


Original Article

A transcriptome-wide association study identified susceptibility genes for hepatocellular carcinoma in East Asia

Jingjing Zhang ^{1,2,†}, Qingrong Zhang^{1,2,†}, Wenyan Hu^{1,2,†}, Yuxuan Liang^{1,2}, Deke Jiang³ and Haitao Chen^{1,2,*}

¹School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou, Guangdong, P. R. China

²School of Public Health (Shenzhen), Shenzhen campus of Sun Yat-sen University, Shenzhen, Guangdong, P. R. China

³Department of Infectious Diseases and Hepatology Unit, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, P. R. China

*Corresponding author: Haitao Chen, School of Public Health (Shenzhen), Sun Yat-sen University, No. 132, East Ring Road, Guangzhou, Guangdong 510006, China. Tel: +86-18321336540; Email: chenht56@mail.sysu.edu.cn

[†]These authors (J.Z., Q.Z. and W.H.) contributed equally to this work.

Abstract

Background: Hepatocellular carcinoma (HCC) is one of the most common cancers worldwide and is prevalent in East Asia. Although genome-wide association studies (GWASs) of HCC have identified 23 risk regions, the susceptibility genes underlying these associations largely remain unclear. To identify novel candidate genes for HCC, we conducted liver single-tissue and cross-tissue transcriptome-wide association studies (TWASs) in two populations of East Asia.

Methods: GWAS summary statistics of 2,514 subjects (1,161 HCC cases and 1,353 controls) from the Chinese Qidong cohort and 161,323 subjects (2,122 HCC cases and 159,201 controls) from the BioBank Japan project were used to conduct TWAS analysis. The single-tissue and cross-tissue TWAS approaches were both used to detect the association between susceptible genes and the risk of HCC. TWAS identified genes were further annotated by Metascape, UALCAN, GEPIA2, and DepMap.

Results: We identified 22 novel genes at 16 independent loci significantly associated with HCC risk after Bonferroni correction. Of these, 13 genes were located in novel regions. Besides, we found 83 genes overlapped in the Chinese and Japanese cohorts with $P < 0.05$, of which, three genes (NUAK2, HLA-DQA1, and ATP6V1G2) were discerned by both single-tissue and cross-tissue TWAS approaches. Among the genes identified through TWAS, a significant proportion of them exhibit a credible role in HCC biology, such as FAM96B, HSPA5, POLRMT, MPHOSPH10, and RABL2A. HLA-DQA1, NUAK2, and HSPA5 associated with the process of carcinogenesis in HCC as previously reported.

Conclusions: Our findings highlight the value of leveraging the gene expression data to identify new candidate genes beyond the GWAS associations and could further provide a genetic insight for the biology of HCC.

Keywords: transcriptome-wide association study; hepatocellular carcinoma; genome-wide association study; single nucleotide polymorphism; genetic mechanism

Introduction

Being the sixth most commonly diagnosed cancer and the second leading cause of cancer-related death, hepatocellular carcinoma (HCC) constitutes one of the major health problems worldwide [1]. There is regional heterogeneity in incidence of HCC; it is estimated that 72% of cases occur in Asia and more than 50% in China [2].

Advancements in high-throughput sequencing have significantly enhanced our comprehension of the intricate molecular landscape of HCC. Over the past decade, genome-wide association studies (GWASs) successfully identified 40 single nucleotide polymorphisms (SNPs) associated with the risk of HCC [3]. These SNPs affected several genes, including DDX18, DEPDC5, MICA, GRIK1, STAT4, KIF1B, PNPLA3, and TM6SF2 [4]. However, the functional

variants and target susceptibility genes at the HCC risk regions have yet to be identified and introduced into clinical practice [4].

Up to 90% of GWAS-associated loci are distributed in intergenic regions or non-coding regions and enriched in predicted transcriptional regulatory regions, called “cis-regulatory elements (CREs)” [5]. Moreover, the GWAS-associated variants were found a high overlap with the SNPs that affect the gene expression identified in the expression quantification traits loci (eQTL) study, which suggests that many GWASs causal variants may influenced disease by altering the downstream target gene expression [5].

As one of the post-GWAS research methods, transcriptome-wide association study (TWAS) could integrate the GWAS signals and eQTLs to identified genes underlying complex traits,

Received: 25 July 2023. Revised: 07 March 2024. Accepted: 30 April 2024

© The Author(s) 2024. Published by Oxford University Press and Sixth Affiliated Hospital of Sun Yat-sen University

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

translating the SNP-disease associations into gene-disease associations, improving the interpretation of the genetic mechanisms underlying GWAS association signals [6]. TWAS has identified hundreds of putative susceptibility genes for many cancers, e.g. pancreatic cancer, prostate cancer, and colorectal cancer [7–10]. However, there have been relatively few TWASs published on HCC to date. [11].

Since HCC has higher incidence in East Asia than in other regions [2], we perform the TWAS in two large-scale HCC GWAS datasets from East Asian populations in this study. One is the Qidong cohort consisting of 1,161 HCC cases and 1,353 non-HCC controls who were all chronic hepatic B virus (CHB) carriers from Eastern China [12]. Another HCC GWAS dataset contains 2,122 diagnosed HCC patients and 159,201 controls from the BioBank Japan project (BBJ) [13]. To further validate our results, both single-tissue and cross-tissue TWAS have been conducted in the two East Asian populations. We hope our study will provide novel clues for the genetic mechanism of HCC in East Asia.

Patients and methods

GWAS data of Chinese

In the initial GWAS stage, 2,689 participants (including 1,212 HCC cases and 1,477 non-HCC controls) were successfully genotyped for a total of 733,202 SNPs. All 2,689 participants were CHB carriers and were recruited from Qidong, Jiangsu province in Eastern China. The determination of the subjects, the genotyping procedure, and details of quality control and imputation were described in the published assay [12]. After quality control, 568,280 SNPs in 2,514 participants (1,161 cases and 1,353 controls) were retained to further analyses. Using the haplotype data from HapMap Han Chinese in Beijing, China and Japanese in Tokyo, Japan subjects, we carried out an imputation analysis to further increase genome coverage using the IMPUTE program. The imputed SNPs were analyzed using the same quality control and filtering criteria used in the previous genotyping steps. At last, 1,672,517 SNPs were retained for the GWAS analysis. We retained the summary statistics of the GWAS results for conducting the TWAS analysis.

GWAS data of Japanese

In order to elucidate the disease biology in the East Asian population, Kazuyoshi Ishigaki et al. [13] conducted a comprehensive GWAS encompassing 42 diseases within the BBJ (<https://biobankjp.org/english/index.html>). They successfully detected 320 independent signals in 276 loci for 27 diseases. The summary results of these GWASs were meticulously preserved in the National Bioscience Database Center (NBDC) Human Database, under the Dataset ID: hum0197.v3.gwas.v1.

For our own subsequent TWASs, we obtained the GWAS summary statistics specifically related to HCC. The summary-level GWAS data comprised 2,122 HCC cases and 159,201 healthy controls. All participants were recruited and followed up between 2003 and 2018 in the BBJ. The diagnosis of HCC cases was confirmed by physicians from the collaborating hospitals of the BBJ consortium. We obtained the HCC GWAS summary data via the following download link: <https://humandbs.biosciencedbc.jp/files/hum0197/hum0197.v3.BBJ.LiC.v1.zip>.

TWASs in single-tissue and cross-tissue

We used the summary-based PrediXcan (S-PrediXcan) method [14] to perform single-tissue TWAS based on GWAS summary statistics. In this process, the European population subset of the 1000 genome project was used as the reference panel to obtain

the linkage disequilibrium information between SNPs within ± 1 Mb of the gene boundaries. We used the publicly available pre-calculated gene expression weights (<https://predictdb.org/>) to perform TWAS statistical inference. We constructed these weights based on gene expression data from 208 liver tissues in the Genotype-Tissue Expression Project version 8 (GTEx v8) datasets [15], following the MASHR-M assumption [16].

For cross-tissue TWAS, we adopted the summary-based MultiXcan (S-MultiXcan) [17] method. Compared with S-PrediXcan, S-MultiXcan capitalizes on the substantial sharing of eQTLs across different tissues and contexts, improving the ability to identify potential target genes. The gene expression prediction models, based on 49 tissues from GTEx v8 and linkage disequilibrium covariance matrix, were also accessible via the PredictDB (<https://predictdb.org/>). Both the single-tissue and cross-tissue TWAS analysis methods were integrated into the MetaXcan software (<https://github.com/hakyimlab/MetaXcan/>).

