Large-scale genome-wide association study identifies HLA class II variants associated with chronic HBV infection: a study from Taiwan Biobank

Yu-Han Huang¹ | Shu-Fen Liao¹ | Seik-Soon Khor² | Yu-Ju Lin¹ | Hsuan-Yu Chen¹ | Ya-Hsuan Chang¹ | Yi-Hsiang Huang¹ | Sheng-Nan Lu³,4 | Hye-Won Lee⁵ | Wen-Ya Ko¹ | Claire Huang¹ | Po-Chun Liu¹ | Yen-Ju Chen¹ | Ping-Feng Wu¹ | Hou-Wei Chu¹ | Pei-Ei Wu¹ | Katsushi Tokunaga² | Chen-Yang Shen¹ | Mei-Hsuan Lee¹

¹Taipei, Taiwan | ²Tokyo, Japan | ³Kaohsiung, Taiwan | ⁴Chiayi, Taiwan | ⁵Seoul, Korea

Correspondence
Mei-Hsuan Lee, Institute of Clinical Medicine, National Yang-Ming University, 155 Li-Nong Street, Section 2, Peitou, Taipei 112, Taiwan. Email: meihlee@ntu.edu.tw; meihlee@ym.edu.tw

Funding information
This study was supported by research grants from the Ministry of Science and Technology, Taipei, Taiwan (grant number: 105-2628-B-010-003-MY4, 107-2314-B-010-004-MY2). None of the funding organisation played a role in the study design and conduct; data collection, management, analysis, and interpretation; data preparation and review; or manuscript approval.

Summary
Background: Chronic hepatitis B virus (HBV) infection is a great health burden with geographical variations.

Aims: To explore genetic variants associated with chronic HBV infection.

Methods: The study included 15 352 participants seropositive for HBV core antibodies in Taiwan Biobank. Among them, 2591 (16.9%) seropositive for HBV surface antigen (HBsAg) were defined as chronic HBV infection. All participants were examined for whole-genome genotyping by Axiom-Taiwan Biobank Array. The human leucocyte antigen (HLA) imputation was performed after identification of the variants within the region. Logistic regressions were used to estimate odds ratios (ORs) with 95% confidence intervals. Correlations of different HLA allele frequencies with HBsAg seroprevalence were evaluated across worldwide populations by Pearson correlation coefficients. Epitope prediction was performed for HLA alleles using NetMHCIIpan method.

Results: Located within a cluster of 450 single nucleotide polymorphisms in HLA class II, rs7770370 ($P = 2.73 \times 10^{-35}$) was significantly associated with HBV chronicity ($P_{corrected} < 8.6 \times 10^{-5}$). Imputation analyses showed that HLA-DPA1*02:02 and HLA-DPB1*05:01 were associated with chronic HBV, with adjusted ORs of 1.43 (1.09-1.89) and 1.61 (1.29-2.01). These allele frequencies were positively correlated with global HBsAg seroprevalence, with R of 0.75 and 0.62 respectively ($P < 0.05$). HLA-DRB1*13:02, HLA-DQA1*01:02 and HLA-DQB1*06:09 associated with HBV chronicity negatively, with adjusted ORs of 0.31 (0.17-0.58), 0.70 (0.56-0.87) and 0.33 (0.18-0.63). These HLA alleles had various binding affinities to the predicted epitopes derived from HBV nucleocapsid protein.

Conclusions: HLA class II variants are relevant for chronicity after HBV acquisition.
1 | INTRODUCTION

Hepatitis B virus (HBV) infection is one of the major risk factors for advanced liver diseases. Individuals with HBV infection are at risk of hepatic decompensation, cirrhosis, or liver cancer, and approximately 800,000 individuals lose their lives each year due to HBV-related complications. Global Burden of Disease estimates show high morbidity and mortality due to chronic HBV infection, despite decreases in chronic infection rate over the past few decades. Although highly effective hepatitis B vaccines have been available, curative anti-virals are still unavailable. Understanding how to control chronic HBV infection is urgently needed.

