GWAS Identifies Novel Susceptibility Loci on 6p21.32 and 21q21.3 for Hepatocellular Carcinoma in Chronic Hepatitis B Virus Carriers

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Published Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>doi:10.1371/journal.pgen.1002791</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Citable link</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:10456187">http://nrs.harvard.edu/urn-3:HUL.InstRepos:10456187</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Terms of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://</td>
</tr>
</tbody>
</table>
GWAS Identifies Novel Susceptibility Loci on 6p21.32 and 21q21.3 for Hepatocellular Carcinoma in Chronic Hepatitis B Virus Carriers

Shengping Li1,3, Ji Qian2,3, Yuan Yang3, Wanting Zhao4, Juncheng Dai5, Jin-Xin Bei1, Jia Nee Foo6, Paul J. McLaren6, Zhiquang Li7, Jingmin Yang2, Feng Shen3, Li Liu8, Jiamei Yang3, Shuhong Li1, Shandong Pan8, Yi Wang3, Wenjin Li7, Xiangjun Zhai9, Boping Zhou10, Lehua Shi3, Xinchun Chen10, Minjie Chu5, Yiqun Yan3, Jun Wang1, Shuqun Cheng3, Jiawei Shen7, Weihua Jia1, Jibin Liu8, Jiahai Yang3, Zujia Wen7, Aijun Li3, Ying Zhang1, Guoliang Zhang10, Xianrong Luo11, Hongbo Qin12, Minshan Chen1, Hua Wang3, Li Jin2, Dongxin Lin13, Hongbing Shen7, Lin He7,14,15, Paul I. W. de Bakker6, Hongyang Wang6, Yi-Xin Zeng1, Mengchao Wu3, Zhibin Hu5, Yongyong Shi7, Weipeng Zhou3*

1 Department of Hepatobiliary Oncology, State Key Laboratory of Oncology in South China, Sun Yat-Sen University Cancer Center, Guangzhou, China, 2 State Key Laboratory of Genetic Engineering, Center for Fudan–VARI Genetic Epidemiology and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China, 3 Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China, 4 Human Genetics, Genome Institute of Singapore, A*STAR, Singapore, Singapore, 5 MOE Key Laboratory of Modern Toxicology, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention, and Treatment, and State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, China, 6 Division of Genetics, Brigham and Women’s Hospital, Harvard Medical School Boston and Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America, 7 Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai, China, 8 Department of Hepatobiliary Surgery, Nantong Tumor Hospital, Nantong, China, 9 Department of Infection Diseases, Jiangsu Province Center for Disease Prevention and Control, Nanjing, China, 10 The Third People’s Municipal Hospital of Shenzhen, Shenzhen, China, 11 No. 458 Hospital of Chinese People’s Liberation Army, Guangzhou, China, 12 The Eighth Municipal People’s Hospital of Guangzhou, Guangzhou, China, 13 State Key Laboratory of Molecular Oncology and Department of Etiology and Carcinogenesis, Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, 14 Institutes for Nutritional Sciences, Shanghai Institute of Biological Sciences, Chinese Academy of Sciences, Shanghai, China, 15 Institutes of Biomedical Sciences, Fudan University, Shanghai, China

Abstract

Genome-wide association studies (GWAS) have recently identified KIF1B as susceptibility locus for hepatitis B virus (HBV)–related hepatocellular carcinoma (HCC). To further identify novel susceptibility loci associated with HBV–related HCC and replicate the previously reported association, we performed a large three-stage GWAS in the Han Chinese population. 523,663 autosomal SNPs in 1,538 HBV–positive HCC patients and 1,465 chronic HBV carriers were genotyped for the discovery stage. Top candidate SNPs were genotyped in the initial validation samples of 2,112 HBV–positive HCC cases and 2,208 HBV carriers and then in the second validation sample of 1,021 cases and 1,491 HBV carriers. We discovered two novel associations at rs9272105 (HLA-DQA1/DRB1) on 6p21.32 (OR = 1.30, 95% CI = 1.21–1.39 (P = 1.25, 95% CI = 1.12–1.40 (P = 0.84, 95% CI = 0.80–0.89. Further, we replicated the previously reported association at rs9272105 in the fourth independent sample of 1,298 cases and 1,026 controls (rs9272105: OR = 1.25, 95% CI = 1.13–1.37 (P = 0.84, 95% CI = 0.80–0.89. We also revealed the associations of HLA-DRB1*0405 and 0901*0405, which could partially account for the association at rs9272105. The association at rs455804 implicates GRIK1 as a novel susceptibility gene for HBV–related HCC, suggesting the involvement of glutamate signaling in the development of HBV–related HCC.


