Title

Genome-wide association study identifies high-impact susceptibility loci for HCC in North America.

Journal

Hepatology, 80(1)

Authors

Hassan, Manal
Li, Donghui
Han, Younghun
et al.

Publication Date

2024-07-01

DOI

10.1097/HEP.0000000000000800

Peer reviewed

Received: 20 September 2023 | Accepted: 18 December 2023

DOI: 10.1097/HEP.0000000000000800

ORIGINAL ARTICLE

OPEN

Genome-wide association study identifies high-impact susceptibility loci for HCC in North America

Manal M. Hassan[1] | Donghui Li[2] | Younghun Han[3] | Jinyoung Byun[3] | Rikita I. Hatia[1] | Erping Long[4] | Jiyeon Choi[4] | Robin Kate Kelley[5] | Sean P. Cleary[6] | Anna S. Lok[7] | Paige Bracci[8] | Jennifer B. Permuth[9,10] | Roxana Bucur[11] | Jian-Min Yuan[12,13] | Amit G. Singal[14] | Prasun K. Jalal[15] | R. Mark Ghobrial[16] | Regina M. Santella[17] | Yuko Kono[18] | Dimpy P. Shah[19] | Mindie H. Nguyen[20] | Geoffrey Liu[21] | Neehar D. Parikh[7] | Richard Kim[9] | Hui-Chen Wu[17] | Hashem El-Serag[22] | Ping Chang[2] | Yanan Li[2] | Yun Shin Chun[23] | Sunyoung S. Lee[2] | Jian Gu[1] | Ernest Hawk[24] | Ryan Sun[25] | Chad Huff[1] | Asif Rashid[26] | Hesham M. Amin[27] | Laura Beretta[28] | Robert A. Wolff[2] | Samuel O. Antwi[29] | Yehuda Patt[30] | Lu-Yu Hwang[31] | Alison P. Klein[32] | Karen Zhang[5] | Mikayla A. Schmidt[33] | Donna L. White[34] | John A. Goss[35] | Saira A. Khaderi[36] | Jorge A. Marrero[14] | Francisco G. Cigarroa[37] | Pankil K. Shah[19] | Ahmed O. Kaseb[2] | Lewis R. Roberts[33] | Christopher I. Amos[3]

[1]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[2]Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[3]Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, Texas, USA

[4]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

[5]Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California, USA

[6]Division of Hepatobiliary and Pancreas Surgery, Mayo Clinic, Rochester, Minnesota, USA

[7]Division of Gastroenterology and Hepatology, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA

[8]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA

[9]Department of Gastrointestinal Oncology, Moffitt Cancer Center, Tampa, Florida, USA

[10]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, Florida, USA

[11]Princess Margaret Cancer Center and Toronto General Hospital, University Health Network, Toronto, Ontario, Canada

[12]Cancer Epidemiology and Prevention Program, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

[13]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

[14]Division of Digestive and Liver Diseases, The University of Texas Southwestern Medical Center, Dallas, Texas, USA

[15]Department of Gastroenterology and Hepatology, Baylor College of Medicine, Houston, Texas, USA

[16]J.C. Walter Jr. Transplant Center, Houston Methodist Hospital, Houston, Texas, USA

[17]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York City, New York, USA

[18]Division of Gastroenterology and Hepatology, University of California San Diego, San Diego, California, USA

[19]Mays Cancer Center, The University of Texas Health Science Center San Antonio MD Anderson, San Antonio, Texas, USA

[20]Division of Gastroenterology and Hepatology, Department of Epidemiology and Population Health, Stanford University Medical Center, Palo Alto, California, USA

[21]Medical Oncology and Hematology, Princess Margaret Cancer Centre, University of Toronto, Toronto, ON, Canada

[22]Department of Medicine, Baylor College of Medicine, Houston, Texas, USA

[23]Division of Surgery, Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[24]Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[25]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[26]Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[27]Department of Hematopathology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[28]Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[29]Division of Epidemiology, Department of Quantitative Health Sciences, Mayo Clinic, Jacksonville, Florida, USA

[30]Division of Hematology/Oncology, Department of Internal Medicine, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA

[31]Department of Epidemiology, Human Genetics, and Environment Science, The University of Texas Health Science Center at Houston, Houston, Texas, USA

[32]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland, USA

[33]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota, USA

[34]Sections of Gastroenterology and Hepatology and Health Services Research, Baylor College of Medicine, Houston, Texas, USA

[35]Division of Abdominal Transplantation, Michael E. DeBakey School of Medicine, Baylor College of Medicine, Houston, Texas, USA

[36]Division of Abdominal Transplantation, Baylor College of Medicine, Houston, Texas, USA

[37]Transplant Center, The University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA

Correspondence
Manal M. Hassan, Division of Cancer Prevention and Population Sciences, Department of Epidemiology, 1155 Pressler Street/ Unit 1340, Houston, TX 77030 USA.
Email: mhassan@mdanderson.org

Christopher I. Amos, Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX 77030 USA.
Email: Chris.Amos@bcm.edu

Abstract

Background and Aims: Despite the substantial impact of environmental factors, individuals with a family history of liver cancer have an increased risk for HCC. However, genetic factors have not been studied systematically by genome-wide approaches in large numbers of individuals from European descent populations (EDP).

Approach and Results: We conducted a 2-stage genome-wide association study (GWAS) on HCC not affected by HBV infections. A total of 1872 HCC cases and 2907 controls were included in the discovery stage, and 1200 HCC cases and 1832 controls in the validation. We analyzed the discovery and validation samples separately and then conducted a meta-analysis. All analyses were conducted in the presence and absence of HCV. The liability-scale heritability was 24.4% for overall HCC. Five regions with significant ORs (95% CI) were identified for non-viral HCC: 3p22.1, *MOBP*, rs9842969, (0.51, [0.40–0.65]); 5p15.33, *TERT*, rs2242652, (0.70, (0.62–0.79]); 19q13.11, *TM6SF2*, rs58542926, (1.49, [1.29–1.72]); 19p13.11 *MAU2*, rs58489806, (1.53, (1.33–1.75]); and 22q13.31, *PNPLA3,* rs738409, (1.66, [1.51–1.83]). One region was identified for HCV-induced HCC: 6p21.31, human leukocyte antigen DQ beta 1, rs9275224, (0.79, [0.74–0.84]). A combination of homozygous variants of *PNPLA3* and *TERT* showing a 6.5-fold higher risk for nonviral-related HCC compared to individuals

Abbreviations: CART, classification and regression tree; EDP, European-descent populations; EQTL, expression quantitative trait loci; GATAD2A, GATA zinc finger domain containing 2A; GWAS, genome-wide association study; HLA-DQB1, human leukocyte antigen DQ beta 1; MOBP, myelin-associated oligodendrocyte basic protein; PARVB, parvin-beta; PNPLA3, patatin-like phospholipase domain-containing 3; SAMM50, sorting and assembly machinery component 50 homolog; sQTL, splicing quantitative trait loci; SUGP1, SURP and G-patch domain containing 1; TERT, telomerase reverse transcriptase.

Manal M. Hassan, Donghui Li, Lewis R. Roberts, and Christopher I. Amos contributed equally.

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.

lacking these genotypes. This observation suggests that gene-gene interactions may identify individuals at elevated risk for developing HCC.

Conclusions: Our GWAS highlights novel genetic susceptibility of nonviral HCC among European descent populations from North America with substantial heritability. Selected genetic influences were observed for HCV-positive HCC. Our findings indicate the importance of genetic susceptibility to HCC development.

INTRODUCTION

HCC is a rising cause of cancer-related mortality.[1] HCC development is influenced by multiple environmental factors, including HBV, HCV, alcohol consumption, and metabolic syndrome.[1] Despite the multifactorial etiology of HCC, metabolic dysfunction–associated fatty liver disease, including its pathological feature of metabolic dysfunction–associated steatohepatitis, is the leading cause of HCC in Western countries.[2] The variation in the prevalence of HCC risk factors among different populations parallels the disease incidence by geographic location, gender, and ethnicity.[3]

Although the magnitude of environmental factors for HCC etiology is high, published studies,[4,5] including ours,[6] showed a positive association between the history of cancer among first-degree relatives and the risk for HCC development. Consequently, the risk of HCC may be influenced by multiple genes in multiple pathways. Thus, the use of a genome-wide association study (GWAS) with comprehensive coverage of the common variants of the entire genome is a more informative approach to examining genetic susceptibility in HCC in the presence and absence of environmental risk factors. GWASs have already identified important genetic variants that confer susceptibility to various chronic liver diseases, including metabolic dysfunction–associated fatty liver disease,[7] viral hepatitis,[8,9] and cirrhosis,[10] and have opened the doors for novel therapeutic targets and algorithms by revealing unexpected functional genotypes associated with treatment response.[11,12]

Moreover, GWASs have been conducted in many common human diseases or traits with substantial impact on personalized medicine and disease prevention.[13] As a result, hundreds of risk alleles have been identified, previously unknown cancer associations have been uncovered, and polygenic risk scores have been applied in risk prediction models.[14] However, for rare cancers, such as HCC, GWASs have been hammered by lack of power due to small sample sizes. Most of the HCC GWASs were conducted in countries with high HCC incidence, including Asian populations, and focused on either HBV-HCC[15–17] or HCV-HCC.[18–22] Two recent large genomic studies were restricted to alcohol-associated HCC among European populations.[23,24] The only GWAS that has been reported in the US white population included 436 HCC cases and 1065 controls.[25] Larger studies with adequate power are needed to reveal genetic factors that contribute to HCC in the US population with consideration of major risk factors.

To identify genetic susceptibility to HCC in a North American population, we performed a 2-stage GWAS in European descent populations (EDP) from the United States and Canada according to the presence or absence of HCV infection. Because HBV is uncommon in North America,[26] we restricted our GWAS to HBV noncarriers.