Antibody to hepatitis B core antigen (anti-HBc) is one of the critical seromarkers that indicates either previous or ongoing HBV infection. Chronic vs acute infection is defined by the presence of hepatitis B surface antigen (HBsAg) for at least 6 vs less than 6 months. Across the world, approximately 257 million people are chronically infected with HBV, 75% of whom reside in the Asian Pacific region. There are large geographical variations in HBsAg seroprevalence across countries. Southeast Asia, sub-Saharan Africa, China, Indonesia and Nigeria are considered highly endemic regions as the prevalence of chronic HBV infection in the population can be greater than 8%, while South and Central America, Eastern and Southern Europe, and Southwest Asia are considered intermittently endemic areas, with prevalence rates between 2% and 7%. Developed countries in Western Europe and North America are classified as low endemic areas, since chronic HBV infection rates range from 0.5% to 2.0%. Persistent HBV infection may also be attributed to the heterogeneity of genetic backgrounds.

Recent genome-wide association studies (GWAS) have identified genetic variants that are associated with persistent HBV infection. However, the previous history of HBV infection among study participants in these studies was unknown, thus, the control group may have included individuals who had never acquired HBV infection. Although three GWAS did utilise both anti-HBc and HBsAg to define chronic HBV infection and spontaneous HBV clearance, the study participants with chronic HBV infection were heterogeneous and consisted of a mixed group that included both healthy HBV carriers and patients with end-stage liver diseases. Therefore, community-based studies conducted in a homogenous population that explore genetic variants associated with persistent HBV infection are still limited. In addition, whether these identified variants were attributable to the geographical variations of HBsAg prevalence globally remained unknown.

This large-scale GWAS carried out in a HBV highly endemic area aimed to explore genetic variants associated with chronic HBV infection. The genetic variants associated with chronic HBV infection were additionally evaluated for their correlations with different prevalence of chronic HBV infection worldwide.

2 | MATERIALS AND METHODS

2.1 | Study population

The individuals were enrolled from the Taiwan Biobank, a large-scale community-based study that was established to enable the investigation of associations between genes, environment and human health. The Taiwan Biobank recruited adults aged 30-70 years from the general population. At study entry, all participants were free of cancers and provided written informed consent for interviews, a physical examination and biospecimen collection. The Taiwan Biobank collected information on sociodemographic characteristics, lifestyle and dietary habits, environmental exposures, and personal or family histories of major diseases using a structured questionnaire. Blood samples were collected for whole-genome single nucleotide polymorphism (SNP) genotyping as well as biochemical and virological testing. This study protocol was approved by the institutional review board of National Yang-Ming University, Taipei, Taiwan.

2.2 | Chronic hepatitis B infection

All study participants were tested for their hepatitis B core antibodies (anti-HBc), hepatitis B surface antigen (HBsAg) and antibodies against hepatitis C virus (anti-HCV). As the Taiwanese government began the implementation of a nationwide vaccination program for newborns in 1984, all participants who had been previously vaccinated were excluded from the study. In addition, those who were seropositive for anti-HCV were also excluded from this study. A total of 15,352 participants who were anti-HBc seropositive were defined as having previously acquired HBV infection. Among those who were seropositive for anti-HBc, individuals seropositive for HBsAg were defined as chronically infected with hepatitis B, while individuals seronegative for HBsAg were defined as having resolved past HBV infection. In our study population, a total of 2,591 (16.9%) individuals had chronic hepatitis B infection, and 12,761 (83.1%) had resolved HBV infection.

2.3 | GWAS power calculation

Statistical power for genetic associations was evaluated using the GAS Power Calculator (http://csg.sph.umich.edu/abecasis/CaTS/gas_power_calculator/index.html), with the assumption of a chronic HBV prevalence of 16.9%, and a significance threshold of $7.7 \times 10^{-8}$. A case-control study with a total of 12,000 study participants would be able to detect genetic variants with a risk allele frequency ≥0.05 and a relative risk ≥1.5 with sufficient statistical power (>80%). In this study, we randomly allocated study participants (N = 15,352) at a three-to-one ratio, resulting in 11,514 individuals for the discovery phase and 3,838 individuals for the replication phase.
2.4 | Genome-wide SNP genotyping and quality control

