrison, E. M. *et al.* Hospital volume and patient outcomes after cholecystectomy in Scotland: Retrospective, national population based study. *British Medical Journal* **344**, e3330 (2012).
2. Everhart, J. E. & Ruhl, C. E. Burden of digestive diseases in the United States Part III: Liver, biliary tract, and pancreas. *Gastroenterology* **136**, 1134–1144 (2009).
3. Jørgensen, T. Prevalence of Gallstones in a Danish Population. *American Journal of Epidemiology* **126**, 912–921 (1987).
4. Everhart, J. E., Khare, M., Hill, M. & Maurer, K. R. Prevalence and ethnic differences in gallbladder disease in the United States. *Gastroenterology* **117**, 632–639 (1999).
5. Katsika, D. *et al.* Genetic and environmental influences on symptomatic gallstone disease: A Swedish study of 43,141 twin pairs. *Hepatology* **41**, 1138–1143 (2005).
6. Buch, S. *et al.* A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nature Genetics* **39**, 995–999 (2007).
7. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nature Genetics* **47**, 435–444 (2015).
8. Joshi, A. D. *et al.* Four Susceptibility Loci for Gallstone Disease Identified in a Meta-analysis of Genome-Wide Association Studies. *Gastroenterology* **151**, 351–363.e28 (2016).
9. Rodriguez, S. *et al.* Lipids, obesity and gallbladder disease in women: Insights from genetic studies using the cardiovascular gene-centric 50K SNP array. *European Journal of Human Genetics* **24**, 106–112 (2016).
10. Ferkingstad, E. *et al.* Genome-wide association meta-analysis yields 20 loci associated with gallstone disease. *Nature Communications* **9**, (2018).
11. Bustos, B. I. *et al.* Variants in ABCG8 and TRAF3 genes confer risk for gallstone disease in admixed Latinos with Mapuche Native American ancestry. *Scientific Reports* **9**, 1–12 (2019).
12. Gellert-Kristensen, H. *et al.* Identification and Replication of Six Loci Associated With Gallstone Disease. *Hepatology* **70**, 597–609 (2019).
13. Portincasa, P., Moschetta, A. & Palasciano, G. Cholesterol gallstone disease. *The Lancet* **368**, 230–239 (2006).
14. van Erpecum, K. J. & Berge-Henegouwen, G. P. V. Gallstones: An intestinal disease? *Gut* **44**, 435–438 (1999).
15. McInnes, G. *et al.* Global Biobank Engine: Enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics* **35**, 2495–2497 (2019).

16. Sudlow, C. *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 medicine* **12**, e1001779 (2015).
17. Smith, B. H. *et al.* Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *International Journal of Epidemiology* **42**, 689–700 (2013).
18. FinnGen. FinnGen research project is an expedition to the frontier of genomics and medicine. *FinnGen* (2020).
19. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
20. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).
21. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291–295 (2015).
22. Buch, S. *et al.* Loci From a Genome-Wide Analysis of Bilirubin Levels Are Associated With Gallstone Risk and Composition. *Gastroenterology* **139**, 1942–1951.e2 (2010).
23. Fertrin, K. Y., Melo, M. B., Assis, Â. M., Saad, S. T. O. & Costa, F. F. UDP-glucuronosyltransferase 1 gene promoter polymorphism is associated with increased serum bilirubin levels and cholecystectomy in patients with sickle cell anemia. *Clinical Genetics* **64**, 160–162 (2003).
24. Chaar, V. *et al.* Association of UGT1A1 polymorphism with prevalence and age at onset of cholelithiasis in sickle cell anemia. *Haematologica* **90**, 188–199 (2005).
25. Wasmuth, H. E. *et al.* Coinheritance of Gilbert syndromeAssociated UGT1A1 mutation increases gallstone risk in cystic fibrosis. *Hepatology* **43**, 738–741 (2006).
26. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).
27. Leeuw, C. A. de, Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology* **11**, e1004219 (2015).
28. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics* **11**, 499–511 (2010).
29. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
30. R Core Team. R: A Language and Environment for Statistical Computing. (2017).
31. Little, J. *et al.* STrengthening the REporting of Genetic Association Studies (STREGA) of the STROBE Statement. *PLOS Medicine* **6**, e1000022 (2009).

32. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics* **19**, 807–812 (2011).
33. Mazzone, A. *et al.* Altered Expression of Ano1 Variants in Human Diabetic Gastroparesis. *The Journal of Biological Chemistry* **286**, 13393–13403 (2011).
34. Cil, O. *et al.* Slowed gastric emptying and improved oral glucose tolerance produced by a nanomolar-potency inhibitor of calcium-activated chloride channel TMEM16A. *The FASEB Journal* **33**, 11247–11257 (2019).
35. Li, Q., Dutta, A., Kresge, C., Bugde, A. & Feranchak, A. P. Bile acids stimulate cholangiocyte fluid secretion by activation of transmembrane member 16A Cl<sup>-</sup> channels. *Hepatology (Baltimore, Md.)* **68**, 187–199 (2018).
36. Rosemond, E. *et al.* Regulation of M<sub>3</sub> muscarinic receptor expression and function by transmembrane protein 147. *Molecular Pharmacology* **79**, 251–261 (2011).
37. Parkman, H. P., Pagano, A. P. & Ryan, J. P. Subtypes of muscarinic receptors regulating gallbladder cholinergic contractions. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **276**, G1243–G1250 (1999).
38. Ko, H. W. *et al.* Broad-minded links cell cycle-related kinase to cilia assembly and Hedgehog signal transduction. *Developmental cell* **18**, 237–247 (2010).
39. Nandadasa, S. *et al.* Secreted metalloproteases ADAMTS9 and ADAMTS20 have a non-canonical role in ciliary vesicle growth during ciliogenesis. *Nature Communications* **10**, 953 (2019).
40. Atorino, E. S., Hata, S., Funaya, C., Neuner, A. & Schiebel, E. CEP44 ensures the formation of bona fide centriole wall, a requirement for the centriole-to-centrosome conversion. *Nature Communications* **11**, 903 (2020).
41. DeCaen, P. G., Dellinger, M., Vien, T. N. & Clapham, D. E. Direct recording and molecular identification of the calcium channel of primary cilia. *Nature* **504**, 315–318 (2013).
42. Amaya, M. J. & Nathanson, M. H. Calcium signaling and the secretory activity of bile duct epithelia. *Cell Calcium* **55**, 317–324 (2014).
43. Taylor, S. P. *et al.* Mutations in DYNC2LI1 disrupt cilia function and cause short rib polydactyly syndrome. *Nature Communications* **6**, 7092 (2015).
44. Dixit, R., Pandey, M., Tripathi, S. K., Dwivedi, A. N. D. & Shukla, V. K. Comparative Analysis of Mutational Profile of Sonic hedgehog Gene in Gallbladder Cancer. *Digestive Diseases and Sciences* **62**, 708–714 (2017).
45. Gradilone, S. A. *et al.* HDAC6 Is Overexpressed in Cystic Cholangiocytes and Its Inhibition Reduces Cystogenesis. *The American Journal of Pathology* **184**, 600–608 (2014).