We used Bonferroni correction to determine the significance threshold. For cross-tissue TWAS, we yielded a threshold of approximately 2.24×10^{-6} (0.05/22,323). In the case of single-tissue TWAS, we yielded a threshold of approximately 4.14×10^{-6} (0.05/12,084). The genes at P -value < 0.05 were considered as suggestive association signals and were subsequently employed in further functional analyses.

Functional annotation for suggestive risk genes

We used various public tools to achieve functional annotations for novel susceptibility genes identified in TWAS, including oncogenesis, prognostic mechanism, and potential treatment targets in HCC. First, we used the Metascape tool (<https://metascape.org/gp/index.html#/>) to explore the function and pathway that the risk genes engaged. Second, we compared the expression level of each gene in HCC tissues with that in normal liver tissues using The University of Alabama at Birmingham CANcer data analysis Portal (UALCAN; <http://ualcan.path.uab.edu/>) [18]. Third, we investigated the prognostic value of HCC using Gene Expression Profiling Interactive Analysis (GEPIA2; <http://gepia2.cancer-pku.cn/#index>) [19]. Finally, we evaluated the gene effect scores of TWAS-suggestive genes in HCC datasets deposited in the DepMap database (<https://depmap.org/portal/>) [20] to identify the potential dependency genes of HCC that could be used as therapeutic targets.

Results

TWAS results in the Chinese population

For the Chinese Han population, our TWASs revealed 434 putative genes in the liver single-tissue analysis and 830 in the cross-tissue analysis, all with a significance level of $P < 0.05$ (Supplementary Table 1 and Supplementary Table 2). Notably, in the liver single-tissue analysis, a single gene ACPPE approached marginal significance ($P_{\text{liver-tissue}} = 4.76 \times 10^{-6}$; Figure 1A and Supplementary Table 1). In the cross-tissue analysis, we identified four genes significantly associated with the risk of HCC at a Bonferroni-corrected threshold of $P < 2.24 \times 10^{-6}$, which were ARHGEF5 ($P_{\text{cross-tissue}} = 2.96 \times 10^{-8}$), AC006272.2 ($P_{\text{cross-tissue}} = 3.61 \times 10^{-7}$), ARNT ($P_{\text{cross-tissue}} = 7.36 \times 10^{-7}$), and MRPL40 ($P_{\text{cross-tissue}} = 9.49 \times 10^{-7}$; Figure 1B and Supplementary Table 2).

TWAS results in the Japanese population

For the Japanese population, TWAS identified 567 putative genes in the liver single-tissue approach and 1,342 in the cross-tissue method ($P < 0.05$; Supplementary Table 1 and Supplementary Table 2). In the liver-tissue analysis, a single

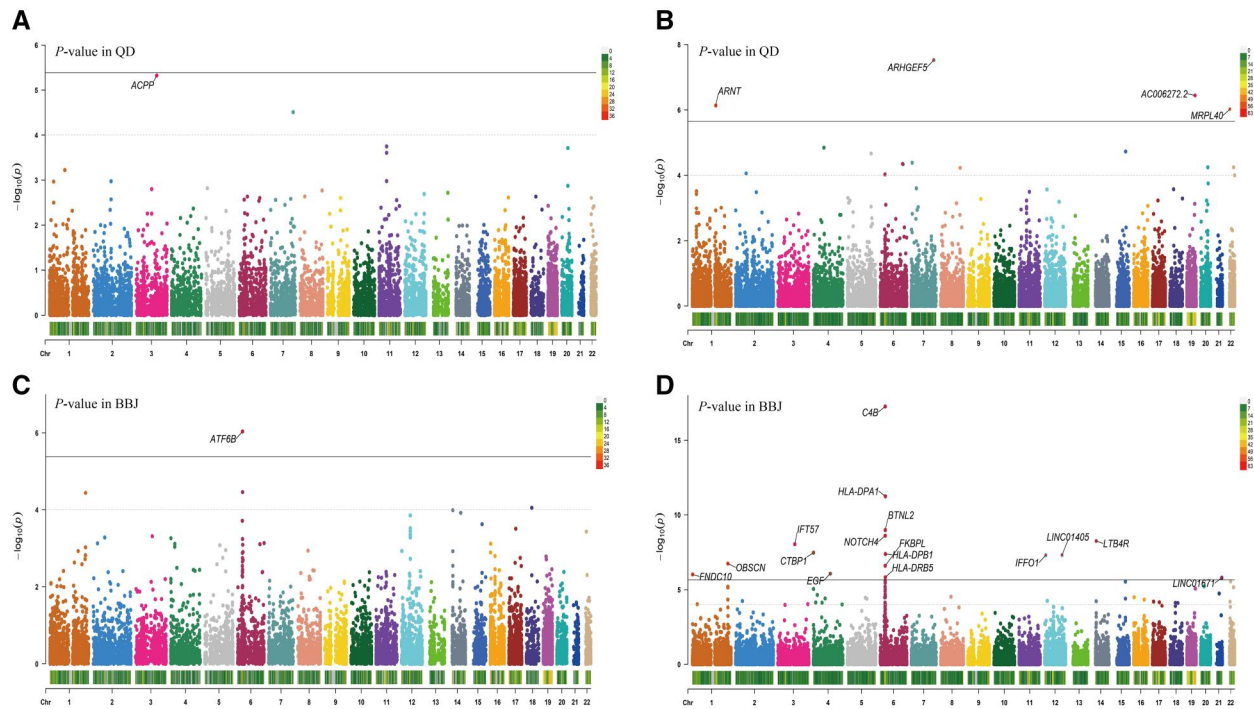


Figure 1. Manhattan plot of the single-tissue and cross-tissue TWAS results in two East Asia populations. (A) Manhattan plot for liver single-tissue TWAS in Chinese. (B) Manhattan plot for cross-tissue TWAS in Chinese. (C) Manhattan plot for liver single-tissue TWAS in Japanese. (D) Manhattan plot for cross-tissue TWAS in Japanese. Each point represents a single gene, the x-axis represents the physical position (chromosome localization) of genes, and the y-axis represents the P-value in $-\log_{10}(p)$. The black solid lines represent a Bonferroni-corrected threshold of $P < 2.24 \times 10^{-6}$ in (A) and (C), $P < 4.14 \times 10^{-6}$ in (B) and (D); grey dotted lines represent a threshold of $P < 1.00 \times 10^{-4}$. QD = Qidong GWAS cohort for Chinese Han population, BBJ = Biobank Japan GWAS cohort for Japanese, Chr = chromosome.

gene, ATF6B ($P_{\text{liver-tissue}} = 9.18 \times 10^{-7}$) was found to be significantly associated with the risk of HCC (Figure 1C and Supplementary Table 1). Additionally, the cross-tissue analysis revealed 16 genes significantly associated with the risk of HCC (Figure 1D and Supplementary Table 2).

In total, we identified 22 genes (Table 1) for which genetically predicted gene expression exhibited significant associations with HCC risk, with 21 genes passing the Bonferroni correction threshold and one coming close to it. Importantly, At three known risk loci (6p21.32, 6p21.33, and 1q42.13) identified in certain GWASs [12, 21–25], our TWAS uncovered nine genes, including ATF6B, C4B, HLA-DPA1, BTNL2, NOTCH4, HLA-DPB1, FKBP1, HLA-DRB5, and OBSCN. Whereas, the remaining 13 genes were previously unidentified.

Overlapped genes between Chinese and Japanese populations

Out of the 434 genes exhibiting $P < 0.05$ in the liver single-tissue analysis for the Chinese population, 14 genes were also identified in the Japanese population with a $P < 0.05$ in single-tissue analysis (Figure 2 and Supplementary Table 3). Furthermore, among the 830 genes with $P < 0.05$ in the cross-tissue analysis for the Chinese population, 72 genes were likewise identified in the Japanese population with a $P < 0.05$ in the cross-tissue analysis (Figure 2 and Supplementary Table 3). Notably, ATP6V1G2 (6p21.33), HLA-DQA1 (6p21.32), and NUA2 (1q32.1) maintained P-values < 0.05 in both TWAS approaches and across both populations (Figure 2 and Supplementary Table 3).

Results from the meta-analysis of Chinese and Japanese populations

In order to enhance the generalizability of these genes, we conducted a meta-analysis of the TWAS results from the two Asian population groups. In single-tissue analyses of liver, ATF6B ($Z_{\text{meta}} = -3.94$; $P_{\text{meta}} = 8.19 \times 10^{-5}$; $P_{\text{fdr}} = 0.98$) emerged as the most significant gene (Supplementary Table 1). However, its P-value did not remain significant after correction for false discovery rate (FDR). In cross-tissue analyses, C4B ($Z_{\text{meta}} = 4.81$; $P_{\text{meta}} = 1.49 \times 10^{-6}$; $P_{\text{fdr}} = 0.03$), HLA-DPA1 ($Z_{\text{meta}} = -4.06$; $P_{\text{meta}} = 4.82 \times 10^{-5}$; $P_{\text{fdr}} = 0.53$), BTNL2 ($Z_{\text{meta}} = -3.95$; $P_{\text{meta}} = 7.96 \times 10^{-5}$; $P_{\text{fdr}} = 0.59$) ranked as the top three most significant genes (Supplementary Table 2). However, only C4B remain statistically significant after FDR correction.

