Genome-wide Association Study and Meta-analysis on Alcohol-Associated Liver Cirrhosis Identifies Genetic Risk Factors

Tae-Hwi Schwantes-An, Rebecca Darlay, Philippe Mathurin, Steven Masson, Suthat Liangpunsakul, Sebastian Mueller, Guruprasad P. Aithal, Florian Eyer, Dermot Gleeson, Andrew Thompson, Beat Mueller, Felix Stickel, Michael Soyka, David Goldman, Tiebing Liang, Lawrence Lumeng, Munir Pirmohamed, Bertrand Nalpas, Jean-Marc Jacquet, Romain Moirand, Pierre Nahon, Sylvie Naveau, Pascal Perney, Greg Botwin, Paul S. Haber, Helmut K. Seitz, Christopher P. Day, Tatiana M. Foroud, Ann K. Daly, Heath J. Cordell, John B. Whitfield, Timothy R. Morgan, and Devanshi Seth, for the GenomALC Consortium

BACKGROUND AND AIMS: Only a minority of heavy drinkers progress to alcohol-associated cirrhosis (ALC). The aim of this study was to identify common genetic variants that underlie risk for ALC.

APPROACH AND RESULTS: We analyzed data from 1,128 subjects of European ancestry with ALC and 614 drinking subjects without known liver disease from Australia, the United States, the United Kingdom, and three countries in Europe. A genome-wide association study (GWAS) was performed, adjusting for principal components and clinical covariates (alcohol use, age, sex, body mass index, and diabetes). We validated our GWAS findings using UK Biobank. We then performed a meta-analysis combining data from our study, the UK Biobank, and a previously published GWAS. Our GWAS found genome-wide significant risk association of rs738409 in patatin-like phospholipase domain containing 3 (PNPLA3) (odds ratio [OR] = 2.19 [G allele], P = 4.93 × 10^{-17}) and rs4607179 near HSD17B13 (OR = 0.57 [C allele], P = 1.09 × 10^{-10}) with ALC. Conditional analysis accounting for the PNPLA3 and HSD17B13 loci identified a protective association at rs374702773 in Fas-associated factor family member 2 (FAF2) (OR = 0.61 [del(T) allele], P = 2.56 × 10^{-5}) for ALC. This association was replicated in the UK Biobank using conditional analysis (OR = 0.79, P = 0.001). Meta-analysis (without conditioning) confirmed genome-wide significance for the identified FAF2 locus as well as PNPLA3 and HSD17B13. Two other previously known loci (SERTPNA1 and SUGP1/TM6SF2) were also genome-wide significant in the meta-analysis. GeneOntology pathway analysis identified lipid droplets as the target for several identified genes. In conclusion, our GWAS identified a locus at FAF2 associated with reduced risk of ALC among heavy drinkers. Like the PNPLA3 and HSD17B13 gene products, the FAF2 product has been localized to fat droplets in hepatocytes.

CONCLUSIONS: Our genetic findings implicate lipid droplets in the biological pathway(s) underlying ALC. (HEPATOL 2021;73:1920-1931).

Chronic alcohol use is a leading cause of cirrhosis in the Western world and is on the rise in many other countries. Although chronic...
and heavy alcohol use is a requirement for development of alcohol-associated liver cirrhosis (ALC), only a minority of heavy drinkers progress to cirrhosis. Reliable risk estimates for cirrhosis among heavy drinkers are difficult to obtain, but it is often estimated at 10%-15% after decades of heavy alcohol use.\(^1,2\) Generally accepted risk factors for the development of ALC include duration and amount of alcohol consumed, female gender, and obesity.

Many authors have suggested genetic variation in vulnerability to alcohol-associated or other types of liver disease. The most reported genetic variant in liver diseases is rs738409 in patatin-like phospholipase domain containing 3 (PNPLA3). The variant’s G allele is associated with increased risk for nonalcoholic fatty liver disease/steatohepatitis (NAFLD/NASH)\(^3\) and with increased risk and severity for alcohol-associated liver diseases.\(^4-6\) A genome-wide association study (GWAS) of ALC, from two cohorts consisting of approximately 700 alcohol-associated cases of cirrhosis and 1,400 drinking controls without known liver disease, reported a genome-wide significant association between rs738409 and ALC.\(^7\) These investigators also reported association of rs58542926 (TM6SF2) and rs641738 (MBOAT7) with ALC.\(^7\) Additionally, Abul-Husn et al. reported an association between rs72613567 (17-β-hydroxysteroid
dehydrogenase type 13 [HSD17B13] and alcohol-associated and nonalcoholic fatty liver–related cirrhosis, although not at a genome-wide significance level. Most recently, Emdin et al. reported an association between rs2642438 (MARC1) and all-cause cirrhosis, including in a subset of patients with alcohol-associated liver diseases. Thus, several analyses have identified single nucleotide polymorphisms (SNPs) associated with both alcohol-associated liver disease and NASH, supporting a genetic predisposition for ALC as well as shared genetic susceptibility with NASH.

