

Genome-wide Association Study Identifies New Loci Associated With Risk Of HBV Infection And Disease Progression

Zheng Zeng (✉ zeng@bjmu.edu.cn)

Department of Infectious Diseases, Peking University First Hospital, Beijing 100034, China; <https://orcid.org/0000-0002-4997-4223>

Liu Hankui

BGI-Shenzhen: BGI Group

Xu Huifang

BGI-Shenzhen: BGI Group

Lu Haiying

Peking University First Hospital

Yu Yanyan

Peking University First Hospital

Xu Xiaoyuan

Peking University First Hospital

Yu Min

Peking University First Hospital

Zhang Tao

BGI-Shenzhen: BGI Group

Tian Xiulan

Peking University First Hospital

Xi Hongli

Peking University First Hospital

Guan Liping

BGI-Shenzhen: BGI Group

Zhang Jianguo

BGI-Shenzhen: BGI Group

O'Brien J. Stephen

Nova Southeastern University

Research article

Keywords: HBV infection, disease progression, GWAS, Host genetic factors, SNPs

Posted Date: January 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-143403/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on March 18th, 2021. See the published version at <https://doi.org/10.1186/s12920-021-00907-0>.

Abstract

Background: Recent studies have identified susceptibility genes of HBV clearance, chronic hepatitis B, liver cirrhosis, hepatocellular carcinoma, and showed the host genetic factors play an important role in these HBV-related outcomes.

Results: In order to discover new susceptibility genes for HBV-related outcomes, we conducted a genome-wide association study in 1031 Chinese participants, including 275 HBV clearance subjects, 92 asymptomatic persistence infection carriers (ASPI), 93 chronic hepatitis B patients (CHB), 188 HBV-related decompensated cirrhosis patients (DC), 214 HBV-related hepatocellular carcinoma patients (HCC) and 169 healthy controls (HC). In the case-control study, we observed novel locus significantly associated with CHB (SNP: rs1264473, Gene: *GRHL2*, $P = 1.57 \times 10^{-6}$) and HCC (SNP: rs2833856, Gene: *EVA1C*, $P = 1.62 \times 10^{-6}$; SNP: rs4661093, Gene: *ETV3*, $P = 2.26 \times 10^{-6}$). In the trend study across progressive stages post HBV infection, one novel locus (SNP: rs1537862, Gene: *LACE1*, $P = 1.85 \times 10^{-6}$), and three MHC loci (*HLA-DRB1*, *HLA-DPB1*, *HLA-DPA2*) showed significant increased progressive risk from ASPI to CHB. Interestingly, underlying the evolutionary study of HBV-related genes in public database, we found that the derived allele of two HBV clearance related locus, rs3077 and rs9277542, are under strong selection in European population.

Conclusions: In this study, we identified several novel candidate genes associated with individual HBV infectious outcomes, progressive stages, and liver enzymes. Moreover, we identified two SNPs that show selective significance (*HLA-DPA1*, *HLA-DPB1*) in non-East Asian (European, American, South Asian) versus East Asian, indicating that host genetic factors contribute to the ethnic disparities of susceptibility of HBV infection. Taken together, these findings provided a new insight into the role of host genetic factors in HBV related outcomes and progression.

Background

Hepatitis B virus (HBV) infection is one of the most common infectious diseases, with about 248 million HBsAg positive individuals worldwide and the largest HBsAg positive population in China [1]. HBV infection can develop a wide spectrum of liver diseases, including chronic hepatitis B, liver cirrhosis, hepatocellular carcinoma [2–4]. Previous studies showed the host genetic factor played a critical role in HBV infection susceptibility and identified associated SNPs with significant contribution, including major histocompatibility complex (MHC) genes, i.e. *HLA-DPA1* (rs3077), *HLA-DPB1* (rs9277535), *HLA-C* (rs3130542), *HLA-DQ* (rs2856718, rs7453920) [5–7], and non-MHC genes, i.e. *UBE2L3* (rs4821116), *INTS10* (rs7000921) [8, 9]. In advanced stages of HBV disease, host genetic factors influence the outcome of HBV infection [7, 10, 11], including *HLA-DQ* (rs9275319), *HLA-DRB1* (rs2647073, rs3997872), *STAT4* (rs7574865), *C2* (rs9267673), *PNPLA3* (rs738408, rs738409), *SLC17A2* (rs80215559), *HFE* (rs1800562) [12, 13] for liver cirrhosis and *KIF1B* (rs17401966), *HLA-DQA1/DRB1* (rs9272105), *HLA-DQ* (rs9275319), *STAT4* (rs7574865) for hepatocellular carcinoma [14–16]. However, these reported HBV-related genes confer relatively small increments in risk and explain a small proportion of heritability. For example, although MHC genes are important for immune response to HBsAg, more than half the heritability is determined by non-MHC genes [17]. Moreover, previous studies showed that the MHC genes share a common influence on HBV infection, liver cirrhosis, hepatocellular carcinoma [6, 12, 15, 16] as well as associating with different risk in these outcomes [18]; i.e. *HLA-DQ*, *STAT4*, *C2*, *HLA-DRB1* for liver cirrhosis and HCC [12], *HLA-DQ* for CHB [6]. These consistent [12] or different [18] risks indicated shared but also modified effects for progressive HBV-related outcomes. These results raised our interest to identify host genetic factor which increases the risk of progressive stages post HBV infection. To reveal new susceptibility genes for HBV infection and the HBV-related outcomes, we performed a genome-wide association study (GWAS) in 1031 participants, including 275 HBV clearance subjects, 92 asymptomatic persistence infection carriers (ASPI), 93 chronic hepatitis B patients (CHB), 188 HBV-related decompensated cirrhosis patients (DC), 214 HBV-related hepatocellular carcinoma patients (HCC) and 169 healthy controls (HC) (Table 1).

Table 1
Characteristics of participants in the genome-wide association cohorts

Disease categories	HBV clearance	ASPI	CHB	DC	HCC	HC
Sample size	275	92	93	188	214	169
Mean age \pm SD	49.56 \pm 8.8	46.89 \pm 6.9	46.46 \pm 5.7	50.65 \pm 8.4	51.34 \pm 10.5	48.82 \pm 7.1
Male/female	105/170	33/59	62/31	148/40	183/31	73/96
ALT, U/L, mean \pm SD	26.95 \pm 30.42	24.5 \pm 8.79	169.53 \pm 243.76	89.57 \pm 115.79	214 \pm 381.67	23.11 \pm 7.97
AST, U/L, mean \pm SD	25.89 \pm 21.89	23.16 \pm 6.45	100.04 \pm 127.54	102.46 \pm 118.36	120.51 \pm 286.05	22.23 \pm 7.00
TBiL, μ mol/L, mean \pm SD	13.68 \pm 9.93	12.28 \pm 2.91	23.12 \pm 35.60	65.87 \pm 83.22	50.11 \pm 90.93	12.95 \pm 4.09
DBiL, μ mol/L, mean \pm SD	5.86 \pm 18.39	3.59 \pm 1.68	9.36 \pm 14.29	35.15 \pm 51.33	25.87 \pm 55.04	4.54 \pm 9.46
ALP, U/L, mean \pm SD	74.30 \pm 35.86	73.57 \pm 44.58	108.89 \pm 42.20	132.22 \pm 61.42	141.63 \pm 104.32	70.85 \pm 25.33
GGT, U/L, mean \pm SD	28.34 \pm 30.77	29.88 \pm 21.19	88.76 \pm 106.63	89.94 \pm 132.10	158.08 \pm 183.92	27.58 \pm 24.88
ALB, g/L, mean \pm SD	43.33 \pm 6.11	44.16 \pm 6.71	42.40 \pm 42.25	32.28 \pm 7.13	37.19 \pm 7.27	43.93 \pm 4.96
AFP, μ g/L, mean \pm SD	21.39 \pm 47.39	5.16 \pm 4.79	34.00 \pm 89.37	87.74 \pm 169.83	7315.22 \pm 37329.94	3.76 \pm 4.34
PTA, %, mean \pm SD	95.22 \pm 52.69	91.67 \pm 14.30	90.06 \pm 21.76	63.97 \pm 28.07	87.31 \pm 33.96	90.84 \pm 18.71
PLT, 10 ⁹ /L, mean \pm SD	147.69 \pm 55.40	146.84 \pm 44.51	150.27 \pm 49.61	73.96 \pm 45.30	145.72 \pm 79.24	148.94 \pm 49.10
Abbreviations: ASPI, asymptomatic persistence infection; CHB, chronic hepatitis B; DC, decompensated cirrhosis; HCC, hepatocellular carcinoma; HC, healthy controls; SD, standard deviation; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TBiL, total bilirubin; DBiL, direct bilirubin; ALP, alkaline phosphatase; GGT, glutamyl transpeptidase.						

