A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population

Hamdi Mbarek¹, Hidenori Ochi²,³, Yuji Urabe¹, Vinod Kumar¹, Michiaki Kubo², Naoya Hosono², Atsushi Takahashi², Yoichiro Kamatani¹, Daiki Miki²,³, Hiromi Abe³, Tatsuhiko Tsunoda², Naoyuki Kamatani², Kazuaki Chayama²,³, Yusuke Nakamura¹,² and Koichi Matsuda¹,*

¹Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, ²Center for Genomic Medicine, RIKEN, Kanagawa, Japan and ³Division of Frontier Medical Science, Department of Medicine and Molecular Science, Programs for Biomedical Research, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima, Japan

Received February 23, 2011; Revised June 23, 2011; Accepted July 7, 2011

Hepatitis B virus (HBV) infection is a major health issue worldwide which may lead to hepatic dysfunction, liver cirrhosis and hepatocellular carcinoma. To identify host genetic factors that are associated with chronic hepatitis B (CHB) susceptibility, we previously conducted a two-stage genome-wide association study (GWAS) and identified the association of HLA-DP variants with CHB in Asians; however, only 179 cases and 934 controls were genotyped using genome-wide single nucleotide polymorphism (SNP) arrays. Here, we performed a second GWAS of 519,747 SNPs in 458 Japanese CHB cases and 2056 controls. After adjustment with the previously identified variants in the HLA-DP locus (rs9277535), we detected strong associations at 16 loci with \( P \)-value of \(<5 \times 10^{-5}\). We analyzed these loci in three independent Japanese cohorts (2209 CHB cases and 4440 controls) and found significant association of two SNPs (rs2856718 and rs7453920) within the HLA-DQ locus (overall \( P \)-value of \(5.98 \times 10^{-28}\) and \(3.99 \times 10^{-37}\)). Association of CHB with SNPs rs2856718 and rs7453920 remains significant even after stratification with rs3077 and rs9277535, indicating independent effect of HLA-DQ variants on CHB susceptibility (\( P \)-value of \(1.52 \times 10^{-21} – 2.38 \times 10^{-36}\)). Subsequent analyses revealed DQA1*0102-DQB1*0604 and DQA1*0101-DQB1*0501 (odds ratios (OR) = 0.16, and 0.39, respectively) as protective haplotypes and DQA1*0102-DQB1*0303 and DQA1*0301-DQB1*0601 (OR = 19.03 and 5.02, respectively) as risk haplotypes. These findings indicated that variants in antigen-binding regions of HLA-DP and HLA-DQ contribute to the risk of persistent HBV infection.

INTRODUCTION

Hepatitis B virus (HBV) is the most common cause of infectious liver diseases, and about 400 million people are suffering from chronic viral infection worldwide. Routes of infection include vertical transmission during neonatal period and horizontal transmission in childhood (bites, lesions and sanitary habits) or adulthood (sexual contact, drug use and medical exposure). In Japan, most of the chronic hepatitis B (CHB) patients were infected through vertical transmission and become HBV carrier (1). Nearly 90% of the HBV carrier will clear HBV (negative for HBsAg and positive for HBc ab) during adolescence, and only 10% of the HBV carrier indicate persistent liver dysfunction and develop chronic hepatitis (2). CHB dramatically increases the risk to progress to liver cirrhosis and hepatocellular carcinoma over a period of several decades (3,4). Currently, CHB is a serious public health problem worldwide, however pathogenesis of HBV-related diseases still remains elusive.
In addition to the viral and environmental factors, host genetic factors are considered to govern the pathology of disease development, progression or regression. Genetic epidemiological studies provide robust evidence that genetic variations contribute to progression from acute to chronic hepatitis (5). In 2009, our group conducted a genome-wide association study (GWAS) in the Asian population and identified a strong association of CHB with variants in the \( \text{HLA-DP} \) genes (6). In addition to our report, several association studies have suggested that genetic factors such as \( \text{HLA} \) (7–9), cytokines (10–12) and immune response-related genes (13–15) could influence the outcomes of HBV infection. However, these susceptibility loci were not identified in our previous study probably due to smaller sample size or smaller phenotypic effects of these loci. Here we conducted a second GWAS in the Japanese population to identify new susceptibility loci for CHB by increasing the number of samples in the screening stage from 179 cases and 934 controls to 458 case and 2056 controls.