46. van Helvoort, A. *et al.* MDR1 P-Glycoprotein Is a Lipid Translocase of Broad Specificity, While MDR3 P-Glycoprotein Specifically Translocates Phosphatidylcholine. *Cell* **87**, 507–517 (1996).
47. Rosmorduc, O. *et al.* ABCB4 gene mutation-associated cholelithiasis in adults. *Gastroenterology* **125**, 452–459 (2003).
48. Harvey, P. R., Upadhyya, G. A. & Strasberg, S. M. Immunoglobulins as nucleating proteins in the gallbladder bile of patients with cholesterol gallstones. *Journal of Biological Chemistry* **266**, 13996–14003 (1991).

RSID	Chr:Pos	Reference / Effect Allele	EAF	OR (95% CI)	P Value	Consequence (Gene)	Hypothesised Functional Gene	Functional Support
<b>Novel Loci</b>								
<b>rs9427114</b>	1:154626705	T/C	0.490	1.05 (1.04-1.07)	$1.47 \times 10^{-10}$	Intergenic (ADAR, KCNN3)	Unclear	Unknown
rs183820313	2:101556478	G/T	0.001	1.38 (1.11-1.71)	$2.00 \times 10^{-08}$	Intron (NPAS2)	NPAS2	Unknown, possible Circadian variation of CYP7A1
rs60360195	2:235937268	A/T	0.200	1.06 (1.04-1.08)	$6.66 \times 10^{-09}$	Intron (SH3BP4)	SH3BP4	Unknown, possible pro-nucleating impact of transferrin
rs71621955	3:71668961	A/AT	0.433	0.95 (0.94-0.97)	$1.08 \times 10^{-08}$	Intergenic (EIF4E3, RP11-154H23.3)	FOXP1	Unknown, possible impact on cholesterol and glucose metabolism
<b>rs4493564</b>	4:77420407	A/G	0.353	1.05 (1.03-1.07)	$5.21 \times 10^{-09}$	Intron (SHROOM3)	SHROOM3	Unknown
rs764777280	5:126045423	T/C	0.001	1.33 (1.06-1.67)	$3.84 \times 10^{-08}$	Intergenic	LMNB1	Unknown
rs190891554	5:140496089	T/C	0.001	1.38 (1.13-1.69)	$2.59 \times 10^{-08}$	Intergenic (PCDHB3, AC005754.8)	Proctocadherins	Unknown, possible impact on biliary composition
rs753948177	5:165762052	A/G	0.001	1.67 (1.34-2.07)	$9.08 \times 10^{-09}$	Intergenic	Unclear	Unknown
rs6936023	6:2303204	G/A	0.181	0.94 (0.92-0.96)	$4.71 \times 10^{-08}$	Intron (GMDS-AS1)	GMDS	Unknown
<b>rs11399580</b>	6:17605699	G/GT	0.453	0.95 (0.93-0.96)	$1.95 \times 10^{-10}$	Intron (FAM8A1)	NUP153	Unknown
<b>rs56339318</b>	6:105404562	A/C	0.104	0.92 (0.89-0.94)	$1.38 \times 10^{-09}$	Upstream (LIN28B)	Unclear	Unknown
rs533560317	6:121468999	A/G	0.001	1.28 (1.05-1.57)	$1.09 \times 10^{-09}$	Intron (TBC1D32)	TBC1D32	Unknown, possible impact on cholangiocyte cilia development
rs7786376	7:73042614	A/G	0.278	1.05 (1.03-1.07)	$4.97 \times 10^{-08}$	Intergenic (MLXIPL, VPS37D)	MLXIPL	Unknown, possible impact on cholesterol and glucose metabolism (encodes carbohydrate-response-element-binding protein)

