New loci associated with chronic hepatitis B virus infection in Han Chinese

Zhibin Hu1,2,16, Yao Liu1,2,16, Xiangjun Zhai3,16, Jun Chen1,2,16, Guangfu Jin1,2,16, Li Wang4,16, Liguo Zhu3, Yuan Yang5, Jibin Liu6, Minjie Chu1,2, Juan Wen1,2, Kaipeng Xie1,2, Guoming Du7, Qungang Wang7, Yuchun Zhou8, Minquan Cao8, Li Liu6, Yisha He1,2, Ying Wang9, Gangqiao Zhou10, Weihua Jia11, Jiachun Lu12, Shenqing Li11, Jianjun Liu13, Haitao Yang3, Yongyong Shi14, Weiping Zhou5,15 & Hongbing Shen1,2

Chronic hepatitis B virus (HBV) infection is a challenging global health problem. To identify genetic loci involved in chronic HBV infection, we designed a three-phase genome-wide association study in Han Chinese populations. The discovery phase included 951 HBV carriers (cases) and 937 individuals who had naturally cleared HBV infection (controls) and was followed by independent replications with a total of 2,248 cases and 3,051 controls and additional replications with 1,982 HBV carriers and 2,622 controls from the general population. We identified two new loci associated with chronic HBV infection: rs3130542 at 6p21.33 (near HLA-C, odds ratio (OR) = 1.33, \(P = 9.49 \times 10^{-14}\)) and rs4821116 at 22q11.21 (in UBE2L3, OR = 0.82, \(P = 1.71 \times 10^{-12}\)). Additionally, we replicated the previously identified associations of HLA-DP and HLA-DQ variants at 6p21.32 with chronic HBV infection. These findings highlight the importance of HLA-C and UBE2L3 in the clearance of HBV infection in addition to HLA-DP and HLA-DQ.

HBV affects more than 2 billion people worldwide1. Epidemiologically, the worldwide pattern of hepatocellular carcinoma (HCC) prevalence is similar to that of chronic HBV infection2-3, with the highest infection rate present in the Asia-Pacific region4. A cohort study in Taiwan showed that the risk for HCC in carriers of hepatitis B surface antigen (HBsAg) is about 200 times higher than in HBV non-carriers5. Thus, there is an urgent healthcare need to understand and control chronic HBV infection. Results from twin studies, family clustering studies and studies of differences between ancestry groups have suggested that host factors are critical in determining the outcome of HBV infection6. Recently, genome-wide association studies (GWAS) in Japanese and Korean populations determined that SNPs at HLA-DPA1, HLA-DPB1, HLA-DQA1–HLA-DQA2 and HLA-DQB2 (proxy SNPs rs3077, rs9277535, rs2856718 and rs7453920) are associated with chronic HBV infection7-9. However, in some of these studies, history of HBV exposure in the control groups was unknown, and the general population controls may have had a fraction of potential cases, possibly reducing effects for loci associated with HBV infection and causing some low-penetrance loci associated with chronic HBV infection to be missed10. Therefore, with an effort to confirm previous findings and to identify additional susceptibility loci for chronic HBV infection, we conducted a GWAS in 951 carriers of chronic HBV infection (cases) and 937 subjects who had naturally cleared HBV after acute infection episodes (controls) in a Han Chinese population. This GWAS phase was followed by two replications of independent sample sets of HBV carriers and controls with HBV clearance (replication Ia: southern Jiangsu, 1,248 cases and 1,248 controls and replication Ib: central Jiangsu, 1,000 cases and 1,803 controls) and another two replications of HBV carriers and general population controls (replication IIa: southern China (Guangdong), 981 cases and 1,417 controls and replication IIb: northern China (Beijing), 1,001 cases and 1,205 controls; Supplementary Table 1).

In the GWAS phase, 490,610 qualified autosomal SNPs were analyzed (detailed in the Online Methods) in 951 cases (478 HBV-positive HCC cases and 473 chronic HBV carriers without HCC diagnosis) and 937 controls. To measure the effect of chronic HBV infection alone, independent of the effect of HCC, both pooled and HCC status–stratified analyses were used. For the pooled analysis...
(independent of HCC status), SNPs showing an association at $P \leq 1 \times 10^{-5}$ were selected for further analysis; for the HCC status–stratified analysis, a threshold of $P \leq 1 \times 10^{-7}$ was used. Twenty-two SNPs met these criteria (Supplementary Fig. 1 and Supplementary Table 2).