METHODS

Study design and participants

The current study's 2-phase design (discovery and validation phases) focused on the genetic susceptibility of HCC in individuals of EDP from North America. Cases were defined as patients with pathological or radiological evidence of HCC without serological evidence of chronic HBV. Case patients were recruited from 14 North American Institutions (Supplemental Tables S1A-E, http://links.lww.com/HEP/I302) that are members of the Hepatocellular Carcinoma Epidemiology Consortium.

We selected 2 groups of controls without evidence of cancer at the time of recruitment. The first group involved healthy individuals without a history of chronic liver diseases, and the second included patients with chronic HCV infection (HCV carriers) (Supplemental Table S1A, http://links.lww.com/HEP/I302). We also included 1960 healthy controls with completed genetic array data from a sister GWAS of the Pancreatic Cancer Case Control Consortium (PanC4)[11] that used the same genotyping platform (Supplemental Methods, http://links.lww.com/HEP/I303). For the discovery phase, a total of 1872 eligible HCC cases, including 738 HCV-positive and 1134 HCV-negative (nonviral HCC), were compared with 2907 non-HCC control subjects, 588 of whom were known to be HCV-carriers (Supplemental Table S2A, http://links.lww.com/HEP/I302).

For the validation phase, genotyping was conducted on 1200 HCC cases (715 HCV-negative and 485

HCV-positive) and 509 control subjects (151 HCV-carrier controls and 358 healthy controls) (Supplemental S2B, http://links.lww.com/HEP/I302). Control data also included data sets from 2 previously conducted GWASs in renal cell carcinoma.[27,28] Because only about 1% of the general adult population tests positive for HCV antibodies,[26] these controls were considered HCV-negative; however, we conducted sensitivity analysis restricted to controls with known HCV-carrier status (Supplemental Table S1A, http://links.lww.com/HEP/I302, Supplemental Figure S1, http://links.lww.com/HEP/I302). Characteristics of the validation phase participants are described in Supplemental Table S2B, http://links.lww.com/HEP/I302.

All participants were self-reported non-Hispanic whites. Each participating site obtained informed written consent from study participants and approval from their Institutional Review Board (IRB) for this study and obtained IRB certification permitting data sharing in accordance with the Policy for Sharing of Data Obtained in National Institutes of Health Supported or Conducted GWAS. The study relies on existing DNA samples collected between 2000 and 2015. Supplemental Figure S1, http://links.lww.com/HEP/I302, summarizes the study design.

Genotyping methods and quality control

Genotyping for the discovery phase was performed at the Center for Inherited Disease Research using the Illumina Infinium OmniExpressExome-8v1-3 platform, with 958,497 single-nucleotide polymorphism (SNP). In all, 3411 DNA samples, including 63 pairs of duplicate samples and 64 HapMap samples, were genotyped. Genotype calling used genome build GRCh37/hg19 and GenomeStudio version 2011.1, Genotyping Module version 1.9.4, and GenTrain Version 1.0. In all, 3366 DNA samples were genotyped and passed the Center for Inherited Disease Research lab's quality control (QC) process (Supplemental Table S3, http://links.lww.com/HEP/I302, Supplemental Figure S2, http://links.lww.com/HEP/I302). Additional QC procedures, data harmonization, and data availability are included in the Supplemental Methods, http://links.lww.com/HEP/I303, and data are available through dbGAP (phs001744.v1.p1).

We conducted ancestry analysis[29] to retain those individuals whose genetic ancestry showed more than 80% alignment with ancestral EDP (Supplemental Figure S3, http://links.lww.com/HEP/I302). We first selected the 652,761 common SNPs retained after QC steps for both HCC GWAS and PanC4. Then, we performed an imputation analysis in a 2-stage procedure using EAGLE (v2.4) to phase genotypes and Minimax 4 to perform imputation by using the Haplotype Reference Consortium (version r1.1 2016)[30] panel of European ancestry on the Michigan Imputation Server. We restricted imputation analysis to SNPs with minor allele frequencies $>$ 0.001. We also performed the same imputation analysis on an additional set of genome controls from a previously conducted renal cell carcinoma GWAS[27] for the validation phase. After imputation, we excluded SNPs with minor allele frequencies of 0.001 or that had an imputation quality score of < 0.5.

Statistical analysis

After imputation to the Haplotype Reference Consortium and following QC, we performed a comprehensive association analysis of 12,651,372 variants on autosomal chromosomes 1-22 and the X chromosome in the discovery phase using logistic regression adjusting for age, gender, and the first 10 eigenvectors from principal component analysis with SNPTEST (v2.5.4)[29,31] and R (v3.6.2). We analyzed the discovery and validation samples separately and then conducted a meta-analysis in a fixed-effect model using METASOFT.[32] We also conducted a stratified analysis according to alcohol use, cigarette smoking status, presence of type 2 diabetes mellitus, and HCV status with R (v3.6.2).

We implemented conditional analyses to identify secondary association signals conditioning on the top-associated lead SNP across the whole genome followed by a stepwise procedure of selecting additional SNPs using SNPTEST (v2.5.4). To account for multiple testing in this GWAS, a fixed p-value threshold of 5×10^{-8} was used to identify the association between a common genetic variant and HCC.

To estimate genome-wide SNP-heritability defined as the proportion of phenotypic variance explained by all SNPs, we computed regional heritability on the liability scale, including the 5 regions that were found to contain genome-wide significant results using genome-wide complex trait analysis with genome-based restricted maximum likelihood method (genome-wide complex trait analysis v1.94.1).[33] We applied 1% for the lifetime risk of HCC and 0.5% for HCV-positive and HCV-negative HCC[1] to compute the regional heritability on the liability scale for a case-control study. The population attributable fraction was estimated and described in the Supplemental Methods, http://links.lww.com/HEP/I303.

Classification trees

The potential role of combinations of alleles in the risk of HCC was evaluated by classification and regression tree (CART) method. Data from the genome-wide significant lead risk–associated variants were subjected to analyses in which a classification tree was built for overall HCC and the HCV-negative HCC. The rpart package in R (v3.6.2) with the Gini index measure was used to identify optimal splits of the data, with the complexity parameter set to 0.001, and the data then

pruned to include only those nodes containing at least 20 observations. ORs in the parsimonious classification tree in the discovery phase were calculated with the following formula[34]: $OR = \frac{\#\ cases\ in\ node\ \ i\ /\ \#\ controls\ in\ node\ i}{\#\ cases\ in\ node\ \ 1\ /\ \#\ controls\ in\ node\ \ 1}$. The Woolf approximation was used to compute SEs and CIs.[35]

Having obtained an optimal classification tree in the discovery data, we further evaluated the fit of this model in the validation data by constructing a logistic regression model containing the same terminal nodes by specifying dummy variables that captured the same structure as the classification tree. The genetically informed model included baseline predictors as well as predictions from the genetic factors in nodes defined by the classification tree. This constructed logistic regression model maintained the same types of interactions as seen in the classification tree but allowed effects from the interactions to be modeled by using all of the data, not just individuals who were subdivided by the tree; this approach allowed us to validate the model in a new data set without rerunning classification tree analysis, which typically yields a new tree.

Colocalization analysis

To identify candidate susceptibility genes from each GWAS locus that contribute to HCC risk, we performed colocalization analyses using expression quantitative trait loci (eQTL) and splicing QTL (sQTL) data of 6 liver cancer-relevant tissue types from the Genotype-Tissue Expression v8, which includes eQTL and sQTL data from normal tissues from 838 donors mainly of EDP. For colocalization analyses, we selected 6 liver cancer-relevant tissues based on the prior knowledge: whole blood (n = 670), liver (n = 208), Epstein-Barr virus-transformed lymphocytes (n = 147), cultured fibroblast cells (n = 483), adipose subcutaneous (n = 581), and adipose visceral omentum (n = 469). Genotype-Tissue Expression eQTL and sQTL association data for variants within +/− 100kb windows of the lead variants present in the GWAS summary statistics were extracted. For the locus overlapping the major histocompatibility complex regions (HCV-positive locus marked by rs9275224), a narrower window of +/− 10kb of the lead variants was used. Colocalization of the GWAS and eQTL or sQTL signals were calculated using coloc.[36] We reported the colocalization results if coloc suggested plausible posterior probability was greater than 0.8 ($PPH_4 > 0.8$).

Integrative multi-omics annotation analysis

We annotated the top 10 lead variants associated with HCC using the Functional Annotation of Variants -Online Resource (FAVOR v2.0; https://favor.genohub.org/) platform, which provides functional annotation information of all possible 8,812,917,339 single nucleotide variants across the human genome and observed insertion and deletions (79,997,898 insertions and deletions) from the Trans-Omics for Precision Medicine BRAVO variant set (Build GRCh38) Further details about FAVOR[37] are presented in the Supplemental Methods, http://links.lww.com/HEP/I303.

RESULTS

We performed univariable analyses of discovery samples comprising 1872 HCC cases (738 HCV-positive HCC; 1134 nonviral HCC) and 2907 non-HCC controls (649 HCV-carrier controls and 2319 healthy controls). The summary statistics with $p < 10^{-5}$ are presented in Supplemental Tables S4A-C, http://links.lww.com/HEP/I302, for overall HCC, HCV-positive, and HCV-negative HCC strata. Table 1 presents the lead SNPs associated with overall HCC susceptibility at the genome-wide significance level of $p < 5 \times 10^{-8}$ in the discovery and validation phases and the meta-analysis. Analysis of the discovery phase population identified loci on chromosomes 3, 5, 19, and 22 that were associated with HCC development (Figure 1A). HCV-negative and HCV-positive HCC GWAS identified 3 and 1 associations, respectively (Table 2 and Figure 1B-C). Test statistic inflation (λ) was estimated to be 1.021 for all subjects, 1.022 for the HCV-positive group, and 1.021 for the HCV-negative group, indicating no inflation of test statistics (Figure 1A-C).

