Human Molecular Genetics, 2011, Vol. 20, No. 19 3884–3892
doi:10.1093/hmg/ddr301
Advance Access published on July 12, 2011

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

Hamdi Mbarek[1], Hidenori Ochi[2,3], Yuji Urabe[1], Vinod Kumar[1], Michiaki Kubo[2], Naoya Hosono[2], Atsushi Takahashi[2], Yoichiro Kamatani[1], Daiki Miki[2,3], Hiromi Abe[3], Tatsuhiko Tsunoda[2], Naoyuki Kamatani[2], Kazuaki Chayama[2,3], Yusuke Nakamura[1,2] and Koichi Matsuda[1,*]

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [2]Center for Genomic Medicine, RIKEN, Kanagawa, Japan and [3]Division 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 5.98×10^{-28} and 3.99×10^{-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 1.52×10^{-21}– 2.38×10^{-30}). 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.

*To whom correspondence should be addressed at: Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, 4-6-1, Shirokanedai, Minato, Tokyo 108-8639, Japan. Tel: +81 354495376; Fax: +81 354495123; Email: koichima@ims.u-tokyo.ac.jp

© The Author 2011. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com

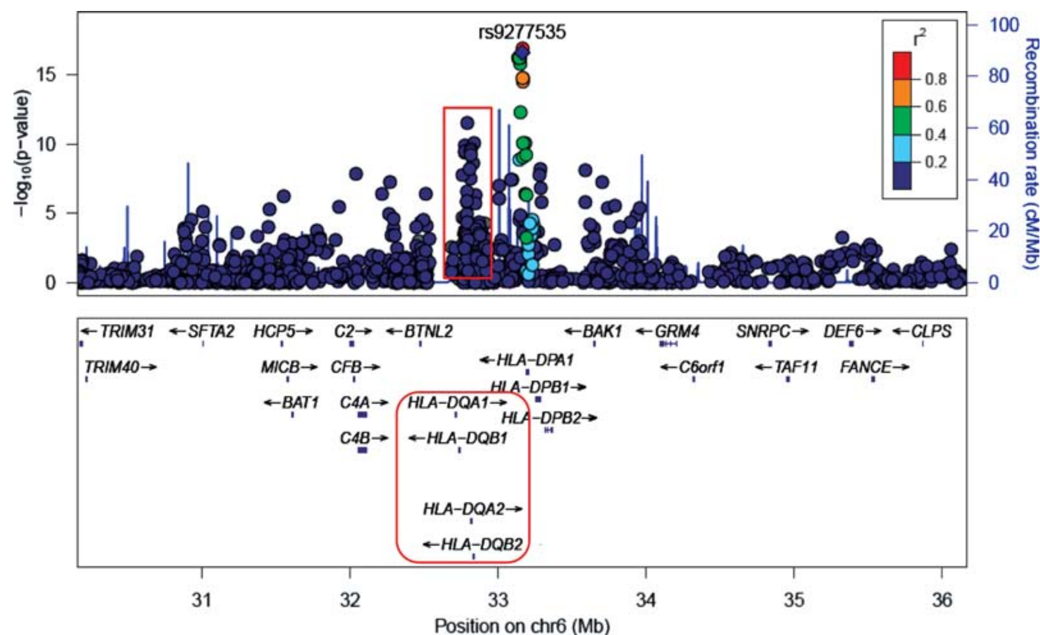

Figure 1. Signal of association with CHB in the *HLA* region of the GWAS stage. This figure shows the regional plots of the negative decadic logarithm trend *P*-values in a ~ 3000 kb window centered on the association peak, located at rs9277535 in *HLA-DPB1*. The top panel shows all SNPs in this region plotted according to the significance of their association with CHB and color coded according to their LD (r^2) with the most significant SNP, rs9277535 (see right corner of the plot). Vertical blue lines indicate local recombination rate. The bottom panel shows the genes in the region. The strongest signal on 6p21.32 localizes to *HLA-DP* genes and the second strongest signal localizes to *HLA-DQ* genes.