<b>rs12532734</b>	7:107472382	G/A,T	0.225	0.92 (0.90-0.94)	$1.18 \times 10^{-16}$	Intergenic (PIGCP2, DLD)	Unclear	Unknown
<b>rs115478735</b>	9:136149711	A/T	0.186	1.05 (1.03-1.08)	$2.33 \times 10^{-08}$	Intron (ABO)	ABO	Unknown, possible impact on bilirubin metabolism
rs192575995	10:101244099	C/T	0.002	1.26 (1.05-1.50)	$8.18 \times 10^{-09}$	Intergenic	GOT1	Unknown (encodes aspartate aminotransferase)
rs528101230	11:65256028	C/T	0.001	1.39 (1.14-1.70)	$4.01 \times 10^{-08}$	Intergenic (AP000769.1, MALAT1)	Unclear	Unknown
<b>rs56363382</b>	11:69834473	C/T	0.080	1.13 (1.09-1.16)	$6.51 \times 10^{-15}$	Intron (RP11-626H12.1)	ANO1	Unknown, possible impact on gallbladder motility
rs10898881	11:72527018	A/G	0.453	0.95 (0.94-0.97)	$1.88 \times 10^{-08}$	Intron (ATG16L2)	Unclear	Unknown
rs185852010	12:44022849	C/T	0.001	1.24 (0.99-1.55)	$2.77 \times 10^{-08}$	Intergenic (RP11-73B8.2, RP11-73B8.3)	ADAMTS20	Unknown, possible impact on ciliogenesis
rs771297649	15:43360509	CCTAAGT AT/C	0.021	1.16 (1.09-1.22)	$3.82 \times 10^{-08}$	Intron (UBR1)	Unclear	Unknown
rs137931761	17:7166254	C/CT	0.174	0.94 (0.91-0.96)	$1.01 \times 10^{-08}$	5'-UTR (CLDN7)	CLDN7	Unknown, possible impact on biliary composition
rs542917327	18:27739443	A/G	0.001	1.33 (1.06-1.66)	$1.70 \times 10^{-08}$	Intergenic (RP11-784B15.1, RP11-675P14.1)	Unclear	Unknown
<b>rs7599</b>	19:36038390	A/G	0.366	0.95 (0.93-0.96)	$2.49 \times 10^{-11}$	3'-UTR (TMEM147)	TMEM147	Unknown, possible impact on gallbladder motility
<b>rs438568</b>	20:12958687	A/G	0.394	0.94 (0.93-0.96)	$4.85 \times 10^{-11}$	Intergenic (RP5-1069C8.2, SPTLC3)	SPTLC3	Unknown, possible impact on cholesterol metabolism
rs367900721	20:33098248	ATT/A	0.478	1.05 (1.03-1.07)	$1.62 \times 10^{-08}$	3'-UTR (ITCH)	ITCH	Unknown
<b>rs11089985</b>	22:40675561	A/G	0.348	1.05 (1.04-1.07)	$8.38 \times 10^{-10}$	Intron (TNRC6B)	Unclear	Unknown
<b>Established Loci</b>								
<b>rs1260326</b>	2:27730940	T/C	0.393	1.08 (1.06-1.1)	$6.80 \times 10^{-18}$	Nonsynonymous (GCKR)	GCKR	Unknown, possible impact on cholesterol and glucose metabolism
<b>rs11887534</b>	2:44066247	G/C	0.066	2.01	$4.96 \times 10^{-324}$	Nonsynonymous	ABCG8	Increased cholesterol

				(1.96-2.07)		us (ABCG8)		efflux into bile
rs2292553	2:219146803	G/A	0.437	1.04 (1.02-1.06)	$1.22 \times 10^{-06}$	Nonsynonymous (TMBIM1)	GPBAR1	Unknown, possible impact on bile acid metabolism
<b>2:234664586_ATC_A</b>	2:234664586	ATC/A	0.312	1.07 (1.05-1.09)	$1.71 \times 10^{-12}$	Intron (UGT1A3 - UGT1A10)	UGT1A1	Possible decreased bilirubin conjugation
<b>rs4681515</b>	3:149212076	A/G	0.437	0.89 (0.88-0.91)	$2.84 \times 10^{-39}$	Intron (TM4SF4)	Unclear	Unknown
<b>rs2290846</b>	4:151199080	G/A	0.290	1.13 (1.11-1.15)	$5.89 \times 10^{-38}$	Nonsynonymous (LRBA)	Unclear	Unknown
<b>rs7810332</b>	7:6486219	G/C	0.262	0.94 (0.92-0.96)	$2.20 \times 10^{-09}$	Intron (DAGLB)	Unclear	Unknown, possible impact on cholesterol and glucose metabolism
<b>rs7802555</b>	7:87130350	A/C	0.139	0.85 (0.83-0.88)	$1.37 \times 10^{-34}$	Intergenic (ABCB1, ABCB4)	ABCB4	Reduced phospholipid content of bile
<b>rs13280055</b>	8:11522353	G/A	0.133	1.08 (1.06-1.11)	$2.19 \times 10^{-11}$	Intergenic (GATA4, LINC00208)	GATA4	Unknown, possible regulation of SLC10A2
<b>rs1993453</b>	8:59347184	A/G	0.337	0.91 (0.89-0.92)	$1.62 \times 10^{-29}$	Intron (UBXN2B)	CYP7A1	Reduced bile acid production and increased cholesterol
rs2469991	8:120347476	A/T	0.289	0.94 (0.92-0.96)	$5.10 \times 10^{-10}$	Intergenic (RP11-4K16, SNORA32)	Unclear	Unknown
<b>rs686030</b>	9:15304782	C/A	0.142	1.13 (1.11-1.16)	$1.96 \times 10^{-21}$	Intron (TTC39B)	TTC39B	Possible altered HDL-C or ABCG8 expression
rs10828250	10:21847178	C/G	0.309	1.06 (1.04-1.08)	$2.30 \times 10^{-10}$	Intron (MLLT10)	Unclear	Unknown
<b>rs3802548</b>	10:45952745	T/A	0.241	1.12 (1.1-1.14)	$2.02 \times 10^{-29}$	3'-UTR (MARCHF8)	MARCHF8	Unknown, possible impact on cholesterol metabolism
<b>rs2393969</b>	10:65140440	A/C	0.470	1.07 (1.06-1.09)	$1.16 \times 10^{-17}$	Intron (JMJD1C)	Unclear	Unknown
<b>rs174574</b>	11:61600342	A/C	0.353	0.94 (0.93-0.96)	$3.04 \times 10^{-10}$	Intron (FADS2)	FADS2	Unknown, possible impact on cholesterol metabolism
<b>rs2393775</b>	12:121424574	G/A	0.378	1.09 (1.07-1.11)	$1.32 \times 10^{-23}$	Intron (HNF1A)	HNF1A	Unknown, possible impact on cholesterol metabolism
rs144846334	13:103737747	G/A	0.010	1.35 (1.25-1.46)	$5.42 \times 10^{-17}$	Intergenic (SLC10A2, RP11-123H22)	SLC10A2	Unknown, possible decreased bile acid reuptake



