Genes that cause severe liver disease in children also influence risk and severity of common liver conditions in adults

Mushi JM[1]*, Sharma PJ[1]*, Schofield A[1,2], Frontino AM[3], Andersson ER[4], Chen VL[5], Cordell HJ[6], Davies SP[2], Gupte GL[7], Hirschfield GM[8,9], Jeyaraj R[2], Jones DE[10], Mells GF[11], Oo YH[2], Sandford RN[12], Siminovitch KA[13,14], Xu J[15†], Zhu K[15], Mašek J[3,17], Trauner M[16], Mann JP[2,7#].

*Joint first authors
[#]Correspondence

Affiliations
[1] Medical school, College of medicine and health, University of Birmingham, UK.
[2] Department of Immunology & Immunotherapy, School of Infection, Inflammation and Immunology, College of Medicine and Health, University of Birmingham, UK.
[3] Department of Cell Biology, Faculty of Science, Charles University, Viničná 7, 128 00 Prague 2, Czech Republic.
[4] Department of Cell and Molecular Biology, Karolinska Institutet, Solna, Sweden
[5] Division of Gastroenterology and Hepatology, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, USA
[6] Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
[7] Liver Unit (including small bowel transplantation), Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK
[8] The Autoimmune and Rare Liver Disease Programme, Division of Gastroenterology and Hepatology, Toronto General Hospital, Toronto, ON, Canada
[9] Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[10] NIHR Newcastle Biomedical Research Centre, Newcastle University, Newcastle, UK
[11] Cambridge Liver Unit, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom
[12] Academic Department of Medical Genetics, University of Cambridge, Cambridge, United Kingdom.
[13] Departments of Medicine, Immunology and Medical Sciences, University of Toronto, Toronto, Ontario, Canada
[14] Mount Sinai Hospital, Lunenfeld-Tanenbaum Research Institute and Toronto General Research Institute, Toronto, Ontario, Canada
[15] Gilead Sciences Inc., Foster City, California, USA
[16] Division of Gastroenterology and Hepatology, Department of Internal Medicine III, Medical University of Vienna, Vienna, Austria
[17] Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czech Republic
[†]Affiliation at the time research was conducted.

Correspondence:
Dr. Jake P. Mann

j.p.mann@bham.ac.uk
Centre for Liver and gastroenterology research, Department of Immunology & Immunotherapy, School of Infection, Inflammation and Immunology, College of Medicine and Health, University of Birmingham, UK
+44 (0)121 414 3344

Funding
AS, SPD, YHO, and JPM: NIHR BRC Birmingham.
JPM is supported by grants from NIHR ACL (CL-2022-09-005), Royal Society (RG\R1\241312), BSPGHAN-Guts (BSPGHAN2023_01), ESPGHAN, Royal College Physicians Dame Sheila Sherlock Bursary, and Little Princess Trust (CCLGA 2024 10 Mann).
JM: Czech Science Foundation (25-15821S), PRIMUS/21/SCI/006 project funded by Charles University Grant Agency, MSCA Fellowships CZ - Charles University CZ.02.01.01/00/22_010/0002902

Conflicts of interest
KZ is an employee of Gilead Sciences Inc. JX was an employee of Gilead Sciences Inc. at the time the research was conducted.

Acknowledgements
We thank Gabriel Cuella Partida for his advice on statistical genetics. We also thank the Birmingham NIHR Biomedical Research Centre, Centre for Liver and Gastrointestinal Research, and University of Birmingham Imaging Facility for their support.

Abstract

Background and aims

Rare, pathogenic variants in multiple genes can cause severe liver disease, requiring transplantation in childhood, but it is unclear whether common variants in the same genes confer risk of liver disease in adults. Here, we aimed to establish whether 'monogenic' liver diseases are associated with a spectrum of liver phenotypes in adulthood.

Methods

We identified 99 genes where pathological mutations cause significant liver disease in children. For each, we used population-level data from over 1.5 million adults to identify associations with biomarkers of liver injury (e.g. ALT). These observations were validated using six external cohorts of adults with clinical liver disease and transcriptomics. Finally, we illustrated the importance of the JAG1-NOTCH pathway on the ductular reaction in adult liver disease using immunohistochemistry and transcriptomics.

Results

We identified 89 genome-wide ($p<5x10^{-8}$) associations with biomarkers of liver injury in 46% (45/99) of genes. Variants in *ABCC2*, *ADK*, *ASL*, *BCS1L*, *HFE*, and *SERPINA1* were also linked with presence of clinical liver disease in adults. For example, rs2862926C>T encoding p.Ile1324= in *ABCC2* (which causes Dubin-Johnson syndrome) was associated with lower ALT ($p=7.1x10^{-12}$) and lower risk of MASLD (p-FDR=0.03). Transcriptional expression of 30% of genes was associated with severity of MASLD. p.Pro871Arg in JAG1 causes a subtle form of Alagille syndrome with higher bilirubin ($9.7x10^{-10}$) and hypertension ($p=4.2x10^{-14}$). JAG1 and NOTCH2 were expressed in injured bile ducts but not adjacent unaffected ducts and transcriptional expression of JAG1 correlated with fibrosis stage in primary sclerosing cholangitis.

Conclusions

Pathways affected by severe childhood liver disease also influence risk and severity of liver disease in adulthood.

Introduction

Germline variants with major functional impact are known to cause rare, apparently monogenic diseases across all body systems. Such disorders (e.g. Alagille syndrome, Wilson disease, progressive familial intrahepatic cholestasis (PFIC)) are frequent indications for liver transplantation in children[1]. However, many conditions (e.g. Alagille syndrome) have wide clinical variation: with some mutations causing severe phenotypes and others that do not seem to be affected at all[2]. This may be due to incomplete penetrance, variable expressivity, and/or influences of environmental factors. Next generation sequencing studies in large population based cohorts have highlighted the gap between higher predicted genetic prevalence and observed clinical prevalence[3]. Some disorders (e.g. progressive familial intrahepatic cholestasis) show close genotype-phenotype correlation, with more severe protein-truncating mutations requiring earlier transplant[4]. Therefore, variants with minor impact on protein function may not manifest unless combined with additional environmental hits[5].

Large, population-scale sequencing studies have also led to the identification of milder phenotypes of monogenic disorders across a range of conditions[6]. Similarly, such studies have demonstrated that variants in HFE (e.g. C282Y) and SERPINA1 confer additional risk towards steatotic liver disease, consistent with an autosomal codominant method of inheritance[7,8]. However, it is not clear whether this principle extends to genes that cause ultra-rare monogenic liver diseases. Particularly as many disorders (e.g. *DCDC2*-associated neonatal sclerosing cholangitis) manifest a severe phenotype with biallelic mutations[9] and are classically believed not to have a phenotype in the haploinsufficient state, though recent data has started to challenge this[10].

Genetic testing for liver disease typically involves large, multigene panels. These often identify coding variants of uncertain significance for infants with neonatal cholestasis. Little is known about whether such variants confer risk later in life or have clinical relevance for the general population. In this study, we leveraged population-level data, patients with clinical liver disease, and functional experiments to show that 'monogenic' disorders show mild phenotypes in adulthood.

Methods

Identification of genes for inclusion

First, we identified a list of genes that are robustly associated with monogenic liver disease. Criteria for inclusion were: (1) the condition causes at least one of: moderate hepatic fibrosis, acute and/or recurrent liver failure, or primary liver cancer; (2) at least two separate case series of patients have been reported; (3) variants in a single gene are regarded as causal according to ACMG criteria[11] review on Online Mendelian Inheritance in Man (OMIM)[12]. Genes were excluded if they were part of a generalised hereditary cancer syndrome (e.g. *TP53*), or were an indication for liver transplantation due to systemic effects without any evidence of clinically significant liver disease (e.g. *PKU*).

To identify genes, first, we searched PubMed and Google Scholar for: ("genetic" or "hereditary" or "monogenic") AND ("liver" or "biliary" or "fibrosis" or "hepatoblastoma" or "hepatocellular"). Relevant articles were reviewed to identify genes for inclusion. Next, we searched databases of genetic disorders (National Organisation for Rare Diseases (https://rarediseases.org/) and OMIM[12]) for the above liver-related terms. Lastly, we reviewed the Genomics England targeted sequencing panels (https://panelapp.genomicsengland.co.uk/) for Ductal plate malformation, Mitochondrial liver disease, including transient infantile liver failure, Neonatal cholestasis, and Polycystic liver disease.

Gene-level association testing

Next, we determined the evidence for association between each of these 99 genes and liver-related traits at a population level. To achieve this, we used genetic-phenotype data from large population-level studies, specifically using the HuGE calculator[13] and GeneBass[14] gene burden testing. For each gene, we obtained HuGE scores for liver-related traits (and associated phenotypes) from the Common Metabolic Diseases Knowledge Portal (CMDKP, cmdkp.org, accessed February 2025). HuGE scores are calculated using the effect of both thousands of common and rare variants in (or near) each gene from multiple GWAS with sample sizes up to 1 million individuals, as described elsewhere in detail[13]. Evidence is graded from 'anecdotal' to 'compelling'.

GeneBass is a curated database of results from UK BioBank whole exome sequencing and methods used for burden testing are described in detail elsewhere[14]. In short, the effects of multiple variants are added together, based upon their predicted impact on protein function. Variants are classified into those predicted to cause loss of function (pLoF), missense mutations, and those without effect on amino acid sequence (synonymous). We extracted pLoF, missense, and synonymous gene burden testing (using SKAT-O p-values) from GeneBass for each of the genes studied, and then filtered for relevant traits.

Variant-trait associations

We then sought to identify variants in (or near (±50kb)) each gene that were associated with liver-related traits at genome-wide significance ($p<5x10^{-8}$). To do this, we obtained variant-trait associations from CMDKP, GeneBass, and OpenTargets Genetics[15]. UCSC LiftOver Tool[16] was used to convert from GRCh37 to GRCh38. We filtered for liver-related traits (including

biochemical indices of liver disease) and genome-wide significant variants. Where duplicates were identified (from different data sources), the association with the smallest p-value was kept.