Editor: Mark R. Thursz, Imperial College London, United Kingdom

Received December 16, 2011; Accepted May 13, 2012; Published July 12, 2012

Copyright: © 2012 Li et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is partly funded by the China National High-Tech Research and Development Program Grant (2009AA022701, 2009AA022705, and 2009AA022706), the National Key Basic Research Program Grant (2010CB529600, 2011CB503802), National Major S&T Projects (2008ZX10002-015, 2008ZX10002-017, 2008ZX10002-018, and 2009ZX1004-904), the National Natural Science Foundation of China (81072344, 81130022, 30800946, 31000553, 30872372), NSF Science Fund for Creative Research Groups (30921006), State Key Laboratory funding of Oncology in South China (30304077008), Foundation for the Author of National Excellent Doctoral Dissertation (grants 201081, 201026), the Program for New Century Excellent Talents in University (NCET-09-0550, NCET-10-0178), a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (public health and preventive medicine), and funds from the Agency for Science, Technology, and Research (A*STAR) of Singapore. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ehphwp@126.com (W Zhou); liujil@gis.a-star.edu.sg (J Liu); shiyongyong@gmail.com (Y Shi); zhibin_hu@njmu.edu.cn (Z Hu)

1 These authors contributed equally to this work.
Introduction

Hepatocellular carcinoma (HCC) is the sixth common cancer and the third common cause of cancer mortality worldwide [1]. The incidence rate of HCC varies considerably in the world, with the highest in East, Southeast Asia and Sub-Saharan Africa, and China alone accounts for approximately half of HCC malignancies [1,2]. Major risk factors for HCC are chronic infections with the hepatitis B or C viruses, and exposure to dietary aflatoxin B1. Hepatitis B virus (HBV) infection is particularly important, because of its coherent distribution with the HCC prevalence [1,2]. However, it is known that only a minority of chronic carriers of HBV develop HCC [3], and the chronic HBV carriers with a family history of HCC have a two-fold risk for HCC than those without the family history [4], strongly suggesting the importance of genetic susceptibility for HBV-related HCC.

A number of candidate genes were investigated by genetic association studies to evaluate their roles in the susceptibility to HCC [5]. However, the findings from these studies are inconclusive due to moderate evidence and lack of independent validation. Recently, a genome-wide association study (GWAS) of HBV-related HCC was performed [6], in which 355 HBV-positive HCC patients and 360 chronic HBV carriers were used for the genome-wide discovery analysis, and the top 45 SNPs from the discovery analysis were further evaluated in additional 1,962 HBV-positive HCC patients and 1,430 controls (both chronic HBV carriers and population controls) as well as 159 trios. The study identified KIF1B as a novel susceptibility locus (top SNP rs17401966) on 1p36.22. Further study with better design and bigger sample size was recommended for identifying additional susceptibility loci for HCC [7,8]. These motivate us to carry out a GWAS with a large sample size in Chinese population to discover novel susceptibility loci for HCC.

Results

We performed a genome-wide discovery analysis by analyzing 523,663 common autosomal SNPs in two independent cohorts of the Han Chinese: 480 cases and 484 controls from central China and 1058 cases and 981 controls from southern China (Table S1 and Figure S1). The principal component analysis (PCA) confirmed all the samples to be Chinese, but indicated moderate genetic mismatch between the cases and controls in the cohort of southern China (Figure S2). To minimize the effect of population stratification, we performed the genome-wide association analysis using PCA-based correction for population stratification. After the adjustment by the first principal component, the age of the genome-wide association results is 1.013 for the cohort of central China, 1.003 for the cohort of southern China and 1.012 for the combined samples. Furthermore, for all the three genome-wide analyses of central, southern and combined samples, the quantile-quantile (QQ) plot of the observed P values revealed a good overall fit with the null distribution (Figure S3). Taken together, these results clearly indicate that the final association results from our genome-wide discovery analysis are free of inflation effect due to population stratification.