rs28929474	14:94844947	C/T	0.020	1.36 (1.29-1.43)	$2.65 \times 10^{-30}$	Nonsynonymous (SERPINA1)	SERPINA1	Unknown, possible hepatic cholesterol synthesis, cholesterol nucleation or gallbladder motility
rs17240268	15:90347814	G/A	0.090	0.89 (0.86-0.92)	$3.03 \times 10^{-13}$	Nonsynonymous (ANPEP)	ANPEP	Unknown, likely to promote cholesterol nucleation
rs3784924	16:11644842	A/G	0.314	1.06 (1.04-1.08)	$2.67 \times 10^{-11}$	Intron (LITAF)	LITAF	Unknown, possible impact on hepatic inflammation
rs17138478	17:36073320	C/A	0.129	1.09 (1.06-1.11)	$2.05 \times 10^{-11}$	Intron (HNF1B)	HNF1B	Unknown, possible impact on cholesterol and glucose metabolism or cholangiocyte cilia
18:55320914_CTT_C	18:55320914	CTT/C	0.181	0.94 (0.92-0.96)	$4.63 \times 10^{-08}$	Intron (ATP8B1, RP11-35G9)	ATP8B1	Progressive familial intrahepatic cholestasis
rs708686	19:5840619	C/T	0.267	1.06 (1.04-1.08)	$4.18 \times 10^{-12}$	Upstream (FUT6)	FUT6	Unknown, possible impact on erythrocyte metabolism
rs34255979	19:46384830	C/T	0.120	1.13 (1.10-1.16)	$8.54 \times 10^{-20}$	Intergenic (IRF2BP1, FOXA3)	FOXA3	Unknown, possible impact on cholesterol metabolism via FOXA3
rs62129966	19:48374950	C/A	0.164	0.85 (0.83-0.87)	$1.06 \times 10^{-40}$	Intron (SULT2A1)	SULT2A1	Unknown, possible decreased bile acid production
rs681343	19:49206462	C/T	0.492	1.12 (1.10-1.14)	$2.76 \times 10^{-32}$	Synonymous (FUT2)	FUT2	Unknown, possible impact on erythrocyte metabolism
rs1800961	20:43042364	C/T	0.031	1.28 (1.23-1.34)	$1.55 \times 10^{-29}$	Nonsynonymous (HNF4A)	HNF4A	Unknown, possible increased gluconeogenesis or ABC8-mediated cholesterol secretion
rs749502751	22:38877775	AC/A	0.367	0.95 (0.93-0.97)	$4.67 \times 10^{-09}$	Intron (KDEL3)	Unclear	Unknown

Chr:Pos: Chromosome:Position, EAF: Effect allele frequency, P Value: P value using allelic model, OR (95% CI): Odds ratio with 95% confidence interval, ABO: Alpha 1-3-N-acetyl-galactos-aminyltransferase and Alpha 1-3-galactosyl-transferase, ADAMTS20: ADAM Metalloproteinase with Thrombospondin Type 1 Motif 20, ADAR: Adenosine Deaminases acting on RNA, ANO1: Anoctamin 1, ATG16L2: Autophagy-related 16 Like 2, CLDN7: Claudin 7, FOXP1: Forkhead Box P1, FAM8A1: Family with sequence similarity 8 member A1, GMDS-DT: GDP-mannose 4,6-dehydratase Divergent Transcript, GOT1: Glutamic-oxaloacetic Transaminase 1, ITC: Itchy E3 Ubiquitin Protein Ligase, LIN28B: Lin-28 Homolog B, LINC02736: Long Intergenic Non-Coding Transcript, LMNB1: Lamin B1, MLXIP: MLX Interacting Protein Like, NPAS2: Neuronal PAS Domain Protein 2, NUP153: Nucleoporin 153, PCDHB4: Protocadherin Beta 4, SH3BP4: SH3 Domain Binding Protein 4, SHROOM3: Shroom Family Member 3, SPTLC3: Serine Palmitoyl-transferase Long Chain Base Subunit 3, TBC1D32: TBC1 Domain Family Member 32, TMEM147: Transmembrane Protein 147, TNRC6B: Trinucleotide Repeat Containing Adaptor 6B, UBR1: Ubiquitin Protein Ligase E3 Component N-Recognin 1, dsRNA: Double-stranded Ribonucleic acid; CYP7A1: Cytochrome P450 Family 7 Subfamily A Member 1, ABCB4: ATP-binding Cassette Subfamily B Member 4, ABCG8: ATP-binding Cassette Subfamily G Member 8, ANPEP: Alanine Aminopeptidase, ATP8B1:ATPase Phospholipid Transporting 8B1, CYP7A1: Cytochrome P450 Family 7 Subfamily A Member 1, DAGLB: Diacylglycerol Lipase Beta, FADS2: Fatty acid desaturase 2, FUT2: Fucosyltransferase 2, FUT6: Fucosyltransferase 6, GATA4: GATA binding protein 4, GSKR: Glucokinase Regulator, HNF1A: Hepatocyte Nuclear Factor 1 Alpha, HNF1B: Hepatocyte Nuclear Factor 1 Beta, HNF4A: Hepatocyte Nuclear Factor 4 Alpha, IRF2BP1: Interferon Regulatory Factor 2 Binding Protein 1, JMJD1C: Jumoni domain containing

1C, KDELR3: KDEL Endoplasmic Reticulum Protein Retention Receptor 3, LITAF: Lipopolysaccharide Induced TNF Factor, LRBA: Lipopolysaccharide Responsive Beige-Like Anchor Protein, MAL2: Mal, T Cell Differentiation Protein 2, MARCHF8: Membrane-associated Ring-CH-type Finger 8, MLLT10: MLLT10 Histone Lysine Methyltransferase DOT1L Cofactor, SERPINA1: Serpin Family A Member 1, SLC10A2: Solute Carrier Family 10 Member 2, SULT2A1: Sulfotransferase Family 2A Member 1, TM4SF4: Transmembrane 4 L Six Family Member 4, TMBIM1: Transmembrane BAX Inhibitor Motif Containing 1, TTC39B: Tetratricopeptide Repeat Domain 39B, UGT1A1: Uridine-5'-diphospho- glucuronosyltransferase 1A1, GPBAR1: G Protein-Coupled Bile Acid Receptor 1, ATP: Adenosine triphosphate, LPS: Lipopolysaccharide, LXR: Liver X receptor, LD: Linkage disequilibrium, TGR5: G Protein-Coupled Bile Acid Receptor 1, HDL-C: High-density lipoprotein cholesterol, SNP: Single nucleotide polymorphism, 'Gene' and 'Consequence' were determined by the nearest gene after annotation by the FUMA application: rs1993453 is an intron within UBXN2B whilst CYP7A1 has a strong plausible biological relation to gallstones. Functional Role is based on assessment of published literature. Chromosome and position based on Genome Reference Consortium Human Build 37. Variants highlighted in bold were replicated in either GS-SFHS or FinnGen.

