Genome-Wide Association Study Identifies a New Locus at 7q21.13 Associated with Hepatitis B Virus-Related Hepatocellular Carcinoma

Yuanfeng Li1,2,3, Yun Zhai1,2,3, Qingfeng Song4, Haitao Zhang1,2,3, Pengbo Cao1,2,3, Jie Ping1,2,3, Xinyi Liu1,2,3, Bingqian Guo1,2,3, Guanjun Liu1,2,3, Jin Song1,2,3, Ying Zhang1,2,3, Aiqing Yang1,2,3, Hongbo Yan1,2,3, Liang Yang4, Ying Cui4, Yilong Ma4, Jiniang Xing5, Xizhong Shen6, Taotao Liu6, Hongxin Zhang7, Jiaze An7, Jin-Xin Bei7, Weihua Jia9,10, Longli Kang11,12, Lijun Liu11,12, Dongya Yuan11,12, Zhibin Hu13, Hongbing Shen13, Lei Lu14, Xuan Wang14, Hua Li15, Fuchu He1,2,3, Hongxing Zhang1,2,3, and Gangqiao Zhou1,2,3

Abstract

Purpose: Hepatocellular carcinoma (HCC) is one of the most common cancers worldwide. In China, chronic hepatitis B virus (HBV) infection remains the major risk factor for HCC. In this study, we performed a genome-wide association study (GWAS) among Chinese populations to identify novel genetic loci contributing to susceptibility to HBV-related HCC.

Experimental Design: GWAS scan is performed in a collection of 205 HBV-related HCC trios (each trio includes an affected proband and his/her both parents), and 355 chronic HBV carriers with HCC (cases) and 360 chronic HBV carriers without HCC (controls), followed by two rounds of replication studies totally consisting of 3,796 cases and 2,544 controls.

Results: We identified a novel association signal within the CDK14 gene at 7q21.13 (index rs10272859, OR = 1.28, P = 9.46 × 10−10). Furthermore, we observed that the at-risk rs10272859[G] allele was significantly associated with higher CDK14 mRNA expression levels in liver tissues. Chromosome conformation capture assays in liver cells confirmed that a physical interaction exists between the promoter region of CDK14 and the risk-associated SNPs in strong linkage disequilibrium with the index rs10272859 at 7q21.13. This index rs10272859 also showed significant association with the survival of HCC patients.

Conclusions: Our findings highlight a novel locus at 7q21.13 conferring both susceptibility and prognosis to HBV-related HCC, and suggest the CDK14 gene to be the functional target of the 7q21.13 locus. Clin Cancer Res; 24(4); 906–15. ©2017 AACR.

Introduction

Hepatocellular carcinoma (HCC) is one of the most common cancers and ranks the third among causes of cancer mortality worldwide (1). It is estimated that about half a million new cases of HCC are reported in China each year (2). The major etiologic factors for HCC include hepatitis B and C virus infection, alcohol consumption, obesity, and diet contamination. In hyper-endemic areas such as China, chronic hepatitis B virus (HBV) infection remains a major risk factor for HCC, contributing to at least 80% of the cases (3). However, only a fraction of chronic HBV carriers develops HCC in their lifetime (4), suggesting that other risk factors may contribute to the interindividual variation in susceptibility to hepatocarcinogenesis. Segregation analyses of familial HCC have strongly suggested that the host genetic susceptibility is...
Translational Relevance

In this study, we performed a novel GWAS on HBV-related HCC among Chinese. We revealed that the variants at 7q21.13 locus (index m10272859) confer both increased susceptibility and worse prognosis to this malignancy. Thus, the variants at 7q21.13 might help to define both high-risk populations of HCC and patients with poor prognosis. It is therefore expected that the polygenic risk scores based on at-risk SNPs, along with environmental and clinical risk factors will potentially improve screening and prognosis prediction of HCC. Furthermore, we demonstrated that the cyclin dependent kinase 14 (CDK14) gene might be the functional target of the 7q21.13 locus. Recently, several inhibitors targeting CDK14 and other CDKs are undergoing clinical trials. Thus, our findings highlight the relevance of CDK14 as a novel therapeutic target for HCC by CDK inhibitors. In summary, our findings expanded our understanding of the genetic susceptibility to HBV-related HCC and would help to improve risk stratification and early treatment decision for this malignancy.

Materials and Methods

Study participants

In this study, we performed a three-stage GWAS among populations of Chinese ancestry, totally consisting of 205 HBV-related HCC trios, 4,151 chronic HBV carriers with HCC (cases) and 2,904 chronic HBV carriers without HCC (controls; Supplementary Table S1A; Supplementary Fig. S1). Specifically, the discovery stage included 205 trios (Guangxi population 1) and 355 cases and 360 controls (Guangxi population 2; ref. 7). Each HBV-related HCC trio consists of a patient with HBV-related HCC (proband) and his or her both parents. After quality control procedures, 189 trios (Guangxi population 1) and 348 cases and 359 controls (Guangxi population 2) were retained. The replication stage I included one case–control population (Shaanxi population), consisting of 1,095 cases and 394 controls. The replication stage II included four case–control populations (Guangdong population, Jiangsu population, Beijing population and Guangxi population 3), totally consisting of 2,701 cases and 2,150 controls. Together, 189 trios, 4,144 cases and 2,903 controls were used in this study.