Functional annotations of TWAS identified genes

In total, we identified 105 genes, comprising 22 genes with significant associations and 83 genes consistently exhibiting $P < 0.05$ in two distinct populations, through the utilization of various TWAS methods. Of these 105 genes, 85 were categorized as protein-coding genes, which will undergo further annotation as part of our subsequent exploration of functional and molecular mechanisms.

For the functional annotation of the above 85 genes, we first carried out a pathway and process enrichment analysis with the following ontology categories using Metascape: KEGG Pathway, GO Biological Processes, GO Molecular Functions, Reactome Gene Sets, Canonical Pathways, Resource for Mammalian Protein Complex (CORUM), and Oncogenic Signatures. Genes in the top

Table 1. Statistically significant genes selected by TWAS for HCC

GWAS cohort	Gene symbol	Region	Lead GWAS SNP ^a	GWAS P ^a	Single-tissue			Cross-tissue			Whether in new loci
					TWAS P	FDR	TWAS Z	TWAS P	FDR	TWAS Z	
Qidong	ACPP	3q22.1	rs6712250	1.73×10^{-4}	4.76×10^{-6}	0.057	-4.57	1.49×10^{-3}	0.563	-4.57 to 4.57	Yes
	MRPL40	22q11.21	rs9606045	3.18×10^{-6}	2.48×10^{-3}	0.999	-3.03	9.49 × 10 ⁻⁷	5.29 × 10 ⁻³	-3.55 to 1.76	Yes
	ARHGEF5	7q35	rs2699503	1.48×10^{-5}	3.08×10^{-5}	0.185	-4.17	2.96 × 10 ⁻⁸	6.60 × 10 ⁻⁴	-4.17 to 3.54	Yes
	ARNT	1q21.3	rs75387976	2.66×10^{-3}	NA	NA	NA	7.36 × 10 ⁻⁷	5.47 × 10 ⁻³	-2.19 to 2.19	Yes
	AC006272.2	19q13.41	rs2659048	2.04×10^{-3}	NA	NA	NA	3.61 × 10 ⁻⁷	4.03 × 10 ⁻³	-0.54 to 2.15	Yes
BioBank Japan	ATF6B	6p21.32	rs9271377	5.29×10^{-9}	9.18 × 10 ⁻⁷	0.011	-4.91	2.52×10^{-6}	3.31×10^{-3}	-5.50 to -2.69	No
	C4B	6p21.33	rs9271377	5.29×10^{-9}	0.147	1.000	1.45	5.45 × 10 ⁻¹⁸	1.22 × 10 ⁻¹³	-3.52 to 3.83	No
	IFFO1	12p13.31	rs11829631	2.98×10^{-3}	0.150	1.000	-1.44	4.90 × 10 ⁻⁸	1.09 × 10 ⁻⁴	-1.91 to 1.78	Yes
	FNDC10	1p36.33	rs181784865	5.71×10^{-4}	0.304	1.000	1.03	9.54 × 10 ⁻⁷	1.52 × 10 ⁻³	-2.25 to 3.06	Yes
	FKBP1	6p21.32	rs9271377	5.29×10^{-9}	0.427	1.000	-0.79	2.45 × 10 ⁻⁷	4.55 × 10 ⁻⁴	-3.41 to 4.15	No
	LTB4R	14q12	rs28418362	8.55×10^{-6}	0.571	1.000	0.57	5.62 × 10 ⁻⁹	2.51 × 10 ⁻⁵	-2.90 to 0.57	Yes
	HLA-DRB5	6p21.32	rs9271377	5.29×10^{-3}	3.47×10^{-5}	0.146	4.14	1.49 × 10 ⁻⁶	2.22 × 10 ⁻³	-2.33 to 4.56	No
	IFT57	3q13.12	rs4855552	6.09×10^{-5}	4.87×10^{-4}	0.346	3.49	9.16 × 10 ⁻⁹	3.41 × 10 ⁻⁵	-3.49 to 4.81	Yes
	NOTCH4	6p21.32	rs9271377	5.29×10^{-9}	5.43×10^{-3}	0.689	-2.78	2.47 × 10 ⁻⁹	1.38 × 10 ⁻⁵	-3.90 to 4.64	No
	CTBP1	4p16.3	rs73069980	4.11×10^{-6}	6.33×10^{-3}	0.697	2.73	3.34 × 10 ⁻⁸	1.06 × 10 ⁻⁴	-3.41 to 4.12	Yes
	OBSCN	1q42.13	rs115344926	8.06×10^{-7}	NA	NA	NA	1.75 × 10 ⁻⁷	3.55 × 10 ⁻⁴	-3.53 to 3.35	No
	EGF	4q25	rs114251637	2.46×10^{-4}	NA	NA	NA	8.42 × 10 ⁻⁷	1.44 × 10 ⁻³	-1.11 to 2.09	Yes
	HLA-DPA1	6p21.32	rs9271377	5.29×10^{-9}	NA	NA	NA	5.57 × 10 ⁻¹²	6.21 × 10 ⁻⁸	-6.02 to 2.99	No
	BTNL2	6p21.32	rs9271377	5.29×10^{-9}	NA	NA	NA	1.04 × 10 ⁻⁹	7.73 × 10 ⁻⁶	-4.03 to 2.75	No
	HLA-DPB1	6p21.32	rs9271377	5.29×10^{-9}	NA	NA	NA	4.10 × 10 ⁻⁸	1.14 × 10 ⁻⁴	-5.88 to 2.85	No
LINC01405	12q24.11	rs11066015	3.80×10^{-7}	NA	NA	NA	4.77 × 10 ⁻⁸	1.18 × 10 ⁻⁴	-3.68 to -2.23	Yes	
LINC01671	21q22.3	rs771148881	1.12×10^{-3}	NA	NA	NA	1.60 × 10 ⁻⁶	2.23 × 10 ⁻³	-2.28 to 2.28	Yes	

TWAS = transcriptome-wide association study, GWAS = Genome-wide association study, SNP = single nucleotide polymorphism.

All the genes listed in the above table (except ACPP) have passed Bonferroni correction threshold with a significance level of $P < 2.24 \times 10^{-6}$ in cross-tissue analysis, or $P < 4.14 \times 10^{-6}$ in single-tissue analysis.

^a The lead GWAS variant and GWAS P value indicate the most statistically significant GWAS variant within ±1 Mb for each gene listed.

^b Genes in reported loci implicated in previous GWASs. The black bold font: the P-value and Z-score of the genes labeled on the Manhattan plot.

