Cancer Management and Research

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Genome-wide association pathway analysis to identify candidate single nucleotide polymorphisms and molecular pathways associated with TP53 expression status in HBV-related hepatocellular carcinoma

This article was published in the following Dove Press journal:
Cancer Management and Research

Xiwen Liao[1,*]
Long Yu[1,*]
Xiaoguang Liu[1,2]
Chuangye Han[1]
Tingdong Yu[1]
Wei Qin[1]
Chengkun Yang[1]
Guangzhi Zhu[1]
Hao Su[1]
Tao Peng[1]

[1]Department of Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, People's Republic of China; [2]Department of Hepatobiliary Surgery, Affiliated Hospital of Guangdong Medical University, Zhanjiang, People's Republic of China

*These authors contributed equally to this work

Correspondence: Tao Peng
Department of Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, Shuang Yong Road 6, Nanning 530021, Guangxi Zhuang Autonomous Region, China
Tel +86 771 535 6528
Fax +86 771 535 0031
Email pengtaogmu@163.com

Background: The aim of this investigation was to identify candidate single nucleotide polymorphisms (SNPs) and molecular pathways associated with tumor protein p53 (TP53) expression status in hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC), clarify their potential mechanisms, and generate SNP-to-gene to pathway hypothesis.

Materials and methods: Identify candidate Causal SNPs and Pathways (ICSNPathway) was used to perform pathway analysis based on the results of our previous genome-wide association study of TP53 expression status in 387 HBV-related HCC patients.

Results: Through the ICSNPathway analysis, we identified 18 candidate SNPs and 10 candidate pathways that are associated with TP53 expression status in HBV-related HCC. The strongest mechanism involved the modulation of major histocompatibility complex, class II, DP beta 1 (human leukocyte antigen [*HLA*]-*DPB1*-rs1042153), major histocompatibility complex, class II, DQ beta 1 (*HLA-DQB1*-rs1130399, *HLA-DQB1*-rs1049056, *HLA-DQB1*-rs1049059, and *HLA-DQB1*-rs1049060), and major histocompatibility complex, class II, DR beta 1 (*HLA-DRB1*-rs35445101). SNPs consequently affected regulatory roles in all the candidate pathways except hematopoietic cell lineage pathways. Association analysis using the GSE14520 data set, Gene Multiple Association Network Integration Algorithm, and Search Tool for the Retrieval of Interacting Genes/Proteins suggests that all genes of the candidate SNPs were associated with *TP53*. Survival analysis showed that the collagen type VI alpha 3 chain (*COL6A3*) rs111231885 and *COL6A3*-rs113155945 and *COL6A3* block 4 CC haplotypes with TP53 negative status may have protective effects in HBV-related HCC patients after hepatectomy.

Conclusion: Our pathway analysis identified 18 candidate SNPs and 10 candidate pathways that were associated with TP53 expression status in HBV-related HCC. Among these candidate SNPs, the genetic variation of *COL6A3* may be a potential prognostic biomarker of HBV-related HCC.

Keywords: TP53, hepatocellular carcinoma, hepatitis B virus, genome-wide association study, pathway analysis

Introduction

Liver cancer is the third leading cause of cancer death in China, with an age-standardized 5-year relative survival rate of 10.1%.[1,2] The majority of liver cancer cases are hepatocellular carcinoma (HCC).[3] A high prevalence of hepatitis B virus (HBV) infection and aflatoxin B1 exposure are the main factors of HCC in the Guangxi

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/CMAR.S163209

province of China.[4–6] Previous studies have demonstrated that the tumor protein p53 (*TP53*) mutation is frequently found in HBV-related HCC in patients of Guangxi province.[7–10] Therefore, in the Guangxi region, there is a representative population in which the associations between HBV infection and the *TP53* gene in HCC can be investigated.

Hepatocarcinogenesis is driven by the interaction of genetic and environmental factors.[11,12] Genome-wide association studies (GWAS) can be used to identify associations between specific single nucleotide polymorphisms (SNPs) and complex diseases or other traits.[13] The roles of the corresponding genes or proteins in the context of the pathway might be altered by trait-related SNPs. Identify candidate Causal SNPs and Pathways (ICSNPathway) is an analytical framework for the comprehensive interpretation of GWAS data by integrating linkage disequilibrium (LD) analysis, functional SNP annotation, and pathway-based analysis (PBA) and can be used to derive the mechanism hypothesis of SNP→gene→pathway(s) for complex disease studies, including cancer.[14] In addition, ICSNPathway also is a tool based on PBA algorithm, which is a method for secondary excavation of GWAS results based on prior biological knowledge on gene function and biological metabolic pathways.[14] By using the PBA algorithm, more information about the pathway and gene sets with same functions which are associated with the diseases or traits from GWAS results could be obtained.

Our previous study has identified several SNPs associated with TP53 expression status in HBV-related HCC in patients of Guangxi by using the GWAS approach.[15] In the present study, we further investigated candidate SNPs and molecular pathways associated with TP53 expression status in HBV-related HCC by using the ICSNPathway web server based on the result of our previous GWAS.

Materials and methods

Study population and GWAS data

Our study was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University with an ethics approval number of 2015 (KY-E-032).[15] Written informed consent was obtained from all the participants enrolled in the study. The primary GWAS data set was extracted from our previous study.[15] Clinicopathological characteristics and prognosis of the HBV-related HCC patients, genotyping, quality control, and GWAS analysis methods have been described and published in our earlier article.[15] A total of 403 patients with serum tests that were HBV surface antigen positive and newly diagnosed with HCC by pathological examination in the First Affiliated Hospital of Guangxi Medical University between 2001 and 2013 were included.[15] TP53 staining in HCC tumor tissues was detected by immunohistochemistry.[15] The SNPs were genotyped by an Illumina Human Exome BeadChip 12 v1-1 system (Illumina Inc, San Diego, CA, USA). Quality control standards were set as follows: samples were excluded if they had 1) an overall genotyping rate of <95%; 2) ambiguous gender; 3) genome-wide identity by-descent >0.1875; 4) outliers in principal component analysis (PCA) for ancestry and population stratification. SNPs had to meet the following criteria: 1) a call rate of >95%; 2) a Hardy–Weinberg equilibrium $P>1\times10^{-6}$; 3) a minor allele frequency >0.01.[15] PCA for ancestry and population stratification suggest that no or mild population stratification was found in the current study population, and similar results were observed in our previous study.[15] A total of 387 patients with 28,952 SNPs passed the quality control filters and were included in further investigations.

Identification of candidate SNPs and pathways

The ICSNPathway (http://icsnpathway.psych.ac.cn, accessed February 20, 2017) web server contains a two-stage analysis: 1) preselect candidate SNPs by LD analysis and functional SNP annotation based on the most significant SNPs and 2) annotate the biological mechanisms for the preselected candidate SNPs by using PBA.[14] A complete list of GWAS SNP *P*-values was input for ICSNPathway analysis. The parameters used in the ICSNPathway were 1) threshold to specify the most significant SNPs: *P*-value $<1\times10^{-2}$; 2) HapMap population: Han Chinese in Beijing, China; 3) LD cutoff: $r^2>0.8$; 4) distance for searching LD neighborhoods: 200 kb; 5) rule of mapping SNPs to genes: 500 kb upstream and downstream of gene; 6) pathway/gene set database: Kyoto Encyclopedia of Genes and Genomes; 7) number of genes in each pathway/gene set: minimum 5 and maximum 100; and 8) false discovery rate cutoff for PBA: 0.1.

Association analysis

Based on the results of the ICSNPathway analysis, haplotype analysis among the candidate SNPs was calculated using Haploview version 4.2 (Broad Institute of MIT and Harvard, Cambridge, MA, USA).[16] Regional LD plots of the candidate SNPs were generated by SNP Annotation and Proxy Search (SNAP) (http://archive.broadinstitute.org/mpg/snap, accessed February 20, 2017), a tool used for the identification and annotation of proxy SNPs using HapMap.[17] Genotype and haplotype distribution of the candidate SNPs in different

TP53 expression status groups were determined using a binary logistic regression model. Co-expression analysis of the *TP53* gene and the genes of candidate SNPs was performed using GSE14520, a Chinese HBV-related HCC mRNA expression chip data set obtained from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo, accessed February 20, 2017).[18,19] In order to eliminate the batch effect of the expression chip, we only included the Affymetrix HT Human Genome U133A Array data set (Thermo Fisher Scientific, Waltham, MA, USA) of GSE14520 in the co-expression analysis. Gene Multiple Association Network Integration Algorithm[20,21] (GeneMANIA; http://genemania.org, accessed February 20, 2017) and Search Tool for the Retrieval of Interacting Genes/Proteins[22,23] (STRING; http://string.embl.de/, accessed February 20, 2017) web servers were used for investigating the gene–gene and protein–protein interactions (PPIs) among genes of the candidate SNPs, respectively.

Survival analysis

We further analyzed the association of the candidate SNPs and clinical outcomes. The TP53 expression status and the candidate SNP interactions were analyzed using a joint effects survival analysis. In addition, we also analyzed haplotypes of the candidate SNPs.