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 *HLA-DP* genes (6). In addition to our report, several association studies have suggested that genetic factors such as *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). We confirmed the most significant association at the *HLA-DP* locus as described in our previous report (rs9277535 and rs3077, $P = 3.72 \times 10^{-17}$ and 1.28×10^{-16}, respectively) (6) and found another significant peak around the *HLA-DQ* locus which is located ~300 kb telomeric to the *HLA-DP* locus (Fig. 1). To identify SNPs that are associated with CHB independently from *HLA-DP* SNPs, we conducted the association analysis after adjustment for a top SNP in the *HLA-DP* locus (rs9277535) using a logistic regression model (Fig. 2). As a result, five SNPs in the MHC region indicated suggestive associations ($P < 5.0 \times 10^{-5}$) even after stratification with rs9277535. Finally, 5 SNPs in the MHC region and 11 SNPs in the non-MHC region were selected for further analysis (Supplementary Material, Table S4).

Subsequently, we analyzed these 16 SNPs in the first replication set consisting of 606 cases and 2022 controls and found 2 SNPs within the MHC region [rs2856718, $P = 1.6 \times 10^{-5}$, odds ratios (OR) =1.33; rs7453920, $P = 5.72 \times 10^{-4}$, OR = 1.43] to be significantly associated with CHB after stratification for rs9277535 ($P_{\text{corrected}} < 3.0 \times 10^{-3}$, Supplementary Material, Table S5). The SNP rs2856718 is located in the intergenic region between *HLA-DQA2* and *HLA-DQB1*, while rs7453920 is located in intron 1 of *HLA-DQB2*

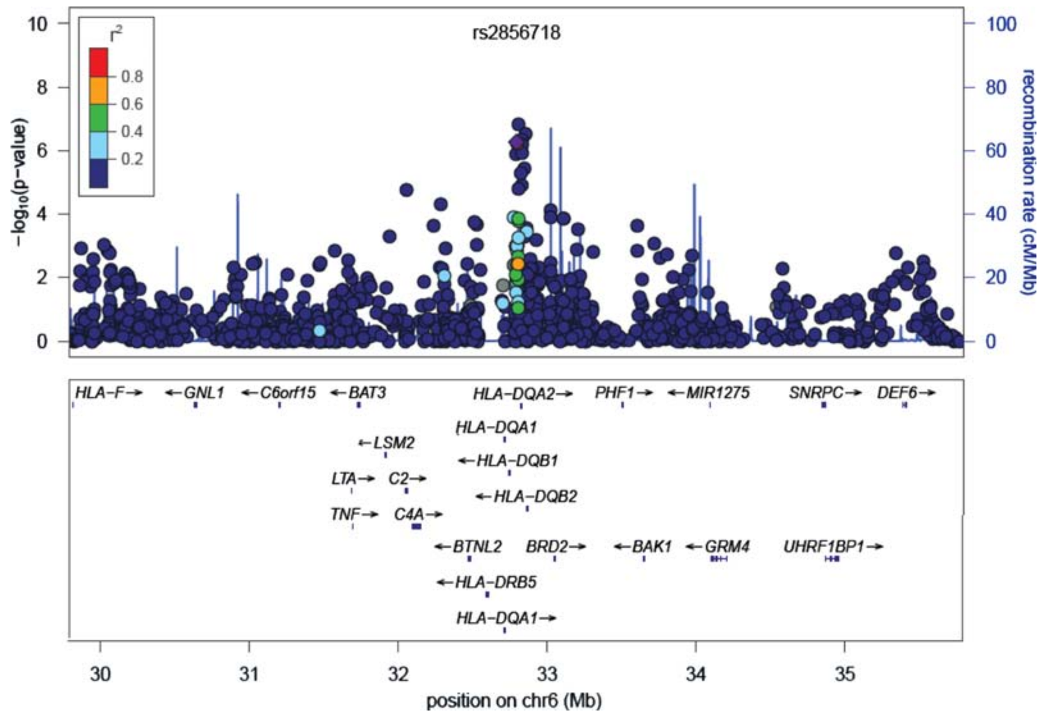

Figure 2. Regional association plot of the 6p21.32 locus after adjustment for the top SNP (rs9277535) in the *HLA-DP* locus in the GWAS stage. This figure shows the evidence of independent association with CHB based on logistic regression analysis. Only one strong peak remained after adjustment for rs9277535. This peak, represented by three top SNPs: rs3892710, rs7453920 and rs2856718, is located in the *HLA-DQ* locus (6p21.32).