rsID	Reference Allele	Effect Allele	P Value	Odds Ratio (95% CI)	Gene
rs41279738	T	G	3.32*10 <sup>-09</sup>	1.13 (1.10-1.15)	GPR61
rs6702549	T	C	3.39*10 <sup>-12</sup>	1.05 (1.05-1.06)	ADAR
rs1260326	T	C	1.22*10 <sup>-23</sup>	1.08 (1.07-1.09)	GCKR
rs11887534	G	C	2.21*10 <sup>-715</sup>	2.07 (2.05-2.10)	ABCG8
rs55932961	A	G	7.65*10 <sup>-09</sup>	1.06 (1.05-1.07)	CPS1
rs736730	C	T	9.46*10 <sup>-11</sup>	1.05 (1.04-1.06)	GPBAR1/TMBIM1
rs62185132	A	G	5.08*10 <sup>-09</sup>	0.95 (0.95-0.96)	AP1S3
rs4148325	C	T	2.26*10 <sup>-18</sup>	1.07 (1.06-1.08)	UGT1A1
rs72987449	T	C	1.11*10 <sup>-10</sup>	1.06 (1.05-1.07)	SH3BP4
rs1476698	A	G	3.47*10 <sup>-08</sup>	1.04 (1.04-1.05)	FARP2
rs7626199	T	C	1.26*10 <sup>-08</sup>	0.96 (0.95-0.96)	ARHGEF3
rs4681515	A	G	8.58*10 <sup>-52</sup>	0.89 (0.88-0.90)	TM4SF4
rs13135475	T	C	1.19*10 <sup>-10</sup>	0.95 (0.94-0.96)	UGDH
rs4493564	A	G	8.84*10 <sup>-11</sup>	1.05 (1.04-1.06)	SHROOM3
rs2290846	G	A	4.85*10 <sup>-46</sup>	1.13 (1.12-1.14)	LRBA
rs6916208	A	C	5.03*10 <sup>-10</sup>	0.94 (0.93-0.95)	GMDS-DT
rs2228375	T	C	5.94*10 <sup>-15</sup>	1.06 (1.06-1.07)	NUP153/FAM8A1
rs735286	T	C	1.54*10 <sup>-08</sup>	0.95 (0.95-0.96)	VEGFA
rs56339318	A	C	1.73*10 <sup>-11</sup>	0.92 (0.90-0.93)	LIN28B
rs113638914	T	C	4.10*10 <sup>-08</sup>	1.05 (1.04-1.06)	SYNJ2
rs3828944	T	C	2.39*10 <sup>-12</sup>	0.94 (0.93-0.95)	DAGLB
rs35368205	C	T	1.89*10 <sup>-09</sup>	1.06 (1.05-1.07)	MLXIPL
rs7802555	A	C	8.59*10 <sup>-49</sup>	0.85 (0.84-0.85)	ABCB4
rs714583	T	A	1.55*10 <sup>-22</sup>	0.92 (0.91-0.92)	Intergenic Chr 7
rs7012637	A	G	5.88*10 <sup>-09</sup>	1.05 (1.04-1.05)	RP11-115J16.1
rs13280055	G	A	3.85*10 <sup>-15</sup>	1.09 (1.08-1.11)	GATA4
rs983812	C	T	1.34*10 <sup>-42</sup>	0.90 (0.89-0.90)	CYP7A1/UBXN2B
rs59893339	C	G	2.14*10 <sup>-10</sup>	1.07 (1.06-1.08)	LOC107986957
rs2469991	A	T	1.01*10 <sup>-09</sup>	0.95 (0.94-0.96)	MAL2
rs28601761	C	G	1.48*10 <sup>-11</sup>	1.05 (1.05-1.06)	TRIB1
rs686030	C	A	4.40*10 <sup>-30</sup>	1.14 (1.13-1.15)	TTC39B
rs115478735	A	T	4.02*10 <sup>-11</sup>	1.07 (1.06-1.08)	ABO
rs138844120	A	AAG	6.11*10 <sup>-10</sup>	1.05 (1.04-1.06)	SKIDA1/MLLT10
rs970548	A	C	1.59*10 <sup>-35</sup>	1.11 (1.10-1.12)	MARCHF8