**RESULTS**

We performed a two-stage GWAS followed by two independent replications as described in the Supplementary Material, Figure S1. In the GWAS stage, we genotyped 458 Japanese patients with CHB and 2056 control individuals using Illumina gene chip and obtained the genotyping results of 423,627 single nucleotide polymorphisms (SNPs) after quality control (QC). Examination of the quantile–quantile plots of the GWAS stage indicated no evidence for inflation of the test statistics, which could occur in the presence of population substructure (\( \lambda = 1.028 \)) and also revealed an enrichment of significant \( P \)-values, suggesting the possible existence of candidates (Supplementary Material, Fig. S2A). The results of genome-wide association analysis are represented in Supplementary Material, Table S2, where a total of 34 SNPs in the major histocompatibility complex (MHC) region satisfied the genome-wide significance level (\( P < 5.0 \times 10^{-8} \)). We also found 54 SNPs (40 in the MHC region and 14 in the non-MHC region) with suggestive associations (\( P < 5.0 \times 10^{-5} \)) (Supplementary Material, Fig. S2B and Tables S2 and S3).

In addition to our report, several association studies have suggested that genetic factors such as \( \text{HLA} \) (7–9), cytokines (10–12) and immune response-related genes (13–15) could influence the outcomes of HBV infection. However, these susceptibility loci were not identified in our previous study probably due to smaller sample size or smaller phenotypic effects of these loci. Here we conducted a second GWAS in the Japanese population to identify new susceptibility loci for CHB by increasing the number of samples in the screening stage from 179 cases and 934 controls to 458 case and 2056 controls.

In addition to the viral and environmental factors, host genetic factors are considered to govern the pathology of disease development, progression or regression. Genetic epidemiological studies provide robust evidence that genetic variations contribute to progression from acute to chronic hepatitis (5). In 2009, our group conducted a genome-wide association study (GWAS) in the Asian population and identified a strong association of CHB with variants in the \( \text{HLA-DP} \) genes (6). In addition to our report, several association studies have suggested that genetic factors such as \( \text{HLA} \) (7–9), cytokines (10–12) and immune response-related genes (13–15) could influence the outcomes of HBV infection. However, these susceptibility loci were not identified in our previous study probably due to smaller sample size or smaller phenotypic effects of these loci. Here we conducted a second GWAS in the Japanese population to identify new susceptibility loci for CHB by increasing the number of samples in the screening stage from 179 cases and 934 controls to 458 case and 2056 controls.

In addition to the viral and environmental factors, host genetic factors are considered to govern the pathology of disease development, progression or regression. Genetic epidemiological studies provide robust evidence that genetic variations contribute to progression from acute to chronic hepatitis (5). In 2009, our group conducted a genome-wide association study (GWAS) in the Asian population and identified a strong association of CHB with variants in the \( \text{HLA-DP} \) genes (6). In addition to our report, several association studies have suggested that genetic factors such as \( \text{HLA} \) (7–9), cytokines (10–12) and immune response-related genes (13–15) could influence the outcomes of HBV infection. However, these susceptibility loci were not identified in our previous study probably due to smaller sample size or smaller phenotypic effects of these loci. Here we conducted a second GWAS in the Japanese population to identify new susceptibility loci for CHB by increasing the number of samples in the screening stage from 179 cases and 934 controls to 458 case and 2056 controls.

In addition to the viral and environmental factors, host genetic factors are considered to govern the pathology of disease development, progression or regression. Genetic epidemiological studies provide robust evidence that genetic variations contribute to progression from acute to chronic hepatitis (5). In 2009, our group conducted a genome-wide association study (GWAS) in the Asian population and identified a strong association of CHB with variants in the \( \text{HLA-DP} \) genes (6). In addition to our report, several association studies have suggested that genetic factors such as \( \text{HLA} \) (7–9), cytokines (10–12) and immune response-related genes (13–15) could influence the outcomes of HBV infection. However, these susceptibility loci were not identified in our previous study probably due to smaller sample size or smaller phenotypic effects of these loci. Here we conducted a second GWAS in the Japanese population to identify new susceptibility loci for CHB by increasing the number of samples in the screening stage from 179 cases and 934 controls to 458 case and 2056 controls.

