Genome-Wide Association Study Identifies a Genetic Prediction Model for Postoperative Survival in Patients with Hepatocellular Carcinoma

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCDEF 1 **Jinwang Wei***
ABCF 1 **Yuanyuan Sheng***
BDF 1 **Jianhua Li**
BCDF 1 **Xiaomei Gao**
BDG 2 **Ning Ren**
ACDEG 1 **Qiongzhu Dong**
ABDEG 1 **Lunxiu Qin**

1 Department of General Surgery, Huashan Hospital; Cancer Metastasis Institute; Institutes of Biomedical Sciences, Fudan University, Shanghai, P.R. China
2 Liver Cancer Institute and Zhongshan Hospital, Fudan University, Shanghai, P.R. China

Corresponding Authors:
Source of support:

* Jinwang Wei and Yuanyuan Sheng contributed equally to this article
Lunxiu Qin, e-mail: qinlx@fudan.edu.cn, Qiongzhu Dong, e-mail: qzhdong@fudan.edu.cn

Background: As an important aspect of tumor heterogeneity, genetic variation may influence susceptibility and prognosis in different types of cancer. By exploring the prognostic value of genetic variation, this study aimed to establish a model for predicting postoperative survival and assessing the impact of variation on clinical outcomes in patients with hepatocellular carcinoma (HCC).

Material/Methods: A genome-wide association study of 367 patients with HCC was conducted to identify single nucleotide polymorphisms (SNPs) associated with prognosis. Identified predictors were further evaluated in 758 patients. Two prognostic models were established using Cox proportional hazards regression and Nomogram strategy, and validated in another 316 patients. The effect of the SNP rs2431 was analyzed in detail.

Results: A prognostic model including 5 SNPs (rs10893585, rs2431, rs34675408, rs6078460, and rs6766361) was established and exhibited high predictive accuracy for HCC prognosis. The panel combined with tumor node metastasis (TNM) stage resulted in a significantly higher c-index (0.723) than the individual c-index values. Stratified by the Nomogram prediction model, the median overall survival for the low-risk and high-risk groups were 100.1 versus 30.8 months ($P < 0.001$) in the training set and 82.2 versus 22.5 months ($P < 0.001$) in the validation set. A closer examination of rs2431 revealed that it may regulate the expression of FNDC3B by disrupting a microRNA-binding site.

Conclusions: This study established prediction models based on genetic factors alone or in combination with TNM stage for postoperative survival in patients with HCC, and identified FNDC3B as a potential therapeutic target for combating HCC metastasis.

MeSH Keywords: **Carcinoma, Hepatocellular • Genome-Wide Association Study • Prognosis**

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/915511

📄 4013   📊 16   📈 12   📖 24

Background

Hepatocellular carcinoma (HCC) is the sixth most common cancer and the second most common cause of death from cancer worldwide [1,2]. Around 50% of newly diagnosed cases occur in China. HCC is highly vascularized with frequent intrahepatic and extrahepatic metastasis, which is responsible for the rapid recurrence and poor survival [3,4]. In the past few decades, although progress has been made in the clinical management of HCC, prognosis has not improved substantially [5,6].

Single nucleotide polymorphisms (SNPs) are the most common form of genetic variation in the human genome. They are related to risk and heterogeneity in human diseases, clinical course, and response to treatment [7,8]. Although studies have demonstrated associations between SNPs in specific candidate genes for HCC and outcome, the prognostic value of these germline variants remains controversial [9–11].

Genome-wide interrogation, which is independent of a priori hypotheses, has emerged as a powerful tool for identifying genetic variants associated with the length of survival for different types of cancer [12–16]. While the candidate gene approach is based on our relatively limited understanding of tumor and host biology, the genome-wide interrogation approach allows a more comprehensive evaluation of germline variants across the whole genome [12]. For HCC, several genome-wide association studies (GWASs) have identified genetic variants at genomic *loci* 1p36.22, 6p21.32, 7q21.13, and 21q21.3, in *STAT4* and *HLA-DQ*, and in the HLA class I region that confer susceptibility in chronic hepatitis B virus carriers [17–21]. However, little is known about genomic *loci* associated with the survival of patients with HCC.

In this study, we first conducted a GWAS of 367 patients with HCC to identify SNPs associated with outcome. With the aim of building a survival prediction model, the prognostic values of SNPs were further evaluated in a replication set of 758 cases. The survival prediction model was then validated in an independent set of 316 patients and combined with TNM stage to construct a new model that included "seed and soil" factors. According to this model, the SNP rs2431, which is located in the 3' untranslated region (3' UTR) of *FNDC3B*, regulates the expression level of this metastasis-promoting gene by affecting the binding of miR-409-3p.

Material and Methods

Study design

A total of 1802 patients with HCC who underwent liver resection at the authors' institutes between 2004 and 2009 were recruited. Among these patients, 1441 were enrolled in the SNP genotyping study: 1) the discovery set consisted of 367 patients for genome-wide interrogation of *loci* related to survival; 2) the replication set consisted of 758 patients to further evaluate the prognostic value of SNPs, and the discovery and replication sets were combined to construct a genetic prediction model for postoperative survival in patients with HCC; 3) the independent validation set consisted of 316 patients to test the prognostic value of the 5-SNP survival prediction model. Formalin-fixed and paraffin-embedded tissues of the remaining 361 patients were used to construct a tissue microarray for immunohistochemical (IHC) analyses. The flow of the study is illustrated in Figure 1.

Patients and specimens

This study was approved by the Research Ethics Committee of Fudan University (2017-041). No patients in this study received preoperative cancer treatment before curative resection for HCC, and all patients were followed up until April 2014. All of the samples included in this study had no visible residual tumor after curative resection. All patients with hepatitis B virus (HBV)-related/hepatitis C virus (HCV)-related HCC received antiviral therapy. No patients included in this study received molecular targeted therapy. Data on survival or death and clinicopathological characteristics were recorded for all participating individuals. Overall survival (OS) was calculated from the date of surgery to the date of last follow-up or death. During the follow-up period, patients were examined every 2–3 months after the operation. Computed tomography or magnetic resonance imaging was performed when tumor recurrence was suspected. Snap-frozen or paraffin-embedded tissue samples and blood were obtained from patients after obtaining written informed consent. All tumors were confirmed to be HCC by 2 independent pathologists. Histological grade was staged using the TNM stage system of the Union for International Cancer Control/American Joint Committee on Cancer (2010).

DNA extraction and genotyping

In the discovery set, genomic DNA of each study participant was extracted from blood samples using the Quick Gene DNA Whole Blood Kit (Fujifilm, Osaka, Japan) and genotyped using GeneChip Axiom Genome-Wide Human ASI (Affymetrix, Santa Clara, CA, USA). A total of 367 samples and 630 573 SNPs were used in the final GWAS. In the replication and validation sets, genomic DNA from each participant was extracted from tumor-adjacent tissue samples using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), and chosen SNPs were genotyped using MassARRAY (Sequenom, San Diego, CA, USA).

Quality control

Systematic quality control was performed on genotyping data from the GWAS and Sequenom MassArray data using PLINK,
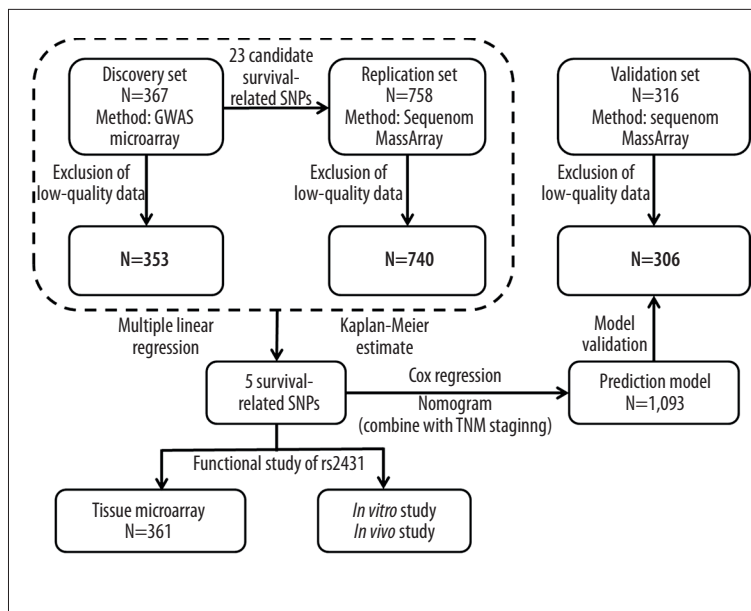
Figure 1. Flow chart illustrating the study design.

including 95% SNP and 90% individual call rate thresholds; the number of samples excluded from analyses is shown in Figure 1. For each sample set, the genotype counts and Hardy-Weinberg equilibrium (HWE) results are shown in Supplementary Table 1. The call rates for 5 selected SNPs in each sample set were all >95%. In the entire cohort, the HWE results were $\chi^2=0.191$, $P=0.662$ for rs10793585; $\chi^2=4.822$, $P=0.028$ for rs2431; $\chi^2=0.890$, $P=0.169$ for rs34675408; $\chi^2=0.233$, $P=0.629$ for rs6078460; $\chi^2=0.077$, $P=0.782$ for rs6766361. No significant deviation from HWE was detected ($P\leq0.001$). To evaluate the genotyping quality, we genotyped the 23 SNPs in 30 randomly selected genome-wide scan samples using Sequenom MassArray, and the concordance between the 2 methods was >95%.

Detection of protein levels by western blotting

Total protein was extracted, separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and transferred onto polyvinylidene fluoride membranes (Millipore, Billerica, MA, USA). The following antibodies were used: anti-FNDC3B (ab135714; Abcam, Cambridge, UK), anti-GAPDH (CST), and goat-anti-rabbit IgG conjugated to horseradish peroxidase (HRP; Cwbiotech), which was used as the secondary antibody. Protein was detected by Image Acquisition using ImageQuant™ LAS 4000 (GE Healthcare Life Sciences, Marlborough, MA, USA).

Dual luciferase reporter assay

HEK293T or Hep3B cells were seeded in a 96-well plate at 50% to 60% confluence. After 24 hours, cells were transfected with 120 ng of pWPI-miR-409-3p vector or control vector, and co-transfected with 30 ng of the 3'UTR of *FNDC3B* vector containing the rs2431[G] or rs2431[A] alleles using 0.45 µL of

Fugene (Promega). Cells were collected 48 hours after transfection, Firefly and Renilla luciferase activity was measured using a dual-luciferase reporter system (Promega). Luciferase reporter assays were performed in duplicate and repeated in 3 independent experiments. Luciferase activity was detected using an Orion II microplate luminometer (Berthold Technologies, Bad Wildbad, Germany).

IHC staining analysis

Formalin-fixed and paraffin-embedded tissues were used for IHC staining. The rabbit polyclonal antibody to FNDC3B (HPA007859; Sigma-Aldrich, St. Louis, MO, USA) was incubated with the tissue sections or microarrays at 4°C overnight and detected using the GTVision™ II Kit (HRP/DAB, Rabbit/Mouse; Genetech, San Francisco, CA, USA). Staining intensity was scored manually by 2 independent experienced pathologists as follows: 0=no staining, 1=weak staining, 2=moderate staining, and 3=strong staining. Tissue sections or microarrays were scored based on the percentage of positively stained cells (0–100%). The final IHC reaction score (IRS) was calculated as the sum of the products of the intensity score and the percentage of positive cells.

Cell lines and cell culture

Four human HCC cell lines (MHCC97-H, HCCLM3, Huh7, and Hep3B), and one human embryonic kidney cell line (HEK293T) were used in this research. MHCC97-H and HCCLM3 were established at the Liver Cancer Institute and Zhongshan Hospital, Fudan University and have high metastatic potential; Huh7, Hep3B, and HEK293T were obtained from the American Type Culture Collection (Manassas, VA, USA). All cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM;

Hyclone, Logan, UT, USA) with 10% fetal bovine serum (FBS; Gibco, Gaithersburg, MD, USA) at 37°C and 5% CO_2.

Statistical analysis

Genotype quality control was performed within each study population using PLINK. For the GWAS and replication of survival data, multiple linear regression was performed for each SNP and case. Linkage disequilibrium (LD) between genotyped SNPs on the same chromosome was determined using Haploview version 4.2 (Broad Institute). TAGster was used for tag SNP selection, and LD was measured by r^2 with a threshold of 0.8 and a maximum distance between SNPs of 250 kb. Genotype data for SNPs were obtained from dbSNP (April 2009) for the Asian population. A principal component analysis (PCA)-based method implemented in the software package GCTA was used to detect population outliers and stratification. Kaplan-Meier and χ^2 tests were implemented in SPSS version 16.0 (SPSS Inc., Chicago, IL, USA). P-values for survival analyses were obtained using log-rank tests. The Cox proportional hazard model was applied for univariate and multivariate analyses to evaluate the effect of each clinical variable, including hazard ratios (HRs) and their 95% confidence intervals (CIs), and to construct a survival predicted model. Nomogram, receiver operating characteristics (ROC) curves, and the calibration curve were plotted using R version 3.3.3, and the nearest neighbor estimate was used for the ROC analysis. HWE was evaluated using χ^2 tests.