rs2393969	A	C	$3.46 \times 10^{-19}$	1.07 (1.06-1.08)	JMJD1C
rs11189138	T	C	$4.98 \times 10^{-08}$	0.96 (0.95-0.97)	FRAT2
rs683854	C	A	$8.10 \times 10^{-10}$	1.05 (1.04-1.06)	GOT1/PKD2L1
rs55939540	A	G	$1.11 \times 10^{-09}$	0.89 (0.87-0.91)	MUC5AC
rs174567	G	A	$5.11 \times 10^{-15}$	0.94 (0.93-0.95)	FADS2
rs56363382	C	T	$7.24 \times 10^{-16}$	1.12 (1.10-1.13)	ANO1
rs7314929	T	C	$1.78 \times 10^{-10}$	0.95 (0.94-0.95)	POC1B
rs2393791	C	T	$8.37 \times 10^{-27}$	1.09 (1.08-1.10)	HNF1A
rs61967462	T	G	$2.08 \times 10^{-14}$	1.13 (1.11-1.15)	SLC10A2
rs9323382	A	G	$1.63 \times 10^{-09}$	0.94 (0.93-0.95)	SYT16
rs28929474	C	T	$1.11 \times 10^{-39}$	1.39 (1.35-1.42)	SERPINA1
rs143875230	G	A	$3.62 \times 10^{-08}$	1.16 (1.13-1.19)	UBR1
rs17240268	G	A	$3.32 \times 10^{-15}$	0.89 (0.88-0.91)	ANPEP
rs11644920	A	T	$2.44 \times 10^{-14}$	1.06 (1.06-1.07)	LITAF
rs8051062	T	C	$1.04 \times 10^{-08}$	0.96 (0.95-0.96)	CYB5B/NFAT5
rs17138478	C	A	$3.60 \times 10^{-15}$	1.09 (1.08-1.10)	HNF1B
rs369298568	T	C	$2.61 \times 10^{-10}$	0.93 (0.92-0.94)	ATP8B1
rs708686	C	T	$9.47 \times 10^{-18}$	1.07 (1.07-1.08)	FUT6
rs7599	A	G	$6.04 \times 10^{-13}$	0.94 (0.94-0.95)	TMEM147
rs7412	T	C	$8.18 \times 10^{-10}$	0.92 (0.90-0.93)	APOE
rs34255979	C	T	$2.70 \times 10^{-22}$	1.13 (1.11-1.14)	IRF2BP1
rs62129966	C	A	$9.12 \times 10^{-52}$	0.85 (0.84-0.86)	SULT2A1
rs601338	G	A	$4.92 \times 10^{-32}$	1.09 (1.09-1.10)	FUT2
rs686548	A	T	$3.77 \times 10^{-14}$	0.94 (0.94-0.95)	SPTLC3
rs2618566	T	G	$2.05 \times 10^{-10}$	0.95 (0.94-0.96)	AL035045.1
rs1800961	C	T	$4.48 \times 10^{-50}$	1.34 (1.31-1.36)	HNF4A
rs17219671	A	G	$1.44 \times 10^{-08}$	1.05 (1.04-1.06)	PTTG1IP
rs1946990	G	A	$1.15 \times 10^{-08}$	0.96 (0.95-0.96)	KDEL3/DMC1
rs41281265	A	G	$9.76 \times 10^{-13}$	1.06 (1.05-1.07)	TNRC6B

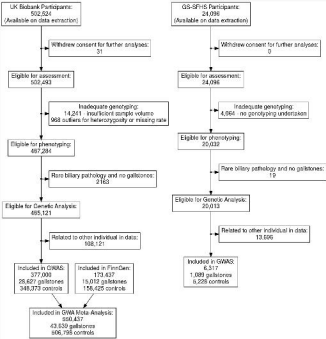
P Value: P value from meta-analysis (allelic model in both studies), OR (95% CI): Odds ratio with 95% confidence interval, ABCB4: ATP-binding Cassette Subfamily B Member 4, ABCG8: ATP-binding Cassette Subfamily G Member 8, ABO: Alpha 1-3-N-acetyl-galactos-aminyltransferase and Alpha 1-3-galactosyl-transferase, ADAR: Adenosine Deaminases acting on RNA, ANO1: Anoctamin 1, ANPEP: Alanine Aminopeptidase, AP1S3: Adaptor Related Protein Complex 1 Subunit Sigma 3, APOE: Apolipoprotein E, ARHGAP3: Rho GTPase-activating protein 3, ATP8B1: ATPase Phospholipid Transporting 8B1, CPS1: carbamoyl-phosphate synthase 1, CYB5B: Cytochrome B5 Type B, NFAT5: Nuclear Factor Of Activated T Cells 5, CYP7A1: Cytochrome P450 Family 7 Subfamily A Member 1, UBXLN2B: UBXLN Domain Protein 2B, DAGLB: Diacylglycerol Lipase Beta, FADS2: Fatty acid desaturase 2, FARP2: FERM, ARH/RhoGEF And Pleckstrin Domain Protein 2, FRAT2: FRAT regulator of WNT signaling pathway 1, FUT2: Fucosyltransferase 2, FUT6: Fucosyltransferase 6, GATA4: GATA binding protein 4, GSK3: Glycogen synthase kinase-3, GMDS-DT: GDP-mannose 4,6-dehydratase Divergent Transcript, GOT1: Glutamic-oxaloacetic Transaminase 1, PKD2L1: Polycystin 2 Like 1, Transient Receptor Potential Cation Channel, GPBAR1: G Protein-Coupled Bile Acid Receptor 1, TM6SF2: Transmembrane 6 Superfamily 2 Member 2, GPR61: G Protein-Coupled Receptor 61, HNF1A: Hepatocyte Nuclear Factor 1 Alpha, HNF1B: Hepatocyte Nuclear Factor 1 Beta, HNF4A: Hepatocyte Nuclear Factor 4 Alpha, IRF2BP1: Interferon Regulatory Factor 2 Binding Protein 1, JMJD1C: Jumonji domain containing 1C, KDEL3: KDEL Endoplasmic Reticulum Protein Retention Receptor 3, DMC1: DNA Meiotic Recombinase 1, LIN28B: Lin-28 Homolog B, LITAF: Lipoteichoic acid binding protein.