Statistical analysis

Pearson correlation coefficient was used to assess co-expression correlation. The odds ratio (OR) and the corresponding 95% confidence interval (CI) of the binary logistic regression model were used to estimate the relative risk of TP53 expression status in HBV-related HCC. Univariate analysis between clinical features and overall survival (OS) were studied using the Kaplan–Meier method with the log-rank test. Cox proportional hazards regression analysis was used to calculate the crude and adjusted hazard ratio (HR) and 95% CI in univariate and multivariate analyses, with adjustment for age, gender, race, body mass index (BMI), smoking status, drinking status, Barcelona Clinic Liver Cancer (BCLC) stage, cirrhosis, radical resection, antiviral therapy, status of tumor capsule, regional invasion, and portal vein tumor thrombus (PVTT). A value of $P<0.05$ was considered statistically significant. All the statistical analyses were conducted with SPSS version 20.0 software (IBM Corporation, Armonk, NY, USA).

Results

Candidate SNPs and pathways

Using the *P*-values of the 28,952 GWAS SNPs as input data, the ICSNPathway identified 18 SNPs as candidate SNPs (Table 1) and 10 pathways as candidate pathways (Table 2) that were associated with TP53 expression status in HBV-related HCC. All of the candidate SNPs were non-synonymous coding SNPs, and the alteration of four SNPs (desmoglein 3 [*DSG3*]-rs16961975, keratin 35 (*KRT35)*-rs2071601, *KRT35*-rs743686, and keratin 36 (*KRT36*)-rs2301354) was deleterious. The strongest mechanism involved the modulation of major histocompatibility complex, class II, DP beta 1 (human leukocyte antigen [*HLA*]-*DPB1*-rs1042153), major histocompatibility complex, class II, DQ beta 1 (*HLA-DQB1*-rs1130399, *HLA-DQB1*-rs1049056, *HLA-DQB1*-rs1049059, and *HLA-DQB1*-rs1049060) and major histocompatibility complex, class II, DR beta 1 (*HLA-DRB1*-rs35445101), consequently affecting their regulatory roles in all the candidate pathways except the hematopoietic cell lineage pathway. The second strongest hypothetical biological mechanism was that SNPs of major histocompatibility complex, class I, C (*HLA-C*-rs1131096 and *HLA-C*-rs1130838) influence the regulatory role of cell adhesion molecules (CAMs), autoimmune thyroid disease, allograft rejection, antigen processing and presentation, type I diabetes mellitus, and graft-versus-host disease pathways. Other multiple mechanisms presented in Tables 1 and 2 indicated that alterations in candidate SNPs of collagen type VI alpha 3 chain (*COL6A3*)-rs111231885 and *COL6A3*-rs113155945, desmoglein 3 (*DSG3*)-rs16961975, keratin 32 (*KRT32*)-rs3744786), keratin 35 (*KRT35*)-rs2071601 and *KRT35*-rs743686, and keratin 36 (*KRT36*)-rs2301354 affect their regulatory roles in cell communication, whereas legumain (*LGMN*)-rs118128989 alterations affected the antigen processing and presentation pathway and interleukin 6 receptor (*IL6R*)-rs2228145, interleukin 7 receptor (*IL7R*)-rs6897932 alterations affected the hematopoietic cell lineage pathway.

Association analysis

In the output of candidate SNPs, *KRT32*-rs3744786 and *KRT35*-rs2071601 were not present in the original GWAS result. So, only 16 candidate SNPs and their corresponding genes were included in further association and survival analysis. Four haplotype blocks were detected in the haplotypes analysis (block 1 pairwise r^2=0.992, constituted by *HLA-C*-rs1130838 and *HLA-C*-rs1131096; block 2 pairwise r^2=0.121–1.0, constituted by *HLA-DRB1*-rs35445101, *HLA-DQB1*-rs1130399, *HLA-DQB1*-rs1049060, *HLA-DQB1*-rs1049059, and *HLA-DQB1*-rs1049056; block 3 pairwise r^2=0.973, constituted by *KRT35*-rs743686 and *KRT36*-rs2301354; block 4 pairwise r^2=0.755, constituted by *COL6A3*-rs111231885 and *COL6A3*-rs113155945;

Table 1 Candidate SNPs identified from ICSNPathway analysis

Candidate SNP	Functional class	Gene	Candidate pathway[a]	–log10(*P*)[b]	In LD with	r^2	D'	–log10(*P*)[c] in original GWAS
rs1042153	non_synonymous_coding	*HLA-DPB1*	1,2,3,4,5,7,9,10	2.45	rs1042153	–	–	2.45
rs1130399	non_synonymous_coding	*HLA-DQB1*	1,2,3,4,5,7,9,10	3.153	rs1130399	–	–	3.153
rs1049056	non_synonymous_coding	*HLA-DQB1*	1,2,3,4,5,7,9,10	2.433	rs1049056	–	–	2.433
rs1049059	non_synonymous_coding	*HLA-DQB1*	1,2,3,4,5,7,9,10	2.357	rs1049059	–	–	2.357
rs1049060	non_synonymous_coding	*HLA-DQB1*	1,2,3,4,5,7,9,10	2.467	rs1049060	–	–	2.467
rs35445101	non_synonymous_coding	*HLA-DRB1*	1,2,3,4,5,7,8,9,10	3.591	rs35445101	–	–	3.591
rs1131096	non_synonymous_coding	*HLA-C*	3,4,5,7,9,10	2.097	rs1131096	–	–	2.097
rs1130838	non_synonymous_coding	*HLA-C*	3,4,5,7,9,10	2.267	rs1130838	–	–	2.267
rs111231885	non_synonymous_coding	*COL6A3*	6	2.541	rs111231885	–	–	2.541
rs113155945	non_synonymous_coding	*COL6A3*	6	2.044	rs113155945	–	–	2.044
rs16961975	non_synonymous_coding (deleterious)	*DSG3*	6	2.12	rs16961975	–	–	2.12
rs3744786	non_synonymous_coding	*KRT32*	6	–	rs2301354	0.847	0.945	3.583
rs2071601	non_synonymous_coding (deleterious)	*KRT35*	6	–	rs2301354	0.901	1	3.583
rs743686	non_synonymous_coding (deleterious)	*KRT35*	6	3.209	rs743686	–	–	3.209
rs2301354	non_synonymous_coding (deleterious)	*KRT36*	6	3.583	rs2301354	–	–	3.583
rs118128989	non_synonymous_coding	*LGMN*	7	2.084	rs118128989	–	–	2.084
rs2228145	non_synonymous_coding	*IL6R*	8	2.081	rs2228145	–	–	2.081
rs6897932	non_synonymous_coding	*IL7R*	8	2.004	rs6897932	–	–	2.004

Notes: [a]The number indicates the index of pathways (listed in Table 2), which are ranked by their statistical significance (FDR). [b]–$\log_{10}(P)$ for candidate SNP in original GWAS. "–" denotes that this SNP is not represented in the original GWAS. [c]–$\log_{10}(P)$ for the SNP (which candidate SNP is in LD with) in original GWAS.

Abbreviations: SNP, single nucleotide polymorphism; ICSNPathway, Identify candidate Causal SNPs and Pathways; LD, linkage disequilibrium; GWAS, genome-wide association studies; *HLA-DPB1*, major histocompatibility complex, class II, DP beta 1; *HLA-DQB1*, major histocompatibility complex, class II, DQ beta 1; *HLA-DRB1*, major histocompatibility complex, class II, DR beta 1; *HLA-C*, major histocompatibility complex, class I, C; *COL6A3*, collagen type VI alpha 3 chain; *DSG3*, desmoglein 3; *KRT32*, keratin 32; *KRT35*, keratin 35; *KRT36*, keratin 36; *LGMN*, legumain; *IL6R*, interleukin 6 receptor; *IL7R*, interleukin 7 receptor; FDR, false discovery rate.

Table 2 Candidate pathways identified from ICSNPathway analysis

Index	Candidate pathway	Description	Nominal *P*-value	FDR
1	hsa05310	Asthma	0.005	0.026
2	hsa05322	Systemic lupus erythematosus	0.005	0.026
3	hsa04514	CAMs	0.034	0.056
4	hsa05320	Autoimmune thyroid disease	0.028	0.058
5	hsa05330	Allograft rejection	0.028	0.058
6	hsa01430	Cell communication	0.024	0.058
7	hsa04612	Antigen processing and presentation	0.041	0.06
8	hsa04640	Hematopoietic cell lineage	0.009	0.069
9	hsa04940	Type I diabetes mellitus	0.032	0.077
10	hsa05332	Graft-versus-host disease	0.032	0.077

Abbreviations: FDR, false discovery rate; CAMs, cell adhesion molecules; ICSNPathway, Identify candidate Causal SNPs and Pathways.

Figure 1A). Distribution of candidate SNPs in different TP53 expression status patients is shown in Table S1. After adjusting for age, gender, race, BMI, smoking status, drinking status, BCLC stage, cirrhosis, radical resection, antiviral therapy, status of tumor capsule, regional invasion, and PVTT, all the SNPs were significantly associated with TP53 expression status in HBV-related HCC. Association analysis in four haplotypes demonstrated that GT in block 1, AGTCC and GATCC/other haplotypes in block 2, GA/other haplotypes in block 3, and TT/other haplotypes in block 4 had significantly decreased risk of TP53 expression status in HBV-related HCC, compared to patients with TC, AGAGA, AG, and CC (Table 3), respectively.