Results

Patient characteristics

Detailed clinicopathological characteristics of participants in the discovery, replication, and validation sets are presented in Table 1. No significant differences in clinicopathologic features were observed among the discovery, replication, and validation sets. Among the selected characteristics, TNM stage, Barcelona Clinic Liver Cancer (BCLC) stage, Child-Pugh stage, serum alpha-fetoprotein (AFP) concentration, tumor size, tumor number, vascular invasion, portal vein tumor thrombus, extrahepatic metastasis, tumor differentiation, and tumor encapsulation were significantly associated with survival (Supplementary Table 2). The characteristics of the 361 cases for tissue microarrays used for IHC analyses are shown in Supplementary Table 3.

Selection of SNPs associated with survival length in HCC

After standard quality filtering, 630 573 SNPs and 367 cases were used for genome-wide interrogation of survival length (Figure 2). PCA showed little evidence of population

stratification in our study populations; therefore, no principal component was adjusted in subsequent association analyses (Supplementary Figure 1A). The effect of genotypes on OS was investigated using multiple linear regression [22]. A quantile-quantile plot revealed a good match between the distributions of observed P-values and those expected by chance (Supplementary Figure 1B).

Among all SNPs, only 6 met the standard threshold ($P<10^{-5}$) for GWAS. Our limited sample size for the discovery set (n=367) might not be sufficient for optimal statistical power. To identify as many significant features for validation as possible, we defined the threshold as $P<10^{-3.5}$ and identified 149 SNPs that were significantly associated with the length of survival. The 6 SNPs that met the $P<10^{-5}$ threshold were directly included in the replication stage. Combined the results of Kaplan-Meier analyses ($P<0.001$), 23 candidate SNPs were selected (Supplementary Table 4) for further evaluation in an independent replication set of 758 samples using MassARRAY. Among them, 8 SNPs were located in functional regions of genes, such as the 3'UTR, 5'UTR, or coding sequence (CDS).

In the replication set, 7 of the 23 SNPs (rs10790836, rs10893585, rs11600756, rs2431, rs34675408, rs6078460, and rs6766361) were significantly associated with OS in HCC ($P<0.05$, Table 2). The association between the 23 SNPs and the time to recurrence was also analyzed, and all 7 SNPs showed significant predictive value in the combined sample set (Supplementary Table 5). Three SNPs (rs10790836, rs10893585, and rs11600756) located on Chr.11q24 were in strong LD and could be tagged by rs10893585 ($r^2=0.8272$ for rs10790836, $r^2=0.8511$ for rs11600756).

According to Kaplan-Meier estimates, genotypes of each SNP could be subdivided into benign and malignant. OS was significantly longer for CT+TT than CC, AA than AG+GG, TT than GT, AG+GG than AA, and AA+CA than CC for rs10893585, rs2431, rs34675408, rs6078460, and rs6766361, respectively (Supplementary Figure 2). To further evaluate the prognostic value of the 5 selected SNPs, multivariate analyses using Cox proportional hazards regression models were performed using the combined cohort. Significant individual features from the univariate analysis were adopted as covariates. The SNPs were independent prognostic indicators of the length of survival (Supplementary Table 6).

Construction and validation of an SNP survival prediction model

To further define an SNP panel that could accurately predict prognosis in patients with HCC after curative resection, we applied the Cox proportional hazards regression model to estimate the risk of long-term survival using an entry procedure on

Table 1. Characteristics of individuals with HCC in the discovery, replication, and validation samples.

Characteristics	Discovery set (N=367)		Replication set (N=758)		Validation set (N=316)	
	Number	(%)	Number	(%)	Number	(%)
Total						
Deceased	167	45.5	350	46.2	154	48.7
Alive	200	54.5	408	53.8	162	51.3
MST, months	62.2	–	58.4	–	58.0	–
Gender						
Male	315	85.8	648	85.5	272	85.9
Female	52	14.2	110	14.5	44	14.1
Age, years						
≤55	211	57.5	460	60.7	188	59.5
>55	156	42.5	298	39.3	128	40.5
AFP, ng/ml						
≤20	147	40.1	281	37.1	124	39.2
>20	218	59.4	477	62.9	192	60.8
ALT, ng/ml						
≤75	316	86.1	610	80.5	272	85.9
>75	50	13.6	120	15.8	44	14.1
Tumor size, cm						
≤5	183	49.9	382	50.4	153	48.4
>5	184	50.1	376	49.6	163	51.6
Tumor number						
Single	305	83.1	651	85.9	270	85.3
Multiple	62	16.9	107	14.1	46	14.7
Vascular invasion						
Yes	143	39.0	318	42.0	122	38.6
No	224	61.0	440	58.0	193	61.1
PVTT						
Yes	38	10.4	84	11.1	30	9.5
No	329	89.6	674	88.9	286	90.5
Extrahepatic metastasis						
Yes	18	4.9	30	4.0	18	5.6
No	349	95.1	728	96.0	298	94.4
TNM stage						
I–II	279	76.0	586	77.3	236	77.1
III–IV	88	24.0	172	22.7	70	22.9

Table 1 continued. Characteristics of individuals with HCC in the discovery, replication, and validation samples.

Characteristics	Discovery set (N=367)		Replication set (N=758)		Validation set (N=316)	
	Number	(%)	Number	(%)	Number	(%)
BCLC stage						
0–A	124	33.8	268	35.4	101	33.0
B–C	243	66.2	490	64.6	205	67.0
Child-Pugh stage						
A	348	94.8	722	95.3	298	94.3
B	19	5.2	36	4.7	18	5.7
Tumor differentiation						
I–II	247	67.3	513	67.7	207	65.4
III–IV	113	30.8	227	29.9	90	28.4
Liver cirrhosis						
Yes	97	26.4	167	22.0	70	22.2
No	270	73.6	590	77.8	238	75.2
Tumor encapsulation						
None	212	57.8	426	56.2	178	56.2
Complete	142	38.7	330	43.5	128	40.5
HBsAg						
Positive	309	84.2	601	79.3	277	87.6
Negative	54	14.7	119	15.7	39	12.4
HBcAb						
Positive	343	93.5	661	87.2	292	92.8
Negative	20	5.5	59	7.8	14	4.2
Anti-HCV						
Positive	3	0.8	6	0.8	4	1.3
Negative	355	96.7	713	94.1	303	96.8

Owing to a lack of information for some samples, the number of subjects in each category may not add up to the total number.
AFP – alpha-fetoprotein; ALT – alanine aminotransferase; BCLC – Barcelona Clinic Liver Cancer; CI – confidence interval; HBsAg – hepatitis B surface antigen; HBcAb – hepatitis B core antibody; HCV – hepatitis C virus; HR: hazard ratio; PVTT – portal vein tumor thrombus; TNM – tumor node metastasis.

the combined discovery and replication sets. The Cox P-values for all 5 SNPs were less than 0.05 (Supplementary Table 7). For each SNP, the Cox coefficient was set as the final value for further calculations to obtain a survival prediction model. The predicted survival value, which could be defined as the Cox index, of different haplotypes was calculated based on the following 5-SNP panel:

$$\text{Cox index} = -0.268 \times \text{rs10893585} - 0.454 \times \text{rs2431} + 0.570 \times \text{rs34675408} + 0.226 \times \text{rs6078460} - 0.288 \times \text{rs6766361}$$

Categorical genotypes were coded as follows: CC as 1 and CT+TT as 0 for rs10893585; AA as 1 and AG+GG as 0 for rs2431; GT as 1 and TT as 0 for rs34675408; AA as 1 and AG+GG as 0 for rs6078460; AA+CA as 1 and CC as 0 for rs6766361. In the model, a higher score indicated poorer prognosis.
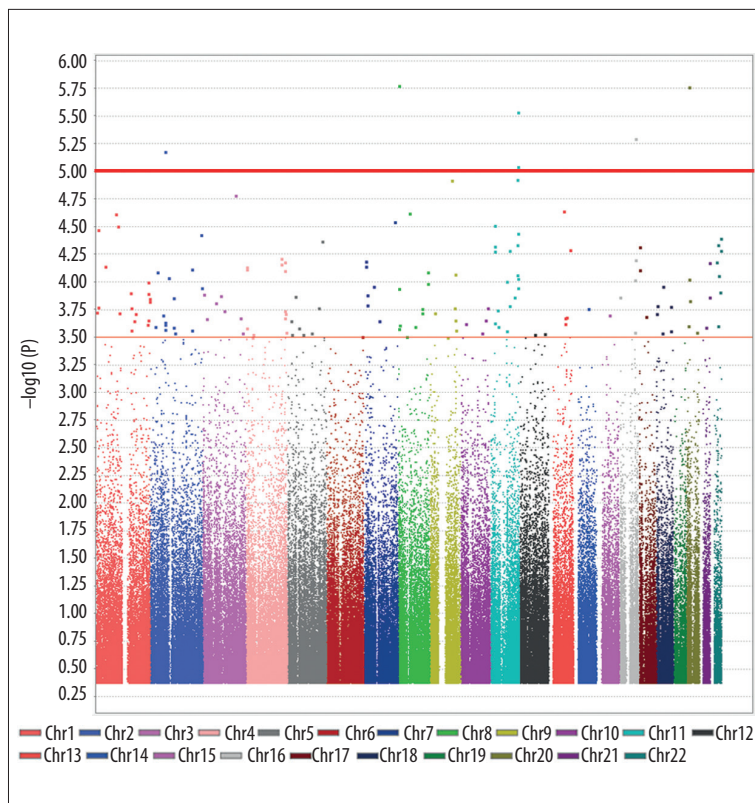
Figure 2. Manhattan plot of genome-wide P-values for associations between SNPs and survival time in patients with HCC. The effects of SNP genotypes on quantitative variables were investigated using multiple linear regression. The $-\log_{10}$P-values (y-axis) for SNPs are presented according to chromosomal positions (x-axis). SNP – single nucleotide polymorphisms; HCC – hepatocellular carcinoma.

The prognostic performance of the SNP scoring system was evaluated by ROC analysis and Harrell's c-index. The area under the ROC curve for 5-year survival was 0.716 (95% CI: 0.683–0.746), better than that for any single SNP (Supplementary Figure 3), and the c-index was 0.666 (95% CI: 0.642–0.689). Using the optimal cutoff value (–0.057) determined from the ROC curve, the sensitivity and specificity of the 5-SNP survival prediction model were 58.8% and 74.8%, respectively. Then, Kaplan-Meier analyses were used to evaluate the prognostic performance of the 5-SNP prediction model using the discovery and replication sets(Figure 3A, 3B). According to the cutoff value of the model, the patients were subdivided into high-risk (higher prediction value) and low-risk (lower prediction value) groups. A significantly longer survival time was found in low-risk patients than in high-risk patients in the combined set, with median OS of 100.1 and 35.5 months, respectively (HR=1.725, P=1.45×10^{-10}; Figure 3C, Supplementary Table 8). To further evaluate the prognostic value of the 5-SNP prediction model, multivariate analyses using Cox proportional hazards regression models were performed for each set. The 5-SNP prediction model was an independent prognostic indicator of OS (Supplementary Table 9).

The prediction model was analyzed using another independent validation set. Kaplan-Meier survival curves provided strong evidence for the reliability of predictions in the validation set, with median OS times of 80.2 and 48.0 months, respectively

(HR=1.481, P=0.002; Figure 3D, 3E, Supplementary Table 8). In addition, multivariate analyses also demonstrated that the 5-SNP prediction model was an independent prognostic factor for OS (Supplementary Table 9). Furthermore, the association between the 5-SNP prediction model and recurrence-free survival (RFS) was also analyzed, and in all sample sets, significantly longer RFS times were found in low-risk patients than in high-risk patients (Supplementary Table 10).

Combining the SNP prediction model with TNM stage

In addition to the 5-SNP prediction model, multivariate analyses showed that various clinicopathologic features, such as tumor number, tumor size, and tumor invasion, were also independent factors for survival in HCC. To explore whether tumor-specific characteristics could complement and strengthen the SNP algorithm, we added TNM stage data, which is a comprehensive summary of these clinicopathological features, to the 5-SNP prediction model using a Nomogram strategy (Figure 4A).

P-values for all variables in the Nomogram model were less than 0.05 (Supplementary Table 11). The predicted survival value for HCC, which could be defined as the Nomogram index, for different haplotypes can be calculated based on the 5 SNPs and the TNM stage as follows:

Table 2. Associations between 23 SNPs and length of survival in the discovery, replication, and combined samples with HCC.

