Received: 26 April 2024 | Accepted: 18 June 2024

DOI: 10.1097/HEP.0000000000001027

ORIGINAL ARTICLE

OPEN

Exome-wide association analysis identifies novel risk loci for alcohol-associated hepatitis

Qiaoping Yuan[1] | Colin Hodgkinson[1] | Xiaochen Liu[2] | Bruce Barton[3] | Nancy Diazgranados[4] | Melanie Schwandt[4] | with DASH, InTEAM, SCAHC, TREAT and Alcohol Hepatitis Genomics consortia | Timothy Morgan[5,6] | Ramon Bataller[7,8,9,10] | Suthat Liangpunsakul[11,12,13] | Laura E. Nagy[14,15] | David Goldman[1,4]

[1]Laboratory of Neurogenetics, National Institute on Alcohol Abuse and Alcoholism, Bethesda, Maryland, USA

[2]Department of Epidemiology and Biostatistics, University of California, Irvine, Irvine, California, USA

[3]Department of Population & Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, Massachusetts, USA

[4]Office of the Clinical Director, National Institute on Alcohol Abuse and Alcoholism, Bethesda, Maryland, USA

[5]Department of Gastroenterology, Long Beach Veterans Healthcare System (VALVE), Long Beach, California, USA

[6]Department of Medicine, University of California, Irvine, CA, USA

[7]Liver Unit, Hospital Clínic de Barcelona, Barcelona, Spain

[8]Facultad de Medicina i Ciències de la Salut, Universitat de Barcelona, Barcelona, Spain

[9]Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[10]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Barcelona, Spain

[11]Division of Gastroenterology, Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA

[12]Department of Biochemistry & Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana, USA

[13]Roudebush Veterans Administration Medical Center, Indianapolis, Indiana, USA

[14]Department of Inflammation & Immunity, Cleveland Clinic Lerner Research Institute, Cleveland, Ohio, USA

[15]Department of Molecular Medicine, Case Western Reserve University, Cleveland, Ohio, USA

Correspondence
David Goldman, Laboratory of Neurogenetics, National Institute on Alcohol Abuse and Alcoholism, Room 3S32, 5625 Fishers Lane, Rockville, MD 20852, USA.
Email: davidgoldman@mail.nih.gov

Abstract

Background and Aims: Alcohol-associated hepatitis (AH) is a clinically severe, acute disease that afflicts only a fraction of patients with alcohol use disorder. Genomic studies of alcohol-associated cirrhosis (AC) have identified several genes of large effect, but the genetic and environmental factors that lead to AH and AC, and their degree of genetic overlap, remain largely unknown. This study aims to identify genes and genetic variations that contribute to the development of AH.

Abbreviations: AC, alcohol-associated cirrhosis; *ADH1B,* alcohol dehydrogenase 1B; AFR, African; AH, alcohol-associated hepatitis; ALD, alcohol-associated liver disease; AMR, Admixed American; AUD, alcohol use disorder; DASH, Defeat Alcohol-associated Steatohepatitis; EAS, East Asian; EUR, European; GWAS, genome-wide association study; *ICOSLG,* inducible T cell costimulatory ligand; LoF, loss-of-function; MAF, minor allele frequency; MAFLD, metabolic-associated fatty liver disease; NIAAA, National Institutes on Alcohol Abuse and Alcoholism; *PNPLA3*, patalin-like phospholipase domain containing 3; SAIGE, scalable and accurate implementation of generalized mixed model; SCAHC, Southern California Alcohol-associated Hepatitis Consortium; SNP, single nucleotide polymorphism; SNV, single nucleotide variant; *TOX4*, TOX high mobility group box family member 4; TREAT, translational research and evolving alcohol-associated hepatitis treatment.

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.hepjournal.com.

Approach and Results: Exome-sequencing of patients with AH (N = 784) and heavy drinking controls (N = 951) identified an exome-wide significant association for AH at patalin-like phospholipase domain containing 3, as previously observed for AC in genome-wide association study, although with a much lower effect size. Single nucleotide polymorphisms (SNPs) of large effect size at inducible T cell costimulatory ligand (*ICOSLG*) (Chr 21) and *TOX4/RAB2B* (Chr 14) were also exome-wide significant. *ICOSLG* encodes a co-stimulatory signal for T-cell proliferation and cytokine secretion and induces B-cell proliferation and differentiation. TOX high mobility group box family member 4 (*TOX4*) was previously implicated in diabetes and immune system function. Other genes previously implicated in AC did not strongly contribute to AH, and the only prominently implicated (but not exome-wide significant) gene overlapping with alcohol use disorder was alcohol dehydrogenase 1B (*ADH1B*). Polygenic signals for AH were observed in both common and rare variant analysis and identified genes with roles associated with inflammation.

Conclusions: This study has identified 2 new genes of high effect size with a previously unknown contribution to alcohol-associated liver disease and highlights both the overlap in etiology between liver diseases and the unique origins of AH.

INTRODUCTION

Excessive alcohol use leads to alcohol use disorder (AUD) and pathology affecting organs throughout the body, including brain, heart, liver, pancreas, gut, and immune system.[1] AUD is a pressing public health concern, with approximately 29.5 million people in the United States meeting the criteria for AUD in 2022.[2] Annually, more than 140,000 people die from alcohol-associated causes in the United States, making alcohol the fourth leading preventable cause of death.[3]

Alcohol can lead to clinically distinct diseases affecting the same organ, and alcohol-associated liver disease (ALD) is an example of how alcohol can affect the same organ in different ways. End-stage liver disease occurs in only some people with AUD and can arise via different routes. The steatosis and accompanying biochemical and metabolic effects of alcohol-associated fatty liver disease are usually fully reversible. On the other hand, alcohol-associated hepatitis (AH) is a unique and clinically perilous condition marked by inflammation of the liver. Acute symptoms of AH may include abdominal pain, nausea, vomiting, fever, and jaundice. AH can occur in patients with underlying alcohol-associated cirrhosis (AC); however, not all cases of AH progress to cirrhosis.[4] AC, which is irreversible, whether preceded by AH or not, is marked by scarring of the liver and may lead to a constellation of other problems, including portal hypertension, liver failure, and HCC.

Nevertheless, the expression of alcohol-associated diseases is highly variable, with only a fraction of heavy drinkers developing severe forms such as AH. Estimates vary, but while approximately 90% of heavy drinkers develop alcohol-associated fatty liver disease, only 10%–20% progress to AC and 10%–35% to AH. AH is distinct both histologically and physiologically. Although steatosis and hepatocellular ballooning, Mallory-Denk bodies, and cell degeneration are not unique or pathognomonic for AH, infiltration of inflammatory cells, including neutrophils, into the liver parenchyma is characteristic, and in severe AH, fibrosis and perivenular necrosis of cells in the centrilobular region of the liver acinus ensues.[5] In AH, an array of biochemical and hematologic findings associated with liver dysfunction can be seen; however, more distinct, although not exclusive to AH, are leukocytosis, reflecting an inflammatory response in the liver, and elevation of acute phase reactants in serum.[6] Thus, the histopathology and biochemistry of AH suggest the importance of processes shared with other ALD, plus additional distinct inflammatory components.

Genome-wide association studies (GWAS) implicating a wide constellation of genes have advanced our

understanding of AC genetics, metabolic-associated fatty liver disease (MAFLD), and AUD.[7] These genomic approaches have been only partially successful in identifying genes and alleles contributing to ALD.[8] The partial success resides in the fact that most of the variance for either AC or AUD remains unexplained.[9] Unusually for GWAS of a common, etiologically complex disease, the genes implicated in AC have much larger effect sizes than typically observed (Table 1). Notably, of the 4 genes implicated in AC by meta-analysis, 3 (patalin-like phospholipase domain containing 3 [*PNPLA3*], *MARC1*, and *FAF2*) are directly connected to lipid droplet formation,[10] and one (*TM6SF2*) is involved in regulating lipid metabolic processes.[11] Although inflammation and hepatitis may precede AC, it is notable that of the top genes linked to AC by GWAS, *FAF2* is the only one potentially proinflammatory in nature.[12]

Based on the histologic and physiologic distinctiveness of AH, we hypothesized that genomic analysis of AH would identify shared genetic liability due to genes common to AH, AC, and AUD as well as reveal genetic vulnerability factors not shared with AC or AUD. To test this, we sequenced exomes of patients with AH and compared them to patients with moderate and severe AUD (The Diagnostic and Statistical Manual of Mental Disorders 5 criteria) without AC and lacking biochemical or inflammatory markers of AH. This was accomplished through a multicenter consortium to accumulate sufficient patients with AH verified via common clinical methods and criteria. We performed 3 types of analysis: a common variant analysis that parallels GWAS in its ability to capture effects of loci proximal to a gene, analyses of polygenic load against a precompiled list of liver disease genes and previously curated gene networks, and a rare variant analysis capable of detecting effects of uncommon alleles that are unlikely to be accurately imputed in GWAS. Identification of genes predisposing individuals to AH can unveil the molecular drivers of AH and potential therapy targets.