Methods

Study Participants

A total of 1104 unrelated, age- and gender- matched, Chinese participants were recruited in the study, enrollment criteria were consistent with a previous report [19]. The population of HBV-related phenotypes was composed of five subgroups: HBV clearance subjects, asymptomatic persistence infection (ASPI) carriers, chronic hepatitis B (CHB) patients, HBV-related decompensated cirrhosis (DC) patients, HBV-related hepatocellular carcinoma (HCC) patients. Healthy controls (HC) who were HBV serum marker-negative (HBsAg, anti-HBc) and had no serological evidence of co-infection with HCV, HDV, and HIV were also included. HBV chronic infection patients were diagnosed based on seropositivity of HBsAg at least 6 months. Then ASPI was defined as HBsAg and anti-HBc positive at least 6 months and serum alanine aminotransferase (ALT), aspartate aminotransferase (AST) in normal values without abnormal before. CHB is defined as HBsAg and anti-HBc positive at least 6 months and ALT, AST abnormal before or at enrollment. DC was defined as HBsAg and anti-HBc positive at least 6 months with decompensated portal hypertension (gastroesophageal bleeding, ascites, edema or encephalopathy) or decompensated liver function (albumin < 35 g/L and total bilirubin > 35 μ mol/L). HCC was defined at least one of following: (a) liver biopsy; or (b) abnormal alpha fetoprotein (AFP) and sonographic, CT or MRI space occupying evidence.

Clinical Parameters

Clinical parameters including serum alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin (TBIL), direct bilirubin (DBIL), alkaline phosphatase (ALP), glutamyl transpeptidase (GGT), albumin (ALB), globulin (Glo), alpha fetoprotein (AFP), prothrombin time activity (PTA), platelets (PLT), HBsAg, anti-HBs, HBeAg, anti-HBe, anti-HBc were collected from hospital information system. Other baseline characteristics were recorded during each patient's clinical examination. In brief, liver biochemistry and virological tests were carried out by Bechman Coulter AU chemistry analyzers, chemiluminescence immunoassays (AxSYM or ARCHITECT I2000, Abbott, USA) or Ortho/Chemi-luminescent assay (Johnson and Johnson Co., USA) with commercially available kits; Anti-HAV IgM antibody, HDV antigen (HDAg) and anti-HDV antibody, and anti-HEV antibody were determined by commercially ELISA kits in China. For HBV DNA level, it was quantified using commercial real-time polymerase chain reaction kit with a lower limit of detection (LLOD) of 100 IU/ml (Daan company, China) or Roche Cobas Ampliprep/Cobas Taqman™ PCR assay with LLOD of 20 IU/ml (Roche, USA).

Genome-wide SNP Genotyping and Quality Control

Genotyping was performed on Affymetrix 500 k Genome-Wide Human SNP Array 6.0 (http://www.affymetrix.com/Auth/analysis/downloads/na35/genotyping/GenomeWideSNP_6.na35.annot.csv.zip). SNPs met the following quality control procedures were excluded: 1) call rate < 95%; 2) minor allele frequency (MAF) < 1%; 3) genotype in controls deviated from the Hardy Weinberg equilibrium (HWE test P-value < 10^{-5}). Overall, 1,031 samples and 607,153 variants on 22 chromosomes were included in the final analysis.

Statistics Analysis

GCTA tool [20] was used to perform principal component analyses for estimating population substructure. The first two eigenvectors, pc1 and pc2, were used to display the population structure. PLINK 1.9 [21] software was used to perform logistic regression for identifying susceptibility SNPs of HBV infection and HBV-related outcomes. Gender and age were used as covariates in logistic regression. Chi-square test for trend in proportions was used to identify SNPs with increased effectiveness on disease progression. We used the Bonferroni method to adjust the false positive rate caused by multiple test. The number of independent LD block was used to represent the number of independent multiple test. We calculated a total of 21,077 independent LD blocks via GEC [22] and then set 0.05/21077 as the threshold of genome-wide significance. The genomic control method was used to measure population stratification by calculating the genomic inflation factor (λ) from median P-value. ANOVA was used to evaluate the significance of the association between biomarkers and genotypes in healthy controls. Using the SNPs in HBV infection-related loci in 1000 Genomes Project [23], we performed evolutionary analyses, including building phylogenetic tree, detecting the signatures of selection, displaying the core haplotypes, estimating effective population size. Derived allele and ancestral allele of SNPs were accessed from Ensemble human ancestral genome (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/ancestral_alignments). PoMo [24], an allele frequency-based approach, was used to build the racial tree based on the allele frequency of SNPs in each population. F_{ST} [25], a classical metrics of population differentiation, was widely employed in detecting signatures of selection [26] in human genome [27, 28] and animal genome [29–31]. In our study, F_{ST} was implemented to detect the selective signature between East Asian population and each other population. VcfTools [32] was used to calculate the F_{ST} statistics of SNPs in paired populations. F_{ST} statistics accesses 0.15 [33] was used as a threshold to detect the signature of selection. Rehh package [34, 35] was used to display the haplotype bifurcation diagrams of the associated SNPs in different populations. Relate [36], a method for genome-wide genealogy estimation for thousands of samples, was used to estimate the historical population size at default setting.

Results

The demographic and clinical characteristics of 1031 study participants included in our association study are presented in Table 1. All participants were genotyped by Affymetrix 500 k SNP Array. A total of 607,153 SNPs passed through quality control (Supplementary Figure S1). These SNPs filtered minor allele frequency of < 1% and a call rate of < 95%.