We undertook a multinational GWAS of European ancestry participants with carefully characterized ALC status to identify predisposing genetic factors. We enrolled subjects at high and low ends of the spectrum of alcohol-associated liver disease, specifically excluding subjects with intermediate stages of disease to maximize statistical power. Cases had a history of high-risk alcohol intake with clinically or histologically defined ALC. Controls were heavy drinkers without clinical evidence of liver disease. We chose drinkers without known liver disease rather than non-drinkers as controls, to avoid the potential of finding genetic risk loci related to alcohol use. Our stringent eligibility criteria for this case-control design allowed for standardization of both the subjects with ALC and those without liver disease across enrollment sites.

Materials and Methods

Subjects

GenomALC Cohort

The GenomALC study participants were recruited at clinical sites in Australia, France, Germany, Switzerland, the United Kingdom, and the United States using a standardized predefined protocol. Enrollment occurred between 2012 and 2017. All participants gave written informed consent. The study was approved by appropriate ethics committees or institutional review board at each site and conformed to the ethical guidelines of the 1975 Declaration of Helsinki. Data and samples were identified by a study-specific code with no personal information. All participants were required to have alcohol consumption of ≥ 80 g/day (males) and ≥ 50 g/day (females) for at least 10 years. Controls were defined as having normal bilirubin, aspartate transaminase (AST), and alanine transaminase (ALT) levels at the time of heavy alcohol use and no prior evidence of liver injury. Those with mildly elevated liver tests were included as controls if their transient elastography was less than 6 kPa. Cases were defined by (1) clinically evident portal hypertension or decompensated cirrhosis (e.g., ascites, esophageal varices); (2) FibroScan stiffness > 22 kPa if AST < 100 IU/L, >32 kPa if AST = 100-200 IU/L, or exclude if AST > 200 IU/L; or (3) liver histology (METAVIR score of F4) on a previously performed liver biopsy. Those with other causes of liver disease (e.g., viral hepatitis, hemochromatosis) or with human immunodeficiency virus were excluded, as described in detail previously.

DNA was obtained from blood samples and genotyped on Global Screening Array v1.0 (Illumina, San Diego, CA). Genotype data were reviewed using the steps detailed in the Supporting Information. In brief, using a GWAS data-cleaning pipeline, genotyped SNPs were filtered for call rate, violation of the Hardy-Weinberg equilibrium, and minor allele frequency (MAF). Samples were checked for genotyping rate, sex, relatedness, and European genetic ancestry. PLINK was used to check for discordance between reported and genetic sex, sample relationship was calculated using the --genome command, and one sample was removed from each pair of related individuals down to second-degree relatives using the Pi-hat value. Genetic ancestry was determined using the SNPRelate package with 1000 Genomes Project as reference. Additional genotypes were imputed using the Michigan Imputation Server. Additional details on GWAS data cleaning are available in the Supporting Information.

Laboratory of Neurogenetics Cohort

Genotype and phenotype data from 860 heavy-drinking subjects who gave consent as part of two National Institutes of Health screening/natural history protocols (98-AA-009 and 05-AA-0121) were provided by the Laboratory for Neurogenetics (LNG) at the National Institutes on Alcohol Abuse and Alcoholism (Rockville, MD) as additional heavy-drinking controls. Participants were required to have alcohol consumption of ≥ 80 g/day (males) and ≥ 50 g/day (females) and were excluded if they had a serious medical condition requiring ongoing treatment, including liver disease. Genotype data (Illumina OmniExpress Bead Chip) were cleaned and imputed using the same pipeline used for GenomALC. After
filtering, 235 participants meeting the inclusion criteria were included as controls.

**UK Biobank Cohort**

Data from the UK Biobank ([https://www.ukbiobank.ac.uk/](https://www.ukbiobank.ac.uk/)) were accessed under approval number 18870. Cirrhosis cases (n = 530) were defined as having International Statistical Classification of Diseases, 10th Revision (ICD-10) code K70.3 (“Alcoholic cirrhosis of liver”) or ICD-9 code 571.2 (“Cirrhosis, liver, alcoholic”), and ICD-10 code K70.1 (“Alcoholic hepatitis without ascites”) or ICD-9 code 571.1 (“Acute alcoholic hepatitis”). Controls (n = 10,222) were defined as having (1) reported alcohol intake of ≥ 80 g/day (males) and ≥ 50 g/day (females) and/or (2) an ICD-10 diagnosis of F10 (Mental and behavioral disorders due to alcohol), but with no recorded diagnosis of any liver disease. Genotype data were cleaned and imputed by the UK Biobank investigators; we used the “White British” ancestry subset of samples provided. After additional data cleaning (see Supporting Information), 8,185,141 SNPs in 439 alcohol-associated cirrhosis cases and 8,364 controls were available for analysis.

Additional sample description and genotype data cleaning information for the three cohorts is available in the Supporting Information.