METHODS

Study cohorts

Patients with AH

Patients with AH (N = 784) were recruited via 4 National Institutes on Alcohol Abuse and Alcoholism (NIAAA)-sponsored AH studies: TREAT (translational research and evolving alcohol-associated hepatitis treatment), DASH (defeat alcohol-associated steatohepatitis), SCAHC (Southern California Alcohol-associated Hepatitis Consortium), and InTeam (Integrated Approaches for Identifying Molecular Targets in Alcoholic Hepatitis). All research was conducted in accordance with both the Declarations of Helsinki and Istanbul. The study was approved by the Cleveland Clinic Institutional Review Board (approval #12-833), and written informed consent was obtained from all participants. A total of 784 AH samples were selected for this study based on clinical attributes and data completeness (Supplemental Table S1, http://links.lww.com/HEP/I582). Heavy drinkers without AH or detectable ALD (N = 951) were recruited at the NIH Clinical Center via the Office of the Clinical Director and Laboratory of Neurogenetics of the NIAAA (Bethesda, MD) and via the TREAT consortium (98-AA-009; 05-AA-0121). Detailed information for each consortium site and diagnostic criteria for AH diagnosis, or heavy drinking without alcohol-related liver disease, is described in Supplemental Materials, http://links.lww.com/HEP/I582.

Demographic data, medical history, and questionnaires were collected at each site to estimate the quantity and pattern of alcohol consumption. Venous blood was analyzed for complete blood counts, comprehensive metabolic panels, hepatic function tests, and coagulation tests at the local pathology and chemistry lab at each site (Table 2).

TABLE 1 A partial listing of genes and SNPs implicated by genome-wide association study of alcohol cirrhosis (AC)

Gene	Chr	Locus	Type	OR	Lipid droplet formation
PNPLA3	22	rs738409	I148M	2.4–5.8	Yes
	—	rs2294915	Intronic	2.07	—
HSD17B13	4	rs72613567	T > TA	0.67	Yes
	—	rs4607179	Downstream	0.57	—
FAF2	5	rs374702773	Intronic T del	0.79	Yes
	—	rs11134977	Intronic	—	—
SERPINA1	14	rs28929474	E366H	2.47	—
SUGP1/ TM6SF2	19	rs10401969	T > C	1.49	Yes
HNRNPUL1	19	rs15052	T > C	1.3	—
MARC1	1	rs2642438	T165A	0.76	Yes
TMC4/MBOAT7	19	rs626283	G > C	1.35	Yes

TABLE 2 Demographic and clinical characteristics of patients from contributing consortium sites

	Heavy drinking controls (N = 951)	AH (N = 784)	Total (N = 1735)
Data source			
DASH	0	313 (39.9)	313 (18.0)
INTEAM	0	185 (23.6)	185 (10.7)
NIAAA LNG	799 (84.0)	0	799 (46.1)
TREAT	152 (16.0)	192 (24.5)	344 (19.8)
SCAHC	0	94 (12.0)	94 (5.4)
Age			
Mean (SE)	39.51 (0.41)	47.37 (0.38)	43.05 (0.30)
Sex (self-reported)			
Male	570 (59.9)	515 (65.7)	1085 (62.6)
Female	381 (40.1)	267 (34.0)	648 (37.3)
N-Miss	0	2 (0.3)	2 (0.1)
Race (self-reported)			
White	516 (54.3)	703 (89.7)	1219 (70.3)
African American/Black	357 (37.5)	46 (5.9)	403 (23.2)
Asian/Asian American	34 (3.6)	4 (0.5)	38 (2.2)
Native American/American Indian	5 (0.5)	9 (1.1)	14 (0.8)
Native Hawaiian/Pacific Islander	2 (0.2)	0	2 (0.1)
Other	21 (2.2)	19 (2.4)	40 (2.3)
Unknown	16 (1.7)	2 (0.3)	18 (1)
N-Miss	0	1 (0.1)	1 (0.1)
Ethnicity			
Hispanic	48 (5.0)	174 (22.2)	222 (12.8)
Not Hispanic	889 (93.5)	598 (76.3)	1487 (85.7)
Unknown	14 (1.5)	11 (1.4)	25 (1.4)
N-Miss	0	1 (0.1)	1 (0.1)
MELD Score			
Mean (SE)	2.59 (0.13)	22.54 (0.27)	12.41 (0.29)
AST, U/L			
Mean (SE)	35.09 (1.31)	136.9 (4.1)	80.75 (2.32)
ALT, U/L			
Mean (SE)	33.94 (1.16)	56.2 (1.83)	43.93 (1.07)
Bilirubin, mg/dL			
Mean (SE)	0.53 (0.01)	14.4 (0.55)	6.75 (0.30)
WBC, k/mcl			
Mean (SE)	6.47 (0.07)	10.71 (0.25)	8.37 (0.13)
Platelets, k/mcl			
Mean (SE)	247.28 (2.10)	100.08 (3.64)	185.24 (2.66)
Creatinine, mg/dL			
Mean (SE)	0.84 (0.01)	1 (0.03)	0.91 (0.01)
INR			
Mean (SE)	1.00 (0.00)	1.79 (0.02)	1.39 (0.01)
Neutrophil k/mcl			
Mean (SE)	3.70 (0.06)	1744.65 (521.68)	451.93 (136.13)
Lymphocyte, k/mcl			
Mean (SE)	1.97 (0.04)	249.35 (38.14)	60.19 (9.53)

TABLE 2. (*continued*)

	Heavy drinking controls (N = 951)	AH (N = 784)	Total (N = 1735)
Monocyte, k/mcl			
Mean (SE)	0.50 (0.01)	139.62 (22.71)	33.24 (5.64)
Basophil, k/mcl			
Mean (SE)	0.04 (0.00)	0.86 (0.08)	0.10 (0.01)
BMI			
Mean (SE)	27.03 (0.18)	28.98 (0.28)	27.8 (0.16)
BMI imputed			
Mean (SE)	27.03 (0.18)	28.98 (0.28)	27.8 (0.16)

Abbreviations: AH, alcohol-associated hepatitis; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; DASH, Defeat Alcohol-associated Steatohepatitis; INR, international normalized ratio; LNG, Laboratory of Neurogenetics; MELD, the Model for End-Stage Liver Disease; NIAAA, National Institutes on Alcohol Abuse and Alcoholism; SCAHC, Southern California Alcohol-associated Hepatitis Consortium; SNP, single nucleotide polymorphism; TREAT, Translational Research and Evolving Alcohol-associated hepatitis Treatment; WBC, white blood cell.

Exome sequencing and quality control

A total of 58 Mb of the genomic target was enriched using the AmpliSeq Exome RDY kit, including 46.3 Mb representing 86.3% of RefSeq coding regions, and sequenced on the Ion Proton platform (Thermo Fisher Scientific, Waltham, MA). Data quality control criteria were >80x average depth and >80% uniformity of coverage. Reads were mapped to UCSC build hg19. Full details of the data processing pipeline are provided in Supplemental Materials, http://links.lww.com/HEP/I582. Subjects were sequenced to a mean depth of 112x (Supplemental Figure S1A, http://links.lww.com/HEP/I582). For the 1,085,982 unique single nucleotide variant (SNVs) called, the average and median nucleotide read depths were 102x and 89x, respectively. (Supplemental Figure S1B, http://links.lww.com/HEP/I582). A call rate cutoff <80% (59,709 SNVs) was used to focus analyses on the most reliably and consistently genotyped variants. An additional 110,851 SNVs were filtered out: 99.6% were rare with an allelic frequency <0.01 and without a matching gnomAD variant, with the remaining 1025 SNVs) having an observed allele frequency ≥ 0.1 but a reported gnomAD allele frequency <0.01, suggesting sequencing error. For the remaining 914,397 SNVs used in the linkage analyses, the correlation of allelic frequencies between our cohort and gnomAD was high ($r=0.997$) (Supplemental Figure S2, http://links.lww.com/HEP/I582) (inclusion of the 1025 likely artifactual SNVs generated a correlation $r=0.996$). Due to the presence of a partial LINE-1 element insertion (rs1229469148) at the *PNPLA3* locus,[13] upstream of the exon containing the ALD relevant rs738409 single nucleotide polymorphism (SNP), sequencing derived genotypes obtained from this amplicon were unreliable (Supplemental Materials, http://links.lww.com/HEP/I582). For this reason, rs738409 was individually genotyped via 5′-exonuclease assay, and these data were used in subsequent analyses. Data for the 3 other variants identified within the affected amplicon were excluded from analyses due to this unique technical problem. Indels were excluded from analyses due to the high error rate of sequence-based genotyping of small indels.[14]

Participants were clustered by ancestry as AFR (African), AMR (Admixed American), EAS (East Asian), or EUR (European) based on 250 ancestry informative exonic SNPs in UT-AIM250.[15] The 1000 Genomes Project (https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/) was used for reference genotypes, and data were analyzed using Structure v2.3.4.[16] Functional annotation of SNVs was performed with Ensembl Variant Effect Predictor[17] (VEP, release 93).

