Received: 14 February 2025 | Accepted: 16 March 2025

DOI: 10.1097/HC9.0000000000000728

ORIGINAL ARTICLE

OPEN

A comprehensive evaluation of candidate genetic polymorphisms in a large histologically characterized MASLD cohort using a novel framework

Aaron Hakim[1,2] | **Kung-Hung Lin**[3] | **Tae-Hwi Schwantes-An**[3,4] | **Marco Abreu**[3,4] | **Jingyi Tan**[5] | **Xiuqing Guo**[5] | **Katherine P. Yates**[6] | **Luca Lotta**[7] | **Niek Verweij**[7] | **Rohit Loomba**[8] | **David E. Kleiner**[9] | **Jeffrey B. Schwimmer**[10] | **Jerome I. Rotter**[5] | **Naga P. Chalasani**[3]

[1]Division of Gastroenterology and Hepatology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

[2]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

[3]Division of Gastroenterology and Hepatology, Indiana University School of Medicine, Indianapolis, Indiana, USA

[4]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana, USA

[5]Department of Pediatrics, The Institute for Translational Genomics and Population Sciences, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, California, USA

[6]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

[7]Regeneron Genetics Center, Tarrytown, New York, USA

[8]Department of Family Medicine and Public Health, Division of Epidemiology, University of California at San Diego, San Diego, California, USA

[9]Laboratory of Pathology, National Cancer Institute, Bethesda, Maryland, USA

[10]Department of Pediatrics, Division of Gastroenterology, Hepatology, and Nutrition, University of California, San Diego School of Medicine, La Jolla, California, USA

Correspondence
Naga P. Chalasani, Division of Gastroenterology and Hepatology, Indiana University School of Medicine, 550 North University Blvd, Indianapolis, IN 46202, USA.
Email: nchalasa@iu.edu

Abstract

Background: There is a substantial heritable component to metabolic dysfunction–associated steatotic liver disease (MASLD), and several genetic variants that promote MASLD development or associate with its severity have been reported. These associations vary in terms of their effect size and degree of replication.

Methods: We developed a framework to classify previously identified MASLD genetic polymorphisms into 4 tiers based on effect size and extent of replication in the literature. We tested the association between "tier 1" single-nucleotide polymorphisms (OR $\geq$1.5, replicated in $>$2 independent studies) and biopsy measures of MASLD severity in a large, well-characterized

Abbreviations: 1KGP, 1000 Genome project; GRS, genetic risk score; GWAS, genome-wide association studies; HWE, Hardy–Weinberg Equilibrium; MASH, metabolic dysfunction–associated steatohepatitis; MASLD, metabolic dysfunction–associated steatotic liver disease; NAS, NAFLD activity score; NASH-CRN, Nonalcoholic Steatohepatitis Clinical Research Network; NHGRI, National Human Genome Research Institute; PC, principal component; RGC, Regeneron Genetics Center; SNP, single-nucleotide polymorphism; WES, whole-exome sequencing.

Aaron Hakim, Kung-Hung Lin, and Tae-Hwi Schwantes-An contributed equally to this work.

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.hepcommjournal.com.

histologic cohort of MASLD patients (n = 3094).

Results: Across 19 "tier 1" variants reflecting 11 genetic loci, only those in the *PNPLA3-SAMM50-PARVB* locus showed significant associations with biopsy-proven fibrosis severity and NAFLD activity score; the highest risk was for the rs738409 p.I148M variant in *PNPLA3*. A genetic risk score based on "tier 1" variants, as well as a previously developed genetic risk score based on variants in *PNPLA3*, *TM6SF2*, and *HSD17B13*, were both associated with fibrosis and NAFLD activity score, but these results were driven entirely by *PNPLA3* rs738409.

Conclusions: Our study provides a framework to prioritize evaluation of genetic polymorphisms for future replication efforts and demonstrates that in a large case-only cohort, histologic severity of MASLD is only robustly associated with the presence of variation in *PNPLA3* among known candidate genes. These findings may have implications for patient risk stratification based on the presence of *PNPLA3* rs738409.

Keywords: fibrosis, genetics, GWAS, steatohepatitis, steatotic liver disease

INTRODUCTION

Metabolic dysfunction–associated steatotic liver disease (MASLD) is a common cause of chronic liver disease characterized by excess fat in the liver.[1] Metabolic dysfunction–associated steatohepatitis (MASH) is the progressive form of MASLD in which patients have hepatic steatosis, inflammation, and cytologic hepatocyte ballooning with varying degrees of fibrosis.[2] The gold standard for the definite diagnosis of MASH and/or evaluation of its severity is a liver biopsy.[2,3] Current regulatory guidance for drug development in MASH also suggests that, for the purposes of accelerated approval, only liver histologic endpoints in clinical trials, such as resolution of steatohepatitis and/or improvement in liver fibrosis scores, are reasonably likely to predict clinical benefit.[4] However, in clinical practice, noninvasive alternatives to liver biopsy, such as circulating biomarkers, imaging methods, and liver stiffness measurements, are often used for diagnosis and risk stratification.[2,3]

Individual susceptibility to MASLD involves both environmental and genetic risk factors. Recent estimates suggest that MASLD heritability ranges from 16% to 50%,[5,6] and genome-wide association studies (GWAS) have identified several risk loci for MASLD.[7,8] For example, genetic polymorphisms in *PNPLA3* were initially found to associate with radiologically assessed hepatic triglyceride content.[9] Subsequently, over 50 independent studies have demonstrated the association of variants in *PNPLA3* with clinically relevant phenotypes, including biopsy-proven steatohepatitis, fibrosis stage, cirrhosis, HCC, and the development of liver decompensation events; this has motivated the development of antisense oligonucleotide therapies to lower the expression of *PNPLA3* in patients homozygous for the risk allele.[8,10,11] Other MASLD genetic variants have been identified in GWAS or in candidate gene studies using physician diagnosis codes,[12] clinical biochemistry parameters such as ALT,[6,13,14] or smaller cohorts of histologically defined disease with genetically matched controls.[15–17] Many of these associations vary in terms of their effect size and degree of replication. Notably, the sample sizes of genetic association studies in steatotic liver disease have been much smaller compared to studies in other cardiometabolic diseases.[18,19]