SNP	*P*-value (multiple linear regression analysis)		
	Discovery set (n=367)	Replication set (n=758)	Combined discovery and replication sets (n=1125)
rs10481370	1.67×10^{-6}	0.873	0.026
rs6078460	1.72×10^{-6}	**0.038**	2.24×10^{-11}
rs11600756	2.93×10^{-6}	**0.007**	4.39×10^{-5}
rs11639771	5.08×10^{-6}	0.985	0.570
rs7579933	6.60×10^{-6}	0.498	0.351
rs10790836	9.10×10^{-6}	**0.042**	4.79×10^{-4}
rs16910200	1.20×10^{-5}	0.455	0.490
rs1893185	1.18×10^{-5}	0.295	0.031
rs6766361	1.65×10^{-5}	**0.042**	3.90×10^{-3}
rs913493	2.30×10^{-5}	0.946	4.70×10^{-20}
rs12897	3.74×10^{-5}	0.334	0.023
rs2239910	4.63×10^{-5}	0.178	5.38×10^{-3}
rs584368	5.19×10^{-5}	0.246	1.81×10^{-14}
rs12504893	6.14×10^{-5}	0.465	0.151
rs1034756	6.54×10^{-5}	0.498	0.284
rs4617227	8.49×10^{-5}	0.473	0.063
rs10893585	9.29×10^{-5}	**0.018**	5.67×10^{-4}
rs3096380	9.63×10^{-5}	0.122	1.34×10^{-17}
rs3748735	9.84×10^{-5}	0.880	0.666
rs2431	2.11×10^{-4}	**0.029**	3.46×10^{-3}
rs7581876	2.41×10^{-4}	0.720	7.10×10^{-4}
rs34675408	2.59×10^{-4}	**0.032**	1.33×10^{-32}
rs6499496	2.85×10^{-4}	0.718	0.146

Nomogram index=−0.255×rs10893585−0.450×rs2431+0.506× rs34675408+ 0.228×rs6078460−0.296×rs6766361+ 0.412(TNM=II)+0.877(TNM=III)+0.926(TNM=IV)

In the model, the higher the score, the worse the prognosis.

Harrell's c-index of the Nomogram for predicting survival was 0.723 (95% CI: 0.701–0.745), which was significantly higher than those of the SNPs or the TNM stage alone (SNPs: 0.666, *P*=2.33×10⁻⁴; TNM stage: 0.654 (95% CI: 0.632–0.676), *P*=5.80×10⁻⁶; Figure 4B (left)). The result was confirmed using the validation set with a Harrell's c-index of the Nomogram of 0.714 (95% CI: 0.673–0.755), which was significantly higher

than those of SNPs or TNM stage alone (SNPs: 0.648 (95% CI: 0.606–0.691), *P*=0.014; TNM stage: 0.652 (95% CI: 0.610–0.694), *P*=0.018; Figure 4B (right)). The calibration plots for the probability of survival at 3 or 5 years after surgery also showed optimal agreement between the prediction by Nomogram and actual observation in the combined set and validation set (Figure 4C).

The area under the ROC curve for 5-year survival was 0.766 (95% CI: 0.735–0.796), which was significantly higher than those for the SNPs or TNM stage alone (SNP: 0.716, *P*=0.011; TNM stage: 0.698 (95%CI: 0.664–0.731), *P*=0.002; Supplementary Figure 4). Using the optimal cutoff value (0.695) determined
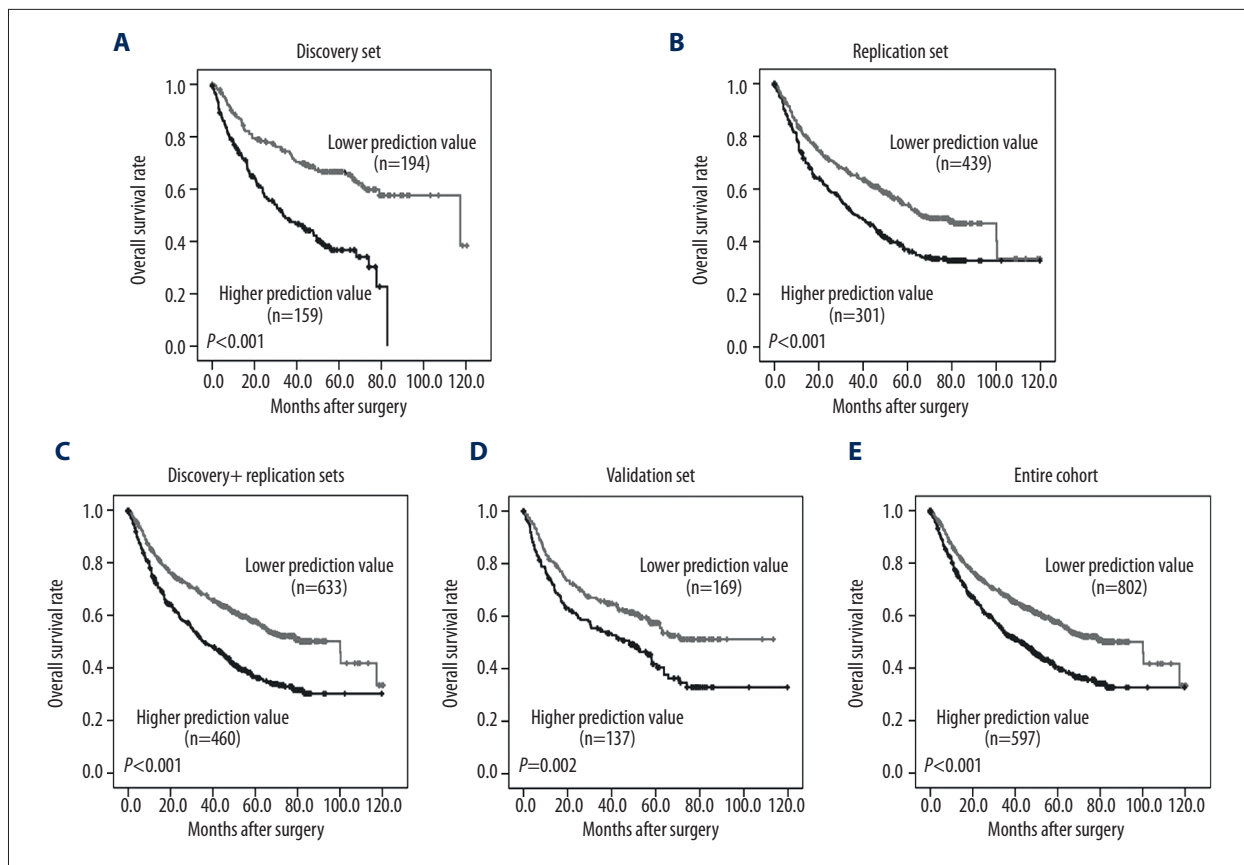
Figure 3. Kaplan-Meier estimates of survival time in the discovery set (**A**), replication set (**B**), combined discovery and replication sets (**C**), validation set (**D**), and the entire cohort (**E**) based on the cutoff prediction value from the 5-SNP prediction model. SNP, single nucleotide polymorphisms.

from the ROC curve, the sensitivity and specificity of the Nomogram survival prediction model were 69.6% and 73.2%, respectively.

Kaplan-Meier survival curves also provided strong evidence for the reliability of predictions in the combined (discovery and replication) set and validation set. Survival time was significantly longer for patients with lower prediction values than for patients with higher prediction values, with median OS times of 100.1 versus 30.8 months and 82.2 versus 22.5 months, respectively (HR=2.512, P=6.52×10^{-24}; HR=2.415, P=2.33×10^{-10}, respectively; Figure 4D, Supplementary Table 12). Similar results were obtained using the discovery set, replication set, and the entire cohort (Supplementary Figure 5, Supplementary Table 12). Furthermore, we compared RFS times between groups with lower and higher prediction values divided by the Nomogram prediction model in each sample set, and found that the Nomogram prediction model was significantly associated with the RFS (Supplementary Table 13.).

To assess the reliability of the 5-SNP prediction model and the Nomogram prediction model for the prediction of HCC

prognosis, we conducted a subgroup analysis of patients with different clinical parameters that were not included in our prognostic model, such as gender, age, serum AFP level, serum alanine aminotransferase level, liver cirrhosis status, and BCLC stage. The Nomogram prediction model showed excellent prognostic performance in all subgroups, while the 5-SNP prediction model showed very good prognostic performance in most subgroups, except in the female subgroup (Table 3).

Functional relevance of rs2431

To understand the mechanism by which genetic variants influence HCC prognosis, we performed functional analyses of the SNPs included in the prediction model. Among the 5 SNPs, rs2431 and rs34675408 were located in the 3' UTR of *FNDC3B* and the exonic region of *OSMR*, respectively. Taking the genotype frequencies into account, we focused on rs2431.

Genetic variants in the 3' UTR might influence gene expression via gain or loss of the binding site of microRNAs involved in regulation [23–27]. Therefore, we first obtained data from SNPinfo (*http://www.niehs.nih.gov/snpinfo*) and found that miR-409-3p
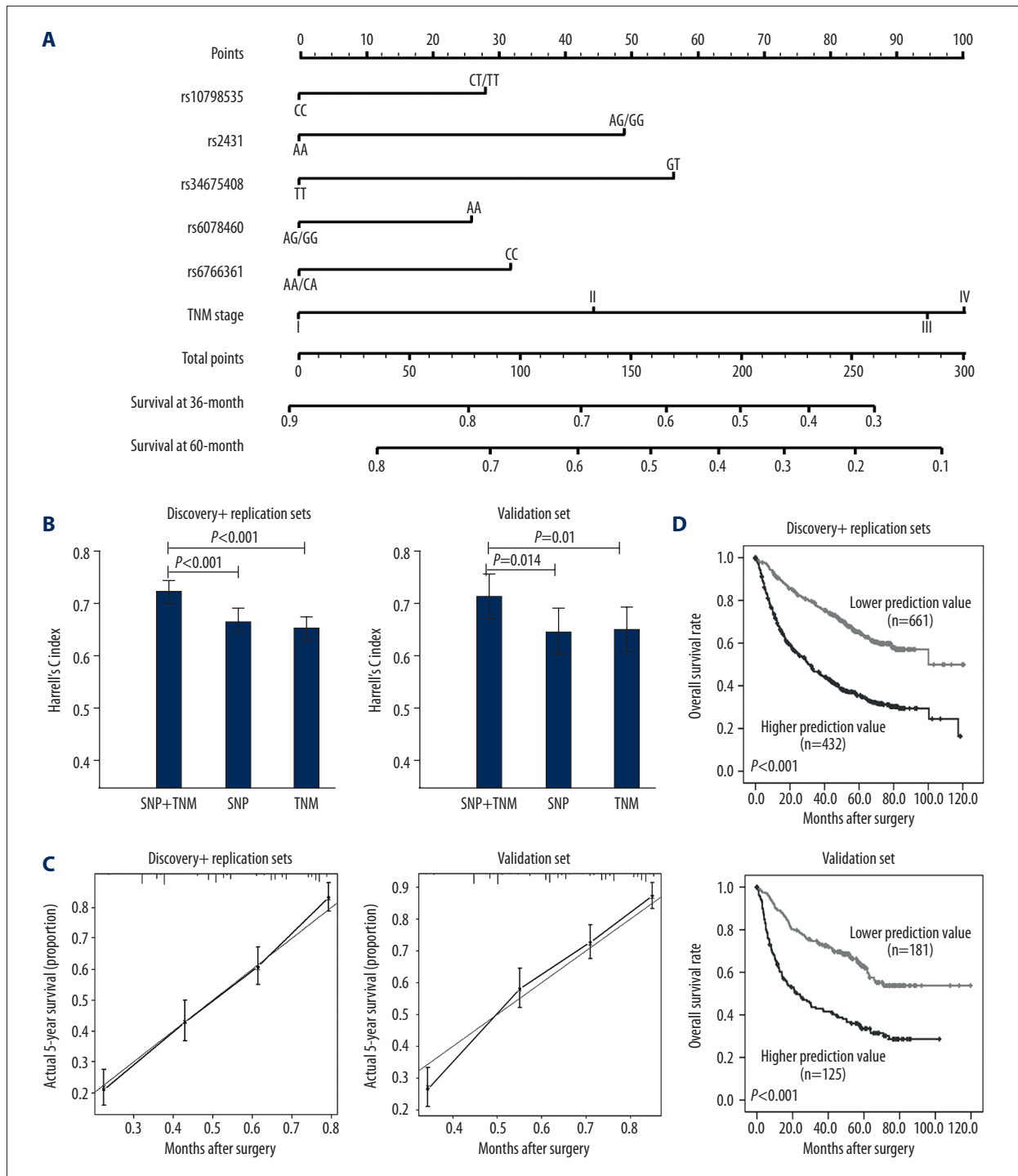
Figure 4. Prognostic value of the Nomogram survival prediction model. (**A**) Genetic variants and TNM stages upon survival in the Nomogram model for the combined set. (**B**) Comparison of Harrell's c-indexes among the Nomogram model, the SNP model, and TNM stage in the combined discovery and replication sets (**left**) and validation set (**right**). The Harrell's c-indexes of the Nomogram model were statistically higher than those of other models for each sample set. (**C**) The calibration curve for predicting patient survival at 5 years in the combined discovery and replication sets (**left**) and validation set (**right**). Nomogram-predicted probability of overall survival is plotted on the x-axis; actual overall survival is plotted on the y-axis. (**D**) Kaplan-Meier estimates for survival time in the combined discovery and replication sets (upper) and validation set (lower), based on the cut-off prediction value from the Nomogram prediction model. TNM – tumor node metastasis; SNP – single nucleotide polymorphisms.