Variant annotation

We annotated variants with functional predictions using Ensembl's variant effect predictor[17]. LocusZoom was used to illustrate locus-trait associations[18]. Variants were further annotated with expression-trait quantitative loci (eQTL) in liver tissue using data from Genotype-Tissue Expression (GTEx) Project via the EBI eQTL catalogue[19].

To determine whether causal variants were shared across traits (e.g. ALT and GGT), we performed colocalisation analysis using summary statistics from the largest available GWAS of liver enzymes[20] (as full bottom-line association results were not available from CMDKP). Summary statistics were standardised using MungeSumstats[21] and converted into variant call format using gwas2vcf[22] for analysis using gwasglue. Evidence of colocalisation due to a single causal variant was defined by $PP.H_4 \geq 0.8$[23].

We manually reviewed evidence from Open Targets Genetics[15] and FORGEdb[24] for every variant to determine whether the look-up gene was likely to be causal. For variants to be included as likely linked to the look-up gene required, at least one of: coding or intronic within the gene of interest, significant eQTL, in strong LD (R^2>0.8) with another variant that fulfilled the above criteria).

To determine the number of unique variants, we subsequently performed clumping using 1000 genomes phase 3[25] reference data with PLINK2[26] --clump-p1 1e-5 --clump-r2 0.05 --clump-kb 250. LeadSNPs were then taken forwards for subsequent validation.

External cohort validation

Next, we sought to replicate these genetic associations in separate genome-wide association studies of patients with clinically significant common liver diseases. We obtained summary statistics from six studies with cases of primary sclerosing cholangitis (PSC)[27], primary biliary cholangitis (PBC)[28], metabolic-dysfunction associated steatotic liver disease (MASLD)[29], and all-cause cirrhosis[7,30,31]. None of these studies were included in the discovery searches, though data on PSC and PBC are also represented at a gene-level using HuGe calculator, as described above. Variants were considered to have demonstrated replication where genome-wide significant variants had false-discovery rate (FDR (Benjamini-Hochberg)) adjusted p-fadr < 0.05 in at least one cohort with clinical liver disease.

LD-score regression heritability analysis

Next, we calculated linkage disequilibrium (LD)-score regression to assess the fraction of heritability of clinical liver disease (PSC, PBC, MASLD, and cirrhosis) that was accounted for by the 98 genes causing monogenic liver disease. We obtained summary statistics for PSC[27], PBC[28], MASLD[32], and cirrhosis[31], standardised using MungeSumstats and calculated LD-score regression (LDSC) heritability (h^2) using precomputed LD scores from 1000 Genomes

European ancestry samples, limited to HapMap3 variants with ldsc v2.0[33]. We subsequently repeated LDSC h^2 estimates after removing variants in or near (±50kb) the 98 genes of interest.

Immunohistochemistry

We assessed the protein-level expression in liver tissue from explants obtained at the time of transplant from patients with PSC, PBC, alcohol-related liver disease (ARLD), and MASLD, as well as healthy donor liver using immunohistochemistry. All liver tissue was obtained following informed consent, under the "Liver Inflammation" study (Research Ethics Committee reference: 18/WA/0214; IRAS reference: 223072). Fresh tissue was formalin-fixed and paraffin-embedded (FFPE). FFPE sections were dewaxed in xylene then rehydrated in graded alcohols before heat antigen retrieval using Tris (Vector, H-3301) for 30 minutes. Endogenous peroxidase activity was blocked using BLOXALL (Vector, SP-6000-100) for 15 minutes at room temperature. Non-specific binding was blocked for 20 minutes at room temperature (Merck Millipore, 20773). Primary antibodies were incubated overnight at 4°C. Primary antibodies used were: JAG1 (Cell Signalling, D4Y1R, at 1/200), NOTCH2 (Cell Signalling, D76A6, at 1/50), and GBA1 (Abcam, EPR5142, at 1/200). Secondary antibodies (MP-7451 and MP-7452, Vector) were incubated for 1 hour at room temperature followed by developing with ImmPACT DAB (Vector, SK-4105) for 5 minutes and counter-stained with Mayers haematoxylin.

Immunofluorescence

Multiplex immunofluorescence was also performed on human liver FFPE samples. Antigen retrieval and blockade was performed as above. Each of four primary antibodies (JAG1, NOTCH2, CK19 (Abcam, Ab220193, at 1/1000), and ACTA2 (□-smooth muscle actin, Abcam, ab5694)) were incubated at room temperature for 1 hour. This was followed by 30 minutes of HRP-conjugated secondary antibody, as above, then 10-minutes development with tyramide signal amplification reagents: fluorescein (Akoya, NEL741001KT), Cy3 (Akoya, NEL744001KT), AZDye594 (CCT-1542), Cy5 (Akoya, NEL745001KT). Antibodies were stripped using Vectaplex (Vector, VRK-1000) for 15 minutes + 15 minutes at room temperature. Then blocking was repeated before the next primary antibody. Autofluorescence was quenched with TruView (Vector, SP-8400-15) and slides mounted with Vectashield with DAPI (Vector, H-1800). Images were acquired on a Zeiss LSM880 Confocal Microscope.

JAG1 variant database and *in silico* saturation mutagenesis

We generated a database of all reported clinically significant JAG1 variants following identification of p.Pro871Arg in JAG1 as a variant of interest with genome-wide significant associations. We compiled all *JAG1* variants from LOVD[34], ClinVar[35], and the most comprehensive previously published database[36], and summarised their reported clinical phenotypes.

We sought to determine whether p.Pro871Arg was in a genomic region particularly susceptible to missense variants. Therefore, we obtained pathogenicity predictions from AlphaFold[37] for all possible amino acids in JAG1 and mapped these onto the predicted domain structure for JAG1 from UniProt[38] and Gilbert et al[36].

Used plasmid vectors and JAG1 p.Pro871Arg variant mutagenesis

Human JAGGED1 cDNA (Genescript), tagged with T2A-EGFP was used as a source vector to recapitulate the P871R mutation. Standard NEB protocol (https://www.neb.com/en/products/e0554-q5-site-directed-mutagenesis-kit) was used for the site-directed mutagenesis using the following primers: JAG1_P871R:1_FWD: GAGTGTGATACgAGATGGGGCCAAATGGGATGATGAC, and JAG1_P871R:1_REV: CCCATGGTGATGCAAGGTCTCCCTG. Whole hJAG1 P871R-T2A-EGFP plasmid sequencing was performed by Plasmidsaurus using Oxford Nanopore Technology with custom analysis and annotation. The following vectors were used in reporter assays: hJAG1P871R-T2A-EGFP in pcDNA3.1 (this work), hJAG1-T2A-EGFP in pcDNA3.1, hNOTCH2-T2A 670 in pcDNA3.1, pcDNA3.1 (all Genescript), PBSKII+ Bluescript (Kind gift from Zbynek Kozmik), 12XCSL luciferase reporter, Renilla luciferase (kind gift from Urban Lendahl).

In vitro functional assessment of JAG1 p.Pro871Arg

To determine whether p.Pro871Arg in JAG1 affects its ability to bind NOTCH2 receptor, we expressed this variant in HEK 293 Low Notch Activity (LNA) cells which were genetically modified to be deficient for the following Notch signaling components *NOTCH1/2/3*$^{-/-}$, *JAGGED1/2*$^{-/-}$, *DLL1*$^{-/in\ frame\ del}$, *DLL3*$^{-/in\ frame\ del}$, *DLL4*$^{+/-}$, as previously described (Buhl *et al.*, in submission). LNA HEK 293 cells were cultured in DMEM (High glucose, Glutamax supplement, pyruvate, Thermo Fisher #31966047), supplemented with 10 % fetal bovine serum (FBS, Gibco™, #A5256701), and 1 % of penicillin/streptomycin (Gibco™, #15140122) at 37°C in a humidified 5% CO□ incubator. On Day 1, cells were seeded into 24-well plates at a density of 1.0–1.25 × 10□ cells per well in 1 mL of media. Transient transfections were performed 24h after seeding using Lipofectamine™ 2000 (Thermo Fisher, #11668019) in Opti-MEM™ medium (Thermo Fisher, #31985062) following the manufacturer's protocol. LNA HEK 293 cells were transfected with 100 ng of hJAG1 (WT or P871R), control pcDNA3.1 plasmid, hNOTCH2-T2A 670, 200 ng of 12xCSL luciferase reporter, and 8 ng of Renilla luciferase. DNA mix, 100ng/well, topped up to 1ug/well with PBSKII+ Bluescript, were first incubated for 20min/RT with Lipofectamine in Opti-MEM™ medium and added dropwise to the wells. Transfected cells were detached 24h after transfection with TrypLE™ Express Enzyme (1X) phenol red (Thermo Fisher, #12605028), resuspended in DMEM, and counted. Ligand-expressing cells (Jag1 WT or mutant) were mixed with Notch2 receptor-expressing cells at a 4:1 ratio, ensuring 40,000 ligand cells and 10,000 receptor cells per well. Co-cultures were plated in 96-well plates in triplicate with 200 µL DMEM per well and incubated for 36–48 hours. On Day 4, 12xCSL reporter activity was assessed using the Dual-Glo® Luciferase Assay System (Promega, #E2940), following the manufacturer's instructions using a Varioskan™ plate reader.

Gene expression in MASLD

Transcriptional expression in candidate genes was assessed in GSE135251[39], which contains RNA sequencing data from 216 individuals across the spectrum of MASLD (control to F4). Raw counts were filtered for low expressed genes then differential gene expression was performed between patients with MASLD and no fibrosis (MASL F0) and steatohepatitis with cirrhosis (MASH F4) with Benjamini–Hochberg adjusted p-values <0.05 as significant. 94 out of 99 candidate genes were available for analysis in the dataset. Violin plots demonstrate normalised

counts. In addition, we performed one-way ANOVAs to look for any variation between groups in candidate genes, with Bonferroni-adjusted p-values (p.adj) <0.05 as significant.