The quality and completeness of SNV genotypes are detailed in Supplemental Results, http://links.lww.com/HEP/I582. Briefly, by a technically independent retest of $>180,000$ loci on SNP arrays, genotype concordance was 0.998, and genotype completeness (the fraction of loci genotyped) was 0.941. The number of SNVs/person (primarily heterozygous) were 45,984, 46,206, 46,957, and 55,184 for people predominantly of EUR, EAS, AMR, and AFR ancestry, respectively (Supplemental Figure S4A, http://links.lww.com/HEP/I582). Overall, the average nucleotide heterozygosity was 5.4×10^{-4} (Supplemental Figure S4B, http://links.lww.com/HEP/I582), consistent with our previous report.[18] Most SNVs detected (0.825) were rare (<0.01), as expected, given the number of people sequenced (Supplemental Table S2, http://links.lww.com/HEP/I582). On an individual basis, and as corrects for the number of people sequenced, the ratio of rare SNVs to common SNPs was 0.023 (1047/45,287) (Supplemental Figure S4C, http://links.lww.com/HEP/I582), in line with previous reports.[19] Higher densities of SNVs, including common SNVs, were observed in intronic regions, and common SNVs in coding regions were more likely to be synonymous (0.165) than missense (0.105) or nonsense (0.055) (Supplemental Table S2, http://links.lww.com/HEP/I582).

Subject identities were confirmed by correlation ≥ 0.95 to genotypes obtained from the Infinium quality

control Array-24 v1.0, OmniExpress, or OmniExpressExome array (Illumina, Inc. San Diego, CA), and agreement for sex by chrX genotypes.

Statistical analyses

SNPs with minor allele frequency (MAF) ≥ 0.01 and existent in gnomAD (v2.1.1) were analyzed with a scalable generalized linear mixed model for additive genotypic effects with sex and ethnicity group as covariates using scalable and accurate implementation of generalized mixed model (SAIGE).[20] The genomic inflation factor (λ) was the ratio of the median of the observed distribution of *p*-values to the expected median. A *p*-value threshold of 7×10^{-7} was considered exome-wide significant.[21]

Rare variants were tested at the gene level using SNP-set kernel association optimal unified test[22] implemented in SAIGE-GENE+[23] using aggregated genotypes (minimum minor allele count >10) and sex and ethnicity groups as covariates. Loss-of-function (LoF) SNVs were weighted as 1, and missense SNVs were weighted based on AlphaMissense pathogenicity scores,[24] with all other SNVs weighted as 0.01. Analysis was performed for 2 variant categories ("LoF + missense" and "LoF + missense + others"). Genes were considered significant after Bonferroni correction for the number of genes in each test and the number of variant categories (2). Rare variants (MAF <0.01) were also analyzed individually for association using SAIGE.[20]

The top 100 significant genes (*p*-value ≤ 0.001) from common variant and gene-wise analysis were analyzed for enrichment in the GWAS Catalog (2023) using Enrichr.[25] Additionally, all tested genes were ranked by *p*-value beta value (direction of effect) and analyzed for gene network enrichment using gene set enrichment analysis (GSEA) (v.4.3.2).[26] For selected SNVs, Open Targets (https://www.opentargets.org/) were used to search for associated traits via PheWAS in large data sets. Multitissue eQTLs were identified for selected SNVs using GTEx (https://gtexportal.org/home/).

RESULTS

Exome-wide association of AH to common variants

For exome-wide association analysis, the 160,074 common (MAF ≥ 0.01) SNPs passing quality control were used. A scalable generalized mixed model implemented in SAIGE adjusting for both case-control imbalance and genetic relatedness was used to analyze the association between the AH/control binary phenotype versus genotype. An additive genetic model with sex and ethnicity group as covariates was used. This identified 5 SNP associations at exome-wide significance ($p<7\times10^{-7}$) and an additional 11 SNPs with *p*-values suggestive of association ($p<10^{-5}$) (Table 3). This included 3 SNPs at *PNPLA3*, a previously identified liver disease risk gene[9] (Figure 1, Table 3): rs738409 ($p=3.75\times10^{-9}$, beta = 0.548), rs4823173 ($p=5.14\times10^{-7}$, beta = 0.499), and rs34879941 ($p=5.30\times10^{-7}$, beta = 0.493). These 3 SNPs, along with several other nearby variants, are in moderate to high linkage disequilibrium, with $r^2>0.5$ (Supplemental Figure S6A, http://links.lww.com/HEP/I582). Several SNPs at nearby *SAMM50* and *PARVB* also generated small but nonsignificant *p*-values ($p\leq 0.0001$). Liver disease associations with *SAMM50* and *PARVB* have previously been reported,[27] but it remains uncertain whether variation at these 2 genes represents risk factors independent of *PNPLA3*.[28]

In addition, 2 novel, genetically independent exome-wide significant (*p*-values $<7\times10^{-7}$) variants were identified implicating inducible T cell costimulatory ligand (*ICOSLG*), rs13052840 ($p=4.59\times10^{-8}$, beta = 0.515) and TOX high mobility group box family member 4 (*TOX4*)/*RAB2B* rs17516362 ($p=1.25\times10^{-7}$, beta = −0.495) (Figure 1, Table 3, Supplemental Figure S6, http://links.lww.com/HEP/I582). rs17516362 lies within an evolutionarily conserved genomic region containing many transcription factor binding sites (Supplemental Figure S7A, http://links.lww.com/HEP/I582), and this region encompasses the bidirectional core promoter shared by *TOX4* and *RAB2B*. *ICOSLG* encodes an inducible T-cell co-stimulator ligand, which is a co-stimulatory signal for T-cell proliferation and cytokine secretion that induces B-cell proliferation and differentiation. In patients with AH, ICOSLG could potentially modulate local tissue responses or responses dependent on memory T-cell function. *ICOSLG* was previously associated with inflammatory bowel disease[29] and the development of primary biliary cirrhosis, an autoimmune liver disease characterized by inflammatory cell infiltration of intralobular biliary ducts, with consequent biliary duct damage.[30]

TOX4 (Chr 14) encodes an insulin receptor-independent regulator of hepatic glucose production and is also involved in T-cell development.[31] Expression of *TOX4* in the liver is increased in humans and experimental animals with diabetes.[32] Diabetes, together with alcohol misuse, strongly contributes to liver disease burden, which led investigators to include diabetes in a recent gene score (burden) formula for AC.[9] RAB2B, a member of Ras protein family, has been reported to be involved in cytosolic DNA-induced innate immune responses.[33] The rs17516362-G allele predicts higher expression for *TOX4* and *RAB2B* in multiple tissues, as observed in GTEx (Supplemental Figure S7C, http://links.lww.com/HEP/I582) and other studies (https://fivex.sph.umich.edu/variant/eqtl/14_21476796), which is consistent with a report that individuals with AH show higher expression of *RAB2B* in CD14+ cells compared to healthy controls.[34] The lower expression *TOX4*/*RAB2B* rs17516362-C allele had frequencies of 0.23 in

TABLE 3 Top common SNVs associated with AH with p-value $< 1\times10^{-5}$

Gene	VEP	Snp151	Chr-Pos-Ref-Alt	*p*	BETA	SE
PNPLA3	I148M	rs738409	chr22-44324727-C-G	**3.75E-09**	0.548	0.095
ICOSLG	Q329Q	rs13052840	chr21-45649848-T-C	**4.59E-08**	0.515	0.096
TOX4/RAB2B	5′-UTR	rs17516362	chr14-21944955-G-C	**1.25E-07**	−0.495	0.096
PNPLA3	Intron	rs4823173	chr22-44328730-G-A	**5.14E-07**	0.499	0.102
PNPLA3	Intron	rs34879941	chr22-44332878-C-T	**5.30E-07**	0.493	0.101
MESP1	Intron	rs2305442	chr15-90293701-G-A	8.06E-07	0.479	0.100
PNPLA3	Intron	rs2072906	chr22-44333172-A-G	1.31E-06	0.474	0.101
ADAMTS7	Intron	rs3971703	chr15-79092481-G-C	1.36E-06	0.442	0.094
SIGLEC15	F273L	rs2919643	chr18-43419003-T-C	4.21E-06	−0.478	0.107
SLC25A19	Intron	rs2306218	chr17-73282299-C-T	4.71E-06	0.423	0.095
CREG2	5′-UTR	rs116287156	chr2-102003962-T-C	5.00E-06	0.467	0.106
CFAP410	5′-UTR	rs73374031	chr21-45759197-G-C	5.07E-06	−0.438	0.099
ADRBK1	Intron	rs10896164	chr11-67052466-G-A	7.57E-06	0.484	0.112
CFAP410	R11R	rs11870	chr21-45759045-C-T	8.89E-06	−0.426	0.099
AP5Z1	Intron	rs3750012	chr7-4828593-T-C	8.94E-06	−0.551	0.128
LRP1B	Intron	rs1429365	chr2-141747249-A-T	9.42E-06	−0.380	0.089