Co-expression analysis in HBV-related HCC expression profile data set from the GSE14520 cohort revealed that the *TP53* gene has a significantly weak negative correlation with *HLA-C* (r=−0.322, P<0.001), *IL6R* (r=−0.132, P=0.007), and *KRT36* (r=−0.186, P<0.001), whereas it had a positive

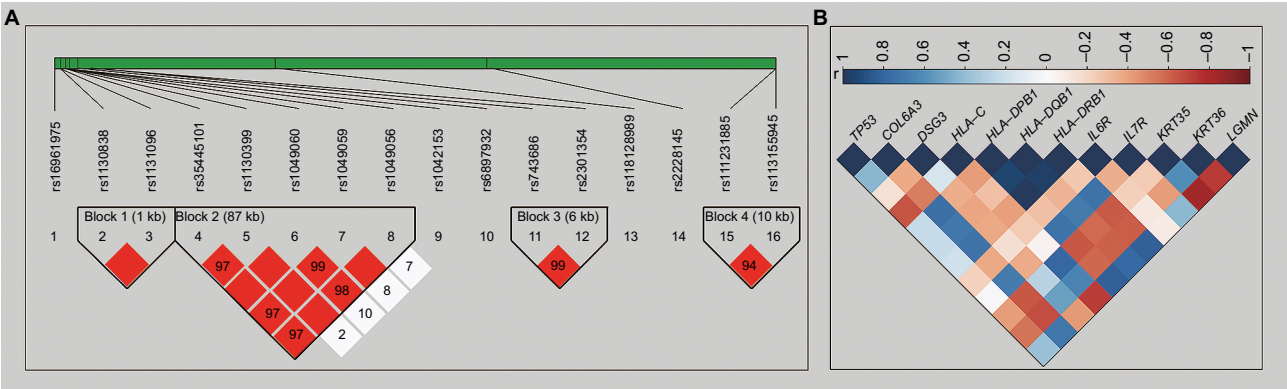

Figure 1 Haplotype association of the candidate SNPs and co-expression heat map for the *TP53* gene and corresponding genes of the candidate SNPs.
Notes: (**A**) Patterns of LD plots for 16 candidate SNPs. (**B**) Co-expression heat map between TP53 and corresponding genes of the candidate SNPs.
Abbreviations: SNP, single nucleotide polymorphism; TP53, tumor protein p53; LD, linkage disequilibrium.

Table 3 Haplotype distribution of the candidate SNPs in patients with different TP53 expression status

Haplotypes	TP53 negative (2n=308)	TP53 positive (2n=466)	Crude OR (95% CI)	Crude *P*-value	Adjusted OR (95% CI)	Adjusted *P*-value[a]
Block 1						
TC	229	387	1		1	
GT	79	79	0.592 (0.416–0.841)	0.003	0.622 (0.432–0.895)	0.011
Block 2						
AGAGA	118	238	1		1	
AGTCC	116	167	0.714 (0.516–0.987)	0.041	0.696 (0.494–0.980)	0.038
GATCC+other haplotypes	74	61	0.409 (0.273–0.612)	<0.001	0.395 (0.259–0.603)	<0.001
Block 3						
AG	156	291	1		1	
GA+other haplotypes	152	175	0.617 (0.461–0.826)	0.001	0.582 (0.427–0.792)	0.001
Block 4						
CC	279	446	1		1	
TT+other haplotypes	29	20	0.431 (0.239–0.777)	0.005	0.376 (0.201–0.703)	0.002

Notes: [a]Adjusted for age, gender, race, body mass index, smoking status, drinking status, BCLC stage, cirrhosis, radical resection, antiviral therapy, status of tumor capsule, regional invasion, and PVTT.
Abbreviations: SNP, single nucleotide polymorphism; TP53, tumor protein p53; BCLC, Barcelona Clinic Liver Cancer; PVTT, portal vein tumor thrombus.

correlation with *COL6A3* (r=0.238, P<0.001), *HLA-DPB1* (r=0.115, P=0.018), *HLA-DQB1* (r=0.176, P=0.0003), and *LGMN* (r=0.157, P=0.001) at the mRNA level in HBV-related HCC of GSE14520; the co-expression heat map is shown in Figure 1B. The remaining genes were not significantly correlated with *TP53* at the mRNA level. Gene and gene co-expression interaction networks constructed by GeneMANIA demonstrated that all the genes of the candidate SNPs exist in a complex gene–gene co-expression interaction network and are directly or indirectly associated with *TP53* (Figure 2A). In addition, PPIs determined experimentally and constructed by STRING showed that HLA-DQB1, HLA-C, KRT35, IL7R, and LGMN were associated with TP53 through ubiquitin C (UBC) (Figure 2B).

Detailed regional LD plots of the four haplotype blocks of the candidate SNPs were generated by SNAP. Regional LD plots for four SNPs of HLA-DQB1 in block 2 are shown in Figure 3A–D, whereas the regional LD plot of *HLA-DRB1*-rs35445101 in block 2 was not available on the SNAP website. Regional LD plots of block 3 (Figure 4A, B) and block 4 (Figure 4C, D) are shown in Figure 4, whereas the regional LD plot for block 1 of *HLA-C*-rs1131096 and *HLA-C*-rs1130838 was not available on the SNAP website. Regional LD plot for these SNPs indicated that there were

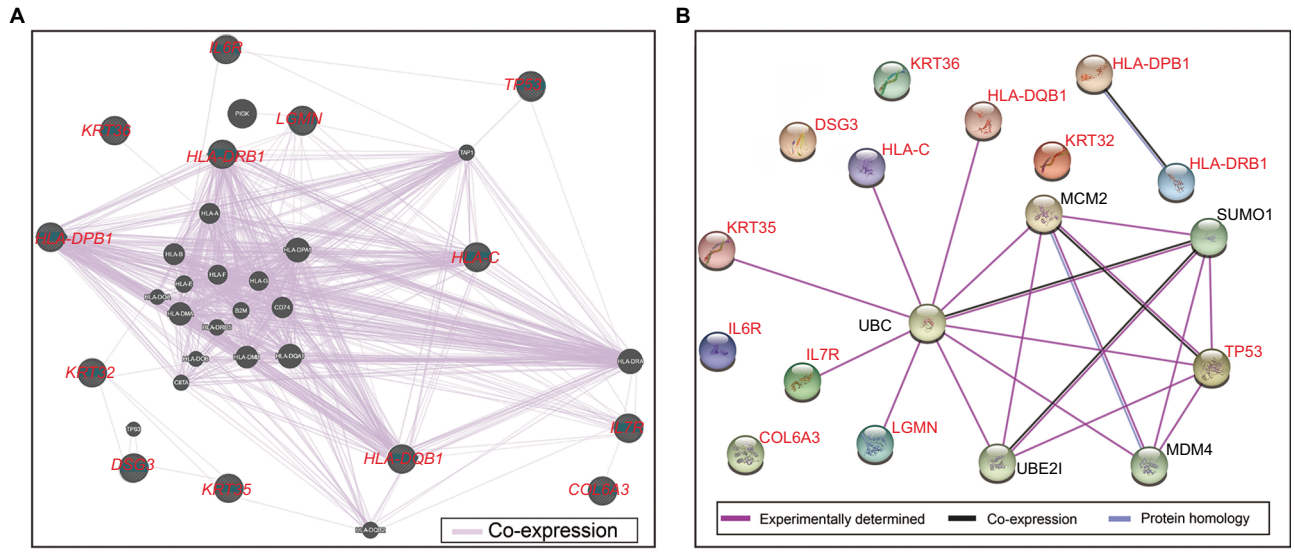

Figure 2 Gene–gene and protein–protein interaction networks.
Notes: (**A**) Gene–gene interaction networks constructed by GeneMANIA. (**B**) Protein–protein interaction networks constructed by STRING.
Abbreviations: GeneMANIA, Gene Multiple Association Network Integration Algorithm; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; TP53, tumor protein p53; UBC, ubiquitin C; *HLA-DPB1*, major histocompatibility complex, class II, DP beta 1; *HLA-DQB1*, major histocompatibility complex, class II, DQ beta 1; *HLA-DRB1*, major histocompatibility complex, class II, DR beta 1; *HLA-C*, major histocompatibility complex, class I, C; *COL6A3*, collagen type VI alpha 3 chain; *DSG3*, desmoglein 3; *KRT32*, keratin 32; *KRT35*, keratin 35; *KRT36*, keratin 36; *LGMN*, legumain; *IL6R*, interleukin 6 receptor; *IL7R*, interleukin 7 receptor.