**STATISTICAL METHODS**

**Genome-wide Association Tests in GenomALC**

A total of 1,128 ALC cases and 614 heavy-drinking controls were available for analysis. To test the association between each SNP and ALC (cases vs. controls), logistic regression between case-control status and dosage value of each SNP was tested using PLINK. The following covariates were included in each regression model: (1) first 10 principal components (PCs), (2) age, (3) sex, (4) years of excessive drinking, (5) alcohol consumed (measured in grams per day), (6) total lifetime alcohol consumed (measured in kilograms), (7) diabetes status (present/absent), and (8) body mass index (BMI). To identify additional genetic associations beyond the two genome-wide significant loci (rs4607179 for HSD17B13 and rs738409 for PNPLA3), we ran a conditional GWAS that included minor allele counts (dosage) for variants at the two loci as covariates. Conditional GWAS also included all covariates used in the primary GWAS along with the two dosage values. The P value threshold for genome-wide statistical significance was set at $5 \times 10^{-8}$.

**GWAS in UK Biobank Data**

The data were analyzed with FaST-LMM, a mixed model method that computes a kinship matrix to adjust for relatedness/population stratification and infers any missing genotypes at an SNP based on the genotypes of other samples at that SNP. Cases (n = 231) and drinking controls (n = 8,364) who had all of the clinical covariates (age, sex, type 2 diabetes mellitus status, BMI, and current daily alcohol intake [in grams]) were included in a GWAS that mirrored the primary analysis performed in the GenomALC cohort. To mirror the GenomALC conditional analysis in the UK Biobank data, these cases and drinking controls were re-analyzed with the minor allele counts at rs738409 and rs4607179 as covariates along with the clinical variables. Odds ratios (ORs) were obtained from logistic regression in PLINK including all relevant covariates, and confidence intervals (CIs) were calculated from back-transformation of FaST-LMM P values and PLINK ORs.

As a sensitivity analysis, we also conducted a GWAS of cases and controls that excluded those who did not meet the criteria of current alcohol intake per day ≥ 80 g/day (males) and ≥ 50 g/day (females), resembling the GenomALC criteria. This GWAS included a total of 231 cases and 5,361 controls and adjusted for the same set of covariates used in the primary GWAS.

**Meta-analysis**

Meta-analysis of our primary GWAS results from GenomALC, the UK Biobank, and the GWAS summary statistics from a published GWAS of ALC was conducted using METAL with the genomic control option. Summary statistics for the published GWAS were obtained from http://gengastro.med.tu-dresden.de/suppl/alc_cirrhosis/.

**Gene-Set Enrichment Analysis**

We used enrichment analysis from Gene Ontology to identify potential pathways shared
by genes identified in our GenomALC GWAS analysis. Using the genes that include or are near the SNPs with $P$ values less than $5 \times 10^{-6}$ from our primary GWAS analysis, we identified 15 genes for enrichment analysis. We performed enrichment analysis for molecular function, cellular component, and biological processes from Gene Ontology.\(^{(17,18)}\)

Results

**PRIMARY GWAS ANALYSIS**

The primary GenomALC GWAS analysis identified two genome-wide significant associations (Fig. 1 and Supporting Fig. S1). rs738409 on chromosome 22 (chr22), located in PNPLA3, was significantly associated with increased risk of ALC, with a $P$ value of $4.93 \times 10^{-17}$ and an OR of 2.19 (95% CI, 1.81-2.54) for each copy of the G allele. rs4607179 on chr4, located near HSD17B13, was associated with lower risk of ALC with a $P$ value of $1.09 \times 10^{-10}$ and OR of 0.57 for each copy of the C allele (95% CI, 0.48-0.62) (Table 1). A list of all SNPs with $P$ values less than $5 \times 10^{-6}$ and their associated test results is given in Supporting Table S1. Age, sex, and 10 PC-adjusted GWAS analysis results of GenomALC and additional controls from LNG are found in Supporting Fig. S2 and Supporting Table S2. Descriptive statistics of the GenomALC cohort are given in Table 2, and the UK Biobank and LNG are found in Supporting Table S3. The primary GWAS analysis results from UK Biobank and sensitivity analysis, performed on controls reporting current alcohol use for >80 g/day for men and >50 g/day for women, are found in Supporting Tables S4 and S5, respectively.

**CONDITIONAL GWAS ANALYSIS**

To identify additional genetic associations, we performed a conditional GWAS that included minor allele counts (dosage) for variants in the two genome-wide significant loci (rs4607179 for HSD17B13 and rs738409 for PNPLA3) from the primary GWAS analysis as covariates along with the previously included covariates. The conditional GWAS analysis identified genome-wide significant association for rs374702773 (FAF2) on chr5 with a $P$ value of $2.56 \times 10^{-8}$ and an OR of 0.61 (95% CI, 0.51-0.73) for each copy of the 7bp deletion allele (Fig. 2, Supporting Fig. S3, and Table 1). Association test results for both primary and conditional GWAS analyses for the FAF2 locus from GenomALC are given in Supporting Table S6. Figure 3 shows a LocusZoom\(^{(19)}\) plot of rs374702773 and nearby variants. rs374702773 was not available in the UK Biobank data; instead, we tested four variants in strong linkage disequilibrium (LD) ($D' \geq 0.9$ and $r^2 \geq 0.6$) with it. The LD values among variants are