To investigate whether the identified SNPs were significantly associated with the survival time of patients with HCC, we also used two independent case-only sample sets (Supplementary Table S1B). To assess the associations between the genotypes of the identified SNPs and mRNA expression levels of nearby genes, two independent sample sets were used (Supplementary Table S1B). In addition, 211 individuals recruited from southern China were used as random naive controls (Supplementary Table S1C).

For detailed descriptions of these populations, see Supplementary Methods. This study was performed with the approval of the Medical Ethical Committee of Beijing Institute of Radiation Medicine (Beijing, China), and was conducted in accordance with International Ethical Guidelines for Biomedical Research Involving Human Subjects (CIOMS). Written informed consent was obtained from each participant, and personal information on demographic factors and clinical data were collected by structured questionnaire.

Genotyping and quality controls in the discovery stage

For detailed descriptions of genotyping and quality controls in the discovery stage, see Supplementary Methods. Briefly, a collection of 205 HCC proband-parent trios (Guangxi population 1) recruited from Guangxi province, a well-known high-risk region for HCC, were genotyped using the HumanOmniguard-8 BeadChip (Illumina), which consists of 400,015 SNPs. Genotyping data for the case–control populations that was also recruited from Guangxi province (Guangxi population 2) in the discovery stage were derived from our previous study (7), which has been genotyped using the Affymetrix Genome-Wide Human SNP Array 5.0. For Guangxi population 1, we performed stringent quality controls for both samples and SNPs, and had a final 189 trios for subsequent association tests. The average genotyping call rate of these trios is 99.7%. Genotyping and quality controls for the Guangxi population 2 have been performed in the previous study, which finally consists of 348 cases and 359 controls (7).

Imputation analyses

To extend the coverage of genomic region in the discovery stage, we performed imputation on the GWAS genotyping data using the IMPUTE2 software (version 2.3.1). The 1,000 Genomes Project data (Version 3, March 2012) was used as the reference dataset. A prephasing strategy performed by SHAPEIT software (version 2) was adopted to improve the imputation performance. The phased haplotypes constructed by SHAPEIT were fed directly into IMPUTE2. A posterior probability of 0.90 was used as a threshold for calling genotypes. We then converted imputed
probabilities into hard genotype calls. In quality controls, we discarded imputed SNPs which had (i) a call rate of < 90%; (ii) a minor allele frequency (MAF) of < 0.01; or (iii) \( P < 1 \times 10^{-5} \) in a Hardy–Weinberg equilibrium (HWE) test among control populations. Finally, in the discovery stage, we had 6,089,459 SNPs in 189 trios (Guangxi population 1), and 4,140,602 SNPs in 348 cases and 359 controls (Guangxi population 2), respectively, for subsequent genome-wide association analyses.

Selection of SNPs for replication studies

For detailed descriptions of selecting candidate SNPs for subsequent replication studies, see Supplementary Methods. Briefly, an association between SNP and risk of HCC was considered significant when \( P \leq 0.1 \) in the trios (Guangxi population 1), or \( P \leq 0.05 \) in the Guangxi case–control population (Guangxi population 2). An SNP was retained whether it showed both a nominal \( P \leq 0.1 \) in the Guangxi population 1 and a nominal \( P \leq 0.05 \) with the same direction of association in the Guangxi population 2. Then the Fisher method was used to combine these two \( P \) values of the SNP. As a result, 10 SNPs showed combined \( P \leq 5 \times 10^{-5} \). Among them, rs17401966 has been reported in our previously published GWAS (7). Therefore, the remaining 9 candidate SNPs which had (i) a call rate of \( > 90\% \); (ii) a minor allele frequency (MAF) of \( < 0.01 \); or (iii) \( P < 1 \times 10^{-5} \) in a Hardy–Weinberg equilibrium (HWE) test among control populations. Finally, in the discovery stage, we had 6,089,459 SNPs in 189 trios (Guangxi population 1), and 4,140,602 SNPs in 348 cases and 359 controls (Guangxi population 2), respectively, for subsequent genome-wide association analyses.

Genotyping and quality controls in replication studies

For detailed descriptions, see Supplementary Methods. Briefly, the Sequenom MassARRAY System (Sequenom) was used to genotype the 14 novel candidate SNPs in the replication stage I. Of them, 2 SNPs (rs11163360 and rs10272859) showing significant association in the replication stage I was further genotyped in the replication stage II using the same platform. Only rs10272859 survived the replication stage II. The mass spectrometers of these SNPs were analyzed by the MassARRAY TYPER software (Sequenom), and visually checked to confirm their good quality. Five percent of the individuals in the replication studies were randomly selected for repeated genotyping, and the results were 100% concordant.

Statistical analyses

Family trios- and case/control–based association analyses were performed using PLINK (v1.07; ref. 14). In the discovery stage, family trios-based association analyses were conducted in Guangxi population 1 using transmission/disequilibrium test (TDT). Case–control association analyses were conducted in Guangxi population 2 using logistic regression under an additive model with adjustment for age, gender, smoking and drinking status, pack-years of smoking and family history of HCC, where appropriate. A lambda (=) inflation factor is given to indicate whether the systematic bias is present. Two \( P \) values of each SNP in the Guangxi population 1 and 2 were combined using the Fisher method to assess the joint association in the discovery stage. In the replication stages, the same case–control association analyses as those used in the discovery stage were conducted. Meta-analyses of data generated from all the family trios and case–control studies of all stages were conducted to assess the pooled genetic effects. We calculated Cochran Q statistic to test for between-group heterogeneity, and the heterogeneity was considered significant when \( Q > 0.05 \). A fixed-effect model was used in the meta-analyses as there was no indication of heterogeneity in this study (\( P > 0.05 \)). These meta-analyses were performed using the METAL software (15). The potential modification effects of gender and age on HCC risk was assessed by addition of interaction terms in the logistic regression model and by separate analyses of subgroups of subjects stratified by these factors.

