Int J Clin Exp Pathol 2019;12(8):3063-3069
www.ijcep.com /ISSN:1936-2625/IJCEP0097889

Original Article

A single nucleotide polymorphism *CTSB* rs12898 is associated with primary hepatic cancer in a Chinese population

Minghua Cui[1,2*], Quanzhu Chen[1,2*], Chaojun He[3], Nan Wang[4], Yong Yu[1,2], Ziyang Sun[1,2], Zhenhua Lin[1,2], Hesong Cui[5], Shengyu Jin[6], Jae Yong Park[8], Guang Jin[1,2], Shin Yup Lee[8], Qingsong Cui[7]

[1]Key Laboratory of The Science and Technology Department of Jilin Province, Yanji, Jilin, China; [2]Department of Pathology & Cancer Research Center, Yanbian University Medical College, Yanji, Jilin, China; [3]Chengdu Second People's Hospital of Sichuan Province, Breast and Vascular Surgery, Chengdu, Sichuan, China; [4]Department of Pathology, Shenyang 242 Hospital, Shenyang, Liaoning, China; Departments of [5]Infections, [6]Hematology, [7]Intensive Care Unit, Yanbian University Hospital, Yanji, Jilin, China; [8]Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea. [*]Equal contributors and co-first authors.

Received June 3, 2019; Accepted June 26, 2019; Epub August 1, 2019; Published August 15, 2019

Abstract: Background & Aims: Primary hepatic cancer (PHC) is a common malignant tumor and the third most frequent cause of cancer-related death worldwide. However, the molecular mechanisms underlying hepatic cancer remain unknown. *CTSB* is considered a biomarker of cancer as it can facilitate tumor progression. We aimed to investigate the association between genetic polymorphisms of potential regulatory SNPs in the *CTSB* gene and PHC. Methods: The relationship between *CTSB* rs12898 and PHC was analyzed in a case-control study with a Chinese population of 608 PHC patients and 608 healthy individuals using SPSS 21.0. Results: PHC was significantly associated with alcohol consumption ($P < 0.001$), history of hepatitis ($P < 0.001$), and liver cirrhosis ($P < 0.001$), but not with smoking ($P = 0.168$), age ($P = 0.175$), or sex ($P = 0.051$). Distribution of three genotypes (GG, GA, and AA) of *CTSB* rs12898 significantly differed between the cases and controls ($P < 0.001$). Compared with the GG genotype, the GA and AA genotype was associated with a significantly increased risk of PHC (OR = 1.425, 95% CI = 1.099-1.848, $P = 0.007$; and OR = 2.220, 95% CI = 1.574-3.132, $P < 0.001$, respectively). *CTSB* rs12898 was associated with a significantly increased risk of PHC under a dominant model (OR = 1.592, 95% CI = 1.243-2.040, $P < 0.001$), and under a recessive model (OR = 1.771, 95% CI = 1.311-2.393, $P < 0.001$) for the variant A allele. Conclusion: Results suggest that *CTSB* rs12898G > A may play a role in the pathogenesis of PHC, and may be a marker for susceptibility to PHC.

Keywords: Case-control study, primary hepatic cancer, single nucleotide polymorphisms, susceptibility

Introduction

Primary hepatic cancer (PHC) is a common malignant tumor correlated with multiple genetic, viral and environmental factors [1]. Hepatic cancer is commonly diagnosed in China and was identified as one of the leading causes of cancer death from 1974 to 2015 [1, 2]. The GOLOBOCAN database shows the incidence of PHC in China is spreading further [3, 4]. A strong correlation between hepatic cancer and underlying chronic liver disease is well known, particularly for hepatitis B virus (HBV) and hepatitis C virus (HCV) [5]. However, the molecular mechanisms underlying liver carcinogenesis remain poorly understood, and traditional treatment approaches for hepatic cancer, such as liver transplantation, percutaneous ablation, chemoembolization, and molecular-targeted therapies all have limitations [6]. Thus, the best chance for long-term, disease-free survival can be achieved by early diagnosis before symptoms develop [7]. Therefore, facilitating early diagnosis of hepatic cancer, and improving access to and practicability of optimal treatments will relieve the existing burden of hepatic cancer in China. To achieve this goal, genetic biomarkers for identifying cancer risk, predict-

ing the prognosis, and enabling prevention are indispensable to reduce the mortality rate of PHC.

Single nucleotide polymorphisms (SNPs) are a type of DNA sequence polymorphism. SNPs are not only genetic markers but can also affect disease susceptibility and prognosis. There have been a number of studies on genetic susceptibility for PHC. A recent study showed that rs3754093 and five other SNPs in *hEXO1* are related to hepatic cancer [8]. Another study suggested that rs1143633, rs1143627 and rs3917356 of the interleukin-1 (*IL1*) family were found to be associated with hepatocellular carcinoma (HCC) [9]. An Egyptian study demonstrated that polymorphisms C626G rs20575 and A683C rs20576 in *DR4* had roles as potential risk factors for HCC development [10]. In China, it was found that rs22670-29, rs28382575, rs738791, rs738792, and rs131451 of the *MMP2* gene may help predict early-stage HCC and may be biomarkers for HCC progression [11].