All participants had their genomic DNA samples whole-genome SNP genotyped using the customised Axiom-Taiwan Biobank Array (TWB chip: Affymetrix, Inc). The customised SNP array included 653 291 SNPs that were specifically selected for the Taiwanese population, and genotyping was performed using standardised procedures at the National Center of Genome Medicine, Academia Sinica, Taipei, Taiwan. The study flowchart is shown in Figure 1. During the GWAS phase, study samples and SNPs were excluded using the quality control procedures as follows: a sample call rate <97%, a SNP call rate <97%, a minor allele frequency <0.01, and if in violation of Hardy-Weinberg equilibrium ($P < 10^{-4}$). Individuals with outlying autosomal heterozygosity rates (beyond a range of mean ± 3 standard deviations) and highly correlated individuals with identity-by-descent >0.1875 were also excluded from subsequent analyses. In the end, 583 383 autosomal SNPs were included in the GWAS phase, and we subsequently examined their associations with chronic HBV infection.

2.5 | HLA imputation

HLA imputation was performed during the replication phase. Imputation is a valuable tool used to investigate the intricate major histocompatibility complex (MHC) region. SNP genotype information within an extended MHC region between 25 759 242 and 33 534 872 base pairs based on hg19 positions were used with a Taiwanese reference for imputation. Two-field HLA imputation was performed on HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1 and HLA-DPB1 using the HIBAG R package, which infers HLA genotypes by averaging posterior probabilities from all classifiers that are built on bootstrapping samples and SNPs subsets. We applied a post-imputation quality control of a call threshold = 0.5 to remove poorly imputed HLA alleles. In the end, 549 chronic HBV infected
cases and 2688 HBV resolved controls with a call threshold ≥0.5 at all of the HLA loci were included in subsequent analyses. HLA haplotypes were examined using Arlequin software version 3.5.

2.6 | Global HLA allele frequencies and HBsAg seroprevalence

In order to evaluate the associations of HLA alleles with chronic hepatitis B infection at the global level, we collected information on HLA allele frequencies and HBsAg seroprevalence. HLA allele frequencies were obtained from the allele frequency net database (AFND, http://www.allelefrequencies.net), an online repository that includes worldwide population frequency information for some specific HLA alleles.15 The database contains various allele frequencies at a given HLA allele that were uploaded by different study groups in the same country. Therefore, to create a standardised frequency for each country, we weighted each specific allele frequency by the reported sample size, then divided this product by the total reported sample size for each country. This quotient was the average HLA frequency of a given country. We obtained the HBsAg seroprevalence of each country from a comprehensive review article on the global epidemiology of HBV infection.6 We used the Pearson correlation coefficient to examine the correlation between a given HLA allele frequency and HBsAg seroprevalence across countries.

To determine the possible prevalence of chronic HBV infection attributable to a specific HLA allele, we calculated population attributable fractions using a fundamental formula \( \frac{P_e(RR – 1) + 1}{P_e(RR – 1) + 1} \) with \( P_e \) being the prevalence of a given HLA risk allele among controls and the adjusted ORs to approximate the relative risk (RR) in the general population. Preventable fractions for protective HLA alleles were obtained using reverse coding.

2.7 | Prediction of T cell epitopes for the HBV nucleocapsid protein

Epitope prediction for the HLA alleles found to be associated with chronic HBV infection was performed using the NetMHCIIpan method in the immune epitope database (IEDB). The HBV nucleocapsid antigen sequence was used for prediction with half maximal inhibitory concentrations (IC50) generated for core peptides. A low IC50 indicated a higher affinity between the HBV antigen and HLA alleles.

2.8 | Statistical methods

Prior to exploring SNP associations with chronic HBV infection, we first evaluated whether there was any existing population stratification. In the quantile-quantile plot of observed vs expected \( P \) values, the genomic inflation factor was found to be 1.039. In addition, using the first two components, principle component analysis showed that the cases and controls were from the same cluster, suggesting that any population stratification was limited and negligible. We examined the 583 383 SNPs and their associations with chronic HBV infection using an additive model, and set a Bonferroni corrected \( P < 8.6 \times 10^{-8} \) as the significance level. We generated a Manhattan plot of \(-\log_{10} P \) values for all SNPs. In addition, we also produced a regional plot of significant SNPs using LocusZoom. (http://csg.sph.umich.edu/locuszoom/). The genome-based restricted maximum likelihood algorithm in the GCTA program was utilised for heritability estimation.

