nature genetics

A genome-wide association study identifies variants in the *HLA-DP* locus associated with chronic hepatitis B in Asians

Yoichiro Kamatani[1,2], Sukanya Wattanapokayakit[3], Hidenori Ochi[4,5], Takahisa Kawaguchi[4], Atsushi Takahashi[4], Naoya Hosono[4], Michiaki Kubo[4], Tatsuhiko Tsunoda[4], Naoyuki Kamatani[4], Hiromitsu Kumada[6], Aekkachai Puseenam[7], Thanyachai Sura[7], Yataro Daigo[1,2], Kazuaki Chayama[4,5], Wasun Chantratita[8], Yusuke Nakamura[1,4] & Koichi Matsuda[1]

Chronic hepatitis B is a serious infectious liver disease that often progresses to liver cirrhosis and hepatocellular carcinoma; however, clinical outcomes after viral exposure vary enormously among individuals[1]. Through a two-stage genome-wide association study using 786 Japanese chronic hepatitis B cases and 2,201 controls, we identified a significant association of chronic hepatitis B with 11 SNPs in a region including *HLA-DPA1* and *HLA-DPB1*. We validated these associations by genotyping two SNPs from the region in three additional Japanese and Thai cohorts consisting of 1,300 cases and 2,100 controls (combined $P = 6.34 \times 10^{-39}$ and 2.31×10^{-38}, OR = 0.57 and 0.56, respectively). Subsequent analyses revealed risk haplotypes (*HLA-DPA10202-*DPB1**0501 and *HLA-DPA1**0202-*DPB1**0301, OR = 1.45 and 2.31, respectively) and protective haplotypes (*HLA-DPA1**0103-*DPB1**0402 and *HLA-DPA1**0103-*DPB1**0401, OR = 0.52 and 0.57, respectively). Our findings show that genetic variants in the *HLA-DP* locus are strongly associated with risk of persistent infection with hepatitis B virus.**

Chronic hepatitis B is one of the most common infectious liver diseases caused by hepatitis B virus (HBV). HBV infection shows a marked regional diversity and is very prevalent in the Asia-Pacific region; HBsAg seropositivity rates are as high as 5–12% in Thai and China, but as low as 0.2–0.5% in North America and Europe[2]. It is estimated that, at present, more than 400 million people worldwide are chronically infected with HBV, and nearly 60% of liver cancers are considered to be related to chronic hepatitis B and subsequent liver cirrhosis[3]. Most HBV carriers are considered to have been infected through maternal transmission in the neonatal period or infancy, particularly in Japan[4]. Although some HBV carriers spontaneously eliminate the virus, 2–10% of individuals with chronic hepatitis B are estimated to develop liver cirrhosis every year, and a subset of these individuals suffer from liver failure or hepatocellular carcinoma[1]. Because clinical outcomes after exposure to HBV are highly variable, identification of genetic and environmental factors that are related to progression of HBV-induced liver diseases is critical.

Several epidemiological factors such as age at infection, sex, chronic alcohol abuse[5] and co-infection with other hepatitis viruses[6] were suspected to affect viral persistence. In addition, a twin study in Taiwan indicated that host genetic background influences infection outcome[7]. Although genetic variants in *IFNG*, *TNF*, *VDR*, *ESR1* and several *HLA* loci were shown to associate with chronic hepatitis B[8–12], none of the associations has been proven to be conclusive. To identify disease-predisposing variants, we carried out a two-stage association study for chronic hepatitis B using genome-wide SNPs as genetic markers.