Cathepsin B (*CTSB*) is a protein-coding gene and plays an important role in protein degradation and processing [12]. *CTSB* is considered a cancer biomarker because it facilitates tumor progression and is upregulated in many tumor types [13]. Overexpression of the encoded protein has been associated with esophageal adenocarcinoma and other tumors [14]. A recent study found that *CTSB* A4383C (rs13332) is significantly associated with HCC risk, and a significantly higher frequency of large-sized tumors was observed in HCC patients, carrying C76G (rs12338) than those carrying the ancestral genotype of *CTSB* [15]. *CTSB* as an oncogene contributes to the development and progression of HCC, thus it may be a valuable prognostic marker for HCC [13].

Regulome-DB is a massive integrated database that contains high-throughput sequencing data from public datasets, genome-regulatory information from ENCODE data, and other vital experimental study data. Regulome-DB has been used to select functional genes and disease-related genes, and ensures the function has great value for accurate diagnosis, treatment, and prognosis. As a useful technology, Regulome-DB has the ability to predict whether certain SNPs have regulatory effects on transcription factor binding and gene expression. It also sorts the functional SNPs by degree of impact, which are divided into categories I-VI (the smaller figure shows a bigger possibility of impact on the transcription factor and gene expression) [16, 17].

In this study, we aimed to extend the knowledge of the association between genetic polymorphisms of *CTSB* gene and PHC by selecting potentially regulatory SNPs using Regulome-DB.

Materials and methods

Subjects

We performed a case-control association study involving 608 PHC patients and 608 healthy individuals. Patients were diagnosed with PHC at Yanbian Hospital and Yanbian Tumor Hospital between June 2011 and August 2017. Control subjects were randomly selected from a pool of healthy volunteers who visited the same hospital for a general health examination.

Patients were diagnosed with PHC according to the *Diagnosis and Treatment Standards of Primary Liver Cancer (2011 Edition)* published by the Ministry of Health of the People's Republic of China [18]. There were no age, sex, stage, or histological restrictions, but patients with a prior history of cancer were excluded. Significant pure alcohol consumption was defined as drinking over 41 g/d and less than 61 g/d of alcohol in males, and over 21 g/d and less than 41 g/d in females [19]. All patients gave informed consent and this study was approved by the ethical review committee of Yanbian University Committee for Medical and Health Research Ethics.

DNA extraction

Five milliliters of human venous blood were collected before treatment (radiotherapy, chemotherapy or interventional therapy), and all blood samples were stored in ethylene diamine tetraacetic acid (EDTA) in an anticoagulant tube (10 mL) at -80°C until use. A nucleic acid isolation system was used to extract DNA from white blood cells. The steps of DNA extraction were based on the manufacturer's instructions of Quick Gene-810 and the matched kit (KURABO, Japan). A Nano Drop 2000 was used to test the purity and concentration of DNA samples; high

Table 1. Primer sequences and sequencing conditions of *CTSB* rs12898

Gene	SNP	Primer sequence	Annealing temperature (°C)	Length of product (bp)
CTSB	rs12898	Forward 5'-GAGGATTCAGCTCATAAAACAAG-3'	54.9	160
		Reverse 5'-CAAACCAGTGGCATACAAATTCA-3'	54.9	

Table 2. Restriction enzyme design for *CTSB* rs12898

Enzyme	Sequence length		
	GG Type	GA Type	AA Type
Hind III	21 bp, 139 bp	21 bp, 139 bp, 160 bp	160 bp

quality samples were selected (concentration > 1.5 Ng/μL; A260/A280 > 1.8; A260/A230 > 1.5) for testing.