Relatively few studies have evaluated the impact of MASLD genetic variants on histologic disease severity, a key prognostic endpoint.[20] We therefore utilized the Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN), one of the largest, multicenter, prospective cohorts of biopsy-confirmed fatty liver disease, including over 3000 participants, to evaluate previously identified MASLD genetic polymorphisms.[21] We first developed a framework to classify previously identified MASLD genetic variants into 4 tiers based on effect size and replication status in the literature, using the National Human Genome Research Institute (NHGRI) GWAS Catalog Database[22] and literature search up to October 2023. We classified "tier 1" single-nucleotide polymorphisms (SNPs) as those having an OR ≥ 1.5 and replicated in >2 independent studies; we identified 19 SNPs reflecting 11 genetic loci meeting these

criteria. We then evaluated the association of these "tier 1" SNPs with 2 clinically relevant phenotypes of interest: fibrosis severity and NAFLD activity score (NAS). We also assessed whether combining "tier 1" SNPs into a genetic risk score (GRS) was associated with fibrosis and NAS, and how the strength of this association compared with a previously reported *PNPLA3-TM6SF2-HSD17B13* GRS,[23] and *PNPLA3* rs738409 alone.

EXPERIMENTAL PROCEDURES

Literature search and prioritization of previously identified genetic variants

To gather polymorphisms with previously reported associations with MASLD, we used the National Human Genome Research Institute (NHGRI) GWAS Catalog Database (https://www.ebi.ac.uk/gwas/) to identify the polymorphisms derived from GWAS.[22] Using the keywords "non-alcoholic fatty liver disease" (EFO_0003095), "hepatic steatosis" (HP_0001397), and "non-alcoholic steatohepatitis" (EFO_1001249), all available data of associations were downloaded on August 26, 2023, resulting in a list of 393 SNPs. Next, we performed a literature search for publications up to October 17, 2023, on PubMed to expand the list to SNPs derived from candidate gene association studies. Using the combination of keywords: (NAFLD) AND (single nucleotide polymorphisms), without limits, restrictions, or published search filters applied, we retrieved 769 articles in English or with English abstracts (until October 17, 2023). SNPs reported to be significantly associated with MASLD and replicated in one or more studies were categorized into "tier 1" through "tier 4", based on the number of replication articles and the effect size of their association with MASLD. "Tier 1" SNPs were defined as OR ≥ 1.5 and replicated in more than 2 articles. "Tier 2" SNPs were defined as OR ≥ 1.5 and replicated in only 2 articles. "Tier 3" SNPs were defined as OR < 1.5 and replicated in more than 2 articles. "Tier 4" SNPs had OR < 1.5 and were replicated in only 2 articles.

Study design, participants, and histologic evaluation

This was a cross-sectional analysis of adults and children enrolled in the Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN); this prospective cohort has been described previously.[21] Briefly, patients with biopsy-confirmed MASLD were prospectively recruited at multiple medical centers across the United States between 2004 and 2020. Patients were included in this cohort based on the presence of NAFLD as defined by the previously published NASH CRN scoring system.[24] Informed consent in writing was obtained from each patient, and the study protocol was confirmed to ethical guidelines of the 1975 Declaration of Helsinki. The diagnosis of MASLD was based on $>5\%$ of hepatocytes containing macrosteatosis and exclusion of significant alcohol consumption (>20 g/d for women, >30 g/d for men) within 2 years of the initial biopsy. All liver biopsies were reviewed in a blinded fashion by the NASH CRN Pathology Committee and scored according to the NASH CRN Scoring System.[24] Fibrosis stage was assessed from 0 (no fibrosis) to 4 (cirrhosis). Steatosis was graded from 0 to 3, ballooning from 0 to 2, and inflammation from 0 to 3; the sum of these measures reflects the NAS.[24] During the study period, the NASH CRN cohort consisted of 28% who were children under the age of 18. The NASH CRN has established a research collaboration with the Regeneron Genetics Center (RGC) through an ancillary study, which was approved by the NASH CRN Steering Committee. This collaboration follows the NIH Genomic Data Sharing policies and procedures. As part of this collaboration, the RGC conducted whole-exome sequencing at no cost to the NASH CRN and returned results to the NASH CRN. An analytical team established by the NASH CRN team at Indiana University conducted analyses described in this paper.

Genotyping

Genotype data were generated by Regeneron Genomic Center (RGC) using RGC's WES (whole-exome sequencing) plus targeted genotyping assay. DNA samples were sequenced for 20× targeted WES and over 1.4 million targeted variants. Genotype data were cleaned using our previously published data QC pipeline.[25] In short, (1) samples and variants with $<95\%$ call rates were dropped, (2) samples with discordant reported and genetic sex were dropped, (3) expected and unexpected relationships between samples were checked, (4) SNPRelate[26] was used to cluster NASH CRN samples with AFR, AMR, and EUR ancestry groups from the 1000 Genome project (1KGP),[27] and (5) genetic ancestry stratified check for Hardy–Weinberg Equilibrium (HWE) was done to remove those variants that deviated from HWE in any of the 3 groups. QC passing genotype data was then imputed to the TOPMED reference panel using the TOPMED Imputation Server.[28] Imputed genotypes were further checked for R^2 (≥ 0.8, a measure of imputation quality) and violation of HWE.