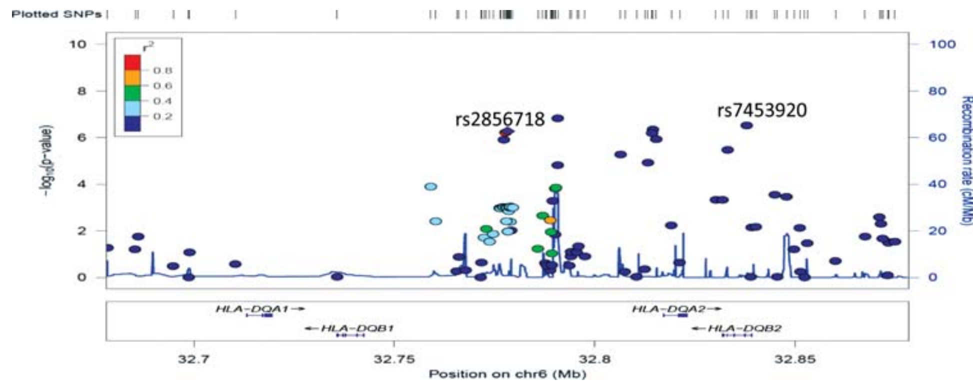

Figure 3. Regional association plot of the *HLA-DQ* locus. This figure indicates a ∼200 kb region centered on the association peak, located between rs2856718 and rs7453920. The middle panel shows the genes in this region including the *HLA-DQ* locus.

(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 $= 3.14 \times 10^{-5}$–3.59×10^{-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 3.99×10^{-37} for rs2856718 [OR = 1.77, 95% confidence interval (CI) = 1.65–1.91], and 5.98×10^{-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 1.57×10^{-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

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

[a]MAF, minor allele frequency.
[b]*P*-value of the Cochrane–Armitage trend test for each stage.
[c]OR and CI are calculated using the non-susceptible allele as reference.
[d]*P*-value of the Breslow–Day test.
[e]Results of meta-analysis were calculated by the Mantel–Haenzel method.

Table 2. Logistic regression results for the top SNPs in HLA-DP and HLA-DQ loci associated with CHB in all stages

SNP	P-value[a]	$P_{\text{adjusted for rs3077}}$	OR (95% CI)	$P_{\text{adjusted for rs9277535}}$	OR (95% CI)	$P_{\text{adjusted for rs2856718}}$	OR (95% CI)	$P_{\text{adjusted for rs7453920}}$	OR (95% CI)
rs3077	1.57×10^{-61}	NA	–	2.05×10^{-10}	1.43 (1.3–1.67)	7.45×10^{-48}	1.7 (1.58–1.83)	9.42×10^{-51}	1.73 (1.61–1.85)
rs9277535	2.55×10^{-54}	1.67E−05	1.25 (1.15–1.45)	NA	–	6.80×10^{-47}	1.67 (1.55–1.79)	9.03×10^{-48}	1.67 (1.56–1.8)
rs2856718	3.99×10^{-37}	8.12E−27	1.43 (1.33–1.54)	2.38×10^{-30}	1.43 (1.37–1.56)	NA	–	6.34×10^{-26}	1.43 (1.34–1.53)
rs7453920	5.98×10^{-28}	1.52E−21	1.66 (1.49–1.85)	2.21×10^{-22}	1.67 (1.51–1.85)	4.96×10^{-18}	1.60 (1.44–1.77)	NA	–

Trend *P*-values are shown with or without adjusting the analysis for the most associated SNPs in *HLA-DP* and *HLA-DQ* loci.
[a]Meta-analysis *P*-value was calculated by the Mantel–Haenzel method.