Regulome-SNP selection and genotyping

In this study, we adopted the scoring criteria of the Regulome-DB and selected SNPs in category I among categories I-VI for the credibility of the experimental results. The category I contains 39,432 potential regulatory SNPs. Because *CTSB* was implicated as a cancer-related biomarker, we selected rs1123140, rs12898, rs1293288, rs1874546, rs2250903, rs2645408, and rs6995787 in this gene based on the Regulome-DB database. Then, by referring the Chinese Han population data Hapmap-SNP (HapMap Data Rel 28 Phase II + III, August 10, on NCBI B36 assembly, dbSNP b126) and National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/), the rs1123140, and rs2250903 were excluded based on linkage diseqilibrium (LD, $r^2 > 0.8$) and minor allele frequency (MAF, < 10%). Therefore, five SNPs in *CTSB* gene (rs12898, rs1293288, rs1874546, rs2645408, and rs6995787) were remained for further test. Moreover, it was found that the over-activation of Akt downregulated CTSB in HepG2/ADM hepatic cancer cells [20]. So we took the SNPs in *AKT* into considerations and tested rs17726963, rs3786527, and rs7144207 in the Training Set. In addition, because CTSB interacts with Hepatitis B spliced protein (HBSP) to activate MAPK/Akt signaling in hepatoma cells, leading to enhanced migration and invasion of hepatoma cells and angiogenesis [21], rs11647753, rs11863174, and rs7202714 in the *MAPK* were selected for genotyping in our study. Finally, a total of eleven SNPs were tested in this association study.

In this study, a three-stage study design was used to select the most significant Regulome-SNPs associated with PHC. In stage one, we built up the Training Set samples, which comprised 382 human venous blood samples from 192 PHC patients and 190 individuals. We analyzed the association between these eleven Regulome-SNPs and PHC, and only *CTSB* rs12898 was determined as relevant to hepatic cancer (Table S1). In stage two, *CTSB* rs12898 was tested further in the Testing Set consisting of 190 hepatic cancer patients and 190 healthy controls (Table S2). In the final stage, the Independent Cohort consisted of 226 PHC patients and 228 healthy persons (Table S3). In total, the combined analysis consisted of 608 PHC patients and 608 healthy individuals (**Table 4**).

All samples were sent to Beijing Genomics Institute (BGI) and Sequenom MassARRAY for genotyping. This study was approved by the Institutional Review Boards of the Yanbian Hospital and Yanbian Tumor Hospital.

PCR-RFLP

Polymerase chain reaction restriction fragment length polymorphism (PCR-RFLP) was used for genotypic accuracy inspection. GG is the wild-type allele and AA is the mutant type allele of *CTSB* rs12898G > A (https://www.ncbi.nlm.nih.gov/pubmed/). Primers were designed using primer design software (http://biotools.nubic.northwestern.edu/OligoCalc.html) and Primer Premier 5.0 (**Table 1**). Restriction enzymes were identified in Endonucleases design software (http://tools.neb.com/NEBcutter2) and the reference book of New England Biolabs (**Table 2**). The DNA sequence products were amplified by polymerase chain reaction (PCR) and the length was 160 bp. Then, the PCR products were cut by Hind *III* and added to the microchip electrophoresis system (MCE-202 MultiNA, SHIMADZU) for DNA analysis. Finally, according to the band sizes, the genotype of *CTSB* rs12898 was determined (GG genotype, 21 bp, 139 bp; GA genotype, 21 bp, 139 bp, 160 bp; AA genotype, 160 bp). To confirm the genotyping results, approximately 10% of the

Table 3. Characteristics of the case-control study population

Variable	Cases (%) (n = 608)	Controls (%) (n = 608)	*P*
Age (years)	62.62 ± 9.94	63.40 ± 10.67	0.189*
Sex			
Male	423 (69.57)	391 (64.31)	0.051**
Female	185 (30.43)	217 (35.69)	
Smoking status			
Smoking	303 (49.84)	279 (45.89)	0.168**
No smoking	305 (50.16)	329 (54.11)	
Significant alcohol consumption			
Yes	305 (50.16)	216 (35.53)	0.001**
No	303 (49.84)	392 (64.47)	
Chronic hepatitis			
Yes	477 (78.45)	25 (4.11)	0.001**
No	131 (21.55)	583 (95.89)	
Liver cirrhosis			
Yes	270 (44.41)	2 (0.33)	0.001**
No	338 (55.59)	606 (99.67)	

*t-test, **χ^2 test.

samples were randomly selected for re-testing by three different investigators, and the results were found to be 100% concordant with the testing genotyping results.

Statistical analysis

Hardy-Weinberg equilibrium was assessed by a goodness-of-fit χ^2 test with 1 degree of freedom using SHEsis online software to test the deviations of the genotype frequencies. The demographic and clinical information of cases and controls were compared using SPSS version 21.0. Continuous variables such as age were tested using the Student's t test. Categorical variables, such as sex, significant alcohol consumption, smoking status, and history of chronic hepatitis and liver cirrhosis were tested using the χ^2 test.

To test the strength of the associations between cases and controls, binary logistic regression models were used to estimate odds ratios (ORs) and 95% confidence intervals (CIs). All statistical analyses were performed using SPSS version 21.0 (IBM SPSS Statistics 21).

Results

Genotyping results confirmed by PCR-RFLP

The successfully amplified PCR products (160 bp) were selected for further studies, following agarose gel electrophoresis. Samples corresponded to the designed sizes of 21 bp, 139 bp, and 160 bp, which indicated the genotype of *CTSB* rs12898 GA. The fragment size of 160 bp represents the genotype of *CTSB* rs12898 AA, and the fragment size of 21 bp and 139 bp indicates the *CTSB* rs12898 GG genotype.