JAG1 transcript expression in PSC

Patients with PSC who were recruited to the phase 2b trial of simtuzumab (NCT01672853)[40] had RNA extracted from FFPE baseline liver biopsies. RNA with DV200 >10% underwent cDNA synthesis using Illumina TruSeq RNA followed by paired-end 150bp read sequencing on Illumina NovaSeq 6000. Only data that passed quality control (Qscore, AT/GC distribution, and error rate distribution) were subject to further analysis. FASTQ files had adapter sequences removed using cutadapt[41] and were then aligned to GRCh38 human reference genome using STAR[42]. Samtools[43] and featureCounts were used for sorting and counting. Log_2 transcripts per million (TPM) were used to compare JAG1, NOTCH1, HEYL, and HES expression between patients with mild (Ishak 0-2), moderate (Ishak 3-4), and severe fibrosis (Ishak 5-6) using Kruskal-Wallis test. These results were replicated using the SteatoSITE MASLD transcriptomic dataset[44] available from https://shiny.igc.ed.ac.uk/SteatoSITE_gene_explorer/.

Statistical analysis

Integration of datasets was performed using R base functions and tidyverse. Genome-wide significance ($p<5x10^{-8}$) was used for initial identification of variant-trait associations with p<0.05 adjusted for multiple testing used throughout otherwise. All analyses were performed using R. Code is available from https://github.com/jmann01/MonogenicLiver.

Results

Gene-level associations

We identified 99 monogenic disorders that cause significant liver disease (Fig 1A & SupTab 1). Of these 63% (62/99) had at least 'moderate' evidence of association with one or more population-level liver-related traits using the HuGE calculator (Fig 1B) whilst 9% (8/99) were associated with more than three traits with 'compelling' evidence (Fig 1D & SupTab 2). This replicated well-established associations between *SERPINA1* (which causes alpha-1-antitrypsin deficiency) and *HFE* (which causes haemochromatosis) on population-level ALT levels (Fig 1C). Genes that cause ductal plate malformations (e.g. *PKD1* and *PKHD1*) clustered together with genes that cause progressive familial intrahepatic cholestasis (e.g. *NR1H4*) due to associations with serum liver enzymes. Multiple genes that perturb hepatocyte (e.g. LIPA, TRMU) and biliary function (e.g. ABCB11) also caused abnormalities in lipid profile and 20% (19/99) influenced markers of insulin resistance (Fig 1E).

These findings were replicated using gene burden testing with data from the UK BioBank: 60% (59/99) demonstrated at least one association (at p-FDR<0.05, SupFig 1A-B & SupTab 3). Similar patterns of gene-trait associations were demonstrated across pLoF, missense, and synonymous burden testing (SupFig 1C-E). Therefore, these observations were not exclusively driven by severe protein-truncating mutations. Overall, the majority of genes that cause severe monogenic liver disease were also associated with liver-related traits at a population level.

Variant-level associations

We identified 89 independent variants associated with liver-related traits in 46% (45/99) of genes that cause 'monogenic' liver disease (Table 1 and SupTab 4). 83% (74/89) were common variants (>1% effect allele frequency). Over 40% of look-up genes were associated with differences in GGT or ALP at a population level (Fig 1F).

Genome-wide significant associations were observed across all studied traits and from multiple different genes (Fig 2A). This replicated well-established associations between variants in *UGT1A1* (which causes Crigler-Najjar syndrome) and bilirubin. Where genes were associated with more than one phenotype (e.g. *ABCB4* were associated with GGT and ALT), colocalisation analyses demonstrated that this was likely due to a single causal variant (Fig 2B & SupTab 5) . For example, a single locus within *LIPA* was associated with ALT, triglycerides, and bilirubin (Fig 2D).

33% (29/89) were protein-coding variants in the gene of interest. For the non-coding variants, 55% (33/60) had a significant eQTL for the gene of interest (Fig 2D). Overall, this demonstrates that common variants in genes that cause monogenic liver disease are also strongly associated with indices of liver injury.

Influence on clinical liver disease

Next, we sought to validate our observations in independent cohorts of adults with clinically significant liver disease.
Using six GWAS in patients with liver disease we established that variants in six genes (*ABCC2*, *ADK*, *ASL*, *BCS1L*, *HFE*, and *SERPINA1*) were associated with at least one clinical liver disease (SupTab 6 & Table 1). This included the well-established all-cause cirrhosis risk variants in *SERPINA1* and *HFE*[30,31]. In addition, rs2633311C>T near ADK (with a blood eQTL for ADK) was negatively associated with GGT ($p=7.0x10^{-13}$) and presence of PBC (p-FDR=.04). However, collectively, these gene regions did not account for a meaningful proportion of PBC, PSC, MASLD, or cirrhosis heritability (SupTab 7).

Next, we determined whether expression in these genes changes with the spectrum of liver disease using n=216 individuals with MASLD. 38% of genes (26/94) demonstrated variance with severity of MASLD (Fig 3 & SupTab 8). Specifically, 30% (28/94) showed differential expression between patients with MASL and MASH cirrhosis. This included upregulation and downregulation of genes across multiple different pathways (e.g. *DCDC2*, *HNF1B*, *HJV*). The top differentially expressed gene was *JAG1* (Fig 3E, log2FC 1.6, p.adj = $2.9x10^{-13}$).

JAG1-NOTCH dysfunction in chronic liver disease
Our analyses found multiple variants in *JAG1* and *NOTCH2* (both of which cause Alagille syndrome) to be associated with indices of liver injury in adults. In particular, we identified p.Pro871Arg (rs35761929G>C) as a common variant (EAF 5.3%) that, at a population-level, replicates an Alagille-like phenotype, including elevated markers of biliary epithelial cell injury, renal function (blood urea nitrogen $p=4.0x10^{-11}$), and hypertension (diastolic blood pressure, $p=4.2x10^{-14}$) (SupFig 2A). We curated a database of 1155 *JAG1* variants (SupTab 9) and did not find JAG1 p.Pro871Arg to be described as pathogenic for Alagille syndrome. Position 871 variant lies in the cysteine-rich domain of JAG1 (Fig 4A & SupFig 2B-D), which is comparatively tolerant of missense mutations from *in silico* saturation mutagenesis. This is in contrast to the DSL domain, terminal EGF-like repeats 15-16, and the trans-membrane domain (SupFig 2C). We also replicated the observation that JAG1 is relatively intolerant of cysteine, proline, and tryptophan, missense mutations[36].

AlphaFold predicts JAG1 p.Pro871Arg to be a well-tolerated variant (SupFig 2D), though it is predicted as deleterious by SIFT. However, when we expressed wild-type and p.Pro871Arg in human cells null for JAG1 and NOTCH1-4 we observed no difference between wild-type and p.Pro871Arg in ability to activate NOTCH2 signalling(Fig 4B). This illustrates that even coding variants identified in this study have subtle effects on signalling pathways, distinct from loss of function pathogenic variants.

Next, we studied how JAG1-NOTCH2 signalling may influence adult liver disease. JAG1 is expressed in arterial structures in healthy adult liver (Fig 4C). However, it becomes strongly expressed in regenerative biliary ductules as part of the ductular reaction, most evident in cholangiopathies but also in ARLD (Fig 4C), as previously reported[45]. NOTCH2 was not readily detectable in healthy liver but widely expressed throughout fibrotic areas and reactive ductules in cirrhosis (Fig 4C). Using multiplex immunohistochemistry, we found that JAG1 and

NOTCH2 are co-expressed on injured biliary epithelial cells (CK19^{+} SMA^{+}) in PSC and PBC but not on adjacent unaffected ducts (Fig 4D). In addition, NOTCH2 was expressed on a subset of SMA^{+} fibroblasts within fibrotic regions (Fig 4D).

Finally, we found that the intensity of expression of *JAG1* (Fig 4E) and *NOTCH* target gene *HEYL* (Fig 4F) correlated with fibrosis stage in PSC. Transcriptional expression of these genes also correlated with non-invasive markers of disease severity (SupFig 3).

Discussion

Since the widespread use of next generation sequencing, it has become clear that onset and severity of acquired liver diseases (e.g. steatotic liver disease, viral hepatitis) are influenced by germline variation[31,46]. However, these were generally thought to be distinct from the pathways perturbed by rare, loss-of-function variants that lead to early onset severe phenotypes. Here, we have demonstrated that common variation in such genes impacts the onset and severity of liver disease in adults. Understanding the pathways implicated in rare, paediatric liver disease may lead to novel therapeutic targets[47] for the top causes of transplantation in adults.

It has long been recognised that there is wide phenotypic variation in apparently monogenic disorders[2]. For example, it is not uncommon to have a child with Alagille syndrome affected with profound cholestasis whilst their parent is asymptomatic, despite both harbouring the same variant. Therefore, such pathways are subject to gene-environment interactions, as are genes not linked to severe liver disease (e.g. *PNPLA3*)[48]. JAG1-NOTCH signalling is a key component of the ductular reaction, being specifically expressed in injured bile ducts, as well as playing a role in liver regeneration in mice[45]. Collectively, the JAG1-NOTCH signalling is one example of how a pathway typically thought of to cause exclusively rare liver disease in early childhood is functionally and prognostically important in adult liver disorders.

The concept of 'monogenic' disease has been progressively eroded across liver disease[5,10] and other fields[49]. For example, variation in genes that cause monogenic developmental disorders cause more mild phenotypes of the same conditions in the general population[50]. Similarly, for coronary artery disease, genome-wide polygenic risk scores have shown that individuals in the top decile of polygenic risk have a disease incidence comparable to that of 'monogenic' familial hypercholesterolaemia[51]. This study has illustrated that a similar principle applies for liver disease: the phenotype that affects patients is a combination of polygenic risk (from both rare and common variants) plus environmental factors, even for conditions typically thought of as exclusively genetic or acquired.

More broadly, we have demonstrated that looking for variants associated with genes that cause severe disease is an effective method for variant prioritisation from genome-wide association studies.