VEP: gene functional annotations predicted by VEP; Snp151: NCBI dbSNP build 151; Chr-Pos-Ref-Alt: UCSC hg19; p-value, Beta, and SE were outputs from SAIGE. Bold p-value numbers are exome-wide significant ($p < 7\times10^{-7}$).

our patients with AH, 0.37 in heavy drinking controls, and an intermediate frequency of 0.297 in gnomAD, suggesting a protective effect (beta = −0.495; OR = 0.61 [95% CI: 0.50–0.74]) against AH. Notably, high TOX4 expression is listed in The Human Protein Atlas as prognostically unfavorable for liver cancer survival (https://www.proteinatlas.org/ENSG00000092203-TOX4/pathology/liver+cancer).

Additionally, 13 other SNPs potentially implicating *MESP1*, *PNPLA3*, *ADAMTS7*, *SIGLEC15*, *SLC25A19*, *CREG2*, *CFAP410*, *ADRBK1*, *AP5Z1*, and *LRP1B* with p-values $< 10^{-5}$ were of interest (Table 3). Several other genomic regions contained clusters of subthreshold ($p < 10^{-4}$) *SNPs.* These included a chr11 region containing *CEND1*, *SLC25A22*, *PIDD*, *PNPLA2*, *EFCAB4A*, a chr7 region containing *AP5Z1*, and *RADIL* (Supplemental Figure S6D, http://links.lww.com/HEP/I582), and a chr11 region containing *ARHGAP42*. Many of these topmost SNPs were also implicated in GWAS of AH-related traits (Supplemental Table S4, http://links.lww.com/HEP/I582). We also identified a subthreshold association to AH for the alcohol dehydrogenase 1B (*ADH1B*) rs1229984-C allele, which encodes the lower activity form of the ADH1B enzyme. This association with AH occurs within the context of heavy drinking and so cannot be explained solely on the basis of alcohol consumption.

Gene and pathway enrichment analysis

To evaluate the sources of polygenic contributions to AH, gene enrichment analysis was performed using GSEA against a 227 gene liver disease set, which included 214 genes parsed from GWAS_Catalog.v1.0.3 and 13 genes from https://pheweb.sph.umich.edu/. We ranked all 18,705 genes represented in this study against these 227 genes, choosing the SNP with the most significant p-value for each gene and using the beta value to observe the directional effects of the reference and nonreference alleles at each locus. In this rank-order test, the significance of enrichment (y-axis of Figure 2) is evaluated via a permutation test against the number of genes in the gene set. As stated, few of these genes were genome-wide significant. Nevertheless, enrichment ($p < 0.0001$) for liver disease–related genes was observed relative to the nonreference allele both at the left (risk) and right (protective) ends of the 18,705 p-value ranked whole gene list (Figure 2A), suggesting a polygenic contribution to AH of many liver disease–related genes. These p-values were determined via simulation via a limited number of trials (ie, 10,000) for the maximum positive or negative enrichment, and therefore, the p-value of 10^{-4} for liver disease–related genes likely represents an underestimate of significance.

The 18,705 p-value ranked genes were also analyzed via GSEA against 791 curated gene sets representing canonical pathways in WikiPathways. Although many of these WikiPathways contain few genes and are thereby unlikely to yield a significant enrichment score unless enrichment is very high, we ranked the 18,705 gene-specific p-values against all 791 WikiPathways genesets to not bias the analysis. Pathways with significant enrichment at the $p < 10^{-4}$ level included “vitamin D sensitive calcium signaling in

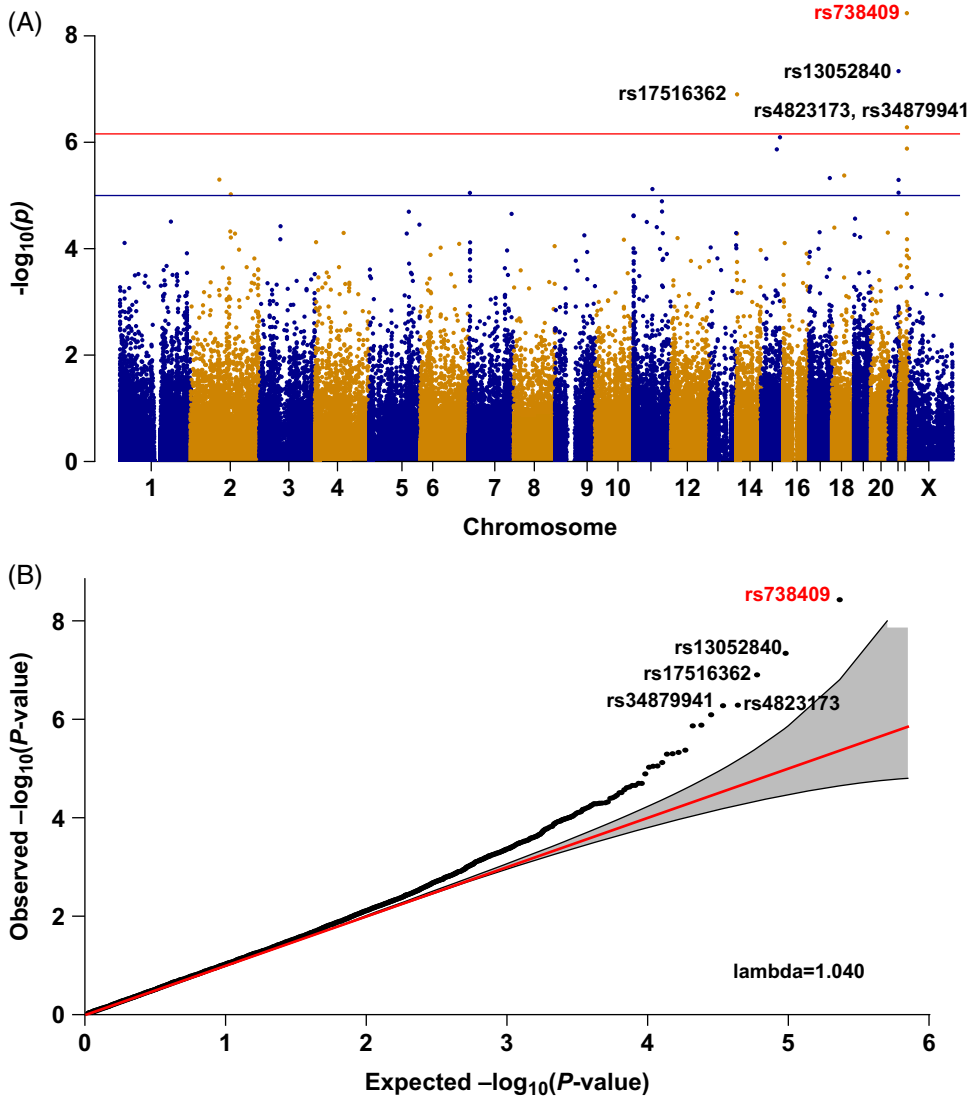

FIGURE 1 Exome wide association to alcohol-associated hepatitis of 160,074 common SNVs from exome sequencing. (A) Manhattan plot of −log10 raw *p*-values. The red line represents the exome-wide significance threshold of $p = 7\times10^{-7}$. The blue line represents suggestive significance at $p = 10^{-5}$; (B) Quantile-quantile (QQ)-plot of randomly expected versus observed *p*-values, with 0.95 CI. $\lambda = 1.040$.