To demonstrate that there is no genetic stratification in the population, we performed a principal component analysis on the SNPs of all participants. The first two principal components show absence of population structure (Supplementary Figure S2). To identify susceptibility SNPs for HBV infection, we performed a GWAS in HBV infection similar with previous design [8, 9]. HBV

clearance was used as a control group versus ASPI, CHB, DC, HCC as HBV chronic infection (case group). We observed associations of two novel MHC loci with progression to certain HBV stages (SNP: rs2395166, Gene: *HLA-DRA*, $P = 1.42 \times 10^{-7}$; SNP: rs615672, Gene: *HLA-DRB1*, $P = 8.54 \times 10^{-7}$) and two reported MHC locus (SNP: rs3077, Gene: *HLA-DPA1*, $P = 6.60 \times 10^{-9}$; SNP: rs9277542, Gene: *HLA-DPB1*, $P = 1.53 \times 10^{-8}$) (Table 2; Fig. 1). These MHC loci variants replicated association results of previous studies affirming that MHC gene alleles confer risks of susceptibility of HBV infection in East Asian. Interestingly, we founded that these reported MHC loci (rs2395166:C, rs615672:G, rs3077:A, rs9277542:T, rs9277341:T) present significant differences in allele frequency between East Asian and non-East Asian population in gnomAD database (Table 3), as well as the differences between HBV infection group and HBV clearance group. Since different groups may not present an identical minor allele, here, we used the derived allele against the ancestral allele for studying the allele frequency across different populations. The derived allele frequencies in East Asian are much closer to the HBV chronic infection group, while other populations, such as European, are much closer to the HBV clearance group. These genetic differences may suggest a selective signal in non-East Asian population versus East Asian population. To confirm this, we firstly build a phylogenetic tree based on these loci and then showed the genetic diversity in world-wide populations, in which the East Asian population is at the root. We set the East Asian as the ancestral group in these loci according to the derived allele frequencies and the phylogenetic tree. Subsequently, we identified two strong phylogenetic signals (*HLA-DPA1*, *HLA-DPB1*) in the European population (Fig. 2) via F_{ST} method. Haplotype bifurcation diagrams of the two core SNPs (rs3077, rs9277542) presented that the resisted allele lead to a long-range, and a high frequency homozygosity in European population (Fig. 3), confirming the natural genetic selection. These evidences revealed that the resisted alleles were under positive selection in European population strongly. We estimated the historic population size and then showed these two loci (*HLA-DPA1*, *HLA-DPB1*) were under selection during the past 26,000 years (Supplementary Figure S3). These results may provide a context for the racking influence of HBV infectious diseases in history.

Table 2
The significance of HBV-related outcomes study

Case-control studies		SNP	Gene	<i>P</i> value	OR	Minor Allele	Minor Allele Frequency		Report	λ
Case (n)	Control (n)						Case	Control		
Infection (587)	Clearance (275)	rs2395166	<i>HLA-DRA</i>	1.42×10^{-7}	0.4534	C	0.1269	0.2182	MHC region	1.003
		rs615672	<i>HLA-DRB1</i>	8.54×10^{-7}	0.5697	G	0.405	0.5276	MHC region	
		rs3077	<i>HLA-DPA1</i>	6.60×10^{-9}	0.5007	A	0.2675	0.4145	(Kamatani et al., 2009)	
		rs9277542	<i>HLA-DPB1</i>	1.53×10^{-8}	0.5353	T	0.3735	0.5347	(Kamatani et al., 2009)	
CHB (93)	ASPI (92)	rs1264473	<i>GRHL2</i>	1.57×10^{-6}	3.931	C	0.4402	0.1957	Novel	1.052
HCC (214)	CHB (93)	rs2833856	<i>EVA1C</i>	1.62×10^{-6}	0.3515	C	0.2243	0.4086	Novel	1.022
HCC (214)	DC (188)	rs4661093	<i>ETV3</i>	2.26×10^{-6}	2.841	A	0.2104	0.0882	Novel	1.022

Abbreviations: OR, odds ratio; ASPI, asymptomatic persistence infection; CHB, chronic hepatitis B; DC, decompensated cirrhosis; HCC, hepatocellular carcinoma; MHC, major histocompatibility complex; λ : statistic of genomic control.

Table 3: Divided allele frequency of significant SNPs in MHC region

SNP	Study	Population	Derived Allele	Derived Allele Frequency					P value
				Case-control studies			gnomAD		
				Infection	Clearance	Healthy	East Asian	non-East Asian	
rs2395166	our study	Chinese	C	0.127	0.218	0.183	0.128	0.367	4.367×10 ⁻¹⁹⁵
rs615672	our study	Chinese	G	0.405	0.528	0.512	0.388	0.590	6.003×10 ⁻¹¹¹
rs3077	our study	Chinese	A	0.268	0.415	0.379	0.283	0.723	0
	Kamatani et al., 2009 [5]	Japanese		0.245	-	0.392			
	Guo et al., 2011 [55]	Chinese		0.314	0.447	0.443			
	Nishida et al., 2012 [56]	Japanese, Korean		0.213	0.393	-			
	Wong et al., 2013 [57]	Southern Chinese		0.206	0.276	0.288			
rs9277341	our study	Chinese	T	0.142	0.242	0.196	0.159	0.582	0
	Guo et al., 2011 [55]	Chinese		0.133	0.237	0.237			
rs9277542	our study	Chinese	T	0.374	0.535	0.482	0.339	0.627	2.005×10 ⁻²²⁵
	Kamatani et al., 2009 [5]	Japanese		0.246	-	0.437			

Abbreviations: East Asian, East Asian population in gnomAD database; non-East Asian, combined all other population excepted East Asian in gnomAD database; *P* value, compared allele frequency between East Asian and non-East Asian population via fisher exact test. Derived Allele was accessed from human ancestral genome (Ensembl-59).

To identify new susceptibility locus for HBV-related outcomes, we performed association studies for CHB, DC, and HCC. Significantly, we observed three associated gene SNP loci: 1) (SNP: rs1264473, Gene: *GRHL2*, $P = 1.57 \times 10^{-6}$) associated with CHB versus ASPI; 2) (SNP: rs2833856, Gene: *EVA1C*, $P = 1.62 \times 10^{-6}$) associated with HCC versus CHB; and 3) (SNP: rs4661093, Gene: *ETV3*, $P = 2.26 \times 10^{-6}$) associated with HCC versus DC (Table 2; Fig. 1). No SNP associated with DC versus CHB were apparent.

HBV clearance, ASPI, CHB, DC, and HCC are progressive stages post HBV infection [4]. We hypothesized that the host genetic factor contributes to the development of outcomes, as well as to the individual outcome. To investigate this hypothesis, we test two progressive stages upon HBV infection: 1.) HBV infection itself (CHB, ASPI, and HBV clearance) and 2.) development of CHB (CHB, DC, and HCC). We performed a chi-square test for trend in proportions of allele to identify SNPs increasing risk of HBV-related outcomes in the progressive stages. We observed association with one novel locus (SNP: rs1537862, Gene: *LACE1*, $P = 1.85 \times 10^{-6}$), one association with a reported locus (SNP: rs9277542, Gene: *HLA-DPB1*, $P = 1.50 \times 10^{-9}$), and two association variants at MHC genes (SNP: rs615672, Gene: *HLA-DRB1*, $P = 1.39 \times 10^{-6}$; SNP: rs3128923, Gene: *HLA-DPA2*, $P = 2.06 \times 10^{-6}$)

with trend test of allele frequency across three outcomes (Table 4; Fig. 4A). The three reported MHC genes were demonstrated to play a critical role in the resistance of HBV infection, and two (*HLA-DPB1*:rs9277542, *HLA-DRB1*:rs9277542) were identified to be associated with HBV clearance (Table 2). We did not observe any SNPs achieve genome-wide significant association with development of CHB; Two additional loci (SNP: rs6942409, Gene: *AC011288.2*, $P = 3.08 \times 10^{-6}$) and the HCC associated locus (SNP: rs2833856, Gene: *EVA1C*, $P = 1.62 \times 10^{-5}$) were associated with increased risk of DC and HCC during the development of CHB (Table 5; Fig. 4B).