In addition to the viral and environmental factors, host genetic factors are considered to govern the pathology of disease development, progression or regression. Genetic epidemiological studies provide robust evidence that genetic variations contribute to progression from acute to chronic hepatitis (5). In 2009, our group conducted a genome-wide association study (GWAS) in the Asian population and identified a strong association of CHB with variants in the \( \text{HLA-DP} \) genes (6). In addition to our report, several association studies have suggested that genetic factors such as \( \text{HLA} \) (7–9), cytokines (10–12) and immune response-related genes (13–15) could influence the outcomes of HBV infection. However, these susceptibility loci were not identified in our previous study probably due to smaller sample size or smaller phenotypic effects of these loci. Here we conducted a second GWAS in the Japanese population to identify new susceptibility loci for CHB by increasing the number of samples in the screening stage from 179 cases and 934 controls to 458 case and 2056 controls.

In addition to the viral and environmental factors, host genetic factors are considered to govern the pathology of disease development, progression or regression. Genetic epidemiological studies provide robust evidence that genetic variations contribute to progression from acute to chronic hepatitis (5). In 2009, our group conducted a genome-wide association study (GWAS) in the Asian population and identified a strong association of CHB with variants in the \( \text{HLA-DP} \) genes (6). In addition to our report, several association studies have suggested that genetic factors such as \( \text{HLA} \) (7–9), cytokines (10–12) and immune response-related genes (13–15) could influence the outcomes of HBV infection. However, these susceptibility loci were not identified in our previous study probably due to smaller sample size or smaller phenotypic effects of these loci. Here we conducted a second GWAS in the Japanese population to identify new susceptibility loci for CHB by increasing the number of samples in the screening stage from 179 cases and 934 controls to 458 case and 2056 controls.
To further validate these results, we analyzed these SNPs in two additional Japanese cohorts consisting of 381 cases and 1539 controls from Biobank Japan as well as 1222 cases and 879 controls from Hiroshima University. Association for these SNPs loci was confirmed in both replication sets ($P$-value = $3.14 \times 10^{-5}$ – $3.59 \times 10^{-12}$; Table 1). To combine these studies, we conducted a meta-analysis with a fixed-effects model using the Mantel–Haenszel method. As shown in Table 1 and Supplementary Material, Figure S3, the OR were quite similar among the four studies and no heterogeneity was observed. Mantel–Haenszel $P$-values for independence were $3.99 \times 10^{-37}$ for rs2856718 (OR = 1.77, 95% confidence interval (CI) = 1.65–1.91), and $5.98 \times 10^{-28}$ for rs7453920 (OR = 1.81, 95% CI = 1.62–2.01). Two previously reported SNPs on the HLA-DP locus (rs9277535 on HLA-DPB1 and rs3077 on HLA-DPA1) were also associated with CHB ($P_{\text{meta-analysis}} = 2.55 \times 10^{-54}$ and $1.57 \times 10^{-61}$) (Table 1).

To test whether the strong association observed in these regions is due to the effect of one of them, we performed logistic regression analysis based on the effect of each top SNP in both HLA-DP and HLA-DQ loci. Notably, rs2856718 and rs7453920 did show strong association with CHB after adjusting for the effect of rs3077 ($P = 8.12 \times 10^{-27}$ and $P = 1.52 \times 10^{-21}$, respectively) and rs9277535 ($P = 2.38 \times 10^{-30}$ and $P = 2.21 \times 10^{-22}$, respectively), indicating variants at the HLA-DQ locus are associated with CHB independent of the effect of HLA-DP polymorphisms (Table 2).

(Fig. 3). To further validate these results, we analyzed these SNPs in two additional Japanese cohorts consisting of 381 cases and 1539 controls from Biobank Japan as well as 1222 cases and 879 controls from Hiroshima University. Association for these SNPs loci was confirmed in both replication sets ($P$-value = $3.14 \times 10^{-5}$ – $3.59 \times 10^{-12}$; Table 1). To combine these studies, we conducted a meta-analysis with a fixed-effects model using the Mantel–Haenszel method. As shown in Table 1 and Supplementary Material, Figure S3, the OR were quite similar among the four studies and no heterogeneity was observed. Mantel–Haenszel $P$-values for independence were $3.99 \times 10^{-37}$ for rs2856718 (OR = 1.77, 95% confidence interval (CI) = 1.65–1.91), and $5.98 \times 10^{-28}$ for rs7453920 (OR = 1.81, 95% CI = 1.62–2.01). Two previously reported SNPs on the HLA-DP locus (rs9277535 on HLA-DPB1 and rs3077 on HLA-DPA1) were also associated with CHB ($P_{\text{meta-analysis}} = 2.55 \times 10^{-54}$ and $1.57 \times 10^{-61}$) (Table 1).