depression" a 36 gene pathway (Figure 2B), and "pyrimidine metabolism and related diseases," a 20 gene pathway (Figure 2C). A glycerolipids pathway (WP4722, 22 genes) was also enriched ($p = 0.0042$) (Figure 2E). The enrichments in these pathways were driven by often rare nonreference alleles at several genes in the pathway, and these could be either AH-promoting or protective. No protective nonreference alleles were seen for the vitamin D calcium signaling, pyrimidine metabolism, or glycerolipid gene pathways (Figure 2B, C, E). Notably, the sterol regulatory element-binding proteins signaling pathway (69 genes, WP1982), a well-known pathway involved in lipid metabolism and immune function,[35] was enriched at the bottom of the ranked genes (Figure 2D, NES = −2.14, $p < 0.0001$), suggesting protective effects of nonreference alleles that were more abundant in controls than cases. In contrast, an alcohol dependence geneset (184 genes) extracted from the GWAS Catalog showed no enrichment (Figure 2H), and we observed that 288 GWAS catalog "Alcohol drinking" genes were not strongly enriched (NES = 1.494, $p = 0.013$), indicating that these "Alcohol drinking" genes are unlikely to have a large polygenic contribution to AH (Figure 2G). Finally, 236 genes implicated in inflammatory disease from The Human Phenotype Ontology (https://hpo.jax.org/app/) demonstrated moderate enrichment (NES = 1.53, $p = 0.006$) (Figure 2F). Supplemental Table S5, http://links.lww.com/HEP/I582, lists additional relevant gene categories in the GWAS Catalog 2023 that are enriched ($p < 0.01$) using Enrichr for the 100 topmost significant AH genes (Supplemental Table S6, http://links.lww.com/HEP/I582) in this study. Prominent among these, and including several with very high odds ratios, are enrichments for such liver-related GWAS catalog

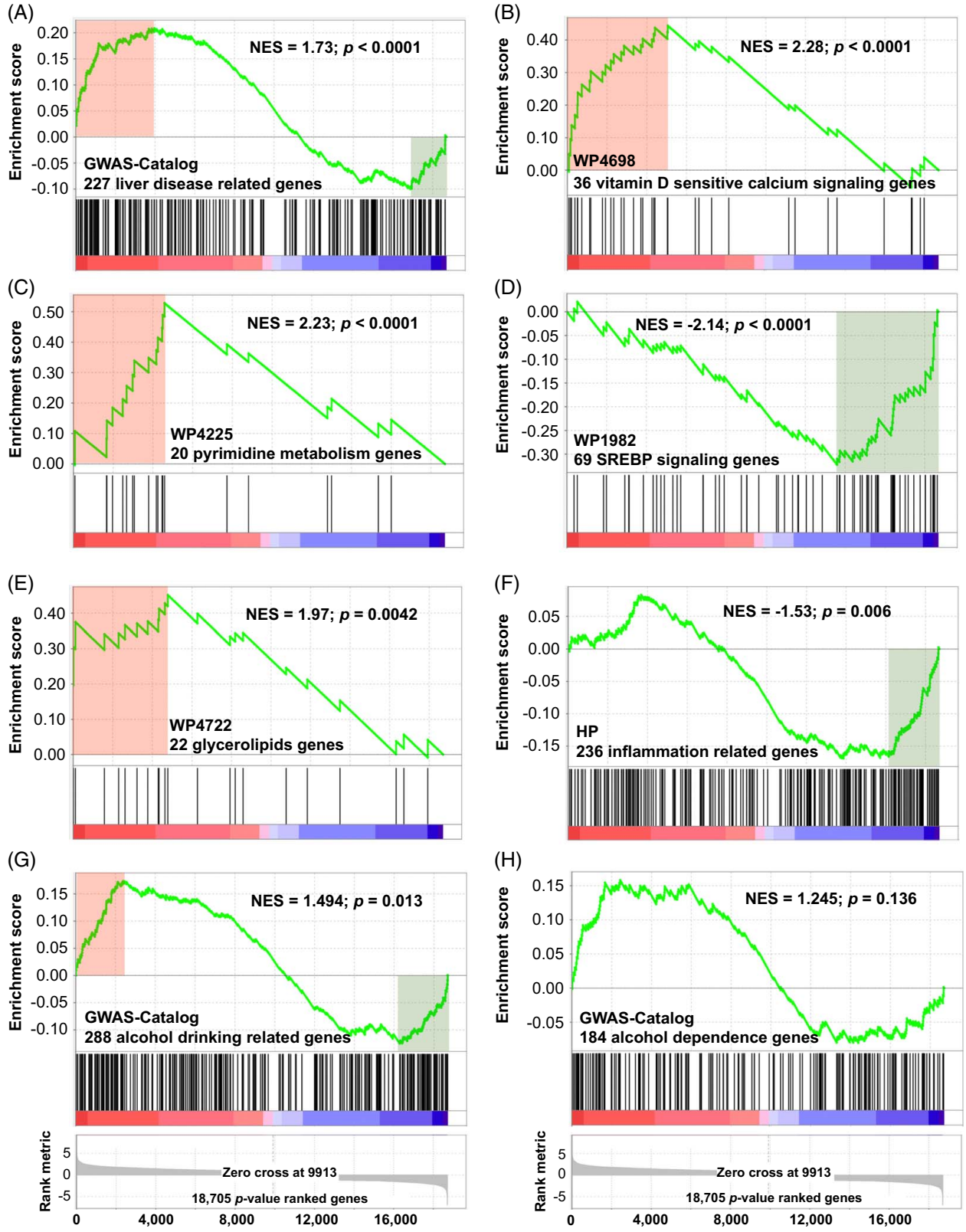

FIGURE 2 Enrichment in alcohol-associated hepatitis (AH) of exome-wide significant and subsignificant genes detecting potential types of polygenic contributions to AH. In each GSEA graph, the top *p*-values from 18,705 genes are rank ordered, and cumulative enrichment in AH for the nonreference (alternative) allele against rank is plotted. Both positive (pink) and negative (green) enrichment values indicate association of the pathway to AH, positive values reflecting greater risk effects of the nonreference alleles, and negative values reflecting protection conferred by the nonreference alleles. *p*-values were estimated empirically via permutation analyses with 10^4 trials, and if *p*-value $< 10^{-4}$, a unidirectional enrichment score of that magnitude was not observed in any of the 10^4 trials. (A) *p*-values of 227 liver disease-related genes precompiled from

PheWeb and genome-wide association study Catalog (v1.0.3) and against all rank-ordered 18,705 genes, showing polygenic enrichment of AH for liver disease genes (NES = 1.73, $p < 10^{-4}$), with symmetrical and thus approximately equivalent contributions of reference and alternative alleles. (B) Enrichment (NES = 2.28, $p < 10^{-4}$) of "vitamin D sensitive calcium signaling in depression," a 36 gene Wikipathway (WP4698). This enrichment is unidirectional, being driven by the risk effects of alternative alleles. (C) Enrichment (NES = 2.23, $p < 10^{-4}$) of "pyrimidine metabolism and related diseases," a 20 gene Wikipathway (WP4225). This enrichment is also unidirectional, being driven by the risk effects of alternative alleles. (D) "SREBP signaling," a 69 Wikipathway (WP1982) was significantly enriched at the bottom of ranked genes (NES = −2.14, $p < 0.0001$). (E) Enrichment (NES = 1.97, $p = 0.0042$) of "glycerolipids and glycerophospholipids," a 22-gene Wikipathway (WP4722). This enrichment is also unidirectional, being driven by risk effects of alternative alleles. (F) Moderate significant enrichment of "Inflammation" genes from The Human Phenotype Ontology (HP:0004386, HP:0012115, HP:0030151, HP:0033196, HP:0200123) in AH (NES = −1.53, $p = 0.006$). The bottom rank metric is shared by each panel. (G) Modest enrichment in AH of "alcohol drinking" (288 genes) implicated by genome-wide association study Catalog (EFO_0004329, NES = 1.494, $p = 0.013$). (H) 184 "alcohol dependence" genes from genome-wide association study Catalog (MONDO_0007079) was not enriched either (NES = 1.245, $p = 0.136$). Abbreviation: NES, Normalized Enrichment Score.

listings as hepatic fat, NAFLD, and serum hepatic enzyme levels.