Identification of such effects requires large studies to demonstrate small effect sizes. Therefore, it is hard to quantify how these observations influence the phenotype of common liver diseases at an individual patient-level. Our heritability analyses suggest that these genes are not critical germline determinants for development of adult liver diseases, though it is harder to assess how they impact on the disease progression. Also, despite use of a sensitive *in vitro* system, we did not demonstrate any significant effect of a JAG1 variant with strong human genetic evidence. Similarly, we observed several variants to reduce the risk of liver diseases, the mechanism of which has not been identified at this time Therefore to elucidate gene- and variant-specific mechanisms will require careful studies with cellular stressors.

Conclusion

The majority of genes that cause severe paediatric liver disease are associated with liver injury in adults. The JAG1-NOTCH pathway is a key component of the ductular reaction that is affected by common variation and associated with disease severity in cirrhosis. These findings challenge the notion of 'monogenic' liver disease and underscore the potential to pharmacologically target the pathways implicated in rare liver diseases for treatment of common diseases in adults.

Figure legends

Fig 1. Genes that cause monogenic liver disease are associated with liver-related traits at a population level using HuGE data. Summary of the study design (A). Proportion of genes with at least 'moderate' evidence for liver-related traits (B). Proportion of genes with each level of evidence for hepatic- and related-traits (C). Heatmaps showing HuGE score and level of evidence for liver-related (D) and non-liver traits (E). Levels of evidence: * = Moderate, ** = Strong, *** = Very strong, **** = Compelling. Bar chart showing the proportion of genes with at least one genome-wide significant variant ($p<5x10^{-8}$) associated with each trait (F). AST, aspartate aminotransferase; BMI, body mass index; CKD, chronic kidney disease; HDL, high-density lipoprotein; ICP, intrahepatic cholestasis of pregnancy; LDL, low density lipoprotein; PSC, primary sclerosing cholangitis; T2DM, type 2 diabetes mellitus; TG, serum triglycerides; UC, ulcerative colitis; WHRadjBMI, waist-to-hip ratio-adjusted BMI.

Fig 2. Common variants in genes that cause monogenic disease are associated with liver-related traits. Dot plot showing individual variants' $log_{10}(-log_{10}(p\text{-values}))$ for liver-related traits, with annotation of top genes (A). LocusZoom plots from Common Metabolic Disease Knowledge Portal Data demonstrating associations for *ABCB4* (B), *HNF1B* (C), *ATP8B1* (D), and *CPS1* (E).

Fig 3. Expression of genes that cause monogenic disease is associated with severity of MASLD. Violin plots showing normalised gene expression across the spectrum of MASLD (n=216) from GSE135251 for *DCDC2* (A), *HJV* (B), *CYP27A1* (C), *GALK1* (D), *JAG1* (E), *NOTCH2* (F), *AKR1D1* (G), and *HNF1B* (H). p.adj represent differential gene expression between MASL and MASH F4 groups.

Fig 4. JAG1-NOTCH expression on injured bile ducts affects severity of PSC. Domain structure of JAG1, illustrating the position of p.Pro871Arg (A). Ability of wild-type (WT) or p.Pro871Arg (P871R) to induce NOTCH-target luciferase reporter signalling. Expression of JAG1 and NOTCH2 in normal liver, PSC, PBC, and ARLD on immunohistochemistry (C, representative of n=5 patients per condition). Multiplex immunofluorescence showing expression of JAG1, NOTCH2, □SMA (ACTA2), and CK19, in human liver tissue (D, representative of n=5 patients per condition). Association of *JAG1* transcripts with fibrosis stage in PSC (E) and MASLD (F).

Table legends

Table 1. Top variant-trait associations in genes that also cause monogenic liver disease. Variants are annotated with: predicted protein functional effect (from Ensembl's Variant Effect Predictor), liver tissue eQTL data from GTEx, effect allele frequency (EAF) from gnomAD, and results from replication across six cohorts of patients with clinical liver disease.

Supplementary data

SupTab 1. Genes included in this study and their respective disorders with Online Mendelian Inheritance in Man (OMIM) identifiers.

SupTab 2. HuGE scores and evidence levels for all genes and traits included in this study.

SupTab 3. GeneBass burden testing (pLoF, missense, and synonymous) for all genes and traits included in this study.

SupFig 1. Genes that cause monogenic liver disease are associated with liver-related traits at a population level using gene burden testing from GeneBass (A). Proportion of genes with at least one association at p-fdr<0.05 on burden testing for liver-related traits (B). Stacked bar charts showing the proportion of genes with each level of significance by traits for predicted loss of function (pLoF (C)), missense (D), and synonymous (E) variants. Correlation heatmaps showing association between traits for pLoF (F), missense (G), and synonymous (H) burden testing.

SupTab 4. Summary of all variant-trait associations identified in this study. Variants are annotated with: predicted protein functional effect (from Ensembl's Variant Effect Predictor), liver tissue eQTL data from GTEx.

SupTab 5. Results from colocalisation analyses using data from Sakaue et al. for plots displayed in Figure 2B.

SupTab 6. Summary of all results from validation of variants in external cohorts.

SupTab 7. Summary of results from LDSC heritability analyses. For each dataset, LDSC heritability was calculated for all variants and then after removal of regions (±50kb) around the 98 genes that cause monogenic liver disease.

SupTab 8. Results from ANOVA and F0 vs F4 from GSE135251: transcriptomic profiling across the spectrum of MASLD.

SupTab 9. A table of all previously reported pathogenic or likely pathogenic variants in *JAG1*, with their reported clinical severity.

SupFig 2. Phenome-wide association plot for JAG1 p.Pro871Arg (A). Domain structure of JAG1 compared with a heatmap showing the effect of missense mutations on protein stability from AlphaFold (B). Zoomed-in heatmap showing *in silico* mutagenesis from AlphaFold around the area of position 871 in JAG1.

SupFig 3. Hepatic expression of JAG1, NOTCH1, HEYL, and HES1 in patients with PSC grouped by Ishak fibrosis stage, collagen deposition, ELF score, and ALP level. Whiskers extend to the minimum and maximum values of the data set. p values are based on the

Kruskal–Wallis test. ALP, alkaline phosphatase; ELF, enhanced liver fibrosis; TPM, transcripts per million.

References

[1] Baumann U, Karam V, Adam R, Fondevila C, Dhawan A, Sokal E, et al. Prognosis of Children Undergoing Liver Transplantation: A 30-Year European Study. Pediatrics 2022;150. https://doi.org/10.1542/peds.2022-057424.

[2] Kamath BM, Baker A, Houwen R, Todorova L, Kerkar N. Systematic review: The epidemiology, natural history, and burden of alagille syndrome. J Pediatr Gastroenterol Nutr 2018;67:148–56.

[3] Gao J, Brackley S, Mann JP. The global prevalence of Wilson disease from next-generation sequencing data. Genet Med 2019;21:1155–63.

[4] van Wessel DBE, Thompson RJ, Gonzales E, Jankowska I, Sokal E, Grammatikopoulos T, et al. Genotype correlates with the natural history of severe bile salt export pump deficiency. J Hepatol 2020;73:84–93.

[5] Nayagam JS, Foskett P, Strautnieks S, Agarwal K, Miquel R, Joshi D, et al. Clinical phenotype of adult-onset liver disease in patients with variants in ABCB4, ABCB11, and ATP8B1. Hepatol Commun 2022;6:2654–64.

[6] Peltonen L, Perola M, Naukkarinen J, Palotie A. Lessons from studying monogenic disease for common disease. Hum Mol Genet 2006;15 Spec No 1:R67-74.

[7] Chen VL, Chen Y, Du X, Handelman SK, Speliotes EK. Genetic variants that associate with cirrhosis have pleiotropic effects on human traits. Liver Int 2020;40:405–15.

[8] Schwantes-An T-H, Darlay R, Mathurin P, Masson S, Liangpunsakul S, Mueller S, et al. Genome-wide Association Study and Meta-analysis on Alcohol-Associated Liver Cirrhosis Identifies Genetic Risk Factors. Hepatology 2021;73:1920–31.

[9] Grammatikopoulos T, Sambrotta M, Strautnieks S, Foskett P, Knisely AS, Wagner B, et al. Mutations in DCDC2 (doublecortin domain containing protein 2) in neonatal sclerosing cholangitis. J Hepatol 2016;65:1179–87.

[10] Hegarty R, Gurra O, Tarawally J, Allouni S, Rahman O, Strautnieks S, et al. Clinical outcomes of ABCB4 heterozygosity in infants and children with cholestatic liver disease. J Pediatr Gastroenterol Nutr 2024;78:339–49.

[11] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–24.

[12] Amberger JS, Bocchini CA, Scott AF, Hamosh A. Omim.org: leveraging knowledge across phenotype-gene relationships. Nucleic Acids Res 2019;47:D1038–43.

[13] Dornbos P, Singh P, Jang D-K, Mahajan A, Biddinger SB, Rotter JI, et al. Evaluating human genetic support for hypothesized metabolic disease genes. Cell Metab 2022;34:661–6.

[14] Karczewski KJ, Solomonson M, Chao KR, Goodrich JK, Tiao G, Lu W, et al. Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. Cell Genom 2022;2:100168.

[15] Ghoussaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res 2021;49:D1311–20.

[16] Perez G, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2025 update. Nucleic Acids Res 2025;53:D1243–9.