Table 4
The significance of progressive HBV infection study

SNP	Gene	P value	Resistant Allele	Resistant Allele Frequency			Related Risk	
				CHB	ASPI	Clearance	CHB (vs ASPI)	Clearance (vs ASPI)
rs615672	<i>HLA-DRB1</i>	1.39×10^{-6}	G	0.3297	0.4185	0.5276	0.82	1.12
rs9277542	<i>HLA-DPB1</i>	1.50×10^{-9}	T	0.3011	0.375	0.5347	0.84	1.17
rs3128923	<i>HLA-DPA2</i>	2.06×10^{-6}	G	0.3804	0.4348	0.5636	0.89	1.14
rs1537862	<i>LACE1</i>	1.85×10^{-6}	C	0.6	0.6374	0.7647	0.92	1.19

Abbreviations: ASPI, asymptomatic persistence infection; CHB, chronic hepatitis B. RR was calculated with the comparison of CHB and ASPI, Clearance and ASPI respectively.

Table 5
The suggestive significance of progressive CHB study

SNP	Gene	P value	Risk Allele	Risk Allele Frequency			Related Risk	
				CHB	DC	HCC	DC (vs CHB)	HCC (vs CHB)
rs6942409	<i>AC011288.2</i>	3.08×10^{-6}	G	0.6129	0.7287	0.7958	1.20	1.37
rs2833856	<i>EVA1C</i>	1.62×10^{-5}	T	0.5914	0.75	0.7757	1.30	1.35

Abbreviations: CHB, chronic hepatitis B; DC, decompensated cirrhosis; HCC, hepatocellular carcinoma. RR was calculated with the comparison of HCC and CHB, DC and CHB respectively.

Host genetic factors were demonstrated to influence concentrations of liver enzymes in plasma, which are widely used to indicate liver disease [37, 38]. Here, to investigate the functional change in liver influenced by the HBV related loci described above, we performed a variance analysis in 10 clinical parameters of serum liver enzymes (ALT, AST, TBIL, DBIL, ALP, GGT, ALB, AFP, PTA, and PLT) between different genotypes in healthy controls (Supplementary Figure S4-9). Six loci (rs1537862, rs3128923, rs9277542, rs9277341, rs9277378, rs4661093) showed modest associations with concentrations of liver enzymes, including ALB, ALP, AFP, and PTA (Fig. 5). These associations suggest pathways linking the host genetic factors, metabolism, and liver function for understanding the mechanisms of infection and disease progression.

In sum, our study identified susceptibility SNPs associated with HBV related outcomes and SNPs increased the risk of progressive outcomes from HBV clearance to HBV chronic infection, DC, and HCC in a Chinese population (Supplementary Figure S10).

Discussion

HBV infection leads to a wide spectrum of clinical outcomes, including spontaneous clearance, asymptomatic carrier, chronic hepatitis B, liver cirrhosis, and hepatocellular carcinoma. Previous studies showed that MHC genes played an important role in outcomes of HBV infection [7]. Alleles associated with HBV infection versus HBV clearance affect infection risk, and a low-risk allele indicated an effect on virus clearance. By contrast loci associated with CHB versus ASPI indicated risk for the severe progression, while a low-risk allele affected tolerance of virus. The tolerance-related gene, *GRHL2*, was demonstrated to influence the inflammation in hepatocytes by regulating microRNA 122 (*MIR122*) and the target of *MIR122*, HIF1 α [39]. Levels of *GRHL2* were increased in liver tissues of patients with alcoholic liver disease and correlated with decreases in levels of *MIR122*. Increased levels of *MIR122* in hepatocytes of mice with ethanol-induced liver disease and advanced fibrosis reduced levels of HIF1 α and reduced serum levels of alanine aminotransferase (ALT). Taken together, we propose that the low-risk allele rs1264473:T at *GRHL2* ablates severe persistent inflammation through increased the levels of *MIR122*.

Our previous studies [40, 41] showed that *NTCP*S267F mutation significantly affected the disease progression to cirrhosis ($P=0.017$), and hepatocellular carcinoma ($P=0.023$) versus CHB [40] and the rs3077:T allele was associated with decreased risk of chronic HBV infection (OR = 0.62, $P=0.001$) [41]. In this study, we searched for host genetic factor with increased risk of the development-related outcomes in GWAS. One novel locus, *LACE1*, and three infection-related MHC loci were associated the progression of HBV infection. These results showed that the host genetic factors, both MHC and non-MHC genes, increased the risk of progressive outcomes post HBV infection, as well as HBV mutation. It is reported that HBV infection altered the mitochondrial metabolism and mitochondrial dynamics, which result in mitochondrial injury and liver disease [42]. *LACE1* was reported to affect mitochondrial protein homeostasis [43]. Knockdown of *LACE1* converted the expression of a crucial component of regulating mitochondrial dynamics, *OPA1* [43–45]. In addition, we found that the risk allele, *LACE1*:rs1537862:T, decreased the level of ALB significantly ($P=0.025$, Fig. 5). ALB is a critical marker decreasing with the deterioration of chronic liver diseases [46–48]. Biosynthesis of ALB was affected by proinflammatory cytokines [49, 50] and excess amounts of oxidative agents released by mitochondria from injured liver [46, 51]. Taken together, we proposed *LACE1* may affect hepatic infection by changing the hepatic mitochondrial metabolism and leading to the progression of HBV infection.

There is a limitation in our study, that is we do not have an additional cohort for replicate study. In spite of that, we showed the reported loci in MHC region are significantly related to HBV infection. These replicate results of previous studies confirm our findings are reliable and provide confidence for our study in this cohort. Here, we provide novel candidate genes related to individual outcomes, progressive stages, and liver enzymes. Moreover, we identified two SNPs that show selective significance (*HLA-DPA1*, *HLA-DPB1*) in non-East Asian (European, American, South Asian) versus East Asian. East Asian populations seem more susceptible to HBV infection than non-EAS Asian, and the differences of susceptibility were affected by HBV genotype [52], immunity [53], and environmental exposure [53, 54]. Even in an identical environment (United States), Asian are more prevalent in chronic HBV infection than non-Asian [53]. It seems likely that host genetic factors contribute to the ethnic disparities of susceptibility of HBV infection. Taken together with the genetic associations and evolutionary signals, our findings provide a new insight for HBV study.

Conclusion

In case-control study, we identified one novel locus (SNP: rs1264473, Gene: *GRHL2*, $P=1.57 \times 10^{-6}$) significantly associated with CHB, two novel locus (SNP: rs2833856, Gene: *EVA1C*, $P=1.62 \times 10^{-6}$; SNP: rs4661093, Gene: *ETV3*, $P=2.26 \times 10^{-6}$) significantly associated with HCC. In trend study across multiple outcomes, we identified one novel locus (SNP: rs1537862, Gene: *LACE1*, $P=1.85 \times 10^{-6}$) and three MHC loci (*HLA-DRB1*, *HLA-DPB1*, *HLA-DPA2*) significantly increased progressive risk from CHB through ASPI to HBV clearance. In evolutionary study, we showed the derived allele of two HBV clearance related locus, rs3077 and rs9277542, are under strong selection in European population. We suggested these selected alleles may play a role in resisting the susceptibility of HBV in Europeans. Our findings provided a new insight into the role of host genetic factors in HBV related outcomes and progression.