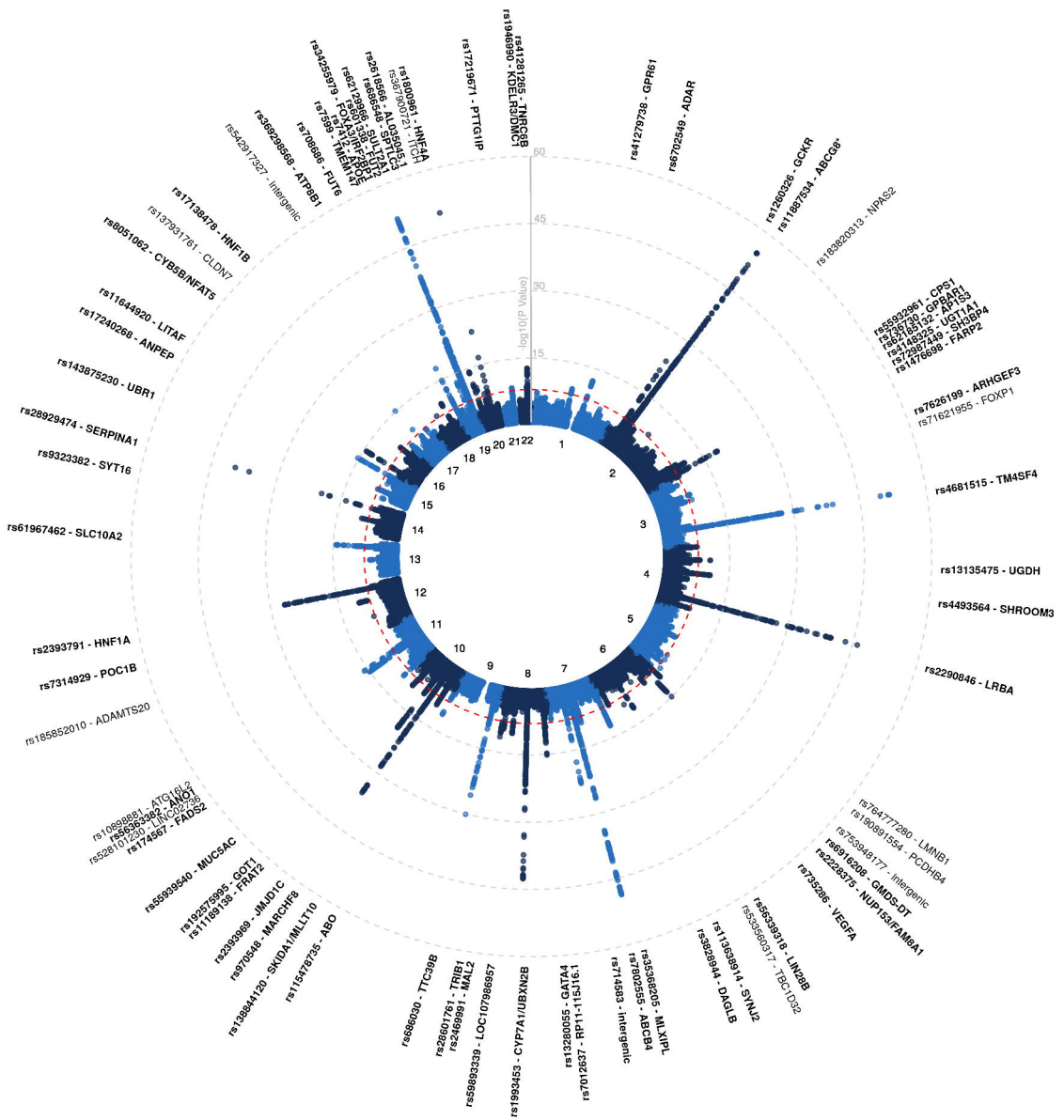
Lipopolysaccharide Induced TNF Factor, LRBA: Lipopolysaccharide Responsive Beige-Like Anchor Protein, MAL2: Mal, T Cell Differentiation Protein 2, MARCHF8: Membrane-associated Ring-CH-type Finger 8, MLXIPL: MLX Interacting Protein Like, MUC5AC: Mucin 5AC, Oligomeric Mucus/Gel-Forming, NUP153: Nucleoporin 153, FAM8A1: Family with sequence similarity 8 member A1, POC1B: Proteome Of Centriole Protein 1B, PTTG1IP: Pituitary Tumor-Transforming Gene 1 Protein-Interacting Protein, SERPINA1: Serpin Family A Member 1, SH3BP4: SH3 Domain Binding Protein 4, SHROOM3: Shroom Family Member 3, SKIDA1: SKI/DACH Domain Containing 1, MLLT10: MLLT10 Histone Lysine Methyltransferase DOT1L Cofactor, SLC10A2: Solute Carrier Family 10 Member 2, SPTLC3: Serine Palmitoyl-transferase Long Chain Base Subunit 3, SULT2A1: Sulfotransferase Family 2A Member 1, SYNJ2: Synaptojanin 2, SYT16: Synaptotagmin 16, TM4SF4: Transmembrane 4 L Six Family Member 4, TMEM147: Transmembrane Protein 147, TNRC6B: Trinucleotide Repeat Containing Adaptor 6B, TRIB1: Tribbles 1, TTC39B: Tetratricopeptide Repeat Domain 39B, UBR1: Ubiquitin Protein Ligase E3 Component N-Recognin 1, UGDH: UDP-Glucose 6-Dehydrogenase, UGT1A1: Uridine-5'-diphospho- glucuronosyltransferase 1A1, ATP: Adenosine triphosphate, VEGFA: Vascular endothelial growth factor A.

Table 3: MAGMA Pathway Analysis - Significant Gene-Sets

Gene Set Name	N Genes	Beta	Beta STD	SE	P	Pbon
GO_bp:go_flavonoid_glucuronidation	9	6.54	0.143	0.79	$4.74 \times 10^{-17}$	$7.35 \times 10^{-13}$
GO_bp:go_xenobiotic_glucuronidation	11	3.73	0.090	0.57	$2.48 \times 10^{-11}$	$3.83 \times 10^{-07}$
GO_bp:go_regulation_of_sterol_transport	44	0.91	0.044	0.16	$3.30 \times 10^{-09}$	$5.10 \times 10^{-05}$
Curated_gene_sets:gutierrez_waldenstroems_macroglobulinemia_1_up	10	1.86	0.043	0.33	$1.40 \times 10^{-08}$	$2.16 \times 10^{-04}$
GO_mf:go_phosphatidylcholine_transporter_activity	10	1.85	0.043	0.34	$2.37 \times 10^{-08}$	$3.67 \times 10^{-04}$
Curated_gene_sets:reactome_glucuronidation	25	1.47	0.053	0.27	$4.13 \times 10^{-08}$	$6.40 \times 10^{-04}$
GO_bp:go_regulation_of_lipid_metabolic_process	363	0.26	0.036	0.05	$9.97 \times 10^{-08}$	$1.54 \times 10^{-03}$
GO_bp:go_lipid_homeostasis	126	0.43	0.035	0.08	$1.46 \times 10^{-07}$	$2.26 \times 10^{-03}$
Curated_gene_sets:biocarta_mrp_pathway	6	1.86	0.033	0.37	$3.46 \times 10^{-07}$	$5.35 \times 10^{-03}$
GO_bp:go_uronic_acid_metabolic_process	19	1.73	0.055	0.35	$5.48 \times 10^{-07}$	$8.48 \times 10^{-03}$
GO_bp:go_regulation_of_steroid_metabolic_process	113	0.42	0.033	0.07	$1.03 \times 10^{-06}$	$1.60 \times 10^{-02}$
Curated_gene_sets:yague_pretumor_drug_resistance_up	7	1.77	0.034	0.39	$2.22 \times 10^{-06}$	$3.45 \times 10^{-02}$
GO_bp:go_phospholipid_transport	87	0.52	0.035	0.11	$2.83 \times 10^{-06}$	$4.39 \times 10^{-02}$