Rare variant analysis

Rare variants from 16,660 genes were tested using 2 variant categories ("LoF + missense" and "LoF + missense + others"). Genes were considered significant at $p = 1.5 \times 10^{-7}$ following Bonferroni correction for the number of genes and for the number of variant categories tested.[36] No genes reached significance at this threshold: The top genes ($p < 0.001$) were *HYAL3*, *RNF144B*, *FAAP20 (C1orf86)*, *RADIL*, *SCTR*, *POU2AF2*, *KIF16B*, *GBA3*, *CAPRIN2*, *EIF4ENIF1* (Supplemental Figure S8, http://links.lww.com/HEP/I582). *RADIL* also had 5 common SNPs with *p*-values (< 0.001), potentially implicating a chromosome 7 region containing both it and *AP5Z1* (Supplemental Figure S6D, http://links.lww.com/HEP/I582). Among the top genes, *HYAL3* encodes a member of the hyaluronidase family of enzymes that degrade hyaluronic acid, a major glycosaminoglycan of the extracellular matrix and a biomarker of liver fibrosis,[37] *RNF144B* (Ring Finger Protein 144B) inhibits LPS-induced inflammatory responses,[38] expression of *SCTR* is strongly correlated with proinflammatory chemokines in a cohort of patients with different stages of ALD,[39] and *GBA3* encodes a cytosolic enzyme that can hydrolyze several types of glycosides and that is involved in sphingolipid metabolism

TABLE 4 Top rare SNVs associated with AH with *p*-value $< 5 \times 10^{-4}$

Gene	VEP	snp151 (Ref/Alt)	AH_af	CK_af	*p*	BETA	SE
GGA1	5′-UTR	rs117099857 (G/A)	0.0102	0.0005	2.81E-05	2.925	0.726
RAD51B	Intron	rs34436700 (A/T)	0.0070	0.0032	8.46E-05	2.272	0.604
ABO	G229D	rs56116432 (C/T)	0.0000	0.0058	1.00E-04	−3.496	0.940
CHST10	V34A	rs138889445 (A/G)	0.0057	0.0037	1.37E-04	2.265	0.622
FAAP20	T120K	rs41315340 (G/T)	0.0108	0.0005	1.47E-04	2.828	0.781
PITPNM2	T926T	rs149074091 (C/T)	0.0006	0.0042	1.53E-04	−2.877	0.797
PITPNM2	Intron	rs375395578 (G/A)	0.0000	0.0026	1.92E-04	−3.907	1.100
PITPNM2	H430H	rs143865800 (G/A)	0.0000	0.0037	2.20E-04	−3.499	0.996
SNAPIN	3′-UTR	rs41305092 (C/T)	0.0019	0.0095	2.71E-04	−1.973	0.571
UBE4B	Intron	rs187355549 (C/T)	0.0089	0.0005	3.02E-04	2.568	0.749
TRIM31	Q245R	rs138900010 (T/C)	0.0070	0.0063	3.07E-04	1.825	0.533
HYAL3	R178H	rs139756596 (C/T)	0.0019	0.0084	3.22E-04	−1.967	0.576
LPIN1	I541V	rs148499322 (A/G)	0.0000	0.0032	3.81E-04	−3.435	1.021
NMRK2	A223T	rs35701787 (G/A)	0.0051	0.0053	3.84E-04	1.971	0.586
RADIL	D307N	rs148761575 (C/T)	0.0045	0.0116	4.18E-04	−1.529	0.458
SEMA3F	Intron	rs182660760 (C/T)	0.0006	0.0058	4.31E-04	−2.533	0.760
UGGT1	A669V	rs145894335 (C/T)	0.0000	0.0037	4.35E-04	−3.188	0.958
SLMAP	Q630R	rs35029175 (A/G)	0.0032	0.0000	4.38E-04	3.380	1.016
CDAN1	Intron	rs117316699 (C/T)	0.0000	0.0068	4.73E-04	−3.226	0.976
SCYL3	L108L	rs147935620 (C/T)	0.0000	0.0037	4.78E-04	−3.131	0.948
C8G	V196I	rs150744928 (G/A)	0.0032	0.0000	4.88E-04	3.398	1.030

VEP: gene functional annotations predicted by VEP; Snp151: NCBI dbSNP build 151; af: allelic frequency; *p*-value, Beta, and SE were outputs from SAIGE. Abbreviation: SNP, single nucleotide polymorphism.

potentially playing an important role in liver diseases.[40] Thus, although none of the top genes identified through aggregate rare variant analysis remained significant following correction for multiple testing, based on function, several could plausibly contribute to AH.

Individual rare variants (MAF <0.01) analyzed for association using a generalized mixed linear model[20] yielded no exome-wide significant hits ($p<7\times10^{-7}$). Rare SNVs with ($p<5\times10^{-4}$) were located at genes *GGA1*, *RAD51B*, *ABO*, *CHST10*, *FAAP20 (C1orf86)*, *PITPNM2*, *SNAPIN*, *UBE4B*, *TRIM31*, *HYAL3*, *LPIN1*, *NMRK2*, *RADIL*, *SEMA3F*, *UGGT1*, *SLMAP*, *CDAN1*, *SCYL3*, and *C8G* (Table 4). Nine other rare missense SNVs ($p<0.01$) had high AlphaMissense predicted pathogenicity scores (>0.5) and beta values >0 (Supplemental Table S7, http://links.lww.com/HEP/I582). In each case, their allelic frequency was higher in AH than in heavy drinking controls and furthermore were higher than the allelic frequencies in gnomAD, implying a functional impact on AH. Via Open Targets (https://www.opentargets.org/) we also noted that several of these rare SNVs have previously shown associations with AH-relevant phenotypes in very large data sets (Supplemental Table S8, http://links.lww.com/HEP/I582), phenotypes including serum alkaline phosphatase and alanine transaminase (*ABO* G229D, *MPDZ*, and *CLIP4*) and also such AH-related phenotypes as body mass, apolipoprotein A1 levels, and immune response phenotypes (Supplemental Table S8, http://links.lww.com/HEP/I582).

DISCUSSION

ALD encompasses a spectrum of hepatic histopathological changes ranging from steatosis to alcohol hepatitis to cirrhosis.[7] The pathogenesis of ALD is multifactorial, and in addition to genetic factors, alcohol-induced hepatocyte damage, reactive oxygen species, and gut-derived microbial components contribute to steatosis, fibrosis, and inflammatory cell recruitment that is most characteristic of AH.[41] AH is characterized by robust and dysregulated liver inflammation after drinking alcohol and may be reversible. On the other hand, AC can be insidious and consequent to AH, is defined as an end state in which scar tissue has replaced healthy tissue, and cirrhosis produced by any mechanism is generally regarded as irreversible. Genetic predisposition plays a powerful role in the consumption of alcohol and, thereby, the amount and pattern of drinking.

It is therefore perhaps surprising that this study on AH did not identify genes associated with AUD or strongly contributing to drinking; however, this is likely due to our experimental design of using heavy drinking/AUD controls whereby the effects of genes associated with compulsive alcohol consumption would be common between the groups. Consequently, although ALD requires alcohol exposure, the genetic vulnerability to ALD appears to mainly be mediated by other innate or acquired pharmacodynamic differences. An important exception is *ADH1B*, which emerges as a genetic risk factor for AH in addition to both AUD and ALD.[42,43] Furthermore, our data indicate that although there are overlaps between AH and AC, AH represents an etiologically distinct disease. Most of the top hits for AC that implicate genes involved in lipid droplet formation were not observed in this study of AH, even at the subthreshold level. Even *PNPLA3*, which appears to play a role in both AC and AH, has a lower effect size in AH (OR 1.7 for AH in this study vs. OR 2.7–7.3 in AC[44]); however, it should be noted that analysis of polygenic contributions of genes implicated in liver function by GWAS shows that such genes do contribute polygenically to AH. In a larger cohort, it is likely that some of these genes would be exome-wide or genome-wide significant for AH, although with a smaller effect size than observed for AC or MAFLD. Meanwhile, our common and rare variant analyses point to genes mediating inflammatory and immune processes as involved in AH and a small polygenic contribution of genes influencing heavy drinking.

In the past decade, genomic studies, including several GWAS, have identified genetic risk factors for AC[8]; however, fewer genetic studies and no GWAS have focused on AH.

We exome-sequenced 784 patients with AH and 951 heavy drinkers without ALD to fill this knowledge gap. This study, although relatively small, generated genome-wide significant findings and, via unbiased enrichment analysis, provides strong evidence for polygenic influences on AH. Critically, the genes implicated in this study point to the genetic distinctiveness of AH, with limited overlap with AC or AUD, and inflammatory and immune genes playing a larger role in AH. Nevertheless, a limitation of this study is the sample size, although this was ameliorated by using heavy drinking controls, which ensured that all subjects were exposed to the exogenous effector, alcohol. A second limitation of this study is the absence of an available second dataset to confirm and validate our findings. All genes detected in this study are genes of large effect. Genes of large effect are of obvious clinical and mechanistic interest; however, as shown by the enrichment analysis, there are also many other genes that contribute polygenically to AH. Larger GWAS, replication studies, and meta-analyses can identify which of these subthreshold genes represent true positives, and it is these statistically robust genes that will ultimately be most useful for gene network and multiomic analyses and potentially to identify new therapeutic targets for AH. A particular strength of this study is the use of heavy drinking/AUD controls. This meant that genes associated with the development of AH were targeted, rather than those for alcohol consumption, which would have been highlighted had nondrinking or light drinking controls been

used. This is illustrated by our observation that it is the Arg48 allele of *ADH1B* (rs1229984C) that is associated with a higher risk of AH. This allele encodes a lower activity ADH1B enzyme, which produces acetaldehyde less rapidly after drinking, and for this reason, is a strong predictor of heavier alcohol consumption and problematic alcohol use. However, even when using heavy drinking controls, we find that the *ADH1B* Arg48 allele is also a risk allele for AH (beta = 1.11, $p = 10^{-5}$) (Supplemental Table S6, http://links.lww.com/HEP/I582), which suggests a previously unrecognized contribution of this locus. A potential explanation for this finding is that the *ADH1B* Arg48 allele has also been associated, via GWAS, with decreased serum vitamin D levels, vitamin D inhibiting TH-1 T-cell inflammatory response while enhancing TH2-mediated anti-inflammatory responses, and thus the Arg48 allele may be associated with an elevated hepatic inflammatory response. The enrichment of vitamin D-sensitive calcium signaling genes in the GSEA analysis (Figure 2B; $p < 0.0001$) is coherent with this hypothesis.