Association of "tier 1" variants with histologic severity, genetic risk score, and statistical analyses

To minimize confounding due to allele frequency differences between genetic ancestry groups, we

performed analyses in samples that clustered with EUR (European continental samples from 1KGP) or AMR (American continental samples from 1KGP) groups. All analyses were adjusted for age, sex, pediatric cohort status, EUR/AMR cluster membership, and the first 2 principal components (PCs) of genotypes. We adjusted analyses for PC1 and PC2 as additional PCs contributed a minimal incremental proportion of variation explained. We assessed the association of each of 19 "tier 1" variants with fibrosis stage and NAS as a continuous variable, using a multivariable linear regression model with genotype fitted with an additive model. A Bonferroni adjusted significance level of $p < 0.0026$ was used for the 19 variants. We also used multivariable logistic regression to test the association of each tier 1 variant with dichotomized fibrosis stage and NAS: presence of advanced fibrosis ($\geq$F3 fibrosis) versus rest and predicted steatohepatitis (NAS ≥ 4) versus rest. We also adjusted for age, sex, pediatric cohort status, EUR/AMR cluster membership, and the first 2 PCs of genotypes. We performed sensitivity analyses adjusting for the same set of covariates in the initial test in (1) EUR adults and pediatric cohort without adjusting for EUR/AMR cluster membership, (2) AMR adults and pediatrics cohort without adjusting for EUR/AMR cluster membership, (3) EUR/AMR adults cohort without adjusting for pediatric status, and (4) EUR/AMR pediatric cohort without adjusting for pediatric status. Next, a GRS was calculated by summing the number of risk alleles each participant carried across the "tier 1" variants (one point for each risk allele, eg, *PNPLA3* G allele, *TM6SF2* T-allele, HSD17B13 T-allele, etc.). We compared the strength of association between NAS or fibrosis with the "tier 1" GRS, a previously reported genetic risk score[23] which counts the number of risk alleles from rs738409 (*PNPLA3*), rs58542926 (*TM6SF2*), and rs72613567 (*HSD17B13*), and compared to rs738409 (*PNPLA3*) alone. Statistical analyses were performed in PLINK2[29] and R.[30]

RESULTS

Participant characteristics

During the study period, 3094 participants of EUR/AMR ancestry, including 900 children, were enrolled into the NASH CRN cohort and genotyped for the 19 "tier 1" variants. The median age of the cohort was 39.6 years, 47.7% of the cohort was male, 27.4% of the cohort had diabetes, 25.9% had advanced fibrosis, and 66.6% had NAS ≥ 4. The average fibrosis score at baseline was 1.5, with a NAS score of 4.2. Table 1 summarizes the baseline characteristics for adults and children included in the study.

Prioritization of variants reported in the literature

We gathered polymorphisms with previously reported associations with MASLD in the literature and identified 19 "tier 1" variants representing 11 genetic loci that were replicated in 2 or more publications, and that carried a large effect size (OR ≥ 1.5) (Figure 1 and Table 2). We identified 33 "tier 2" variants representing 29 genetic loci that were replicated in 2 articles in the literature and carried a large effect size. There were 21 "tier 3" variants from 13 loci with small effect size (OR < 1.5) but were replicated in more than 2 papers, and 14 "tier 4" variants representing 12 loci were replicated in 2 articles in the literature and had small effect size. For "tier 1" variants, the median OR for the association with MASLD phenotype was 1.74. There were 12 candidate gene studies that provided support for "tier 1" SNPs, which included studies with a sample size as low as 152 participants, and 40 independent GWAS studies or meta-analyses that provided support for "tier 1" SNPs. Associated MASLD phenotypes for "tier 1 SNPs" included physician-diagnosed MASLD or MASH by ICD-10 codes, elevated hepatic fat by imaging, elevated biomarkers of liver injury, and some liver histologic cohorts (Supplemental Tables S1–S4, http://links.lww.com/HC9/C6).

Association of "tier 1" variants with histologic disease severity

We first evaluated the association of "tier 1" variants with histopathological fibrosis score as a continuous variable. Across the 19 variants, only those in the *PNPLA3–SAMM50–PARVB* gene complex showed statistically significant associations with fibrosis score (Figure 2). The highest risk was for the *PNPLA3* rs738409 C > G SNP, encoding the p.I148M variant ($\beta = 0.19$, $p = 9.49e^{-10}$). The *HSD17B13* rs72613567 polymorphism, encoding for a protein-truncating variant, approached nominal significance for protection against fibrosis ($\beta = -0.08$, $p = 0.051$). We next evaluated the association of "tier 1" variants with NAS as a continuous variable. Only variants in the PNPLA3–SAMM50–PARVB complex showed statistically significant associations, with the strongest association for *PNPLA3* rs738409 ($\beta = 0.27$, $p = 3.74e^{-10}$). Similar findings were noted in logistic regression analyses assessing the associations of "tier 1" variants with the presence of advanced fibrosis ($\geq$F3 fibrosis) or predicted steatohepatitis (NAS ≥ 4) (Table 3). Additionally, sensitivity analyses stratified by pediatric (< 18 years of age) versus adult subjects, and in EUR versus AMR subjects, yielded similar conclusions (Supplemental Table S5, http://links.lww.com/HC9/C6). Specifically, in adults, only variants in the *PNPLA3* genetic locus and in

TABLE 1 Cohort baseline characteristics stratified by EUR (European continental samples from 1KGP), AMR (American continental samples from 1KGP), adult, or pediatric subjects; Yrs (years), peds (pediatric participants), BMI (body mass index, kg/m^2), DM (type 2 diabetes), HTN (hypertension), SD, NAS (NAFLD activity score), and vitE (vitamin E)