The genome-wide discovery analysis revealed multiple suggestive associations (P<10⁻⁵) on 2q22.1, 6p21.32, 11p15.1 and 20q12 (Figure S4 and Table S2). To validate these findings, 39 SNPs were selected according to their overall association evidence in three GWAS analyses as well as their consistencies of association between the two independent GWAS samples (Central and Southern China) (see the Methods for the selection criteria). The 39 SNPs were genotyped in additional 2,112 HBV-positive HCC cases and 2,208 HBV carriers (Phase I validation) (Table S1). Of the 39 SNPs, only 3 (rs9272105 on 6p21.32, rs11148740 on 13q21.32 and rs455804 on 21q21.3) were validated, showing consistent association between the GWAS discovery and Phase I validation samples (Table S3). These 3 SNPs were then genotyped in additional 1,026 HBV-positive HCC cases and 1,491 HBV carriers (Phase II validation). The Phase II validation analysis (Table 1) confirmed the associations at rs9272105 on 6p21.32 (OR = 1.41, P = 7.63×10⁻⁶) and rs455804 on 21q21.3 (OR = 0.83, P = 3.63×10⁻⁵), but not the association at rs11148740 on 13q21.32 (Table S3).

For both rs9272105 and rs455804, no heterogeneity of associations were observed among the GWAS and validation samples (P>0.05), and the associations in the combined GWAS and validation samples achieved genome-wide significance (P<5.0×10⁻⁸) (rs9272105: OR = 1.30, P = 1.13×10⁻¹⁹ and rs455804: OR = 0.84, P = 1.86×10⁻¹⁰) (Table 1). As a replication, these two SNPs were genotyped in the fourth independent samples of 1,298 cases and 1,026 controls from central China, which further confirmed the associations at rs9272105 (OR = 1.25, P = 1.71×10⁻⁸) and rs11148740 (OR = 0.84, P = 6.92×10⁻⁶) (Table 1). When combining all the five groups of samples, the two SNPs resulted in a 28% increased, and a 16% decreased risk (Table S4A). Furthermore, stratification analysis by age, gender, smoking and drinking status revealed similar ORs for rs9272105 and rs455804 among subgroups, except that the association at rs9272105 showed a stronger effect in the non-smoking group than the smoking group (OR = 1.38 vs. 1.19, P for heterogeneity = 0.0044) (Table S4B). Pair-wise interaction analysis among these two SNPs, smoking and drinking status did not reveal any significant interaction (data not shown). The samples used in the GWAS, validation and replication analyses are summarized in Table S1, and the multi-stage design of the whole study is shown in Figure S5.

We further investigated the association of HLA alleles in our GWAS samples through imputation. After QC filtering (see the Methods), 37 HLA alleles were successfully imputed, and 5 alleles showed nominal association (P<0.05) (Table S5 and Table 2). Further stepwise conditional analysis revealed that only two DRB1...
alleles showed independent associations (DRB1*0405: OR = 0.69, \(P = 6.18 \times 10^{-3}\); DRB1*0901: OR = 0.82, \(P = 3.62 \times 10^{-3}\)) (Table 2). Conditioning on rs9272105 could abolish the associations of the DRB1 alleles, and conditioning on the DRB1 alleles could weaken, but not eliminate, the association at rs9272105 (Table 2). The haplotype analysis of rs9272105 and the two DRB1 alleles revealed consistent result, showing that both the DRB1 alleles sit on the haplotypes carrying the protective G allele of rs9272105 (Table S6). Taking together, there seems to be additional risk effect beyond the ones carried by the DRB1 alleles.

We further explored whether the SNPs rs9272105 and rs455804 play any role in HBV infection. First, we compared the frequencies of these 2 SNPs between 408 non-symptomatic HBV carriers and 521 symptomatic chronic HBV patients from southern China (GWA scanned). The analysis revealed a protective association at rs9272105 (OR = 0.82, \(P = 7.9 \times 10^{-3}\)). However, there is no linkage disequilibrium (LD) between rs9272105 and our SNP rs9272105 (*rs2596542) on 6p21.33 and DEPDC5 (rs1012068) on 22q12.3 which were also identified as independent susceptibility loci for HCV-related HCC in Japanese population.