---

**FIG. 1.** Manhattan plot of primary GWAS analysis of GenomALC. The plot shows significant association of SNPs on chromosomes 4 (HSD17B13) and 22 (PNPLA3). Vertical axis shows -log10 transformed $P$ value of each tested SNP after adjusting for age, sex, diabetes (yes/no), BMI, years of excessive drinking, alcohol consumed in grams per day, total lifetime alcohol consumed in kilograms, and 10 PCs. Horizontal solid gray line shows the genome-wide significance level ($P = 5 \times 10^{-8}$); dotted gray line shows the suggestive level ($P = 5 \times 10^{-6}$).
In the conditional analysis of the UK Biobank, rs11134977 showed the strongest association with ALC (OR = 0.79, \(P = 0.001\)); other SNPs (rs11027, rs12514451, and rs34152523) in LD with rs374702773 showed similar OR and \(P\) value (Supporting Table S6). We tested for interaction between \(\text{FAF2}\) locus (rs374702773 and four SNPs in LD) and \(\text{PNPLA3}\) (rs738409), and \(\text{FAF2}\) locus and \(\text{HSD17B13}\) (rs4607179) in GenomALC. None of the interaction terms were statistically significant (\(P > 0.44\)) (Supporting Table S8).

Meta-analysis of results from GenomALC, UK Biobank, and a previously published GWAS of alcohol-associated liver cirrhosis provided in Supporting Table S7. In the conditional analysis of the UK Biobank, rs11134977 showed the strongest association with ALC (OR = 0.79, \(P = 0.001\)); other SNPs (rs11027, rs12514451, and rs34152523) in LD with rs374702773 showed similar OR and \(P\) value (Supporting Table S6). We tested for interaction between \(\text{FAF2}\) locus (rs374702773 and four SNPs in LD) and \(\text{PNPLA3}\) (rs738409), and \(\text{FAF2}\) locus and \(\text{HSD17B13}\) (rs4607179) in GenomALC. None of the interaction terms were statistically significant (\(P > 0.44\)) (Supporting Table S8).

META-GWAS OF THREE STUDIES

Meta-analysis of results from GenomALC, UK Biobank, and a previously published GWAS of alcohol-associated liver cirrhosis identified five independent genome-wide significant associations with ALC (Fig. 4 and Supporting Fig. S4). Summaries of top associated SNPs from each locus are provided in Table 1, and all SNPs with \(P\) values less than \(5 \times 10^{-6}\) are given in Supporting Table S9. In the meta-analysis, the most significant association at \(\text{FAF2}\) was for rs11134977 (\(P = 1.56 \times 10^{-5}\)), which is in LD with rs374702773 in \(\text{FAF2}\). In addition to associations at the \(\text{HSD17B13}, \text{PNPLA3},\) and \(\text{FAF2}\) loci, we identified two additional genome-wide significant associations: rs28929474 (chr14:94,844,947, \(\text{SERPINA1}\).
OR = 1.90; 95% CI, 1.52–2.38; \( P = 1.99 \times 10^{-8} \) and rs10401969 (chr19:19,407,718, SUGP1; OR = 1.49; 95% CI, 1.31–1.70; \( P = 2.40 \times 10^{-7} \)) (Table 1). LocusZoom plots of these five loci are available in Supporting Fig. S5. Meta-analysis of age, sex, and 10 PC-adjusted GWASs in GenomALC + LNG and

FIG. 2. Manhattan plot of secondary GWAS analysis of GenomALC. After accounting for minor allele counts (dosage) for variants in rs4607179 near \textit{HSD17B13} and rs738409 in \textit{PNPLA3}, secondary analysis in GenomALC shows significant association of SNPs on chromosome 5 (\textit{FAF2}). Vertical axis shows \(-\log_{10}\) transformed \( P \) value of each tested SNP after adjusting for rs4607179, rs738409, age, sex, diabetes (yes/no), BMI, years of excessive drinking, alcohol consumed in grams per day, total lifetime alcohol consumed in kilograms, and 10 PCs. Horizontal solid gray line shows the genome-wide significance level (\( P = 5 \times 10^{-8} \)); dotted gray line shows the suggestive level (\( P = 5 \times 10^{-6} \)).

FIG. 3. LocusZoom plot of rs374702773 and nearby SNPs on chromosome 5. From the conditional analysis of GenomALC data, we identified the locus rs374702773 (purple diamond) on Chr 5 associated with ALC and nearby SNPs in LD. Chr, Chromosome.
SCHWANTES-ÁN, DARLAY, ET AL.

UK Biobank and published GWASs from the Buch et al. cohorts are available in Supporting Fig. S6 and Supporting Table S10.