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 1.87×10^{-5}, 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 5.38×10^{-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 ($P = 6.3 \times 10^{-4}$ 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

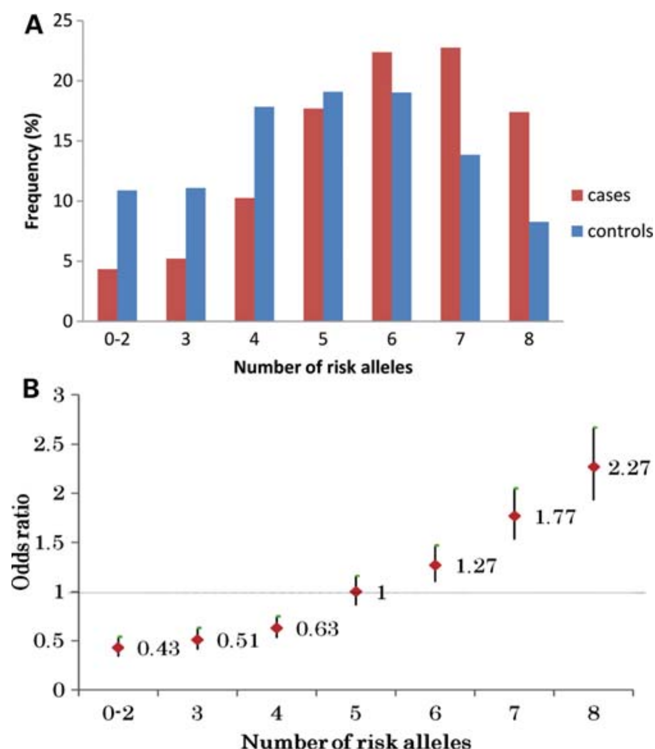

Figure 4. Cumulative effects of CHB risk alleles. (**A**) Distribution of risk alleles in CHB cases (red bars) and controls (blue bars). (**B**) Plot of the increasing OR for CHB with increasing number of risk alleles. The OR are relative to the median number of four risk alleles (rs3077, rs9277535, rs2856718 and rs7453920). Vertical bars correspond to 95% CIs. Horizontal line marks the null value (OR = 1).

causality of *HLA-DQ* and *HLA-DP* alleles and their independent effects on the CHB susceptibility. We further performed haplotype analysis and found four haplotypes showing the highest association ($8.39 \times 10^{-5} - 3.42 \times 10^{-13}$); *DQA1*0102-DQB1*0604* and *DQA1*0101-DQB1*0501* were considered to have protective effects ($P = 3.42 \times 10^{-13}$, OR = 0.16 and $P = 1.06 \times 10^{-5}$, OR = 0.39, respectively), whereas *DQA1*0102-DQB1*0303* and *DQA1*0301-DQB1*0601* increased a risk of CHB ($P = 8.39 \times 10^{-5}$, OR = 19.03, and $P = 7.34 \times 10^{-5}$, OR = 5.02, respectively, Table 3). Furthermore, we performed integrated analysis to test the haplotypic relationship between HLA-DP and DQ. We found seven associated haplotypes: *DQA1*0501-DQB1*0301-DPA1*0202-DPB1*0501*, *DQA1*0301-DQB1*0401 DPA1*0103DPB1*0201*, *DQA1*0301-DQB1*0302-DPA1*0202-DPB1*0501* and *DQA1*0102-DQB1*0604-DPA1*0103-DPB1*0401* showed protective effects ($P = 1.90 \times 10^{-4}$, OR = 0.18; $P = 5.30 \times 10^{-3}$, OR = 0.27; $P = 5.90 \times 10^{-3}$, OR = 0.43 and $P = 9.70 \times 10^{-3}$, OR = 0.41, respectively), whereas *DQA1*0301-DQB1*0301-DPA1*0103-DPB1*0201*, *DQA1*0102-DQB1*0602-DPA1*0202-DPB1*0501* and *DQA1*0301-DQB1*0601-DPA1*0202-DPB1*0501* were associated with susceptibility to CHB ($P = 2.30 \times 10^{-3}$, OR = 4.9; $P = 9.30 \times 10^{-4}$, OR = 4.8 and $P = 3.30 \times 10^{-5}$, OR = 11, respectively, Supplementary Material, Table S9). Taken together, our findings strongly implicated the significant association of *HLA-DQ-DP* haplotypes with CHB.