Patient demographics and characteristics

The demographics and characteristics of the 608 PHC patients and 608 controls enrolled in this study are listed in **Table 3**. The distribution of age in case and control groups obeys a normal distribution (F = 3.844, P = 0.051). The average age of the patients was 62.62 ± 9.94 years and that of the controls was 63.40 ± 10.67 years; there was no significant difference between the two groups (P = 0.189). There was also no significant difference between two groups for sex (P = 0.051). However, the patient group had a higher prevalence of significant alcohol consumption (P < 0.001), chronic hepatitis (P < 0.001), and liver cirrhosis (P < 0.001). There was no significant difference in smoking between cases and controls (P = 0.168).

Genotype frequencies and PHC risk

The genotypic distribution of the control group was in Hardy-Weinberg equilibrium (P = 0.054), demonstrating these data are from the same Mendelian population and had satisfactory genetic equilibrium. The distribution of the *CTSB* rs12898 genotypes and alleles among the patient group was significantly different from that of the control group (P < 0.001 and P < 0.001, respectively, **Table 4**). Logistic regression analyses showed that compared with the GG genotype, the AA genotype was associated with a significantly increased risk of PHC (OR = 2.220, 95% CI = 1.574-3.132, P < 0.001). Moreover, compared with the GG genotype, the GA genotype showed an increased risk of PHC (OR = 1.425, 95% CI = 1.099-1.848, P = 0.007). *CTSB* rs12898 was associated with a significantly increased risk of PHC under a recessive

Table 4. Logistic regression analyses for *CTSB* rs12898 genotypes and alleles in cases and controls

Genotype	*CTSB* rs12898		*P**	Minor allele frequency		*P**	OR (95% CI)	*P***
	Cases n = 608 (%)	Controls n = 608 (%)		Cases	Controls			
GG	153 (25.16)	212 (34.87)	< 0.001	0.48	0.39	< 0.001	1.000 (Reference)	
GA	322 (52.96)	313 (51.48)					1.425 (1.099-1.848)	0.007
AA	133 (21.88)	83 (13.65)					2.220 (1.574-3.132)	< 0.001
GA+AA vs. GG							1.592 (1.243-2.040)	< 0.001
AA vs. GG+GA							1.771 (1.311-2.393)	< 0.001

*Two-sided χ^2 test for either genotype distribution or allele frequencies between the cases and controls. **OR (95% CI) and corresponding *P* values were calculated by binary logistic regression analysis. CI, confidence interval; OR, odds ratio.

model (OR = 1.771, 95% CI = 1.311-2.393, $P < 0.001$), and under a dominant model for the variant A allele (OR = 1.592, 95% CI = 1.243-2.040, $P < 0.001$).

Discussion

The human genome project has revealed that only 2% of the human genome contains protein-coding genes, with the vast majority of human genome remained as 'junk DNA' [22]. However, it has been proven that so-called junk DNA functions as an essential gene regulator in many biologic processes [23]. Since the existence of ncRNAs has been tested further, the concept of junk DNA is long gone [24, 25]. The ENCODE project has demonstrated that 80% of the genome, mostly outside of protein-coding regions, harbors biochemically functional elements (i.e., transcription factor binding), which led to a new insight into the regulatory mechanisms of genome [26]. ENCODE also provided evidence that genetic variation in non-coding DNA affects its regulatory function in gene expression, suggesting possible roles of those variations in carcinogenesis.

CTSB has been implicated as a cancer-related biomarker, necessitating the exploration of *CTSB* polymorphism in the susceptibility to PHC [27]. *CTSB* encodes a pre-proprotein to generate multiple protein products. These products include the light and heavy chains of cathepsin B, which dimerize to form the double chain of the enzyme [28]. This enzyme is a type of lysosomal cysteine protease with both endopeptidase and exopeptidase activity and has a great affect on protein turnover and intracellular proteolysis [28]. Cathepsin B is normally associated with the lysosomes involved in autophagy and immune response, but its aberrant expression has been shown to lead to cancer [29]. Numerous studies have shown that abnormal regulation of cathepsin B overexpression is correlated with invasive and metastatic phenotypes in cancers [29]. Researches demonstrated that the regulation of cathepsin B can be altered at multiple levels during tumor expansion, thereby resulting in its overexpression and export outside the cell. This suggested the role of cathepsin B in alterations leading to carcinogenesis [29, 30]. It was found that *CTSB* is a potential pharmacologic target for colorectal tumor therapy, because it was a significant factor in colorectal tumor development, invasion, and metastatic spread [31]. In the present study, *CTSB* rs12898 was identified as a novel susceptibility marker for hepatic cancer. By replicating the results in two successive validation studies, we provided credibility of *CTSB* rs12898 as a susceptibility marker of PHC. The results suggested that rs12898, a potential regulatory SNP in *CTSB* gene, may play important roles in the pathogenesis of PHC. Testing for this SNP may help identifying individuals at high risk for PHC, and screening programs for the high risk individuals may facilitate early diagnosis of PHC. However, the role of *CTSB* in the pathogenesis of PHC remains unclear. Therefore, further studies are needed to investigate how CTSB participates in the pathogenesis of PHC, and to confirm the association between the SNP and the risk of PHC.