Results

Genome-wide association analyses for the HCC trios

To identify novel loci conferring genetic susceptibility to HBV-related HCC, we carried out a three-stage GWAS (Supplementary Fig. S1). In the discovery stage, a collection of 205 HBV-related HCC proband-parent trios (Guangxi population 1) were genotyped by the HumanOmniZhongHua-8 BeadChip (Illumina). After quality controls, 694,794 autosomal SNPs in 189 trios survived, with an average genotyping call rate in these trios being 99.7% (Table 1; Supplementary Tables S1A and S2). No outlier was present using principal component analyses (PCA; Supplementary Fig. S2). By further imputing the genotypes of SNPs to extend the coverage of the genomic region, we finally achieved a total of 6,089,459 SNPs in 189 trios (Supplementary Table S3). We then performed a TDT test to assess the genotype-phenotype association. A Manhattan plot was created to show the associations between the SNPs and the risk of HBV-related HCC (Supplementary Fig. S3A). Quantile-quantile plots showed a good match between the distributions of observed \( P \) values and those expected by chance (inflation factor \( \lambda = 1.04 \); Supplementary Fig. S4A), suggesting minimal overall inflation of the genome-wide statistical results for these trios. Overall, 634,355 SNPs showed association with risk of HBV-related HCC in the 189 trios (\( P \leq 0.1 \); Supplementary Table S4).

Previously published GWAs have identified several SNPs relevant to risk of HBV-related HCC (7–10). In the present study, we confirmed the genetic effects of rs17401966 at KIF1B (1p36.22), rs9272105 at HLA-DQA1/DRB1 (6p21.32), rs9275319 at HLA-DQB1/DQA2 (6p21.32), and rs12663434 at BACH2 (6q15; \( P = 0.061, 0.053, 0.037 \) and 0.047, respectively; Supplementary Table S5). However, the rs455804 at GRIK1 (21q21.3) and rs7574865 at STAT4 (2q32.3) failed to be replicated (both \( P \) values > 0.1; Supplementary Table S5). The inconsistent associations may be due to the different study designs or population diversity and need further investigation. Taken together, we replicated several previously reported SNPs, although the statistical power of our family-based GWAS was modest due to limited sample size.

Recent GWAs have also identified several SNPs that were associated with other HBV-related phenotypes (16–18). Many of these SNPs were also shown to be significantly associated with HBV-related HCC in the present trios (Supplementary Table S6).
For examples, SNPs at HLA-DPA1 (rs3077), HLA-DPB1 (rs9277535), and NOTCH4 (rs422951) have been reported to be significantly associated with chronic HBV infection (16, 17). In the present study, the at-risk alleles of these SNPs were also shown to be risk factors for the increased susceptibility to HBV-related HCC ($P < 0.012, 0.0010$ and $0.023$, respectively; Supplementary Table S6). In addition, rs11866328 at GRIN2A has been reported to be associated with the disease progression of chronic HBV infection (18). In this study, the at-risk allele of rs11866328 was shown to contribute a significantly increased risk to HBV-related HCC ($P = 0.039$; Supplementary Table S6). Collectively, these results suggest that there might exist shared genetic risk factors among these HBV-related phenotypes.

Combined analyses of two GWAS datasets in the trios and case/control population

To increase statistical power to detect candidate loci for subsequent replication studies, we combined the present GWAS data in trios (Guangxi population 1) with our previously published GWAS data in case/control population (Guangxi population 2; ref. 7). For the published GWAS data passing quality filtering, we further performed imputation analyses to extend the coverage of the genomic region, which were not conducted in the previous study. Totally, 1,140,602 SNPs were obtained in 348 cases and 359 controls (Supplementary Table S3). A logistic regression analysis under additive model with adjustment for age and gender was then used to assess the genetic associations for these SNPs. Manhattan plot and Quantile-quantile plots were shown in Supplementary Figs. S3B and S4B, respectively. Then, we combined these two GWAS datasets and used the Fisher method to perform joint association analyses (see Materials and Methods). Totally, we identified 15 candidate SNPs to be significantly associated with HBV-related HCC in the discovery stage (combined $P$ values $< 5 \times 10^{-8}$), including the previously identified rs17401966 and other 14 novel candidate SNPs (Supplementary Table S7).

A novel susceptibility locus at 7q21.13 was identified

We then selected the 14 novel candidate SNPs for subsequent replication studies. We genotyped these 14 SNPs using Sequenom MassARRAY System in an independent case–control population recruited at Shaanxi province located in northwestern China (replication stage I, including 1,095 cases and 394 controls; Table 1; Supplementary Table S8). Of these 14 SNPs, two SNPs, rs11163360 and rs10272859, were replicated with the same direction of association as in the discovery stage ($P = 0.0089$ and 0.0083, respectively; Table 2; Supplementary Table S9). We further genotyped these 2 SNPs using Sequenom assay or SNP array in additional four independent case–control populations in replication stage II, totally consisting of 2,701 cases and 2,150 controls (Table 1; Supplementary Table S8). Only the rs10272859 at 7q21.13 survived this stage of replication ($P = 9.29 \times 10^{-5}$; Table 2; Supplementary Table S9). A meta-analysis combining all the family trios and case–control studies of all stages gave a joint $P$ value of $9.46 \times 10^{-10}$ (joint OR = 1.28; Fig. 1A–C; Table 2; Supplementary Fig. S5), which reached the genome-wide significant threshold of $5 \times 10^{-8}$. No evidence for heterogeneity of OR values for rs10272859 was observed across all sample sets ($P_{hetogeneity} = 0.49$; Table 2). In association analyses in all case–control populations, gender and age have been adjusted to reduce the confounding derived from gender- or age-unmatching. Furthermore, the potential modification effects of gender and age on association between rs10272859 and HBV-related HCC was assessed by stratification