Characteristics of each cohort group are shown in **Supplementary Table 1** online. We carried out a two-stage genome-wide association approach as described in the Methods. In the first stage, we genotyped 179 Japanese individuals with chronic hepatitis B and 934 control individuals using Illumina HumanHap550 BeadChip (**Fig. 1a**). For the second stage, we selected the top 12,000 SNPs that had the smallest *P* values on the basis of minimum *P* value considering three genetic models: allelic, dominant or recessive. Analysis of an independent set of 607 cases and 1,267 controls using these sub-selected SNPs showed 11 SNPs to be significantly associated ($P = 3.62 \times 10^{-8} \sim 1.16 \times 10^{-13}$) with chronic hepatitis B after Bonferroni correction (**Fig. 1b** and **Supplementary Table 2** online). Application of the Cochrane-Armitage

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, the University of Tokyo, Tokyo, Japan. [2]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, the University of Tokyo, Tokyo, Japan. [3]Medical Genetic Section, National Institute of Health, Department of Medical Sciences, Ministry of Public Health, Nonthaburi, Thailand. [4]Center for Genomic Medicine, RIKEN, Kanagawa, Japan. [5]Department of Medicine and Molecular Science, Division of Frontier Medical Science, Programs for Biomedical Research, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima, Japan. [6]Department of Hepatology, Toranomon Hospital, Tokyo, Japan. [7]Department of Medicine, Faculty of Medicine and [8]Virology and Molecular Microbiology Unit, Department of Pathology, Faculty of Medicine, Ramathidi Hospital, Mahidol University, Bangkok, Thailand. Correspondence should be addressed to Y.N. (yusuke@ims.u-tokyo.ac.jp).

Received 10 November 2008; accepted 16 January 2009; published online 6 April 2009; doi:10.1038/ng.348

test to all the tested SNPs indicated that the genetic inflation factor lambda was 1.02 for the second stage (**Supplementary Fig. 1a** online), implying a low possibility of false positive associations due to population stratification. All 11 SNPs are located within or around the *HLA-DPA1* and *HLA-DPB1* locus (**Fig. 2**). We also conducted age- and sex-adjusted analysis using a logistic regression model, and confirmed similar association after adjustment (data not shown).

To validate the result of the discovery-phase analysis, we carried out replication analyses using three independent cohorts. We selected the most or second-most strongly associated SNPs from each *HLA-DP* locus (rs9277535 on *HLA-DPB1* and rs3077 on *HLA-DPA1*, respectively), as we failed to design a Taqman or Invader probe for rs2395309 on *HLA-DPA1*. We first examined two independent sets of Japanese case–control samples comprising 274 cases and 274 controls (age-, sex- and alcohol consumption–matched cohort from BioBank Japan) as well as 718 cases and 1,280 controls. We found significant associations at two SNP loci in both studies ($P = 1.06 \times 10^{-16} \sim 1.96 \times 10^{-6}$; **Table 1**). We also genotyped 308 individuals with chronic hepatitis B and 546 healthy controls in Thailand, and further confirmed the association at the two loci, rs3077 ($P = 6.53 \times 10^{-6}$) and rs9277535 ($P = 6.52 \times 10^{-8}$).

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 Figure 1b**, the odds ratios (OR) were quite similar across the four studies (the second stage of GWAS and three replication studies) and no heterogeneity was observed. Mantel-Haenszel *P* values for independence were 2.31×10^{-38} for rs3077 (OR = 0.56, 95% confidence interval (CI) = 0.51–0.61), and 6.34×10^{-39} for rs9277535 (OR = 0.57, 95% CI = 0.52–0.62).

The 11 SNPs showing significant associations are located within a 50-kb region including *HLA-DPA1* and *HLA-DPB1* (**Fig. 2**). Although the *HLA* region is known to show extensive linkage disequilibrium (LD) spanning over 7 Mb, the LD block including these 11 SNPs (surrounded by a bold line in **Fig. 2a**) was not in strong LD with the other *HLA* loci. In accordance with the extent of LD, only SNPs around the *HLA-DPA1* and *HLA-DPB1* genes showed very strong associations with chronic HBV (surrounded by a bold line in **Fig. 2b**), and SNPs outside of this particular LD block did not have significant association.