Abbreviations

AFP	Alpha Fetoprotein
AFR	African
ALB	Albumin
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AMR	American
anti-HBc	Antibody to Hepatitis B core Antigen
anti-HBe	Antibody to Hepatitis B e Antigen
anti-HBs	Antibody to Hepatitis B surface Antigen
anti-HDV	Antibody to Hepatitis D Virus
anti-HEV	Antibody to Hepatitis E Virus
ASPI	Asymptomatic Persistence Infection
AST	Aspartate Aminotransferase
CHB	Chronic Hepatitis B
DBIL	Direct Bilirubin
DC	Decompensated Cirrhosis
EAS	East Asian
EUR	European
Glo	Globulin
GGT	Glutamyl Transpeptidase
GWAS	Genome Wide Association Study
HBeAg	Hepatitis B e Antigen
HBsAg	Hepatitis B surface Antigen
HBV	Hepatitis B Virus
HCC	Hepatocellular Carcinoma
HC	Healthy Controls
HCV	Hepatitis C Virus
HDV	Hepatitis D Virus
HIV	Human Immunodeficiency Virus
HWE	Hardy-Weinberg Equilibrium
MAF	Minor Allele Frequency
MHC	Major Histocompatibility Complex
PI	Persistence Infection
PLT	Platelets
PTA	Prothrombin Time Activity

AFP	Alpha Fetoprotein
SAS	South Asian
SD	Standard Deviation
SNP	Single Nucleotide Polymorphism
TBIL	Total Bilirubin

Declarations

Ethics approval and consent to participate

This study was carried out following the guidelines of the 1975 Declaration of Helsinki. Informed written consent was obtained from all participants and the study was performed with the approval of the Ethics Committee of Peking University First Hospital and local cooperation hospitals.

Consent to publish

All authors had full access to the final version of the manuscript and agreed to its submission for publication..

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available as they are still being investigated but are available from the corresponding author (Dr. Zheng Zeng) on reasonable request.

Competing interests

The authors declared that they have no conflict of interest.

Funding

This study was supported by the National Natural Science Foundation of China (No. 61972007, No. 30671855), the International Science & Technology Cooperation Program of China (No. 2014DFR31200), Federal funds from the National Cancer Institute, National Institutes of Health, USA (No. N01-CO-12400), Shenzhen Municipal of Government of China (JCYJ20170412153248372, JCYJ20180507183615145).

Authors' contributions

ZZ, HKL, HYL, HFX contributed equally. SJO and ZZ designed the study. ZZ, HYL, MY, XLT, HLX, and HBV study consortium collected samples and clinical data. ZZ, XLT maintained database. HYL and ZZ analyzed clinical data. HKL, ZZ, HFX, TZ, LPG, and JGZ performed bioinformatic analyses. ZZ, SJO, and JGZ reviewed and revised the manuscript. SJO and ZZ supervised the research and made final decision. All authors had full access to the final version of the manuscript and agreed to the submission.

Acknowledgements

We thank the participants for their involvement in this study.

References

1. Schweitzer A, et al. Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. *The Lancet*. 2015;386(10003):1546–55.

2. Fattovich G, et al. Occurrence of hepatocellular carcinoma and decompensation in western European patients with cirrhosis type B. *Hepatology*. 1995;21(1):77–82.
3. Liaw Y-F, Chu C-M. Hepatitis B virus infection. *The Lancet*. 2009;373(9663):582–92.
4. Sundaram V, Kowdley K. Management of chronic hepatitis B infection. *Bmj*. 2015;351:h4263.
5. Kamatani Y, 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(5):591.
6. Mbarek H, et al. A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population. *Human molecular genetics*. 2011;20(19):3884–92.
7. Zeng Z. Human genes involved in hepatitis B virus infection. *World journal of gastroenterology: WJG*. 2014;20(24):7696.
8. Hu Z, et al. New loci associated with chronic hepatitis B virus infection in Han Chinese. *Nat Genet*. 2013;45(12):1499.
9. Li Y, et al. Genome-wide association study identifies 8p21. 3 associated with persistent hepatitis B virus infection among Chinese. *Nature communications*. 2016;7(1):1–11.
10. Frodsham AJ. Host genetics and the outcome of hepatitis B viral infection. *Transplant immunology*. 2005;14(3–4):183–6.
11. He Y-L, et al. Host susceptibility to persistent hepatitis B virus infection. *World journal of gastroenterology: WJG*. 2006;12(30):4788.
12. Jiang D-K, et al. Genetic variations in STAT4, C2, HLA-DRB1 and HLA-DQ associated with risk of hepatitis B virus-related liver cirrhosis. *Scientific reports*. 2015;5:16278.
13. Chen VL, et al. Genetic variants that associate with cirrhosis have pleiotropic effects on human traits. *Liver International*. 2020;40(2):405–15.
14. Zhang H, 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(9):755–8.
15. Li S, et al., *GWAS identifies novel susceptibility loci on 6p21. 32 and 21q21. 3 for hepatocellular carcinoma in chronic hepatitis B virus carriers*. *PLoS genetics*, 2012. 8(7).
16. Jiang D-K, et al. Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus–related hepatocellular carcinoma. *Nat Genet*. 2013;45(1):72.
17. Höhler T, et al. Differential genetic determination of immune responsiveness to hepatitis B surface antigen and to hepatitis A virus: a vaccination study in twins. *The Lancet*. 2002;360(9338):991–5.
18. Ji X, et al. Impacts of human leukocyte antigen DQ genetic polymorphisms and their interactions with hepatitis B virus mutations on the risks of viral persistence, liver cirrhosis, and hepatocellular carcinoma. *Infection Genetics Evolution*. 2014;28:201–9.
19. Zeng Z, et al. A population-based study to investigate host genetic factors associated with hepatitis B infection and pathogenesis in the Chinese population. *BMC Infect Dis*. 2008;8(1):1.
20. Yang J, et al. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*. 2011;88(1):76–82.
21. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*. 2007;81(3):559–75.
22. Li M-X, et al. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human genetics*. 2012;131(5):747–56.
23. Consortium GP. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.
24. De Maio N, Schrempf D, Kosiol C. PoMo: an allele frequency-based approach for species tree estimation. *Syst Biol*. 2015;64(6):1018–31.
25. Weir BS, Cockerham CC. *Estimating F-statistics for the analysis of population structure*. *evolution*, 1984: p. 1358–1370.
26. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting F ST. *Nat Rev Genet*. 2009;10(9):639–50.