The strongest association was found on chromosome 22 in the same LD block containing the patatin-like phospholipase domain-containing 3 (*PNPLA3*) gene, the sorting and assembly machinery component 50 homolog (*SAMM50*) gene, and parvin-beta (*PARVB*) gene (Figure 2A). In all, 61 variants on/near *PNPLA3*, 29 variants on/near *SAMM50*, and 5 variants on/near *PARVB* had *p* values of $< 1 \times 10^{-9}$ in this locus. The top hit was rs738409 in the *PNPLA3* gene (Supplemental Table S4A, http://links.lww.com/HEP/I302). Our meta-analysis showed a 1.65-fold increased risk of HCC among *G* allele carriers (Table 1). Conditional analysis showed that the variants encompassing *SAMM50* and *PARVB* were no longer significant after adjusting for the effect of *PNPLA3* (Supplemental Table S5, http://links.lww.com/HEP/I302).

The second significant locus for HCC was found within an LD block on chromosome 19 containing transmembrane 6 superfamily member 2 (*TM6SF2*), SURP and G-patch domain containing 1 (*SUGP1*), the sister chromatid cohesion factor (*MAU2*), and GATA zinc finger domain containing 2A (*GATAD2A*) genes (Figure 2B). Two SNPs of *MAU2*, 3 SNPs of *SUGP1*, 1 SNP each of the *TM6SF2* and *GATAD2A* genes, and an

TABLE 1 Top significant and validated SNPs associated with overall HCC development

SNP identifier	Allele NEA[b]/EA[c]	Gene name	Chromosome position	GWAS analytic-phase	EAF[a]		OR (95% CI)[d]	*p*
					Cases	Controls		
rs9842969	T/C	*MOBP*	3p22.1	Discovery	0.924	0.946	0.51 (0.40–0.65)	2.70E-08
	—	—	39600177	Validation	0.956	0.969	0.67 (0.51–0.88)	4.26E-03
	—	—	—	Meta-analysis	0.937	0.955	0.57 (0.48–0.69)	1.36E-09
rs2242652	G/A	*TERT*	5p15.33	Discovery	0.150	0.199	0.70 (0.62–0.79)	1.55E-09
	—	—	1280028	Validation	0.147	0.197	0.70 (0.61–0.81)	6.32E-07
	—	—	—	Meta-analysis	0.148	0.198	0.70 (0.64–0.77)	8.27E-15
rs9275224	A/G	*HLA-DQB1*	6p21.31	Discovery	0.464	0.488	0.90 (0.83–0.98)	1.80E-02
	—	—	32659878	Validation	0.345	0.460	0.66 (0.59–0.73)	2.81E-16
	—	—	—	Meta-analysis	0.417	0.477	0.79 (0.74–0.84)	1.45E-12
rs58542926	C/T	*TM6SF2*	19p13.11	Discovery	0.114	0.075	1.49 (1.29–1.72)	4.02E-08
	—	—	19379549	Validation	0.111	0.069	1.59 (1.33–1.90)	4.29E-07
	—	—	—	Meta-analysis	0.113	0.073	1.53 (1.36–1.70)	1.03E-13
rs10401969	T/C	*SUGP1*	19p13.11	Discovery	0.129	0.083	1.53 (1.33–1.75)	1.16E-09
	—	—	19407718	Validation	0.122	0.083	1.47 (1.24–1.73)	7.73E-06
	—	—	—	Meta-analysis	0.126	0.083	1.50 (1.35–1.67)	4.87E-14
rs58489806	C/T	*MAU2*	19p13.11	Discovery	0.129	0.083	1.53 (1.33–1.75)	1.16E-09
	—	—	19456917	Validation	0.122	0.083	1.47 (1.24–1.73)	7.73E-06
	—	—	—	Meta-analysis	0.126	0.083	1.50 (1.35–1.67)	4.87E-14
rs3794991	C/T	*GATAD2A*	19p13.11	Discovery	0.123	0.083	1.46 (1.27–1.67)	7.72E-08
	—	—	19610596	Validation	0.128	0.083	1.48 (1.26–1.73)	1.05E-06
	—	—	—	Meta-analysis	0.125	0.083	1.47 (1.32–1.62)	3.99E-13
rs738409	C/G	*PNPLA3*	22q13.31	Discovery	0.337	0.225	1.66 (1.51–1.83)	1.07E-25
	—	—	44324727	Validation	0.341	0.236	1.62 (1.45–1.82)	5.38E-17
	—	—	—	Meta-analysis	0.338	0.229	1.65 (1.53–1.77)	1.22E-40
rs2294922	G/C	*SAMM50*	22q13.31	Discovery	0.320	0.221	1.57 (1.43–1.73)	5.72E-21
	—	—	44379565	Validation	0.308	0.220	1.51 (1.35–1.69)	1.53E-12
	—	—	—	Meta-analysis	0.315	0.220	1.55 (1.44–1.66)	1.21E-31
rs2401514	T/A	*PARVB*	22q13.31	Discovery	0.252	0.185	1.45 (1.31–1.60)	1.57E-12
	—	—	44394019	Validation	0.242	0.185	1.39 (1.22–1.57)	4.05E-07
	—	—	—	Meta-analysis	0.248	0.185	1.42 (1.31–1.54)	4.13E-18

[a]Effect allele frequency.
[b]Noneffect allele.
[c]Effect allele (tested allele).
[d]OR (95% CI).
Abbreviations: GATAD2, GATA zinc finger domain containing 2A; GWAS, GWAS, genome-wide association study; HLA−DQB1, human leukocyte antigen DQ beta 1; MAU2, sister chromatid cohesion factor; MOBP, myelin-associated oligodendrocyte basic protein; PARVB, parvin-beta; PNPLA3, patatin-like phospholipase domain-containing 3; SAMM50, sorting and assembly machinery component 50 homolog; SNP, single-nucleotide polymorphism; SUGP1, SURP and G-patch domain containing 1; TERT, telomerase reverse transcriptase; TM6SF2, transmembrane 6 superfamily member 2.

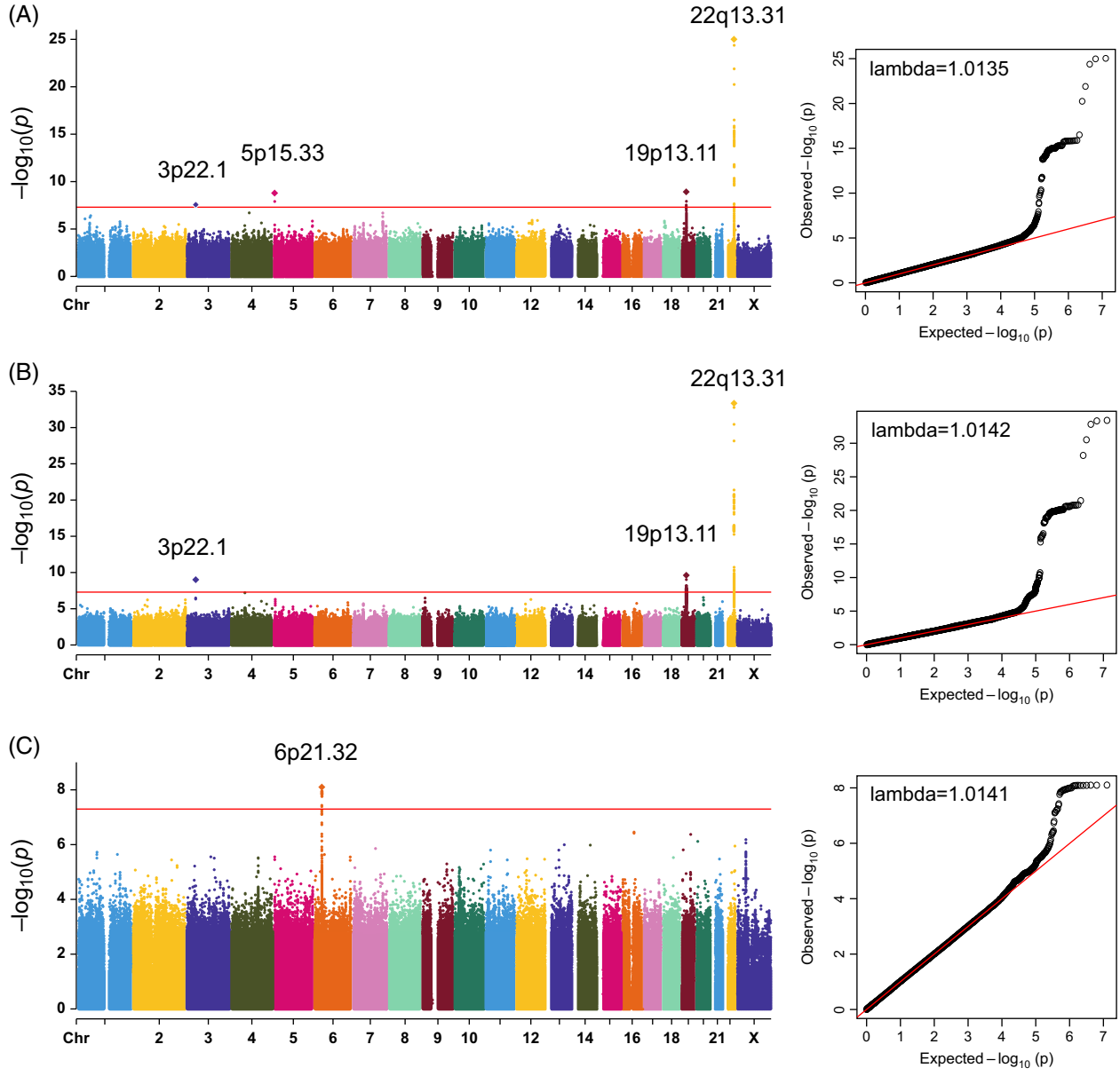

FIGURE 1 Manhattan plots and quantile-quantile (Q-Q) plots for the genome-wide association study results of HCC. (A) All populations; (B) non-HCV–induced HCC; (C) HCV-induced HCC. X-axis represents SNP positions across the entire genome by chromosome, and the y-axis is the negative logarithm *p*-value, -log10 (p). The red horizontal line denotes $p = 5 \times 10{-}8$. For Q-Q plots, the x-axis represents Expected -log10 (p), and the y-axis is Observed -log10 (p) of each SNP. Abbreviation: SNP, single-nucleotide polymorphism.