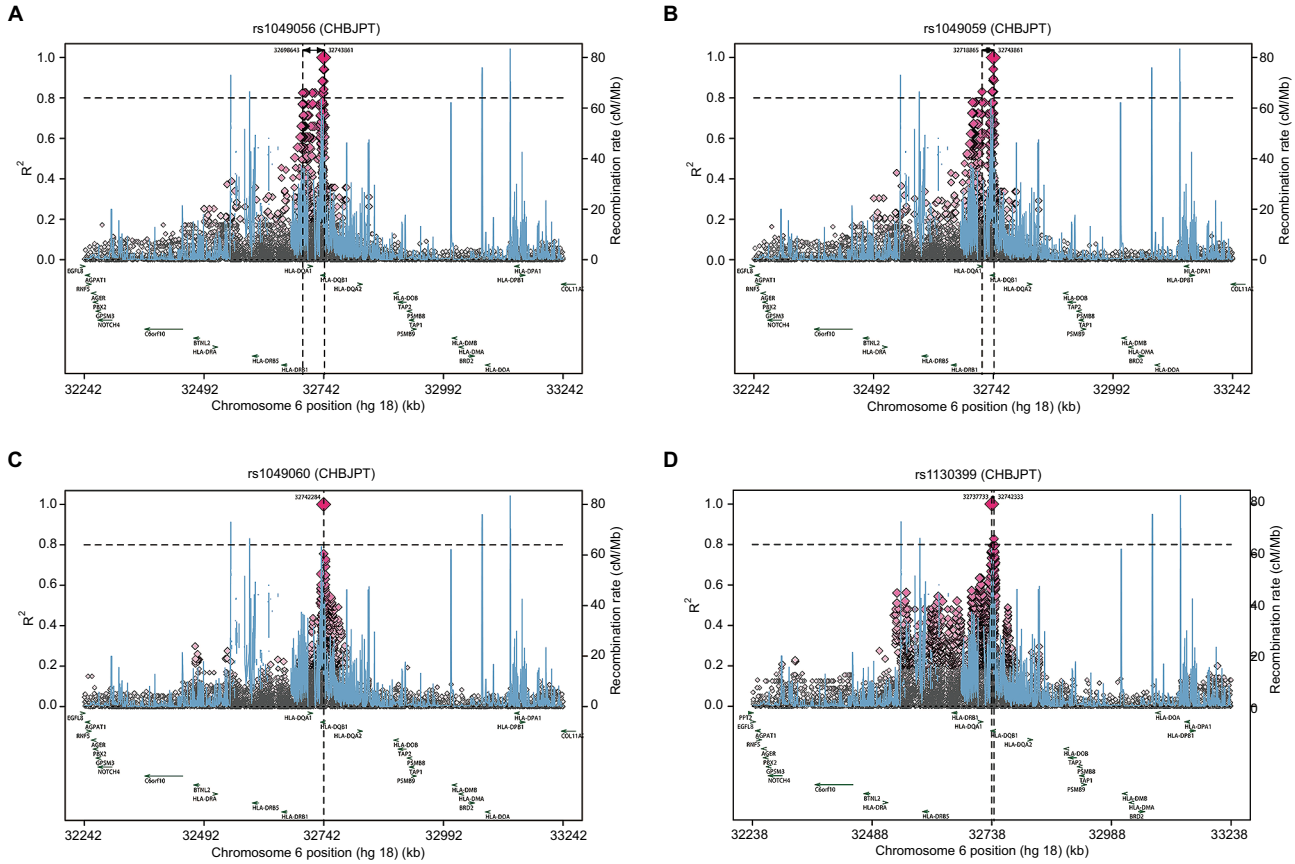

Figure 3 Regional LD plots of block 2 (*HLA-DQB1*).
Notes: Regional LD plots of (**A**) *HLA-DQB1*-rs1049056, (**B**) *HLA-DQB1*-rs1049059, (**C**) *HLA-DQB1*-rs1049060, and (**D**) *HLA-DQB1*-rs1130399.
Abbreviations: LD, linkage disequilibrium; HLA-DQB1, major histocompatibility complex, class II, DQ beta 1; CHB, chronic hepatitis B.

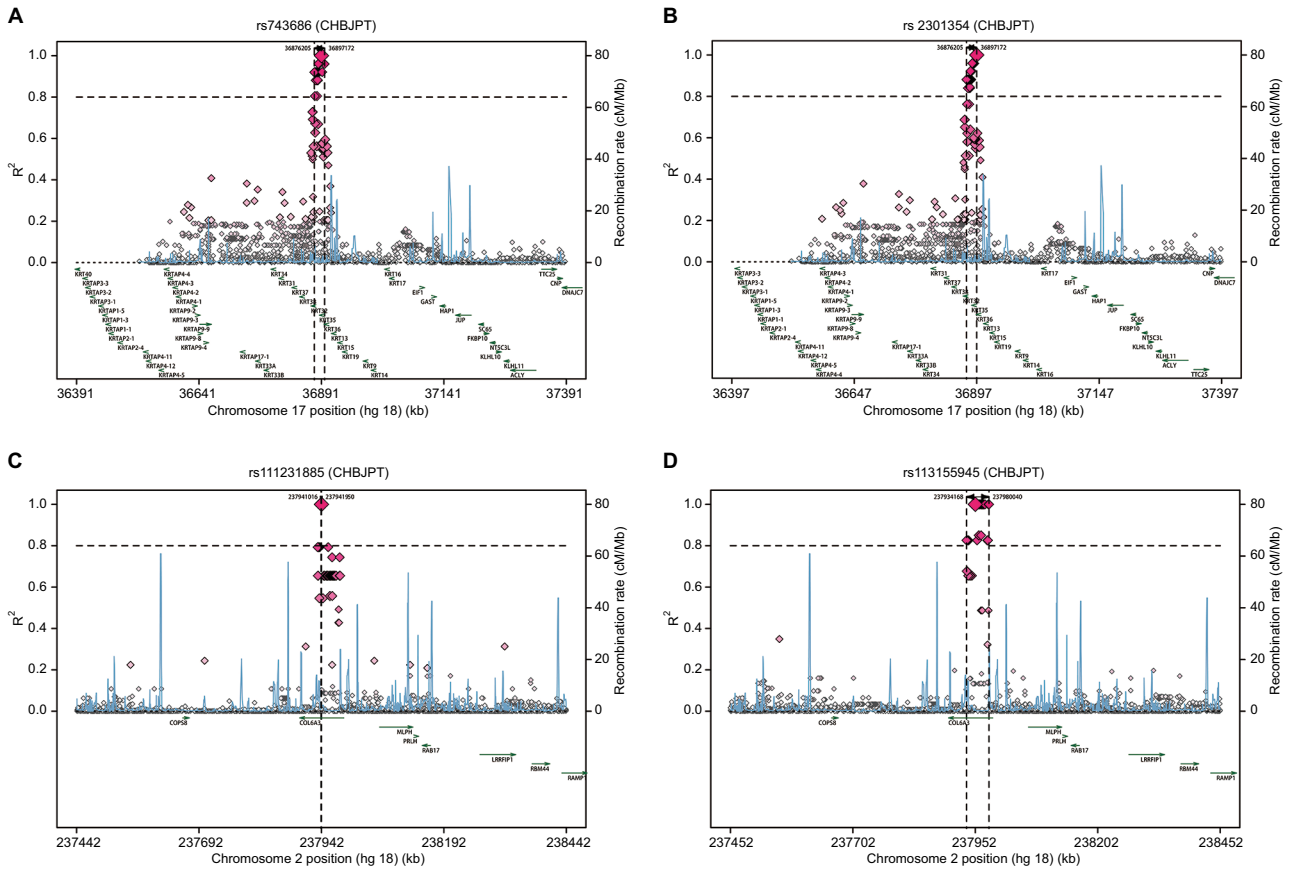

Figure 4 Regional LD plots of block 3 (*KRT35* and *KRT36*) and block 4 (*COL6A3*).
Notes: Regional LD plots of (**A**) *KRT35*-rs743686, (**B**) *KRT36*-rs2301354, (**C**) *COL6A3*-rs111231885, and (**D**) *COL6A3*-rs113155945.
Abbreviations: LD, linkage disequilibrium; KRT35, keratin 35; KRT36, keratin 36; COL6A3, collagen type VI alpha 3 chain.

strong LD loci of these blocks detectable in the region nearby them.

Survival analysis

Survival analysis was used to further investigate the associations between the candidate SNPs and haplotypes with HBV-related HCC prognosis. Clinicopathological characteristics and prognosis information of patients with HBV-related HCC have been described and published in an earlier article[15] and shown in Table 4. Survival analysis of *COL6A3*-rs111231885 showed that patients with the T allele had a shorter median survival time (MST) than those with the C allele (51 vs 33 months for CC vs TT/TC, log-rank *P*=0.012; Table S2). After adjusting for age, gender, race, BMI, smoking status, drinking status, BCLC stage, cirrhosis, radical resection, antiviral therapy, status of tumor capsule, regional invasion, and PVTT in the Cox proportional hazards regression model, patients with the T allele had a significantly increased risk of death compared to those with the C allele (adjusted *P*=0.043, HR=1.64, 95% CI=1.015–2.647; Table S2). Similar results were also observed with *COL6A3*-rs113155945; patients with the TT genotype had a significantly increased risk of death compared to the CC genotype (adjusted *P*=0.047, HR=4.281, 95% CI=1.017–18.022; Table S2). In addition, joint effects analysis was also used to explore the SNPs and TP53 interaction in HBV-related HCC prognosis. TP53-negative patients with the *COL6A3*-rs111231885 T allele carriers had a significantly increased risk of death (adjusted *P*=0.034, HR=1.994, 95% CI=1.052–3.778; Table S3) and a poor clinical outcome (MST: 33 vs 68 months for TT/TC vs CC, log-rank *P*=0.031; Table S3) in HBV-related HCC, compared to TP53-negative patients with C allele carriers. No other genotypes were significantly associated with OS in single and joint effects analysis.