After HLA imputation, we used Fisher’s exact tests to compare differences in allele frequencies at a given HLA gene between those with chronic HBV infection and those with resolved HBV infection. Alleles or haplotypes with frequencies <1% were combined as rare alleles or haplotypes. Logistic regression models were used to obtain odds ratios (ORs) with 95% confidence intervals for the associations between SNPs, HLA alleles or haplotypes with chronic HBV infection. We also adjusted for other confounding factors including age, sex, serum alanine aminotransferase (ALT) level, body mass index, cigarette smoking and alcohol consumption in the multiple logistic regression models. Statistical analyses were performed with PLINK, R packages, and SAS version 9.4 (SAS Institute, Inc, Cary, NC).

3 | RESULTS

The baseline characteristics of study participants in the GWAS and replication phases are shown in Table 1. There were a total of 1943 chronic HBV infected cases and 9571 resolved controls in the GWAS phase, and 648 chronic HBV infected cases and 3190 resolved controls in the replication phase. There were no significant differences in age, sex, HBsAg status, cigarette smoking and alcohol consumption between individuals in the two phases. However, participants in the replication phase had significantly higher proportions of high body mass index, and elevated serum ALT levels (≥45 U/L). Participants with chronic HBV infection tended to be younger in age, male, and have elevated serum ALT levels compared to those with resolved HBV infection (\( P < 0.05 \)).

3.1 | SNPs on HLA genes associated with chronic HBV infection

After quality control measures, we examined a total of 583 383 SNPs among 1810 chronic HBV infected cases and 9002 HBV resolved controls in the GWAS phase. The heritability was 0.62 for chronic HBV infection. A Manhattan plot showed a large cluster of SNPs that were associated with HBV chronic infection (Figure S1), with 450 SNPs that reached significance (\( P < 8.6 \times 10^{-8} \)). These SNPs were located on chromosome 6. Extensive linkage disequilibrium patterns were found among the 450 SNPs, corresponding with three significant peaks around the HLA-DPA1/
### TABLE 1 Baseline characteristics of study participants among the two phases

<table>
<thead>
<tr>
<th>Baseline characteristics</th>
<th>GWAS phase (n = 11,514)</th>
<th></th>
<th>Replication phase (n = 3,838)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Chronic infected cases (n = 1,943)</td>
<td>Resolved controls (n = 9,571)</td>
<td>P</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30-39</td>
<td>2,049 (17.8%)</td>
<td>541 (27.8%)</td>
<td>1,508 (15.8%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>40-49</td>
<td>3,174 (27.6%)</td>
<td>623 (32.1%)</td>
<td>2,551 (26.7%)</td>
<td></td>
</tr>
<tr>
<td>50-59</td>
<td>3,510 (30.5%)</td>
<td>483 (24.9%)</td>
<td>3,027 (31.6%)</td>
<td></td>
</tr>
<tr>
<td>60+</td>
<td>2,781 (24.2%)</td>
<td>296 (15.2%)</td>
<td>2,485 (26.0%)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>5,903 (51.3%)</td>
<td>1,094 (56.3%)</td>
<td>4,809 (50.2%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Female</td>
<td>5,611 (48.7%)</td>
<td>849 (43.7%)</td>
<td>4,762 (49.8%)</td>
<td></td>
</tr>
<tr>
<td>HBsAg status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seronegative</td>
<td>9,571 (83.1%)</td>
<td>0 (0.0%)</td>
<td>9,571 (100.0%)</td>
<td></td>
</tr>
<tr>
<td>Seropositive</td>
<td>1,943 (16.9%)</td>
<td>1943 (100.0%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Body mass index</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18.5</td>
<td>257 (2.2%)</td>
<td>54 (2.8%)</td>
<td>203 (2.1%)</td>
<td>0.1387</td>
</tr>
<tr>
<td>18.5-24.9</td>
<td>6,774 (58.8%)</td>
<td>1,166 (60.0%)</td>
<td>5,608 (58.6%)</td>
<td></td>
</tr>
<tr>
<td>25.0-29.9</td>
<td>3,748 (32.6%)</td>
<td>606 (31.2%)</td>
<td>3,142 (32.8%)</td>
<td></td>
</tr>
<tr>
<td>&gt;30</td>
<td>735 (6.4%)</td>
<td>117 (6.0%)</td>
<td>618 (6.5%)</td>
<td></td>
</tr>
<tr>
<td>Serum ALT level, U/L</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;45</td>
<td>10,589 (92.0%)</td>
<td>1,712 (88.1%)</td>
<td>8,877 (92.7%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>≥45</td>
<td>925 (8.0%)</td>
<td>231 (11.9%)</td>
<td>694 (7.3%)</td>
<td></td>
</tr>
<tr>
<td>Cigarette smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>9,949 (86.4%)</td>
<td>1,670 (85.9%)</td>
<td>8,279 (86.5%)</td>
<td>0.5179</td>
</tr>
<tr>
<td>Yes</td>
<td>1,565 (13.6%)</td>
<td>273 (14.1%)</td>
<td>1,292 (13.5%)</td>
<td></td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>10,634 (92.4%)</td>
<td>1,793 (92.3%)</td>
<td>8,841 (92.4%)</td>
<td>0.8883</td>
</tr>
<tr>
<td>Yes</td>
<td>880 (7.6%)</td>
<td>150 (7.7%)</td>
<td>730 (7.6%)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Abbreviations: ALT, alanine aminotransferase; GWAS, genome-wide association study; HBsAg, hepatitis B surface antigen.