intergenic variant rs58489806 showed genome-wide significant associations with HCC risk. The most significant SNP in the meta-analysis overall was rs10401969 in *SUGP1,* but variants rs58489806 in *MAU2* and rs58542926 in *TM6SF2* (Table 1) had similarly strong levels of support. In analyses restricted to HCV-negative cases, SNP rs58542926 in *TM6SF2* showed the most significant result in this region ($p = 3.8 \times 10^{-20}$). Analyses conditioning on rs58542926 in *TM6SF2* (Supplemental Table S5, http://links.lww.com/HEP/I302) revealed that the rs58489806 variant in *MAU2* remained nominally significant in the overall discovery and meta-analysis data sets but not in the absence of HCV, and other SNPs in this region were not significant after conditioning.

Next, we identified a novel locus 3p22.1 (rs9842969) of the myelin-associated oligodendrocyte basic protein (*MOBP*) on chromosome 3 (Figure 2C) and 3 SNPs (rs2242652, rs10069690, and rs72709458) of the telomerase reverse transcriptase (*TERT*) gene on chromosome 5 (Figure 2D) that were significantly associated with HCC. Genomic regional association plots for the HCV-stratified analysis are shown in Figure 2E-F. Results from these SNPs replicated very well with consistent ORs in the discovery and validation

TABLE 2 Top significant and validated SNPs associated with HCV-negative and HCV-positive HCC

SNP identifier	Allele NEA[b]/EA[c]	Gene name	Chromosome position	GWAS analytic-phase	EAF[a]		OR (95% CI)[d]	*p*
					Cases	Controls		
HCV-negative HCC								
rs9842969	T/C	*MOBP*	3p22.1	Discovery	0.921	0.951	0.42 (0.32–0.55)	9.45E-10
	—	—	39600177	Validation	0.954	0.969	0.63 (0.46–0.88)	6.09E-03
	—	—	—	Meta-analysis	0.934	0.959	0.50 (0.40–0.62)	1.19E-10
rs2242652	G/A	*TERT*	5p15.33	Discovery	0.150	0.202	0.69 (0.60–0.80)	4.51E-07
	—	—	1280028	Validation	0.139	0.199	0.65 (0.54–0.77)	9.21E-07
	—	—	—	Meta-analysis	0.146	0.201	0.67 (0.60–0.75)	3.60E-12
rs58542926	C/T	*TM6SF2*	19p13.11	Discovery	0.134	0.076	1.69 (1.44–1.99)	3.79E-10
	—	—	19379549	Validation	0.139	0.069	2.01 (1.65–2.46)	7.55E-12
	—	—	—	Meta-analysis	0.136	0.073	1.81 (1.60–2.06)	3.83E-20
rs10401969	T/C	*SUGP1*	19p13.11	Discovery	0.136	0.077	1.70 (1.44–2.00)	2.37E-10
	—	—	19407718	Validation	0.141	0.074	1.83 (1.52–2.21)	3.09E-10
	—	—	—	Meta-analysis	0.138	0.076	1.75 (1.55–1.98)	4.67E-19
rs58489806	C/T	*MAU2*	19p13.11	Discovery	0.146	0.084	1.67 (1.43–1.96)	2.30E-10
	—	—	19456917	Validation	0.148	0.081	1.82 (1.51–2.20)	4.86E-10
	—	—	—	Meta-analysis	0.147	0.083	1.73 (1.54–1.96)	7.29E-19
rs3794991	C/T	*GATAD2A*	19p13.11	Discovery	0.140	0.085	1.61 (1.37–1.88)	6.14E-09
	—	—	19610596	Validation	0.150	0.080	1.79 (1.50–2.15)	2.42E-10
	—	—	—	Meta-analysis	0.144	0.083	1.69 (1.50–1.90)	1.11E-17
rs738409	C/G	*PNPLA3*	22q13.31	Discovery	0.375	0.217	2.03 (1.81–2.28)	5.39E-34
	—	—	44324727	Validation	0.407	0.233	2.12 (1.86–2.43)	1.72E-28
	—	—	—	Meta-analysis	0.387	0.224	2.07 (1.90–2.26)	8.17E-60
rs2294922	G/C	*SAMM50*	22q13.31	Discovery	0.357	0.213	1.91 (1.70–2.14)	7.13E-29
	—	—	44379565	Validation	0.363	0.216	1.92 (1.67–2.19)	2.62E-21
	—	—	—	Meta-analysis	0.359	0.214	1.91 (1.75–2.08)	6.10E-48
rs2401514	T/A	*PARVB*	22q13.31	Discovery	0.276	0.177	1.68 (1.49–1.90)	9.33E-17
	—	—	44394019	Validation	0.276	0.185	1.65 (1.42–1.91)	3.51E-11
	—	—	—	Meta-analysis	0.276	0.181	1.67 (1.52–1.83)	2.34E-26
HCV-positive HCC								
rs9275224	A/G	*HLA-DQB1*	6p21.31	Discovery	0.407	0.488	0.71 (0.63–0.80)	7.94E-09
	—	—	32659878	Validation	0.325	0.460	0.59 (0.51–0.68)	7.86E-13
	—	—	—	Meta-analysis	0.374	0.477	0.66 (0.60–0.72)	2.97E-19

[a]Effect Allele Frequency.
[b]Noneffect allele.
[c]Effect allele (tested allele).
[d]OR (95% CI).

Abbreviations: GATAD2, GATA zinc finger domain containing 2A; GWAS, GWAS, genome-wide association study; HLA−DQB1, human leukocyte antigen DQ beta 1; MAU2, sister chromatid cohesion factor; MOBP, myelin-associated oligodendrocyte basic protein; PARVB, parvin-beta; PNPLA3, patatin-like phospholipase domain-containing 3; SAMM50, sorting and assembly machinery component 50 homolog; SNP, single-nucleotide polymorphism; SUGP1, SURP and G-patch domain containing 1; TERT, telomerase reverse transcriptase; TM6SF2, transmembrane 6 superfamily member 2.