To test whether the strong association observed in these regions is due to the effect of one of them, we performed logistic regression analysis based on the effect of each top SNP in both HLA-DP and HLA-DQ loci. Notably, rs2856718 and rs7453920 did show strong association with CHB after adjusting for the effect of rs3077 ($P = 8.12 \times 10^{-27}$ and $P = 1.52 \times 10^{-21}$, respectively) and rs9277535 ($P = 2.38 \times 10^{-30}$ and $P = 2.21 \times 10^{-22}$, respectively), indicating variants at the HLA-DQ locus are associated with CHB independent of the effect of HLA-DP polymorphisms (Table 2).
Table 1. Summary of results for GWAS and replication study

<table>
<thead>
<tr>
<th>Chr (position)</th>
<th>SNP</th>
<th>Nearest gene</th>
<th>Allele (1/2)</th>
<th>Stage</th>
<th>Case 11</th>
<th>12</th>
<th>22</th>
<th>MAFa</th>
<th>Control 11</th>
<th>12</th>
<th>22</th>
<th>MAFa</th>
<th>P-valueb</th>
<th>ORc (95% CI)</th>
<th>Pcontrol</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 (33141000)</td>
<td>rs3077</td>
<td>HLA-DPA1</td>
<td>A/G</td>
<td>GWAS</td>
<td>38</td>
<td>156</td>
<td>264</td>
<td>0.25</td>
<td>330</td>
<td>991</td>
<td>735</td>
<td>0.40</td>
<td>1.28 × 10⁻¹⁶</td>
<td>1.98 (1.68–2.32)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>First replication</td>
<td>42</td>
<td>240</td>
<td>324</td>
<td>0.27</td>
<td>313</td>
<td>947</td>
<td>762</td>
<td>0.39</td>
<td>1.93 × 10⁻¹⁴</td>
<td>1.74 (1.51–2.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Second replication</td>
<td>36</td>
<td>139</td>
<td>204</td>
<td>0.28</td>
<td>268</td>
<td>742</td>
<td>529</td>
<td>0.42</td>
<td>9.52 × 10⁻¹²</td>
<td>1.84 (1.55–2.19)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Third replication</td>
<td>115</td>
<td>430</td>
<td>681</td>
<td>0.27</td>
<td>155</td>
<td>420</td>
<td>304</td>
<td>0.42</td>
<td>1.53 × 10⁻²¹</td>
<td>1.93 (1.69–2.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Meta-analysis</td>
<td>1.57 × 10⁻⁶¹</td>
<td>1.87 (1.73–2.01)</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 (33162839)</td>
<td>rs9277535</td>
<td>HLA-DPB1</td>
<td>A/G</td>
<td>GWAS</td>
<td>40</td>
<td>179</td>
<td>239</td>
<td>0.28</td>
<td>384</td>
<td>1020</td>
<td>652</td>
<td>0.43</td>
<td>3.72 × 10⁻¹⁷</td>
<td>1.95 (1.67–2.28)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>First replication</td>
<td>58</td>
<td>254</td>
<td>294</td>
<td>0.31</td>
<td>364</td>
<td>963</td>
<td>696</td>
<td>0.42</td>
<td>3.70 × 10⁻¹²</td>
<td>1.63 (1.42–1.87)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Second replication</td>
<td>42</td>
<td>145</td>
<td>192</td>
<td>0.30</td>
<td>301</td>
<td>758</td>
<td>480</td>
<td>0.44</td>
<td>5.43 × 10⁻¹²</td>
<td>1.83 (1.54–2.17)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Third replication</td>