27. Sabeti PC, et al. Positive natural selection in the human lineage. *science*. 2006;312(5780):1614–20.
28. Voight BF, et al. A map of recent positive selection in the human genome. *PLoS Biol*. 2006;4(3):e72.
29. Kijas JW, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol*. 2012;10(2):e1001258.
30. Wang K, et al. Detection of selection signatures in Chinese Landrace and Yorkshire pigs based on genotyping-by-sequencing data. *Frontiers in genetics*. 2018;9:119.
31. Zhang Z, et al. Whole-genome resequencing reveals signatures of selection and timing of duck domestication. *GigaScience*. 2018;7(4):gij027.
32. Danecek P, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
33. Frankham R, Briscoe DA, Ballou JD, *Introduction to conservation genetics*. 2002: Cambridge university press.
34. Sabeti PC, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 2002;419(6909):832–7.
35. Gautier M, Klassmann A, Vitalis R. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular ecology resources*. 2017;17(1):78–90.
36. Speidel L, et al. A method for genome-wide genealogy estimation for thousands of samples. *Nat Genet*. 2019;51(9):1321–9.
37. Chambers JC, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet*. 2011;43(11):1131–8.
38. Masuda M, et al. Interaction of genetic markers associated with serum alkaline phosphatase levels in the Japanese population. *Human genome variation*. 2015;2(1):1–6.
39. Satishchandran A, et al., MicroRNA 122, regulated by GRLH2, protects livers of mice and patients from ethanol-induced liver disease. *Gastroenterology*, 2018. 154(1): 238–52. e7.
40. An P, Zeng Z, Winkler CA. The Loss-of-Function S267F variant in HBV receptor NTCP reduces human risk for HBV infection and disease progression. *The Journal of infectious diseases*. 2018;218(9):1404–10.
41. An P, et al. A common HLA–DPA1 variant is a major determinant of hepatitis B virus clearance in Han Chinese. *The Journal of infectious diseases*. 2011;203(7):943–7.
42. Kim S-J, et al., *Hepatitis B virus disrupts mitochondrial dynamics: induces fission and mitophagy to attenuate apoptosis*. *PLoS pathogens*, 2013. 9(12).
43. Cesnekova J, et al. The mammalian homologue of yeast Afg1 ATPase (lactation elevated 1) mediates degradation of nuclear-encoded complex IV subunits. *Biochem J*. 2016;473(6):797–804.
44. Frezza C, et al. OPA1 controls apoptotic cristae remodeling independently from mitochondrial fusion. *Cell*. 2006;126(1):177–89.
45. Anand R, et al. The i-AAA protease YME1L and OMA1 cleave OPA1 to balance mitochondrial fusion and fission. *J Cell Biol*. 2014;204(6):919–29.
46. Kumar PA, SuBramanian K. The role of ischemia modified albumin as a biomarker in patients with chronic liver disease. *Journal of clinical diagnostic research: JCDR*. 2016;10(3):BC09.
47. Yuen M-F, et al. Prognostic determinants for chronic hepatitis B in Asians: therapeutic implications. *Gut*. 2005;54(11):1610–4.
48. Walayat S, et al. Role of albumin in cirrhosis: from a hospitalist's perspective. *Journal of community hospital internal medicine perspectives*. 2017;7(1):8–14.
49. Castell JV, et al. Interleukin-6 is the major regulator of acute phase protein synthesis in adult human hepatocytes. *FEBS Lett*. 1989;242(2):237–9.
50. Carvalho JR, Machado MV. New insights about albumin and liver disease. *Annals of hepatology*. 2018;17(4):547–60.
51. Cichoż-Lach H, Michalak A. Oxidative stress as a crucial factor in liver diseases. *World journal of gastroenterology: WJG*. 2014;20(25):8082.

52. Teshale EH, et al., *Genotypic distribution of hepatitis B virus (HBV) among acute cases of HBV infection, selected United States counties, 1999–2005*. *Clinical infectious diseases*, 2011. 53(8): p. 751–756.
53. Kim H, et al. Racial/ethnic disparities in the prevalence and awareness of hepatitis B virus infection and immunity in the United States. *Journal of viral hepatitis*. 2017;24(11):1052–66.
54. Tan AT, et al. Host ethnicity and virus genotype shape the hepatitis B virus-specific T-cell repertoire. *Journal of virology*. 2008;82(22):10986–97.
55. Guo X, et al., *Strong influence of HLA-DP gene variants on development of persistent chronic HBV carriers in the Han Chinese population*. *Hepatology (Baltimore, Md.)*, 2011. 53(2): p. 422.
56. Nishida N, et al. Genome-wide association study confirming association of HLA-DP with protection against chronic hepatitis B and viral clearance in Japanese and Korean. *PLoS one*. 2012;7(6):e39175.
57. Wong DK-H, et al., *Role of HLA-DP polymorphisms on chronicity and disease activity of hepatitis B infection in Southern Chinese*. *PLoS One*, 2013. 8(6).

Figures

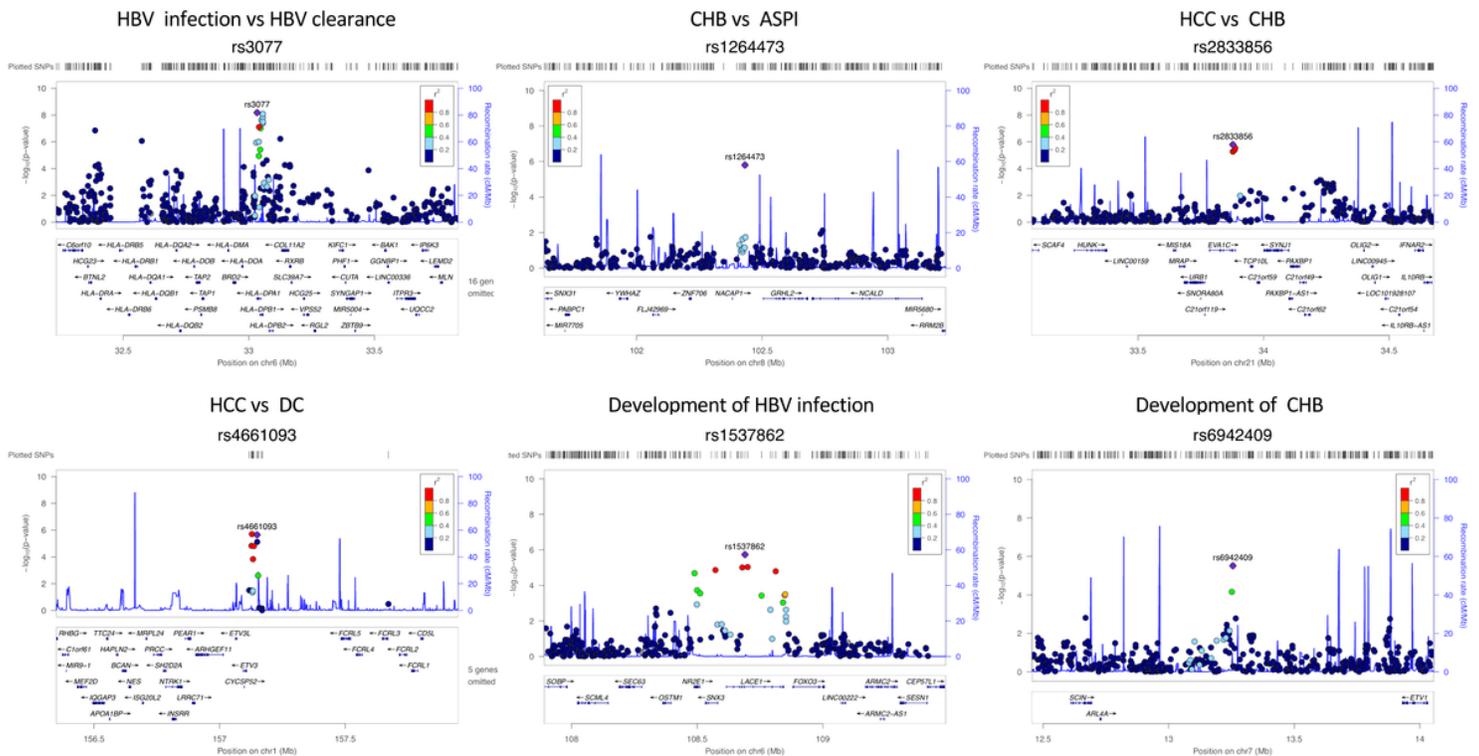


Figure 1

Regional plots shown $-\log_{10}$ P-values of SNPs in association study. Marker SNPs are shown as purple diamonds, other SNPs are shown as dots. R-square of Marker SNPs and other SNPs are shown against dark blue, blue, green, yellow and red colors,

indicating the linkage disequilibrium. The structure of genes within the region are shown as rectangles and arrows. Abbreviation: ASPI, asymptomatic persistence infection; CHB, chronic hepatitis B; DC, decompensated cirrhosis; HCC, hepatocellular carcinoma.