	Overall (EUR/AMR) N = 3049	EUR (Adults/Peds) N = 2010	AMR (Adults/Peds) N = 1039	Adults (EUR/AMR) N = 2149	Peds (EUR/AMR) N = 900
Age at enrollment in Yrs, mean (SD)	39.6 (19.9)	47.5 (16)	24.4 (17.8)	50.7 (12)	13.1 (2.8)
% Male (N/Total)	47.7% (1455/3049)	40.8% (820/2010)	61.1% (635/1039)	37.1% (798/2149)	73% (657/900)
BMI, mean (SD)	33.9 (6.5)	34.8 (6.6)	32.1 (6.1)	34.6 (6.5)	32.1 (6.3)
% DM diagnosed (N/Total)	27.4% (833/3045)	34.3% (689/2010)	13.9% (144/1035)	37.1% (798/2149)	3.9% (35/896)
% HTN (N/Total)	39.9% (1214/3045)	52% (1045/2010)	16.3% (169/1035)	54.1% (1163/2149)	5.7% (51/896)
Fibrosis grade, mean (SD)	1.5 (1.3)	1.7 (1.3)	1.2 (1.1)	1.7 (1.3)	1.1 (1)
NAS, mean (SD)	4.2 (1.7)	4.3 (1.7)	4.1 (1.6)	4.3 (1.7)	4.1 (1.5)
% Advanced fibrosis (N/Total)	25.9% (787/3039)	30.5% (610/2002)	17.1% (177/1037)	31% (663/2140)	13.8% (124/899)
% NAS $\geq$ 4 (N/Total)	0% (0/3049)	0% (0/2010)	0% (0/1039)	0% (0/2149)	0% (0/900)
% Hyperlipidemia (N/Total)	42% (1280/3045)	53.7% (1080/2010)	19.3% (200/1035)	56.6% (1217/2149)	7% (63/896)
FIB-4 score, mean (SD)	1.2 (1.2)	1.5 (1.3)	0.7 (1)	1.6 (1.3)	0.3 (0.1)
% VitE use (N/Total)	15.7% (478/3045)	19.8% (398/2010)	7.7% (80/1035)	20.1% (433/2149)	5% (45/896)
% Pioglitazone use (N/Total)	2.4% (72/3041)	3.2% (64/2010)	0.8% (8/1031)	3.4% (72/2149)	0% (0/892)
Steatosis grade, mean (SD)	1.9 (0.9)	1.8 (0.9)	1.9 (0.9)	1.8 (0.9)	2.1 (0.9)
% Lobular inflammation (N/Total)	44.5% (1358/3049)	43.2% (869/2010)	47.1% (489/1039)	45.5% (977/2149)	42.3% (381/900)
% Portal inflammation, (N/Total)	88.3% (2691/3048)	89% (1788/2010)	87% (903/1038)	88.4% (1899/2149)	88.1% (792/899)
% Ballooning (N/Total)	56.1% (1710/3049)	61.9% (1245/2010)	44.8% (465/1039)	64.4% (1385/2149)	36.1% (325/900)
% Mallory hyaline (N /Total)	0.2 (0.4)	0.3 (0.5)	0.1 (0.3)	0.3 (0.5)	0 (0.2)

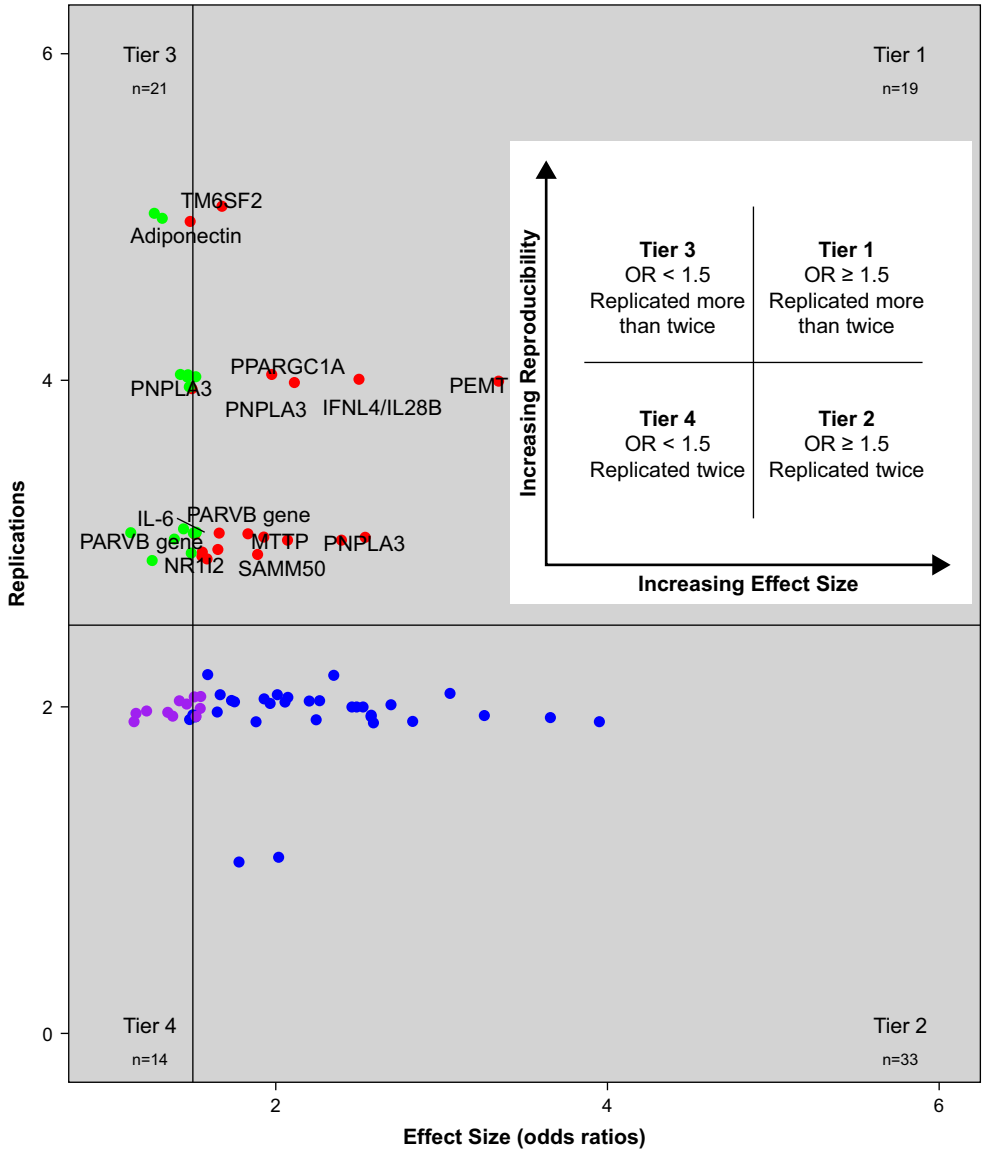

FIGURE 1 Framework used to prioritize candidate MASLD variants based on effect size and degree of replication in the published literature, and a number of variants and genetic loci within each tier. Gene names are omitted in tiers 2–4. Abbreviation: MASLD, metabolic dysfunction–associated steatotic liver disease.

HSD17B13 demonstrated an association with continuous fibrosis, and only variants in *PNPLA3* were associated with the continuous MAS score (Supplemental Table S5, http://links.lww.com/HC9/C6). In pediatrics, only variants in *PNPLA3* were robustly associated with continuous MAS and continuous fibrosis (Supplemental Table S5, http://links.lww.com/HC9/C6).