The *PNPLA3* missense variant rs738409 (I148M) is one of the most significant genetic factors for AC and other liver diseases, including MAFLD and HCC, and rs738409 G/G homozygotes were observed to be at higher risk of mortality due to AH.[45] For AH, we again find that *PNPLA3* is the gene of the largest effect and significance, although the OR for rs738409 is somewhat lower than observed in AC but of similar magnitude as for ALD and MAFLD (Supplemental Figure S9, http://links.lww.com/HEP/I582).

Unlike most of the genes associated with AC (Table 1), many of the genes implicated in our study of AH are potentially involved in immune function. The exome-wide significant variants implicating 3 novel genes (*ICOSLG*, *TOX4*, and *RAB2B*) for AH were also of large effect (rs13052840 OR = 1.67, 95% CI: 1.39–2.02 and rs1751636 OR = 0.61, 95% CI: 0.50–0.74, respectively) (Supplemental Figure S9, http://links.lww.com/HEP/I582). *TOX4* has important immune functions but is also involved in hepatic glucose production and is activated in diabetic liver. *ICOSLG* is predicted to be involved in T-cell receptor signaling pathway and positive regulation of IL-4 production, and the same SNP (rs13052840) we associate with AH has been associated with inflammatory bowel disease[46] ($p = 5.9 \times 10^{-9}$) and has a strong downregulatory effect on *ICOSLG* expression (https://fivex.sph.umich.edu/variant/eqtl/21_44229965). In GWAS, the strongest association for *ICOSLG* is to eosinophilia ($p = 10^{-34}$, 0.025 SD unit increase), which is sometimes observed in hepatobiliary and gastrointestinal disorders and in the liver can be a feature of primary biliary cirrhosis.[47] Furthermore, many other top genes that we identified for AH, albeit subthreshold for most, may also modulate immune responses. These include *MESP1*, *CREG2*, *ADRBK1*, *AP5Z1*, *PITPNM2*, *UGGT1*, *CDAN1*, *SCYL3*, *C8G*, *CEP128*, and *CTRL* (Supplemental Tables S4, http://links.lww.com/HEP/I582 and S8, http://links.lww.com/HEP/I582), and we also observe evidence for polygenic loading of subthreshold genes to inflammatory disease. Genetically predisposed dysregulation of immune genes can lead to a heightened inflammatory response to alcohol-induced injury, potentially explaining the acuity and severity of AH.

In this study, we used whole-exome sequencing, as it allowed us to focus on the contribution of common functional cSNPs and rare cSNPs, which might contribute to the elusive "missing heritability" observed in many complex diseases. In line with previous sequencing-based studies and large genomic data resources such as gnomAD, most SNVs detected in our study of 1735 individuals are rare (82.5%), and on a per-individual basis, 2.3% SNVs were rare, and 97.7% were common. An important caveat to any association study is that a linked SNP or SNV does not necessarily implicate a particular gene but can implicate several, and other evidence may be required to refine findings to the genic level. For example, the exome-implicated *TOX4/RAB2B* locus is a *cis* eQTL for both *TOX4* and *RAB2B.*

In this study, no individual rare variants or gene-aggregated variants reached exome-wide significance for association to AH, probably due to sample size. However, several of the top genes with subthreshold SNVs are both functionally plausible and associated with relevant phenotypes in large PheWAS (Supplemental Table S8, http://links.lww.com/HEP/I582). The rare variant data generated here thereby represents a resource that can be used cumulatively or meta-analytically with future data sets.

In summary, our results demonstrate that there is only limited genetic overlap between AH and other alcohol-related liver diseases. We have identified novel genes associated with AH and significant polygenic loading for a series of liver-related and immune genes, improving our understanding of the genetic architecture of AH.

DATA AVAILABILITY STATEMENT

Sequence data available at dbGAP—accession number is phs003659. Additional data is available in Supplemental Materials, http://links.lww.com/HEP/I582.

AUTHOR CONTRIBUTIONS

All authors made substantial contributions to study conception and design, acquisition of data, or analysis and interpretation of data. Qiaoping Yuan, Colin Hodgkinson, Laura E. Nagy, and David Goldman participated in drafting the article or revising it critically for important intellectual content. Qiaoping Yuan, Colin Hodgkinson, Xiaochen Liu, Bruce Barton, Nancy Diazgranados, Melanie Schwandt, Timothy Morgan, Ramon Battaler, Suthat Liangpunsakul, Laura E. Nagy, and David Goldman gave final approval for the version to be published.

ACKNOWLEDGMENTS

The authors thank Trina Norden-Krichmar for contribution to methodology and funding acquisition; Daniel M. Rotroff for funding acquisition; Tae-Hwi Schwantes for funding acquisition, data curation, and resources; Marco Antonio Abreu for data curation and formal analysis; Annette Bellar and Vai Pathak for formal analysis.

COLLABORATORS

Joseph Alcorn, Aliya Asghar, Annette Bellar, Divya Birudaraju, Andy Borst, Gregory Botwin, Hema Buddha, Naga P. Chalasani, Daniel Chen-Kang Chao, Lavanya Cherukuri, Megan Comerford, Ryan Cook, David W. Crabb, Sheena Cruz, Sherrie Cummings, Srinivasan Dasarathy, John Donovan, Lakye Edwards, Wayne Fleischman, Monique French, Ted George, Gregory J. Gores, Jessica Clare Gozum, Sajad Hamal, Yvonne Horneffer, Carol Jones, Patrick S. Kamath, Neil Kaplowitz, Barry Katz, Spencer Lourens, Lauren MacHarg, Craig J. McClain, Arthur J McCullough, Cheryl Marietta, Susan Milstein, Mack Mitchell, David Nelson, Amy Olofson, Andras Orosz, Yuxin Ouyang, Kyle L Poulsen, Puneet Puri, Svetlana Radaeva, Christy Rico, Arun J. Sanyal, Romil Saxena, Amanda Schimek, Vijay H. Shah, Andrew Stolz, Hui Sun, Gyongyi Szabo, John Tayek, Vikas Verma, Susan Walker, Sarah Wilder, Qigui Yu.

FUNDING INFORMATION

This work was supported the Alcohol Hepatitis Genomics Consortium, NIAAA UO1 grant 5U01AA026264 (MPI Laura Nagy and Ramon Bataller) and NIAAA Division of Intramural Clinical and Biological Research (Grants Z1A AA000130 and Z1A AA000466). Additional support was from R01AA030312, U01 AA026917, UH2/UH3 AA026903, U01AA026817, and VA Merit Award 1I01CX000361 (Suthat Liangpunsakul) and U01AA026938, P50AA024333 (Laura E. Nagy).

CONFLICTS OF INTEREST

Ramon Bataller consults for GlaxoSmithKline and Nov Nordisk. He is on the speakers' bureau for AbbVie, Boehringer Ingelheim, and Gilead. Suthat Liangpunsakul consults for Durect, Korro Bio, and Surrozen. The remaining authors have no conflicts to report.