Survival analysis for haplotypes of candidate SNPs is shown in Figure 5A–D, and indicated that patients with TT/other haplotypes in block 4 had a significantly shorter OS (MST: 33 vs 51 months for TT/other haplotypes vs CC, log-rank *P*=0.019; Table 5, Figure 5D). Multivariate analyses of the Cox proportional hazards regression model suggest that patients with TT/other haplotypes in block 4 had increased risk of death in HBV-related HCC (adjusted

Table 4 Clinicopathological characteristics of patients with HBV-related HCC after data quality control

Variable	GWAS				Survival analysis			
	TP53 negative (n=154)	TP53 positive (n=233)	OR (95% CI)	*P*-value	Patients (n=387)	MST (months)	HR (95% CI)	Log-rank *P*
Age (years)								0.313
≤60	133	211	1		344	51	1	
>60	21	22	0.66 (0.35–1.247)	0.201	43	41	1.268 (0.795–2.023)	
Sex								0.563
Male	141	207	1		348	48	1	
Female	13	26	0.734 (0.365–1.477)	0.386	39	42	0.856 (0.504–1.455)	
Race								0.837
Han	103	142	1		245	47	1	
Minority	51	91	1.294 (0.845–1.983)	0.236	142	50	0.968 (0.711–1.318)	
BMI								0.933
≤25	121	181	1		302	48	1	
>25	33	52	1.053 (0.643–1.725)	0.836	85	47	0.985 (0.694–1.399)	
Smoking status								0.054
None	97	147	1		244	61	1	
Ever	57	86	0.996 (0.653–1.518)	0.984	143	39	1.342 (0.992–1.816)	
Drinking status								0.153
None	91	134	1		225	51	1	
Ever	63	99	1.067 (0.706–1.613)	0.758	162	41	1.239 (0.921–1.666)	
Child–Pugh[a]								0.009
A	116	189	1		305	51	1	
B	28	30	0.658 (0.374–1.156)	0.146	58	31	1.665 (1.131–2.452)	
Cirrhosis								0.117
No	12	27	1		39	88	1	
Yes	142	206	0.645 (0.316–1.315)	0.228	348	48	1.503 (0.897–2.518)	
Radical resection[b]								0.095
Yes	82	133	1		215	71	1	
None	67	94	0.865 (0.57–1.313)	0.496	161	41	1.286 (0.955–1.733)	
Portal hypertension[c]								0.347
No	69	124	1		193	52	1	
Yes	75	87	0.645 (0.421–0.989)	0.044	162	42	1.163 (0.847–1.598)	
Pathological grade[d]								0.665
Well differentiated	14	10	1		24	79	1	
Moderately differentiated	126	190	2.111 (0.909–4.901)	0.082	316	48	1.350 (0.688–2.648)	
Poorly differentiated	1	10	14 (1.536–127.621)	0.019	11	NA	1.203 (0.370–3.908)	
Serum AFP[e]								0.499
≤400 (ng/mL)	87	110	1		197	51	1	
>400 (ng/mL)	58	105	1.432 (0.935–2.193)	0.099	163	43	1.111 (0.817–1.512)	
Antiviral therapy								0.002
No	95	157	1		252	40	1	
Yes	59	76	0.779 (0.51–1.192)	0.25	135	NA	0.574 (0.398–0.828)	
Tumor behavior								
Tumor size								<0.001
≤5 cm	72	85	1		157	75	1	
>5 cm	82	148	1.529 (1.011–2.313)	0.044	230	39	1.757 (1.278–2.415)	
Tumor number								<0.001
Single	111	169	1		280	58	1	
Multiple	43	64	0.978 (0.62–1.54)	0.922	107	27	1.788 (1.310–2.439)	
Status of tumor capsule								0.131
Complete	126	201	1		327	50	1	
Incomplete	28	32	0.716 (0.412–1.247)	0.238	60	35	1.335 (0.914–1.949)	
Regional invasion								0.421
Absence	131	199	1		330	51	1	
Presence	23	34	0.973 (0.549–1.726)	0.926	57	40	1.185 (0.781–1.800)	

(*Continued*)

Table 4 (*Continued*)

Variable	GWAS				Survival analysis			
	TP53 negative (n=154)	TP53 positive (n=233)	OR (95% CI)	*P*-value	Patients (n=387)	MST (months)	HR (95% CI)	Log-rank *P*
BCLC stage								<0.001
A	87	129	1		216	95	1	
B	28	38	0.915 (0.523–1.6)	0.756	66	39	2.080 (1.395–3.102)	
C	39	66	1.141 (0.706–1.845)	0.59	105	27	2.672 (1.919–3.720)	
PVTT								<0.001
No	130	189	1		319	73	1	
Yes	24	44	1.261 (0.731–2.175)	0.404	68	27	2.233 (1.615–3.088)	

Notes: [a]Child–Pugh class was unavailable in 24 patients. Information of [b]radical resection was unavailable in 11 patients, [c]portal hypertension was unavailable in 32 patients, [d]pathological diagnosis was unavailable in 36 patients, [e]serum AFP level was unavailable in 27 patients.
Abbreviations: GWAS, genome-wide association study; BMI, body mass index; AFP, α-fetoprotein; BCLC, Barcelona Clinic Liver Cancer; PVTT, portal vein tumor thrombus; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; NA, not applicable; TP53, tumor protein p53; MST, median survival time.

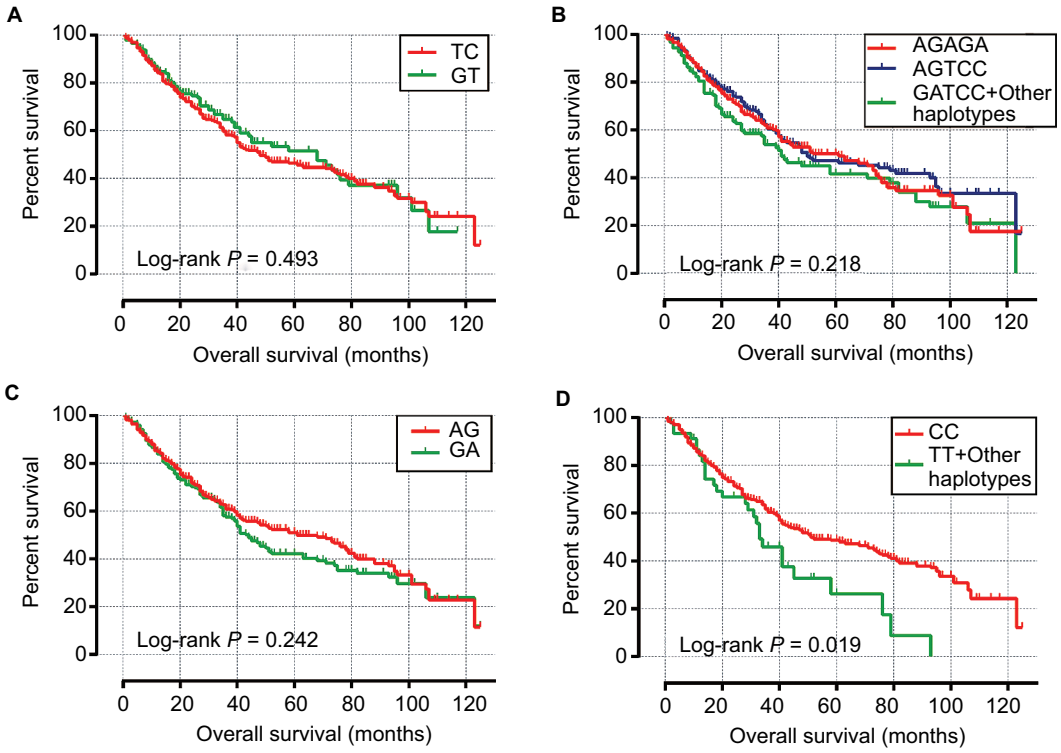

Figure 5 Survival curves of different haplotypes.
Notes: OS stratified by (**A**) block 1 haplotypes, (**B**) block 2 haplotypes, (**C**) block 3 haplotypes, (**D**) block 4 haplotypes.
Abbreviations: OS, overall survival.

P=0.07, HR=1.450, 95% CI=0.970–2.167; Table 5), with a critically significant *P*-value, compared to patients with CC haplotypes. Joint effects analysis (Figure 6A–D) indicated that TP53-negative patients with TT/other haplotypes in block 4 had a significantly poor prognosis (MST: 33 vs 68 months for TT/other haplotypes vs CC haplotypes, log-rank *P*=0.047; Table 6, Figure 6D) and increased risk of death in HBV-related HCC (adjusted *P*=0.047, HR=1.713, 95% CI =1.006–2.918; Table 6), compared to TP53-negative patients with CC haplotypes. No other haplotypes were significantly associated with OS in single or joint effects analysis.