This work is licensed under Creative Common Attribution-
NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

2461

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System]
[ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica]
[Chemical Abstracts/CAS]

Table 3. Comparison of overall survival between lower prediction value and higher prediction value groups divided by the cut-off value from the 5-SNP or Nomogram prediction model in each subgroup of the entire cohort.

Subgroup	Prediction value	5-SNP prediction model				Nomogram prediction model			
		Median time	95.0% CI for median time		P-value	Median time	95.0% CI for median time		P-value
			Lower	Upper			Lower	Upper	
Female (n=200)	Lower	100.1	50.735	149.465	0.061	100.1	82.703	117.497	1.21×10⁻⁷
	Higher	52.0	23.841	80.159		33.3	17.540	49.127	
Male (n=1196)	Lower	100.4	79.641	121.092	1.51×10⁻¹⁶	NA	NA	NA	2.38×10⁻²²
	Higher	33.7	26.919	40.548		32.5	26.121	38.945	
Age ≤55 (n=832)	Lower	117.3	57.128	177.539	2.01×10⁻⁷	NA	NA	NA	6.26×10⁻²²
	Higher	35.1	24.792	45.341		27.2	21.908	32.425	
Age >55 (n=564)	Lower	78.9	62.626	95.107	8.96×10⁻⁶	81.1	66.444	95.756	4.26×10⁻⁸
	Higher	44.5	34.714	54.286		42.9	35.863	50.004	
AFP ≤20 (n=535)	Lower	NA	NA	NA	3.20×10⁻⁴	NA	NA	NA	4.01×10⁻⁸
	Higher	56.0	NA	NA		58.0	48.056	67.944	
AFP >20 (n=859)	Lower	64.5	50.263	78.803	3.82×10⁻⁸	100.1	67.752	132.448	9.93×10⁻¹⁹
	Higher	29.2	23.884	34.45		23.5	18.450	28.617	
ALT ≤75 (n=1161)	Lower	100.1	80.601	119.599	5.39×10⁻¹²	100.1	NA	NA	1.22×10⁻²⁶
	Higher	36.3	28.233	44.3		30.8	25.299	36.234	
ALT >75 (n=207)	Lower	81.1	40.164	122.036	0.025	81.1	43.595	118.605	0.004
	Higher	31.8	19.317	44.35		29.4	10.532	48.268	
Cirrhosis (n=324)	Lower	100.1	67.891	132.309	5.26×10⁻¹²	100.1	NA	NA	3.24×10⁻²⁶
	Higher	32.4	24.954	39.912		28.0	22.763	33.237	
Non-cirrhosis (n=1064)	Lower	81.1	NA	NA	0.028	81.1	NA	NA	0.007
	Higher	52.0	43.303	60.697		58.0	45.729	70.271	
BCLC 0–A (n=478)	Lower	100.1	NA	NA	3.60×10⁻⁴	100.1	NA	NA	0.005
	Higher	64.5	NA	NA		28.7	25.319	32.014	
BCLC B–C (n=909)	Lower	53.8	43.991	63.542	5.26×10⁻⁹	61.9	41.736	81.998	5.35×10⁻⁴
	Higher	23.7	17.58	29.82		32.8	26.565	38.968	

Owing to failure of genotyping in several of the 5 SNPs for some samples, the number of subjects in each set may not equal to the total number. AFP – alpha-fetoprotein; ALT – alanine aminotransferase; BCLC – Barcelona Clinic Liver Cancer; CI – confidence interval; HBsAg – hepatitis B surface antigen; HBcAb – hepatitis B core antibody; HCV – hepatitis C virus; HR – hazard ratio; PVTT – portal vein tumor thrombus; TNM – tumor node metastasis; NA – not achieved.

is the only miRNA that exerts diverse effects when binding to different alleles. Dual-luciferase reporter assays showed that the binding between miR-409-3p and rs2431[G] was significantly weaker than binding to rs2431[A] (Figure 5A). In addition, we investigated the genotype of rs2431 in four HCC cell lines, and found the GG genotype in MHCC97-H and HCCLM3

and AA genotype in Huh7 and Hep3B. Interestingly, the overexpression of miR-409-3p induced much greater suppression of FNDC3B expression in cell lines harboring the rs2431[A] allele than in those with the rs2431[G] allele (Figure 5B). By IHC staining in 60 HCC tissues randomly selected from the 1125 genotyped samples, we found that HCC tissues with the
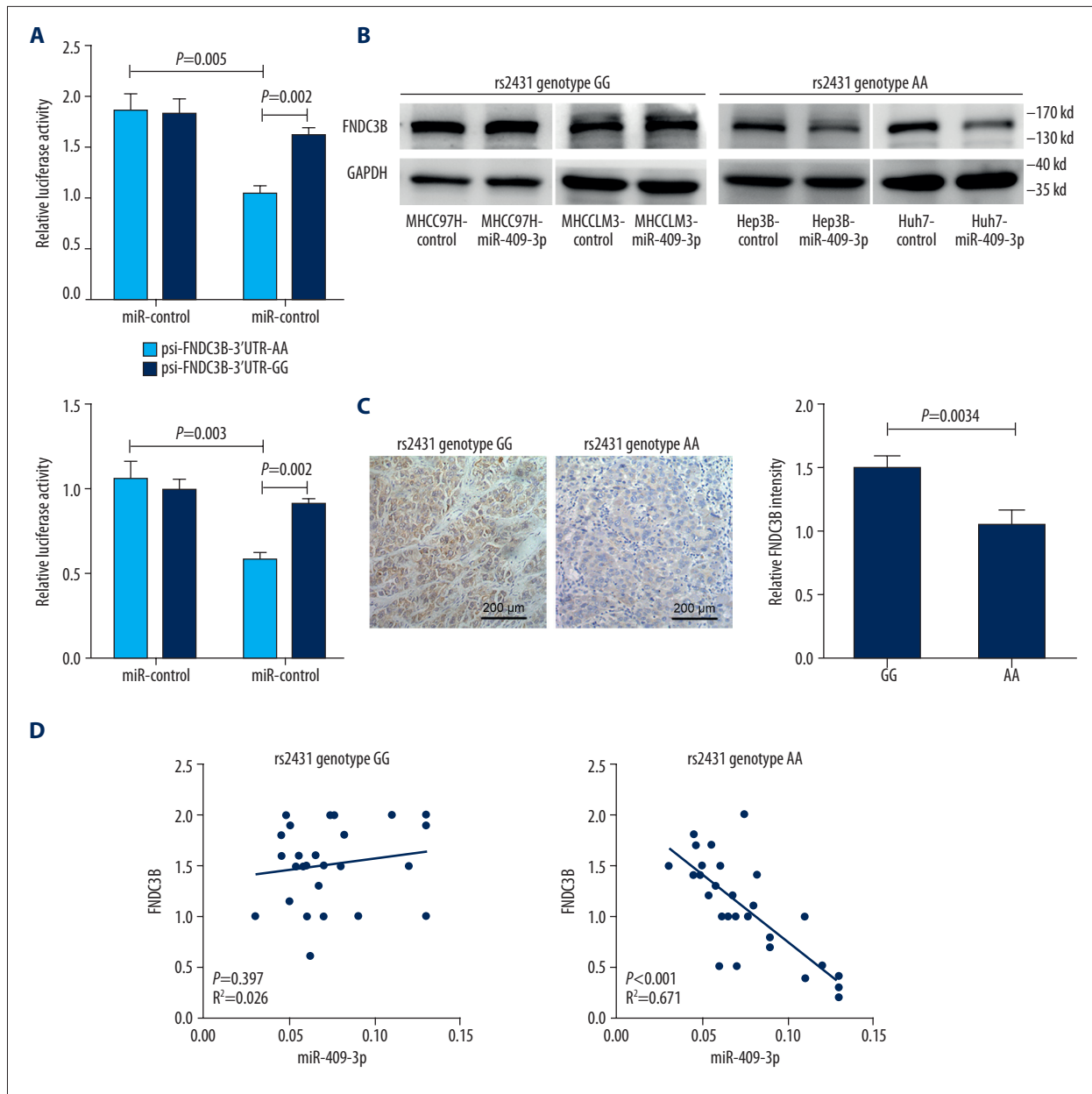
Figure 5. Functional characterization of the rs2431 SNP. (**A**) Dual-luciferase reporter assays with the FNDC3B 3'UTR containing rs2431[G] or rs2431[A] in Hep3B (upper) and HEK293T (lower) cells. The regulatory activity of miR-409-3p on the 3' UTR containing rs2431[G] was significantly weaker. (**B**) The suppression of FNDC3B expression in HCC cell lines containing the rs2431[A] allele was substantially stronger than that in HCC cell lines containing the rs2431[G] allele after miR-409-3p overexpression. (**C**) Representative images of IHC staining. Much stronger positive staining for FNDC3B was detected in HCC tissues with the rs2431[G] allele than with the rs2431[A] allele, and the signal was mainly detected in the cytoplasm. (**D**) A linear regression analysis showed a negative relationship between miR-409-3p and FNDC3B in HCC tissues with the rs2431[A] allele, but not in those with the rs2431[G] allele. Scale bar represents 200 μm. SNP – single nucleotide polymorphisms; IHC – immunohistochemistry; HCC – hepatocellular carcinoma.
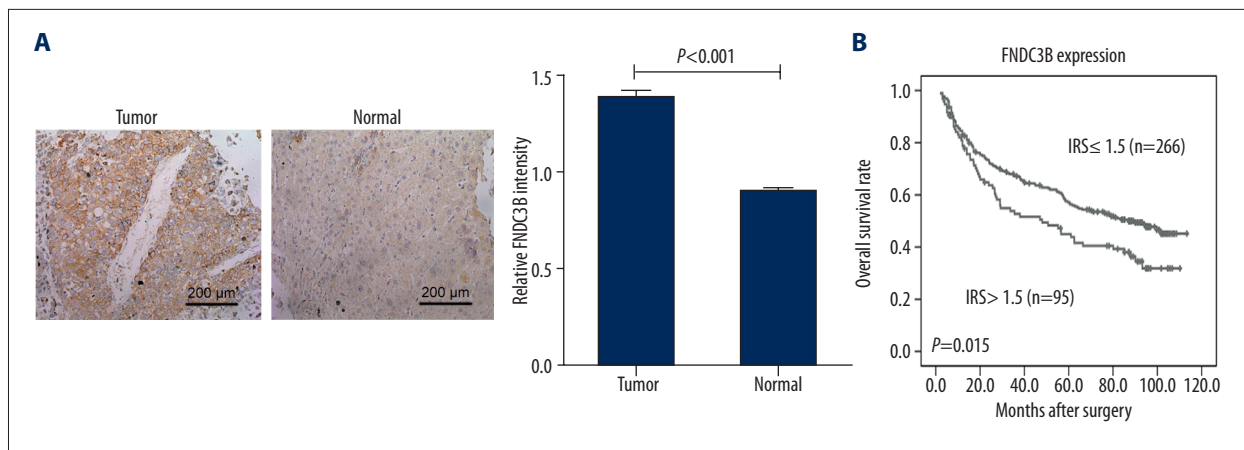
Figure 6. Functional characterization of FNDC3B. (**A**) Representative images of IHC staining. Much stronger positive staining for FNDC3B was detected in HCC tissues than in adjacent non-tumor liver tissues. (**B**) Kaplan–Meier estimate of the length of survival for individuals with HCC, stratified by FNDC3B expression. Patients with lower FNDC3B expression in tumor tissues exhibited longer survival times. Scale bar represents 200 μm. IHC – immunohistochemistry; HCC – hepatocellular carcinoma.

rs2431[G] allele had significantly higher FNDC3B IHC reaction scores (IRS) (1.500±0.097; n=30) than those of tissues with the rs2431[A] allele (1.050±0.124; n=30) (P=0.0034; Figure 5C). Moreover, a linear regression analysis showed a negative relationship between miR-409-3p and FNDC3B in HCC tissues with the rs2431[A] allele, but showed no relationship in HCC tissues with the rs2431[G] allele with a similar expression level of miR-409-3p (Figure 5D). These results suggested that *FNDC3B* is a target gene of miR-409-3p in HCC cells harboring the rs2431[A] allele but not rs2431[G].

Furthermore, the median time to recurrence after surgery for individuals with the GG or AG genotype at rs2431 (37.5 months) was substantially shorter than that of individuals with the AA genotype (62.8 months); the HR for recurrence was 1.314 (95% CI=1.042–1.657; P=0.023). A significant association was also found between the rs2431 genotype and tumor vascular invasion (P=0.046; Supplementary Figure 6).

To evaluate the role of *FNDC3B* in HCC, we also investigated the expression level of FNDC3B in tissue microarrays that contained HCC tissues from 361 patients by IHC staining. The staining density for FNDC3B in HCC tissues was significantly higher than that in adjacent non-tumor tissues (P<0.0001) (Figure 6A). IRS>1.5 was observed in 95 of 361 HCC tissues but only in 28 of the corresponding adjacent non-tumor tissues. Consistent with the rs2431 genotypes, the FNDC3B level was significantly associated with the length of survival (P=0.015; Figure 6B) and tumor vascular invasion (P=0.036). Additionally, FNDC3B levels in tumor tissues were significantly associated with the sex, age, tumor size, and TNM stage, as shown in Supplementary Table 3.