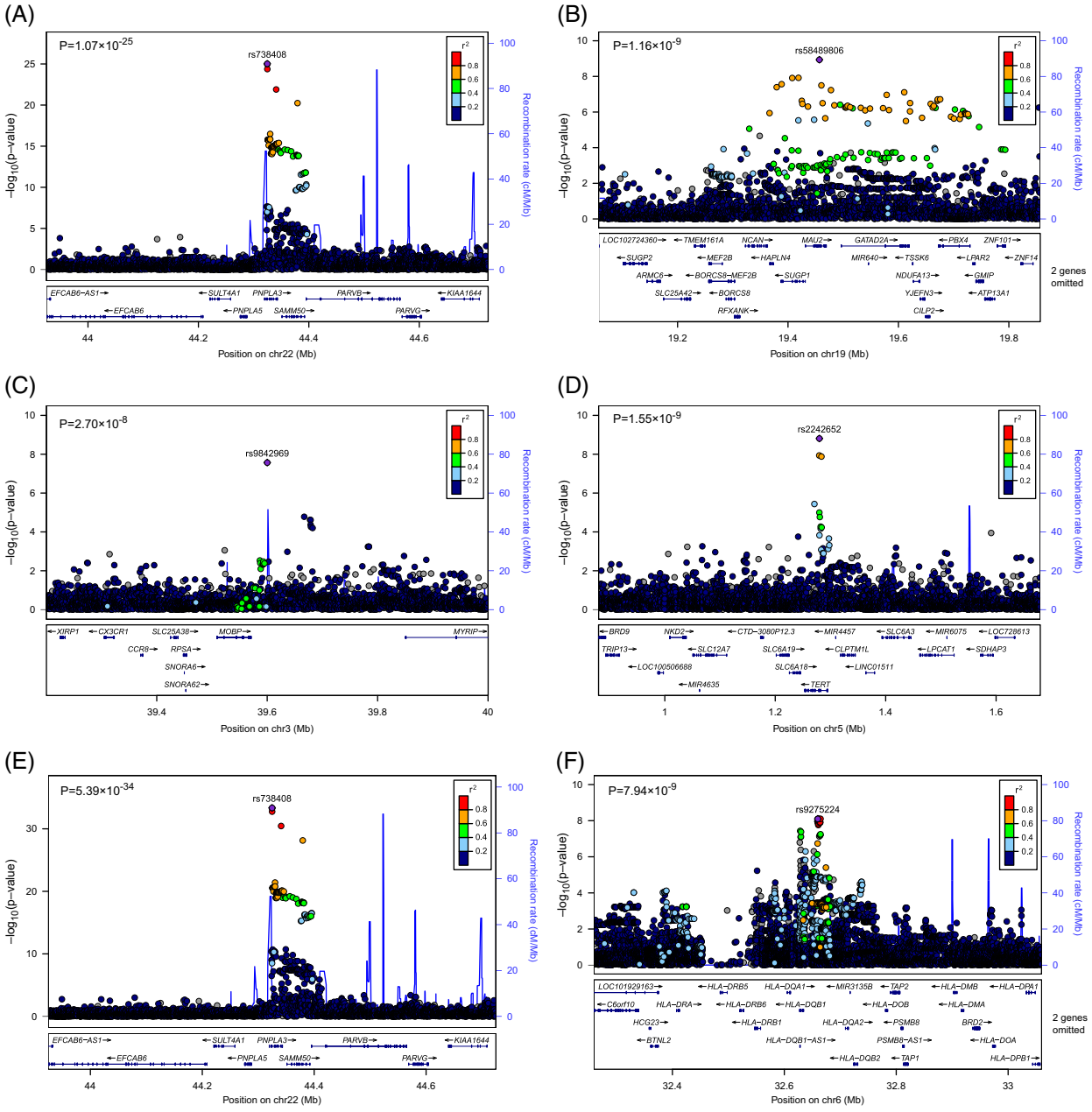

FIGURE 2 Regional visualization of the GWAS of –log10 of the *p*-value by genomic locations: (A) *PNPLA3* in overall HCC (rs738408); (B) *MAU2* in overall HCC (rs58489806); (C) *MOBP* in overall HCC (rs9842969); (D) *TERT* in overall HCC (rs2242652); (E) *PNPLA3* in non-HCV–induced HCC (rs738408); (F) *HLA−DQB1* in HCV-induced HCC (rs9275224). In all figures, each dot represents a SNP, with the location depicted on the corresponding genes at the bottom. Abbreviations: GWAS, genome-wide association study; HLA−DQB1, human leukocyte antigen DQ beta 1; MAU2, sister chromatid cohesion factor; MOBP, myelin-associated oligodendrocyte basic protein; PNPLA3, patatin-like phospholipase domain-containing 3; SNP, single-nucleotide polymorphism; TERT, telomerase reverse transcriptase.

phases and overall *p* values in the meta-analysis of 1.36E-09 for *MOBP* and 8.27E-15 for *TERT*.

Adjusting for known covariates or stratified analyses by alcohol, smoking, type 2 diabetes mellitus, and HCV infection did not substantially affect our results (Figure 3). However, the ORs for lipid-related genetic factors all increased after these adjustments (Supplemental Table S6, http://links.lww.com/HEP/I302). Restricting analyses to the HCV-negative population with further adjustment also increased the magnitude of the association between lipid-related genes and HCC risk.

For the HCV-positive group, 1223 HCV-associated HCC and 739 HCV-positive controls were included in the discovery and validation phases. We identified 35 SNPs within a 34-kb region on chromosome 6, reaching genome-wide significant associations with HCC (Supplemental Table S4B, http://links.lww.com/HEP/I302). Most of these SNPs were intergenic or regulatory region

MOBP 3p22.1 rs9842969

		OR	95%CI	p-value
Alcohol Drinking	Yes	0.57	0.43-0.75	5.16×10^{-5}
	No	0.52	0.36-0.74	3.19×10^{-4}
Diabetes Mellitus	Yes	0.42	0.29-0.72	6.92×10^{-4}
	No	0.57	0.46-0.70	2.86×10^{-7}
Cigarette Smoking	Yes	0.61	0.48-0.78	9.76×10^{-5}
	No	0.50	0.38-0.67	1.81×10^{-6}
HCV Status	Positive	0.66	0.52-0.83	5.83×10^{-4}
	Negative	0.51	0.34-0.77	1.27×10^{-3}

TERT 5p15.33 rs2242652

		OR	95%CI	p-value
Alcohol Drinking	Yes	0.67	0.59-0.77	4.81×10^{-9}
	No	0.73	0.62-0.87	2.33×10^{-4}
Diabetes Mellitus	Yes	0.65	0.53-0.80	3.47×10^{-5}
	No	0.70	0.63-0.78	4.53×10^{-10}
Cigarette Smoking	Yes	0.64	0.57-0.73	1.91×10^{-11}
	No	0.78	0.68-0.89	3.38×10^{-4}
HCV Status	Positive	0.72	0.63-0.81	2.88×10^{-7}
	Negative	0.67	0.60-0.75	3.6×10^{-12}

HLA-DQB1 6p21.32 rs9275224

		OR	95%CI	p-value
Alcohol Drinking	Yes	0.76	0.69-0.84	2.41×10^{-8}
	No	0.86	0.76-0.97	1.14×10^{-2}
Diabetes Mellitus	Yes	0.78	0.68-0.91	1.37×10^{-3}
	No	0.78	0.72-0.85	2.3×10^{-9}
Cigarette Smoking	Yes	0.75	0.68-0.82	1.08×10^{-9}
	No	0.83	0.75-0.92	2.32×10^{-4}
HCV Status	Positive	0.65	0.54-0.77	1.3×10^{-6}
	Negative	0.83	0.53-1.29	4.01×10^{-1}

TM6SF2 19p13.11 rs58542926

		OR	95%CI	p-value
Alcohol Drinking	Yes	1.47	1.25-1.74	3.95×10^{-6}
	No	1.62	1.30-2.03	2.16×10^{-5}
Diabetes Mellitus	Yes	1.77	1.38-2.28	7.77×10^{-6}
	No	1.35	1.17-1.55	3.4×10^{-5}
Cigarette Smoking	Yes	1.38	1.17-1.62	1.01×10^{-4}
	No	1.76	1.49-2.07	3.07×10^{-11}
HCV Status	Positive	1.09	0.92-1.30	2.95×10^{-1}
	Negative	1.83	1.54-2.17	3.34×10^{-12}

PNPLA3 22q13.31 rs738409

		OR	95%CI	p-value
Alcohol Drinking	Yes	1.70	1.52-1.89	1.91×10^{-21}
	No	1.44	1.26-1.66	1.73×10^{-7}
Diabetes Mellitus	Yes	2.02	1.70-2.40	1.76×10^{-15}
	No	1.47	1.34-1.60	8.36×10^{-17}
Cigarette Smoking	Yes	1.58	1.42-1.76	2.05×10^{-17}
	No	1.68	1.50-1.87	1.68×10^{-20}
HCV Status	Positive	1.17	0.95-1.44	1.36×10^{-1}
	Negative	2.07	1.90-2.26	8.17×10^{-60}

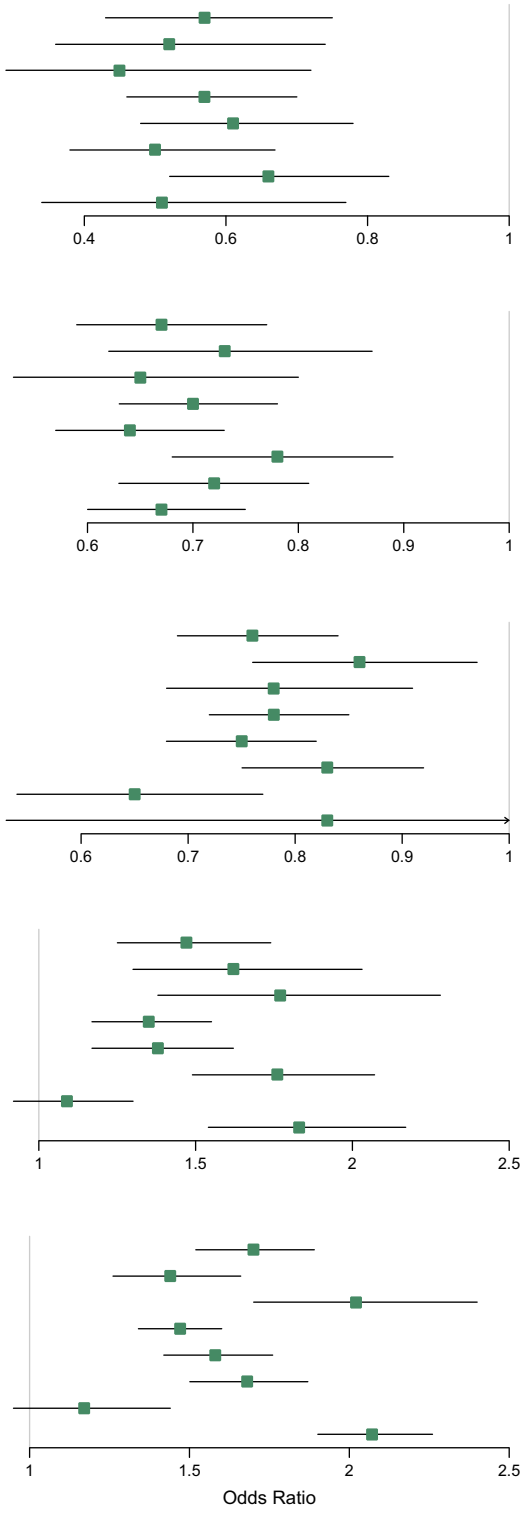

FIGURE 3 Graphical display (Forest Plot) of the estimated ORs (95% CIs) of the significant SNPs stratified by the main nongenetic risk factors for HCC, including alcohol drinking, type 2 diabetes mellitus, cigarette smoking, and HCV infection. Abbreviation: SNP, single-nucleotide polymorphism.

variants, and 4 SNPs were within the human leukocyte antigen DQ beta 1(HLA-DQB1) locus. The top association of *HLA-DQB1* is rs9275224 (Table 2). In a sensitivity analysis of genetic factors in known HCV-negative cases compared to HCV-controls (Supplemental Table S7, http://links.lww.com/HEP/I302), the results show consistent results compared to including all controls.