ORCID

Qiaoping Yuan https://orcid.org/0000–0002–5880–8702
Colin Hodgkinson https://orcid.org/0000–0002–5365–8119
Xiaochen Liu https://orcid.org/0000–0001–6710–2771
Bruce Barton https://orcid.org/0000–0001–7878–8895
Nancy Diazgranados https://orcid.org/0000–0002–7415–8455
Melanie Schwandt https://orcid.org/0000–0002–8355–4900
Timothy Morgan https://orcid.org/0000–0003–1328–0307
Ramon Bataller https://orcid.org/0000–0002–1119–7799
Suthat Liangpunsakul https://orcid.org/0000–0002–6504–8123
Laura E. Nagy https://orcid.org/0000–0002–0580–2809
David Goldman https://orcid.org/0000–0002–1724–5405

REFERENCES

1. Ohashi K, Pimienta M, Seki E. Alcoholic liver disease: A current molecular and clinical perspective. Liver Res. 2018;2:161–72.
2. SAMHSA, Center for Behavioral Health Statistics and Quality. 2022 National Survey on Drug Use and Health. Alcohol use disorder in past year: among people aged 12 or older; by age group and demographic characteristics, numbers in thousands, 2021 and 2022.
3. Esser MB, Leung G, Sherk A. Estimated deaths attributable to excess alcohol use among US adults aged 20 to 64 years, 2016 to 2019. JAMA Network Open. 2022;5:e2239485.
4. Wu X, Fan X, Miyata T, Kim A, Cajigas-Du Ross CK, Ray S, et al. Recent advances in understanding of pathogenesis of alcohol-associated liver disease. Annu Rev Pathol. 2023;18:411–38.
5. Altamirano J, Miquel R, Katoonizadeh A, Abraldes JG, Duarte–Rojo A, Louvet A, et al. A histologic scoring system for prognosis of patients with alcoholic hepatitis. Gastroenterology. 2014;146:1231–9.
6. Avitabile E, Díaz A, Montironi C, Pérez-Guasch M, Gratacós-Ginès J, Hernández-Évole H, et al. Adding inflammatory markers and refining National Institute on Alcohol Abuse and Alcoholism Criteria improve diagnostic accuracy for alcohol-associated hepatitis. Clin Gastroenterol Hepatol. 2023;21:3080–8.
7. Konkwo C, Chowdhury S, Vilarinho S. Genetics of liver disease in adults. Hepatol Commun. 2024;8:e0408.
8. Norden-Krichmar TM, Rotroff D, Schwantes-An TH, Bataller R, Goldman D, Nagy LE, et al. Genomic approaches to explore susceptibility and pathogenesis of alcohol use disorder and alcohol-associated liver disease. Hepatology. 2023. doi:10.1097/HEP.0000000000000617. [Online ahead of print].
9. Whitfield JB, Schwantes-An TH, Darlay R, Aithal GP, Atkinson SR, Bataller R, et al. the GenomALC Consortium. A genetic risk score and diabetes predict development of alcohol-related cirrhosis in drinkers. J Hepatol. 2022;76:275–82.
10. Schwantes-An TH, Darlay R, Mathurin P, Masson S, Liangpunsakul S, Mueller S, et al. Genom, A. L. C. C. Genome-wide association study and meta-analysis on alcohol-associated liver cirrhosis identifies genetic risk factors. Hepatology. 2021;73:1920–31.
11. Mahdessian H, Taxiarchis A, Popov S, Silveira A, Franco-Cereceda A, Hamsten A, et al. TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic droplet content. Proc Natl Acad Sci USA. 2014;111:8913–8.
12. Imai N, Suzuki M, Ishizu Y, Kuzuya T, Honda T, Hayashi K, et al. Hepatocyte-specific depletion of ubiquitin regulatory X domain containing protein 8 accelerates fibrosis in a mouse non-alcoholic steatohepatitis model. Histochem Cell Biol. 2017;148:219–7.

13. Leníček M, Šmíd V, Pajer P, Nazarova A, Dvořák K, Subhanová I, et al. Clinically silent LINE 1 insertion in the PNPLA3 gene may impede genotyping of the p.I148M variant. Sci Rep. 2021;11: 20924.
14. Gampawar P, Saba Y, Werner U, Schmidt R, Muller-Myhsok B, Schmidt H. Evaluation of the performance of AmpliSeq and SureSelect exome sequencing libraries for ion proton. Front Genet. 2019;10:856.
15. Wang LJ, Zhang CW, Su SC, Chen HIH, Chiu YC, Lai Z, et al. An ancestry informative marker panel design for individual ancestry estimation of Hispanic population using whole exome sequencing data. BMC Genomics. 2019;20:1007.
16. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155: 945–59.
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. Yuan Q, Zhou Z, Lindell SG, Higley JD, Ferguson B, Thompson RC, et al. The rhesus macaque is three times as diverse but more closely equivalent in damaging coding variation as compared to the human. BMC Genet. 2012;13:52.
19. 1000 Genomes Project Consortium. A global reference for human variation. Nature. 2015;526:68–74.
20. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet. 2018;50:1335–41.
21. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. Eur J Hum Genet. 2016;24:1202–5.
22. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. Biostatistics. 2012; 13:762–5.
23. Zhou W, Bi W, Zhao Z, Dey KK, Jagadeesh KA, Karczewski KJ, et al. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. Nat Genet. 2022;54: 1466.
24. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science. 2023;381:eadg7492.
25. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with Enrichr. Curr Protoc. 2021;1:e90.
26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS. 2005;102:15545–50.
27. Kitamoto T, Kitamoto A, Yoneda M, Hyogo H, Ochi H, Nakamura T, et al. Genome-wide scan revealed that polymorphisms in the PNPLA3, SAMM50, and PARVB genes are associated with development and progression of nonalcoholic fatty liver disease in Japan. Hum Genet. 2013;132:783–92.
28. Nischalke HD, Schmalz F, Buch S, Fischer J, Möller C, Matz-Soja M, et al. Genetic variation of SAMM50 is not an independent risk factor for alcoholic hepatocellular carcinoma in Caucasian patients. Int J Mol Sci. 2022;23:15353.
29. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet. 2017;49:256–61.
30. Aiba Y, Yamazaki K, Nishida N, Kawashima M, Hitomi Y, Nakamura H, et al. Disease susceptibility genes shared by primary biliary cirrhosis and Crohn's disease in the Japanese population. J Hum Genet. 2015;60:525–31.
31. Wang T, Zhao R, Zhi J, Liu Z, Wu A, Yang Z, et al. Tox4 regulates transcriptional elongation and reinitiation during murine T cell development. Commun Biol. 2023;6:613.
32. Wang L, Yu J, Zhou Q, Wang X, Mukhanova M, Du W, et al. TOX4, an insulin receptor-independent regulator of hepatic glucose production, is activated in diabetic liver. Cell Metab. 2022;34:158–70 e155.
33. Takahama M, Fukuda M, Ohbayashi N, Kozaki T, Misawa T, Okamoto T, et al. The RAB2B-GARIL5 complex promotes cytosolic DNA-induced innate immune responses. Cell Rep. 2017;20:2944–54.
34. Wu X, Fan X, McMullen MR, Miyata T, Kim A, Pathak V, et al. Macrophage-derived MLKL in alcohol-associated liver disease: Regulation of phagocytosis. Hepatology. 2023;77:902–19.
35. DeBose-Boyd RA, Ye J. SREBPs in lipid metabolism, insulin signaling, and beyond. Trends Biochem Sci. 2018;43:358–68.
36. Wightman DP, Savage JE, de Leeuw CA, Jansen IE, Posthuma D. Rare variant aggregation in 148,508 exomes identifies genes associated with proxy dementia. Sci Rep. 2023;13:2179.
37. Orasan OH, Ciulei G, Cozma A, Sava M, Dumitrascu DL. Hyaluronic acid as a biomarker of fibrosis in chronic liver diseases of different etiologies. Clujul Med. 2016;89:24–31.
38. Zhang Z, Zhang L, Wang B, Zhu X, Zhao L, Chu C, et al. RNF144B inhibits LPS-induced inflammatory responses via binding TBK1. J Leukoc Biol. 2019;106:1303–1.
39. Aguilar-Bravo B, Rodrigo-Torres D, Ariño S, Coll M, Pose E, Blaya D, et al. Ductular reaction cells display an inflammatory profile and recruit neutrophils in alcoholic hepatitis. Hepatology. 2019;69:2180–95.
40. Mari M, Fernandez-Checa JC. Sphingolipid signalling and liver diseases. Liver Int. 2007;27:440–50.
41. Bataller R, Arab JP, Shah VH. Alcohol-associated hepatitis. N Engl J Med. 2022;387:2436–48.
42. Kranzler HR, Zhou H, Kember RL, Vickers Smith R, Justice AC, Damrauer S, et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat Commun. 2019;10:1499.
43. 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.
44. 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;110:846–56.
45. Atkinson SR, Way MJ, McQuillin A, Morgan MY, Thursz MR. Homozygosity for rs738409:G in PNPLA3 is associated with increased mortality following an episode of severe alcoholic hepatitis. J Hepatol. 2017;67:120–7.
46. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015;47:979–86.
47. Zuo L, Rothenberg ME. Gastrointestinal eosinophilia. Immunol Allergy Clin North Am. 2007;27:443–55.

How to cite this article: Yuan Q, Hodgkinson C, Liu X, Barton B, Diazgranados N, Schwandt M, et al. Exome-wide association analysis identifies novel risk loci for alcohol-associated hepatitis. Hepatology. 2025;81:1304–1317. https://doi.org/10.1097/HEP.0000000000001027