Discussion

Despite recent advancements in diagnostic techniques and multimodal treatments, HCC remains one of the most aggressive and lethal diseases [1,6]. In this study, using a genome-wide interrogation and replication approach, we identified 5 SNPs as independent genetic factors that affect the survival of patients with HCC. Using a Cox proportional hazards regression model, we established a 5-SNP model to predict survival in HCC. We validated the model using another independent sample set. We further showed that the prediction efficiency can be enhanced by incorporating TNM stage data into a new Nomogram model.

Our results indicate that individual genetic background plays an important role in both susceptibility and the outcome of HCC; this may explain observed variation in the response to the same treatment among patients with HCC. The SNP prediction model may provide a tool for the improved prediction of HCC progression and for guiding HCC monitoring or intervention, even before surgery. Additionally, the prediction model may provide new insight into the determinants of prognosis for patients with HCC. Further studies of the SNPs may identify new biological pathways affecting postoperative survival in HCC.

Since the progression of HCC is complicated, it is unlikely that a single genetic variant will have a drastic effect on clinical outcomes in HCC. However, several synergistic SNPs may have a combined effect on progression [7,8]. Therefore, using a Cox proportional hazards regression model, we established a model that combined 5 SNPs to predict the long-term survival of patients with HCC after surgery. Strikingly, the survival rate 60 months after surgery calculated using our 5-SNP panel could serve as a reliable and accurate prognostic biomarker.

We chose rs2431, which has a MAF of 0.620 and is located in the 3'UTR of *FNDC3B* for detailed investigations. *FNDC3B* is located at Chr.3q26 and is a member of the fibronectin gene family [28]. The disruption of FNDC3B inhibits adipocyte differentiation and proliferation, and the loss of FNDC3B expression causes a notable reduction in stress fiber formation, delaying cell adhesion, spread, and migration [29]. Interestingly, FNDC3B over-expression could result in the malignant transformation of mammary and kidney epithelial cells and hepatocytes [30–32]. Our functional experiments confirmed this at the biochemical level. Taken together, the rs2431 genotype may be an invaluable target for guiding anti-metastasis therapeutic strategies.

The sample sizes of the discovery set (n=367), replication set (n=758), and validation set (n=316) might not be sufficient with respect to statistical power. In the near future, the association between the 5 genetic variants and survival time for individuals with HCC should be confirmed using larger sample sizes. To identify as many significant features as possible, we defined the threshold as $10^{-3.5}$. Using this threshold, 17 of 23 SNPs chosen for replication had a *P*-value that does not meet the median threshold for SNPs in previous GWASs (usually 10^{-5}) [33,34]. Multicenter studies should be conducted to further optimize the current model.

Conclusions

In conclusion, using a genome-wide analysis, we developed survival prediction models for postoperative survival in patients with HCC. Hopefully, the performance and practicability of these models may facilitate clinical management, outcome prediction, and decision-making by physicians.

Conflict of interest

None.

Supplementary Files

Supplementary Table 1. Comparison of the allele frequencies of each SNP among discovery, replication and validation samples.

SNP	Genotype	Discovery set (N=367)	Replication set (N=758)	Validation set (N=316)
rs10893585	CC	29.1%	34.2%	30.1%
	CT	48.4%	47.3%	52.3%
	TT	22.5%	18.5%	17.6%
	HWE	0.583	0.401	0.277
rs2431	AA	12.8%	17.7%	12.7%
	AG	45.5%	43.1%	44.8%
	GG	41.7%	39.2%	42.5%
	HWE	0.892	0.009	0.757
rs34675408	GT	7.7%	6.4%	7.8%
	TT	92.3%	93.6%	92.2%
	HWE	0.446	0.365	0.475
rs6078460	AA	57.4%	52.9%	56.2%
	AG	37.0%	38.6%	38.2%
	GG	5.6%	8.4%	5.6%
	HWE	0.808	0.310	0.617
rs6766361	AA	6.3%	7.9%	5.2%
	CA	39.5%	39.1%	39.5%
	CC	54.2%	52.9%	55.2%
	HWE	0.639	0.624	0.341

SNP – single nucleotide polymorphisms; HWE – Hardy-Weinberg equilibrium.

Supplementary Table 2. Univariate analyses of factors associated with overall survival of HCC in samples combined discovery and replication sets (n=1125).

Variable	P-value	HR	95% CI	
			Lower	Upper
Gender (Male *vs.* Female)	0.483	1.088	0.860	1.376
Age, years (>55 *vs.* ≤55)	0.579	0.954	0.808	1.126
AFP, ng/ml (>20 *vs.* ≤20)	$5.36×10^{-8}$	1.636	1.370	1.954
ALT, ng/ml (>75 *vs.* ≤75)	0.177	1.164	0.934	1.451
Tumor size, cm (>5 *vs.* ≤5)	$5.64×10^{-23}$	2.333	1.972	2.760
Tumor number (multiple *vs.* single)	0.021	1.290	1.040	1.600
Vascular invasion (yes *vs.* no)	$8.96×10^{-30}$	2.594	2.200	3.059
PVTT (yes *vs.* no)	$1.33×10^{-40}$	4.322	3.486	5.359
Extrahepatic metastasis (yes *vs.* no)	$3.33×10^{-6}$	2.305	1.621	3.278
TNM stage (III–IV *vs.* I–II)	$2.52×10^{-34}$	2.973	2.496	3.541
BCLC stage (B–C *vs.* 0–A)	$2.42x10^{-21}$	2.569	2.114	3.122
Child-Pugh stage (B *vs.* A)	$1.91×10^{-6}$	2.299	1.632	3.240
Tumor differentiation (III–IV *vs.* I–II)	$1.19×10^{-6}$	1.532	1.290	1.820
Liver cirrhosis (yes *vs.* no)	0.099	0.843	0.688	1.033
Tumor encapsulation (none *vs.* complete)	$1.70×10^{-7}$	1.574	1.328	1.866
HBsAg (negative *vs.* positive)	0.544	1.073	0.855	1.347
HBcAb (negative *vs.* positive)	0.540	0.909	0.670	1.233
Anti-HCV (negative *vs.* positive)	0.354	1.423	0.675	3.000

HCC – hepatocellular carcinoma; AFP – alpha fetoprotein; ALT – alanine aminotransferase; BCLC – Barcelona Clinic Liver Cancer; CI – confidence interval; HBsAg – hepatitis B surface antigen; HBcAb – hepatitis B core antibody; HCV – hepatitis C virus; HR – hazard ratio; PVTT – portal vein tumor thrombus; TNM – tumor node metastasis.

Supplementary Table 3. Patient characteristics and distributions of immunoreactive scoring (IRS) by immunohistochemical staining of FNDC3B in 361 HCC tumor tissues on tissue microarrays.

Variable	Number	IRS ≤1.5	IRS>1.5	P-value*
Total				
Deceased	193	133	60	0.027
Alive	168	133	35	
MST, months	76.2	87.1	48.0	
Gender				
Male	298	211	87	0.007
Female	63	55	8	
Age, years				
≤55	231	162	69	0.041
>55	130	104	26	

Variable	Number	IRS ≤1.5	IRS>1.5	P-value*
AFP, ng/ml				
≤20	135	101	34	0.706
>20	226	165	61	
ALT, ng/ml				
≤75	316	234	82	0.675
>75	45	32	13	
Tumor size, cm				
≤5	276	214	62	**0.003**
>5	85	52	33	
Tumor number				
Single	339	249	90	0.693
Multiple	22	17	5	
Vascular invasion				
Yes	117	78	39	0.036
No	244	188	56	
Extrahepatic metastasis				
Yes	5	5	0	0.178
No	356	261	95	
TNM stage				
I–II	296	227	69	**0.006**
III–IV	65	39	26	
Tumor differentiation				
I–II	272	205	67	0.184
III–IV	88	60	28	
Liver cirrhosis				
Yes	79	63	16	0.170
No	279	201	78	
Tumor encapsulation				
None	186	136	50	0.801
Complete	175	130	45	
HBsAg				
Positive	331	244	87	0.801
Negative	29	22	7	
HBcAb				
Positive	340	249	91	0.244
Negative	20	17	3	

Variable	Number	IRS $\leq$1.5	IRS>1.5	*P*-value*
Anti-HCV				
Positive	2	1	1	0.444
Negative	356	263	93	

Owing to a lack of information for some samples, the number of participants in each category may not add up to the total number.
HCC – hepatocellular carcinoma; AFP – alpha-fetoprotein; ALT – alanine aminotransferase; CI – confidence interval; HBsAg – hepatitis B surface antigen; HBcAb – hepatitis B core antibody; HCV – hepatitis C virus; HR – hazard ratio; TNM – tumor node metastasis.
* χ^2 – test between different IRS groups and clinical variables.

Supplementary Table 4. The information of 23 survival related SNPs selected from discovery set.

SNP	*P*-value of multiple linear regression	*P*-value of Kaplan-Meier estimate	Alleles DBSNP	Function	Amino acids	Gene list	Chr.
rs10481370	1.67×10^{-6}	1.42×10^{-4}	A/G	Intron	None	CSMD1	8
rs6078460	1.72×10^{-6}	1.70×10^{-3}	A/G	Intergenic	None	None	20
rs11600756	2.93×10^{-6}	1.10×10^{-4}	C/T	Intron	None	KIRREL3	11
rs11639771	5.08×10^{-6}	4.91×10^{-5}	C/T	Intergenic	None	None	16
rs7579933	6.60×10^{-6}	4.52×10^{-3}	A/G	Intron	None	AAK1	2
rs10790836	9.10×10^{-6}	3.42×10^{-3}	C/T	Intron	None	KIRREL3	11
rs16910200	1.20×10^{-5}	9.37×10^{-4}	C/T	Intergenic	None	None	9
rs1893185	1.18×10^{-5}	1.58×10^{-4}	C/T	Intron	None	PKNOX2	11
rs6766361	1.65×10^{-5}	2.42×10^{-4}	A/C	Intergenic	None	None	3
rs913493	2.30×10^{-5}	4.54×10^{-5}	A/G	Intron	None	PCDH9	13
rs12897	3.74×10^{-5}	6.89×10^{-4}	A/G	Utr-3	None	FNDC3B	3
rs2239910	4.63×10^{-5}	7.46×10^{-7}	A/C	Utr-3	None	SARM1, SLC46A1	17
rs584368	5.19×10^{-5}	4.63×10^{-6}	C/T	Intergenic	None	None	11
rs12504893	6.14×10^{-5}	5.00×10^{-4}	A/C	Intron	None	FSTL5	4
rs1034756	6.54×10^{-5}	4.27×10^{-4}	A/G	Intergenic	None	None	7
rs4617227	8.49×10^{-5}	3.06×10^{-4}	A/G	Intergenic	None	None	9
rs10893585	9.29×10^{-5}	1.39×10^{-4}	A/G	Intron	None	KIRREL3	11
rs3096380	9.63×10^{-5}	1.89×10^{-5}	A/G	Missense	LEU, PHE	FTSJD1	16
rs3748735	9.84×10^{-5}	9.16×10^{-4}	C/T	Missense	ALA, THR	PLXNA2	1
rs2431	2.11×10^{-4}	5.42×10^{-4}	A/G	Utr-3	None	FNDC3B	3
rs7581876	2.41×10^{-4}	4.62×10^{-9}	A/G	Utr-3	None	AAK1	2
rs34675408	2.59×10^{-4}	1.92×10^{-4}	G/T	Missense	HIS, GLN	OSMR	5
rs6499496	2.85×10^{-4}	1.91×10^{-5}	A/G	Utr-3	None	FTSJD1	16

SNP – single nucleotide polymorphisms; UTR – untranslated region.

Supplementary Table 5. Associations between 23 SNPs and time to recurrence in the discovery, replication and combined samples with HCC.

SNP	*P*-value (multiple linear regression analysis)		
	Discovery set (n=367)	Replication set (n=758)	Combined discovery and replication sets (n=1125)
rs10481370	$6.82×10^{-6}$	0.941	0.021
rs6078460	**0.012**	**0.139**	**$2.28×10^{-8}$**
rs11600756	**$1.02×10^{-4}$**	**0.029**	**$5.92×10^{-4}$**
rs11639771	$7.33×10^{-4}$	0.278	0.133
rs7579933	$2.21×10^{-3}$	0.574	0.460
rs10790836	**$8.35×10^{-4}$**	**0.031**	**$1.53×10^{-3}$**
rs16910200	$2.59×10^{-5}$	0.703	0.340
rs1893185	0.016	0.108	0.025
rs6766361	**$3.71×10^{-3}$**	**0.120**	**0.011**
rs913493	$2.39×10^{-5}$	0.989	$4.61×10^{-14}$
rs12897	0.020	0.153	0.028
rs2239910	$1.50×10^{-3}$	0.182	$7.91×10^{-3}$
rs584368	$1.10×10^{-4}$	0.392	$8.13×10^{-10}$
rs12504893	$5.18×10^{-5}$	0.629	0.141
rs1034756	$9.74×10^{-5}$	0.875	0.585
rs4617227	$2.89×10^{-5}$	0.655	0.080
rs10893585	**$1.59×10^{-4}$**	**0.036**	**$1.58×10^{-3}$**
rs3096380	$2.34×10^{-3}$	0.104	$8.10×10^{-12}$
rs3748735	$6.43×10^{-3}$	0.670	0.469
rs2431	**0.023**	**0.059**	**$3.34×10^{-3}$**
rs7581876	$9.93×10^{-3}$	0.904	$6.64×10^{-3}$
rs34675408	**$4.96×10^{-3}$**	**0.051**	**$1.60×10^{-22}$**
rs6499496	$8.07×10^{-3}$	0.752	0.410

SNP – single nucleotide polymorphisms; HCC – hepatocellular carcinoma.