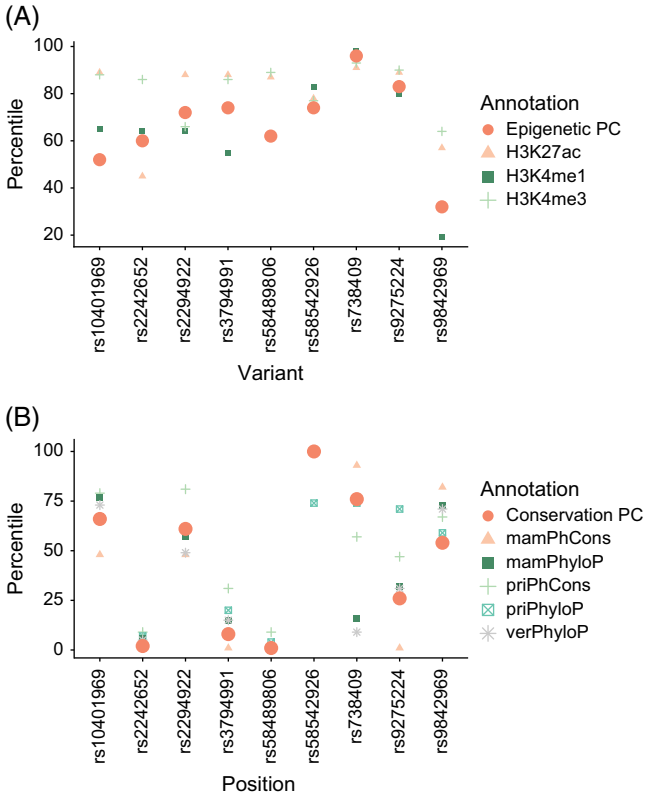

FIGURE 4 Integrative multi-omnics annotation analysis: (A) Plot for epigenetic annotation values. Instead of plotting raw values, for example, H3K27ac peaks, we showed the percentiles of scores of each attribute when ranked against findings across the entire genome. A higher percentile indicates a more functional value. The large red circle is an epigenetic PC that integrates all epigenetic annotation values, including some that were not plotted. (B) Plot for evolutionary conservation annotations. These annotations use data from many different organisms to describe whether certain genetic sequences are conserved through many evolutionary processes. More conserved sequences are believed to carry more functional importance. For instance, the mamPhCons annotation shows the level of conservation in mammals according to the phastCons algorithm. The large red circle is a PC (conservation PC) that integrates all conservation annotation values, including some not shown. We can see that rs58542926 is highly prioritized by these conservation annotations. Abbreviation: PC, principal component.

We estimated the SNP-heritability on the liability scale of HCC from discovery data with the use of genome-wide complex trait analysis-genome-based restricted maximum likelihood.[33] Assuming a lifetime risk of 1% for HCC, the heritability was estimated to be 0.221 ± 0.039. Partitioning the heritability around the 5 genome-wide significant regions, using the sentinel SNP ± 200 kilobases to define regions, indicated the known heritability values were 0.038 ± 0.008 for the regional heritability and 0.183 ± 0.039 for the remainder of the genome. Array heritability estimates for overall heritability in HCV-negative and HCV-positive HCC were similar, yielding estimates respectively of 0.258 ± 0.05 and 0.298 ± 0.067. For overall HCC, HCV-negative and HCV-positive HCC, 17.1%, 19.7%, and 5.9%, respectively, of the genome-wide heritability was explained by the regions around the SNPs that reach genome-wide significance.

Using data from the meta-analysis shown in Table 1, the population attributable fraction per loci were *PNPLA3*, 0.13; *TERT*, 0.26; *TM6SF2*, 0.04; *HLA*, 0.11; and *MOBP*, 0.03 (Supplemental Table S8, http://links.lww.com/HEP/I302, Supplemental Methods, http://links.lww.com/HEP/I303). For the combined population attributable fraction, we estimated that 46% of attributable risk for HCC may be removed if modulation of the effects from these loci becomes possible.

Using a colocalization tool, coloc (colocalization posterior probability, $PPH_4 > 0.8$),[36] we identified 1 or more colocalized genes for all but one GWAS loci in at least 1 tissue type (Supplemental Table S9, http://links.lww.com/HEP/I302). A total of 5 candidate genes were identified by eQTL and 13 candidate genes by sQTL colocalization. From the locus on chromosome 22, *PNPLA3* sQTL signals from liver tissue displayed colocalization ($PPH_4 = 0.86$). For the locus at 19q13.11, *MAU2* ($PPH_4 = 0.84$; whole blood) and ATP 13A1 ($PPH_4 = 0.90$; adipose) exhibited the strongest colocalization based on eQTL (Supplemental Figure S4, http://links.lww.com/HEP/I302). For the HCV-positive locus on chromosome 6, e/sQTL colocalization identified multiple genes, including HLA genes (*HLA-DQB1/2* sQTL; $PPH_4 = 0.98$; lymphocytes).

We also performed CART analysis among the most significant 5 SNPs to identify any higher-order genetic interactions influencing disease risk (Supplemental Figure S5, http://links.lww.com/HEP/I302, Supplemental Tables 10-11, http://links.lww.com/HEP/I302). This analysis identified a subset of individuals at-risk who have the genetic combination of the minor variant homozygote of *PNPLA3* and homozygote common variant of *TERT* showing a 4.12-fold increased risk of developing HCC (95% CI = 3.18–5.33) compared with those who have only the homozygote of *PNPLA3*. Results were similar after adjusting for known covariates (Supplemental Table S12, http://links.lww.com/HEP/I302). The risk from this joint genotype was 6.49 (95% CI = 4.78–8.80) and 4.56 (95% CI = 3.16–6.58) for nonviral-related HCC, before covariate adjustment and after adjusting for age, gender, alcohol drinking, and type 2 diabetes mellitus status.

We conducted integrative multi-omics annotation analysis to explore the functional significance of the top 10 hits (Supplemental Tables S13-14, http://links.lww.com/HEP/I302). A selection of annotation values is plotted in Figure 4. Specifically, Figure 4A shows epigenetic annotation values known to predict regulatory function. We can see that the selected variants generally show above-average evidence of epigenetic functionality. Figure 4B shows evolutionary conservation annotations. These annotations use data from many different organisms to describe whether certain genetic sequences are conserved through many

evolutionary processes. We can see that rs58542926 is highly prioritized by these conservation annotations.

DISCUSSION

This large GWAS is the first to highlight novel genetic susceptibility to nonviral HCC among EDP from North America. We specifically demonstrated substantial heritability in nonviral HCC. The strongest hit was a previously established functional variant, rs738409 (I148M), of *PNPLA3*.

PNPLA3 mRNA is expressed in the liver and the adipose tissue.[38] Its expression is positively correlated with body mass index and carbohydrate intake.[39,40] The function of the *PNPLA3* gene has not been well described, but it has been reported to have lipase-like activity to promote triglyceride hydrolysis in the liver.[41] In addition to its lipase activity, *PNPLA3* may be involved in fatty acid transfer to monoacylglycerol or diacylglycerol.[42] The direct role of *PNPLA3* in cancer development has not been examined, yet its association with chronic fatty liver disorders and cirrhosis may indirectly explain its role in HCC development.

The second top hit is the rs58542926 (E167K) variant of the gene encoding *TM6SF2*, which has a relatively low minor allele frequency. Our results showed that the frequency of the effect allele in our controls was consistent with the global allele frequency of 0.07.[43] Previous studies suggested that the variant might regulate liver transcript and protein expression in an allele-specific manner in animals and humans.[43] Results from previous meta-analyses[43,44] supported our findings for the positive association between this missense variant and HCC. *TM6SF2* is a key regulator of liver fat metabolism, influencing serum levels of total cholesterol, LDL cholesterol, and triglyceride and might modulate hepatic fat accumulation.[43] It has been suggested that the rs58542926 (E167K) variant might influence the cell cycle in HCC HEPA1–6 cells through cyclin D1 and P53 upregulation and P27 downregulation.[44] Moreover, this variant may promote the inflammatory response and provoke cell injury observed in HCC.[44]

TERT variants have been associated with risk of various types of human cancers. The A allele of SNP rs2242652, which confers increased *TERT* expression, has been associated with reduced risk of HCC in the Chinese Han population[45] and alcohol-induced HCC among Europeans.[24] These results highlight the importance of telomere function for liver cancer development.

The significance of rs738409, rs58542926, and rs2242652 variants of *PNPLA3, TM6SF2,* and *TERT*, respectively, was consistent with the recently published European alcohol-associated HCC GWASs.[23,24,46] In addition, our study highlights 2 novel variants of *MAU2* and *MOBP* for the susceptibility on nonviral HCC, which were not previously reported.

The *MAU2* variant is a novel HCC susceptibility locus identified in this study, and eQTL colocalization identified *MAU2* as a candidate susceptibility gene in this locus. Variants in *MAU2* have been associated with lipid traits,[47] and this gene is known for its function in cohesin binding to chromatin, sister chromatid cohesion, and mitotic progression. Thus, it is possible that *MAU2* may contribute to HCC development through mechanisms of increased chromosome instability and lipid metabolism. Our findings highlight the role of fatty liver pathways in HCC development in the absence of HCV and HBV.

The second novel finding for HCC susceptibility is the *MOBP* variant. Although the *MOBP* function has not been well elucidated, its *MOBP* expression is altered in the brain of patients with alcoholism who tend to experience alcohol-induced brain atrophy,[48] which may implicate a role of *MOBP* in the maintenance and integrity of myelin sheath. The observed protected effect of rs39600177 was independent of alcohol consumption by conditional analysis (Supplemental Methods, http://links.lww.com/HEP/I303). Meanwhile, *MOBP* has been suggested to play a role in lipid transport and cholesterol metabolism.[49]

Genetic susceptibility of the immune response is a possible pathway for HCC development among HCV-related HCC. This is exemplified by the association between HCV-positive HCC and our novel finding of association with the *HLA-DQB1* variant. Moreover, sQTL colocalization in lymphocytes supported this gene and other HLA genes in this locus. Multiple studies have reported a significant association of *HLA-DQB1* alleles with HCV spontaneous clearance and persistence in patients with chronic HCV infection.[50] A previous study found that *HLA-DQB1* was independently associated with HCC, and HCV genotypes modified the effects of *HLA-DQB1* on the risk of HCC.[22] Further studies are needed to refine these estimates and to evaluate the role of specific HLA alleles on HCC risk, especially in the presence of HCV.