<td>133</td>
<td>464</td>
<td>628</td>
<td>0.30</td>
<td>160</td>
<td>429</td>
<td>290</td>
<td>0.43</td>
<td>1.02 × 10⁻¹⁶</td>
<td>1.75 (1.54–1.99)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Meta-analysis</td>
<td>2.55 × 10⁻⁵⁴</td>
<td>1.77 (1.65–1.91)</td>
<td>0.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 (32778233)</td>
<td>rs2856718</td>
<td>HLA-DQB1</td>
<td>A/G</td>
<td>GWAS</td>
<td>158</td>
<td>226</td>
<td>73</td>
<td>0.41</td>
<td>477</td>
<td>1001</td>
<td>568</td>
<td>0.48</td>
<td>4.41 × 10⁻¹⁰</td>
<td>1.59 (1.37–1.85)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>First replication</td>
<td>209</td>
<td>266</td>
<td>127</td>
<td>0.43</td>
<td>484</td>
<td>966</td>
<td>572</td>
<td>0.48</td>
<td>1.07 × 10⁻⁷</td>
<td>1.43 (1.27–1.64)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Second replication</td>
<td>128</td>
<td>191</td>
<td>62</td>
<td>0.41</td>
<td>325</td>
<td>746</td>
<td>468</td>
<td>0.45</td>
<td>7.49 × 10⁻¹¹</td>
<td>1.72 (1.45–2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Third replication</td>
<td>465</td>
<td>530</td>
<td>227</td>
<td>0.40</td>
<td>216</td>
<td>420</td>
<td>243</td>
<td>0.48</td>
<td>3.59 × 10⁻¹²</td>
<td>1.59 (1.39–1.79)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Meta-analysis</td>
<td>3.99 × 10⁻³⁷</td>
<td>1.56 (1.45–1.67)</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 (32837990)</td>
<td>rs7453920</td>
<td>HLA-DQB2</td>
<td>A/G</td>
<td>GWAS</td>
<td>4</td>
<td>72</td>
<td>382</td>
<td>0.09</td>
<td>67</td>
<td>582</td>
<td>1407</td>
<td>0.17</td>
<td>1.27 × 10⁻¹⁰</td>
<td>2.20 (1.73–2.81)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>First replication</td>
<td>5</td>
<td>127</td>
<td>471</td>
<td>0.11</td>
<td>50</td>
<td>575</td>
<td>1397</td>
<td>0.17</td>
<td>5.47 × 10⁻⁶</td>
<td>1.56 (1.28–1.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Second replication</td>
<td>4</td>
<td>75</td>
<td>302</td>
<td>0.11</td>
<td>53</td>
<td>422</td>
<td>1064</td>
<td>0.17</td>
<td>3.14 × 10⁻⁵</td>
<td>1.69 (1.32–2.17)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Third replication</td>
<td>14</td>
<td>198</td>
<td>1011</td>
<td>0.09</td>
<td>19</td>
<td>245</td>
<td>615</td>
<td>0.16</td>
<td>2.21 × 10⁻¹¹</td>
<td>1.88 (1.56–2.27)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Meta-analysis</td>
<td>5.98 × 10⁻²⁸</td>
<td>1.81 (1.62–2.01)</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aMAF, minor allele frequency.
bP-value of the Cochrane–Armitage trend test for each stage.
cOR and CI are calculated using the non-susceptible allele as reference.
dP-value of the Breslow–Day test.
eResults of meta-analysis were calculated by the Mantel–Haenzel method.
Logistic regression results for the top SNPs in HLA-DP and HLA-DQ loci associated with CHB in all stages