Discussion

The GWAS approach is increasingly being used to discover the association between genes and disease. However, most GWAS have focused on SNPs with high statistical

Table 5 Survival analysis of different haplotypes

Haplotypes	Patients (2n=774)	MST (months)	Crude HR (95% CI)	Crude *P*-value	Adjusted HR (95% CI)	Adjusted *P*-value[a]
Block 1						
TC	616	47	1		1	
GT	158	68	0.913 (0.702–1.187)	0.497	0.934 (0.714–1.221)	0.618
Block 2						
AGAGA	356	51	1		1	
AGTCC	283	48	0.930 (0.735–1.176)	0.544	0.915 (0.717–1.170)	0.48
GATCC+other haplotypes	135	41	1.198 (0.909–1.580)	0.199	1.153 (0.867–1.533)	0.327
Block 3						
AG	447	58	1		1	
GA+other haplotypes	327	41	1.132 (0.918–1.396)	0.247	1.121 (0.903–1.392)	0.299
Block 4						
CC	725	51	1		1	
TT+other haplotypes	49	33	1.566 (1.070–2.292)	0.021	1.450 (0.970–2.167)	0.07

Notes: [a]Adjusted for age, gender, race, body mass index, smoking status, drinking status, BCLC stage, cirrhosis, radical resection, antiviral therapy, status of tumor capsule, regional invasion, and PVTT.
Abbreviations: BCLC, Barcelona Clinic Liver Cancer; PVTT, portal vein tumor thrombus; MST, median survival time.

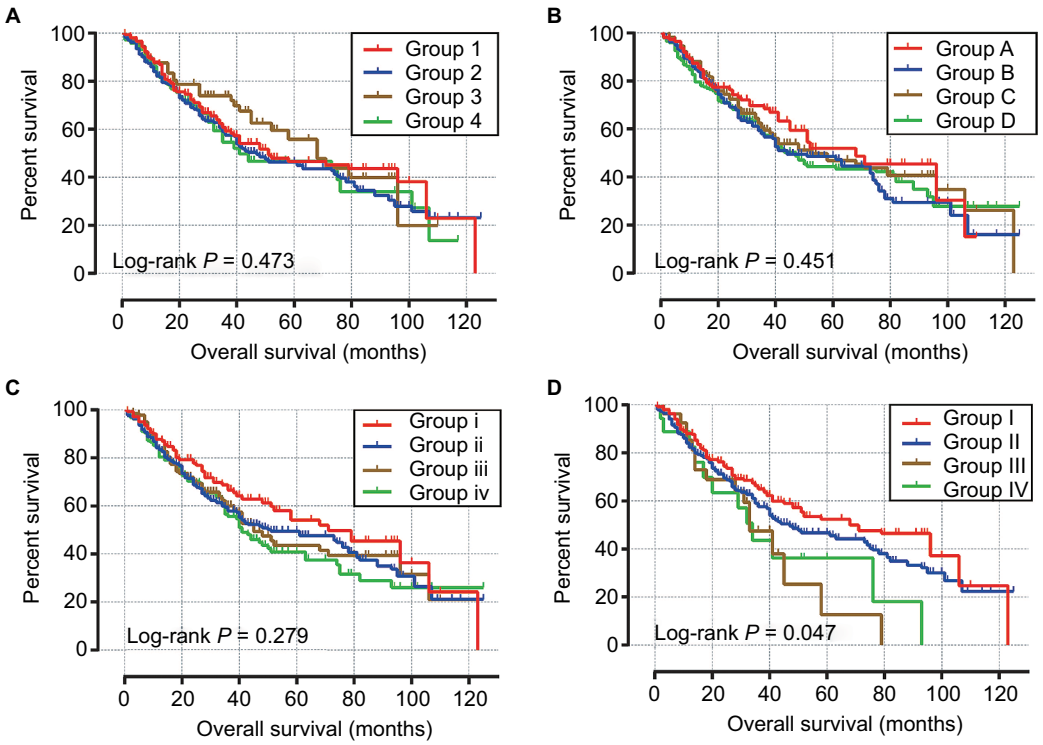

Figure 6 Joint effects survival analysis of different haplotypes and TP53 expression status.
Notes: OS stratified by (**A**) block 1 haplotypes and TP53 expression status, (**B**) block 2 haplotypes and TP53 expression status, (**C**) block 3 haplotypes and TP53 expression status, (**D**) block 4 haplotypes and TP53 expression status.
Abbreviations: OS, overall survival; TP53, tumor protein p53.

significance, whereas many other SNPs have received little attention and the full potential of these data have not been fully exploited.[24,25] Therefore, it is necessary to perform in-depth data mining with GWAS results. Genome-wide pathway analysis can investigate the GWAS SNPs through a SNP→gene→pathway approach to discover the overrepresented pathways in the GWAS data, which would consider rare variants, multi-omics, and interactions. In the current study, ICSNPathway analysis identified 18 candidate SNPs and 10 candidate pathways that are associated with TP53

Table 6 Joint effects survival analysis of different haplotypes and TP53 expression status

Group	Haplotypes	TP53 status	Patients (2n=774)	MST (months)	Crude HR (95% CI)	Crude *P*-value	Adjusted HR (95% CI)	Adjusted *P*-value[a]
	Block 1							
Group 1	TC	Negative	229	51	1		1	
Group 2	TC	Positive	387	43	1.107 (0.868–1.413)	0.413	1.104 (0.855–1.425)	0.449
Group 3	GT	Negative	79	68	0.845 (0.567–1.260)	0.409	0.857 (0.570–1.289)	0.459
Group 4	GT	Positive	79	41	1.126 (0.771–1.644)	0.539	1.158 (0.787–1.706)	0.456
	Block 2							
Group A	AGAGA	Negative	118	68	1		1	
Group B	AGAGA	Positive	238	41	1.302 (0.928–1.827)	0.126	1.206 (0.854–1.704)	0.288
Group C	AGTCC+other haplotypes	Negative	190	48	1.171 (0.818–1.675)	0.389	1.057 (0.733–1.524)	0.765
Group D	AGTCC+other haplotypes	Positive	228	43	1.256 (0.892–1.769)	0.191	1.199 (0.841–1.709)	0.316
	Block 3							
Group i	AG	Negative	156	71	1		1	
Group ii	AG	Positive	291	48	1.243 (0.920–1.678)	0.156	1.329 (0.976–1.811)	0.071
Group iii	GA+other haplotypes	Negative	152	45	1.235 (0.875–1.743)	0.23	1.341 (0.943–1.905)	0.102
Group iv	GA+other haplotypes	Positive	175	40	1.380 (0.993–1.918)	0.055	1.361 (0.971–1.908)	0.074
	Block 4							
Group I	CC	Negative	279	68	1		1	
Group II	CC	Positive	446	44	1.196 (0.953–1.501)	0.123	1.261 (0.995–1.599)	0.055
Group III	TT+other haplotypes	Negative	29	33	1.677 (0.992–2.834)	0.054	1.713 (1.006–2.918)	0.047
Group IV	TT+other haplotypes	Positive	20	32	1.855 (1.045–3.294)	0.035	1.651 (0.872–3.126)	0.123

Notes: [a]Adjusted for age, gender, race, body mass index, smoking status, drinking status, BCLC stage, cirrhosis, radical resection, antiviral therapy, status of tumor capsule, regional invasion, and PVTT.

Abbreviations: TP53, tumor protein p53; BCLC, Barcelona Clinic Liver Cancer; PVTT, portal vein tumor thrombus; MST, median survival time.

expression status in HBV-related HCC. Five hypothetical biological mechanisms can be obtained from ICSNPathway analysis.

The strongest hypothetical biological mechanism found that the candidate SNPs of *HLA-DPB1*, *HLA-DQB1*, and *HLA-DRB1* affected their regulatory roles in all the candidate pathways except hematopoietic cell lineage pathway. The major histocompatibility complex class II molecule is a heterodimer consisting of an alpha (DQA) and a beta chain (DQB), both anchored in the membrane. Previous studies have demonstrated that *HLA-DPB1* polymorphisms were significantly associated with the risk of HBV infection susceptibility,[26,27] whereas the distribution of the SNPs genotype frequencies was similar in HCC and chronic hepatitis B patients.[26,28] A case–control study that compared persistence and natural clearance of HBV infection in a population indicated that the *HLA-DPB1*-rs9277535 A allele has a major effect on the risk of persistent HBV infection.[29] Subsequently, another study also reported that *HLA-DPB1*-rs9277535 was significantly related to HBV infection risks and increased HBV clearance possibility in a dose-dependent manner.[30] Furthermore, the polymorphisms of another HLA class II molecule, *HLA-DQB1*, were also associated with the development of chronic HBV infection and liver cirrhosis,[31] as well as the risk factor of HCC.[32,33] In addition, our previous study also showed that *HLA-DQB1* polymorphisms have a prognosis predictive value in HBV-related HCC patients undergoing hepatic resection.[34] Similar genetic susceptibility research on *HLA-DRB1* also demonstrated that polymorphisms in the *HLA-DRB1* gene were significantly associated with HCC risk, HBV infection, and progression from CHB to HCC.[35–38]

The second strongest mechanism was that a candidate SNP of *HLA-C*-rs1131096 and *HLA-C*-rs1130838 influenced the regulatory role of CAMs, autoimmune thyroid disease, allograft rejection, antigen processing and presentation, type I diabetes mellitus, and graft-versus-host disease pathways. HLA-C belongs to the HLA class I heavy chain paralogues and its genetic variation can influence the risk of HBV-related HCC development;[39] furthermore, HLA-C*15 is also an important host immunogenetic factor that negatively associates with hepatitis C virus viral load in chronic hepatitis C patients.[40]

Studies of major histocompatibility complex class I and II gene polymorphisms demonstrate a risk factor for hepatitis virus and HCC. In the current study, our findings suggest that genetic variation in *HLA-DPB1*, *HLA-DQB1*, *HLA-DRB1*, and *HLA-C* were associated with TP53 expression status in

HBV-related HCC. These results contribute to a better understanding of the heritability of *HLA-DPB1*, *HLA-DQB1*, *HLA-DRB1*, and *HLA-C* in HBV-related HCC and subsequently provide hypotheses to clarify their potential mechanisms in HBV and HCC genetic susceptibility.