Since the effect of genetic variations of the validated genes on HCC risk is possibly mediated through fatty liver infiltration, fatty liver disease is considered an intermediate factor in the HCC carcinogenic pathway. Once steatosis has developed in genetically susceptible patients, lipid peroxidation, inflammation, and generated free oxygen radicals may play a central role in pathological progression of steatosis to steatohepatitis, during which the initiation phase of the HCC mechanism may take place. The balance between apoptotic and anti-apoptotic factors and the disturbance in the growth factors may facilitate oval cell proliferation and the promotion phase of hepatocarcinogenesis. Adjustment for the major etiological factors related to fatty liver disease did not change our findings, indicating the additional value of considering these genetic factors for risk analyses of HCC.

The heritability of HCC (> 20%) is similar to the heritability of other gastrointestinal cancers like colon and pancreatic cancer.[51] Because the estimated heritability is substantial, genetic studies are likely to continue to provide insights into the pathways that contribute to increased risk for developing HCC. Moreover, the CART analysis identified a subset of individuals with the combination of homozygous variants of *PNPLA3* and *TERT*, showing a 6.5-fold higher risk for nonviral-related HCC compared to individuals lacking these genotypes. This observation suggests that gene-gene interactions may identify individuals at elevated risk for developing HCC.

Our study has some limitations. First is the lack of information about HCV treatment among HCV-related HCC cases and HCV-positive controls. All cases and controls were enrolled prior to the initiation of the current curative direct-acting-antiviral treatment of HCV. It is possible that some participants had been exposed to a single or combination of interferon-based treatments that has a lower response rate. Nevertheless, this may lead to an equal misclassification (nondifferential) of HCV treatment exposure with toward the null bias. Despite this possible null bias, we still observed very significant associations between *HLA-DQB1*, rs9275224 and HCV-related HCC. In addition, both cases and controls were matched by cirrhosis. There is no evidence that SNP *HLA-DQB1*, rs9275224, is a significant predictor for HCV treatment response. Therefore, we do not believe that prior HCV treatment is a substantial confounding factor in this study.

The second limitation is missing HCV testing in some of the healthy controls. Serological markers for HBV and HCV were only available for 6% of the healthy controls. According to the NHANES report, the highest HCV prevalence among the white US population was 0.9% in subjects aged 50 years or older, and the highest HBV prevalence among this group was 0.05%.[26] Given that the mean age of the controls in our study was 60 years, we estimated that only 36 healthy controls may have been HCV-positive but were coded as HCV-negative during analysis. Such a small number of HCV-positive controls would not have a meaningful impact on the estimated final ORs in this large study. In sensitivity analyses, the estimates of the ORs from the study comparing known HCV-negative cases to all controls were consistent with findings restricted to those participants known to be HCV-negative. Finally, the lack of functional characterization of our top variants may limit the generalization of our results, but functional annotation indicates that all of the most significant variants are likely to be influenced by epigenetic remodeling.

To the best of our knowledge, this is the largest HCC GWAS in EDP from North America with stratified analyses by HCV with the addition of CART, conditional, and colocalization analyses. A recently published US GWAS[25] did not detect the significance of the above genetic variants reported by this GWAS and the European GWASs. As suggested by the authors, this may be explained by the study limitations, including a small sample size, the multiethnic population, and the use of smaller microarray chips. On the other hand, there are inherent differences in etiology, genetics, and clinical outcome between HCC from Eastern and Western populations.[52] Accordingly, we did not anticipate similar results between our GWAS and the Asian GWASs.[15,19] Further studies are warranted to identify common genetic predisposition factors that contribute to susceptibility to HCC development in minorities, in whom the incidence of HCC and the prevalence of metabolic syndrome are higher than in EDP. Additional GWASs on HCC are needed to confirm our findings and to allow further gene-environment and gene-gene interaction analyses to assess previously reported variants in association with HCC, for example, hydroxysteroid 17-beta dehydrogenase 13 variant (rs72613567 or rs6834314) in alcohol-associated HCC.[53–56] Supplemental Table S15, http://links.lww.com/HEP/I302 shows that rs6834314 is protective for this variant, yielding an overall OR of 0.87 per allele ($p = 0.0006$) for HCC carriers per allele, but fails to reach genome-wide significance. This OR was similar to the value of 0.85 previously found in meta-analysis across several studies for cirrhosis.[10] Confirmation of our findings may lead to novel strategies to identify high-risk individuals for HCC, especially among patients with fatty liver diseases, for whom we have already identified several loci that substantially contribute to the risk for HCC.

DATA AVAILABILITY STATEMENT

Data that support the findings of this study are available from the corresponding author, upon request.

AUTHOR CONTRIBUTIONS

Manal M. Hassan, Donghui Li, and Lewis R. Roberts conceived the project. Younghun Han, Jinyoung Byun, and Christopher I. Amos performed the main analyses. Donghui Li, Manal M. Hassan, Younghun Han, and Christopher I. Amos wrote the first draft of the manuscript. Yanan Li and Ping Chang conducted the DNA extraction, QC, and Taqman assays. All authors participated in data interpretation, funding support, and data collection (ie, sample and epidemiological data collection). All authors also reviewed and approved the manuscripts.

FUNDING INFORMATION

This work was supported by NIH/NCI grant RO1CA186566 (Manal M. Hassan, Donghui Li, Lewis R. Roberts, and Christopher I. Amos) and the Center for Inherited Disease Research grant X01 HG007490 (Manal M. Hassan). Christopher I. Amos is a Research Scholar of the Cancer Prevention Research Interest of Texas

(CPRIT) award (RR170048). Lewis R. Roberts was supported in part by the Mayo Clinic SPORE in Hepatobiliary Cancers NIH/NCI P50CA210964.

CONFLICTS OF INTEREST

Robin Kate Kelley advises and received grants from Agios, AstraZeneca, Excelixis, Ipsen, and Merck. She advises Compass, Kinnate, Regeneron, and Tyra Biosciences. She received grants from Bayer, Bristol Myers Squibb, Eli Lilly, EMD Serono, Genentech, Loxo Oncology, Partner Therapeutics, Roche, QED, Relay Therapeutics, Servier, Surface Oncology, and Taiho. Jian-Min Yuan received grants from the National Institutes of Health. R. Mark Ghobrial consults for TransMedics. Yuko Kono advises Lantheus Medical Imaging. She received grants from Bracco Diagnostics and Canon Medical Systems. Dimpy P. Shah received grants from Anthos, AstraZeneca, and Guardant. Mindie H. Nguyen advises and received grants from Exact, Gilead, and Intercept. She received grants from AstraZeneca, CurveBio, Delfi, Enanta, Innogen, Pfizer, and Vir. Geoffrey Liu advises and received grants from AstraZeneca, Pfizer, and Takeda. He advises EMD Serono, Jazz, Merck, and Novartis. Richard Kim consults and is on the speakers' bureau for AstraZeneca. He consults for AbbVie, Eisai, Excelixis, Ipsen, Pfizer, and Taiho. He is on the speakers' bureau for Incyte. Ernest Hawk received grants from Exact. Alison P. Klein received grants from the National Cancer Institute. Mikayla A. Schmidt is employed by the Mayo Clinic. Saira A. Khaderi consults for AstraZeneca. Ahmed O. Kaseb consults, advises, and received grants from Bristol Myers Squibb, Eisai, Exelixis, Genentech, Merck, and Roche. He received grants from Adaptimmune and Tvardi. Lewis R. Roberts advises and received grants from Bayer, Exact, and Gilead. He advises Grail, QED Therapeutics, and Tavec. He received grants from Ariad, BTG/Boston Scientific, Fujifilm, Glycotest, RedHill, and Target. The remaining authors have no conflicts to report.

ORCID

Manal M. Hassan https://orcid.org/0000–0002–8413–449X
Donghui Li https://orcid.org/0000–0003–4241–5522
Younghun Han https://orcid.org/0000–0001–5048–8479