"Tier 1" variant genetic risk score

Prior studies have analyzed the impact of combining genetic variants on the severity and phenotype of MASH. We found that combining all 19 "tier 1" genetic variants in a "risk allele" model was associated with continuous fibrosis ($p = 1.19e^{-05}$, $r^2 = 0.097$), NAS ($p = 1.01e^{-06}$, $r^2 = 0.02$), advanced fibrosis ($p = 1.93e^{-03}$, $r^2 = 0.082$), and NAS ≥ 4 ($p = 3.58e^{-05}$, $r^2 = 0.014$). This was also true of a previously described 3-allele GRS using variants in PNPLA3, TM6SF2, and HSD17B13 (Table 4). Notably, however, the strength and magnitude of these associations were entirely driven by *PNPLA3* rs738409 (Table 4 and Supplemental Table S6, http://links.lww.com/HC9/C6). Excluding the *PNPLA3* genetic locus, an 18-SNP GRS was not associated with NAS, fibrosis, advanced fibrosis, or NAS ≥ 4. Similarly, a 2 SNP GRS, excluding the *PNPLA3* genetic locus, was not associated with NAS; the strength of the relationship was also diminished for the association with fibrosis, advanced fibrosis, or NAS ≥ 4 (Table 4 and Supplemental Table S6, http://links.lww.com/HC9/C6). These GRS

TABLE 2 List of tier 1 variants, closest gene, rsID, effect allele, coded allele, and estimated allele frequency (EAF) in the EUR/AMR cohort

Tier 1 variant	Gene	rsID	Effect allele	Coded allele	EAF
chr1.11794419.T.G_T	MTHFR	rs1801131	G	T	0.26
chr3.119806650.A.G_A	NR1I2	rs7643645	G	A	0.43
chr3.186841685.C.G_C	Adiponectin	rs266729	G	C	0.28
chr3.186853770.A.G_A	Adiponectin	rs3774261	A	A	0.41
chr4.23814039.C.T_C	PPARGC1A	rs8192678	T	C	0.32
chr4.87310240.T.TA_T	HSD17B13	rs72613567	TA	T	0.17
chr4.99574331.G.T_G	MTTP	rs1800591	T	G	0.20
chr7.22727026.C.G_C	IL-6	rs1800795	C	C	0.33
chr17.17506246.C.T_C	PEMT	rs7946	C	C	0.33
chr19.19268740.C.T_C	TM6SF2	rs58542926	T	C	0.11
chr19.39241143.A.G_A	IL28B	rs12980275	G	A	0.35
chr19.39248147.C.T_C	IFNL4/IL28B	rs12979860	T	C	0.36
chr22.43932850.G.A_G	PNPLA3	rs4823173	A	G	0.42
chr22.43936690.G.A_G	PNPLA3	rs2281135	A	G	0.42
chr22.43937814.T.G_T	PNPLA3	rs2896019	A	T	0.43
chr22.43958231.C.T_C	SAMM50	rs738491	G	C	0.42
chr22.43997195.C.T_C	PARVB	rs6006473	T	C	0.54
chr22.44002644.A.G_A	PARVB	rs5764455	A	A	0.47
chr22_43928847_C_G_C	PNPLA3	rs738409	G	C	0.54

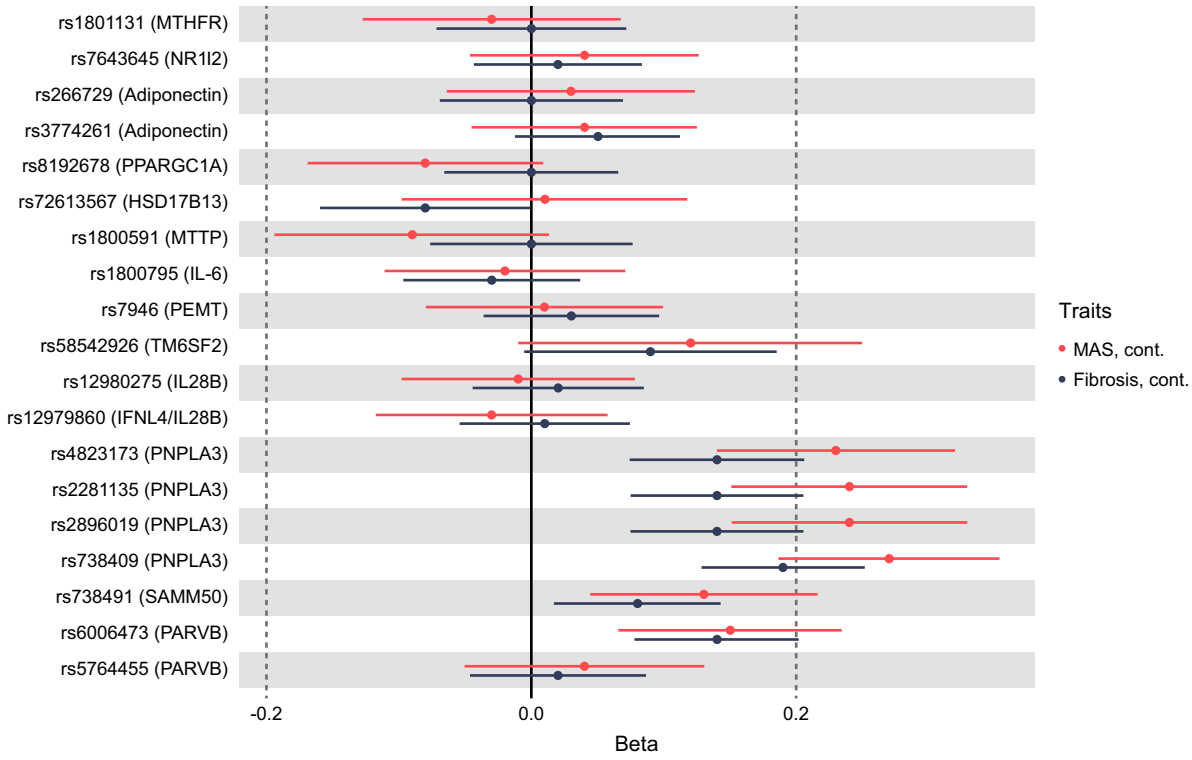

FIGURE 2 Association of genetic variants with NAFLD activity score and fibrosis severity as continuous variables, using multivariable linear regression adjusted for age, sex, ancestry, and 2 principal components of ancestry.