The HLA-DQ locus has also been shown to be associated with HCV-related HCC in a Japanese GWAS (rs92775572, OR = 1.30, \(P = 9.38 \times 10^{-3}\)) [13]. SNPs rs92775572 and rs9272105 are 79 kb away from each other and in weak LD (\(D' = 0.43, r^2 = 0.08\) in the HapMap CHB samples). The SNP rs92775572 did not show any association with HBV-related HCC in our GWAS discovery samples revealed additional SNPs showing association, but rs9272105 remained to be the top SNP within the region (Figure 1A). The residual association at rs9272105 after conditioning the association effects of the HLA alleles DRB1*0403 and *0901 suggests that there may be additional risk effect beyond the DRB1 alleles in Chinese population. The associations of the DRB1 alleles revealed by this study are consistent with the previous reports that HLA-DQ/DR alleles associated with HCC risk [9,10].

### Discussion

SNP rs9272105 is located between HLA-DQA1 and HLA-DRB1 on 6p21.32 (Figure 1A). SNP imputation in the GWAS discovery samples revealed additional SNPs showing association, but rs9272105 remained to be the top SNP within the region (Figure 1A). The residual association at rs9272105 after conditioning the association effects of the HLA alleles DRB1*0403 and *0901 suggests that there may be additional risk effect beyond the DRB1 alleles in Chinese population. The associations of the DRB1 alleles revealed by this study are consistent with the previous reports that HLA-DQ/DR alleles associated with HCC risk [9,10]. In addition, we investigated the previously reported HBV infection-associated SNPs rs3077, rs9277535, rs7453920, and rs2586718 within the HLA DP/DQ region [11,12] with HCC development in our GWAS samples. By imputation, we found the evidence of the association at rs9277535 with HCC (rs9277535: OR = 0.85, \(P = 7.9 \times 10^{-3}\)). However, there is no linkage disequilibrium (LD) between rs9277535 and our SNP rs9272105 (*rs2596542) on 6p21.33 and DEPDC5 (rs1012068) on 22q12.3 which were also identified as independent susceptibility loci for HCV-related HCC in Japanese population.
Table 2. Summary of the association results of five imputed HLA alleles in the GWAS discovery samples.

<table>
<thead>
<tr>
<th>CHR</th>
<th>HLA-allele</th>
<th>position</th>
<th>OR*</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>HLA_DQA1_0301</td>
<td>32716284</td>
<td>0.85</td>
<td>4.39E-03</td>
</tr>
<tr>
<td>6</td>
<td>HLA_DQA1_0601</td>
<td>32726642</td>
<td>1.24</td>
<td>7.12E-03</td>
</tr>
<tr>
<td>6</td>
<td>HLA_DQB1_0401</td>
<td>32739039</td>
<td>0.71</td>
<td>4.49E-03</td>
</tr>
<tr>
<td>6</td>
<td>HLA_DRB1_0405</td>
<td>32660042</td>
<td>0.69</td>
<td>6.18E-04</td>
</tr>
<tr>
<td>6</td>
<td>HLA_DRB1_0901</td>
<td>32660042</td>
<td>0.85</td>
<td>1.46E-02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional on HLA_DRB1_0405</th>
<th>6</th>
<th>rs9272105</th>
<th>32707977</th>
<th>1.28</th>
<th>8.99E-06</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>HLA_DRB1_0901</td>
<td>32660042</td>
<td>0.82</td>
<td>3.62E-03</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>HLA_DQA1_0301</td>
<td>32716284</td>
<td>0.90</td>
<td>8.78E-02</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>HLA_DQA1_0601</td>
<td>32716284</td>
<td>1.22</td>
<td>1.35E-02</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>HLA_DQB1_0401</td>
<td>32739039</td>
<td>0.89</td>
<td>4.77E-01</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional on HLA_DRB1_0405 and HLA_DRB1_0901</th>
<th>6</th>
<th>rs9272105</th>
<th>32707977</th>
<th>1.26</th>
<th>7.58E-04</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>HLA_DQA1_0301</td>
<td>32716284</td>
<td>1.14</td>
<td>2.35E-01</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>HLA_DQA1_0601</td>
<td>32716284</td>
<td>1.17</td>
<td>5.88E-02</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>HLA_DQB1_0401</td>
<td>32739039</td>
<td>0.87</td>
<td>4.13E-01</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional on the five HLA alleles</th>
<th>6</th>
<th>rs9272105</th>
<th>32707977</th>
<th>1.25</th>
<th>2.08E-03</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Conditional on rs9272105</th>
<th>6</th>
<th>HLA_DRB1_0405</th>
<th>32660042</th>
<th>0.78</th>
<th>3.13E-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>HLA_DRB1_0901</td>
<td>32660042</td>
<td>1.02</td>
<td>7.90E-01</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>HLA_DQA1_0301</td>
<td>32716284</td>
<td>1.02</td>
<td>8.22E-01</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>HLA_DQA1_0601</td>
<td>32716284</td>
<td>1.08</td>
<td>4.04E-01</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>HLA_DQB1_0401</td>
<td>32739039</td>
<td>0.78</td>
<td>4.95E-02</td>
<td></td>
</tr>
</tbody>
</table>