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 ($P = 5.62 \times 10^{-2}$ for rs2856718 and $P = 4.88 \times 10^{-2}$ 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 ($P = 1.26 \times 10^{-10}$). The SNP rs3892710 which is in strong LD with rs11752643 ($r^2 = 0.8$, $D' = 1$) and showed higher association in the GWAS stage ($P = 4.49 \times 10^{-12}$) was selected for replication in the first independent replication set. However, rs3892710 failed to clear Bonferroni correction

Table 3. Haplotype analysis

No.	Haplotype *HLA.DQA1*	*HLA.DQB1*	Haplotype frequencies Case (%)	Control (%)	*P*-value[a]	OR[a] (95% CI)
1	*0102	*0604	1.22	6.59	3.42×10^{-13}	0.16 (0.09–0.29)
2	*0101	*0501	1.68	4.77	1.06×10^{-5}	0.39 (0.24–0.65)
3	*0501	*0301	3.06	5.79	1.52×10^{-3}	0.53 (0.35–0.79)
4	*0301	*0401	9.73	13.40	2.98×10^{-3}	0.76 (0.57–1.02)
5	*0301	*0302	5.08	7.56	1.67×10^{-2}	0.72 (0.50–1.02)
6	*0301	*0402	2.55	3.49	1.73×10^{-1}	0.74 (0.45–1.22)
7	*0401	*0402	1.31	1.62	4.91×10^{-1}	0.72 (0.36–1.44)
8	*0101	*0503	4.23	4.34	8.69×10^{-1}	0.94 (0.62–1.42)
9	*0103	*0601	18.70	18.90	9.11×10^{-1}	Reference
10	*0601	*0301	1.38	0.89	2.53×10^{-1}	1.46 (0.68–3.11)
11	*0301	*0503	1.48	0.95	2.06×10^{-1}	1.65 (0.74–3.68)
12	*0301	*0301	2.46	1.79	1.97×10^{-1}	1.33 (0.76–2.33)
13	*0101	*0502	2.09	1.39	1.89×10^{-1}	1.67 (0.90–3.11)
14	*0301	*0303	16.90	13.10	7.50×10^{-3}	1.32 (1–1.74)
15	*0102	*0602	3.39	1.55	3.47×10^{-3}	2.24 (1.28–3.92)
16	*0102	*0303	1.91	0.25	8.39×10^{-5}	19.03 (2.53–143.39)
17	*0301	*0601	2.45	0.42	7.34×10^{-5}	5.02 (1.87–13.45)

[a]*P*-values, OR and its 95% CIs of each haplotype were calculated as described in Materials and Methods.

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 2.52×10^{-12}, respectively). Therefore, our findings clearly indicated that hepatitis B is associated with the variants on *HLA-DP* 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, and other 314 haplotypes were missed due to low haplotype frequency (<1% in both cases and controls). 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/gudelines/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 Toranomon 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

$\geq$0.99 in both cases and controls and *P*-value of the Hardy–Weinberg equilibrium test of $\geq 1.0 \times 10^{-6}$ in controls. SNPs with minor allele frequency of $\leq$0.01 in both case and control samples were excluded from the further analysis. In the first replication, we genotyped an additional panel of 616 cases by multiplex polymerase chain reaction (PCR)-based Invader assay (Third Wave Technologies, Madison, WI, USA) (30). After excluding 10 cases with the call rate of $<$0.95, all cluster plots were visually analyzed by trained staffs and SNPs with ambiguous calls were excluded. Randomly selected 94 case samples in the GWAS stage were re-genotyped in the first replication and SNPs with concordance rates of $<$98% between two assays (Illumina and Invader) were excluded. In the subsequent replication analyses, we used the TaqMan genotyping system (Applied Biosystems, Foster City, CA, USA) or the multiplex PCR-based Invader assay.

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 Cochrane–Armitage trend test and logistic regression analysis adjusted with top SNP (rs9277535) in the *HLA-DP* locus. Significance levels after Bonferroni correction for multiple testing were $P = 3.0 \times 10^{-3}$ (0.05/16) in the first replication and $P = 0.025$ (0.05/2) in second and third replication. OR and CIs were calculated using the non-susceptible allele as a reference. The meta-analysis was conducted using the Mantel–Haenszel method. Heterogeneity among studies was examined by the Breslow–Day test. To assess the association of each *HLA* allele, we used Fisher's exact test on two-by-two contingency tables with or without each *HLA* allele. To analyze the association of haplotypes, we used R package haplo.stats. *P*-values for each haplotype were given by the results of a score test, and OR and 95% CIs were calculated from coefficients of the generalized linear model. OR of each haplotype were calculated relative to the major haplotype. All of these statistical values were calculated by function haplo.cc.