In this study, we observed that alcohol consumption, history of hepatitis (HBV/HCV), and liver cirrhosis are significantly associated with the risk of PHC. Previous studies have shown that hepatic cancer is related to HBV or HCV infection, [32] which is in line with our findings. We also verified again that alcohol consumption and liver cirrhosis were significant risk

factors for PHC ($P < 0.001$, $P < 0.001$, respectively), which is similar to previous studies [33-36]. All of the above results confirmed the known risk factors for PHC and showed the reliability of our study cohort. However, because this research included only people in the Yanbian area of China, further studies with people from different provinces of China are warranted for validating *CTSB* rs12898 as a susceptibility SNP for PHC in China.

Conclusion

These findings suggest that the functional polymorphism rs12898 in *CTSB* may contribute to PHC susceptibility, and that the variant A allele could increase the risk of PHC. High-risk groups who carry the *CTSB* rs12898 GA or AA genotype are candidates for cancer prevention and early detection.

Acknowledgements

The work reported here is supported by the National Natural Science Foundation of China and the Science and Technology Projects of "12th Five-Year Plan" of Jilin Provincial. We thank H Nikki March, PhD, from Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript. This work was supported by Chinese National Nature Science Foundation [grant number 31460284]; and the Science and Technology Projects of "12th Five-Year Plan" of Jilin Provincial [grant number 4130-10025].

Disclosure of conflict of interest

None.

Abbreviations

CIs, confidence intervals; EDTA, ethylene diamine tetraacetic acid; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; LD, linkage diseqilibrium; miRNAs, microRNAs; NAFLD, nonalcoholic fatty liver disease; ORs, Odds ratios; PCR, polymerase chain reaction; PCR-RFLP, polymerase chain reaction restriction fragment length polymorphism; PHC, primary hepatic cancer; SNPs, single nucleotide polymorphisms; 3' UTRs, 3' untranslated regions.

Address correspondence to: Dr. Guang Jin, Department of Pathology & Cancer Research Center, Yanbian University, Yanji 133002, Jilin, China. E-mail: jinguang1@ybu.edu.cn; Dr. Shin Yup Lee, Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea. E-mail: shinyup@knu.ac.kr; Dr. Qingsong Cui, Department of Intensive Care Unit, Yanbian University Hospital, Yanji, Jilin, China. E-mail: qingsong-0206@126.com

References

[1] Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. CA Cancer J Clin 2016; 66: 115-132.

[2] Li M, Wang S, Han X, Liu W, Song J, Zhang H, Zhao J, Yang F, Tan X, Chen X, Liu Y, Li H, Ding Y, Du X, Yin J, Zhang R, Cao G. Cancer mortality trends in an industrial district of Shanghai, China, from 1974 to 2014, and projections to 2029. Oncotarget 2017; 8: 92470-92482.

[3] Habib A, Desai K, Hickey R, Thornburg B, Lewandowski R, Salem R. Transarterial approaches to primary and secondary hepatic malignancies. Nat Rev Clin Oncol 2015; 12: 481-489.

[4] Song P, Hai Y, Ma W, Zhao L, Wang X, Xie Q, Li Y, Wu Z, Li Y, Li H. Arsenic trioxide combined with transarterial chemoembolization for unresectable primary hepatic carcinoma: a systematic review and meta-analysis. Medicine 2018; 97: e0613.

[5] Balogh J, Victor D 3rd, Asham EH, Burroughs SG, Boktour M, Saharia A, Li X, Ghobrial RM, Monsour HP Jr. Hepatocellular carcinoma: a review. J Hepatocell Carcinoma 2016; 3: 41-53.

[6] Forner A, Llovet JM, Bruix J. Chemoembolization for intermediate HCC: is there proof of survival benefit? J Hepatol 2012; 56: 984-986.

[7] Forner A, Bruix J. Biomarkers for early diagnosis of hepatocellular carcinoma. Lancet Oncol 2012; 13: 750-751.

[8] Tan S, Qin R, Zhu X, Tan C, Song J, Qin L, Liu L, Huang X, Li A, Qiu X. Associations between single-nucleotide polymorphisms of human exonuclease 1 and the risk of hepatocellular carcinoma. Oncotarget 2016; 7: 87180-87193.