<table>
<thead>
<tr>
<th>SNP</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>P-adjusted for rs3077</th>
<th>OR (95% CI)</th>
<th>P-adjusted for rs9277535</th>
<th>OR (95% CI)</th>
<th>P-adjusted for rs2856718</th>
<th>OR (95% CI)</th>
<th>P-adjusted for rs7453920</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3077</td>
<td>2.61E-10</td>
<td>1.57 (1.3–1.8)</td>
<td>1.43 (1.3–1.67)</td>
<td>1.43 (1.3–1.67)</td>
<td>7.45 × 10^-9</td>
<td>1.67 (1.55–1.79)</td>
<td>9.42 × 10^-8</td>
<td>1.73 (1.56–1.75)</td>
<td>1.48 (1.34–1.65)</td>
<td>10^-6</td>
</tr>
<tr>
<td>rs9277535</td>
<td>2.48</td>
<td>1.35 (1.45–1.75)</td>
<td>1.6 (1.5–1.76)</td>
<td>1.6 (1.5–1.76)</td>
<td>6.8 (10^-6</td>
<td>1.67 (1.51–1.85)</td>
<td>8.12E-10</td>
<td>1.67 (1.55–1.8)</td>
<td>4.96 × 10^-18</td>
<td>10^-18</td>
</tr>
<tr>
<td>rs2856718</td>
<td>2.54</td>
<td>1.43 (1.45–1.75)</td>
<td>1.6 (1.5–1.76)</td>
<td>1.6 (1.5–1.76)</td>
<td>6.8 (10^-6</td>
<td>1.67 (1.51–1.85)</td>
<td>8.12E-10</td>
<td>1.67 (1.55–1.8)</td>
<td>4.96 × 10^-18</td>
<td>10^-18</td>
</tr>
<tr>
<td>rs7453920</td>
<td>2.51</td>
<td>1.43 (1.34–1.65)</td>
<td>1.6 (1.5–1.76)</td>
<td>1.6 (1.5–1.76)</td>
<td>6.8 (10^-6</td>
<td>1.67 (1.51–1.85)</td>
<td>8.12E-10</td>
<td>1.67 (1.55–1.8)</td>
<td>4.96 × 10^-18</td>
<td>10^-18</td>
</tr>
</tbody>
</table>

Subsequently, we examined the interaction of four SNPs in HLA-DP and HLA-DQ genes on CHB susceptibility. We only found evidence for interactive effects between HLA-DP SNPs and also between HLA-DQ SNPs (Supplementary Material, Table S6). For all other pairwise combinations, each locus had an independent role in CHB ($P_{\text{interaction}} > 0.10$). CHB risk increases with increasing number of risk alleles for four SNPs (Fig. 4 and Supplementary Material, Table S7). Individuals with seven or eight risk alleles have more than 5-fold higher CHB risk than those with two or less risk alleles. Taken together, our findings clearly indicated the additive effects of variants in HLA-DP and HLA-DQ loci on CHB susceptibility.

HLA-DQ molecules function as a heterodimer of α and β subunits, those are encoded by the HLA-DQA1 and the HLA-DQB1 genes, respectively. The SNP rs2856718 is located in a linkage disequilibrium (LD) block including HLA-DQB1 and HLA-DQA1 genes, and rs7453920 and rs2856718 are in LD with $r^2$ of 0.1 and $D'$ of 0.73 (Fig. 3 and Supplementary Material, Fig. S4). Similar to HLA-DPs, HLA-DQs are highly polymorphic especially in exon 2 which encode antigen-binding sites. We therefore considered that the association of these SNPs with CHB might reflect variations in antigen-binding sites of HLA-DQA1 and DQB1 that would affect the immune response to HBV. Hence, we genotyped HLA-DQA1 and DQB1 alleles by direct sequencing of exon 2 (cases and controls from the GWAS and first replication sets) and found HLA-DQB1*0303 and DQB1*0602 were significantly associated with CHB susceptibility ($P = 1.49 \times 10^{-6}$ and $1.87 \times 10^{-7}$, OR = 1.64 and 2.51, respectively), while DQB1*0501 and DQB1*0604 were significantly associated with protection from persistent HBV infection ($P = 3.61 \times 10^{-4}$ and $5.38 \times 10^{-16}$, OR = 0.50 and 0.22, respectively) (Supplementary Material, Table S8). To further investigate the relationship between HLA-DQ alleles and CHB susceptibility, we performed logistic regression analysis using SNPs rs2856718 and rs7453920 as covariates. Interestingly, HLA-DQB1*0303 and*0604 showed strong association with CHB after adjustment for rs2856718 and rs7453920 ($P = 6.5 \times 10^{-7}$ and $P = 2.59 \times 10^{-8}$, respectively). In addition, we performed logistic regression analysis using the top HLA-DQ alleles that show the strongest association (DQB1*0303, *0602, *0501, *0604) as covariate. As expected, HLA-DQ SNPs rs2856718 and rs7453920 failed to find the association between CHB and those SNPs ($P = 0.36$, and $P = 0.08$, respectively). Finally, we performed conditional analysis of the DQB1, DPA1 and DPB1 alleles together. As a result, HLA-DP SNPs rs3077 and rs9277535 as well as HLA-DQ SNPs rs2856718 and rs7453920 did not show any further association beyond these HLA-DQ and DP alleles (rs9277535, $P = 0.55$, OR = 0.88; rs3077, P = NA; rs2856718, P = 0.63, OR = 0.95 and rs7453920, P = 0.30, OR = 0.85). We also performed conditional analysis of the DPA1 and DPB1 and we found that HLA-DQ alleles *0303, *0602 and *0604 still showed strong association ($P = 0.0006$, OR = 1.5; $P = 0.00047$, OR = 2.28 and $P = 6.66 \times 10^{-7}$, OR = 0.31) except for DQB1*0501 ($P = 0.35$, OR = 0.81) which already showed weak association before adjustment as shown in Supplementary Material, Table S8. Collectively, these results together confirmed our findings for the
Recent GWASs have identified several SNPs that are associated with viral and non-viral liver diseases as well as response to HBV vaccination and liver function test (16–18). More recently, Zhang et al. (19) performed a GWAS of hepatocellular carcinoma in chronic HBV carriers of Chinese ancestry. They successfully identified one intronic SNP rs17401966 in KIF1B on chromosome 1p36.22 that was highly associated with HBV-related hepatocellular carcinoma. We analyzed those loci in our GWAS data, but failed to find the association between CHB and those SNPs (Supplementary Material, Table S10).