In linkage disequilibrium (LD) analysis (Supplementary Table 3), 8 of the SNPs showed strong LD ($r^2 > 0.8$); the remaining 14 SNPs were subsequently evaluated in an independent case-control study performed in southern Jiangsu province (replication Ia). Three of the 14 SNPs (rs7453920 at 6p21.32, rs3130542 at 6p21.33 and rs4821116 at 22q11.21) demonstrated significant association ($P < 0.05$) in the same direction as in the GWAS (Supplementary Table 4). These three SNPs were then genotyped in three case-control studies in central Jiangsu province (replication Ib), southern China (Guangdong; replication IIA) and northern China (Beijing; replication IIB) and were all found to be consistently associated with chronic HBV infection. In the combined analysis, all three loci reached genome-wide significance for association with chronic HBV infection ($OR = 0.53, P_{meta} = 4.93 \times 10^{-14}$ for rs7453920 at 6p21.32; OR = 1.33, $P_{meta} = 9.49 \times 10^{-14}$ for rs3130542 at 6p21.33 and OR = 0.82, $P_{meta} = 1.71 \times 10^{-12}$ for rs4821116 at 22q11.21) (Table 1). The rs7453920 SNP is located 1.5 Mb away from rs3130542 in the human leukocyte antigen (HLA) region; conditional analysis showed that these two loci had independent effects in all studies (Supplementary Table 5).

Using imputation analyses based on data from the 1000 Genomes Project (Phase I integrated variant set, v2, March 2012), we tested the associations of SNPs (imputed info > 0.5, minor allele frequency (MAF) > 0.05) located in 150-kb windows centered on the three marker SNPs. We observed a series of significant signals around all three regions (Fig. 1), and 278 SNPs showed association ($P < 1.0 \times 10^{-5}$) with chronic HBV infection (Supplementary Table 6).

Previously, several studies have explored the association between HLA classical alleles and chronic HBV infection; however, the results of these studies were controversial11–13. Given the unique LD pattern in the major histocompatibility complex (MHC) region, we performed HLA allele genotyping in silico on the basis of known SNP genotypes using the R package HIBAG (see URLs). We then further analyzed associations between HLA alleles and chronic HBV infection, finding that six alleles (HLA-*C702, HLA-*DPB1*201, HLA-*DQA1*301, HLA-*DQB1*301, HLA-*DQB1*302 and HLA-*DRB1*1202) were associated with chronic HBV infection ($P < 1.0 \times 10^{-5}$; Supplementary Table 7). To assess whether rs3130542 and the other four HLA SNPs (rs3077, rs9277535, rs2856718 and rs7453920) could tag HLA alleles, we performed LD analysis, finding moderate LD ($r^2 > 0.3$) for rs3130542 and HLA-*C702 ($r^2 = 0.78$), for rs3077 and HLA-*DPB1*501 ($r^2 = 0.49$), and for rs7453920 and HLA-*DQB1*302 ($r^2 = 0.33$) and HLA-*DQA1*301 ($r^2 = 0.32$) (Supplementary Table 7), suggesting that some HLA alleles may be partially tagged by the identified SNPs.