In the remaining genes with candidate SNPs, we only found that *IL6R* and *IL7R* were associated with HCC among the previous studies, whereas associations between the other genes and human HCC have not been reported. IL6R encodes a subunit of the *IL6R* complex, and its dysregulation is related to the pathogenesis of many diseases, including cancer. A study by Deng et al revealed that the *IL6R*-rs6684439 T allele is associated with a lower susceptibility of HBV-related HCC in the Guangxi population,[41] whereas miR-451 plays a suppressive role in tumor angiogenesis via the regulation of the IL6R-signal transducer and activator of transcription 3-vascular endothelial growth factor signaling pathway.[42] Research by Midorikawa et al confirms that IL7R is downregulated in well-differentiated tumor tissue in HCC and can serve as a predictor gene of HCC dedifferentiation.[43] Our results contribute to a better understanding of genes associated with different HBV-related HCC subgroups.

Our association analysis demonstrated that seven genes with candidate SNPs were correlated to *TP53* at the HBV-related HCC mRNA level, whereas PPI networks showed that five genes with candidate SNPs were associated with TP53 via UBC through experiments. However, the GeneMANIA gene–gene interaction networks showed complex co-expression networks among those genes, and all genes were directly or indirectly related to the *TP53* gene. In the present study, we confirmed that *LGMN* and *TP53* are positively correlated at the mRNA level in HBV-related HCC based on the GSE14520 data set, and our bioinformatics analysis by GeneMANIA also suggests that *LGMN* and *TP53* were co-expressed via the *TAP1* gene, whereas *LGMN* was also related to *TP53* via the *UBC* gene in the PPI networks that were constructed by STRING. A study by Murthy et al has reported that *LGMN* was significantly upregulated in tumor tissue and its low expression showed a better prognosis in colorectal cancer (CRC); meanwhile, it has a positive correlation with *TP53*.[44] *LGMN* expression and its enzyme activity can also be regulated by *TP53*, and knockdown experiments suggest that *LGMN* and *TP53* have a positive correlation in HCT116 cells.[45] Our bioinformatics analysis also suggests that IL7R is associated with TP53 in GeneMANIA and PPI networks, and *IL7*/*IL7R* prevents apoptosis by regulating *bcl-2* expression and the *TP53* pathway in A549 and human bronchial epithelial cells.[46] Among the 10 candidate pathways, CAMs and cell communication pathways were the most common hypothetical biological mechanisms that involved the majority of candidate SNPs. CAMs play an important role in cell communication[47] and are associated with HCC diagnosis and survival prediction.[48,49] Our findings suggest a novel hypothetical biological mechanism between CAMs and TP53 expression status in HBV-related HCC.

Survival analysis in the current study indicates that the C allele of *COL6A3*-rs111231885 and *COL6A3*-rs113155945, and *COL6A3* block 4 CC haplotypes with TP53 negative status significantly decrease the risk of death in HBV-related HCC patients after hepatectomy. Previous studies have confirmed that high *COL6A3* expression was significantly associated with poor prognosis and its mutation can be used for survival prediction in CRC.[50,51] Furthermore, *COL6A3* was markedly upregulated in the tumor tissue of gastric cancer,[52,53] pancreatic cancer,[54,55] and CRC[50,56] and can serve as a potential diagnostic biomarker in these cancers. This evidence suggests that *COL6A3* may be a potential diagnosis and prognosis marker in CRC and may serve as an oncogene of CRC. Our findings demonstrate that several SNPs of *COL6A3* have a prediction value for HBV-related HCC prognosis and provide insight into the clinical utility of HBV-related HCC prognosis. Once validated, *COL6A3* may be used for prognosis prediction and decision-making in HCC management.

There were limitations in our study that need to be recognized. First, our study evaluates the association between TP53 expression status and candidate SNPs using the GWAS approach, and validates the association between the genes of the candidate SNPs and *TP53* using the Gene Expression Omnibus data set, GeneMANIA, and STRING bioinformatics tools that lack confirmation by in vivo and in vitro experiments. Second, all patients in the present study were exclusively from a Guangxi population of HBV-related HCC; therefore, in order to generalize our findings, additional external validation in cohorts from other ethnic populations is necessary to confirm our results.

Despite these limitations, our study is the first to explore the association between the SNPs and molecular pathways associated with TP53 expression status in HBV-related HCC by using the genome-wide association pathway analysis approach, and that might have etiology or clinical implications.

Conclusion

Genome-wide association pathway analysis in the current study identified 18 candidate SNPs and 10 candidate

pathways that are associated with TP53 expression status in HBV-related HCC and generated five novel SNP-to-gene to pathway hypotheses. These results contribute to a better understanding of the heritability of HBV-related HCC in different TP53 expression subgroups and provide evidence for personalized treatment strategies of different TP53 expression subgroups in HBV-related HCC patients. Additional in vivo and in vitro experimental studies will be necessary to elucidate the role of these pathways in different TP53 expression subgroups of HBV-related HCC. Among these candidate SNPs, the C allele of *COL6A3*-rs111231885 and *COL6A3*-rs113155945, and *COL6A3* block 4 CC haplotypes with TP53 negative status may have protective effects in HBV-related HCC patients after hepatectomy and can serve as a potential prognostic biomarker. Further well-designed and larger sample size studies are needed to validate the associations between *COL6A3* genetic variation and HBV-related HCC prognosis.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (no. 81560535, 81072321, 30760243, 30460143, and 30560133), 2009 Program for New Century Excellent Talents in University, Guangxi Natural Science Foundation (no. GuiKeGong 1104003A-7), Guangxi Health Ministry Medicine Grant (Key-Scientific Research-Grant Z201018), and Self-raised Scientific Research Fund of the Health and Family Planning Commission of Guangxi Zhuang Autonomous Region (Z2016318). We also acknowledge the support offered by the National Key Clinical Specialty Programs (General Surgery and Oncology) and the Key Laboratory of Early Prevention and Treatment for Regional High-Incidence-Tumor (Guangxi Medical University), Ministry of Education, China. The authors thank Prof Minhao Peng, Lequn Li, Xiao Qin, Kaiyin Xiao, Xigang Chen, Bin Chen, Zhixiong Su, Ming Su, Zhang Wen, Jingning Lu, Ning Peng, and Hai Zhu, from the Department of Hepatobiliary Surgery, First Affiliated Hospital of Guangxi Medical University, for providing part of the HCC samples for this study. Thanks also go to Prof Xue Qin (Department of Clinical Laboratory, First Affiliated Hospital of Guangxi Medical University) for providing patient's serum AFP assay results, and Prof Zengnan Mo (Center for Genomic and Personalized Medicine, Guangxi Medical University) for help in the GWAS analysis. We would like to acknowledge the assistance offered by researcher Jiaquan Li and Ying Gui from Guangxi Medical University for their contribution to specimen management. In addition, we also would like to acknowledge the contributors of GSE14520 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14520) for sharing their data set on open access. We thank our reviewers for helpful comments on this article.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66(2):115–132.
2. Zeng H, Zheng R, Guo Y, et al. Cancer survival in China, 2003–2005: a population-based study. *Int J Cancer.* 2015;136(8):1921–1930.
3. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology.* 2007;132(7):2557–2576.
4. Schweitzer A, Horn J, Mikolajczyk RT, Krause G, Ott JJ. Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. *Lancet.* 2015;386(10003):1546–1555.
5. Qi LN, Li LQ, Chen YY, et al. Genome-wide and differential proteomic analysis of hepatitis B virus and aflatoxin B1 related hepatocellular carcinoma in Guangxi, China. *PLoS One.* 2013;8(12):e83465.
6. Liu ZM, Li LQ, Peng MH, et al. Hepatitis B virus infection contributes to oxidative stress in a population exposed to aflatoxin B1 and high-risk for hepatocellular carcinoma. *Cancer Lett.* 2008;263(2):212–222.
7. Yu L, Liu X, Han C, et al. XRCC1 rs25487 genetic variant and TP53 mutation at codon 249 predict clinical outcomes of hepatitis B virus-related hepatocellular carcinoma after hepatectomy: a cohort study for 10 years' follow up. *Hepatol Res.* 2016;46(8):765–774.
8. Stern MC, Umbach DM, Yu MC, London SJ, Zhang ZQ, Taylor JA. Hepatitis B, aflatoxin B(1), and p53 codon 249 mutation in hepatocellular carcinomas from Guangxi, People's Republic of China, and a meta-analysis of existing studies. *Cancer Epidemiol Biomarkers Prev.* 2001;10(6):617–625.
9. Qi LN, Bai T, Chen ZS, et al. The p53 mutation spectrum in hepatocellular carcinoma from Guangxi, China: role of chronic hepatitis B virus infection and aflatoxin B1 exposure. *Liver Int.* 2015;35(3):999–1009.
10. Long XD, Ma Y, Huang HD, Yao JG, Qu DY, Lu YL. Polymorphism of XRCC1 and the frequency of mutation in codon 249 of the p53 gene in hepatocellular carcinoma among Guangxi population, China. *Mol Carcinog.* 2008;47(4):295–300.
11. Farazi PA, DePinho RA. Hepatocellular carcinoma pathogenesis: from genes to environment. *Nat Rev Cancer.* 2006;6(9):674–687.
12. Llovet JM, Zucman-Rossi J, Pikarsky E, et al. Hepatocellular carcinoma. *Nat Rev Dis Primers.* 2016;2:16018.
13. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med.* 2010;363(2):166–176.
14. Zhang K, Chang S, Cui S, Guo L, Zhang L, Wang J. ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. *Nucleic Acids Res.* 2011;39(Web Server issue):W437–W443.
15. Liao X, Han C, Qin W, et al. Genome-wide association study identified PLCE1-rs2797992 and EGFR-rs6950826 were associated with TP53 expression in the HBV-related hepatocellular carcinoma of Chinese patients in Guangxi. *Am J Transl Res.* 2016;8(4):1799–1812.
16. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21(2):263–265.
17. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 2008;24(24):2938–2939.