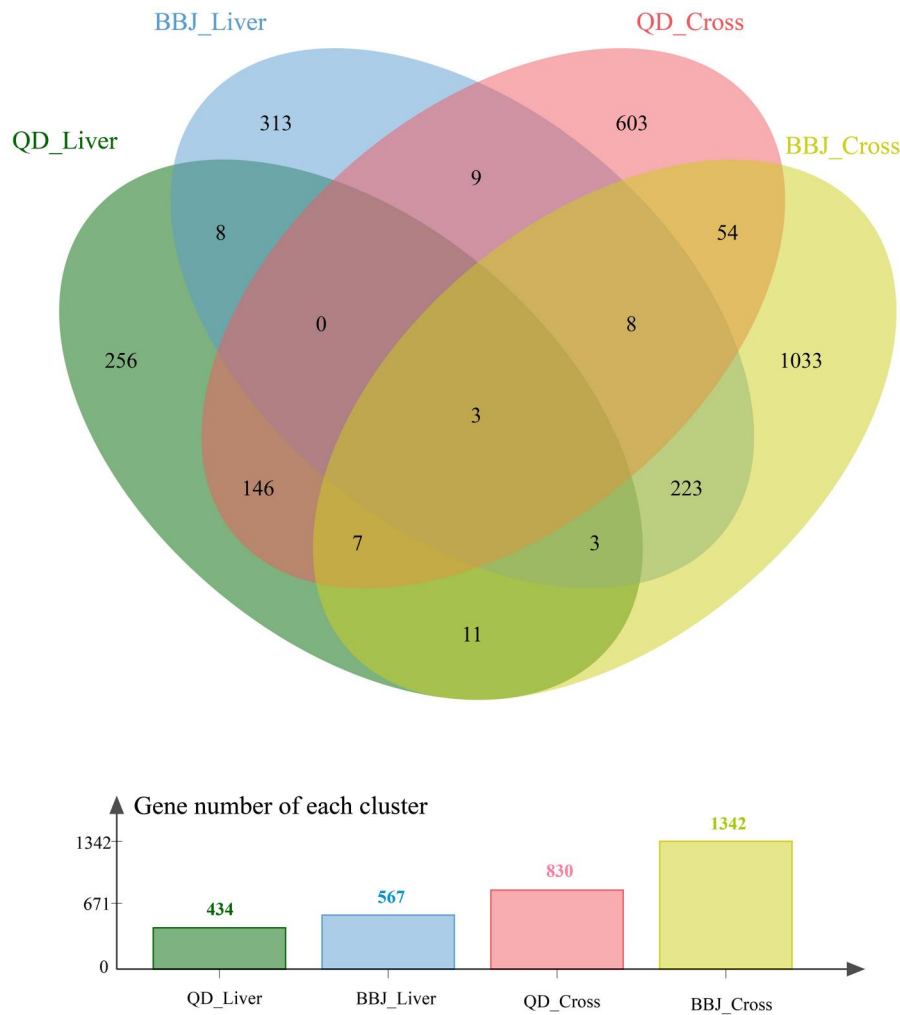


Figure 2. Venn diagram of the TWAS identified genes in different populations and approaches. Different colors represent the number of genes with $P < 0.05$ in different clusters. QD = Qidong cohort of Chinese, BBJ = Biobank Japan, Liver = liver single-tissue TWAS analysis (SPrediXcan), Cross = cross-tissue TWAS analysis (SMultiXcan).

18 significant clusters primarily engaged in immune response and transcription regulation processes, such as *Staphylococcus aureus* infection, initial triggering of complement, and over-expressing an important regulator of the cell cycle, Cyclin D1 (Figure 3).

Subsequently, we utilized TCGA data from UALCAN to compare the mRNA expression level of 85 selected protein-coding genes in HCC tumor tissue ($n = 371$) and normal tissues ($n = 50$). Our analysis revealed that 50 genes, including *NUAK2*, *ATP6V1G2*, *HLA-DRB5*, *ATF6B*, *FAM96B*, and *HSPA5*, exhibited significantly higher expression in HCC tumor tissues than in normal tissues, whereas 11 genes, such as *CAT*, *PMAIP1*, and *RXRG*, displayed significantly lower expression in HCC tumor tissue than in normal tissues ($P < 0.01$; Table 2).

We then employed GEPIA2 tool to perform survival analyses for the above 85 genes. In the overall survival (OS) analyses, we found that patients with high expression levels of *NUAK2*, *HSPA5*, *FKBPL*, *OBSCN*, *EGF*, *CSRNP2*, *SLC25A39*, *STXBP4*, *PMAIP1*, *HCG20*, and *CYP21A2* had worse prognoses than those with low expression levels of these 11 genes. Conversely, patients with high expression levels of *PKD1L3*, *CAT*, *CCDC38*, *CFB*, and *ITGAD* had better prognoses than those with low expression levels of these five genes (Table 2; $P < 0.05$). In the disease-free survival (DFS)

analyses, patients with low expression of 10 genes (*FAM96B*, *FKBPL*, *UNG*, *SLC25A39*, *STXBP4*, *EGFL8*, *MICB*, *POLRMT*, *MPHOSPH10*, and *RABL2A*) and high expression of three genes (*PKD1L3*, *GP1BA*, and *RXRG*) exhibited better prognoses compared with other patients (Table 2).

Finally, we assessed the dependency of the 85 genes in 23 HCC cell lines for both CRISPR-Cas9 and RNAi data in the DepMap database (Supplementary Table 4). We employed the gene effect scores to determine the gene's dependency on a specific cell line or tissue. Negative gene effect scores indicated that the cell line's growth would slow down after experimental manipulation of a specific gene, with more negative scores indicating stronger dependency on the cell line. We set the cut-off value at -0.5 according to the DepMap website (<https://forum.depmap.org/t/depmap-genetic-dependencies-faq/131>). The results revealed that for *FAM96B*, *HSPA5*, *XAB2*, *MPHOSPH10*, and *GRPEL*, 20 out of 23 (86.96%) cell lines fell below the cut-off value (-0.5) for the gene effect score in CRISPR-Cas9 data (Table 2). This suggests that HCC cells exhibited a high sensitivity to these five genes, and the growth of cancer cells was significantly constrained following CRISPR-Cas9 editing, highlighting their potential as promising drug targets for HCC.

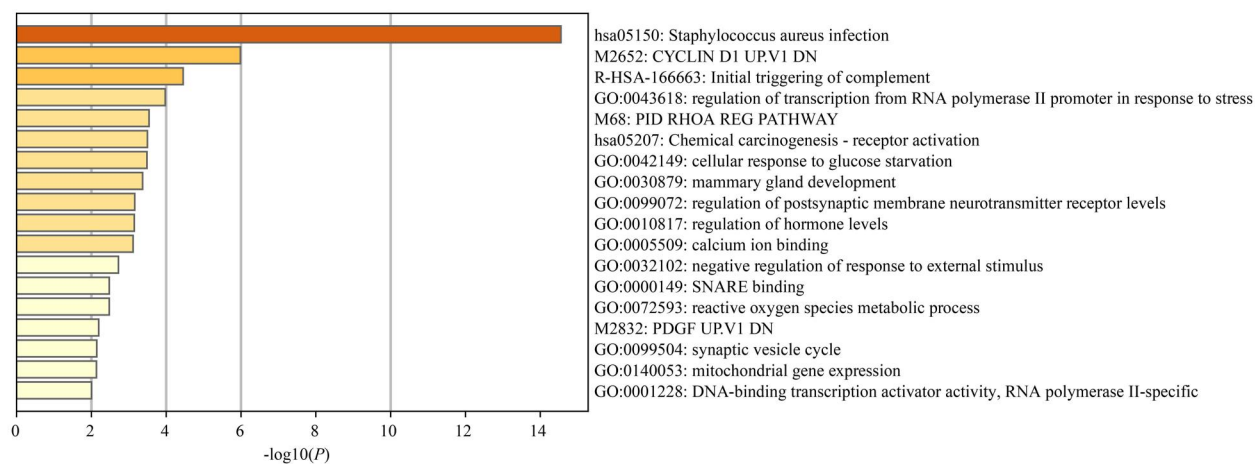


Figure 3. Visualization of the top statistically enriched terms of 85 HCC-associated genes identified by TWAS. The functional group terms were clustered based on their kappa score (≥ 0.3). The size of the bar represents the significance of term enrichment, and the darker the color, the more significant it is.

Discussion

Utilizing large-scale GWAS data of our cohort of Chinese CHB patients with HCC and BBJ HCC consortium, we conducted the TWAS analysis in East Asian population to explore novel susceptibility loci and putative causally relevant genes for HCC risk. In our current study, we identified 22 putative susceptibility genes for HCC after multiple testing corrections. Notably, 13 of these genes were located in the regions previously unexplored in GWASs. Furthermore, we conducted an overlapped analysis of TWAS-identified genes with a significance level of $P < 0.05$. Ultimately, 83 genes were identified in both Chinese and Japanese populations. Moreover, our functional annotations revealed some potentially pathogenic genes that may serve as underlying risk factors for HCC. Specifically, *FAM96B* at 16q22.1, *HSPA5* at 9q33.3, *POLRMT* at 19p13.3, *MPHOSPH10* at 2p13.3, and *RABL2A* at 2q14.1 displayed consistent effect directions in various functional aspects.

Our findings provide further support for several genes that have been previously implicated in GWASs or other researches. Earlier studies have identified 31 susceptibility loci associated with hepatitis virus-related HCC, nine of which were located in the major histocompatibility complex (MHC) region [12, 21–24, 26], suggesting the pivotal role of the anti-infectious immunity in the progression of HCC. In our TWAS results, we reported that eight of the 22 significant genes (*ATF6B*, *C4B*, *HLA-DPA1*, *BTNL2*, *NOTCH4*, *HLA-DPB1*, *FKBP1*, and *HLA-DRB5*) were situated within the MHC region, indicating a potential impact on the pathogenesis of HCC through the immune processes related to viral infection. Additionally, six of the 22 genes (*ARNT*, *LTB4R*, *IFT57*, *CTBP1*, *OBSCN*, *EGF*) have previously been shown to be associated with HCC through bioinformatics methods or experimental studies [27–33].