Supplementary Table 6. Multivariate Cox regression analyses of each SNP (rs10893585, rs2431, rs34675408, rs6078460, and rs6766361) for overall survival of HCC samples combined discovery and replication sets (n=1125).

Variable	*P*-value	HR	95.0% CI for HR	
			Lower	Upper
For rs10893585				
AFP, ng/ml (>20 *vs.* ≤20)	0.019	1.249	1.037	1.504
Tumor size, cm (>5 *vs.* ≤5)	$9.63×10^{-10}$	2.162	1.689	2.768
Tumor number (multiple *vs.* single)	0.010	1.340	1.073	1.675
Vascular invasion (yes *vs.* no)	$7.72×10^{-10}$	1.987	1.596	2.473

Variable	*P*-value	HR	95.0% CI for HR	
			Lower	Upper
Extrahepatic metastasis (yes *vs.* no)	5.35×10^{-3}	1.699	1.170	2.467
Tumor differentiation (III–IV *vs.* I–II)	0.011	1.262	1.056	1.508
Tumor encapsulation (none *vs.* complete)	1.25×10^{-4}	1.412	1.184	1.683
Genotype (CC *vs.* CT+TT)	**0.019**	0.802	0.667	0.964
For rs2431				
AFP, ng/ml (>20 *vs.* ≤20)	0.032	1.227	1.018	1.478
Tumor size, cm (>5 *vs.* ≤5)	3.35×10^{-10}	2.226	1.734	2.858
Tumor number (multiple *vs.* single)	7.16×10^{-3}	1.360	1.087	1.702
Vascular invasion (yes *vs.* no)	4.03×10^{-9}	1.938	1.555	2.416
Extrahepatic metastasis (yes *vs.* no)	9.46×10^{-3}	1.639	1.129	2.381
Tumor differentiation (III–IV *vs.* I–II)	9.54×10^{-3}	1.267	1.059	1.516
Tumor encapsulation (none *vs.* complete)	7.14×10^{-5}	1.433	1.200	1.711
Genotype (AA *vs.* AG+GG)	**8.61×10^{-3}**	0.713	0.554	0.918
For rs34675408				
AFP, ng/ml (>20 *vs.* ≤20)	0.046	1.210	1.004	1.458
Tumor size, cm (>5 *vs.* ≤5)	1.47×10^{-9}	2.145	1.675	2.747
Tumor number (multiple *vs.* single)	0.016	1.315	1.053	1.644
Vascular invasion (yes *vs.* no)	2.29×10^{-10}	2.033	1.632	2.531
Extrahepatic metastasis (yes *vs.* no)	4.29×10^{-3}	1.722	1.186	2.499
Tumor differentiation (III–IV *vs.* I–II)	4.45×10^{-3}	1.294	1.084	1.547
Tumor encapsulation (none *vs.* complete)	1.04×10^{-4}	1.417	1.188	1.689
Genotype (GT *vs.* TT)	**2.37×10^{-5}**	1.863	1.396	2.486
For rs6078460				
AFP, ng/ml (>20 *vs.* ≤20)	0.051	1.205	0.999	1.452
Tumor size, cm (>5 *vs.* ≤5)	7.91×10^{-11}	2.289	1.783	2.938
Tumor number (multiple *vs.* single)	3.15×10^{-3}	1.402	1.120	1.754
Vascular invasion (yes *vs.* no)	6.03×10^{-10}	2.001	1.607	2.493
Extrahepatic metastasis (yes *vs.* no)	4.20×10^{-3}	1.726	1.188	2.507
Tumor differentiation (III–IV *vs.* I–II)	0.015	1.248	1.044	1.493
Tumor encapsulation (none *vs.* complete)	2.85×10^{-4}	1.393	1.165	1.666
Genotype (AA *vs.* AG+GG)	**8.95×10^{-3}**	1.258	1.059	1.494
For rs6766361				
AFP, ng/ml (>20 *vs.* ≤20)	0.034	1.224	1.016	1.474
Tumor size, cm (>5 *vs.* ≤5)	4.17×10^{-10}	2.219	1.728	2.849
Tumor number (multiple *vs.* single)	4.19×10^{-3}	1.387	1.109	1.736
Vascular invasion (yes *vs.* no)	3.66×10^{-9}	1.935	1.554	2.410
Extrahepatic metastasis (yes *vs.* no)	9.68×10^{-3}	1.635	1.127	2.373
Tumor differentiation (III–IV *vs.* I–II)	9.25×10^{-3}	1.268	1.060	1.517
Tumor encapsulation (none *vs.* complete)	4.66×10^{-5}	1.444	1.210	1.724
Genotype (AA+CA *vs.* CC)	**1.04×10^{-3}**	0.751	0.633	0.891

SNP – single nucleotide polymorphisms; HCC – hepatocellular carcinoma; AFP – alpha-fetoprotein; CI – confidence interval; HR – hazard ratio.

Supplementary Table 7. genetic variants upon survival in the equation of Cox proportional hazard regression model of combined set.

| SNP | Cox coefficient | P-value | HR | 95.0% CI for HR | |
				Lower	Upper
rs10893585	−0.268	4.28×10^{-3}	0.765	0.637	0.919
rs2431	−0.454	0.019	0.635	0.493	0.818
rs34675408	0.570	3.10×10^{-4}	1.768	1.326	2.356
rs6078460	0.226	0.010	1.254	1.060	1.484
rs6766361	−0.288	2.50×10^{-3}	0.750	0.633	0.888

SNP – single nucleotide polymorphisms; CI – confidence interval; HR – hazard ratio.

Supplementary Table 8. Comparison of overall survival between lower prediction value and higher prediction value groups divided by the Cutoff value from the 5-SNP prediction model.

| Sample | Prediction value | Median time | Std. error | 95.0% CI for median time | | P-value | HR | 95.0% CI for HR | |
				Lower	Upper			Lower	Upper
Discovery set (n=353)	Lower	117.3	31.846	54.915	179.751	1.10×10^{-7}	2.363	1.720	3.246
	Higher	33.7	6.325	21.270	46.063				
Replication set (n=740)	Lower	65.6	7.385	51.126	80.074	2.62×10^{-5}	1.529	1.254	1.863
	Higher	37.4	4.399	28.745	45.988				
Combined set (n=1093)	Lower	100.1	12.228	76.134	124.066	1.45×10^{-10}	1.725	1.460	2.038
	Higher	35.5	3.575	28.493	42.507				
Validation set (n=306)	Lower	80.2	8.702	63.11	97.223	0.002	1.481	1.078	2.033
	Higher	48.0	9.695	29.032	67.035				
Entire cohort (n=1399)	Lower	81.1	5.868	69.599	92.601	3.96×10^{-9}	1.577	1.355	1.836
	Higher	42.2	3.522	35.297	49.103				

Owing to failure of genotyping in several of the 5 SNPs for some samples, the number of participants in each set may not equal to the total number. SNP – single nucleotide polymorphisms; CI – confidence interval; HR – hazard ratio.

Supplementary Table 9. Multivariate Cox regression analyses of 5-SNP prediction model for overall survival of HCC patients in each set.

| Variable | P-value | HR | 95.0% CI for HR | |
			Lower	Upper
For discovery set				
AFP, ng/ml (>20 vs. ≤20)	0.547	1.117	0.779	1.604
Tumor size, cm (>5 vs. ≤5)	8.05×10^{-5}	2.613	1.621	4.212
Tumor number (multiple vs. single)	0.098	1.403	0.940	2.093
Vascular invasion (yes vs. no)	2.10×10^{-3}	1.909	1.264	2.881
Extrahepatic metastasis (yes vs. no)	0.615	1.185	0.611	2.299
Tumor differentiation (III–IV vs. I–II)	0.047	1.409	1.004	1.976
Tumor encapsulation (none vs. complete)	1.60×10^{-3}	1.748	1.236	2.473
Prediction value (higher vs. lower)	**9.13×10^{-8}**	2.452	1.764	3.407

Variable	P-value	HR	95.0% CI for HR	
			Lower	Upper
For replication set				
AFP, ng/ml (>20 *vs.* ≤20)	0.407	1.097	0.881	1.367
Tumor size, cm (>5 *vs.* ≤5)	$1.50×10^{-7}$	2.319	1.694	3.174
Tumor number (multiple *vs.* single)	0.249	1.193	0.883	1.611
Vascular invasion (yes *vs.* no)	$6.04×10^{-5}$	1.760	1.335	2.319
Extrahepatic metastasis (yes *vs.* no)	0.032	1.691	1.045	2.735
Tumor differentiation (III–IV *vs.* I–II)	0.039	1.265	1.012	1.582
Tumor encapsulation (none *vs.* complete)	0.317	1.117	0.899	1.389
Prediction value (higher *vs.* lower)	**0.018**	1.294	1.045	1.601
For combined discovery and replication sets				
AFP, ng/ml (>20 *vs. vs.*20)	0.058	1.196	0.994	1.440
Tumor size, cm (>5 *vs. vs.*5)	$3.20×10^{-9}$	1.879	1.525	2.315
Tumor number (multiple *vs.* single)	$2.54×10^{-3}$	1.426	1.133	1.797
Vascular invasion (yes *vs.* no)	$6.52×10^{-5}$	1.611	1.275	2.036
Extrahepatic metastasis (yes *vs.* no)	0.013	1.621	1.107	2.375
Tumor differentiation (III–IV *vs.* I–II)	0.036	1.162	1.010	1.336
Tumor encapsulation (none *vs.* complete)	0.015	1.253	1.044	1.503
Prediction value (higher *vs.* lower)	**$1.80×10^{-7}$**	1.597	1.339	1.903
For validation set				
AFP, ng/ml (>20 *vs.* ≤20)	$5.56×10^{-3}$	1.672	1.163	2.406
Tumor size, cm (>5 *vs.* ≤5)	$1.80×10^{-3}$	2.236	1.349	3.707
Tumor number (multiple *vs.* single)	0.518	1.151	0.751	1.763
Vascular invasion (yes *vs.* no)	$3.24×10^{-7}$	2.892	1.924	4.346
Extrahepatic metastasis (yes *vs.* no)	$1.75×10^{-4}$	3.148	1.729	5.730
Tumor differentiation (III–IV *vs.* I–II)	0.173	1.288	0.895	1.855
Tumor encapsulation (none *vs.* complete)	0.315	1.183	0.852	1.643
Prediction value (higher *vs.* lower)	**$4.65×10^{-3}$**	1.603	1.156	2.223
For entire cohort				
AFP, ng/ml (>20 *vs.* ≤20)	$2.53×10^{-3}$	1.288	1.093	1.518
Tumor size, cm (>5 *vs.* ≤5)	$1.23×10^{-12}$	1.977	1.638	2.386
Tumor number (multiple *vs.* single)	$1.44×10^{-3}$	1.390	1.135	1.702
Vascular invasion (yes *vs.* no)	$1.83×10^{-10}$	1.910	1.565	2.330
Extrahepatic metastasis (yes *vs.* no)	$5.77×10^{-5}$	1.923	1.398	2.644
Tumor differentiation (III–IV *vs.* I–II)	$8.23×10^{-3}$	1.184	1.045	1.342
Tumor encapsulation (none *vs.* complete)	$7.83×10^{-3}$	1.240	1.058	1.452
Prediction value (higher *vs.* lower)	**$7.57×10^{-9}$**	1.573	1.349	1.834

SNP – single nucleotide polymorphisms; HCC –hepatocellular carcinoma; AFP – alpha-fetoprotein; CI – confidence interval; HR – hazard ratio.

Supplementary Table 10. Comparison of recurrence-free survival between lower prediction value and higher prediction value groups divided by the Cutoff value from the 5-SNP prediction model.