REFERENCES

1. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. Nat Rev Dis Primers. 2021;7:6.
2. Rinella ME. Nonalcoholic fatty liver disease: A systematic review. JAMA. 2015;313:2263–73.
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
4. Chen CH, Huang GT, Lee HS, Yang PM, Chen DS, Sheu JC. Clinical impact of screening first-degree relatives of patients with hepatocellular carcinoma. J Clin Gastroenterol. 1998;27:236–9.
5. Turati F, Edefonti V, Talamini R, Ferraroni M, Malvezzi M, Bravi F, et al. Family history of liver cancer and hepatocellular carcinoma. Hepatology. 2012;55:1416–25.
6. Hassan MM, Spitz MR, Thomas MB, Curley SA, Patt YZ, Vauthey JN, et al. The association of family history of liver cancer with hepatocellular carcinoma: A case-control study in the United States. J Hepatol. 2009;50:334–41.
7. 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.
8. Li Y, Si L, Zhai Y, Hu Y, Hu Z, Bei JX, et al. Genome-wide association study identifies 8p21.3 associated with persistent hepatitis B virus infection among Chinese. Nat Commun. 2016;7:11664.
9. Miki D, Ochi H, Takahashi A, Hayes CN, Urabe Y, Abe H, et al. HLA-DQB1*03 confers susceptibility to chronic hepatitis C in Japanese: A genome-wide association study. PLoS One. 2013;8:e84226.
10. Emdin CA, Haas M, Ajmera V, Simon TG, Homburger J, Neben C, et al. Association of genetic variation with cirrhosis: A multitrait genome-wide association and gene-environment interaction study. Gastroenterology. 2021;160:1620–33.e13.
11. Teufel A, Staib F, Kanzler S, Weinmann A, Schulze-Bergkamen H, Galle PR. Genetics of hepatocellular carcinoma. World J Gastroenterol. 2007;13:2271–82.
12. Reay WR, Cairns MJ. Advancing the use of genome-wide association studies for drug repurposing. Nat Rev Genet. 2021;22:658–71.
13. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: Biology, function, and translation. Am J Hum Genet. 2017;101:5–22.
14. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. Hum Mol Genet. 2019;28(R2):R133–r42.
15. Li Y, Zhai Y, Song Q, Zhang H, Cao P, Ping J, et al. Genome-wide association study identifies a new locus at 7q21.13 associated with Hepatitis B virus-related hepatocellular carcinoma. Clin Cancer Res. 2018;24:906–15.
16. Sawai H, Nishida N, Khor SS, Honda M, Sugiyama M, Baba N, et al. Genome-wide association study identified new susceptible genetic variants in HLA class I region for hepatitis B virus-related hepatocellular carcinoma. Sci Rep. 2018;8:7958.
17. Zhang H, Zhai Y, Hu Z, Wu C, Qian J, Jia W, et al. Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. Nat Genet. 2010;42:755–8.
18. Yu XJ, Fang F, Tang CL, Yao L, Yu L, Yu L. dbHCCvar: A comprehensive database of human genetic variations in hepatocellular carcinoma. Hum Mutat. 2011;32:E2308–16.
19. Miki D, Ochi H, Hayes CN, Abe H, Yoshima T, Aikata H, et al. Variation in the DEPDC5 locus is associated with progression to hepatocellular carcinoma in chronic hepatitis C virus carriers. Nat Genet. 2011;43:797–800.
20. Urabe Y, Ochi H, Kato N, Kumar V, Takahashi A, Muroyama R, et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. J Hepatol. 2013;58:875–82.
21. Matsuura K, Sawai H, Ikeo K, Ogawa S, Iio E, Isogawa M, et al. Genome-wide association study identifies TLL1 variant associated with development of hepatocellular carcinoma after eradication of hepatitis C virus infection. Gastroenterology. 2017;152:1383–94.
22. Lee MH, Huang YH, Chen HY, Khor SS, Chang YH, Lin YJ, et al. Human leukocyte antigen variants and risk of hepatocellular

carcinoma modified by hepatitis C virus genotypes: A genome-wide association study. Hepatology. 2018;67:651–1.
23. Trépo E, Caruso S, Yang J, Imbeaud S, Couchy G, Bayard Q, et al. Common genetic variation in alcohol-related hepatocellular carcinoma: A case-control genome-wide association study. Lancet Oncol. 2022;23:161–71.
24. Buch S, Innes H, Lutz PL, Nischalke HD, Marquardt JU, Fischer J, et al. Genetic variation in TERT modifies the risk of hepatocellular carcinoma in alcohol-related cirrhosis: Results from a genome-wide case-control study. Gut. 2023;72:381–91.
25. Wang Z, Budhu AS, Shen Y, Wong LL, Hernandez BY, Tiirikainen M, et al. Genetic susceptibility to hepatocellular carcinoma in chromosome 22q13.31, findings of a genome-wide association study. JGH Open. 2021;5:1363–72.
26. Roberts H, Kruszon-Moran D, Ly KN, Hughes E, Iqbal K, Jiles RB, et al. Prevalence of chronic hepatitis B virus (HBV) infection in U.S. households: National Health and Nutrition Examination Survey (NHANES), 1988-2012. Hepatology. 2016;63:388–97.
27. Wu X, Scelo G, Purdue MP, Rothman N, Johansson M, Ye Y, et al. A genome-wide association study identifies a novel susceptibility locus for renal cell carcinoma on 12p11.23. Hum Mol Genet. 2012;21:456–62.
28. Chen M, Tsai CW, Chang WS, Xiong GY, Xu Y, Bau DT, et al. High circulating insulin-like growth factor-1 reduces the risk of renal cell carcinoma: A Mendelian randomization study. Carcinogenesis. 2021;42:826–30.
29. Byun J, Han Y, Gorlov IP, Busam JA, Seldin MF, Amos CI. Ancestry inference using principal component analysis and spatial analysis: A distance-based analysis to account for population substructure. BMC Genomics. 2017;18:789.
30. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48:1279–83.
31. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007;39:906–13.
32. Han B, Eskin E. Interpreting meta-analyses of genome-wide association studies. PLoS Genet. 2012;8:e1002555.
33. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011;88:294–305.
34. Merkel PA, Xie G, Monach PA, Ji X, Ciavatta DJ, Byun J, et al. Identification of functional and expression polymorphisms associated with risk for antineutrophil cytoplasmic autoantibody-associated vasculitis. Arthritis Rheumatol. 2017;69:1054–66.
35. Woolf B. On estimating the relation between blood group and disease. Ann Hum Genet. 1955;19:251–3.
36. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10:e1004383.
37. Zhou H, Arapoglou T, Li X, Li Z, Zheng X, Moore J, et al. FAVOR: functional annotation of variants online resource and annotator for variation across the human genome. Nucleic Acids Res. 2023;51(D1):D1300–d11.
38. BasuRay S, Wang Y, Smagris E, Cohen JC, Hobbs HH. Accumulation of PNPLA3 on lipid droplets is the basis of associated hepatic steatosis. Proc Natl Acad Sci U S A. 2019;116:9521–6.
39. Romeo S, Huang-Doran I, Baroni MG, Kotronen A. Unravelling the pathogenesis of fatty liver disease: Patatin-like phospholipase domain-containing 3 protein. Curr Opin Lipidol. 2010;21:247–52.
40. Davis JN, Lê KA, Walker RW, Vikman S, Spruijt-Metz D, Weigensberg MJ, et al. Increased hepatic fat in overweight Hispanic youth influenced by interaction between genetic variation in PNPLA3 and high dietary carbohydrate and sugar consumption. Am J Clin Nutr. 2010;92:1522–7.
41. Starley BQ, Calcagno CJ, Harrison SA. Nonalcoholic fatty liver disease and hepatocellular carcinoma: A weighty connection. Hepatology. 2010;51:1820–32.
42. Huang Y, Cohen JC, Hobbs HH. Expression and characterization of a PNPLA3 protein isoform (I148M) associated with nonalcoholic fatty liver disease. J Biol Chem. 2011;286:37085–93.
43. Pirola CJ, Sookoian S. The dual and opposite role of the TM6SF2-rs58542926 variant in protecting against cardiovascular disease and conferring risk for nonalcoholic fatty liver: A meta-analysis. Hepatology. 2015;62:1742–56.
44. Tang S, Zhang J, Mei TT, Guo HQ, Wei XH, Zhang WY, et al. Association of TM6SF2 rs58542926 T/C gene polymorphism with hepatocellular carcinoma: A meta-analysis. BMC Cancer. 2019;19:1128.
45. Zhang Y, Wang S, Shi Y, Cao Y, He H, Zhou S, et al. Associations of TERT polymorphisms with hepatocellular carcinoma risk in a Han Chinese population. Int J Clin Exp Pathol. 2017;10:7776–83.
46. Stickel F, Buch S, Nischalke HD, Weiss KH, Gotthardt D, Fischer J, et al. Genetic variants in PNPLA3 and TM6SF2 predispose the development of hepatocellular carcinoma in individuals with alcohol-related cirrhosis. Am J Gastroenterol. 2018;113:1475–83.
47. Below JE, Parra EJ, Gamazon ER, Torres J, Krithika S, Candille S, et al. Meta-analysis of lipid-traits in Hispanics identifies novel loci, population-specific effects, and tissue-specific enrichment of eQTLs. Sci Rep. 2016;6:19429.
48. Mayfield RD, Lewohl JM, Dodd PR, Herlihy A, Liu J, Harris RA. Patterns of gene expression are altered in the frontal and motor cortices of human alcoholics. J Neurochem. 2002;81:802–13.
49. Poitelon Y, Kopec AM, Belin S. Myelin fat facts: An overview of lipids and fatty acid metabolism. Cells. 2020;9:812.
50. Vergara C, Thio CL, Johnson E, Kral AH, O'Brien TR, Goedert JJ, et al. Multi-ancestry genome-wide association study of spontaneous clearance of hepatitis C virus. Gastroenterology. 2019;156:1496–507.e7.
51. Chen F, Childs EJ, Mocci E, Bracci P, Gallinger S, Li D, et al. Analysis of heritability and genetic architecture of pancreatic cancer: A PanC4 study. Cancer Epidemiol Biomarkers Prev. 2019;28:1238–45.
52. Choo SP, Tan WL, Goh BKP, Tai WM, Zhu AX. Comparison of hepatocellular carcinoma in Eastern versus Western populations. Cancer. 2016;122:3430–46.
53. Stickel F, Lutz P, Buch S, Nischalke HD, Silva I, Rausch V, et al. Genetic Variation in HSD17B13 reduces the risk of developing cirrhosis and hepatocellular carcinoma in alcohol misusers. Hepatology. 2020;72:88–102.
54. Yang J, Trépo E, Nahon P, Cao Q, Moreno C, Letouzé E, et al. A 17-beta-hydroxysteroid dehydrogenase 13 variant protects from hepatocellular carcinoma development in alcoholic liver disease. Hepatology. 2019;70:231–40.
55. 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.
56. Kogiso T, Ogasawara Y, Horiuchi K, Taniai M, Tokushige K. Analysis of genetic factors associated with fatty liver disease-related hepatocellular carcinoma. Cancer Med. 2023;12:17798–807.

How to cite this article: Hassan MM, Li D, Han Y, Byun J, Hatia RI, Long E, et al. Genome-wide association study identifies high-impact susceptibility loci for HCC in North America. Hepatology. 2024;80:87–101. https://doi.org/10.1097/HEP.0000000000000800