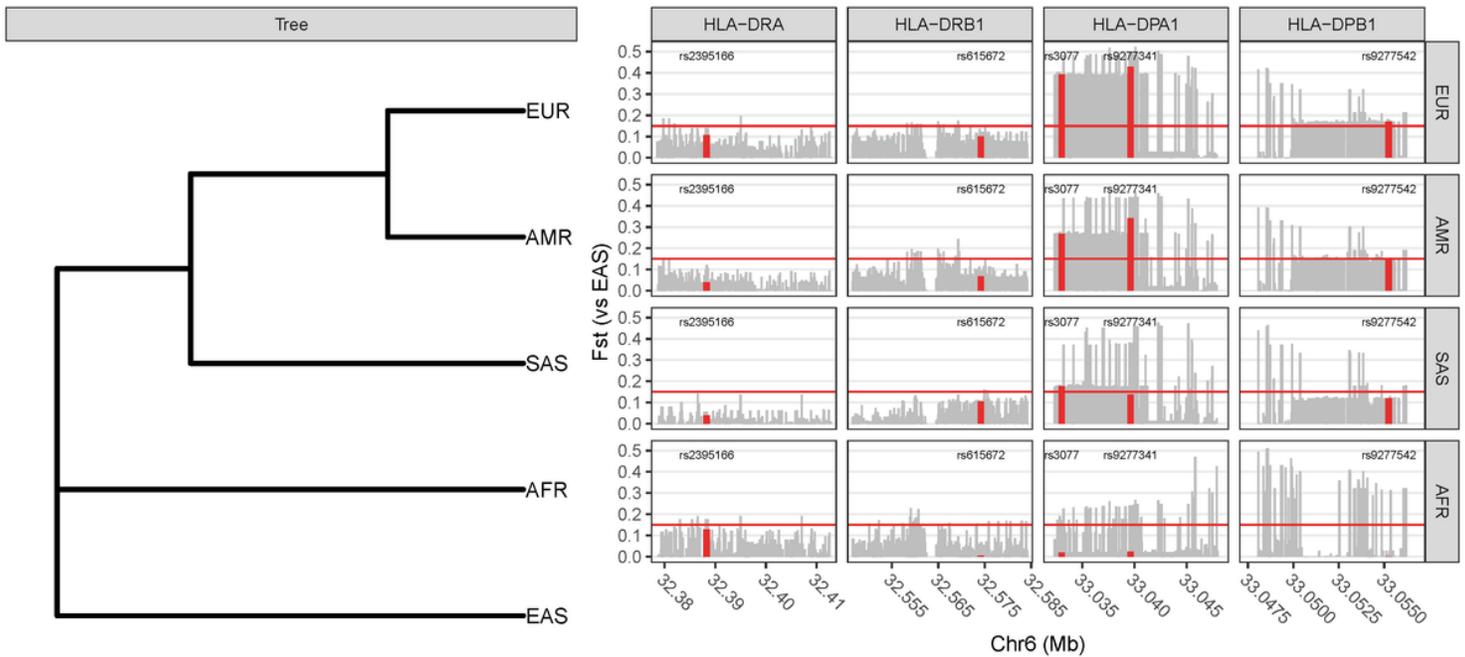


Figure 2

The racial tree (Left) were based on the SNPs in HBV-infection related genes, including HLA-DRA, HLA-DRB1, HLA-DPA1 and HLA-DPB1. The genotype and minor allele frequency of each SNPs were accessed from 1000 Genome Project. EAS, AFR, SAS, EUR, AMR refer to East Asian, African, South Asian, European and American of 1000 Genome Project, respectively. The racial tree indicated a genetic difference in HBV-infection related genes among EAS, AFR, SAS, EUR, AMR. The genetic difference (Right) of each SNPs was evaluated by FST value. X-axis refer to physical position in chromosome 6. Y-axis refer to FST value of paired SNP. FST values of all paired SNPs of AFR, SAS, AMR, EUR versus EAS were displayed in grey bar. FST values accessed 0.15 (Red horizontal line) indicated the signal of selective even. Red bars and rs IDs showed the reported HBV infection-related SNPs. The FST values of European versus East Asian showed the genetic difference in HLA-DPA1 and HLA-DPB1, indicating a genetic selection against the HBV infection. The racial tree showed that the East Asian population is at the root, indicating that why we used East Asian population as a comparative population but not the other population, and compared other four populations with East Asian population.

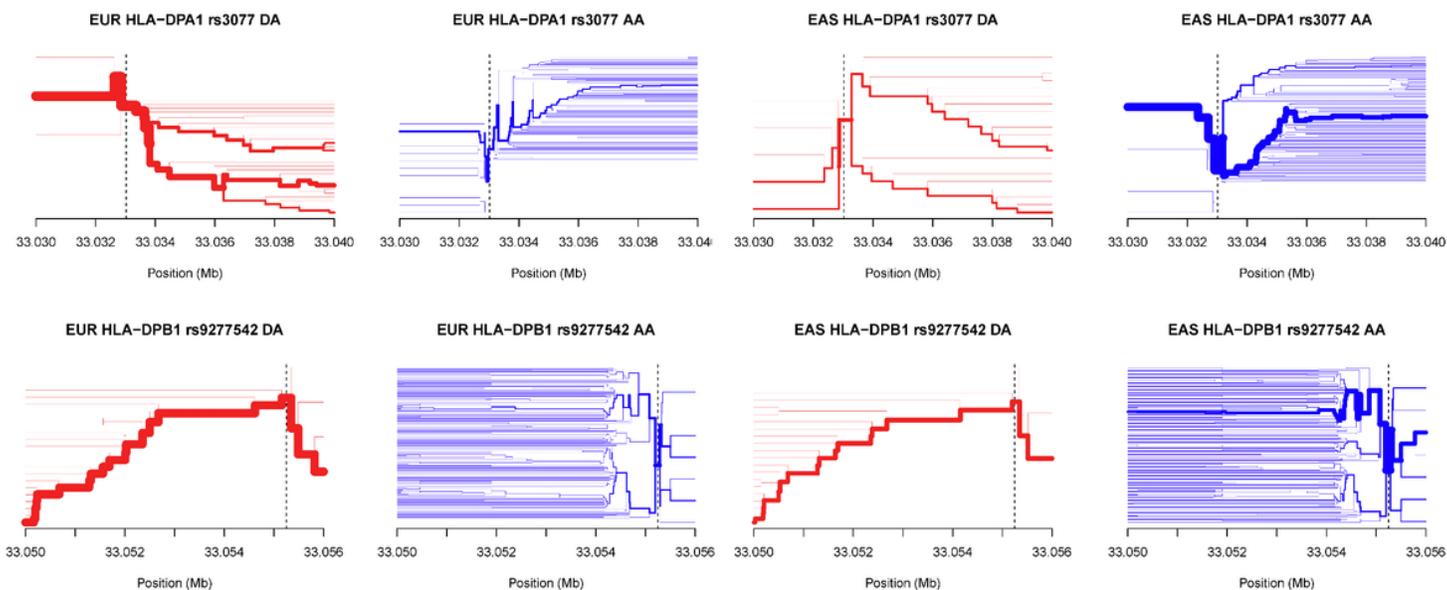


Figure 3

Haplotype bifurcation diagrams of infection-related SNPs, including rs3077 on HLA-DPA1 (upper) and rs9277542 on HLA-DPB1 (lower), in European (left) and East Asian (right). EUR and EAS in plot title refer to European and East Asian. DA and AA in plot title refer to derived allele (red) and ancestral allele (blue). Black dash line refers to the position of core SNP. Each node refers to a haplotype. The edge width reflects the population-specific frequency. Haplotype bifurcation diagrams were displayed via Rehh package. Haplotype bifurcation diagrams showed that the derived allele lead to a long-range high frequency haplotype in European population and the ancestral allele lead to a high frequency haplotype in East Asian population; the ancestral allele lead to more haplotypes than the derived allele. The long-range high frequency haplotype confirms the genetic selection in HLA-DPA1 and HLA-DPB1 in European population.