| Table 1. Summary of the populations used in the discovery and replication stages of the GWAS |
|---|
| **Probands** | **Sample size** | **Mean age (SD)** | **Female/male** | **Controls** | **Sample size** | **Mean age (SD)** | **Female/male** |
| Discovery stage (Guangxi population 1) | 189 | 35.70 (6.87) | 18/177 | 378 | 64.52 (8.23) | 189/189 |
| Replication stage I (Shaanxi population) | 1,095 | 50.00 (10.52) | 121/974 | 394 | 40.88 (14.09) | 151/243 |
| Replication stage II (Guangdong population) | 1,562 | 49.62 (11.68) | 150/1,412 | 580 | 48.09 (12.44) | 62/518 |
| Replication stage II (Beijing population) | 625 | 55.91 (9.61) | 91/534 | 266 | 55.91 (12.22) | 41/225 |
| Replication stage II (Jiangsu population) | 229 | 50.32 (9.39) | 22/207 | 681 | 50.62 (10.44) | 80/601 |
| Replication stage II (Guangxi population 3) | 285 | 47.84 (11.27) | 3/254 | 623 | 40.96 (10.36) | 158/465 |
| **Total** | 4,144 | 50.56 (11.01) | 460/3,684 | 2,903 | 46.01 (12.46) | 541/2,362 |

Abbreviations: GWAS, genome-wide association study; s.d., standard deviation.

| Table 2. Association results at the discovery and replication stage, and their joint analyses for rs10272859 |
|---|
| **GWAS stages** | **Cases** | **Controlsa** | **ORs (95% CIs)** | **$P$** | **$P_{hetogeneity}$** |
| **Trios** | | | | | |
| Discovery stage (Guangxi population 1) | 142 (1.05–1.92) | 3.20 $\times 10^{-7}$ | | | |
| Replication stage I (Shaanxi population) | 47/179/120 | 30/75/169 | 1.61 (1.27–2.04) | 9.27 $\times 10^{-5}$ | | |
| Replication stage II (Guangdong population) | 120/496/488 | 41/143/194 | 1.31 (1.07–1.61) | 8.30 $\times 10^{-5}$ | | |
| Replication stage II (Beijing population) | 196/763/599 | 62/261/257 | 1.21 (1.10–1.34) | 9.29 $\times 10^{-5}$ | | |
| Replication stage II (Jiangsu population) | 74/267/282 | 21/100/26 | | | | |
| Replication stage II (Guangxi population 3) | 30/703/89 | 69/280/319 | | | | |
| Totalb | 37/140/105 | 62/288/267 | | | | |

Abbreviation: CI, confidence interval.

Counts of GG/CC/CC genotypes for rs10272859 in the cases and controls, respectively. The frequencies for genotypes of rs10272859 conform to Hardy–Weinberg equilibrium in each population ($P > 0.05$).

Meta-analyses of data generated from all the family trios and case–control studies were conducted to assess the pooled genetic effects. ORs were calculated on the basis of the rs10272859[G] allele.
with these factors. We found no appreciable variation of the effect across all case–control populations stratified by age or gender (nominal \( P_{\text{heterogeneity}} = 0.028 \sim 0.87 \), respectively; Supplementary Table S10), indicating that these confounding factors had no modification effect on the risk of HCC related to rs10272859 in these populations. The interaction effects between rs10272859

**Figure 1.**
Regional plot for associations in the region surrounding rs10272859 in the discovery stage. Results are shown for SNPs in the region 150 Kb up- or downstream of the marker SNP rs10272859. Genomic positions are based on NCBI Build 37. The \( P \) value of rs10272859 in the discovery stage is shown as a purple dot. The LD values \( (r^2) \) to rs10272859 for the other SNPs are indicated by marker’s color. Red signifies \( r^2 \geq 0.8 \), with orange \( 0.6 \leq r^2 < 0.8 \), green \( 0.4 \leq r^2 < 0.6 \), light blue \( 0.2 \leq r^2 < 0.4 \) and blue \( r^2 < 0.2 \). Estimated recombination rates (from the phase I (version 3) release of the 1000 Genomes Project, March 2012, Asian population) are plotted in light blue. Chr., chromosome. **A**, Regional plot surrounding rs10272859 for Guangxi population 1. **B**, Regional plot surrounding rs10272859 for Guangxi population 2. **C**, Regional plot surrounding rs10272859 for meta-analysis of Guangxi population 1 and 2. The \( P \) value of rs10272859 in the meta-analysis of discovery and replication stages shown as a purple diamond.
and other confounding factors, including cirrhosis status and viral factors (HBV genotypes and mutations, viral load and age at infection) were not investigated because these data were not fully available in the participants. Therefore, the possibility that the association signal detected by rs10272859 reflects some other aspects of disease biology related to HCC risk cannot be completely ruled out.