[9] Tak KH, Yu GI, Lee MY, Shin DH. Association between polymorphisms of interleukin 1 family genes and hepatocellular carcinoma. Med Sci Monit 2018; 24: 3488-3495.

[10] Alsalawy NF, Darwish RK, Kamal MM, ElTaweel AE, Shousha HI, Elbaz TM. Evaluation of trail receptor 1 (DR4) polymorphisms C626G and

A683C as risk factors of hepatocellular carcinoma. J Med Virol 2018; 90: 490-496.
[11] Wang B, Hsu CJ, Lee HL, Chou CH, Su CM, Yang SF, Tang CH. Impact of matrix metalloproteinase-11 gene polymorphisms upon the development and progression of hepatocellular carcinoma. Int J Med SCI 2018; 15: 653-658.
[12] Li YY, Fang J, Ao GZ. Cathepsin B and L inhibitors: a patent review (2010 - present). Expert Opin Ther Pat 2017; 27: 643-656.
[13] Chen TP, Yang SF, Lin CW, Lee HL, Tsai CM, Weng CJ. A4383C and C76G SNP in cathepsin B is respectively associated with the high risk and tumor size of hepatocarcinoma. Tumour Biol 2014; 35: 11193-11198.
[14] Ruan J, Zheng H, Rong X, Rong X, Zhang J, Fang W, Zhao P, Luo R. Over-expression of cathepsin B in hepatocellular carcinomas predicts poor prognosis of HCC patients. Mol Cancer 2016; 15: 17.
[15] Ma W, Ma L, Zhe H, Bao C, Wang N, Yang S, Wang K, Cao F, Cheng Y, Cheng Y. Detection of esophageal squamous cell carcinoma by cathepsin B activity in nude mice. PLoS One 2014; 9: e92351.
[16] Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012; 22: 1790-1797.
[17] Liu Z, Sun J, Liu B, Zhao M, Xing E, Dang C. MiRNA222 promotes liver cancer cell proliferation, migration and invasion and inhibits apoptosis by targeting BBC3. Int J Mol Med 2018; 42: 141-148.
[18] The Ministry of Health of the People's Republic of China. Diagnosis and treatment standards of primary liver cancer (2011 edition). Chinese Clinical Oncology 2011; 16: 929-946.
[19] Chinese Center for Disease Control and Prevention. Surveillance report on chronic diseases and the risk factors in China (2010). Beijing: Military Medical Science Press; 2012. pp. 12.
[20] Sun H, Huang M, Yao N, Hu J, Li Y, Chen L, Hu N, Ye W, Chi-Shing Tai W, Zhang D, Chen S. The cycloartane triterpenoid ADCX impairs autophagic degradation through akt overactivation and promotes apoptotic cell death in multidrug-resistant HepG2/ADM cells. Biochem Pharmacol 2017; 146: 87-100.
[21] Chen WN, Chen JY, Jiao BY, Lin WS, Wu YL, Liu LL, Lin X. Interaction of the hepatitis b spliced protein with cathepsin b promotes hepatoma cell migration and invasion. J Virol 2012; 86: 13533-13541.
[22] Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB. Annotating non-coding regions of the genome. Nat Rev Genet 2010; 11: 559-571.
[23] Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, Biechele S, Huang B, Shen X, Ramalho-Santos M. A LINE1-nucleolin partnership regulates early development and ESC identity. Cell 2018; 174: 391-405.e19.
[24] Uchida S, Bolli R. Short and long noncoding RNAs regulate the epigenetic status of cells. Antioxid Redox Signal 2018; 29: 832-845.
[25] Singer RA, Sussel L. Islet long noncoding RNAs: a playbook for discovery and characterization. Diabetes 2018; 67: 1461-1470.
[26] Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489: 57-74.
[27] Sloane BF. Cathepsin b and cystatins: evidence for a role in cancer progression. Semin Cancer Biol 1990; 1: 137-152.
[28] Tang Y, Cao G, Min X, Wang T, Sun S, Du X, Zhang W. Cathepsin b inhibition ameliorates the non-alcoholic steatohepatitis through suppressing caspase-1 activation. J Physiol Biochem 2018; 19: 1-8.
[29] Gondi CS, Rao JS. Cathepsin b as a cancer target. Expert Opin Ther Targets 2013; 17: 281-291.
[30] Yan S, Sloane BF. Molecular regulation of human cathepsin b: implication in pathologies. Biological Chemistry 2003; 384: 845-854.
[31] Bian B, Mongrain S, Cagnol S, Langlois MJ, Boulanger J, Bernatchez G, Carrier JC, Boudreau F, Rivard N. Cathepsin b promotes colorectal tumorigenesis, cell invasion, and metastasis. Mol Carcinog 2016; 55: 671-87.
[32] Hooks KB, Audoux J, Fazli H, Lesjean S, Ernault T, Dugot-Senant N, Leste-Lasserre T, Hagedorn M, Rousseau B, Danet C, Branchereau S, Brugières L, Taque S, Guettier C, Fabre M, Rullier A, Buendia MA, Commes T, Grosset CF, Raymond AA. New insights into diagnosis and therapeutic options for proliferative hepatoblastoma. Hepatology 2018; 68: 89-102.
[33] Singal AK, Anand BS. Mechanisms of synergy between alcohol and hepatitis c virus. J Clin Gastroenterol 2007; 41: 761-772.
[34] International Agency for Research on Cancer (IARC). Monographs on the evaluation of carcinogenic risks to humans. Alcohol Drinking 1998; 44: 207-215.
[35] Chen G, Lin W, Shen F, Iloeje UH, London WT, Evans AA. Past HBV viral load as predictor of mortality and morbidity from HCC and chronic liver disease in a prospective study. Am J Gastroenterol 2006; 101: 1797-1803.
[36] Gao C, Fang L, Zhao HC, Li JT, Yao SK. Potential role of diabetes mellitus in the progression of cirrhosis to hepatocellular carcinoma: a cross sectional case-control study from Chinese patients with HBV infection. Hepatobiliary Pancreat Dis Int 2013; 12: 385-393.