*Adjusted by the first principal component.

doi:10.1371/journal.pgen.1002791.t002

[13,14]. But, our GWAS discovery analysis did not reveal any supportive evidence for these two loci (rs2596542: OR = 1.06, P = 0.36; and rs1012068: OR = 1.06, P = 0.37) (Table S8 and Figure S6C and S6D). We also evaluated the power of our GWAS discovery samples and found that our samples should have sufficient power for detecting the previously reported association at rs9275572 (power = 94%), rs2596542 (power = 92%) and rs1012068 (power = 94%). Taken together, the disparity of association at rs17401966, but the strength of association in our GWAS discovery sample (OR = 0.90) is much weaker than the previously reported one (OR = 0.61) (Table S8). SNP imputation in our GWAS discovery samples did not reveal any stronger association than the association at rs17401966 within the LD region surrounding the 1p36.22 locus (Figure S6A).

Previous studies have clearly shown the existence of subpopulation structure of Chinese Han population along the north-south axis, and further demonstrated that geographic matching can be used as a good surrogate for genetic matching, and PCA-based correction is very effective in controlling the inflation effect of population stratification [26]. In the current study, all the cases and controls were matched by their geographic origin of residence. Moreover, the GWAS discovery samples were from central and southern China, while all the validation and replication samples were from central China. Our PCA analysis indicates that while there was mild population stratification in the sample of southern China, the cases and controls from central China were well matched without any indication of population stratification. In our study, the PCA-based correction was used in the GWAS analysis, and all the validation and replication analyses were from central China. Therefore, our findings should be free of adverse effect of population stratification in Chinese population.

In conclusion, the current GWAS identified two biologically plausible, novel loci on 6p21.32 and 21q21 for HBV-related HCC. These findings highlight the importance of HLA-DQ/DR molecules and glutamate signaling in the development of HBV-related HCC.

Methods

Patient samples

The genome-wide discovery analysis was performed by genotyping 731,442 SNPs in 1,575 HBV positive HCC patients...
and 1,490 HBV positive controls derived from two independent case-control cohorts of 500 cases and 500 controls from Central China (Shanghai) and 1,075 cases and 990 controls from Southern China (Guangdong). The first stage validation samples included 2,112 HBV-positive cases and 2,208 HBV-positive controls recruited from Jiangsu. The second stage validation samples consisted of 1,021 HBV-positive cases and 1,491 HBV carriers recruited from Shanghai. The replication samples of 1,298 HBV-positive cases and 1,026 HBV carriers were recruited from Central China (Shanghai and Jiangsu). Table S1 and Figure 1) All the samples are Han Chinese and partially participated in the previously published studies [27,28]. The diagnosis of HCC was confirmed by a pathological examination and/or α-fetoprotein elevation (>400 ng/ml) combined with imaging examination (Magnetic resonance imaging, MRI and/or computerized tomography, CT). Because HCV infection is rare in Chinese, we excluded HCC with HCV infection. Cancer-free HBV+ control subjects from central China were recruited from those receiving routine physical examinations in local hospitals or those participating in the community-based screening for the HBV/HCV markers and frequency-matched for age, gender, and geographic regions to each set of the HCC patients. Almost all these community-based controls are asymptomatic HBV carriers. Similarly, cancer-free control subjects from southern China are all HBV+ and 408 of them were asymptomatic HBV carriers and 521 were symptomatic chronic hepatitis B patients. All the HBV+ controls were positive for both HBsAg and antibody to hepatitis B core antigen (anti-HBc), and negative for anti-HCV.

We also recruited a HBV natural clearance cohort from Jiangsu Province (Zhangjiagang and Changzhou cities) through a population-based screening for the HBV/HCV markers and frequency-matched for age, gender, and geographic regions to each set of the HCC patients. Most of these community-based controls are asymptomatic HBV carriers. Similarly, cancer-free control subjects from southern China are all HBV+ and 408 of them were asymptomatic HBV carriers and 521 were symptomatic chronic hepatitis B patients. All the HBV+ controls were positive for both HBsAg and antibody to hepatitis B core antigen (anti-HBc), and negative for anti-HCV.