Sample	Prediction value	Median time	Std. error	95.0% CI for median time		P-value	HR	95.0% CI for HR	
				Lower	Upper			Lower	Upper
Discovery set (n=353)	Lower	38.6	9.849	19.262	57.872	0.001	1.561	1.197	2.036
	Higher	16.4	3.286	9.926	22.808				
Replication set (n=740)	Lower	34.5	4.122	26.453	42.613	5.54×10^{-4}	1.369	1.146	1.637
	Higher	16.9	2.346	12.336	21.531				
Combined set (n=1093)	Lower	37.0	3.981	29.198	44.802	3.38×10^{-6}	1.419	1.224	1.645
	Higher	16.6	1.711	13.246	19.954				
Validation set (n=306)	Lower	39.0	5.365	28.485	49.515	0.032	1.382	1.029	1.856
	Higher	19.1	4.027	11.208	26.992				
Entire cohort (n=1399)	Lower	37.3	3.570	30.270	44.264	3.63×10^{-7}	1.409	1.235	1.608
	Higher	17.0	1.599	13.865	20.135				

Owing to failure of genotyping in several of the 5 SNPs for some samples, the number of participants in each set may not equal to the total number. SNP – single nucleotide polymorphisms; CI – confidence interval; HR – hazard ratio.

Supplementary Table 11. Genetic variants and the TNM stages upon survival in the nomogram model of combined set.

SNP	Nomogram coefficient	P-value	HR	95.0% CI for HR	
				Lower	Upper
rs10893585	−0.255	6.21×10^{-3}	0.775	0.645	0.930
rs2431	−0.450	4.86×10^{-4}	0.638	0.495	0.821
rs34675408	0.506	5.66×10^{-4}	1.658	1.244	2.211
rs6078460	0.228	8.16×10^{-3}	1.256	1.061	1.487
rs6766361	−0.296	6.25×10^{-4}	0.744	0.628	0.881
TNM stage (II)	0.412	7.40×10^{-4}	1.510	1.229	1.856
TNM stage (III)	0.877	6.73×10^{-9}	2.404	1.939	2.980
TNM stage (IV)	0.926	8.15×10^{-7}	2.524	1.723	3.698

SNP – single nucleotide polymorphisms; CI – confidence interval; HR – hazard ratio; TNM – tumor node metastasis.

Supplementary Table 12. Comparison of overall survival between lower prediction value and higher prediction value groups divided by the Cutoff value from the nomogram prediction model.

Sample	Prediction value	Median time	Std. error	95.0% CI for median time		P-value	HR	95.0% CI for HR	
				Lower	Upper			Lower	Upper
Discovery set (n=353)	Lower	NA	NA	NA	NA	1.39×10^{-10}	3.311	2.297	4.773
	Higher	32.8	4.322	24.296	41.238				
Replication set (n=740)	Lower	100.1	NA	NA	NA	3.25×10^{-15}	2.300	1.870	2.829
	Higher	30.2	3.909	22.572	37.895				
Combined set (n=1093)	Lower	100.1	NA	NA	NA	6.52×10^{-24}	2.512	2.100	3.005
	Higher	30.8	2.801	25.277	36.257				

| Sample	Prediction value	Median time	Std. error	95.0% CI for median time		P-value	HR	95.0% CI for HR	
				Lower	Upper			Lower	Upper
Validation set (n=306)	Lower	82.2	NA	NA	NA	2.33×10^{-10}	2.415	1.823	3.201
	Higher	22.5	5.31	12.093	32.907				
Entire cohort (n=1399)	Lower	100.1	NA	NA	NA	5.04×10^{-22}	2.503	2.176	2.886
	Higher	28.1	2.856	22.535	33.731				

Owing to failure of genotyping in several of the 5 SNPs for some samples, the number of participants in each set may not equal to the total number. CI – confidence interval; HR – hazard ratio; NA – not achieved.

Supplementary Table 13. Comparison of recurrence-free survival between lower prediction value and higher prediction value groups divided by the Cutoff value from the nomogram prediction model.

Sample	Prediction value	Median time	Std. error	95.0% CI for median time		P-value	HR	95.0% CI for HR	
				Lower	Upper			Lower	Upper
Discovery set (n=353)	Lower	61.4	7.906	45.904	76.896	1.35×10^{-6}	1.997	1.508	2.643
	Higher	14.0	2.239	9.644	18.422				
Replication set (n=740)	Lower	55.6	5.099	45.607	65.593	3.73×10^{-15}	2.079	1.732	2.494
	Higher	12.5	1.087	10.403	14.664				
Combined set (n=1093)	Lower	56.0	4.332	47.543	64.524	4.10×10^{-20}	2.044	1.755	2.381
	Higher	13.0	1.192	10.664	15.336				
Validation set (n=306)	Lower	56.2	7.680	41.180	71.286	2.57×10^{-4}	1.751	1.297	2.365
	Higher	16.8	1.877	13.087	20.446				
Entire cohort (n=1399)	Lower	56.0	3.944	48.303	63.764	5.42×10^{-23}	1.984	1.732	2.273
	Higher	14.0	1.065	11.913	16.087				

Owing to failure of genotyping in several of the 5 SNPs for some samples, the number of participants in each set may not equal to the total number. CI – confidence interval; HR – hazard ratio.
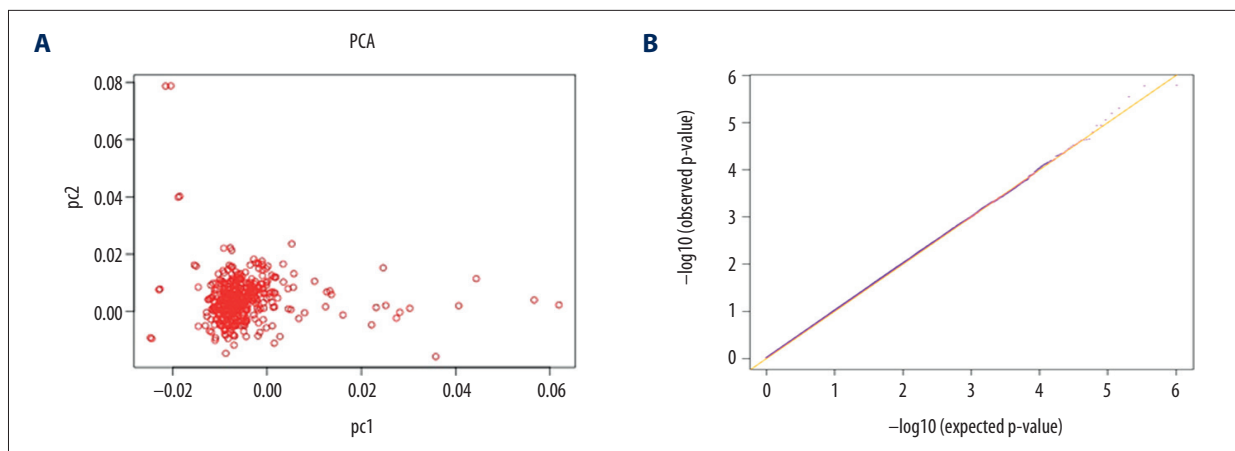
Supplementary Figure 1. Principal component analysis (PCA) and quantile-quantile plot. (**A**) Plot of the first 2 components derived from a principal component analysis implemented in the software package GCTA, PCA showed little evidence of population stratification in our study populations. (**B**) Quantile-quantile plot of observed *P*-values. The purple crosses represent the distribution of *P*-values for the association of 630 573 autosomal SNPs in 367 patients with HCC. The observed versus expected χ^2 test statistics showed no evidence for inflation of χ^2 tests. SNP, single nucleotide polymorphisms.
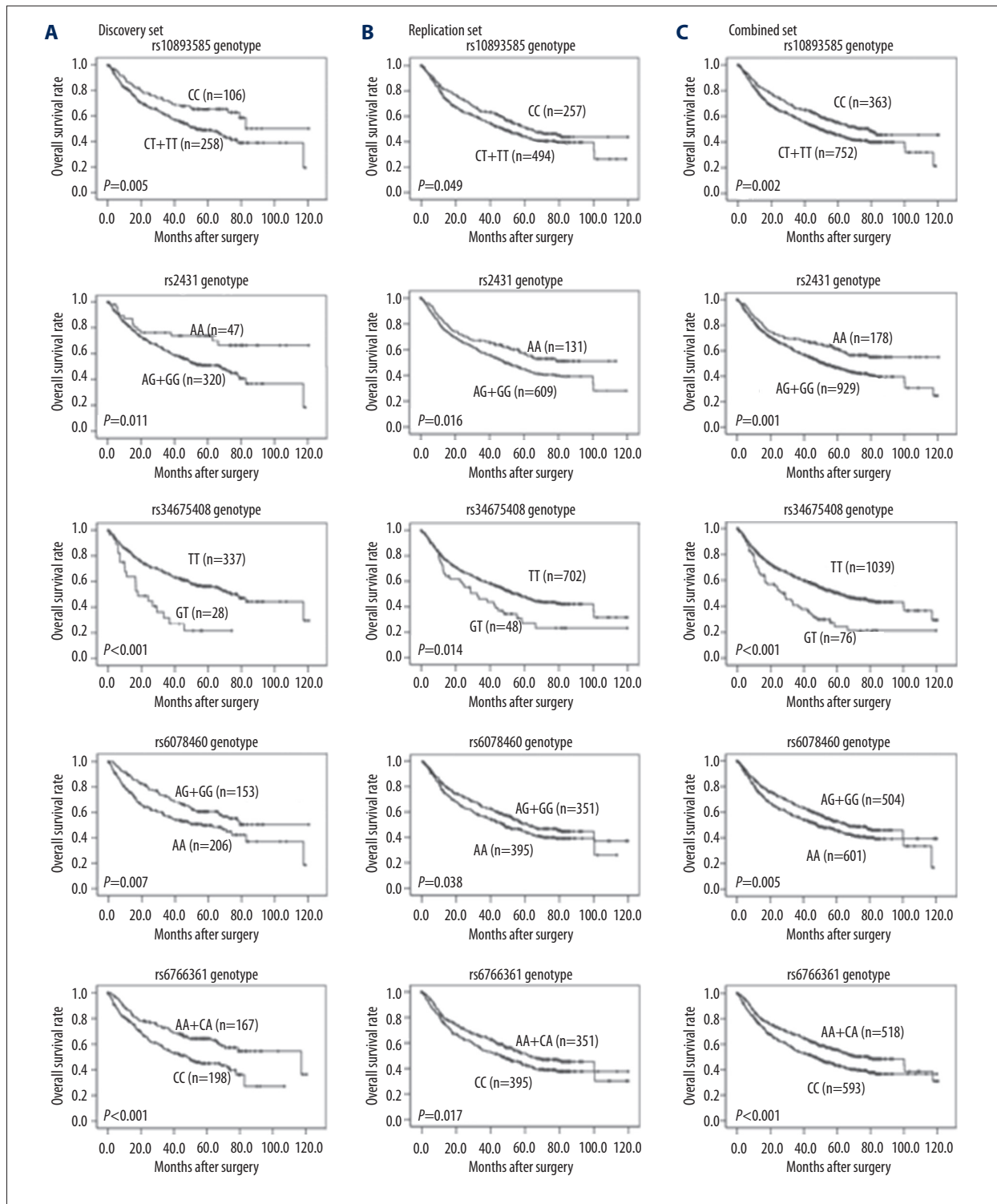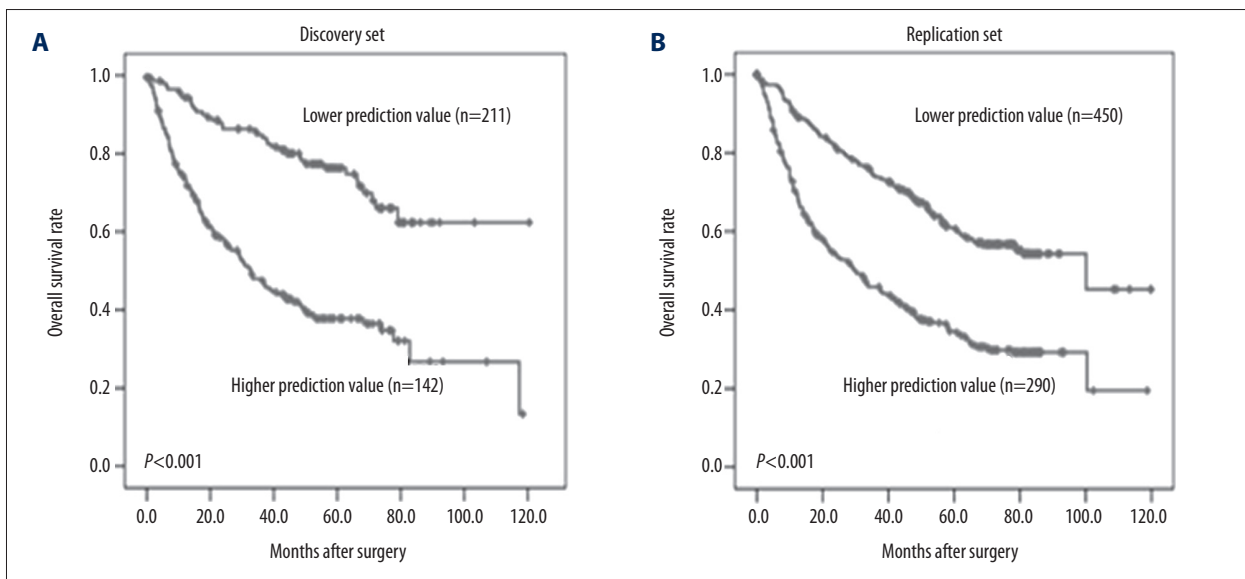
Supplementary Figure 2. Kaplan-Meier curves of survival times in HCC patients according to different genotypes at rs10893585, rs2431, rs34675408, rs6078460 and rs6766361. (**A**) Discovery set, (**B**) replication set, (**C**) combined set. Because of genotyping failure for some samples, the number of participants in each category may not add up to the total number. HCC – hepatocellular carcinoma.
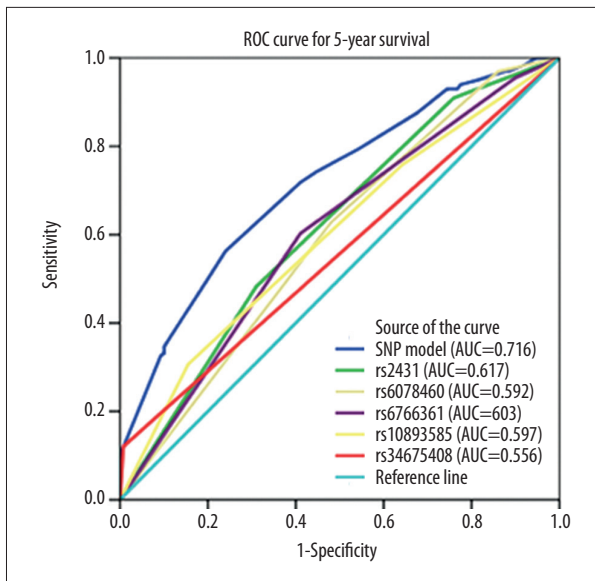
Supplementary Figure 3. ROC analyses of the SNP survival prediction system in the time point of 60 months (5-year) after surgery of HCC patients. ROC – receiver operating characteristics; SNP – single nucleotide polymorphisms; HCC – hepatocellular carcinoma.
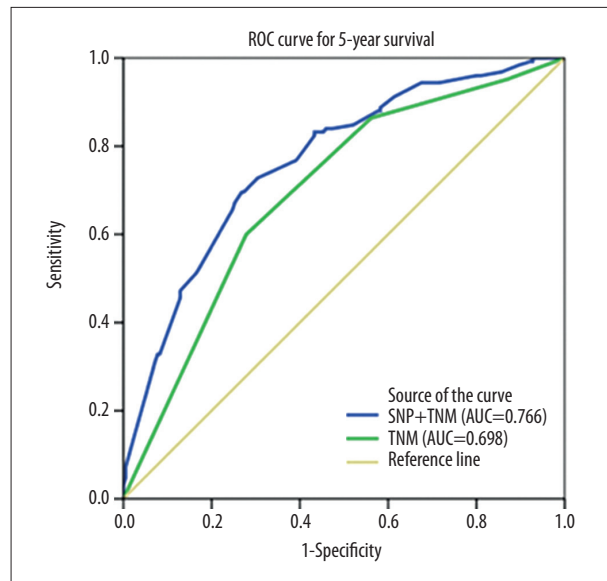
Supplementary Figure 4. ROC analyses of the Nomogram survival prediction model and TNM staging in the time point of 60 months (5-year) after surgery of HCC patients. ROC – receiver operating characteristics; SNP – single nucleotide polymorphisms; TNM – tumor node metastasis.
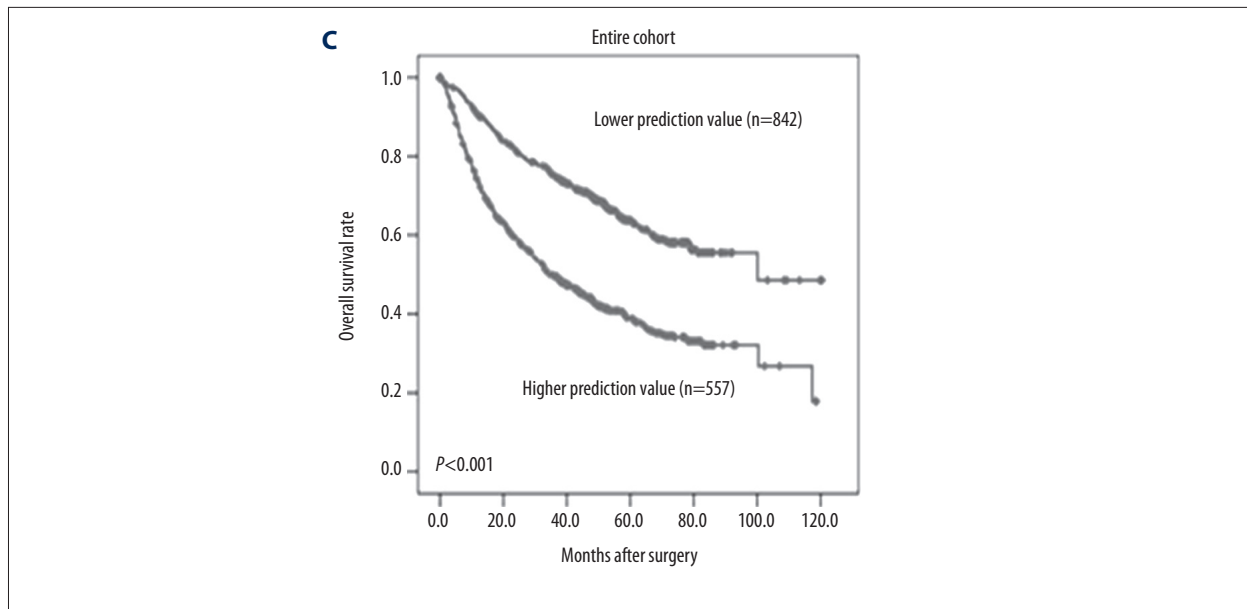
Supplementary Figure 5. Kaplan-Meier estimates of overall survival time in the discovery set (**A**), replication set (**B**) and entire cohort (**C**) based on the cut-off prediction value from the Nomogram prediction model.
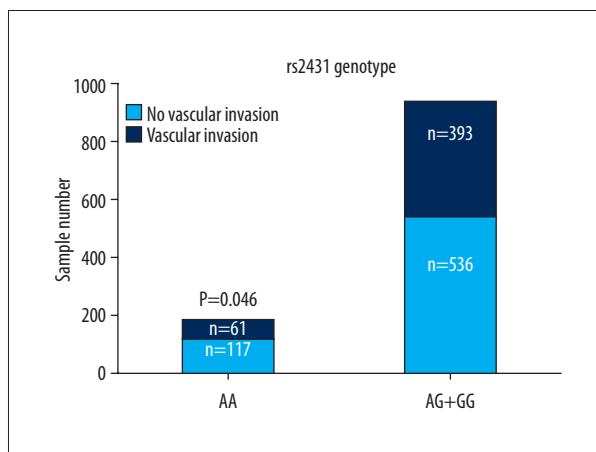
Supplementary Figure 6. Pearson chi-square estimate of tumor vascular invasion for rs2431 genotypes. Columnar section shows vascular invasion ratio in individuals with A or G genotypes.