Through a comprehensive TWAS across diverse populations and using different TWAS analysis methods, we identified 83 susceptible genes that exhibited overlap in at least one TWAS approach within the two populations. Among these, compelling evidence supported the association of three genes (*NUAK2*, *HLA-DQA1*, and *ATP6V1G2*) with the risk of HCC, as they were consistently validated in both two populations using distinct TWAS analysis approaches. Notably, variants such as rs9272105 in *HLA-DQA1* have been proven to be associated with the susceptibility of HCC in previous report of GWAS [21]. *NUAK2* functions as a

stress response kinase involved in multiple processes, including cellular response to glucose starvation, negative regulation of apoptotic process, and protein phosphorylation [34]. In HCC, it serves as an essential mediator of liver proliferation and tumorigenesis by targeting Hippo/YAP signaling pathway [35]. Known as ATPase H⁺ Transporting V1 Subunit G2, *ATP6V1G2* encodes an enzyme transporter V-ATPase, which regulates the acidification of intracellular compartments in eukaryotic cells [36]. A previous study indicated that high expression of the ATPase-V0a3 subunit (*TCIRG1*) was observed in HCC and promoted tumor cell growth and proliferation [37, 38]. Although *ATP6V1G2* has been proven to be a risk gene for colorectal cancer [39] and glioma [38], there is currently no evidence linking *ATP6V1G2* to HCC progression.

In our analysis, we found *C4B* (6p21.33) may be the potential candidate risk gene in HCC. *C4B* encodes the basic form of complement factor 4, and together with the *C4A* gene, is part of the classical activation pathway [40]. *C4A/C4B* up-regulation is thought to contribute to HCC development through inflammatory and immunosuppressive mechanisms [41]. In the context of HCC, the specific role of the *C4B* gene may be related to its involvement in the immune response against cancer cells. But the exact role of *C4B* in HCC can vary and may be influenced by various factors, including the specific characteristics of the tumor and the individual's immune response. We need further research to fully understand the precise mechanisms by which genes may influence the HCC.

The pathway enrichment analysis revealed several candidate pathways for HCC, some of which have been reported to be directly or indirectly involved in HCC development. Notably, the cell cycle, immune response, complement reactions, metabolic processes, transcription factor activity, and the oxidative stress pathways have been widely recognized as significant contributors to HCC [42–45]. Additionally, researchers have provided evidence of the enrichment of the *Staphylococcus aureus* infection pathway in HCC tumor samples [46].

Functional annotation revealed that five genes (*FAM96B*, *HSPA5*, *POLRMT*, *MPHOSPH10*, and *RABL2A*) consistently displayed significant effects across various methods. For instance, their expression was significantly higher in HCC tissues than in normal controls. Patients with high expression of these genes experienced shorter survival times and a poorer prognosis, compared with others. Furthermore, most of the HCC cell lines demonstrated sensitivity to these five genes. And their knockout

Table 2. Functional examination of the TWAS-identified genes associated with HCC

Gene symbol	Region	Whether in new loci	Meta-analysis		Differential expression analysis		Survival analysis			Number of gene effect score < -0.5	
			TWAS Z	TWAS P	P value	Expression in HCC compared to control ^a	OS; HR (P)	DFS; HR (P)	High expression and survival time ^b	CRISPR	RNAi
NUAK2	1q32.1	Yes	2.86	4.23 × 10 ⁻³	4.50 × 10 ⁻⁶	Up	1.50 (0.035)	1.20 (0.160)	Shorter	0	0
HLA-DOA1	6p21.32	No	3.27	1.07 × 10 ⁻³	7.82 × 10 ⁻²	Down	0.75 (0.110)	0.74 (0.053)	NA	0	/
ATP6V1G2	6p21.33	No	-3.27	1.06 × 10 ⁻³	4.67 × 10 ⁻¹⁰	Up	1.10 (0.700)	0.95 (0.260)	NA	0	0
HLA-DRB5	6p21.32	No	3.01	2.58 × 10 ⁻³	2.30 × 10 ⁻³	Up	0.84 (0.320)	0.84 (0.240)	NA	0	1
HLA-DRA	6p21.32	No	1.65	9.96 × 10 ⁻²	2.09 × 10 ⁻²	Down	0.93 (0.690)	0.076 (0.079)	NA	0	0
FAM96B	16q22.1	Yes	3.18	1.47 × 10 ⁻³	1.62 × 10 ⁻¹²	Up	1.30 (0.120)	1.40 (0.018)	Shorter	20	0
UGT2B17	4q13.2	Yes	-1.61	0.110	NA	NA	0.90 (0.560)	1.10 (0.660)	NA	0	0
VTT1A	10q25.2	No	-2.58	9.75 × 10 ⁻³	<1 × 10 ⁻¹²	Up	1.20 (0.420)	1.20 (0.280)	NA	0	0
LTBP3	11q13.1	Yes	-1.62	0.110	1.43 × 10 ⁻⁹	Up	1.10 (0.680)	0.80 (0.160)	NA	1	0
SNX27	1q21.3	Yes	-2.21	0.027	<1 × 10 ⁻¹²	Up	1.30 (0.140)	1.20 (0.310)	NA	0	0
KLHL12	1q32.1	Yes	-2.07	0.038	<1 × 10 ⁻¹²	Up	1.40 (0.055)	1.10 (0.540)	NA	0	0
HSPA5	9q33.3	Yes	0.43	0.660	<1 × 10 ⁻¹²	Up	1.50 (0.017)	1.10 (0.680)	Shorter	20	1
AGPAT1	6p21.32	No	-1.88	0.061	1.62 × 10 ⁻¹²	Up	1.30 (0.130)	1.30 (0.120)	NA	0	0
PRMT7	16q22.1	Yes	-0.59	0.560	1.62 × 10 ⁻¹²	Up	1.00 (0.810)	1.10 (0.370)	NA	1	1
PKDIL3	16q22.2	Yes	2.28	0.023	8.31 × 10 ⁻¹	Down	0.61 (0.020)	0.66 (0.025)	Longer	/	0
CAT	11p13	Yes	2.79	5.32 × 10 ⁻³	1.90 × 10 ⁻¹²	Down	0.63 (0.011)	0.84 (0.250)	Longer	0	0
DUSP3	17q21.31	Yes	-2.11	0.035	<1 × 10 ⁻¹²	Up	1.20 (0.250)	1.30 (0.120)	NA	0	0
SOC56	18q22.2	Yes	-2.50	0.013	4.12 × 10 ⁻³	Down	1.40 (0.051)	1.20 (0.280)	NA	0	0
ACPP	3q22.1	Yes	-1.73	0.084	2.61 × 10 ⁻⁴	Down	0.97 (0.870)	0.87 (0.400)	NA	0	0
MRPL40	22q11.21	Yes	-1.47	0.141	1.62 × 10 ⁻¹²	Up	1.10 (0.690)	1.30 (0.140)	NA	5	4
ARHGEF5	7q35	Yes	-2.30	0.021	1.90 × 10 ⁻¹¹	Up	1.20 (0.430)	1.10 (0.400)	NA	4	1
ARNT	1q21.3	Yes	-1.80	0.072	1.62 × 10 ⁻¹²	Up	1.30 (0.150)	1.10 (0.580)	NA	0	0
ATF6B	6p21.32	No	-3.94	8.19 × 10 ⁻⁵	<1 × 10 ⁻¹²	Up	1.40 (0.060)	1.20 (0.230)	NA	0	0
C4B	6p21.33	No	4.81	1.49 × 10 ⁻⁶	NA	NA	0.85 (0.360)	0.84 (0.270)	NA	/	0
IFFO1	12p13.31	Yes	2.77	5.55 × 10 ⁻³	<1 × 10 ⁻¹²	Up	0.89 (0.520)	1.00 (0.750)	NA	0	0
FNDC10	1p36.33	Yes	2.91	3.56 × 10 ⁻³	NA	NA	NA	NA	NA	1	/
FKBP1	6p21.32	No	-3.10	1.93 × 10 ⁻³	<1 × 10 ⁻¹²	Up	1.50 (0.018)	1.70 (0.001)	Shorter	0	0
LTB4R	14q12	Yes	-3.53	4.23 × 10 ⁻⁴	1.62 × 10 ⁻¹²	Up	1.10 (0.450)	1.30 (0.056)	NA	2	0
HLA-DRB5	6p21.32	No	3.01	2.58 × 10 ⁻³	2.30 × 10 ⁻³	Up	0.84 (0.320)	0.84 (0.240)	NA	/	/
IFT57	3q13.12	Yes	3.40	0.001	7.13 × 10 ⁻⁸	Up	1.10 (0.470)	0.97 (0.850)	NA	1	0
NOTCH4	6p21.32	No	3.43	6.06 × 10 ⁻⁴	1.11 × 10 ⁻¹⁶	Up	0.90 (0.550)	1.20 (0.170)	NA	0	0
CTBP1	4p16.3	Yes	3.53	4.12 × 10 ⁻⁴	<1 × 10 ⁻¹²	Up	1.20 (0.390)	1.00 (0.890)	NA	0	0
OBSCN	1q42.13	No	-3.07	2.13 × 10 ⁻³	1.86 × 10 ⁻¹²	Up	1.50 (0.015)	1.10 (0.680)	Shorter	0	0
EGF	4q25	Yes	-3.30	9.54 × 10 ⁻⁴	4.72 × 10 ⁻¹²	Up	1.60 (0.015)	0.93 (0.630)	Shorter	0	0
HLA-DPA1	6p21.32	No	-4.06	4.82 × 10 ⁻⁵	1.97 × 10 ⁻¹	Down	0.93 (0.670)	0.80 (0.150)	NA	0	0
BTNL2	6p21.32	No	-3.95	7.96 × 10 ⁻⁵	NA	NA	NA	NA	NA	/	0
HLA-DPB1	6p21.32	No	-2.89	3.84 × 10 ⁻³	8.72 × 10 ⁻²	Down	1.00 (0.930)	0.80 (0.150)	NA	0	0
PHYH	10p13	Yes	-2.07	0.038	4.39 × 10 ⁻⁴	Down	0.96 (0.820)	1.10 (0.730)	NA	0	0
HABP2	10q25.3	Yes	-0.80	0.423	2.85 × 10 ⁻⁶	Down	0.86 (0.410)	0.89 (0.440)	NA	0	0
SSSCA1	11q13.1	Yes	-2.58	9.81 × 10 ⁻³	1.62 × 10 ⁻¹²	Up	1.10 (0.520)	1.10 (0.630)	NA	/	/
ARHGEF17	11q13.4	Yes	-1.02	0.306	1.79 × 10 ⁻¹²	Up	1.20 (0.240)	0.92 (0.580)	NA	0	0
H3F3C	12p11.21	Yes	-2.17	0.030	1.40 × 10 ⁻⁴	Up	0.90 (0.570)	0.96 (0.820)	NA	/	/
CSRNP2	12q13.12	Yes	2.30	0.022	1.62 × 10 ⁻¹²	Up	1.60 (0.008)	1.30 (0.065)	Shorter	0	/
TFCP2	12q13.12-q13.13	Yes	2.41	0.016	1.62 × 10 ⁻¹²	Up	1.40 (0.074)	1.20 (0.330)	NA	0	/
CCDC38	12q23.1	Yes	-1.13	0.260	1.16 × 10 ⁻¹	Down	0.68 (0.030)	0.81 (0.160)	Longer	0	0
UNG	12q24.11	Yes	1.98	0.047	<1 × 10 ⁻¹²	Up	1.40 (0.082)	1.70 (0.001)	Shorter	0	0