We performed standard quality control on the raw genotyping samples. We detected population outliers and stratification using a principal component analysis (PCA) based method. After removing MHC SNPs on chromosome 6 from 25–37 Mb, PCA was performed by using common autosomal SNPs with low LD (r^2<0.2) in the reference samples of the HapMap project (YRI (n = 90), CEU (n = 90), CHB (n = 45) and JPT (n = 44)) as the internal controls and our 3,010 participants of the GWAS discovery samples (after removal of samples with low call rates, ambiguous gender, and familial relationships). Projection onto the two multidimensional scaling axes is shown in Figure S2A. 7 outliers (more than 6 standard deviations) were identified and excluded. Finally, 523,663 autosomal SNPs in 1,538 cases and 1,465 controls, consisting of 480 cases and 484 controls from Central China and 1,058 cases and 981 controls from Southern China, were retained for association testing (Table S1).

SNPs selection and genotyping in validation phases

SNPs for the first stage validation were selected based on the following criteria: (i) SNP had Pjoint≤1.0×10^-4 in the analysis of the combined GWA samples or either the Central China sample or the Southern China sample, and had a consistent association in the two participant studies, meaning that the ORs from the two samples are both either above or below 1; (ii) only SNP with the lowest P value was selected when multiple SNPs showed a strong LD (r^2≥0.8). As a result, a total of 39 SNPs were included in the first stage validation. 3 SNPs that were significantly associated with HCC risk in the first validation stage were further genotyped in the second stage validation samples. Genotyping in the two validation samples were done by using the iPLEX platform (Sequenom) or the TaqMan assays (Applied Biosystems). The primers and probes were available upon request (Table S9). Laboratory technicians who performed genotyping experiments were blinded to case/control status. For TaqMan assay, ten percent of random samples were repeated, and the reproducibility was 100%. The 2 validated SNPs were genotyped in another independent replication using the same method.

Statistical analysis

Population structure was evaluated by the PCA in the software package EIGENSTRAT 3.0 [26]. PCA revealed one significant (P<0.05) eigenvector which was included in the logistic regression with other covariates of age, gender, smoking and drinking status for both the genome-wide discovery analysis and the joint analysis of the combined discovery and replication samples. Ancestral origin checking by PCA confirmed all the samples to be Han Chinese and further demonstrated moderate genetic stratification.
between the cases and the controls of the Southern cohort (Figure S2). The genome-wide association analysis was therefore performed in logistic regression using PCA-based correction for population stratification and by treating the samples of two cohorts as independent studies. The genome-control inflation factor \(\hat{\lambda}_G\) after adjustment by the first PC was calculated for the Central cohort samples \(\hat{\lambda}_G = 1.013\), the Southern cohort samples \(\hat{\lambda}_G = 1.003\) and the combined GWAS discovery samples \(\hat{\lambda}_G = 1.012\). Consistently, the QQ plot of the observed \(P\) values also showed a minimal inflation of genome-wide association results due to population stratification (Figure S3).

Statistical analyses were performed by using PLINK 1.07 [29] and R 2.11.1. The Manhattan plot of \(-\log_{10}P\) was generated using Haplovlew (v4.1) [30]. Untyped genotypes were imputed in the GWAS discovery samples by using IMPUTE2 [31] and the haplotype information from the 1000 Genomes Project (ASN samples as the reference set) and HapMap3 (CHB and JPT samples as the reference samples). The regional plot of association was created by using an online tool, LocusZoom 1.1. \(P\) value was two-sided, and OR presented in the manuscript was estimated by using additive model and logistic regression analyses if not specified.

To impute classical HLA alleles, we used 180 phased haplotypes from the HapMap CHB and JPT samples as our reference panel. This panel comprised dense SNP data and HLA allele types at 4-digit resolution for the HLA class I (HLA-A, B, C) and II (DQA1, DQB1 and DRB1) genes as previously described [32]. Genotypes, probability and allelic dosages were then imputed separately in the two discovery samples of Central and Southern Chinese using the BEAGLE program. Association testing was performed by using a logistic regression model on the best-guessed genotypes and allelic dosages. The results were checked for consistency between the two methods, and the results from best-guessed genotypes were presented.