**CDK14** was identified as the candidate susceptibility gene at 7q21.13

The rs10272859 maps to chromosome 7q21.13 and lies in an approximately 110 Kb linkage disequilibrium (LD) block containing part of the *CDK14* gene (Fig. 1). *CDK14* encodes a member of the family of cyclin-dependent kinase (CDK), which plays critical role in cell-cycle progression and cell proliferation (19).

To assess whether *CDK14* was the functional target gene conferring susceptibility to HCC at 7q21.13, we performed expression quantitative trait locus (eQTL) analyses to assess the associations between the genotypes of rs10272859 and *CDK14* mRNA levels in liver tissues. In a sample set (Sample set 1) consisting of 88 HCC cases from the Jiangsu population in replication stage II (Materials and Methods), we observed that mRNA expression of *CDK14* in HBV-related HCC tissues was significantly higher than that in adjacent non-tumor tissues (*P* = 0.018, paired t test; Fig. 2A).

Similar result was observed in a GEO dataset GSE25097. The at-risk rs10272859[G] allele was associated with higher expression of *CDK14* in liver tissues. A, The at-risk rs10272859[G] allele was significantly associated with higher *CDK14* mRNA levels in both HCC tissues and adjacent non-tumor liver tissues of the Sample set 1 under an additive model (*P* = 0.013 and 0.075, respectively) or a dominant model (*P* = 0.0052 and 0.031, respectively; Fig. 2A). Similarly, the eQTL results were replicated in another independent dataset of HCC tissues, which consists of 82 HCC patients of Asian ancestry from The Cancer Genome Atlas (TCGA; *P* = 0.045 under additive model and *P* = 0.011 under dominant model, respectively; Supplementary Fig. S7).

Taken together, these results suggest that *CDK14* might be the functional target gene of 7q21.13.

The rs10272859 locates in the fifth intron of *CDK14*, and may have no functional relevance. Thus, to identify potential causative variants at 7q21.13, we performed functional annotation for the genetic variants that are tagged by the index SNP rs10272859. At the approximately 110 Kb region of 7q21.13, there were 75 SNPs in strong LD with rs10272859 based on Asian population in 1000 Genomes Project (all *r*² > 0.7; Supplementary Table S12). We annotated these SNPs with respect to exons, introns and UTRs as well as the epigenetic marks from the Encyclopedia of DNA Elements (ENCODE) data. None of these SNPs locate in exons or UTRs. However, 22 (29.3%) of them are in enhancer elements within F2 were colored as green. Genomic positions were based on NCBI Build 37. **Figure 2.** *CDK14* may be the susceptibility gene at 7q21.13. **A.** The at-risk rs10272859[G] allele was associated with higher expression of *CDK14* in liver tissues. The effects of rs10272859 genotypes on *CDK14* mRNA expression levels were examined in tumor tissues from 88 HBV-related HCC patients (Sample set 1). The expression data of *CDK14* mRNA were log₂ transformed. The SNP-expression association analyses were performed under an additive model (GG vs. GC vs. CC) or a dominant model, respectively. Paired t test was used to compare the *CDK14* expression levels between the tumor tissues and paired adjacent non-tumor tissues. Error bars indicate standard error of the mean (SEM). The *P* values were considered to be significant when below 0.05. **B.** Hi-C data of human blood CD34+ cells showed the candidate SNPs at 7q21.13 interacted physically with the promoter of *CDK14*. The Hi-C data of CD34+ cells were obtained from the previously published study (Materials and Methods). Each red arc represents an interaction between the promoter of *CDK14* and one DNA fragment. All the SNPs in strong LD with index rs10272859 (*r*² > 0.7 in Asian population based on the phase 1 (version 3) release of the 1000 Genomes Project data) were defined as candidate SNPs and shown as vertical short lines. The SNP rs10272859 was highlighted in red. According to the Hi-C data, a DNA fragment F2 containing 3 SNPs (rs885973, rs11356019, and rs4728919) was shown to be interacted with the promoter of *CDK14*, and the interaction was highlighted as a red square. The three candidate SNPs located within F2 were colored as green. Genomic positions were based on NCBI Build 37. **C.** Chromosome conformation capture (3C) analyses confirmed the interaction between the promoter of *CDK14* and F2 in HepG2 cells. Agarose gel electrophoresis of PCR products of predicted size confirmed the interaction between the *CDK14* promoter and the DNA fragment F2 with the addition of ligase (+). The interactions were not observed in controls without the addition of ligase (−). The DNA sequencing chromatogram of the indicated band by a red arrow shows the promoter of *CDK14* (blue sequence) ligated to the DNA fragment F2 (red sequence) and the intervening Dpn II restriction enzyme site (black box). Chr., chromosome; L, DNA ladder.
of multiple types of cells (Supplementary Table S12). Furthermore, using genome-wide chromosome conformation capture (Hi-C) data from two types of cells (human blood cells GM12878 and CD34+), we checked whether any of the 22 SNPs at 7q21.13 interact physically with the promoter regions of nearby genes (20). Indeed, in GM12878 cells, we found an interaction for the promoter of CDK14 with a DNA fragment (F1) containing the SNP rs10953011 (Supplementary Fig. S8), which locates at an enhancer element in the human hepatoma HepG2 cells, and also showed significant association with HCC (\(P = 1.9 \times 10^{-5}\); Supplementary Table S12). In addition, in both GM12878 and CD34+ cells, there exists an interaction between the promoter of CDK14 and a DNA fragment (F2) containing 3 of the 22 SNPs (rs885973, rs11356019 and rs4728919; Fig. 2B and Supplementary Fig. S8). All of the 3 SNPs were shown to be significantly associated with susceptibility to HBV-related HCC (all \(P\) values < \(2.3 \times 10^{-3}\); Supplementary Table S12). Because these 3 SNPs have not been revealed to locate at enhancer elements in liver-derived cells according to the ENCODE data, whether or not the interactions observed in GM12878 and CD34+ cells exist in liver-derived cells warrants further investigation. Therefore, we performed 3C assays and Sanger sequencing in HepG2 cells, and successfully confirmed the physical interaction between the DNA fragment F2 and the CDK14 promoter (Fig. 2C). We also checked other genes located within 1 Mb surrounding rs10272859, and found no interaction for them with any of the 75 SNPs.