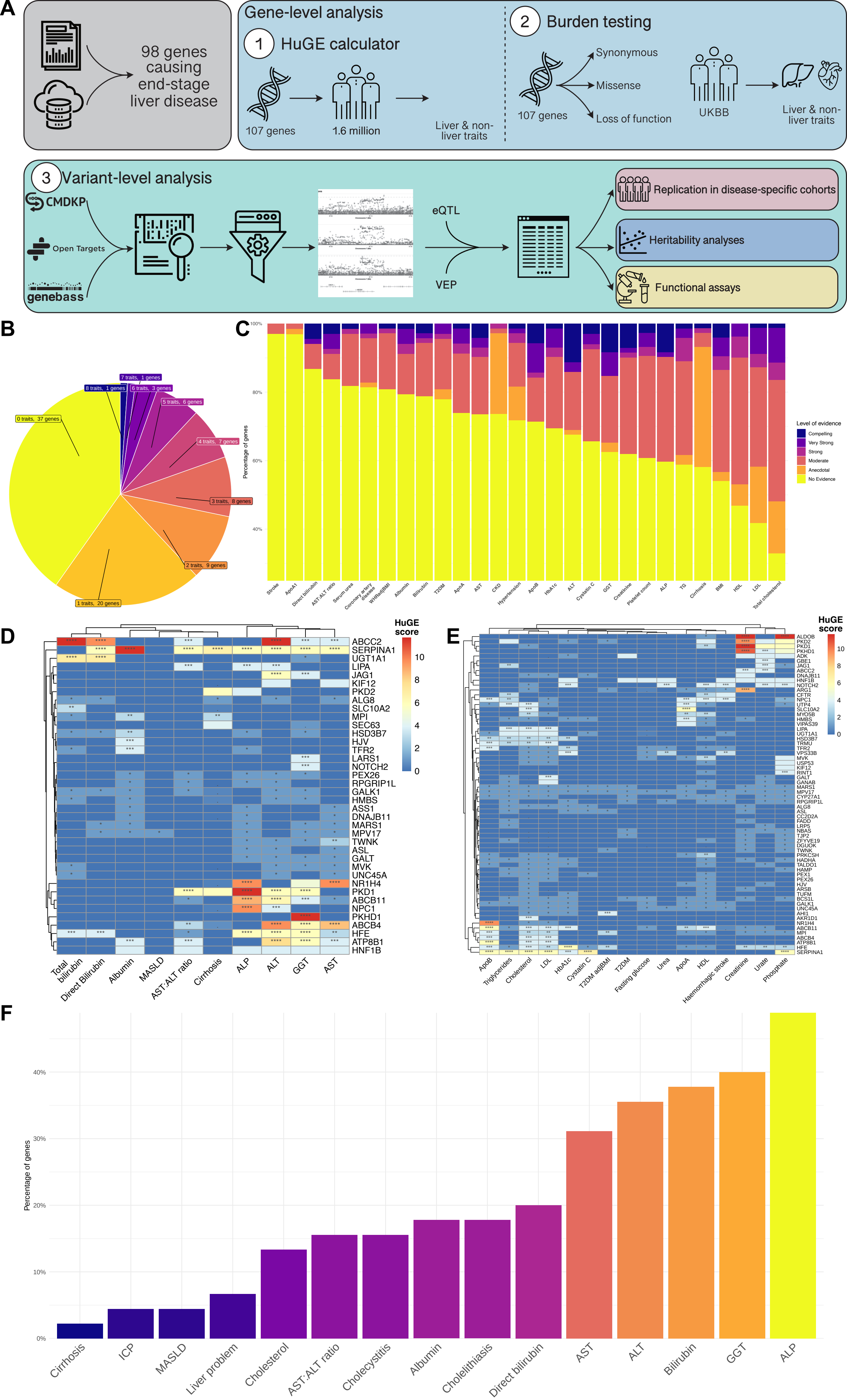
[17] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;17:122.

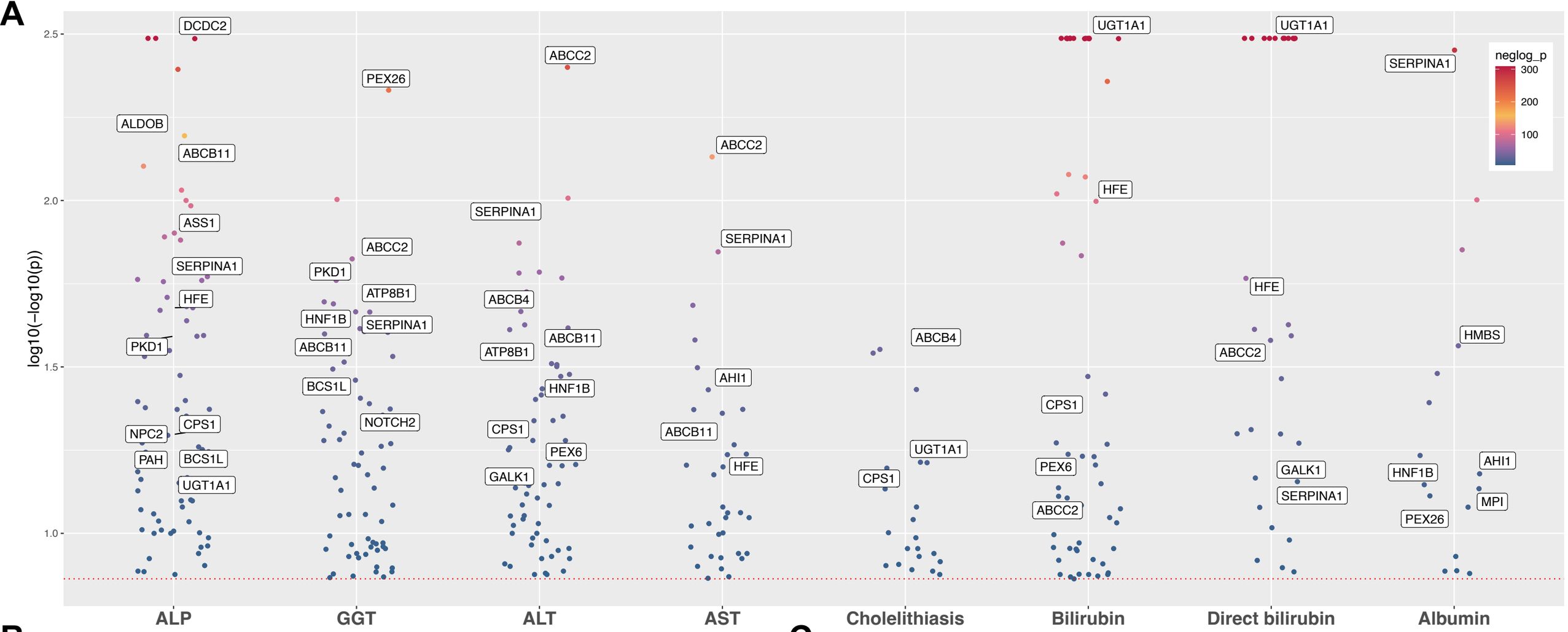
[18] Boughton AP, Welch RP, Flickinger M, VandeHaar P, Taliun D, Abecasis GR, et al. LocusZoom.js: interactive and embeddable visualization of genetic association study results. Bioinformatics 2021;37:3017–8.

[19] Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, et al. A

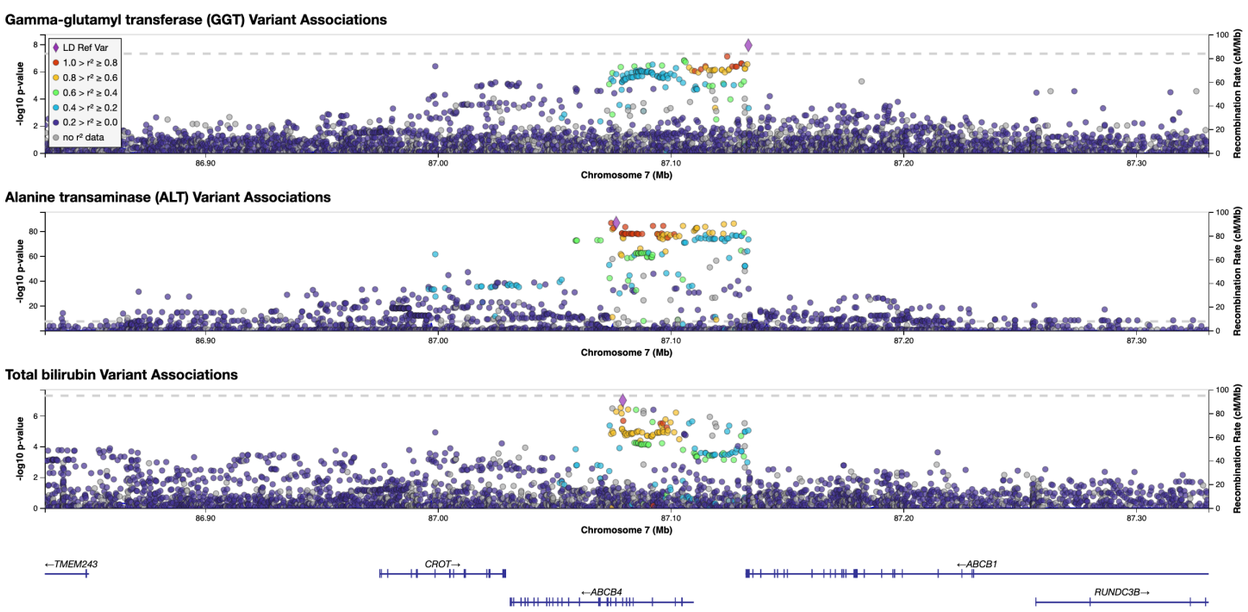
compendium of uniformly processed human gene expression and splicing quantitative trait loci. Nat Genet 2021;53:1290–9.
[20] Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet 2021;53:1415–24.
[21] Murphy AE, Schilder BM, Skene NG. MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics. Bioinformatics 2021;37:4593–6.
[22] Lyon MS, Andrews SJ, Elsworth B, Gaunt TR, Hemani G, Marcora E. The variant call format provides efficient and robust storage of GWAS summary statistics. Genome Biol 2021;22:32.
[23] Guo H, Fortune MD, Burren OS, Schofield E, Todd JA, Wallace C. Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. Hum Mol Genet 2015;24:3305–13.
[24] Breeze CE, Haugen E, Gutierrez-Arcelus M, Yao X, Teschendorff A, Beck S, et al. FORGEdb: a tool for identifying candidate functional variants and uncovering target genes and mechanisms for complex diseases. Genome Biol 2024;25:3.
[25] The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 2015;526:68–74.
[26] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7.
[27] Ji S-G, Juran BD, Mucha S, Folseraas T, Jostins L, Melum E, et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. Nat Genet 2017;49:269–73.
[28] Cordell HJ, Fryett JJ, Ueno K, Darlay R, Aiba Y, Hitomi Y, et al. An international genome-wide meta-analysis of primary biliary cholangitis: Novel risk loci and candidate drugs. J Hepatol 2021;75:572–81.
[29] Chen Y, Du X, Kuppa A, Feitosa MF, Bielak LF, O'Connell JR, et al. Genome-wide association meta-analysis identifies 17 loci associated with nonalcoholic fatty liver disease. Nat Genet 2023;55:1640–50.
[30] Emdin CA, Haas M, Ajmera V, Simon TG, Homburger J, Neben C, et al. Association of Genetic Variation With Cirrhosis: A Multi-Trait Genome-Wide Association and Gene-Environment Interaction Study. Gastroenterology 2021;160:1620-1633.e13.
[31] Ghouse J, Sveinbjörnsson G, Vujkovic M, Seidelin A-S, Gellert-Kristensen H, Ahlberg G, et al. Integrative common and rare variant analyses provide insights into the genetic architecture of liver cirrhosis. Nat Genet 2024;56:827–37.
[32] Ghodsian N, Abner E, Emdin CA, Gobeil É, Taba N, Haas ME, et al. Electronic health record-based genome-wide meta-analysis provides insights on the genetic architecture of non-alcoholic fatty liver disease. Cell Rep Med 2021;2:100437.
[33] Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47:291–5.
[34] Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. Hum Mutat 2011;32:557–63.
[35] Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 2014;42:D980-5.
[36] Gilbert MA, Bauer RC, Rajagopalan R, Grochowski CM, Chao G, McEldrew D, et al. Alagille syndrome mutation update: Comprehensive overview of JAG1 and NOTCH2 mutation frequencies and insight into missense variant classification. Hum Mutat