References:

1. Ferlay J, Soerjomataram I, Dikshit R et al: Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer, 2015; 136: E359–86

2. Chen W, Zheng R, Baade PD et al:Cancer statistics in China, 2015. Cancer J Clin, 2016; 66(2): 115–32

3. Fernandez M, Semela D, Bruix J et al: Angiogenesis in liver disease. J Hepatol, 2009; 50: 604–20

4. Shen H, Zhong F, Zhang Y et al: Transcriptome and proteome of human hepatocellular carcinoma reveal shared metastatic pathways with significant genes. Proteomics, 2015; 15(11): 1793–800

5. Forner A, Llovet JM, Bruix J: Hepatocellular carcinoma. Lancet, 2012; 379: 1245–55

6. Reig M, Darnell A, Forner A et al: Systemic therapy for hepatocellular carcinoma: The issue of treatment stage migration and registration of progression using the BCLC-refined RECIST. Semin Liver Dis, 2014; 34: 444–55

7. Altshuler DM, Gibbs RA, Peltonen L et al: Integrating common and rare genetic variation in diverse human populations. Nature, 2010; 467: 52–58

8. Shastry BS: SNPs: Impact on gene function and phenotype. Methods Mol Biol, 2009; 578: 3–22

9. Dong QZ, Zhang XF, Zhao Y et al: Osteopontin promoter polymorphisms at locus –443 significantly affect the metastasis and prognosis of human hepatocellular carcinoma. Hepatology, 2013; 57: 1024–34

10. Hodo Y, Honda M, Tanaka A et al: Association of interleukin-28B genotype and hepatocellular carcinoma recurrence in patients with chronic hepatitis C. Clin Cancer Res, 2013; 19: 1827–37

11. Jung SW, Park NH, Shin JW et al: Prognostic impact of telomere maintenance gene polymorphisms on hepatocellular carcinoma patients with chronic hepatitis B. Hepatology, 2014; 59: 1912–20

12. Wu C, Li D, Jia W et al: Genome-wide association study identifies common variants in SLC39A6 associated with length of survival in esophageal squamous-cell carcinoma. Nat Genet, 2013; 45(6): 632–38

13. Han JY, Lee YS, Shin ES et al: A genome-wide association study of survival in small-cell lung cancer patients treated with irinotecan plus cisplatin chemotherapy. Pharmacogenomics J, 2014; 14: 20–27

14. Wu X, Wang L, Ye Y et al: Genome-wide association study of genetic predictors of overall survival for non-small cell lung cancer in never smokers. Cancer Res, 2013; 73: 4028–38

15. Wu C, Kraft P, Stolzenberg-Solomon R et al: Genome-wide association study of survival in patients with pancreatic adenocarcinoma. Gut, 2014; 63: 152–60

16. Shu XO, Long J, Lu W et al: Novel genetic markers of breast cancer survival identified by a genome-wide association study. Cancer Res, 2012; 72: 1182–89

17. Zhang H, Zhai Y, Hu Z et al: Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. Nat Genet, 2010; 42: 755–58

18. 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

19. 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: 72–75

20. Sawai H, Nishida N, Khor S 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(1): 7958

21. Li Y, Zhai Y, Song Q et al: Genome-wide association study identifies a new locus at 7q21.13 associated with hepatitis B virus – related hepatocellular carcinoma. Clin Cancer Res, 2018; 24: 906–15

22. Schwaederle M, Zhao M, Lee JJ et al: Impact of precision medicine in diverse cancers: A meta-analysis of phase II clinical trials. J Clin Oncol, 2015; 33(32): 3817–25

23. Wang K, Li J, Guo H et al: MiR-196a binding-site SNP regulates RAP1A expression contributing to esophageal squamous cell carcinoma risk and metastasis. Carcinogenesis, 2012; 33: 2147–54

24. Vaishnavi V, Manikandan M, Munirajan AK: Mining the 3'UTR of autism-implicated genes for SNPs perturbing microRNA regulation. Genomics Proteomics Bioinformatics, 2014; 12: 92–104

25. Yang B, Liu C, Diao L et al: A polymorphism at the microRNA binding site in the 3' untranslated region of C14orf101 is associated with non-Hodgkin lymphoma overall survival. Cancer Genet, 2014; 207: 141–46

26. Ahangari F, Salehi R, Salehi M, Khanahmad H: A miRNA-binding site single nucleotide polymorphism in the 3'-UTR region of the NOD2 gene is associated with colorectal cancer. Med Oncol, 2014; 31: 173

27. Maiti GP, Ghosh A, Mondal P et al: SNP rs1049430 in the 3'-UTR of SH3GL2 regulates its expression: Clinical and prognostic implications in head and neck squamous cell carcinoma. Biochim Biophys Acta, 2015; 1852: 1059–67

28. Tominaga K, Kondo C, Johmura Y et al: The novel gene fad104, containing a fibronectin type III domain, has a significant role in adipogenesis. Febs Lett, 2004; 577: 49–54

29. Nishizuka M, Kishimoto K, Kato A et al: Disruption of the novel gene fad104 causes rapid postnatal death and attenuation of cell proliferation, adhesion, spreading and migration. Exp Cell Res, 2009; 315: 809–19

30. Cai C, Rajaram M, Zhou X et al: Activation of multiple cancer pathways and tumor maintenance function of the 3q amplified oncogene FNDC3B. Cell Cycle, 2012; 11: 1773–81

31. Chen CF, Hsu EC, Lin KT et al: Overlapping high-resolution copy number alterations in cancer genomes identified putative cancer genes in hepatocellular carcinoma. Hepatology, 2010; 52: 1690–701

32. Zaravinos A, Kanellou P, Lambrou GI, Spandidos DA: Gene set enrichment analysis of the NF-kappaB/Snail/YY1/RKIP circuitry in multiple myeloma. Tumour Biol, 2014; 35: 4987–5005

33. Bjorkegren JL, Kovacic JC, Dudley JT, Schadt EE: Genome-wide significant *loci*: How important are they? Systems genetics to understand heritability of coronary artery disease and other common complex disorders. J Am Coll Cardiol, 2015; 65: 830–45

34. Jannot AS, Ehret G, Perneger T: P<5×10(–8) has emerged as a standard of statistical significance for genome-wide association studies. J Clin Epidemiol, 2015; 68: 460–65