The rs7453920 SNP, located in the first intron of HLA-DQB2, was previously reported in a GWAS of chronic HBV infection in a Japanese population8. A transcriptome study showed that the A allele of rs7453920 was associated with higher HLA-DQ mRNA levels in circulating monocytes, which are critical for mounting immune responses14. Additionally, the A allele in HLA-DQB2 is consistent with a PPARγ-binding site, but the G allele is not15. Previous GWAS also identified HLA variants—rs3077, rs9277535 and rs2856718—that were, in addition to rs7453920, associated with chronic HBV infection16–8. In the present study, we also genotyped these three HLA variants in replications Ia and Ib, finding that they were significantly associated with chronic HBV infection (nominal $P < 0.05$ for known loci), except for rs2856718 in replication Ib (Supplementary Table 8). Specifically, rs3077 at HLA-*DPA1 reached genome-wide significance in the combined analysis ($P = 6.50 \times 10^{-14}$). Differences in association for some loci in Chinese and Japanese populations is probably due to population heterogeneity, and such heterogeneity was even observed among our three study data sets for rs9277535 and rs2856718 (Supplementary Table 8). Such differences may also reflect the different HBV prevalence rates in Chinese and Japanese populations1. In addition, disparate study designs may have also contributed to differences in association, as HBV-exposed controls were used in the current study, whereas general population
controls were used in the Japanese study. So far, several studies in various populations have assessed the association between these HLA SNPs and chronic HBV infection. A meta-analysis of these studies and our new data confirmed that the A alleles of rs3077 and rs9277535 and the G allele of rs2856718 were associated with decreased risk of chronic HBV infection (per-allele OR = 0.63 for rs3077, 0.65 for rs9277535 and 0.73 for rs2856718) (Supplementary Fig. 2). At rs3077 and rs9277535, genotypes were associated with expression of HLA-DPA1 and HLA-DPB1, respectively. Both HLA-DP and HLA-DQ encode type II HLA molecules that have been implicated in the response to HBV infection. HLA-DP and HLA-DQ encode cell surface glycoproteins that bind and present short peptide epitopes to CD4+ T cells. These T cells exhibit a weaker or undetectable activation response in individuals with chronic HBV infection compared to in individuals with resolved infection.

In addition to the known rs7453920 locus in HLA-DQB2, we identified two new loci associated with chronic HBV infection at 6p21.33 (lead SNP: rs3130542, near HLA-C) and 22q11.21 (lead SNP: rs4821116, in UBE2L3). Both HLA-C and UBE2L3 are associated with the risk of developing certain autoimmune diseases, including celiac disease, Crohn’s disease, psoriatic arthritis, systemic lupus erythematosus and rheumatoid arthritis. HLA-C encodes a type I HLA molecule. Increased HLA-C expression in hepatocytes stimulates attack by cytotoxic T cells (CTLs), a key component of virus elimination in chronic HBV carriers. Individuals who successfully clear the virus usually have strong CTL responses. The A allele of rs3130542 was associated with lower expression of HLA-C, which may lead to increased risk of developing chronic HBV infection.

The rs4821116 SNP is located in the last intron of UBE2L3, which encodes the ubiquitin-conjugating enzyme UBC7. Included among the known UBC7 substrates are p53 and nuclear factor (NF)-κB. During the development of chronic HBV infection, NF-κB signaling is frequently stimulated. The A allele of rs4821116, which is associated with natural HBV clearance after acute infection, is associated with higher expression of UBE2L3 (ref. 14). Overexpression of UBE2L3 has been shown to accelerate the degradation of NF-κB and may result in an attenuated immune response to HBV infection.

To identify candidate causal variants, we performed functional annotation for genetic variants that are tagged by the marker SNP rs4821116 (r² > 0.8) on the basis of publically available data sets or tools (Online Methods and Supplementary Table 9). Of the 22 SNPs highly correlated with rs4821116, 17 were significantly associated with the mRNA levels of UBE2L3 (in expression quantitative trait locus (eQTL) analysis). Subsequently, we evaluated whether these SNPs modulate UBE2L3 mRNA levels through a transcriptional or post-transcriptional mechanism. On the basis of a data set for DNase I–hypersensitive site sequencing (DNase-seq) in T H, T H, and T helper T cell lines and T regulatory T cell lines that are frequently activated in the immune response against HBV infection, we found that two SNPs, rs2266959 and rs73166630, were in open chromatin regions, associated with gene regulatory elements. Furthermore, three SNPs (rs5749495, rs5749502 and rs73166630) may affect microRNA binding. In summary, these analyses suggest that genetic variants tagged by rs4821116 may transcriptionally or post-transcriptionally modulate the expression of UBE2L3, which may function in the clearance of HBV infection. However, these results are very preliminary and merit further investigation.

In conclusion, the current study confirmed the previously reported association of chronic HBV infection with genetic variants in HLA-DPA1, HLA-DPB1 and HLA-DQB2 and identified two new associated loci at 6p21.33 (HLA-C) and 22q11.21 (UBE2L3). These findings highlight the importance of both
type I and type II MHC molecules and other immunological molecules, such as UBE2L3, in chronic HBV infection. However, because we only focused on top signals in this study, large-scale evaluation of the remaining potential signals will be required in further studies.


METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank all study subjects, research staff and students who participated in this work and Z. Wang (US National Cancer Institute) for statistical help with imputation. This work was funded by the National Key Basic Research Program (2013CB911400), the project supported by the National Science Foundation for Distinguished Young Scholars of China (81225020), the project supported by the State Key Program of National Natural Science of China (81230067), the Foundation for the Program for New Century Excellent Talents in University (NCET-10-0178), the Author of National Excellent Doctoral Dissertation (201081), the National Natural Science Foundation of China (30800946 and 81072344), the State Key Infectious Disease Project of China (2012ZX10002010 and 2012ZX10002016), the National Major Science & Technology Project (2011ZX00040-902), the National Science Fund for Creative Research Groups (30921006), the National Key Basic Research Program (2011CB910304), a China National High-Tech Research and Development Program grant (2012AA02A315), the Foundation for Distinguished Young Scholars of Jiangsu Province (BK2012042 and BK20130042) and the Priority Academic Program for the Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine).

AUTHOR CONTRIBUTIONS

H.S. and Z.H. directed the study; obtained financial support and were responsible for study design, interpretation of the results and manuscript writing together with W.Z. and Y.S. Z.H. performed overall project management along with Y.L., G.J. and X.Z. and drafted the initial manuscript. J.D. performed statistical analyses along with Y.L., M. Chu, J.W., Y.H. and X.K. were responsible for sample processing and managing the genotyping data. Y.Y. and Y.H. were responsible for subject recruitment and sample preparation for the Shanghai samples. X.Z., L.Z., G.D., Q.W., H.Y. and M. Cao were responsible for subject recruitment and sample preparation for the Zhongshan hospital samples. Jbin Liu, L.L. and X.K. were responsible for subject recruitment and sample preparation for the Nantong samples. L.W., G.Z. and Y.W. were responsible for subject recruitment and sample preparation for the Guangdong samples. Jianjun Liu provided technical and material support. All authors approved the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

ONLINE METHODS

Study subjects. In this study, we performed a three-phase GWAS analysis, including phase I GWAS, phase II replications and phase III replications. The GWAS phase included 951 HBV carriers (478 HBV-positive HCC cases and 473 carriers of chronic HBV infection without HCC from Shanghai) who were included in our previously published study and 937 controls from southern Jiangsu (Zhangjiagang). Zhangjiagang is a suburb about 100 km away from Shanghai. People in these two cities have similar lifestyles and HBV infection rates. Replication Ia samples included 1,248 HBV carriers and 1,248 controls from southern Jiangsu (Zhangjiagang and Changzhou). Replication Ib samples consisted of 1,000 HBV carriers and 1,803 controls from central Jiangsu (Taizhou and Nantong). Replication Ia and Ib samples consisted of 981 HBV carriers and 1,417 controls from southern China (Guangdong) and 1,001 HBV carriers and 1,205 controls from northern China (Beijing), respectively. Cases in all phases were HBV carriers, i.e., positive for HBsAg and antibody specific for the hepatitis B core antigen and were negative for antibody to HCV. Controls in the GWAS and replications Ia and Ib were those who declared no HBV vaccination and naturally cleared HBV, i.e., negative for HBsAg and antibody specific for HCV and positive for antibodies specific for HBsAg and the hepatitis B core antigen. Controls in replications Ia and Ib were from the general population with unknown HBV and HCV status (Supplementary Table 1).

A 5-ml sample of whole blood was obtained from each participant as a source of genomic DNA for further genotyping analyses and serological testing for HBV and HCV markers. Informed consent was obtained from each subject at recruitment, and this study was approved by the institutional review boards of all of the participating institutions.