**DISCUSSION**

Here, we present the results of the two-stage GWAS followed by two independent replications on a total of 2667 cases with CHB and 6496 controls in Japanese population. In this study, we genotyped additional 279 cases and 1122 controls by using Illumina Human610-Quad BeadChip. As a result, we increased the number of samples in the first screening from 179 cases and 934 controls in the previous study to 458 cases and 2056 controls in current study. As a result, the statistic power to detect SNPs with moderate effects (i.e. OR of 1.4 and risk allele frequency of 0.2) increased from 23 to 85% at a significance threshold of 5 × 10^{-5}. Indeed, two SNPs in HLA-DQ locus did not indicate significant association in the GWAS stage of our previous GWAS (P = 5.62 × 10^{-2} for rs2856718 and P = 4.88 × 10^{-2} for rs7453920), confirming the importance of sample size in GWAS (20).

Most of significant SNPs with P-value of smaller than 5 × 10^{-5} (74 among 88 SNPs) are located in the MHC region which encompasses a large number of genes involved in our immunological response.

Three groups of HLA class II genes produce cell-surface Ag, designated HLA-DR, HLA-DQ and HLA-DP. It is suggested that the host immune response to HBV is under T lymphocyte control, and this response has been shown to be HLA-restricted (21). The HLA-DQ locus is located ~300 kb telomeric of the HLA-DP locus in a different LD block. Indeed, the analysis of the HLA complex revealed several recombination hot spots distributing across the HLA complex, including two hot spots near DP and DQ genes (22,23). The result of conditional analyses also demonstrated that the association of the HLA-DQ locus with CHB is independent from that of the HLA-DP locus.

Previous reports showed an association of HLA class II alleles with susceptibility of persistent HBV infection (24–27), but the results were inconsistent even within the same population except for HLA-DR13. HLA-DR13 (corresponding to HLA-DRB1*1301 and *1302 alleles) was consistently associated with HBV clearance across the population, and we found that rs11752643 which is strongly linked with HLA-DR13 (28) showed a strong association in the GWAS stage (P = 1.26 × 10^{-10}). The SNP rs3892710 which is in strong LD with rs11752643 (r^2 = 0.8, D’ = 1) and showed higher association in the GWAS stage (P = 4.49 × 10^{-12}) was selected for replication in the first independent replication set. However, rs3892710 failed to clear Bonferroni correction.
for multiple testing after adjustment for rs9277535 ($P = 4.73 \times 10^{-2}$). In addition, the association of hepatitis B with HLA-DQ SNPs rs2856718 and rs7453920 remarkably attenuated after adjustment for rs11752643 using the logistic regression model ($P = 2.53 \times 10^{-6}$ and $P = 5.84 \times 10^{-4}$, respectively). Unlike HLA-DP SNPs, rs3077 and rs9277535 remained highly significant ($P = 7.74 \times 10^{-13}$ and $2.52 \times 10^{-12}$, respectively). Therefore, our findings clearly indicated that hepatitis B is associated with the variants on HLA-DQ loci independent of the association with SNP rs11752643 that is closely linked with HLA-DR13 and also reinforce the previous report of HLA-DQ-DR linkage. Thus, our study demonstrated that the association of CHB with the variants in the HLA-DQ locus was more prominent and consistent than those with HLA-DR13 in the Japanese population. However, the 19 major haplotypes shown in Supplementary Material, Table S9 accounted for only 51.80% of cases and 57.92% of controls, respectively. Therefore, the result of DP-DQ haplotype analyses should be carefully interpreted. Subsequently, further functional analysis including HLA-DR, DQ and DP is essential to fully elucidate the molecular mechanism whereby these variations confer CHB susceptibility.