**GENE-ENRICHMENT ANALYSIS**

Gene-enrichment analysis for biological process and cellular component showed a significant enrichment for lipid droplet and its organization among the 15 genes (Supporting Table S11). Enrichment for biological process was observed for lipid droplet organization (>100-fold, false discovery rate [FDR] = 0.003) and cellular component (>70-fold, FDR = 0.020) (Table 3).

**Discussion**

To date, this is the largest GWAS for alcohol-associated cirrhosis, and the only multicenter GWAS of prospectively enrolled subjects meeting strict, protocol-defined criteria for heavy drinking, and for alcohol-associated cirrhosis or no known liver disease. We conducted the GWAS meta-analysis in ALC, contrasting clinically defined groups of cases and heavy drinking controls from GenomALC, the UK Biobank, and a previously published GWAS on ALC. We report a genome-wide association with ALC: rs374702773 located in FAF2 with the C (minor) allele shows a protective effect. Additionally, we replicated the
association between rs738409 (PNPLA3) and ALC, strengthening previous reports.\(^{(8,20)}\) We also replicated an association between \textit{HSD17B13} locus and ALC by identifying a genome-wide significant association between ALC and rs4607179 (\(P = 1.09 \times 10^{-10}\)). Meta-analysis also identified two additional genome-wide significant associations: rs28929474, a missense variant in \textit{SERPINA1}, and rs10401969, located in an LD block that spans multiple genes including \textit{SUGP1} and \textit{TM6SF2}. However, we did not find genome-wide significance for rs641738 or any other SNPs in \textit{MBOAT7} (primary GWAS \(P = 5.31 \times 10^{-03}\), meta-analysis \(P = 4.03 \times 10^{-05}\), as reported by Buch et al.\(^{(7)}\) A recently reported association of rs2642438 in \textit{MARCI} gene and cirrhosis\(^{(9)}\) was also not genome-wide significant in our meta-analysis (OR = 0.81; \(P = 7.54 \times 10^{-06}\)) nor in GenomALC GWAS (OR = 0.79; \(P = 0.006\)), but the effect was in the same direction.

\textit{FAF2} represents a locus affecting the genetic risk for ALC. This association became genome-wide significant for rs374702773 in our GenomALC cohort after accounting for the effects of rs738409 (PNPLA3) and rs4607179 (near \textit{HSD17B13}) (Supporting Table S6). We also replicated this association using the UK Biobank cohort (OR = 0.79; \(P = 0.001\) for rs11134977). Using the larger sample size available in the meta-analysis of GenomALC, UK Biobank, and Buch et al.,\(^{(7)}\) this association was found without conditional analysis (OR = 0.79; \(P = 1.56 \times 10^{-08}\) for rs11134977). We did not observe interaction between \textit{FAF2} locus and PNPLA3 or HSD17B13 loci (Supporting Table S8). Although this may indicate that these loci act independently, our study was not powered enough to detect interactions. These variants in \textit{FAF2} are in strong LD and represent the same association. Also known as UBDX8, \textit{FAF2} is a sensor of intracellular levels of long-chain unsaturated fatty acids.\(^{(21)}\) In the presence of unsaturated fatty acid, UBDX8 binds to adipose triglyceride lipase (ATGL), the major triglyceride hydrolizing enzyme, and dissociates it from its activator \(\alpha/\beta\) hydrolase domain 5 (ABHD5, also known as comparative gene identification 58 [CGI-58]). This dissociation from ABHD5/CGI-58 inhibits triacylglycerol hydrolysis in lipid droplets.\(^{(21,22)}\) This biochemical pathway (binding to ABHD5/CGI-58 and inhibiting triglyceride hydrolysis in lipid droplets) is believed to be one mechanism by which PNPLA3 (I148M) contributes to hepatic steatosis.\(^{(23,24)}\) Thus, \textit{FAF2} fits into the current proposed genetic pathways that affect hepatocyte lipid metabolism. However, \textit{FAF2} also regulates sterol regulatory element binding protein 1 (SREBP-1) activation and triacylglycerol synthesis from diacylglycerol (DAG) by DAG acyltransferase.\(^{(21)}\) The role and molecular function of the \textit{FAF2}-associated variant and those SNPs in LD with it, are yet to be explored. Thus, the actual role of \textit{FAF2} in ALC risk will require a better understanding of the pathophysiology of liver injury by \textit{FAF2} and other implicated genes.