2019;40:2197–220.
[37] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9.
[38] UniProt Consortium. UniProt: The universal protein knowledgebase in 2025. Nucleic Acids Res 2025;53:D609–17.
[39] Govaere O, Cockell S, Tiniakos D, Queen R, Younes R, Vacca M, et al. Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. Sci Transl Med 2020;12:eaba4448.
[40] Muir AJ, Levy C, Janssen HLA, Montano-Loza AJ, Shiffman ML, Caldwell S, et al. Simtuzumab for Primary Sclerosing Cholangitis: Phase 2 Study Results With Insights on the Natural History of the Disease. Hepatology 2019;69:684–98.
[41] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetJournal 2011;17:10–2.
[42] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.
[43] Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience 2021;10. https://doi.org/10.1093/gigascience/giab008.
[44] Kendall TJ, Jimenez-Ramos M, Turner F, Ramachandran P, Minnier J, McColgan MD, et al. An integrated gene-to-outcome multimodal database for metabolic dysfunction-associated steatotic liver disease. Nat Med 2023;29:2939–53.
[45] Boulter L, Govaere O, Bird TG, Radulescu S, Ramachandran P, Pellicoro A, et al. Macrophage-derived Wnt opposes Notch signaling to specify hepatic progenitor cell fate in chronic liver disease. Nat Med 2012;18:572–9.
[46] Tamaki N, Ahlholm N, Luukkonen PK, Porthan K, Sharpton SR, Ajmera V, et al. Risk of advanced fibrosis in first-degree relatives of patients with nonalcoholic fatty liver disease. J Clin Invest 2022;132. https://doi.org/10.1172/JCI162513.
[47] Brinkman RR, Dubé M-P, Rouleau GA, Orr AC, Samuels ME. Human monogenic disorders - a source of novel drug targets. Nat Rev Genet 2006;7:249–60.
[48] Stender S, Kozlitina J, Nordestgaard BG, Tybjærg-Hansen A, Hobbs HH, Cohen JC. Adiposity amplifies the genetic risk of fatty liver disease conferred by multiple loci. Nat Genet 2017:EPub 24 April 2017.
[49] Jacob E, Hegele RA. Monogenic versus polygenic forms of hypercholesterolemia and cardiovascular risk: Are there any differences? Curr Atheroscler Rep 2022;24:419–26.
[50] Kingdom R, Beaumont RN, Wood AR, Weedon MN, Wright CF. Genetic modifiers of rare variants in monogenic developmental disorder loci. Nat Genet 2024;56:861–8.
[51] Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;50:1219–24.

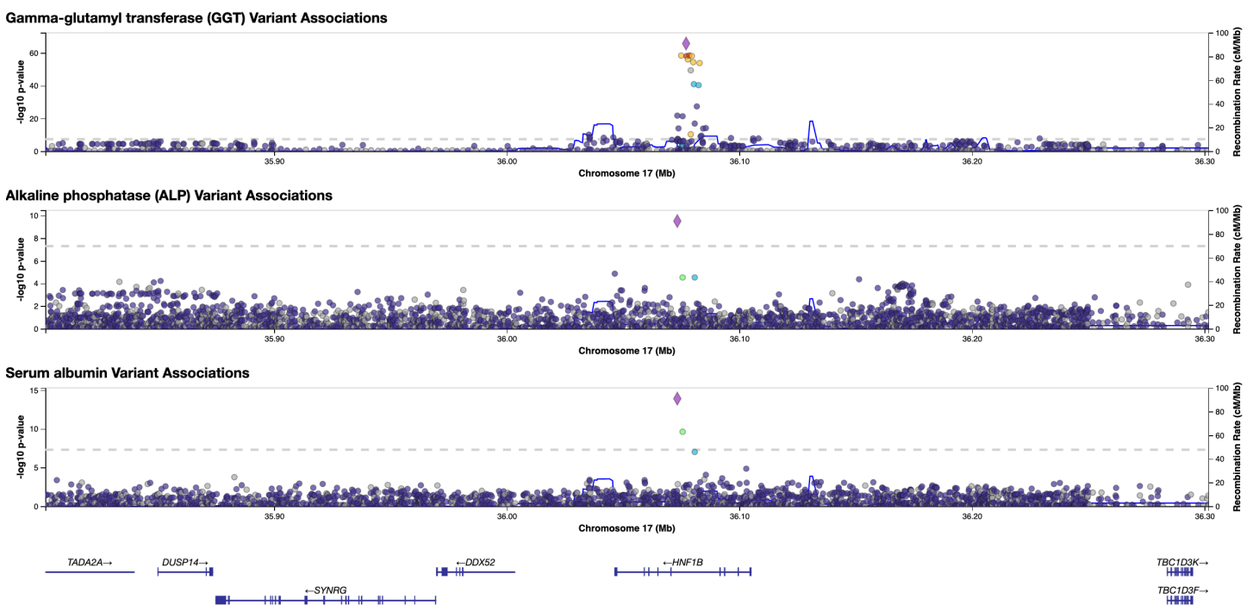

A
98 genes causing end-stage liver disease
Gene-level analysis
1 HuGE calculator
107 genes
1.6 million
Liver & non-liver traits
2 Burden testing
107 genes
Synonymous
Missense
Loss of function
UKBB
Liver & non-liver traits
3 Variant-level analysis
CMDKP
Open Targets
genebass
eQTL
VEP
Replication in disease-specific cohorts
Heritability analyses
Functional assays
B
7 traits, 1 genes
8 traits, 1 genes
6 traits, 3 genes
5 traits, 6 genes
4 traits, 7 genes
3 traits, 8 genes
2 traits, 9 genes
1 traits, 20 genes
0 traits, 37 genes
C
Percentage of genes
100%
80%
60%
40%
Stroke
ApoA1
Direct bilirubin
AST:ALT ratio
Serum urea
Coronary artery disease
WHRadjBMI
Albumin
Bilirubin
T2DM
ApoA
AST
CKD
Hypertension
ApoB
HbA1c
ALT
Cystatin C
GGT
Creatinine
Platelet count
ALP
TG
Cirrhosis
BMI
HDL
LDL
Total cholesterol
Level of evidence
Compelling
Very Strong
Strong
Moderate
Anecdotal
No Evidence
D
HuGE score
10
8
6
4
2
0
ABCC2
SERPINA1
UGT1A1
LIPA
JAG1
KIF12
PKD2
ALG8
SLC10A2
MPI
SEC63
HSD3B7
HJV
TFR2
LARS1
NOTCH2
PEX26
RPGRIP1L
GALK1
HMBS
ASS1
DNAJB11
MARS1
MPV17
TWNK
ASL
GALT
MVK
UNC45A
NR1H4
PKD1
ABCB11
NPC1
PKHD1
ABCB4
HFE
ATP8B1
HNF1B
Total bilirubin
Direct Bilirubin
Albumin
MASLD
AST:ALT ratio
Cirrhosis
ALP
ALT
GGT
AST
E
HuGE score
10
8
6
4
2
0
ALDOB
PKD2
PKD1
PKHD1
ADK
GBE1
JAG1
ABCC2
DNAJB11
HNF1B
NOTCH2
ARG1
CFTR
NPC1
UTP4
SLC10A2
MYO5B
HMBS
VIPAS39
LIPA
UGT1A1
HSD3B7
TRMU
TFR2
VPS33B
MVK
USP53
KIF12
RINT1
GALT
GANAB
MARS1
MPV17
CYP27A1
RPGRIP1L
ALG8
ASL
CC2D2A
FADD
LRP5
NBAS
TJP2
ZFYVE19
DGUOK
TWNK
PRKCSH
HADHA
TALDO1
HAMP
PEX1
PEX26
HJV
ARSB
TUFM
BCS1L
GALK1
UNC45A
AHI1
AKR1D1
NR1H4
ABCB11
MPI
ABCB4
ATP8B1
HFE
SERPINA1
ApoB
Triglycerides
Cholesterol
LDL
HbA1c
Cystatin C
T2DM adjBMI
T2DM
Fasting glucose
Urea
ApoA
HDL
Haemorrhagic stroke
Creatinine
Urate
Phosphate
F
Percentage of genes
40%
30%
20%
10%
0%
Cirrhosis
ICP
MASLD
Liver problem
Cholesterol
AST:ALT ratio
Cholecystitis
Albumin
Cholelithiasis
Direct bilirubin
AST
ALT
Bilirubin
GGT
ALP