(continued)

Table 2. (continued)

Gene symbol	Region	Whether in new loci	Meta-analysis		Differential expression analysis		Survival analysis			Number of gene effect score < -0.5	
			TWAS Z	TWAS P	P value	Expression in HCC compared to control ^a	OS; HR (P)	DFS; HR (P)	High expression and survival time ^b	CRISPR	RNAi
GP1BA	17p13.2	Yes	2.53	0.011	1.99 × 10 ⁻¹	Down	0.73 (0.080)	0.70 (0.023)	Longer	0	0
KRT24	17q21.2	Yes	-0.70	0.483	NA	NA	NA	NA	NA	0	1
SLC25A39	17q21.31	Yes	-1.04	0.298	1.62 × 10 ⁻¹²	Up	1.90 (0.001)	1.50 (0.015)	Shorter	0	0
EPX	17q22	Yes	1.34	0.179	5.66 × 10 ⁻²	Up	NA	NA	NA	0	0
STXBP4	17q22	Yes	-0.61	0.543	1.62 × 10 ⁻¹²	Up	1.50 (0.018)	1.40 (0.018)	Shorter	0	0
PMAIP1	18q21.32	Yes	1.96	0.050	1.32 × 10 ⁻³	Down	1.50 (0.020)	1.30 (0.120)	Shorter	0	0
XAB2	19p13.2	Yes	-1.00	0.315	1.62 × 10 ⁻¹²	Up	1.10 (0.740)	1.20 (0.350)	NA	20	16
HPCAL4	1p34.2	Yes	-2.12	0.034	2.19 × 10 ⁻¹	Down	NA	NA	NA	0	0
C1QC	1p36.12	Yes	-2.14	0.033	7.74 × 10 ⁻²	Down	1.20 (0.320)	0.86 (0.340)	NA	0	0
WFDC13	20q13.11	Yes	-2.12	0.034	3.20 × 10 ⁻¹	NA	NA	NA	NA	0	0
SHROOM3	4q21.1	Yes	-3.26	1.10 × 10 ⁻³	3.12 × 10 ⁻³	Down	0.95 (0.770)	1.10 (0.720)	NA	0	0
NR3C1	5q31.3	Yes	-2.39	0.017	9.78 × 10 ⁻⁵	Up	1.10 (0.780)	1.30 (0.090)	NA	0	0
PCDHB1	5q31.3	Yes	0.57	0.570	3.70 × 10 ⁻²	NA	NA	NA	NA	0	0
HCG20	6p21	Yes	-0.90	0.369	NA	NA	1.50 (0.021)	1.10 (0.480)	Shorter	/	/
PBX2	6p21.32	No	-3.10	1.94 × 10 ⁻³	1.62 × 10 ⁻¹²	Up	1.10 (0.460)	1.10 (0.480)	NA	1	0
EGFL8	6p21.32	No	1.95	0.051	<1 × 10 ⁻¹²	Up	0.85 (0.370)	1.40 (0.031)	Shorter	0	0
HLA-DRB1	6p21.32	No	1.65	0.099	6.62 × 10 ⁻⁴	Up	0.84 (0.310)	0.87 (0.380)	NA	0	0
CFB	6p21.33	No	1.93	0.054	4.59 × 10 ⁻²	Down	0.66 (0.020)	0.79 (0.130)	Longer	0	0
CYP21A2	6p21.33	No	-3.25	1.17 × 10 ⁻³	5.33 × 10 ⁻¹³	Up	1.50 (0.037)	1.10 (0.670)	Shorter	0	1
DDR1	6p21.33	No	2.35	0.019	1.02 × 10 ⁻⁹	Up	1.20 (0.220)	1.10 (0.530)	NA	0	0
MICB	6p21.33	No	-2.59	9.51 × 10 ⁻³	1.62 × 10 ⁻¹²	Up	1.30 (0.130)	1.60 (0.002)	Shorter	0	0
MRPS18B	6p21.33	No	1.29	0.199	<1 × 10 ⁻¹²	Up	1.20 (0.250)	1.00 (1.000)	NA	10	0
NFKBIL1	6p21.33	No	1.08	0.281	<1 × 10 ⁻¹²	Up	1.20 (0.270)	1.40 (0.470)	NA	0	0
SIM1	6q16.3	Yes	2.32	0.021	2.66 × 10 ⁻³	Down	0.91 (0.600)	0.91 (0.570)	NA	0	0
GRB10	7p12.1	Yes	1.35	0.175	<1 × 10 ⁻¹²	Up	1.10 (0.610)	0.94 (0.700)	NA	0	0
GRID2IP	7p22.1	Yes	-0.23	0.822	5.83 × 10 ⁻⁴	Down	0.90 (0.640)	0.93 (0.700)	NA	0	/
DEFA4	8p23.1	Yes	-2.06	0.040	7.69 × 10 ⁻¹	NA	NA	NA	NA	0	0
FAM86B2	8p23.1	Yes	-2.38	0.017	4.24 × 10 ⁻³	Down	1.10 (0.590)	1.30 (0.085)	NA	0	/
RABEPK	9q33.3	Yes	0.38	0.707	2.74 × 10 ⁻¹⁴	Up	0.91 (0.600)	1.00 (0.830)	NA	0	0
CCDC62	12q24.31	Yes	3.32	8.97 × 10 ⁻⁴	4.62 × 10 ⁻⁴	Up	1.20 (0.390)	1.20 (0.260)	NA	0	0
HCAR2	12q24.31	Yes	3.16	1.57 × 10 ⁻³	7.90 × 10 ⁻²	Down	1.20 (0.350)	0.94 (0.710)	NA	0	0
ITGAD	16p11.2	Yes	1.14	0.254	6.20 × 10 ⁻¹	Down	0.70 (0.045)	0.75 (0.061)	Longer	0	0
PRRT2	16p11.2	Yes	-0.74	0.461	<1 × 10 ⁻¹²	Up	1.10 (0.530)	1.20 (0.240)	NA	0	0
POLRMT	19p13.3	Yes	2.73	6.33 × 10 ⁻³	1.62 × 10 ⁻¹²	Up	1.30 (0.120)	1.40 (0.032)	Shorter	9	/
XARG	1q23.3	Yes	1.04	0.300	2.31 × 10 ⁻⁵	Down	0.77 (0.150)	0.68 (0.013)	Longer	0	1
MPHOSPH10	2p13.3	Yes	0.71	0.480	<1 × 10 ⁻¹²	Up	1.20 (0.300)	1.40 (0.027)	Shorter	20	6
SH3YL1	2p25.3	Yes	2.74	6.24 × 10 ⁻³	8.80 × 10 ⁻²	Down	1.00 (0.830)	0.87 (0.380)	NA	0	0
RABL2A	2q14.1	Yes	3.05	2.27 × 10 ⁻³	1.62 × 10 ⁻¹²	Up	1.40 (0.079)	1.60 (0.004)	Shorter	4	1
GRPEL1	4p16.1	Yes	1.02	0.306	1.90 × 10 ⁻¹	Down	1.00 (1.000)	0.83 (0.220)	NA	20	6