TABLE 3 Logistic regression for the association of tier 1 variants with advanced fibrosis and steatohepatitis (NAS ≥4)

		Advanced fibrosis		NAS ≥4	
Gene	rsID	OR (95% CI)	*p*	OR (95% CI)	*p*
MTHFR	rs1801131	1.05 (0.92–1.21)	4.72E−01	1 (0.88–1.13)	9.97E−01
NR1I2	rs7643645	0.98 (0.87–1.11)	7.60E−01	1.03 (0.92–1.15)	5.91E−01
Adiponectin	rs266729	0.98 (0.86–1.13)	7.88E−01	1.02 (0.91–1.16)	6.99E−01
Adiponectin	rs3774261	1.05 (0.93–1.19)	4.26E−01	1.04 (0.93–1.17)	4.46E−01
PPARGC1A	rs8192678	1.01 (0.89–1.15)	8.54E−01	0.93 (0.83–1.05)	2.36E−01
HSD17B13	rs72613567	0.85 (0.73–0.99)	3.97E−02	0.96 (0.83–1.1)	5.38E−01
MTTP	rs1800591	0.99 (0.85–1.14)	8.50E−01	0.88 (0.77–1)	5.54E−02
IL-6	rs1800795	0.96 (0.85–1.1)	5.88E−01	1.01 (0.9–1.13)	8.93E−01
PEMT	rs7946	1.09 (0.95–1.24)	2.21E−01	1.05 (0.93–1.18)	4.31E−01
TM6SF2	rs58542926	1.05 (0.88–1.27)	5.87E−01	1.22 (1.02–1.45)	2.67E−02
IL28B	rs12980275	1.05 (0.92–1.19)	4.88E−01	0.98 (0.87–1.09)	6.88E−01
IFNL4/IL28B	rs12979860	1.03 (0.91–1.17)	6.33E−01	0.96 (0.86–1.07)	4.77E−01
PNPLA3	rs4823173	1.2 (1.05–1.36)	6.90E−03	1.26 (1.12–1.42)	1.00E−04
PNPLA3	rs2281135	1.2 (1.06–1.37)	4.84E−03	1.3 (1.16–1.46)	1.19E−05
PNPLA3	rs2896019	1.2 (1.06–1.37)	5.06E−03	1.3 (1.15–1.46)	1.30E−05
SAMM50	rs738491	1.12 (0.99–1.26)	8.35E−02	1.13 (1.01–1.26)	3.07E−02
PARVB	rs6006473	1.32 (1.16–1.49)	1.18E−05	1.12 (1.01–1.25)	3.51E−02
PARVB	rs5764455	1.04 (0.91–1.19)	5.39E−01	1.03 (0.92–1.16)	6.25E−01
PNPLA3	rs738409	1.34 (1.18–1.51)	2.93E−06	1.33 (1.19–1.48)	4.14E−07

Abbreviation: NAS, NAFLD activity score.

findings were consistent when stratifying by adult versus pediatric subjects, and by EUR versus AMR subjects (Supplemental Table S5, http://links.lww.com/HC9/C6).

DISCUSSION

Recent advances in genetic and genomic technology have led to the identification of hundreds of disease-associated loci across various cardiometabolic conditions.[18,31] These genetic associations may help improve patient risk stratification or highlight underlying molecular drivers of disease progression. Several genetic polymorphisms have been previously described to associate with MASLD; these polymorphisms have been identified using a variety of phenotypes, including biomarkers of liver injury,[6,13] ICD-10 physician diagnosis codes,[12] intrahepatic triglyceride content,[9] or relatively small histologic cohorts.[15,16] These polymorphisms vary in terms of reported effect size and degree of replication in the literature. We first developed a novel framework to classify 87 previously described MASLD SNPs according to effect size and degree of replication; we prioritized 19 variants from 11 genetic loci for further

TABLE 4 Association of a "tier 1" variant genetic risk score, previously described 3 SNP genetic risk score, and *PNPLA3* rs738491 with continuous fibrosis and NAS

Phenotype	Genetic risk score	OR (95% CI)	*p*	Delta R^2
Fibrosis	Tier 1	1.02 (1.01–1.03)	1.19×10^{-05}	
Fibrosis	Tier 1 without PNPLA3	1.01 (0.99–1.03)	0.48	−5.70%
Fibrosis	Three SNP scores	1.15 (1.10–1.20)	6.23×10^{-10}	
Fibrosis	Three SNP scores without PNPLA3	1.09 (1.02–1.15)	8.71×10^{-03}	−9.40%
NAS	Tier 1	1.03 (1.02–1.05)	1.01×10^{-06}	
NAS	Tier 1 without PNPLA3	1.01 (0.99–1.04)	0.4	−44.90%
NAS	Three SNP scores	1.17 (1.10–1.24)	2.69×10^{-07}	
NAS	Three SNP scores without PNPLA3	1.04 (0.96–1.13)	0.3	−48.80%

Abbreviations: NAS, NAFLD activity score; SNP, single-nucleotide polymorphism.

evaluation. As the number of human biobank cohorts with associated genetic data continues to grow, our framework for prioritizing the evaluation of disease-associated variants may be useful in other disease states.