Meta-analysis found genome-wide significance for association between rs28929474, a missense variant of \textit{SERPINA1}, and ALC. \textit{SERPINA1} codes for the \(\alpha\)-antitrypsin (AAT) protein, with rs28929474 (Glu366His) coding for the Z-variant of AAT. The Z-variant causes conformational rearrangement of AAT, affecting its degradation by proteases.\(^{(25,26)}\) Homozygosity for the Z-variant of \textit{SERPINA1} leads to AAT deficiency\(^{(27,28)}\) and increased risk for cirrhosis. Although this is a variant with low MAF, there is independent evidence that it contributes to the risk of alcohol-associated liver disease and cirrhosis. Heterozygosity of this variant has also been associated with portal hypertension\(^{(29)}\) and cystic fibrosis–associated liver disease.\(^{(30)}\) Abul-Husn et al.\(^{(8)}\) and Chen et al.\(^{(31)}\) reported rs28929474 association with alcohol-associated liver disease/cirrhosis and nonalcoholic liver diseases using candidate variant analysis and a GWAS-level significance with liver transaminases. Strmad et al. reported an association between the Z-variant heterozygous carriers and cirrhosis in NAFLD (OR = 7.3; \(P < 0.0001\)) and in alcohol misusers (OR = 5.8; \(P < 0.0001\)).\(^{(32)}\) Although none of these associations approached GWAS-level significance, their findings are in the same direction as our report. This appears to be a disease risk associated with the AAT Z-variant gene product; until now it was thought that AAT-related liver disease only occurs in ZZ homozygotes.

We extended an association between \textit{HSD17B13} locus and liver disease, reported previously. Although not at genome-wide significance level, these authors reported that the A allele (minor allele) of rs72613567 showed protective effect against NAFLD and/or alcohol-associated liver diseases including cirrhosis and hepatocellular carcinoma (HCC).\(^{(8)}\) However, most recently, another group reported GWAS significance for \textit{HSD17B13} associated with protection against NAFLD.\(^{(33)}\) In our discovery GWAS of GenomALC, we found genome-wide significance
for rs4607179 near HSD17B13. Our meta-analysis confirmed significance for SNPs in HSD17B13 and risk for ALC; we observe a protective effect of the C allele of rs4607179 on ALC at a genome-wide significance level (OR = 0.80; P = 1.39 × 10^{-08}). Although the detailed biological function of HSD17B13 is not fully understood, it too is known to be associated with hepatic lipid droplets and circulating triglyceride and high density lipoprotein levels, suggesting overlapping mechanisms in ALC and other complex liver diseases.

We replicated the association between ALC and rs738409, which codes for a missense mutation in PNPLA3. rs738409 had the strongest statistical association and the highest OR among the genome-wide significant SNPs (OR = 2.19; P = 4.93 × 10^{-17}). This locus has been associated with risk for NAFLD, alcohol-associated liver disease, and HCC in several meta-analyses. PNPLA3 is a lipase found on the surface of triglyceride droplets in hepatocytes. The variant PNPLA3 protein may increase hepatic steatosis by increasing dissociation of ABDH5/CGI-58 from adipose triglyceride lipase, thereby inhibiting the triglyceride catabolism by ATGL. In addition, the half-life of the rs738409 variant PNPLA3 protein on the lipid droplet appears prolonged, extending its ability to inhibit triglyceride hydrolysis.

Our meta-analysis found genome-wide significance for rs10401969 near SUGP1 on chr19 (OR = 1.49 for C allele; P = 2.40 × 10^{-09}). rs10401969 has been proposed to modulate activity of 3-hydroxy-3-methylglutaryl-CoA reductase (also known as HMG-CoA reductase), the rate-limiting enzyme for cholesterol synthesis. rs10401969 has been reported to show associations with plasma low-density lipoprotein cholesterol level, coronary artery disease, hepatic fat in obese patients, suggesting that SUGP1 may be directly related to genetic risk for ALC. However, SUGP1 is part of a locus with multiple genes, and this intron variant is in strong LD (r^2 ~ 0.95 in Europeans) with variants located in nearby TM6SF2 (Supporting Fig. S5), a locus that has been associated with ALC, NAFLD, cirrhosis/steatosis/fibrosis among patients with chronic hepatitis C, and impaired lipid synthesis. Thus, it is not clear whether this variant implicates SUGP1 as a potential underlying gene for risk of cirrhosis or is a marker for TM6SF2.

One common thread among our findings is the involvement of genetic variants in lipid biology, particularly triglyceride metabolism. Enrichment analysis using GeneOntology identified lipid droplet cellular components and lipid droplet organization biological processes among the genes identified by GWAS. The genes involved in the process from our list of candidate genes were HSD17B13, PNPLA3, and FAF2. However, these genes have pleotropic effects: FAF2 also regulates SREBP-1 activation and triacylglycerol synthesis, whereas HSD17B13 has retinol dehydrogenase activity. Thus, defining the actual role of these genes in ALC will depend on better understanding of lipid droplet biology and other biochemical pathways in the pathophysiology of liver injury, especially for non-alcohol-related disease, which shares several underlying genetic factors and pathophysiological mechanisms.

This GWAS has potential limitations. Despite being the largest available cohort for ALC, the size of our sample for a GWAS of common variants imposes a limit on the power of detecting associations with weaker effect sizes. Second, although the trend for the ORs in the UK Biobank cohort is in the same direction as for GenomALC, the strength of the association was weaker. We suspect the weaker association may be due to the smaller number of cases in UK Biobank and expanded definition of heavy-drinking controls in the UK Biobank as compared with GenomALC. Using stricter drinking criteria for controls, in our secondary and sensitivity analysis in the smaller UK Biobank cohort, we observed similar association results (Supporting Fig. S8).