18. Roessler S, Long EL, Budhu A, et al. Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. *Gastroenterology.* 2012;142(4):957–966.
19. Roessler S, Jia HL, Budhu A, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res.* 2010;70(24):10202–10212.
20. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010;38(Web Server issue):W214–W220.
21. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.* 2008;9(Suppl 1):S4.
22. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43(Database issue):D447–D452.
23. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 2003;31(1):258–261.
24. Ramanan VK, Shen L, Moore JH, Saykin AJ. Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trends Genet.* 2012;28(7):323–332.
25. Elbers CC, van Eijk KR, Franke L, et al. Using genome-wide pathway analysis to unravel the etiology of complex diseases. *Genet Epidemiol.* 2009;33(5):419–431.
26. Posuwan N, Payungporn S, Tangkijvanich P, et al. Genetic association of human leukocyte antigens with chronicity or resolution of hepatitis B infection in Thai population. *PLoS One.* 2014;9(1):e86007.
27. Kim YJ, Kim HY, Lee JH, et al. A genome-wide association study identified new variants associated with the risk of chronic hepatitis B. *Hum Mol Genet.* 2013;22(20):4233–4238.
28. Donaldson PT, Ho S, Williams R, Johnson PJ. HLA class II alleles in Chinese patients with hepatocellular carcinoma. *Liver.* 2001;21(2):143–148.
29. Akgollu E, Bilgin R, Akkiz H, et al. Association between chronic hepatitis B virus infection and HLA-DP gene polymorphisms in the Turkish population. *Virus Res.* 2017;232:6–12.
30. Yu L, Cheng YJ, Cheng ML, et al. Quantitative assessment of common genetic variations in HLA-DP with hepatitis B virus infection, clearance and hepatocellular carcinoma development. *Sci Rep.* 2015;5:14933.
31. Liu C, Cheng B. Association of polymorphisms of human leucocyte antigen-DQA1 and DQB1 alleles with chronic hepatitis B virus infection, liver cirrhosis and hepatocellular carcinoma in Chinese. *Int J Immunogenet.* 2007;34(5):373–378.
32. El-Chennawi FA, Auf FA, Metwally SS, Mosaad YM, El-Wahab MA, Tawhid ZE. HLA-class II alleles in Egyptian patients with hepatocellular carcinoma. *Immunol Invest.* 2008;37(7):661–674.
33. De Re V, Caggiari L, Talamini R, et al. Hepatitis C virus-related hepatocellular carcinoma and B-cell lymphoma patients show a different profile of major histocompatibility complex class II alleles. *Hum Immunol.* 2004;65(11):1397–1404.
34. Liu X, Yu L, Han C, et al. Polymorphisms of HLA-DQB1 predict survival of hepatitis B virus-related hepatocellular carcinoma patients receiving hepatic resection. *Clin Res Hepatol Gastroenterol.* 2016;40(6):739–747.
35. Shi Y, Zhai W, Wang B, et al. Genetic susceptibility of eight nonsynonymous polymorphisms in HLA-DRB1 gene to hepatocellular carcinoma in Han Chinese. *Oncotarget.* 2016;7(49):80935–80942.
36. Jiang DK, Ma XP, Wu X, et al. Genetic variations in STAT4,C2,HLA-DRB1 and HLA-DQ associated with risk of hepatitis B virus-related liver cirrhosis. *Sci Rep.* 2015;5:16278.
37. 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(7):e1002791.
38. Jin YJ, Shim JH, Chung YH, et al. Relationship of HLA-DRB1 alleles with hepatocellular carcinoma development in chronic hepatitis B patients. *J Clin Gastroenterol.* 2012;46(5):420–426.
39. Pan N, Jiang W, Sun H, et al. KIR and HLA loci are associated with hepatocellular carcinoma development in patients with hepatitis B virus infection: a case-control study. *PLoS One.* 2011;6(10):e25682.
40. Tseng KC, Tseng CW, Hsieh YH, et al. Effect of human leukocyte antigen class I and II alleles on hepatitis C viral load among chronic hepatitis C patients in Southern Taiwan. *Hum Immunol.* 2013;74(8):978–982.
41. Deng Y, Li M, Wang J, et al. Susceptibility to hepatocellular carcinoma in the Chinese population: associations with interleukin-6 receptor polymorphism. *Tumour Biol.* 2014;35(7):6383–6388.
42. Liu X, Zhang A, Xiang J, Lv Y, Zhang X. miR-451 acts as a suppressor of angiogenesis in hepatocellular carcinoma by targeting the IL-6R-STAT3 pathway. *Oncol Rep.* 2016;36(3):1385–1392.
43. Midorikawa Y, Tsutsumi S, Taniguchi H, et al. Identification of genes associated with dedifferentiation of hepatocellular carcinoma with expression profiling analysis. *Jpn J Cancer Res.* 2002;93(6):636–643.
44. Murthy RV, Arbman G, Gao J, Roodman GD, Sun XF. Legumain expression in relation to clinicopathologic and biological variables in colorectal cancer. *Clin Cancer Res.* 2005;11(6):2293–2299.
45. Yamane T, Murao S, Kato-Ose I, et al. Transcriptional regulation of the legumain gene by p53 in HCT116 cells. *Biochem Biophys Res Commun.* 2013;438(4):613–618.
46. Liu ZH, Wang MH, Ren HJ, et al. Interleukin 7 signaling prevents apoptosis by regulating bcl-2 and bax via the p53 pathway in human non-small cell lung cancer cells. *Int J Clin Exp Pathol.* 2014;7(3):870–881.
47. Linhardt RJ, Toida T. Role of glycosaminoglycans in cellular communication. *Acc Chem Res.* 2004;37(7):431–438.
48. Iliaz R, Akyuz U, Tekin D, et al. Role of several cytokines and adhesion molecules in the diagnosis and prediction of survival of hepatocellular carcinoma. *Arab J Gastroenterol.* 2016;17(4):164–167.
49. Shimizu Y, Minemura M, Tsukishiro T, et al. Serum concentration of intercellular adhesion molecule-1 in patients with hepatocellular carcinoma is a marker of the disease progression and prognosis. *Hepatology.* 1995;22(2):525–531.
50. Qiao J, Fang CY, Chen SX, et al. Stroma derived COL6A3 is a potential prognosis marker of colorectal carcinoma revealed by quantitative proteomics. *Oncotarget.* 2015;6(30):29929–29946.
51. Yu J, Wu WK, Li X, et al. Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer. *Gut.* 2015;64(4):636–645.
52. Xie X, Liu X, Zhang Q, Yu J. Overexpression of collagen VI α3 in gastric cancer. *Oncol Lett.* 2014;7(5):1537–1543.
53. Sun H. Identification of key genes associated with gastric cancer based on DNA microarray data. *Oncol Lett.* 2016;11(1):525–530.
54. Arafat H, Lazar M, Salem K, et al. Tumor-specific expression and alternative splicing of the COL6A3 gene in pancreatic cancer. *Surgery.* 2011;150(2):306–315.
55. Kang CY, Wang J, Axell-House D, et al. Clinical significance of serum COL6A3 in pancreatic ductal adenocarcinoma. *J Gastrointest Surg.* 2014;18(1):7–15.
56. Snezhkina AV, Krasnov GS, Zaretsky AR, et al. Differential expression of alternatively spliced transcripts related to energy metabolism in colorectal cancer. *BMC Genomics.* 2016;17(Suppl 14):1011.