Note: Baseline characteristics of study participants were compared between chronic HBV infected cases and resolved controls within GWAS and replication phase using the chi-squared test. Comparison was also performed between two phases.
DPB1, HLA-DQB1/DRB1 and BTNL2 genes (Figure 2). The most significantly associated SNP was rs7770370, with an OR of 0.61 (0.56-0.66; \(P = 2.73 \times 10^{-35}\)). During the replication phase, the 450 significant SNPs were further validated among the 648 chronic HBV infected cases and 3190 HBV resolved controls. After correction for multiple testing, 242 of 450 SNPs remained significantly associated with chronic HBV infection \((P < 0.0001)\). The frequencies and ORs of the 450 SNPs examined in the two phases are displayed in Table S1.

### 3.2 | HLA class II genes and the associated risk of HBV chronicity

We identified 17 HLA-DRB1 alleles, 12 HLA-DQA1 alleles, 12 HLA-DQB1 alleles, 5 HLA-DPA1 alleles and 13 HLA-DPB1 alleles with allele frequencies >1%. There were 549 chronic HBV infected cases and 2688 resolved controls that had adequate imputation quality (call threshold >0.5) at all of the five HLA class II genes. The alleles HLA-DRB1*13:02, HLA-DQA1*01:02, HLA-DQB1*06:09, HLA-DPA1*02:02...
and HLA-DPB1*05:01 showed significantly different frequencies between chronic HBV infected cases and resolved controls (P < 0.0008) (Tables S2-S6). Compared to resolved controls, a high proportion of chronic HBV infected cases carried HLA-DPA1*02:02 and HLA-DPB1*05:01. After adjustment for potential confounders, HLA-DPA1*02:02 and HLA-DPB1*05:01 still showed significant positive associations with chronic HBV infection, with adjusted ORs of 1.43 (1.09-1.89; P = 0.0105) and 1.61 (1.29-2.01; P < 0.0001) respectively (Table 2). On the other hand, three HLA alleles showed negative associations with chronic HBV infection; the adjusted ORs were 0.31 (0.17-0.58; P = 0.0003) for HLA-DRB1*13:02, 0.70 (0.56-0.87; P = 0.0013) for HLA-DQA1*01:02, and 0.33 (0.18-0.63; P = 0.0006) for HLA-DQB1*06:09 (Table 2).

Because of the strong linkage disequilibrium patterns within the HLA-DPA1/HLA-DPB1 and HLA-DRB1/HLA-DQB1 gene loci, we also performed haplotype analyses. Individuals who carried the HLA-DPA1*02:02-DPB1*05:01 haplotype were found to have a higher likelihood of having chronic HBV infection, with an adjusted OR of 1.59 (1.28-1.97; P < 0.0001). On the other hand, HLA-DRB1*13:02-DQB1*06:09 showed a negative association, with an adjusted OR of 0.33 (0.18-0.63; P = 0.0006) (Table 2, Tables S7 and S8).