Table S1. Logistic regression analyses for genotype and alleles in cases and controls in Training Set

Gene	SNP	Genotype	Cases n = 192 (%)	Controls n = 190 (%)	$P^{\#}$	Minor allele frequency		$P^{\#}$	OR (95% CI)	P^{*}
						Cases	Controls			
CTSB	rs12898	GG	44 (22.92)	64 (33.69)	0.001	0.51	0.39	< 0.001	1.000 (Reference)	
	G > A	GA	100 (52.08)	105 (55.26)					1.385 (0.864-2.220)	0.176
		AA	48 (25.00)	21 (11.05)					3.325 (1.752-6.309)	< 0.001
		GA+AA vs. GG							1.709 (1.088-2.683)	0.020
		AA vs. GG+GA							2.683 (1.534-4.691)	0.001
CTSB	rs1293288	TT	48 (25.00)	46 (24.21)	0.509	0.50	0.48	0.611	1.000 (Reference)	
	T > C	TC	95 (49.48)	104 (54.74)					0.875 (0.536-1.430)	0.595
		CC	49 (25.52)	40 (21.05)					1.174 (0.656-2.100)	0.589
		TC+CC vs. TT							0.958 (0.602-1.527)	0.858
		CC vs. TT+TC							1.285 (0.798-2.069)	0.302
CTSB	rs1874546	CC	61 (31.77)	54 (28.42)	0.648	0.44	0.45	0.783	1.000 (Reference)	
	C > G	CG	92 (47.92)	100 (52.63)					0.385 (0.513-1.294)	0.385
		GG	39 (20.31)	36 (18.95)					0.959 (0.536-1.717)	0.888
		CG+GG vs. CC							0.853 (0.550-1.321)	0.476
		GG vs. CC+CG							1.090 (0.658-1.807)	0.737
CTSB	rs2645408	AA	69 (35.94)	55 (28.95)	0.099	0.44	0.45	0.674	1.000 (Reference)	
	A > G	AG	78 (40.62)	98 (51.58)					0.634 (0.399-1.008)	0.054
		GG	45 (23.44)	37 (19.47)					0.969 (0.553-1.699)	0.914
		AG+GG vs. AA							0.726 (0.472-1.117)	0.145
		GG vs. AA +AG							1.266 (0.775-2.067)	0.346
CTSB	rs6995787	GG	52 (27.08)	60 (31.58)	0.219	0.49	0.43	0.108	1.000 (Reference)	
	G > C	GC	92 (47.92)	96 (50.53)					1.106 (0.692-1.767)	0.674
		CC	48 (25.00)	34 (17.89)					1.629 (0.916-2.896)	0.096
		GC+CC vs. GG							1.243 (0.799-1.932)	0.335
		CC vs. GG+GC							1.529 (0.933-2.507)	0.092
AKT	rs17726963	AA	67 (34.90)	56 (29.47)	0.053	0.44	0.43	0.869	1.000 (Reference)	
	A > C	AC	82 (42.71)	104 (54.74)					0.659 (0.417-1.042)	0.074
		CC	43 (22.39)	30 (15.79)					1.198 (0.667-2.152)	0.546
		AC+CC vs. AA							0.780 (0.507-1.199)	0.257
		CC vs. AA+AC							1.539 (0.918-2.581)	0.102
AKT	rs3786527	GG	96 (50.00)	106 (55.79)	0.498	0.30	0.26	0.231	1.000 (Reference)	
	G > A	GA	77 (40.10)	69 (36.32)					1.232 (0.804-1.887)	0.337
		AA	19 (9.90)	15 (7.89)					1.399 (0.673-2.905)	0.368