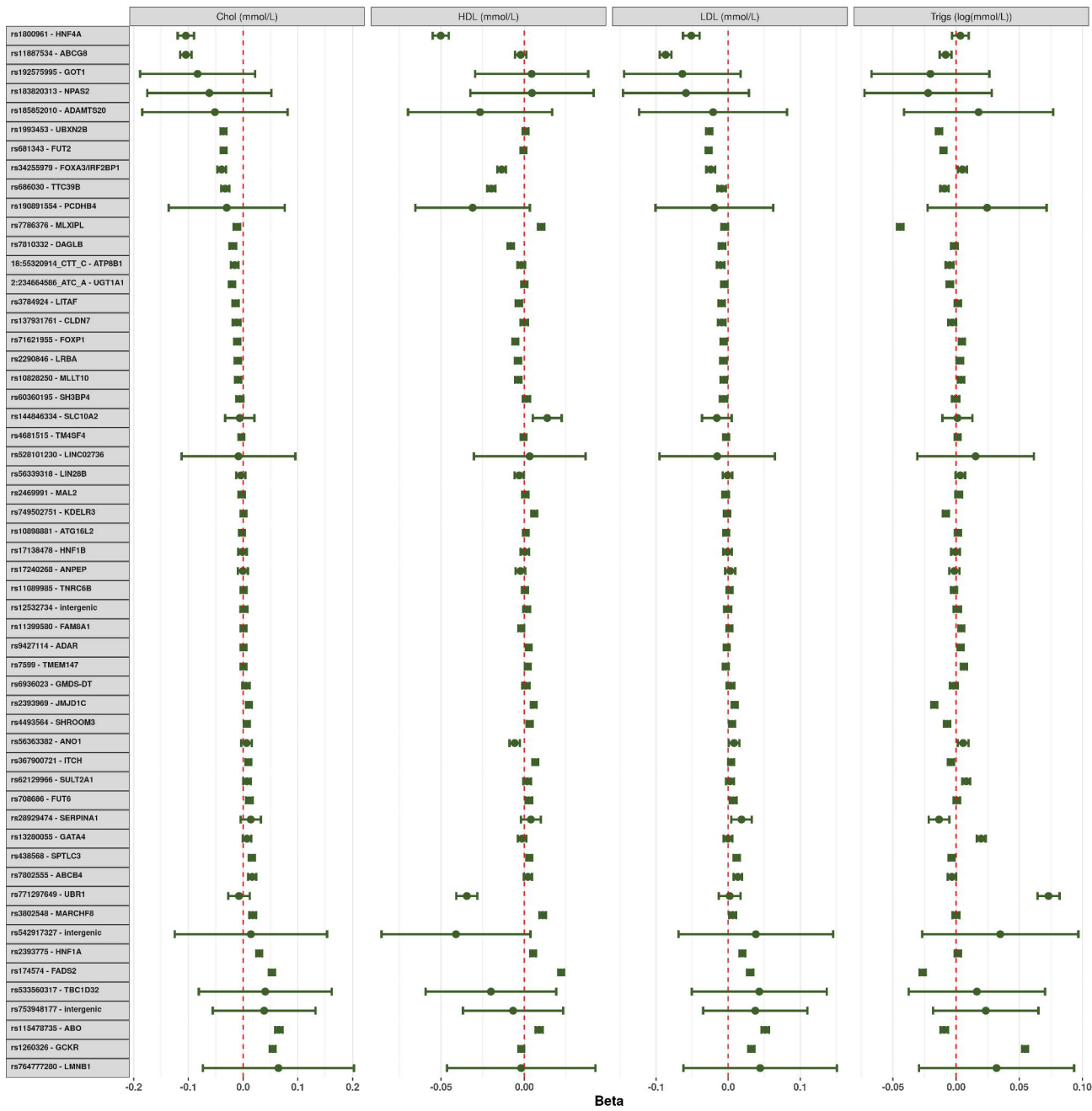
N Genes: the number of genes in the gene-set. Beta: the regression coefficient of the gene set / gene covariate, Beta STD: the semi-standardized regression coefficient, corresponding to the predicted change in Z-value given a change of one standard deviation in the predictor gene set / gene covariate, SE: the standard error of the regression coefficient, P: the competitive gene-set P value, Pbon: the empirically-corrected P value corrected for all gene-sets tested.







SNP



# Gallstone Disease: OR (95% CI, p-value)

Polygenic Risk Score Decile	OR (95% CI, p-value)
1	1.00 (reference)
2	1.30 (1.20-1.40, p<0.001)
3	1.52 (1.41-1.63, p<0.001)
4	1.76 (1.64-1.89, p<0.001)
5	2.00 (1.86-2.15, p<0.001)
6	2.34 (2.18-2.51, p<0.001)
7	2.75 (2.57-2.94, p<0.001)
8	3.08 (2.88-3.30, p<0.001)
9	3.83 (3.59-4.09, p<0.001)
10	5.97 (5.60-6.36, p<0.001)