### 3.3 HLA allele frequencies correlated with global HBsAg seroprevalence

Figure S2 shows that both allele frequencies for HLA-DPA1*02:02 and HLA-DPB1*05:01 were positively correlated with HBsAg seroprevalence, with Pearson's R coefficients of 0.75 (P = 0.00031) and 0.62 (P = 0.032) respectively. After excluding the data in Taiwan, where HBV is prevalent, the HLA-DPB1*05:01 still showed a positive correlation with global HBsAg seroprevalence (R = 0.69, P = 0.0021). The population attributable fractions for HLA-DPA1*02:02 and HLA-DPB1*05:01 in chronic HBV infection were 26.3% and 29.5% respectively. On the other hand, the prevalent fractions were 67.7% for HLA-DRB1*13:02, 22.7% for HLA-DQA1*01:02 and 65.7% for HLA-DQB1*06:09. There were no significant correlations between allele frequencies of HLA-DRB1*13:02, HLA-DQA1*01:02 and HLA-DQB1*06:09 and HBsAg seroprevalence worldwide (R = -0.077, 0.22, and 0.38).

### 3.4 T cell epitope prediction for the HBV antigen

We examined potential peptide binding repertoires to investigate immune responses driven by the HLA class II alleles that we found to be associated with chronic HBV infection. A set of 514 predicted epitopes were generated for HLA-DRB1*13:02, HLA-DQA1*01:02/HLA-DQB1*06:09 and HLA-DPA1*02:02/HLA-DPB1*05:01. Table S9 lists the top 10th percentile (IC50 ≤ 500 nM) of predicted epitopes that were recommended by the immune epitope database for selecting binders with sufficient affinity. Among those predicted epitopes, the protective HLA-DRB1*13:02 allele had the most potential epitopes with better binding affinities (IC50 < 100 nM). The predicted epitopes for HLA-DQA1*01:02/HLA-DQB1*06:09 had mostly intermediate binding affinities (100 nM ≤ IC50 < 250 nM). On the other hand, the risk HLA-DPA1*02:02/HLA-DPB1*05:01 genotype showed a larger proportion of predicted epitopes with relatively lower binding affinities. (250 nM ≤ IC50 < 500 nM) (Table S10).

### 4 DISCUSSION

We conducted a GWAS and found a significant cluster of SNPs within the HLA class II region that was associated with chronic HBV infection. Due to extensive polymorphisms and linkage disequilibrium patterns among HLA class II loci, we used imputation to examine HLA class II region. We found that the HLA-DRB1*13:02, HLA-DQA1*01:02/HLA-DQB1*06:09 and HLA-DPA1*02:02/HLA-DPB1*05:01 did not show significant correlations with global HBsAg seroprevalence.
Abbreviations: CI, confidence interval; HBV, hepatitis B virus; HLA, human leukocyte antigen; OR, odds ratio.

<table>
<thead>
<tr>
<th>HLA haplotypes</th>
<th>Chronic infected cases (n = 549)</th>
<th>Resolved controls (n = 2688)</th>
<th>Crude OR (95% CI)</th>
<th>P value</th>
<th>Adjusted OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N  %</td>
<td>N  %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non DRB1<em>13:02-DQB1</em>06:09</td>
<td>538 98.0</td>
<td>2.538 94.4</td>
<td>1.00 (referent)</td>
<td></td>
<td>1.00 (referent)</td>
<td></td>
</tr>
<tr>
<td>DRB1<em>13:02-DQB1</em>06:09</td>
<td>11 2.0</td>
<td>150 5.6</td>
<td>0.35 (0.19-0.64)</td>
<td>0.0008</td>
<td>0.33 (0.18-0.63)</td>
<td>0.0006</td>
</tr>
<tr>
<td>Non DPA1<em>02:02-DPB1</em>05:01</td>
<td>129 23.5</td>
<td>882 32.8</td>
<td>1.00 (referent)</td>
<td></td>
<td>1.00 (referent)</td>
<td></td>
</tr>
<tr>
<td>DPA1<em>02:02-DPB1</em>05:01</td>
<td>420 76.5</td>
<td>1.806 67.2</td>
<td>1.59 (1.29-1.97)</td>
<td>&lt;0.0001</td>
<td>1.59 (1.28-1.97)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

TABLE 3  HLA Class II haplotypes and their associated ORs for chronic HBV infection

Abbreviations: CI, confidence interval; HBV, hepatitis B virus; HLA, human leukocyte antigen; OR, odds ratio.