Supporting Information

Figure S1 The map of China. The regions of the sample collection were highlighted in red. (DOCX)

Figure S2 Plots of principal components from the PCA for genetic matching. (A) a-b: plot of the first two PCs from the PCA of GWAS (central and southern) samples and the HapMap individuals. (B) plots between the 1st–8th PCs, which derived from PCA of 964 Central samples. (C) plots between the 1st–8th PCs, which derived from PCA of 2039 Southern samples. (D) plots between the 1st–8th PCs, which derived from PCA of 3033 Central and Southern samples. (DOCX)

Figure S3 Quantile-Quantile plot. (A) QQ plot of Central Samples. (B) QQ plot of Southern Samples. (C) QQ plot of combined GWAS Samples. (DOCX)

Figure S4 Manhattan plot of the genome-wide \(P\) values of association. Association was assessed using logistic regression analysis with adjustment for the first principal components of population stratification. (A) Manhattan plot for Central Samples. (B) Manhattan plot for Southern Samples. (C) Manhattan plot for combined GWAS Samples. (DOCX)

Figure S5 Workflow of the study. (DOCX)

Figure S6 Regional plots of 4 interested regions. (DOCX)

Table S1 Summary description of the samples used in this study. (DOCX)

Table S2 Summary of all the SNPs with \(P\) values less than \(10^{-4}\). (DOCX)

Table S3 A: Associations of the 39 fast-track replicated SNPs from the GWAS scan. B: Associations of 39 SNPs in GWAS scan and validations. (DOCX)

Table S4 Adjusted and stratified analyses of the 2 validated SNPs. (DOCX)

Table S5 Results of all the HLA alleles that have been successfully imputed after quality controls. (DOCX)

Table S6 The haplotype analysis of the two DRB1 alleles and rs9272105. (DOCX)

Table S7 A: Association of two SNPs in the subjects with asymptomatic and symptomatic HBV infection from the southern Chinese cohort. B: Association of rs9272105 and rs455804 in asymptomatic HBV carriers and HBV natural clearance samples from central China. (DOCX)

Table S8 Association results of rs17401966, rs2596542, rs1012068, and rs9275572 in the GWAS samples. (DOCX)

Table S9 Information of primers and probes for the 39 fast-track replicated SNPs of the GWAS scan. (DOCX)

Author Contributions

Study concept and design: Weiping Zhou, Shengping Li, Jianjun Liu, Zhibin Hu, Yongyong Shi, Ji Qian, Yuan Yang. Subject recruitment and sample preparation: Weiping Zhou, Yuan Yang, Ji Qian, Jingmin Yang, Feng Shen, Jiamei Yang, Yi Wang, Lehua Shi, Yiqun Yan, Shuqun Cheng, Jiaye Yang, Ajun Li, Li Jin (Shanghai samples); Shengping Li, Suishoung Li, Jun Wang, Boping Zhou, Weihua Jia, Ying Zhang, Xinshun Chen, Guofang Zhang, Xianrong Luo, Hongbo Qin, Minshan Chen (Guangzhou samples); Zhibin Hu, Li Liu, Shandong Pan, Minjie Chu, Xiangjun Zhai, Jibin Liu, Huawang Yang (Jiangsu samples); GWA data processing: Zhiquiang Li, Jiawei Sheng, Lehua Shi. Sample genotyping: Wenjin Li, Zujia Wen (GWAS Scan); Li Liu, Shandong Pan, Minjie Chu (Validation stages). Analysis and interpretation of data: Weiping Zhou, Shengping Li, Jianjun Liu, Zhibin Hu, Yongyong Shi, Ji Qian, Yuan Yang, Juncheng Dai, Wanting Zhao. Drafting of the manuscript: Jianjun Liu, Zhibin Hu. Critical revision of the manuscript for important intellectual content: Weiping Zhou, Shengping Li, Jianjun Liu, Zhibin Hu, Yongyong Shi, Ji Qian, Yuan Yang, Juncheng Dai, Wanting Zhao. Administrative, technical, or material support: Hongbing Shen, Lin He, Hongyang Wang, Yi-Xin Zeng, Mengchao Wu. Study supervision: Weiping Zhou, Shengping Li, Jianjun Liu, Zhibin Hu, Yongyong Shi, Ji Qian.
References