Taken together, the results from the SNP-expression association analyses, epigenetic marks prediction, Hi-C and 3C analyses suggest that the genetic variants tagged by rs10272859 may transcriptionally modulate the expression of CDK14. Further studies will be needed to confirm the roles of these SNPs and CDK14 in the development of HCC.

Index rs10272859 at 7q21.13 was also associated with prognosis of patients with HBV-related HCC

Previous studies have revealed that ectopic expression of CDK14 induced substantial cellular invasion and migration of HCC cells (21). In addition, increased CDK14 expression was associated with advance tumor grading to poorly differentiated state and the histological presence of microvascular invasion in patients with HCC (22). Similarly, we observed in Sample set 1 that up-regulated CDK14 was significantly associated with the decreased overall survival time of patients with HCC (Supplementary Fig. S9). Given that the newly identified rs10272859 in this study correlates with CDK14 expression and that CDK14 expression in turn correlates with tumor severity, we further hypothesized that genotype at rs10272859 was directly associated with the survival of patients with HBV-related HCC. We used two case-only sample sets to test this priori hoc hypothesis (Supplementary Table S1B). The first sample set is the Sample set 1 which has been used in the eQTL analysis, for whom death or survival information was available. Log-rank test showed that the at-risk rs10272859[G] allele was significantly associated with shorter overall survival of HCC patients under an additive model \(P = 2.7 \times 10^{-3}\) for GG vs. GC vs. CC; HR = 2.28; Fig. 3A) or a dominant model \(P = 0.013\) for GG + GC vs. CC; HR = 2.43; Fig. 3B; Supplementary Table S13). Furthermore, another sample set (Sample set 2) consisting of 104 HCC cases from the Guangxi population 3 in replication stage II, for whom death or survival information was available, was also assessed. Again, similar results were observed under either additive model \(P = 0.048\);

Discussion

In the present study, we conducted a new GWAS on HBV-related HCC, which combined family trios-based and case-control-based designs in the discovery stage. We confirmed several previously reported SNPs, including those at \(KIF1B\), \(HLA-DQA1/DRB1\), \(HLA-DQB1/DQA2\), and \(BACH2\), and identified a novel association signal at 7q21.13 (index rs10272859), although the power of this study is limited (Supplementary Table S16). We further observed that the at-risk rs10272859[G] allele was associated with poor prognosis of patients with HBV-related HCC. Moreover, SNP-expression association analyses, epigenetic marks prediction, Hi-C and 3C analyses suggested that CDK14 may be the functional target gene at 7q21.13 for this malignancy.

In a separate study, we genotyped the rs10272859 in 211 random control subjects of Chinese origin without information on HBV infection status (Supplementary Table S1C). We found that the rs10272859[G] allele frequency in our chronic HBV carriers (0.319) was similar to that in this naive control set (0.340) and in Asians of the 1000 Genomes Project (0.324), but significantly less than that in Europeans, Africans and Americans of the 1000 Genomes Project (0.657, 0.623 and 0.500, respectively; all \(P < 0.05\), \(x^2\) test; Supplementary Table S17). It remains to be investigated whether these differences between ethnic groups may result in different susceptibility to HCC occurrence.

To examine whether rs10272859 is also associated with the risk for persistent HBV infection, we compared 1,251 persistently HBV infected subjects and 1,057 spontaneously recovered subjects from our previous GWAS on persistent HBV infection (23). We observed that the genetic association between rs10272859 and persistent HBV infection was far from statistical significance \(P = 0.84\), suggesting an irrelevant role of CDK14 in susceptibility to persistent HBV infection among Chinese.

The genetic association between the rs10272859 in CDK14 region at 7q21.13 and HBV-related HCC is biologically plausible. First, frequent DNA amplification at 7q21-22 was observed in
HCC tissues, and CDK14 has been highlighted to be the candidate oncogene within this region (21). Besides, a recurrent somatic mutation in CDK14 was identified in HBV-related HCCs, which was shown to promote colony formation of HCC cell lines (24). Moreover, analyses of transposon insertion sites from a Sleeping Beauty (SB) transposon mutagenesis screen in a HBV-related HCC mouse model revealed a significant enrichment of common insert sites (CISs) in CDK14 gene (25).