Overall, we report a genetic association between FAF2 and ALC. This locus is functionally linked to two previously reported loci (PNPNA3 and HSD17B13) by a biological pathway that governs lipid droplet organization. We also report increased risk for ALC in heterozygous carriers of the SERPINA1 Z variant. Taken together, our findings suggest that risk of developing ALC is, in part, related to genetic factors, especially genetic factors regulating lipid homeostasis, and heavy alcohol exposure is a necessary but not sufficient precondition for ALC.

**REFERENCES**

1) Mann RE, Smart RG, Govoni R. The epidemiology of alcoholic liver disease. Alcohol Res Health 2003;27:209-219.
2) Askgaard G, Kjar MS, Tolstrup JS. Opportunities to prevent alcoholic liver cirrhosis in high-risk populations: a systematic review with meta-analysis. Am J Gastroenterol 2019;114:221-232.

3) Romeo S, Kozlitina J, Xing C, Pertsemidis A, Cox D, Penachcio LA, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2008;40:1461-1465.

4) Tian C, Stokowski RP, Kershenobich D, Ballinger DG, Hinds DA. Variant in PNPLA3 is associated with alcoholic liver disease. Nat Genet 2010;42:21-23.

5) Seth D, Daly AK, Haber PS, Day CP. Patatin-like phospholipase domain containing 3: a case in point linking genetic susceptibility for alcoholic and nonalcoholic liver disease. Hepatology 2010;51:1463-1465.

6) Salameh H, Raff E, Erwin A, Seth D, Nischalke HD, Falleti E, et al. PNPLA3 gene polymorphism is associated with predisposition to and severity of alcoholic liver disease. Am J Gastroenterol 2015;110:846-856.

7) Buch S, Stickel F, Trépo E, Way M, Herrmann A, Nischalke HD, et al. A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. Nat Genet 2015;47:1443-1448.

8) Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, et al. A protein-truncating HSD17B13 variant and protection from chronic liver disease. N Engl J Med 2018;378:1096-1106.

9) Emdin CA, Haas ME, Khera AV, Aragam K, Chaffin M, Klarin D, et al. A missense variant in mitochondrial amidoxime reducing component 1 gene and protection against liver disease. PLoS Genet 2020;16:e1008629.

10) Scott E, Anstee QM. Genetics of alcoholic liver disease and non-alcoholic steatohepatitis. Clinical Med 2018;18:s54-59.

11) Whittfield JB, Rahman K, Haber PS, Day CP, Masson S, Daly AK, et al. Brief report: genetics of alcoholic cirrhosis—GenomALC multinational study. Alcohol Clin Exp Res 2015;39:836-842.

12) Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotyping imputation service and methods. Nat Genet 2016;48:1284-1287.

13) 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.

14) Lippert C, Listjarten J, Liu Y, Kadie CM, Davidson RJ, Ahekerman D. FaST linear mixed models for genome-wide association studies. Nat Methods 2011;8:833-835.

15) Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190-2191.

16) Mi H, Huang X, Murugunajan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res 2017;45:D183-D189.

17) Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genet 2000;25:25-29.

18) The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still going strong. Nucleic Acids Res 2018;47:D130-D138.

19) Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics (Oxford, England) 2010;26:236-237.

20) Stickel F, Lutz P, Buch S, Nischalke HD, Silva I, Rausch V, et al. Genetic variation in HSD17B13 reduces the risk of developing cirrhosis and hepatocellular carcinoma in alcohol misusers. Hepatology 2020;72:88-102.

21) Kim H, Ye J. Cellular responses to excess fatty acids: focus on ubiquitin regulatory X domain-containing protein 8. Curr Opin Lipidol 2014;25:118-124.

22) Olzmann JA, Richter CM, Kopito RR. Spatial regulation of UBxD8 and p97/VCP controls ATGL-mediated lipid droplet turnover. Proc Natl Acad Sci USA 2013;110:1345-1350.

23) Wang Y, Kory N, BasuRay S, Cohen JC, Hobbs HH. PNPLA3, CGI-58, and inhibition of hepatic triglyceride hydrolysis in mice. Hepatology 2019;69:2427-2441.

24) Yang A, Mottillo EP, Mladenovic-Lucas L, Zhou L, Granneman JG. Dynamic interactions of ABHD5 with PNPLA3 regulate triacylglycerol metabolism in brown adipocytes. Nat Metab 2019;1:560-569.

25) Lomas DA, Li-Evans D, Finch JT, Carrell RW. The mechanism of Z alpha-1-antitrypsin accumulation in the liver. Nature 1992;357:605-607.

26) Kopito RR, Ron D. Conformational disease. Nat Cell Biol 2000;2:E207-E209.

27) de Serres FJ, Blanco I, Fernandez-Bustillo E, Pi S and Pi Z. alpha-1 antitrypsin deficiency worldwide. A review of existing genetic epidemiological data. Monaldi Arch Chest Dis 2007;67:184-208.