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. Haploview 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

1. Pungpapong, S., Kim, W.R. and Poterucha, J.J. (2007) Natural history of hepatitis B virus infection: an update for clinicians. Mayo Clin. Proc., **82**, 967–975.
2. Okada, K., Kamiyama, I., Inomata, M., Imai, M. and Miyakawa, Y. (1976) e Antigen and anti-e in the serum of asymptomatic carrier mothers as indicators of positive and negative transmission of hepatitis B virus to their infants. N. Engl. J. Med., **294**, 746–749.
3. Chemin, I. and Zoulim, F. (2009) Hepatitis B virus induced hepatocellular carcinoma. Cancer Lett., **286**, 52–59.
4. McMahon, B. (2009) The natural history of chronic hepatitis B virus infection. *Hepatology*, **49**, S45–S55.
5. Lin, T., Chen, C., Wu, M., Yang, C., Chen, J., Lin, C., Kwang, T., Hsu, S., Lin, S. and Hsu, L. (1989) Hepatitis B virus markers in Chinese twins. Anticancer Res., **9**, 737–741.
6. Kamatani, Y., Wattanapokayakit, S., Ochi, H., Kawaguchi, T., Takahashi, A., Hosono, N., Kubo, M., Tsunoda, T., Kamatani, N., Kumada, H. *et al.* (2009) A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nat. Genet., **41**, 591–595.
7. Thio, C., Carrington, M., Marti, D., O'Brien, S., Vlahov, D., Nelson, K., Astemborski, J. and Thomas, D. (1999) Class II HLA alleles and hepatitis B virus persistence in African Americans. J. Infect. Dis., **179**, 1004–1006.
8. Kummee, P., Tangkijvanich, P., Poovorawan, Y. and Hirankarn, N. (2007) Association of HLA-DRB1*13 and TNF-alpha gene polymorphisms with clearance of chronic hepatitis B infection and risk of hepatocellular carcinoma in Thai population. J. Viral. Hepat., **14**, 841–848.
9. Hwang, S., Sohn, Y., Oh, H., Hwang, C., Lee, S., Shin, E. and Lee, K. (2007) Human leukocyte antigen alleles and haplotypes associated with chronicity of hepatitis B virus infection in Koreans. Arch. Pathol. Lab. Med., **131**, 117–121.
10. Ben-Ari, Z., Mor, E., Papo, O., Kfir, B., Sulkes, J., Tambur, A., Tur-Kaspa, R. and Klein, T. (2003) Cytokine gene polymorphisms in patients infected with hepatitis B virus. Am. J. Gastroenterol., **98**, 144–150.
11. Höhler, T., Kruger, A., Gerken, G., Schneider, P., Meyer zum Büschenefelde, K. and Rittner, C. (1998) A tumor necrosis factor-alpha (TNF-alpha) promoter polymorphism is associated with chronic hepatitis B infection. Clin. Exp. Immunol., **111**, 579–582.
12. Migita, K., Maeda, Y., Abiru, S., Nakamura, M., Komori, A., Miyazoe, S., Nakao, K., Yatsuhashi, H., Eguchi, K. and Ishibashi, H. (2007) Polymorphisms of interleukin-1beta in Japanese patients with hepatitis B virus infection. J. Hepatol., **46**, 381–386.
13. Chong, W., To, Y., Ip, W., Yuen, M., Poon, T., Wong, W., Lai, C. and Lau, Y. (2005) Mannose-binding lectin in chronic hepatitis B virus infection. *Hepatology*, **42**, 1037–1045.