		GA+AA vs. GG							1.262 (0.844-1.887)	0.257
		AA vs. GG+GA							1.281 (0.631-2.603)	0.493
AKT	rs7144207	AA	53 (27.60)	49 (25.79)	0.169	0.46	0.51	0.170	1.000 (Reference)	
	A > G	AG	102 (53.13)	89 (46.84)					1.060 (0.655-1.715)	0.814
		GG	37 (19.27)	52 (27.37)					1.520 (0.857-2.697)	0.152
		AG+GG vs. AA							0.911 (0.579-1.435)	0.689
		GG vs. AA+AG							0.633 (0.392-1.024)	0.062
MAPK	rs11647753	GG	22 (11.46)	17 (8.95)	0.642	0.67	0.70	0.362	1.000 (Reference)	
	G > A	GA	84 (43.75)	81 (42.63)					0.801 (0.397-1.618)	0.537
		AA	86 (44.79)	92 (48.42)					0.72 (0.359-1.451)	0.361
		GA+AA vs. GG							0.759 (0.390-1.480)	0.419
		AA vs. GG+GA							1.157 (0.774-1.730)	0.477
MAPK	rs11863174	AA	85 (44.27)	96 (50.53)	0.434	0.34	0.29	0.193	1.000 (Reference)	
	A > C	AC	84 (43.75)	76 (40.00)					1.248 (0.815-1.911)	0.308
		CC	23 (11.98)	18 (9.47)					1.443 (0.729-2.855)	0.292
		AC+CC vs. AA							1.286 (0.860-1.923)	0.221
		CC vs. AA+AC							1.300 (0.677-2.497)	0.430
MAPK	rs7202714	CC	12 (6.25)	8 (4.21)	0.668	0.32	0.78	0.541	1.000 (Reference)	
	C > T	CT	67 (34.90)	67 (35.26)					0.667 (0.256-1.735)	0.406
		TT	113 (58.85)	115 (60.53)					0.655 (0.258-1.663)	0.373
		CT+TT vs. CC							0.659 (0.263-1.651)	0.374
		TT vs. CC +CT							0.933 (0.620-1.404)	0.739

[#]Two-sided χ^2 test for either genotype distribution or allele frequencies between the cases and controls. [*]OR (95% CI) and corresponding *P* values were calculated by binary logistic regression analysis. CI, confidence interval; OR, odds ratio.

Table S2. Logistic regression analyses for *CTSB* rs12898 genotype and allele in cases and controls in Testing Set

Gene	SNP	Genotype	Cases n = 190 (%)	Controls n = 190 (%)	$P^{\#}$	Minor allele frequency		$P^{\#}$	OR (95% CI)	P^{*}
						Cases	Controls			
CTSB	rs12898	GG	48 (25.26)	65 (34.21)	0.020	0.50	0.40	0.070	1.000 (Reference)	
	G > A	GA	94 (49.48)	97 (51.05)					1.312 (0.821-2.097)	0.256
		AA	48 (25.26)	28 (14.74)					2.321 (1.278-4.217)	0.006
		GA+AA vs. GG							1.538 (0.987-2.398)	0.057
		AA vs. GG+GA							1.956 (1.165-3.282)	0.011

$^{\#}$Two-sided χ^2 test for either genotype distribution or allele frequencies between the cases and controls. *OR (95% CI) and corresponding *P* values were calculated by binary logistic regression analysis. CI, confidence interval; OR, odds ratio.

Table S3. Logistic regression analyses for *CTSB* rs12898 genotype and allele in cases and controls in Independent Cohort

Gene	SNP	Genotype	Cases n = 226 (%)	Controls n = 228 (%)	$P^{\#}$	Minor allele frequency		$P^{\#}$	OR (95% CI)	P^{*}
						Cases	Controls			
CTSB	rs12898	GG	61 (26.99)	83 (36.40)	0.096	0.45	0.39	0.097	1.000 (Reference)	
	G > A	GA	128 (56.64)	111 (48.68)					1.569 (1.034-2.381)	0.034
		AA	37 (16.37)	34 (14.92)					1.481 (0.837-2.621)	0.178
		GA+AA vs. GG							1.548 (1.039-2.307)	0.032
		AA vs. GG+GA							1.117 (0.673-1.854)	0.669

$^{\#}$Two-sided χ^2 test for either genotype distribution or allele frequencies between the cases and controls. *OR (95% CI) and corresponding *P* values were calculated by binary logistic regression analysis. CI, confidence interval; OR, odds ratio.