*Adjusted for age, sex, serum alanine aminotransferase level, body mass index, cigarette smoking and alcohol consumption.

Hence, the allele frequencies of HLA-DPB1*05:01 across different ethnic populations. The population attributable
conclusive. Most works only included a limited number of study
DPA1*02:02 relevant for outcomes after HBV acquisition.

Haplotypes highlighting the fact that previous reports in different ethnic groups. The population frequencies

works did not comprehensively examine the polymorphic HLA genes, which made difficulties for interpretation. However, by using a high-resolution HLA imputation with satisfactory accuracy, we were able to effectively analyze the associations with chronic HBV infection across the whole spectrum of HLA class II genes. Our study consistently found that HLA-DPA1*02:02 and HLA-DPB1*05:01 were positively associated with chronic HBV infection. On the other hand, HLA-DRB1*13:02, HLA-DQA1*01:02 and HLA-DQB1*06:09 showed negative associations with chronic HBV infection, which was consistent with previous reports in different ethnic groups. The haplotype patterns HLA-DPA1*02:02-DPB1*05:01 and HLA-DRB1*13:02-DQB1*06:09 had ever been identified in previous studies. Our findings highlight the fact that HLA imputation may provide an advantage in the quest to exhaustively understand the disease associations.

HBV endemic regions had considerable higher frequencies of HLA-DPA1*02:02 and HLA-DPB1*05:01 compared to regions with low prevalence of HBV. The positive correlation of HLA-DPB1*05:01 remained significant after excluding the data of the high prevalent area, suggesting that the relevance of HLA-DPB1*05:01 on HBV chronicity worldwide. Interestingly, the frequencies of these two risk alleles were still prevalent in Japan, which is a low endemic region for HBV. One possible explanation is that the homozygotes of the HLA-DPB1 gene, which consist of the allele *05:01 and any of the *02:01, *04:01 and *04:02 alleles, are common in Japan. The HLA-DPB1*02:01, HLA*04:01 and HLA*04:02 alleles were found to be associated with a higher likelihood of resolving HBV infection. These protective alleles may have a stronger ability to drive immune cells to clear HBV infection. Thus, the protective effects from chronic HBV infection may dominate the risk effect of HLA-DPB1*05:01, which was found to be associated with HBV persistent. On the other hand, there were no obvious correlations between the frequencies of HLA-DRB1*13:02, HLA-DQA1*01:02, HLA-DQB1*06:09 and HBsAg seroprevalence. The epidemiological characteristics of HBV, HBV transmission routes and predominant HBV genotypes among various populations are distinctively different. Our findings provide a clue for future comparative analyses of HLA allele frequencies and HBV patterns across worldwide populations. International collaborative meta-analysis will be helpful for validating these findings. Several GWASs have been conducted to examine the associations between genetic variants and chronic HBV infection. However, the different definitions of chronic HBV infection lead to conflicting findings. Most of the previous studies did not consider the history of past HBV acquisition, as defined by anti-HBc and HBsAg. In addition, including chronic HBV patients that already have liver cancer may result in findings that are difficult to interpret. Previously reported SNPs, that included rs3077 (HLA-DPA1), rs9277535 (HLA-DPB1), rs2856718 and rs7453920 (HLA-DQ) were validated in our study.

HLA variants positively associated with chronic HBV infection suggest that individuals may not have sufficient immune responses to clear the virus; persistent liver inflammation could then increase the risk for end-stage liver diseases. The risk allele G of rs3077, which is in linkage with HLA-DPA1*02:02, was found to be strongly associated with both chronic HBV infection and hepatocellular carcinoma development. The variants might increase advanced liver diseases by mediating HBV controls. Individuals who carry HLA-DPA1*02:02 might be unable to clear the virus efficiently then lead to long-term chronic inflammation of the liver. All participants were healthy adults free of cancers at enrolment in this study. It would be interesting to examine these HLA variants and liver diseases after follow-up.