A
log10(−log10(p))
neglog_p
300
200
100
DCDC2
ALDOB
ABCB11
ASS1
SERPINA1
HFE
PKD1
NPC2
CPS1
PAH
BCS1L
UGT1A1
PEX26
ABCC2
PKD1
ATP8B1
HNF1B
SERPINA1
ABCB11
BCS1L
NOTCH2
ABCC2
SERPINA1
ABCB4
ATP8B1
ABCB11
HNF1B
CPS1
PEX6
GALK1
ABCC2
SERPINA1
AHI1
ABCB11
HFE
ABCB4
CPS1
UGT1A1
UGT1A1
HFE
CPS1
PEX6
ABCC2
UGT1A1
HFE
ABCC2
GALK1
SERPINA1
SERPINA1
HMBS
HNF1B
AHI1
MPI
PEX26
ALP
GGT
ALT
AST
Cholelithiasis
Bilirubin
Direct bilirubin
Albumin

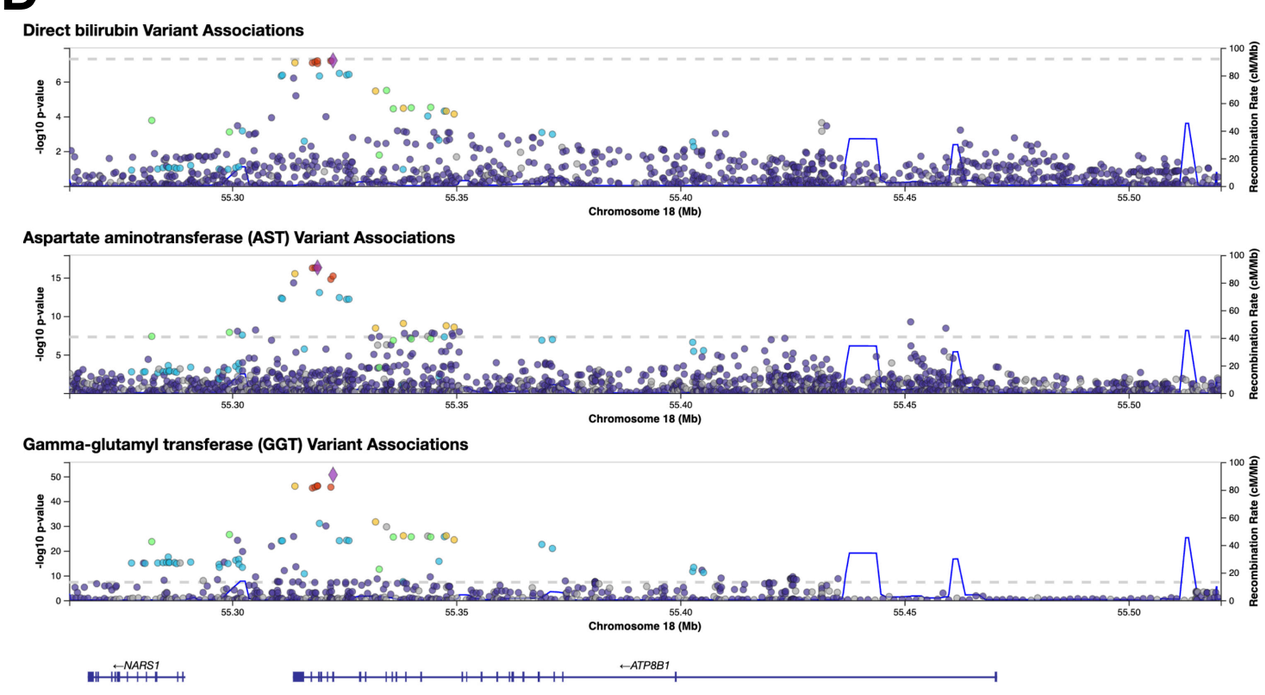
B


Gamma-glutamyl transferase (GGT) Variant Associations
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 7 (Mb)
Alanine transaminase (ALT) Variant Associations
Total bilirubin Variant Associations
←TMEM243
CROT→
←ABCB4
←ABCB1
RUNDC3B→

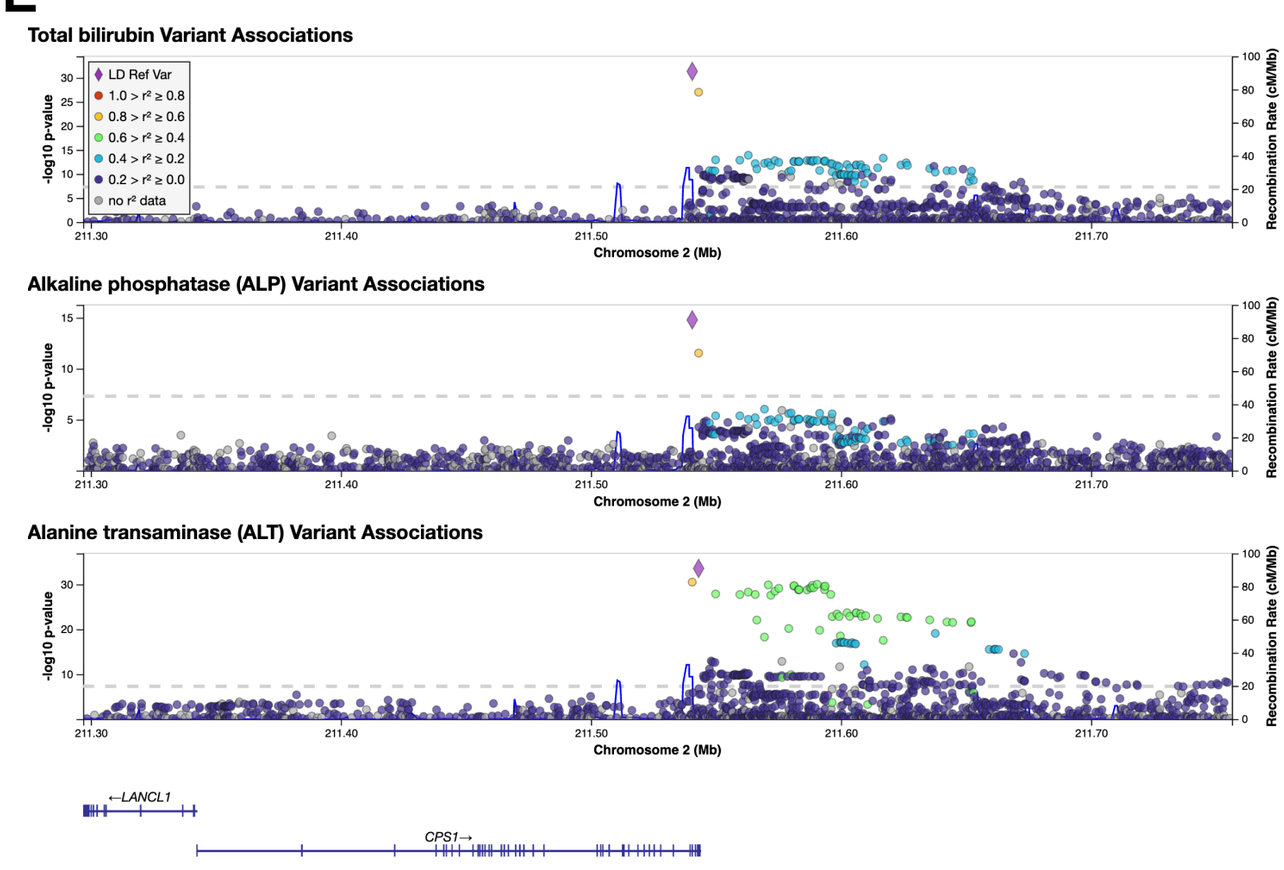
C

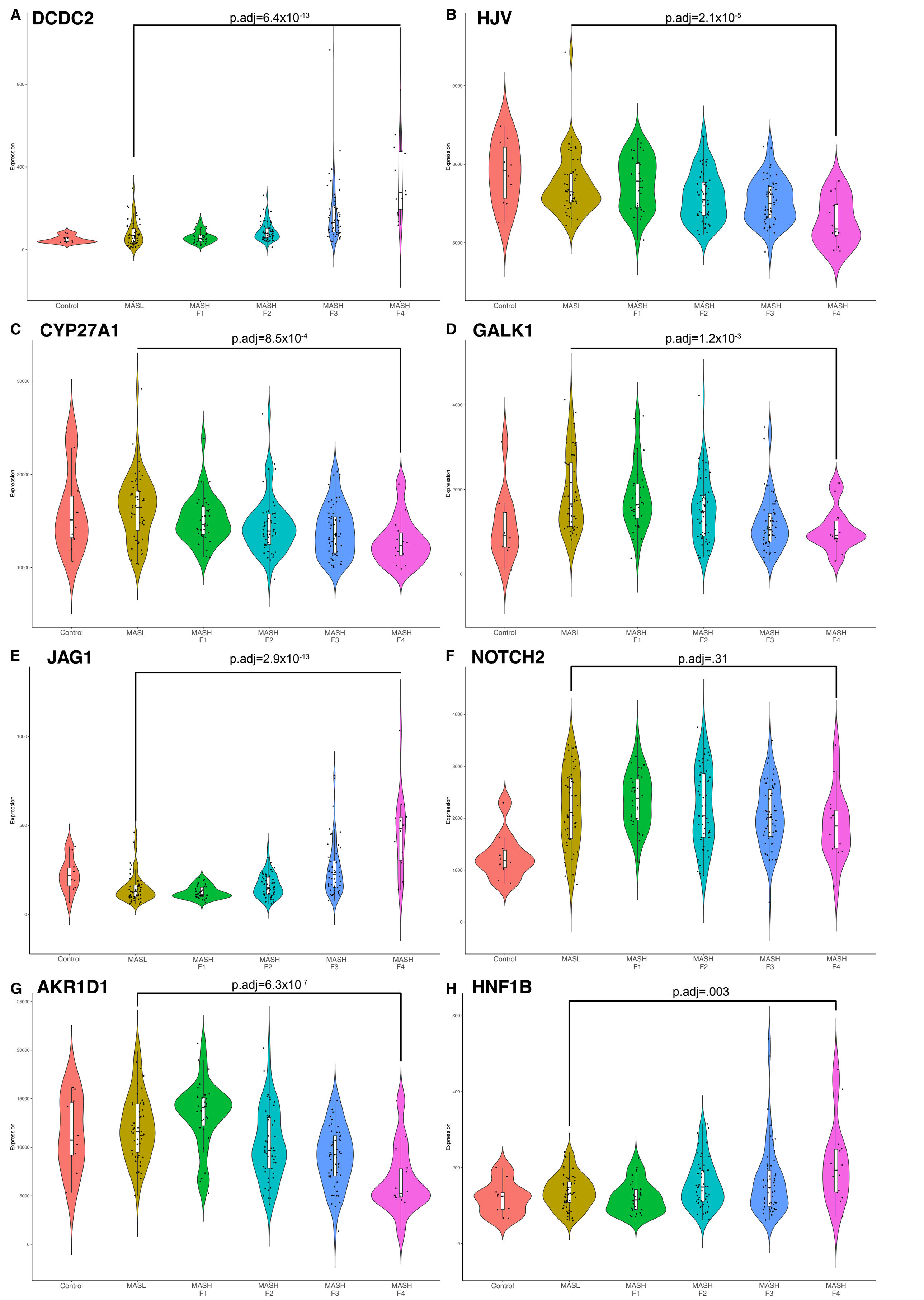

Gamma-glutamyl transferase (GGT) Variant Associations
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 17 (Mb)
Alkaline phosphatase (ALP) Variant Associations
Serum albumin Variant Associations
TADA2A→
DUSP14→
←DDX52
←HNF1B
←SYNRG
TBC1D3K→
TBC1D3F→