Second, previous studies have shown that the CDK14 abnormalities could affect cell growth, invasion, migration and motility in various cancers cell lines, including HCC (21, 24, 26–29). Third, CDK14 was shown to interact with cyclin Y to activate non-canonical Wnt signaling in HCC (29). CDK14 could also inactivate a tumor suppressor gene TAGLN2 via weakening its phosphorylation in HCC cells (22). Finally, in clinical samples, CDK14 expression was upregulated in HCC tissues as shown in the present study (Fig. 2A), in GEO dataset GSE25097 (Supplementary Fig. S6), and in previous study (21). CDK14 expression was also up-regulated in other cancers, including breast cancer and glioma (26, 27). Furthermore, upregulated CDK14 was highly associated with the poor prognosis of patients with breast cancer (26), or HCC (Supplementary Fig. S9; refs. 21, 22).

Given the role of CDK14 in the development of HCC, together with the genetic results from this study, one might expect that the individuals who carry the at-risk rs10272859[G] allele have elevated expression of CDK14 through the transcriptional regulation by DNA elements containing causative SNPs tagged by rs10272859, and subsequently have an increased risk of developing HBV-related HCC. If carrying the G allele of rs10272859 is regarded as a risk factor for the development of HBV-related HCC, then the population attributable fraction (PAF) implies that 8.7% (95% CI, 5.5%–11.6%) of elevation in the risk of developing this malignancy may be attributed to the susceptible effect of rs10272859. With more susceptibility loci of HCC being identified in future studies, polygenic risk scores based on these loci together with rs10272859 will potentially improve HCC screening and prevention.
In addition to the eQTL analysis of rs10272859 in the Sample set 1 and in TCGA data, we also performed eQTL analysis of rs10272859 according to the data from Genotype-Tissue Expression Project (GTEx). However, we found no significant association between rs10272859 and CDK14 in normal liver tissues of the GTEx data. The donors of normal liver samples involved in the GTEx are European ancestry, of which people usually have low rate of HBV infection. In contrast, the non-tumor liver samples used in the present study are from Chinese HCC patients infected with HBV. As suggested in a recent review (30), a key feature of regulatory genetic effects is its context-specificity, that is, varying effects of a given variant due to differences in the surrounding cellular or genomic environment. Therefore, the difference between the effects of rs10272859 in the GTEx data and those in the present study may be related with the context of HBV infection. In addition, eQTLs can be observed for arteries and several other tissues in the GTEx data, suggesting pleiotropic effects across multiple tissues of functional SNPs targeted by rs10272859. The functional consequence of the SNPs in these tissues warrant(s) further investigation.

In summary, our GWAS identified a novel locus at 7q21.13 conferring both susceptibility and prognosis to HBV-related HCC in Chinese populations. The variants at 7q21.13 may help to define high-risk populations of HBV-related HCC and patients with poor prognosis. We also demonstrated that CDK14 may be the susceptibility gene at the 7q21.13 locus. These findings might expand our understanding of the genetic susceptibility to HBV-related HCC and would help to improve risk stratification and early therapeutic decision making for this malignancy.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: X. Shen, F. He, H. Zhang, G. Zhou
Development of methodology: H. Zhang, H. Zhang

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): Y. Zhai, Q. Song, H. Zhang, X. Liu, B. Guo, G. Liu, J. Song, Y. Zhang, A. Yang, H. Yan, L. Yang, Y. Cui, Y. Ma, J. Xing, X. Shen, T. Liu, H. Zhang, J. An, J. X. Bai, W. Jia, L. Kang, L. Liu, D. Yuan, Z. Hu, H. Shen, L. Lu, X. Wang, H. Li, H. Zhang

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Y. Li, Y. Zhai, H. Zhang, P. Cao, J. Jing, X. Shen, H. Zhang, H. Zhang

Writing, review, and/or revision of the manuscript: Y. Li, H. Zhang, F. He, H. Zhang, G. Zhou

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Q. Song, T. Liu, H. Zhang, H. Shen, H. Li, H. Zhang

Study supervision: Z. Hu, H. Shen, F. He, G. Zhou

Acknowledgments
The authors thank all the patients participating in this study. The authors also thank Dr. Xiangdong Fang and Mr. Yadong Yang for the help in examining the association between rs10272859 genotypes and the mRNA expression levels of CDK14 based on the TCGA data. This work was supported by grants from the Major Project of Chinese National Programs for Fundamental Research and Development (http://program.most.gov.cn/, no. 2013CB910301, to G. Zhou), National Natural Science Foundation of China (http://www.nsfc.gov.cn/, no. 91440206 and 81125017, to G. Zhou; 81222027, to H. Zhang; 31260252 and 31460286, to L. Kang), Program of Beijing Municipal Commission for Science and Technology for Frontier Technology of Life Science (http://www.bjkw.gov.cn/, no. Z1411000001214014, to G. Zhou), Chinese Key Project for Infectious Diseases (http://program.most.gov.cn/, no. 2017ZX10032022, to G. Zhou; 2017ZX10033025, for H. Zhang), Program of International S&T Cooperation (http://www.nsfc.gov.cn/publish/portal2/, 2014DFB30020, to G. Zhou), and the National Science Foundation of Xizang (Tibet) Autonomous Region (http://tibetss.gov.cn/index.aspx/22014A09C2-3, to L. Kang).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 30, 2017; revised October 20, 2017; accepted December 5, 2017; published OnlineFirst December 15, 2017.