14. Thio, C., Mosbruger, T., Kaslow, R., Karp, C., Strathdee, S., Vlahov, D., O'Brien, S., Astemborski, J. and Thomas, D. (2004) Cytotoxic T-lymphocyte antigen 4 gene and recovery from hepatitis B virus infection. J. Virol., **78**, 11258–11262.
15. Zhou, J., Lu, L., Yuen, M., Lam, T., Chung, C., Lam, C., Zhang, B., Wang, S., Chen, Y., Wu, S. *et al.* (2007) Polymorphisms of type I interferon receptor 1 promoter and their effects on chronic hepatitis B virus infection. J. Hepatol., **46**, 198–205.
16. Davila, S., Froeling, F., Tan, A., Bonnard, C., Boland, G., Snippe, H., Hibberd, M. and Seielstad, M. (2010) New genetic associations detected in a host response study to hepatitis B vaccine. Genes Immun., **11**, 232–238.
17. Kamatani, Y., Matsuda, K., Okada, Y., Kubo, M., Hosono, N., Daigo, Y., Nakamura, Y. and Kamatani, N. (2010) Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat. Genet., **42**, 210–215.
18. Romeo, S., Kozlitina, J., Xing, C., Pertsemlidis, A., Cox, D., Pennacchio, L., Boerwinkle, E., Cohen, J. and Hobbs, H. (2008) Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat. Genet., **40**, 1461–1465.
19. Zhang, H., Zhai, Y., Hu, Z., Wu, C., Qian, J., Jia, W., Ma, F., Huang, W., Yu, L., Yue, W. *et al.* (2010) Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. Nat. Genet., **42**, 755–758.
20. Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
21. Milich, D. (1988) T- and B-cell recognition of hepatitis B viral antigens. *Immunol. Today*, **9**, 380–386.
22. Jeffreys, A., Kauppi, L. and Neumann, R. (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat. Genet., **29**, 217–222.
23. Cullen, M., Perfetto, S., Klitz, W., Nelson, G. and Carrington, M. (2002) High-resolution patterns of meiotic recombination across the human major histocompatibility complex. Am. J. Hum. Genet., **71**, 759–776.
24. Liu, C. and Cheng, B. (2007) Association of polymorphisms of human leucocyte antigen-DQA1 and DQB1 alleles with chronic hepatitis B virus infection, liver cirrhosis and hepatocellular carcinoma in Chinese. Int. J. Immunogenet., **34**, 373–378.
25. Xun, Y., Guo, J., Shi, W., Shi, J. and Liu, C. (2009) Association between HLA-DQA1 gene polymorphism and the outcomes of hepatitis B virus infection. *Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi*, **23**, 430–433.
26. Thio, C., Thomas, D., Karacki, P., Gao, X., Marti, D., Kaslow, R., Goedert, J., Hilgartner, M., Strathdee, S., Duggal, P. *et al.* (2003) Comprehensive analysis of class I and class II HLA antigens and chronic hepatitis B virus infection. J. Virol., **77**, 12083–12087.
27. Singh, R., Kaul, R., Kaul, A. and Khan, K. (2007) A comparative review of HLA associations with hepatitis B and C viral infections across global populations. World J. Gastroenterol., **13**, 1770–1787.
28. Thursz, M., Kwiatkowski, D., Allsopp, C., Greenwood, B., Thomas, H. and Hill, A. (1995) Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. N. Engl. J. Med., **332**, 1065–1069.
29. Nakamura, Y. (2007) The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.*, **5**, 696–697.
30. Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. and Nakamura, Y. (2001) A high-throughput SNP typing system for genome-wide association studies. J. Hum. Genet., **46**, 471–477.
31. van der Zwan, A., Griffith, B., Rozemuller, E., Williams, T. and Tilanus, M.G.J. (2002) Sequence-based typing for HLA-DQB1 strategy for ABI sequencing equipment. In Tilanus, M.G.J. (ed.), International Histocompatibility Working Group, Seattle, Washington, 2002.
32. van Dijk, A., Melchers, R., Tilanus, M. and Rozemuller, E. (2007) HLA-DQB1 sequencing-based typing updated. *Tissue Antigens*, **69**(Suppl. 1), 64–65.
33. Witter, K., Zahn, R., Volgger, A. and Reininger, A.J. (2010) HLA-DQB1*0404, a novel DQB1-allele detected in a volunteer blood platelet donor typed for HLA. *Tissue Antigens*, **76**, 256–258.
34. Schaid, D.J., Rowland, C.M., Tines, D.E., Jacobson, R.M. and Poland, G.A. (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am. J. Hum. Genet., **70**, 425–434.