TWAS = transcriptome-wide association study, OS = overall survival, DFS = disease-free survival, HR (P) = hazard ratio (P-value), RNAi = RNA interference.

^a "Up" means that the expression levels of these genes were significantly up-regulated in hepatocellular carcinoma tissues compared to normal tissues; "Down" means the genes were down-regulated in HCC tissues.

^b "shorter" means that the patients with high miRNA expression levels of these genes had shorter survival time compared with those with low expression level; "longer" means that patients with high expression levels of genes had longer survival time.

led to a restriction in the growth of HCC cells, highlighting the potential role of these five genes as oncogenes (Table 2). FAM96B, also called cytosolic iron-sulfur assembly component 2B (CIAO2B), is a highly conserved homologous protein belonging to the MIP18 family. An important paralog of FAM96B is FAM96A/CIAO2A. FAM96A and FAM96B participate in regulating tumor cell apoptosis by interacting with different proteins [47]. Some studies have shown that FAM96A can act as a tumor suppressor gene in various human cancers, including liver cancer [48]. But there is currently no detailed research report on the functional role of FAM96B in tumor progression at present. In one Chinese research, FAM96B expression was found to be down-regulated in HCC [47]. But in our study, we observed an up-regulation of FAM96B expression in TCGA-LIHC datasets. Further *In vitro/In vivo* experiments are needed to elucidate the molecular regulatory mechanisms and functional roles of FAM96B in the occurrence and development of HCC in different populations. Located in the lumen of the endoplasmic reticulum, HSPA5 is a member of the heat shock protein 70 (HSP70) family, and it is involved in endoplasmic reticulum protein folding and assembly. HSPA5 plays a role in endoplasmic reticulum stress and is implicated in various human diseases, including virus infection and tumorigenesis [49, 50]. In HCC, HSPA5 interacts with CD5L, a soluble scavenger cysteine-rich protein that modulates inflammatory responses, and promotes the proliferation of liver cancer cells, playing an antiapoptotic role [51]. In both Hispanic and Asian populations, HSPA5 could be the potential diagnostic biomarkers for HCC patients [52]. As an essential primer synthesis to initiate mitochondrial DNA (mtDNA) replication, POLRMT play a key role in the replication of mammalian mtDNA and is part of a transcriptional mechanism that controls the switch between primer formation for mtDNA replication and mitochondrial gene expression [53]. However, there is currently no direct evidence or functional experiments conducted on HCC samples. The role of POLRMT in HCC deserves further investigation. MPHOSPH10 is a M-Phase Phosphoprotein involved in rRNA processing in the nucleus and cytosol. Several GWASs for neuropsychiatric diseases have identified several SNPs on MPHOSPH10 [54, 55]. But the role of MPHOSPH10 in HCC has not been experimentally confirmed and requires further exploration. The RABL2A gene is a member of the RAS oncogene family-like 2 and plays an essential role in the proper function of flagella and cilia [56]. For HCC, one study illustrated that the expression of RABL2A was significantly up-regulated in sorafenib-resistant HCC cells, suggesting the involvement of the RABL2A–CCDC34 axis in mediating p38/MAPK and JNK/MAPK signaling pathways, thereby contributing to acquired sorafenib resistance in HCC [57].

It should be noted that there are several limitations in this study. Firstly, of the 22 genes with significant threshold values, the genes identified in the Chinese population were not validated in the Japanese cohort, even with a nominal significance. This discrepancy may be attributed the differing primary causes of HCC in the two populations. The main cause of HCC in China is hepatitis B viral (HBV) infection, while in Japan it is the hepatitis C viral (HCV) infection [58, 59]. Secondly, our HCC GWAS data were both derived for Asian populations, but the TWAS reference weights were based on European populations. To validate our results, it is essential to establish TWAS reference weights specifically tailored to Asian populations. Thirdly, though the TWAS results in Chinese and Japanese found three overlapping genes using different methods, these associations were only nominally significant and may include false positives. Further investigation into the functions of these genes in HCC is warranted.

In summary, we conducted a comprehensive TWAS in diverse populations using different TWAS approaches, identified some novel candidate genes and pathways associated with HCC. These findings contribute to a better understanding of the genetic mechanisms underlying HCC. But still need further biological studies to confirm our findings in the future.

Conclusion

TWASs have the potential ability to uncover new candidate genes that go beyond the associations identified in GWASs. Our research offers a valuable genetic insight into the biology of HCC.

Supplementary data

Supplementary data is available at *Gastroenterology Report* online.

Authors' Contributions

H.C. and D.J. were involved in the study design and supervision. D.J., W.H., and J.Z. were involved in the acquisition and interpretation of data. J.Z., Y.L. and W.H. were involved in the analysis of data, interpretation of data. J.Z., W.H., and Q.Z. drafted the initial manuscript. J.Z., Q.Z. and H.C. revised the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China [grant number: 82272312]; the 100 Top Talent Programs of Sun Yat-sen University [grant number: 58000–12230029]; and the Shenzhen-Hong Kong-Macao Science and Technology Project (Category C project) [grant number: SGDX20220530111403024].

Acknowledgements

This study utilized the publicly available data for analysis. The authors thank everyone at their institution who helped with this study.

Conflicts of interest

The authors declare that there is no conflict of interest in this study.

References

1. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet* 2018; **391**:1301–14.
2. Singal AG, Lampertico P, Nahon P. Epidemiology and surveillance for hepatocellular carcinoma: new trends. *J Hepatol* 2020; **72**:250–61.
3. Sollis E, Mosaku A, Abid A et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Res* 2023; **51**:D977–D985.
4. Caruso S, O'Brien DR, Cleary SP et al. Genetics of hepatocellular carcinoma: approaches to explore molecular diversity. *Hepatology* 2021; **73** Suppl 1:14–26.
5. Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. *Am J Hum Genet* 2018; **102**:717–30.