28) Zorzetto M, Russi E, Senn O, Imboden M, Ferrarotti I, Tinelli C, et al. SERPINA1 gene variants in individuals from the general population with reduced alpha1-antitrypsin concentrations. Clin Chem 2008;54:1331-1338.

29) Mandorfer M, Bucisics T, Hutya V, Schmid-Scherzer K, Schaefer B, Zoller H, et al. Liver disease in adults with alpha-1-antitrypsin deficiency. United Eur Gastroenterol J 2018;6:710-718.

30) Boelle P-Y, Debray D, Guillot L, Corvol H. On behalf of the French CFMOSI. SERPINA1 Z allele is associated with cystic fibrosis liver disease. Genet Med 2019;21:2151-2155.

31) 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;40:405-415.

32) Strnad P, Buch S, Hamesch K, Fischer J, Rosendahl J, Schmelz R, et al. Heterozygous carriage of the alpha-1-antitrypsin PrZ variant increases the risk to develop liver cirrhosis. Gut 2019;68:1099-1107.

33) Anstee QM, Darlay R, Cockell S, Meroni M, Govaere O, Tinakos D, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically-characterised cohort. J Hepatol 2020;73:505-515.

34) Su W, Mao Z, Liu Y, Zhang X, Zhang W, Gustafsson J-A, et al. Role of HSD17B13 in the liver physiology and pathophysiology. Mol Cell Endocrinol 2019;489:119-125.

35) Rotroff DM, Pijut SS, Marvel SW, Jack JR, Havener TM, Pujol J, et al. Genetic variants in HSD17B3, SMAD3, and IPO11 impact circulating lipids in response to fenofibrate in individuals with type 2 diabetes. Clin Pharmacol Ther 2018;103:712-721.

36) Su W, Wang Y, Jia X, Wu W, Li L, Tian X, et al. Comparative proteomic study reveals 17β-HSD13 as a pathogenic protein in nonalcoholic fatty liver disease. Proc Natl Acad Sci USA 2014;111:11437-11442.

37) Sookoian S, Pirola CJ. Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of non-alcoholic fatty liver disease. Genet Med 2019;21:2151-2155.

38) Singal AG, Manjunath H, Yopp AC, Beg MS, Marrero JA, Gopal P, et al. The effect of PNPLA3 on fibrosis progression and development of hepatocellular carcinoma: a meta-analysis. Am J Gastroenterol 2014;109:325-334.

39) Zhang L, You W, Zhang H, Peng R, Zhu Q, Yao A, et al. PNPLA3 polymorphisms (rs738409) and non-alcoholic fatty liver disease risk and related phenotypes: a meta-analysis. J Gastroenterol Hepatol 2015;30:821-829.

40) Chamorro AJ, Torres JL, Miron-Canelo JA, Gonzalez-Sarmiento R, Laso FJ, Marcos M. Systematic review with meta-analysis: the I148M variant of patatin-like phospholipase domain-containing
3 gene (PNPLA3) is significantly associated with alcoholic liver cirrhosis. Aliment Pharmacol Ther 2014;40:571-581.

41) DiStefano JK, Kingsley C, Craig Wood G, Chu X, Argyropoulos G, Still CD, et al. Genome-wide analysis of hepatic lipid content in extreme obesity. Acta Diabetol 2015;52:373-382.

42) Whitfield JB, Zhu G, Madden PAF, Montgomery GW, Heath AC, Martin NG. Biomarker and genomic risk factors for liver function test abnormality in hazardous drinkers. Alcohol Clin Exp Res 2019;43:473-482.

43) Li Y, Liu S, Gao Y, Ma H, Zhan S, Yang Y, et al. Association of TM6SF2 rs58542926 gene polymorphism with the risk of non-alcoholic fatty liver disease and colorectal adenoma in Chinese Han population. BMC Biochem 2019;20:3.

44) Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjærg-Hansen A, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2014;46:352-356.

45) Liu Y-L, Reeves HL, Burt AD, Tiniakos D, McPherson S, Leathart JBS, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. Nat Commun 2014;5:4309.

46) Liu Z, Que S, Zhou L, Zheng S, Romeo S, Mardinoglu A, et al. The effect of the TM6SF2 E167K variant on liver steatosis and fibrosis in patients with chronic hepatitis C: a meta-analysis. Sci Rep 2017;7:9273.

47) Loukkonen PK, Zhou Y, Nidhina Haridas PA, Dwivedi OP, Hyötyläinen T, Ali A, et al. Impaired hepatic lipid synthesis from polyunsaturated fatty acids in TM6SF2 E167K variant carriers with NAFLD. J Hepatol 2017;67:128-136.

48) Romeo S, Sanyal A, Valenti L. Leveraging human genetics to identify potential new treatments for fatty liver disease. Cell Metab 2020;31:35-45.

49) Trépo E, Valenti L. Update on NAFLD genetics: from new variants to the clinic. J Hepatol 2020;72:1196-1209.

Author names in bold designate shared co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.31535/suppinfo.