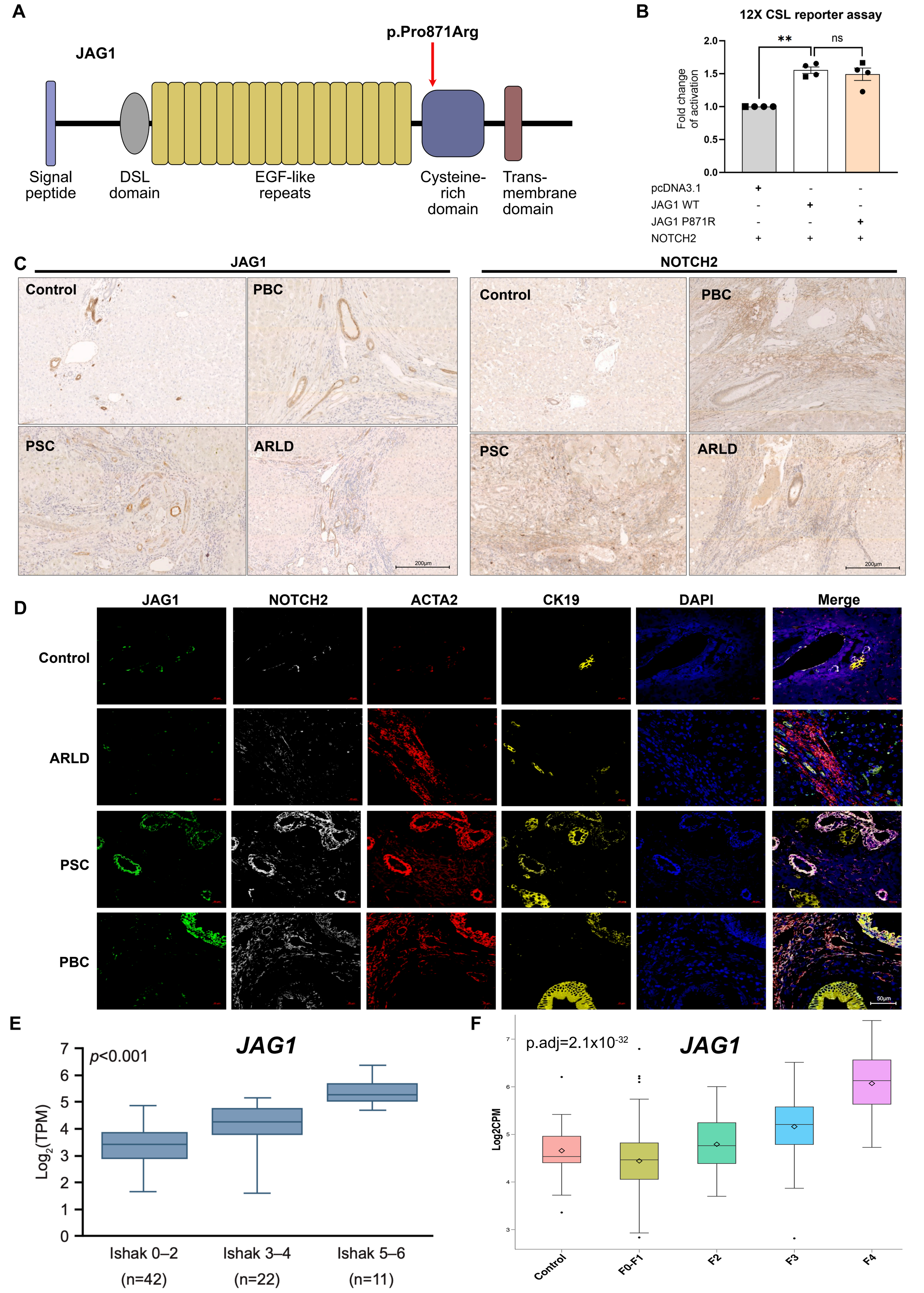
D


Direct bilirubin Variant Associations
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 18 (Mb)
Aspartate aminotransferase (AST) Variant Associations
Gamma-glutamyl transferase (GGT) Variant Associations
←NARS1
←ATP8B1

E


Total bilirubin Variant Associations
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 2 (Mb)
Alkaline phosphatase (ALP) Variant Associations
Alanine transaminase (ALT) Variant Associations
←LANCL1
CPS1→


A
DCDC2
p.adj=6.4x10-13
B
HJV
p.adj=2.1x10-5
C
CYP27A1
p.adj=8.5x10-4
D
GALK1
p.adj=1.2x10-3
E
JAG1
p.adj=2.9x10-13
F
NOTCH2
p.adj=.31
G
AKR1D1
p.adj=6.3x10-7
H
HNF1B
p.adj=.003
Expression
Control
MASL
MASH F1
MASH F2
MASH F3
MASH F4


A
JAG1
p.Pro871Arg
Signal peptide
DSL domain
EGF-like repeats
Cysteine-rich domain
Trans-membrane domain
B
12X CSL reporter assay
Fold change of activation
2.0
1.5
1.0
0.5
0.0
**
ns
pcDNA3.1 + - -
JAG1 WT - + -
JAG1 P871R - - +
NOTCH2 + + +
C
JAG1
NOTCH2
Control
PBC
PSC
ARLD
200µm
D
JAG1
NOTCH2
ACTA2
CK19
DAPI
Merge
Control
ARLD
PSC
PBC
50µm
E
JAG1
p<0.001
Log2(TPM)
Ishak 0–2 (n=42)
Ishak 3–4 (n=22)
Ishak 5–6 (n=11)
F
p.adj=2.1x10-32
JAG1
Log2CPM
Control
F0-F1
F2
F3
F4

Gene	SNP	Variant	EAF	Variant	eQTL	Phenotypes
ABCB4	rs17149637	7_87462869_G_A	0.252396	ABCB4 (p.Leu59=)	Multiple	ALT: - (5.5x10-26), AST: - (4.6x10-8)
ABCC2	rs17222723	10_99836239_T_A	0.0373403	ABCC2 (p.Val1188Glu)	-	ALT: - (3.3x10-75), Bilirubin: - (1.7x10-13), GGT: - (1.7x10-67)
ADK	rs2633311	10_73838341_C_T	0.327476	Intergenic	Blood	GGT: - (7.0x10-13), PBC: - (3.8e-02)
AHI1	rs9376090	6_135090090_T_C	0.146166	Intergenic	Liver	Cholesterol: - (2.6x10-9)
ATP8B1	rs12968116	18_57655270_C_T	0.0451278	ATP8B1 (p.Arg952Gln)	Blood	ALT: - (7.5x10-33), AST: - (1.1x10-10), Cholecystitis: (2.4x10-8), DirectBilirubin: - (4.0x10-11), GGT: - (1.2x10-49)
BCS1L	rs2292553	2_218282080_G_A	0.449481	Intergenic	Liver	ALP: - (6.5x10-19), Cholelithiasis: (1.0x10-8)
CPS1	rs2193617	2_210708835_G_A	0.139177	Intergenic	Liver	ALT: + (1.0x10-14)
DCDC2	rs2296538	6_24289365_C_T	0.101238	DCDC2, intronic_variant	-	ALP: - (7.0x10-146)
GALK1	rs8081539	17_75796291_G_T	0.458866	Intergenic	Liver	DirectBilirubin: - (5.0x10-15)
HNF1B	rs17138478	17_37713312_C_A	0.138778	HNF1B, intronic_variant	Liver	Alb: + (2.4x10-14), ALT: - (2.4x10-30), AST_ALT_ratio: + (4.0x10-15), AST: - (2.0x10-9), Cholecystitis: (3.2x10-8), Cholelithiasis: + (1.0x10-11), GGT: - (5.5x10-19)
JAG1	rs35761929	20_10641853_G_C	0.0529153	JAG1 (p.Pro871Arg)		Bilirubin: + (9.7x10-10), DirectBili: + (2.2x10-8) Diastolic blood pressure: + (4.2x10-14)
LIPA	rs2250781	10_89247713_C_A	0.370407	LIPA, intronic_variant	Liver	ALP: + (3.6x10-12), ALT: - (2.7x10-11)

NOTCH 2	rs10923931	1_119915381_A_T	0.192093	NOTCH2 (p.Gly2447=)	Liver	GGT: + (7.4x10-20)
NPC2	rs8014204	14_74856091_G_A	0.334864	Intergenic	Liver	Bilirubin: - (8.5x10-10)
PKD1	rs34197769	16_2094175_G_A	0.0662939	PKD1 (p.Ala3512Val)	Liver	ALP: - (8.8x10-26), GGT: - (1.3x10-8)