In summary, we have demonstrated that genetic variations in the HLA-DQ genes were strongly associated with CHB in the Japanese population, and this association was independent from the HLA-DP genes which we reported previously. Considering the importance of the MHC region in the clearance after the infection of HBV, our findings should provide a novel insight that the antigen presentation on the HLA-DP and HLA-DQ molecules might be critical for virus elimination and play an important role in the development of CHB. We are confident that our findings would serve to allow better understanding of the pathogenesis of hepatitis B and contribute to better clinical outcome of the disease.

### MATERIALS AND METHODS

#### Study population

A total of 2667 cases and 6496 control subjects were analyzed in this study. Characteristics of each cohort are shown in Supplementary Material, Table S1. DNA samples from both CHB patients and non-HBV controls used in this study were obtained from the BioBank Japan at the Institute of Medical Science, the University of Tokyo (29) except for samples for the third replication. Among the BioBank Japan samples, we selected HBsAg-seropositive CHB patients with elevated serum aminotransferase levels for more than six months, according to the guideline for diagnosis and treatment of chronic hepatitis from The Japan Society of Hepatology (http://www.jsh.or.jp/medical/guidelines/index.html). The control groups for the GWAS and first replication as well as for the second replication consisted of subjects with diseases other than CHB (uterine cancer, esophageal cancer, hematological cancer, pulmonary tuberculosis, ovarian cancer, keloid, peripheral artery disease and ischemic stroke) that were also negative for HBsAg. Case and control samples for the third replication cohort were collected from hospitals participating to the Hiroshima Liver Study Group (listing of participating doctors in this study group can be obtained at http://home.hiroshima-u.ac.jp/naika1/research_profile/pdf/liver_study_group_e.pdf) and Tanonomon Hospital. All the participants provided written informed consent. This project was approved by the ethical committees at each institute.

#### SNP genotyping and QC

In the GWAS stage, 458 patients with CHB and 2056 non-HBV controls were genotyped using Illumina Infinium HumanHap550v3 or Illumina Infinium Human610-Quad DNA Analysis Genotyping BeadChip. SNP QC for all sets of samples was applied as follows: SNP call rate of
We analyzed HLA-DQA1 and HLA-DQB1 genotyping. We analyzed HLA-DQ genotypes using 748 cases and 614 controls (from GWAS and first replication sets). The second exons of the HLA-DQA1 and HLA-DQB1 genes were amplified and directly sequenced according to the protocol reported previously (31–33). HLA-DQA1 and DQB1 alleles were determined based on the alignment database of dbMHC.

Statistical analysis
In the GWAS stage and replication analyses, statistical significance of the association with each SNP was assessed using 1-df Cochran–Armitage trend test and logistic regression. We also thank the technical staff of the Laboratory for Genotyping Development at RIKEN.

Software
For general statistical analysis, we used R statistical environment version 2.11.1 (http://cran.r-project.org) or plink-1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/). Estimation of haplotype frequencies and analysis of haplotype association were performed by R package haplo.stats (34). Sequence variants in the second exons of HLA-DQA1 and HLA-DQB1 were analyzed by Sequencher 4.8. Haplovieview software was employed to analyze LD values and draw LD map.

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS
We would like to acknowledge members of Hiroshima Liver Study Group and the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan for supporting our study. We also thank the technical staff of the Laboratory for Genotyping Development at RIKEN.

Conflict of Interest statement. None declared.

FUNDING
This work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government.

REFERENCES