References
1. Yang JD, Roberts JR. Hepatocellular carcinoma: a global view. Nat Rev Gastroenterol Hepatol 2010;7:488–8.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115–32.
3. Bosch FX, Ribes J, Diaz M, Cleries R. Primary liver cancer: worldwide incidence and trends. Gastroenterology 2004;127:S5–S16.
4. Yu MW, Hsu FC, Sheen IS, Chu CM, Lin DY, Chen CJ, et al. Prospective study of hepatocellular carcinoma and liver cirrhosis in asymptomatic chronic hepatitis B virus carriers. Am J Epidemiol 1997;145:1039–44.
5. Shen FM, Lee MK, Gong HM, Cai XQ, King MC. Complex segregation analysis of primary hepatocellular carcinoma in Chinese families: interaction of inherited susceptibility and hepatitis B viral infection. Am J Hum Genet 1993;49:88–93.
6. Yu MW, Chang HC, Liaw YF, Lin SM, Lee SD, Liu CJ, et al. Familial risk of hepatocellular carcinoma in chronic hepatitis B carriers and their relatives. J Natl Cancer Inst 2000;92:1159–64.
7. Zhang H, Zhai Y, Hu Z, Wu C, Qian J, Jia W, 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;42:755–8.
8. Li S, Qian J, Yang Y, Zhao W, Dai J, Bei JX, et al. GWAS identifies novel susceptibility loci on 6p21.32 and 2q11.23 for hepatocellular carcinoma in chronic hepatitis B virus carriers. PLoS Genet 2012;8:e1002791.
9. Jiang DK, Sun J, Cao G, Liu Y, Lin D, Cao YZ, et al. Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus-related hepatocellular carcinoma. Nat Genet 2013;45:72–5.

10. Clifford RJ, Zhang J, Meerzaman DM, Luu MS, Hu Y, Culturero CM, et al. Genetic variations at loci involved in the immune response are risk factors for hepatocellular carcinoma. Hepatology 2010;52:2034–43.
11. Laird NM, Lange C. Family-based designs in the age of large-scale gene-association studies. Nat Rev Genet 2006;7:385–94.
12. Ott J, Kamatani Y, Lathrop M. Family-based designs for genome-wide association studies. Nat Rev Genet 2011;12:465–74.
13. Leslie EL, Liu H, Carlson JC, Shaffer JR, Feingold E, Welbey G, et al. A genome-wide association study of nonsyndromic cleft palate identifies an etiologic missense variant in GRHL3. Hum Mol Genet 2015;24:4394–402.
14. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
15. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–1.
16. Kamatani Y, Watanapokasikit S, Ochi H, Kagawuchi T, Takahashi A, Hosono N, et al. A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nat Genet 2009;41:591–5.
17. Jiang DK, Ma XP, Yu H, Cao G, Ding DL, Chen H, et al. Genetic variants in five novel loci including CFB and CD40 predispose to chronic hepatitis B. Hepatology 2015;62:118–28.
18. Liu L, Li Y, Jao J, Yu J, Zhang J, Ning Q, et al. A genome-wide association study with DNA pooling identifies the variant rs11866328 in the GRINA2 gene that affects disease progression of chronic HBV infection. Viral Immunol 2011;24:397–402.
19. Shu F, Lv S, Qin Y, Ma X, Wang X, Peng X, et al. Functional characterization of human PFTK1 as a cyclin-dependent kinase. Proc Natl Acad Sci U S A 2007;104:9248–53.

20. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet 2015;47:598–606.

21. Pang EY, Bai AH, To KF, Sy SM, Wong NL, Lai PB, et al. Identification of PFTAIRE protein kinase 1, a novel cell division cycle-2 related gene, in the motile phenotype of hepatocellular carcinoma cells. Hepatology 2007;46:436–45.

22. Leung WK, Ching AK, Chan AW, Poon TC, Mian H, Wong AS, et al. A novel interplay between oncogenic PFTK1 protein kinase and tumor suppressor TAGLN2 in the control of liver cancer cell motility. Oncogene 2011;30:4464–75.

23. Li Y, Si L, Zhai Y, Hu Y, Hu Z, Bei IX, et al. Genome-wide association study identifies 8p21.3 associated with persistent hepatitis B virus infection among Chinese. Nat Commun 2016;7:11664.

24. Huang J, Deng Q, Wang Q, Li KY, Dai JH, Li N, et al. Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma. Nat Genet 2012;44:1117–21.

25. Bard-Chapeau EA, Nguyen AT, Rust AG, Sayadi A, Lee P, Chua BQ, et al. Transposon mutagenesis identifies genes driving hepatocellular carcinoma in a chronic hepatitis B mouse model. Nat Genet 2014;46:24–32.

26. Gu X, Wang Y, Wang H, Ni Q, Zhang C, Zhu J, et al. Upregulated PFTK1 promotes tumor cell proliferation, migration, and invasion in breast cancer. Med Oncol 2015;32:195.

27. Fan S, Zhao C, Zhang L, Dai S, Ren J, Zhang X, et al. Knockdown of PFTK1 inhibits the migration of glioma cells. J Mol Neurosci 2015;57:257–64.

28. Jiang M, Gao Y, Yang T, Zhu X, Chen J. Cyclin Y, a novel membrane-associated cyclin, interacts with PFTK1. FEBS Lett 2009;583:2171–8.

29. Sun T, Co NN, Wong N. PFTK1 interacts with cyclin Y to activate non-canonical Wnt signaling in hepatocellular carcinoma. Biochem Biophys Res Commun 2014;449:163–8.

30. Lappalainen T. Functional genomics bridges the gap between quantitative genetics and molecular biology. Genome Res 2015;25:1427–31.