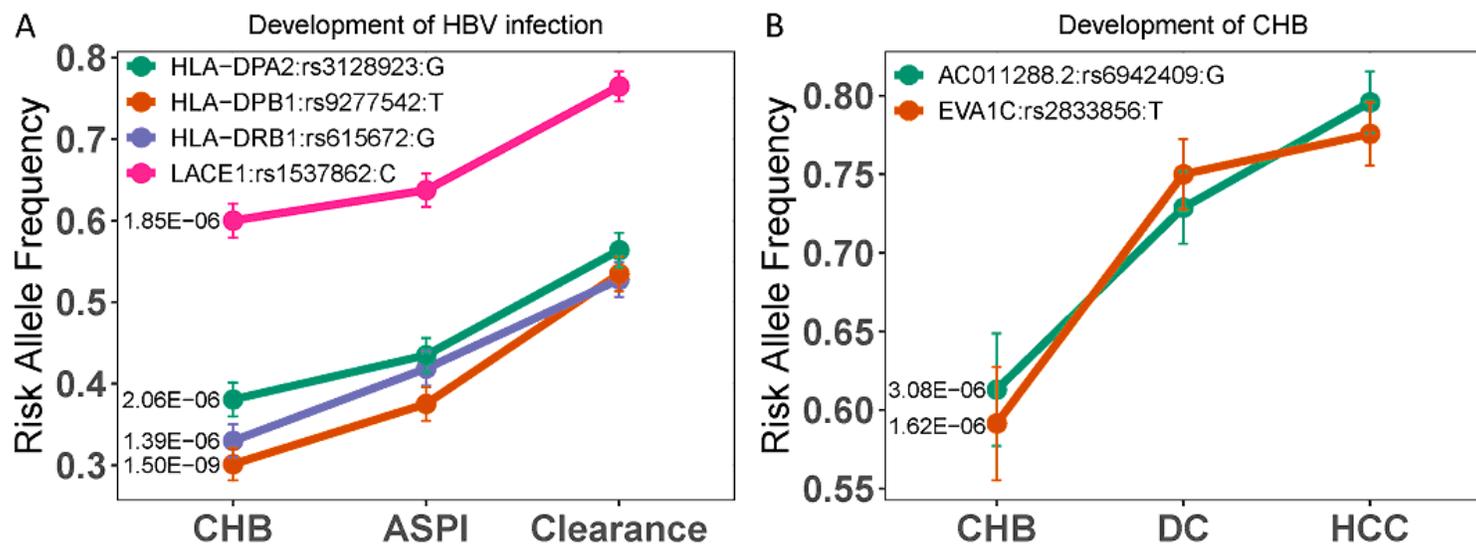


Figure 4

The raising allele frequency in HBV related outcomes during the progression. Four SNPs with increased resistance in CHB, ASPI, HBV clearance during HBV infection (A) and two SNPs with increased risk in the CHB, DC, HCC during the development of CHB (B). Abbreviation: ASPI, asymptomatic persistence infection; CHB, chronic hepatitis B; DC, decompensated cirrhosis; HCC, hepatocellular carcinoma.

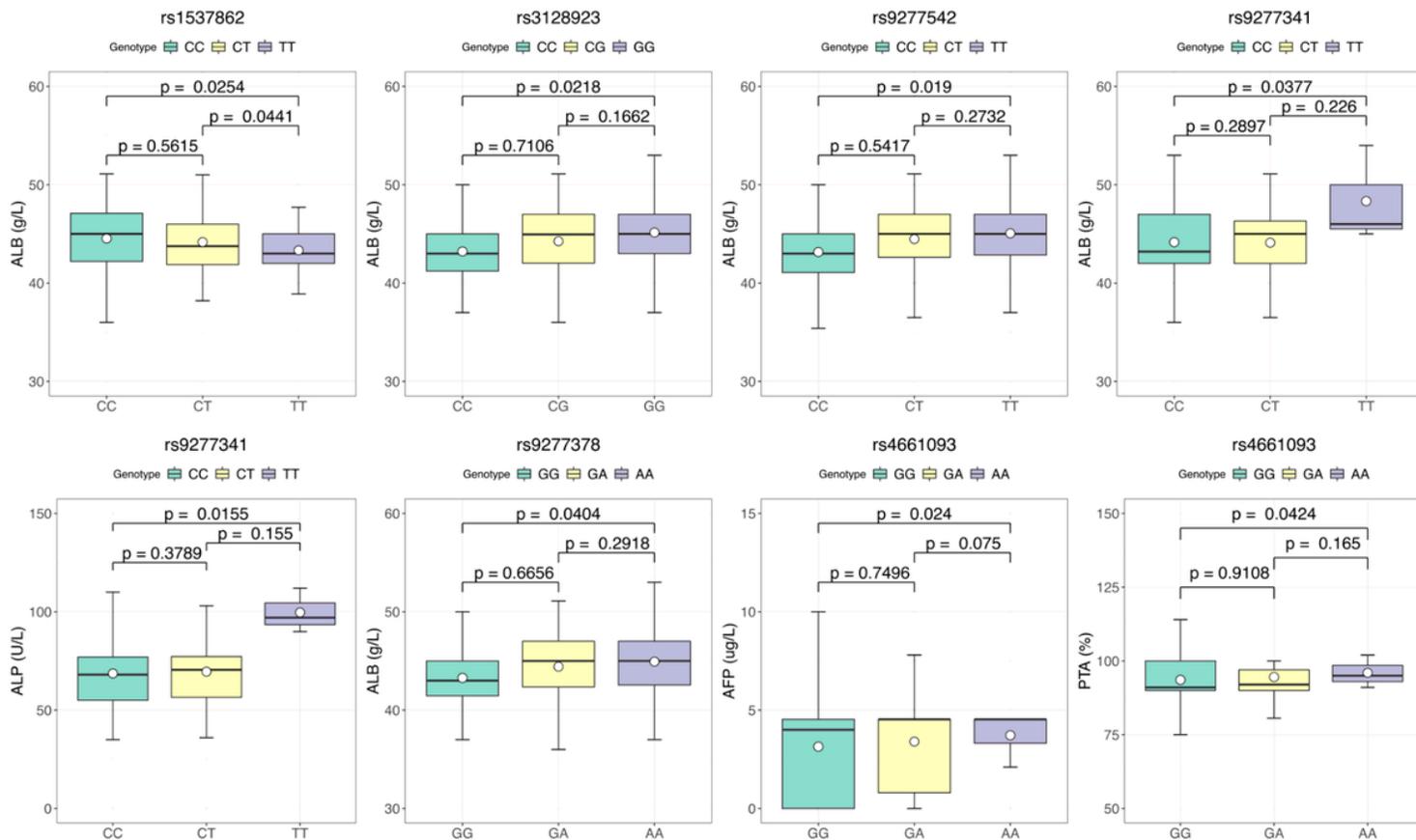


Figure 5

The association between HBV related locus and serum liver enzyme levels in health controls. P values were calculated by ANOVA test. White-circle refer to the mean liver enzymes level with different genotypes. The significant differences indicate that these SNPs contribute to liver enzyme activity.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigures.docx](#)