HLA class II molecules are mainly encoded by three isotypes, including HLA-DR, HLA-DQ and HLA-DP. These molecules function as a heterodimer and consist of alpha and beta chains on the antigen-presenting cells. Foreign antigen peptides that bind to HLA class II molecules with high binding affinities may strengthen CD4+ T cell recognition and initiate a cascade of adaptive immune responses to clear HBV. Hepatitis B nucleocapsid protein elicited more vigorous
CD4+ T cell responses than the envelope protein.32,33 We found that HLA-DRB1*13:02 had better binding affinity to peptide epitopes derived from the HBV nucleocapsid protein thus it was associated with a higher probability of HBV clearance.21,23,27,28,32 On the other hand, HLA-DPB1*05:01 had relatively low binding affinity, which was found to be associated with HBV vaccine nonresponse,34 and in agreement with our findings. Our findings suggest that various HLA variants may initiate different nucleocapsid-specific CD4+ T cell responses to clear HBV.

Our study has some limitations. We did not have serial test data on HBV DNA and serum ALT levels, and therefore, we could not define inactive HBV infection. The findings may have clinical impact if the alleles could be validated in clinical settings to identify patients with inactive HBV infection. Due to the lack of serum HBV DNA measurements, occult HBV infection was unknown. However, the associations between HLA variants and HBV chronicity might be underestimated if there were subjects with occult HBV but were defined as HBV resolve in the current study. The prevalence of occult HBV infection in blood donors and unvaccinated individuals was quite low, with 0.1% and 1.7% respectively.35,36 Therefore, occult infection may influence the estimates with limited impacts. Second, whether the HLA alleles we found may be determinants for the spontaneous or anti-viral driven HBsAg loss needs further investigations. Finally, the most predominant genotypes among infected individuals in Taiwan are genotypes B and C. The prevalence of HBV genotypic differences may result in various attributable fractions of these HLA alleles on chronic HBV infection in other areas.

In conclusion, this community-based study found HLA class II variants were associated with chronic HBV infection. In addition, the frequencies of HLA-DPA1*02:02 and HLA-DPB1*05:01 were associated with HBsAg seroprevalence worldwide. Our findings provide information to future immunologic studies for current or innovative immune modulating regimens.

ACKNOWLEDGEMENTS

We thank all the study subjects who participated in the Taiwan Biobank.

Declaration of personal interests: The authors disclose no conflict of interest.

AUTHORSHIP

Guarantor of the article: Mei-Hsuan Lee.

Author contributions: Mei-Hsuan Lee and Yu-Han Huang were involved in study concept and design and drafting of the manuscript. Mei-Hsuan Lee and Hsuan-Yu Chen were involved in acquisition of data. Yu-Han Huang, Shu-Fen Liao, Seik-Soon Khor, Yu-Ju Lin, Hsuan-Yu Chen and Ya-Hsuan Chang were involved in analysis and interpretation of data. Shu-Fen Liao, Seik-Soon Khor, Yu-Ju Lin, Hsuan-Yu Chen, Yi-Hsiang Huang, Sheng-Nan Lu, Hye-Won Lee, Wen-Ya Ko, Claire Huang, Po-Chun Liu, Yen-Ju Chen, Ping-Feng Wu, Hou-Wei Chu, Pei-Ei Wu, Chen-Yang Shen and Katsushi Tokunaga were involved in critical revision of the manuscript for important intellectual content. Mei-Hsuan Lee obtained funding and study supervision.

All authors approved the final version and author list of the article.

ORCID

Yi-Hsiang Huang https://orcid.org/0000-0001-5241-5425
Sheng-Nan Lu https://orcid.org/0000-0002-5493-4752
Hye-Won Lee https://orcid.org/0000-0002-3552-3560
Mei-Hsuan Lee https://orcid.org/0000-0002-0222-2251

REFERENCES

15. Fan CT, Lin JC, Lee CH. Taiwan Biobank: a project aiming to aid Taiwan’s transition into a biomedical island. Pharmacogenomics. 2008;9:235-246.