Quality control in the GWAS. We performed systematic quality control analysis on the raw genotyping data before conducting the association analysis. HBV carriers were genotyped using Illumina HumanOmniExpress12v1 chips (731,442 SNPs) as described previously, and controls were genotyped with OmniZhongHua chips (900,015 SNPs; see flow chart in Supplementary Fig. 3). SNPs available on both chips (595,310 SNPs; 81.4% for the HumanOmnExpress12v1 chip) were further analyzed. SNPs were excluded if they (i) did not map to autosomal chromosomes, (ii) had a call rate of <95%, (iii) had a MAF of <0.05 or (iv) had a genotype distribution in the controls that deviated from the expected Hardy-Weinberg equilibrium ($P < 1.0 \times 10^{-5}$). Samples with overall genotype completion rates of less than 95% were excluded from further analysis (41 subjects). No subjects were excluded for sex discrepancies between our records and the genetically inferred data. We detected 21 unexpected duplicates or probable relatives, which were excluded on the basis of identity-by-descent (IBD) analysis implemented with PLINK1.07 (ref. 45) (all PL_HAT > 0.25). We detected population outliers and stratification using EIGENSTRAT 3.0 software to carry out a principal-component analysis (PCA). A set of 48,413 common autosomal SNPs (MAF > 0.25) with low LD ($r^2 < 0.05$) were used to identify population outliers ($n = 33$) in the samples that had passed quality control, using the founders of the HapMap trios of Yoruba in Ibadan, Nigeria (YRI; $n = 90$), Utah residents of Northern and Western European ancestry (CEU; $n = 90$), Han Chinese in Beijing, China (CHB; $n = 45$) and Japanese in Tokyo, Japan (JPT; $n = 44$) as internal controls (Supplementary Fig. 4a). PCA showed that the cases and controls were genetically matched and homogeneous (Supplementary Fig. 4b) and that the genomic control inflation factor ($\lambda$) was 1.07. After quality control procedures had been performed, a total of 951 cases and 937 controls with 490,610 SNPs were included in the final analysis.

SNPs genotyping in the replication phases. Genotyping for replication Ia was performed using the iPLEX platform (Sequenom), and replications Ib, Iia and Iib were performed using TaqMan assays (Applied Biosystems). Laboratory technicians who performed genotyping were blinded to the case-control status of samples. In each 384-well plate, there were similar numbers of case and control samples.

Functional annotation based on publically available databases. We performed eQTL analysis using the University of Chicago eQTL Browser (see URLs), which contains significant eQTLs that were identified in recent studies across multiple cell lines and populations. Open chromatin regions recognized by DNase-seq are associated with gene regulatory elements, including promoters, enhancers, silencers, insulators and locus control regions. We determined whether SNPs were located in the DNase-seq peaks of $\text{T}_2\theta$, $\text{T}_2\lambda_1$, $\text{T}_2\lambda_2$ and $\text{T}_{\text{reg}}$ cell lines by using the ENCODE database (see URLs). The extent to which SNPs affected microRNA binding was predicted by Patrocles (see URLs).

Statistical analysis. We used PLINK 1.07 (ref. 45) and R 2.15.1 (see URLs) for general statistical analysis.

As determined using PGA software (see URLs), all SNPs identified in our study could meet the 80% level of statistical power (given 2,000 cases and 2,000 controls, MAF = 0.05 and OR = 1.31). Population structure was evaluated by PCA using the EIGENSTRAT 3.0 software package. This analysis identified one significant ($P < 0.05$) eigenvector, which was included as a covariate in the logistic regression analysis along with age and sex. SNPs for all phases were combined using a meta-analysis–based method, which used a fixed-effect model with inverse variance weighting when there was no indication of heterogeneity ($P$ for Cochran’s $Q$ statistic > 0.05). Otherwise, we adopted a random-effect model (DerSimonian-Laird) for corresponding SNPs. The Manhattan plot of −log$_{10}$ ($P$ values) was generated using R 2.15.1. Ungenotyped SNPs were imputed in the GWAS discovery samples using ShapeIt 1.0 (ref. 47) (phasing step), IMPUTE2 (imputation step) and haplotype information from the 1000 Genomes Project (Phase I integrated variant set across all 1,092 individuals, v2, March 2012; see URLs). Regional plots were generated using an online tool, LocusZoom 1.1 (see URLs). HLA alleles were predicted from dense SNP genotypes using the R package HIBAG (see URLs). ORs and 95% CIs were calculated with an additive model by logistic regression analyses, except where otherwise specified.