Our study evaluated whether previously identified MASLD polymorphisms were associated with the severity of histopathologic measures of disease in a large, well-characterized cohort. Relatively few studies have evaluated the impact of MASLD genetic variants on the histologic grade of steatohepatitis or stage of hepatic fibrosis in a disease-only setting. Histopathologic fibrosis severity and NAS are highly disease-relevant metrics; these measures are linked to hepatic decompensation events[20] and improvements in these histologic measures of disease are currently the only acceptable regulatory endpoint to support an accelerated approval for new MASH therapeutics.[4] We find strong evidence for the association of variants in the *PNPLA3* locus with liver fibrosis and NAS, especially *PNPLA3* rs738409. This is consistent with prior studies. Specifically, a prior GWAS in a European cohort of 1483 individuals with biopsy-proven MASLD confirmed *PNPLA3* as a risk factor for the full histological spectrum of NAFLD at genome-wide significance levels.[17] Our study across ~3000 participants confirms the relevance of *PNPLA3* genetic variation to disease severity in the largest histology-based cohort to date, with multicenter recruitment, with blinded and central histologic assessment, and in both adults and children. Our findings are also consistent with a recent study demonstrating that between 21% and 56% population attributable fraction for advanced fibrosis is attributable to the PNPLA3 rs738409 *G* allele,[21] and other studies demonstrating that this polymorphism is linked to major adverse liver outcomes in MASLD.[11,32] While our candidate gene score based on 19 “tier 1” variants showed significant association with fibrosis and NAS in a large cohort of patients with biopsy-proven MASLD, these results were driven by the *PNPLA3* rs738409 allele. Our study helps add to an emerging consensus that *PNPLA3* genotyping may improve prognostication and allow for prioritization of intensive intervention.

Surprisingly, our study failed to show a robust association between histologic disease severity and other previously described MASLD variants; other than the *PNPLA3* locus, only the *HSD17B13* rs72613567 polymorphism approached nominal significance for protection against fibrosis. There are several possible reasons for the discrepant findings. Our results suggest that apart from *PNPLA3*, previously identified MASLD variants may predispose to the development of fatty liver disease, but not influence the severity of histologic disease once diagnosed. It is possible that a distinct set of variants will emerge that influence solely the progression of steatohepatitis, fibrosis, or cirrhosis without promoting the initial occurrence of steatosis. Additionally, a histology-based analysis, while addressing the most clinically relevant phenotypic characteristic, reduces statistical power due to the limited number of cases in particular histologic categories. It is possible that increased phenotypic breadth of samples, including individuals with no steatosis or just steatosis, would be required to replicate previously identified variants. Finally, in contrast to other MASLD phenotypes, such as intrahepatic triglyceride content by imaging or measurement of liver biochemistries, the histologic scoring systems to evaluate steatohepatitis and fibrosis only provide nonlinear, semiquantitative assessments of disease and are subject to intra-observer and inter-observer variations. Although a clear diagnostic consensus among pathologists is not always feasible, our study attempted to address this by a centralized, blinded committee review.[4]

To our knowledge, this is the largest and most well-powered assessment of the association of histology-based MASLD severity and previously identified MASLD candidate genetic variants. Our results reinforce that variation at *PNPLA3* strongly influences the severity of fibrosis and steatohepatitis. Our results call into question whether other previously identified candidate MASLD variants are associated with the histology severity of the disease. Further work is required to evaluate whether polygenic risk scores beyond *PNPLA3* will assist with patient risk stratification and whether therapeutics that target the *PNPLA3* gene can decrease the histologic severity of MASLD.

AUTHOR CONTRIBUTIONS

All authors had full access to all the data and approved the final version of this manuscript. All authors take responsibility for the integrity of the data and the accuracy of the data analysis.

ACKNOWLEDGMENTS AND FINANCIAL SUPPORT

The authors thank the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) for its support of the Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) and this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank the NASH CRN investigators and the Ancillary Studies Committee for providing clinical samples and relevant data from the Nonalcoholic Fatty Liver Disease (NAFLD) Databases 1 and 2 (Adult: NCT01030484 and Pediatric: NCT01061684), PIVENS (NCT00063622), FLINT (NCT01265498), TONIC (NCT00063635), CyNCh (NCT01529268), and STOP-NAFLD (NCT03467217) trials. The authors also thank the participants involved in the NASH CRN studies. This study was conducted by the Ancillary Study Investigators. The biospecimens

from the NASH CRN reported here were supplied by the NIDDK Central Repository. This manuscript was not prepared in collaboration with the NIDDK Central Repository and does not necessarily reflect the opinions or views of the NIDDK Central Repository. The Regeneron Genomic Center is acknowledged for whole-exome sequencing and targeting genotyping through a collaborative agreement with Indiana University. This work was also supported in part by the NIDDK Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) grants (U01DK061730, U24DK061730, U01DK061737), the National Center for Advancing Translational Sciences, CTSI grants (UL1TR001881, UL1TR000006), the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. This work is also supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute. Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute (NHLBI) grant R01HL105756.

CONFLICTS OF INTEREST

Aaron Hakim consults for Deep Track Capital. Rohit Loomba serves as a consultant to Aardvark Therapeutics, Altimmune, Arrowhead Pharmaceuticals, AstraZeneca, Cascade Pharmaceuticals, Eli Lilly, Gilead, Glympse Bio, Inipharma, Intercept, Inventiva, Ionis, Janssen Inc., Lipidio, Madrigal, Neurobo, Novo Nordisk, Merck, Pfizer, Sagimet, 89 bio, Takeda, Terns Pharmaceuticals, and Viking Therapeutics. Rohit Loomba has stock options in Sagimet Biosciences. In addition, his institution received research grants from Arrowhead Pharmaceuticals, AstraZeneca, Boehringer-Ingelheim, Bristol-Myers Squibb, Eli Lilly, Galectin Therapeutics, Gilead, Intercept, Hanmi, Intercept, Inventiva, Ionis, Janssen, Madrigal Pharmaceuticals, Merck, Novo Nordisk, Pfizer, Sonic Incytes, and Terns Pharmaceuticals. Rohit Loomba is a co-founder of LipoNexus Inc. Jeffrey B. Schwimmer consults for Merck and receives grants from Intercept and Seraphina. Naga P. Chalasani has ongoing or recent (within 12 months) consulting agreements with Madrigal, Zydus, BioMea, GSK, Akira, Pfizer, Merck, Ipsen, and Altimmune. He receives research support from Exact Sciences. He has equity interest in Avant Sante, Inc. (a contract research organization) and Heligenics (a drug discovery start-up). These relationships are not significantly or directly related to this paper. Luca Lotta and Niek Verweij are employees of Regeneron Genetics Center. The remaining authors have no conflicts to report.