6. Li B, Ritchie MD. From GWAS to gene: transcriptome-wide association studies and other methods to functionally understand GWAS discoveries. *Front Genet* 2021;**12**:713230.
7. Liu D, Zhou D, Sun Y et al. A transcriptome-wide association study identifies candidate susceptibility genes for pancreatic cancer risk. *Cancer Res* 2020;**80**:4346–54.
8. Liu D, Zhu J, Zhou D et al. A transcriptome-wide association study identifies novel candidate susceptibility genes for prostate cancer risk. *Int J Cancer* 2022;**150**:80–90.
9. Zhong J, Jermusyk A, Wu L et al. A transcriptome-wide association study identifies novel candidate susceptibility genes for pancreatic cancer. *J Natl Cancer Inst* 2020;**112**:1003–12.
10. Guo X, Lin W, Wen W et al. Identifying novel susceptibility genes for colorectal cancer risk from a transcriptome-wide association study of 125,478 subjects. *Gastroenterology* 2021;**160**:1164–78.e6.
11. Lu M, Zhang Y, Yang F et al. TWAS Atlas: a curated knowledge-base of transcriptome-wide association studies. *Nucleic Acids Res* 2023;**51**:D1179–D1187.
12. Jiang DK, Sun J, Cao G et al. Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus–related hepatocellular carcinoma. *Nat Genet* 2013;**45**:72–5.
13. Ishigaki K, Akiyama M, Kanai M et al. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet* 2020;**52**:669–79.
14. Barbeira AN, Dickinson SP, Bonazzola R, GTEx Consortium et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 2018;**9**:1825.
15. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020;**369**:1318–30.
16. Barbeira AN, Bonazzola R, Gamazon ER, GTEx Consortium et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol* 2021;**22**:49.
17. Barbeira AN, Pividori M, Zheng J et al. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet* 2019;**15**:e1007889.
18. Chandrashekar DS, Karthikeyan SK, Korla PK et al. UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia* 2022;**25**:18–27.
19. Tang Z, Kang B, Li C et al. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* 2019;**47**:W556–W560.
20. Tsherniak A, Vazquez F, Montgomery PG et al. Defining a Cancer Dependency Map. *Cell* 2017;**170**:564–76.e16.
21. Li S, Qian J, Yang Y et al. GWAS identifies novel susceptibility loci on 6p21.32 and 21q21.3 for hepatocellular carcinoma in chronic hepatitis B virus carriers. *PLoS Genet* 2012;**8**:e1002791.
22. Kumar V, Kato N, Urabe Y et al. Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat Genet* 2011;**43**:455–8.
23. Clifford RJ, Zhang J, Meerzaman DM et al. Genetic variations at loci involved in the immune response are risk factors for hepatocellular carcinoma. *Hepatology* 2010;**52**:2034–43.
24. Sawai H, Nishida N, Khor SS 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.
25. Trépo E, Caruso S, Yang J et al. Common genetic variation in alcohol-related hepatocellular carcinoma: a case-control genome-wide association study. *Lancet Oncol* 2022;**23**:161–71.
26. Lee MH, Huang YH, Chen HY, REVEAL-HCV Cohort Study Group 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**:651–61.
27. Liang Y, Li WW, Yang BW et al. Aryl hydrocarbon receptor nuclear translocator is associated with tumor growth and progression of hepatocellular carcinoma. *Int J Cancer* 2012;**130**:1745–54.
28. Gesualdo A, Passerini F, D'Amore S et al. (13 August 2021) Genomic expression in peripheral blood mononuclear cells (Pbmcs) of patients with hepatocellular carcinoma (Hcc): effect of ablative treatment [PREPRINT (Version 1)]. *Research Square*, [10.21203/rs.3.rs-765972/v1](https://doi.org/10.21203/rs.3.rs-765972/v1).
29. Xu Y, Liu R. Analysis of the role of m6A and lncRNAs in prognosis and immunotherapy of hepatocellular carcinoma. *Heliyon* 2022;**8**:e10612.
30. Zhang XL, Huang CX, Zhang J et al. CtBP1 is involved in epithelial-mesenchymal transition and is a potential therapeutic target for hepatocellular carcinoma. *Oncol Rep* 2013;**30**:809–14.
31. Liu X, Sun B, Yao Y et al. Identification of copper metabolism and cuproptosis-related subtypes for predicting prognosis tumor microenvironment and drug candidates in hepatocellular carcinoma. *Front Immunol* 2022;**13**:996308.
32. Kim T, Issa D, Onyshchenko M. Analyzing TCGA Data to Identify Gene Mutations Linked to Hepatocellular Carcinoma in Asians. *Gastrointest Tumors* 2022;**9**:43–58.
33. Liu Z, Chen D, Ning F et al. EGF is highly expressed in hepatocellular carcinoma (HCC) and promotes motility of HCC cells via fibronectin. *J Cell Biochem* 2018;**119**:4170–83.
34. Bonnard C, Navaratnam N, Ghosh K et al. A loss-of-function NUA2 mutation in humans causes anencephaly due to impaired Hippo-YAP signaling. *J Exp Med* 2020;**217**:e20191561.
35. Yuan WC, Pepe-Mooney B, Galli GG et al. NUA2 is a critical YAP target in liver cancer. *Nat Commun* 2018;**9**:4834.
36. UniProt Consortium. UniProt: the universal protein knowledge-base in 2021. *Nucleic Acids Res* 2021;**49**:D480–9.
37. Yang HD, Eun JW, Lee KB et al. T-cell immune regulator 1 enhances metastasis in hepatocellular carcinoma. *Exp Mol Med* 2018;**50**:e420.
38. Qi C, Lei L, Hu J et al. Identification of a five-gene signature deriving from the vacuolar ATPase (V-ATPase) sub-classifies gliomas and decides prognoses and immune microenvironment alterations. *Cell Cycle* 2022;**21**:1294–315.
39. Consoli V, Fallica AN, Sorrenti V et al. Novel insights on ferroptosis modulation as potential strategy for cancer treatment: when nature kills. *Antioxid Redox Signal* (3 May 2023);**40**:40–85. [10.1089/ars.2022.0179](https://doi.org/10.1089/ars.2022.0179).
40. Qin X, Gao B. The complement system in liver diseases. *Cell Mol Immunol* 2006;**3**:333–40.
41. Barefoot ME, Varghese RS, Zhou Y et al. Multi-omic pathway and network analysis to identify biomarkers for hepatocellular carcinoma. *Annu Int Conf IEEE Eng Med Biol Soc* 2019;**2019**:1350–4.
42. Zucman-Rossi J, Villanueva A, Nault JC et al. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology* 2015;**149**:1226–39.e4.
43. Wu SY, Lan SH, Wu SR et al. Hepatocellular carcinoma–related cyclin D1 is selectively regulated by autophagy degradation system. *Hepatology* 2018;**68**:141–54.
44. Malik A, Thanekar U, Amarachintha S et al. “Complimenting the Complement”: mechanistic insights and opportunities for therapeutics in hepatocellular carcinoma. *Front Oncol* 2020;**10**:627701.

45. D'souza S, Lau KC, Coffin CS et al. Molecular mechanisms of viral hepatitis induced hepatocellular carcinoma. *World J Gastroenterol* 2020;**26**:5759–83.
46. Liang J, Lv J, Liu Z. Identification of dysfunctional biological pathways and their synergistic mechanism in hepatocellular carcinoma process. *Exp Mol Pathol* 2015;**98**:540–5.
47. Zhang DD, Sun XL, Liang ZY et al. FAM96A and FAM96B function as new tumor suppressor genes in breast cancer through regulation of the Wnt/ β -catenin signaling pathway. *Life Sci* 2022;**308**:120983.
48. Zhang MY, Wang JP. A multi-target protein of hTERTR-FAM96A presents significant anticancer potent in the treatment of hepatocellular carcinoma. *Tumour Biol* 2017;**39**:1010428317698341.
49. Wang Q, Ke S, Liu Z et al. HSPA5 promotes the proliferation, metastasis and regulates ferroptosis of bladder cancer. *Int J Mol Sci* 2023;**24**:5144.
50. Shi Y, Jin X, Wu S et al. Release of hepatitis B virions is positively regulated by glucose-regulated protein 78 through direct interaction with preS1. *J Med Virol* 2023;**95**:e28271.
51. Aran G, Sanjurjo L, Bárcena C et al. CD5L is upregulated in hepatocellular carcinoma and promotes liver cancer cell proliferation and antiapoptotic responses by binding to HSPA5 (GRP78). *FASEB J* 2018;**32**:3878–91.
52. Ma Y, Qiu C, Wang B et al. Autoantibody against tumor-associated antigens as diagnostic biomarkers in hispanic patients with hepatocellular carcinoma. *Cells* 2022;**11**:3227.
53. Nouws J, Goswami A, Bestwick SM et al. Mitochondrial ribosomal protein L12 is required for POLRMT stability and exists as two forms generated by alternative proteolysis during import. *J Biol Chem* 2015;**291**:989–97.
54. Pan H, Shen JY, Du JJ et al. Lack of association between DNMT3B polymorphisms and sporadic Parkinson's disease in a Han Chinese population. *Neurosci Bull* 2018;**34**:867–9.
55. Wilfert AB, Turner TN, Murali SC, SPARK Consortium et al. Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nat Genet* 2021;**53**:1125–34.
56. Ding X, Fragoza R, Singh P et al. Variants in RABL2A causing male infertility and ciliopathy. *Hum Mol Genet* 2020;**29**:3402–11.
57. Zhou M, Chen X, Bai H et al. RABL2A–CCDC34 axis promotes sorafenib resistance in hepatocellular carcinoma. *DNA Cell Biol* 2021;**40**:1418–27.
58. Ukawa S, Okada E, Nakamura K et al. Characteristics of patients with liver cancer in the BioBank Japan project. *J Epidemiol* 2017;**27**:S43–8.
59. Goh GBB, Chang PE, Tan CK. Changing epidemiology of hepatocellular carcinoma in Asia. *Best Pract Res Clin Gastroenterol* 2015;**29**:919–28.