ORCID

Aaron Hakim https://orcid.org/0000–0002–9752–8826

REFERENCES

1. Moon AM, Singal AG, Tapper EB. Contemporary epidemiology of chronic liver disease and cirrhosis. Clin Gastroenterol Hepatol. 2020;18:2650–66.
2. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, Abdelmalek MF, Caldwell S, Barb D, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. Hepatology. 2023;77:1797–835.
3. Tacke F, Horn P, Wai-Sun Wong V, Ratziu V, Bugianesi E, Francque S, et al. EASL–EASD–EASO Clinical Practice Guidelines on the management of metabolic dysfunction–associated steatotic liver disease (MASLD). J Hepatol. 2024;81:492–542.
4. Anania FA, Dimick-Santos L, Mehta R, Toerner J, Beitz J. Nonalcoholic steatohepatitis: Current thinking from the Division of Hepatology and Nutrition at the Food and Drug Administration. Hepatology. 2021;73:2023–7.
5. Loomba R, Schork N, Chen C-H, Bettencourt R, Bhatt A, Ang B, et al. Heritability of hepatic fibrosis and steatosis based on a prospective twin study. Gastroenterology. 2015;149:1784–93.
6. Vujkovic M, Ramdas S, Lorenz KM, Guo X, Darlay R, Cordell HJ, et al. A multiancestry genome-wide association study of unexplained chronic ALT elevation as a proxy for nonalcoholic fatty liver disease with histological and radiological validation. Nat Genet. 2022;54:761–71.
7. Sookoian S, Rotman Y, Valenti L. Genetics of metabolic dysfunction-associated steatotic liver disease: The state of the art update. Clin Gastroenterol Hepatol. 2024;22:2177–187.e3.
8. 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.
9. Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet. 2008;40:1461–5.
10. Lindén D, Ahnmark A, Pingitore P, Ciociola E, Ahlstedt I, Andréasson AC, et al. PNPLA3 silencing with antisense oligonucleotides ameliorates nonalcoholic steatohepatitis and fibrosis in PNPLA3 I148M knock-in mice. Mol Metab. 2019;22:49–61.
11. Chen VL, Oliveri A, Miller MJ, Wijarnpreecha K, Du X, Chen Y, et al. PNPLA3 genotype and diabetes identify patients with nonalcoholic fatty liver disease at high risk of incident cirrhosis. Gastroenterology. 2023;164:966–977.e17.
12. Fairfield CJ, Drake TM, Pius R, Bretherick AD, Campbell A, Clark DW, et al. Genome-Wide Association Study of NAFLD using electronic health records. Hepatol Commun. 2022;6:297–308.
13. Hakim A, Moll M, Brancale J, Liu J, Lasky-Su JA, Silverman EK, et al. Genetic variation in the mitochondrial glycerol-3-phosphate acyltransferase is associated with liver injury. Hepatology. 2021;74:3394–408.
14. Jamialahmadi O, Mancina RM, Ciociola E, Tavaglione F, Luukkonen PK, Baselli G, et al. Exome-Wide Association Study on alanine aminotransferase identifies sequence variants in the GPAM and APOE associated with fatty liver disease. Gastroenterology. 2021;160:1634–646.e7.
15. Chalasani N, Guo X, Loomba R, Goodarzi MO, Haritunians T, Kwon S, et al. Genome-wide association study identifies variants associated with histologic features of nonalcoholic Fatty liver disease. Gastroenterology. 2010;139:1567–76, 1576.e1–6.
16. Speliotes EK, Butler JL, Palmer CD, Voight BF, Hirschhorn JN. PNPLA3 variants specifically confer increased risk for histologic nonalcoholic fatty liver disease but not metabolic disease. Hepatology. 2010;52:904–12.
17. Anstee QM, Darlay R, Cockell S, Meroni M, Govaere O, Tiniakos D, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort (☆). J Hepatol. 2020;73:505–15.

18. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. Nat Genet. 2022;54:1803–15.
19. Suzuki K, Hatzikotoulas K, Southam L, Taylor HJ, Yin X, Lorenz KM, et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. Nature. 2024;627:347–57.
20. Sanyal AJ, Van Natta ML, Clark J, Neuschwander-Tetri BA, Diehl A, Dasarathy S, et al. Prospective study of outcomes in adults with nonalcoholic fatty liver disease. N Engl J Med. 2021;385:1559–69.
21. Jarasvaraparn C, Vilar-Gomez E, Yates KP, Wilson LA, Neuschwander-Tetri B, Loomba R, et al. Age, BMI, and type 2 diabetes modify the relationship between PNPLA3 and advanced fibrosis in children and adults with NAFLD. Clin Gastroenterol Hepatol. 2024;22:1024–36.e2.
22. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017;45:D896–901.
23. Paternostro R, Staufer K, Traussnigg S, Stättermayer AF, Halilbasic E, Keritam O, et al. Combined effects of PNPLA3, TM6SF2 and HSD17B13 variants on severity of biopsy-proven non-alcoholic fatty liver disease. Hepatol Int. 2021;15:922–33.
24. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology. 2005;41:1313–21.
25. Schwantes-An T-H, Darlay R, Mathurin P, Masson S, Liangpunsakul S, Mueller S, et al. Genome-wide association study and meta-analysis on alcohol-associated liver cirrhosis identifies genetic risk factors. Hepatology. 2021;73:1920–31.
26. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28:3326–8.
27. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.
28. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48:1284–7.
29. 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.
30. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2021. https://www.R-project.org/
31. Ingelsson E, McCarthy MI. Human genetics of obesity and type 2 diabetes mellitus: Past, present, and future. Circ Genom Precis Med. 2018;11:e002090.
32. Stender S, Loomba R. PNPLA3 genotype and risk of liver and all-cause mortality. Hepatology. 2020;71:777–9.

How to cite this article: Hakim A, Lin K, Schwantes-An T, Abreu M, Tan J, Guo X, et al. A comprehensive evaluation of candidate genetic polymorphisms in a large histologically characterized MASLD cohort using a novel framework. Hepatol Commun. 2025;9:e0728. https://doi.org/10.1097/HC9.0000000000000728