HLA-DPA1 and *HLA-DPB1* encode the HLA-DP α and β chains, respectively. HLA-DPs belong to the HLA class II molecules that form heterodimers on the cell surface and present antigens to CD4-positive T lymphocytes. HLA-DPs are highly polymorphic, especially in exon 2, which encodes antigen-binding sites. We thus considered that the association of these SNPs with chronic HBV might reflect variations in antigen-binding sites that might affect the immune response to HBV. We genotyped *HLA-DPA1* and *HLA-DPB1* alleles by direct sequencing of exon 2 (cases at second stage and controls at first stage) and found significant association of chronic hepatitis B with *HLA-DPA1**0103, *DPA1**0202, *DPB1**0402 and *DPB1**0501 ($P = 2.93 \times 10^{-11}$, 4.45×10^{-8}, 2.27×10^{-7} and 6.98×10^{-7}, respectively; **Supplementary Table 3** online). Because sequence variants in exon 2 of *HLA-DPA1* and *HLA-DPB1* could be linked to individual nucleotide variants, we inferred haplotypes using the 11 SNPs and variants in exon 2, and found very strong LD among them (**Supplementary Fig. 2**

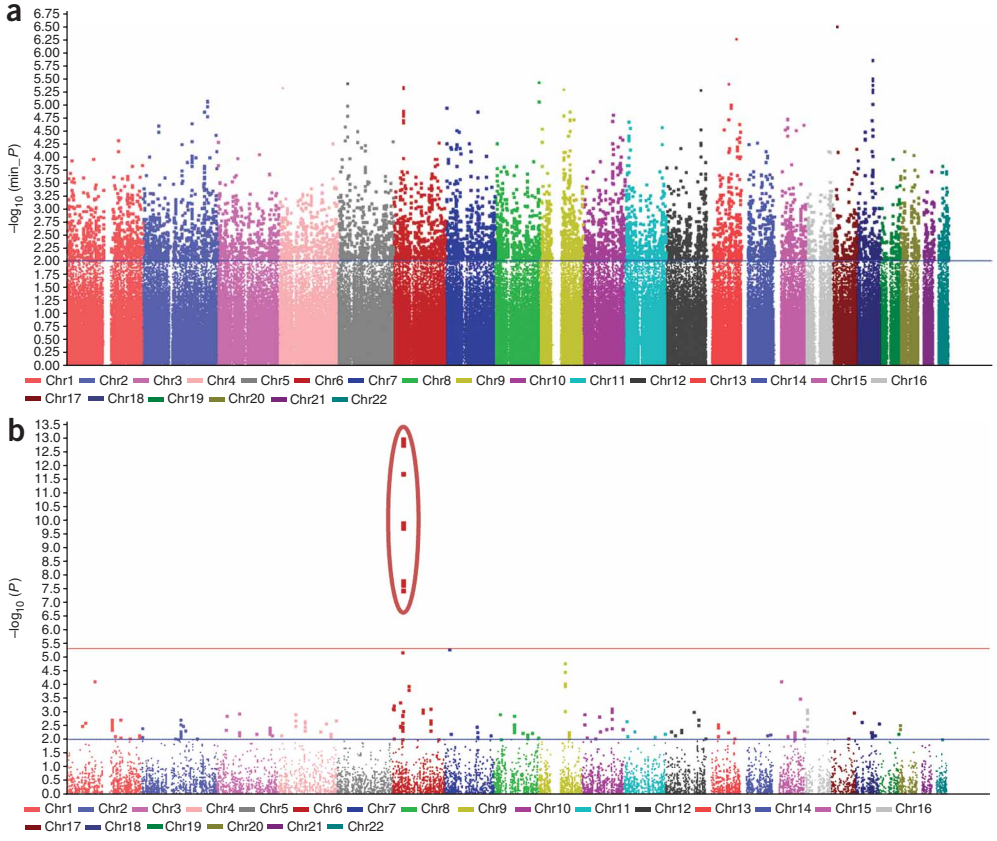

Figure 1 Results from a two-stage genome-wide association study. (**a**) $-\log_{10}$ *P* value plot at the first stage. Each *P* value is the minimum of Fisher's exact tests for three models: dominant, recessive and allele frequency model. (**b**) $-\log_{10}$ *P* value plot at the second stage. *P* values were calculated by 1-d.f. Cochrane-Armitage trend test. The large dots circled by red on the chromosome 6 showed significant associations ($P < 5.06 \times 10^{-6}$) with chronic hepatitis B.

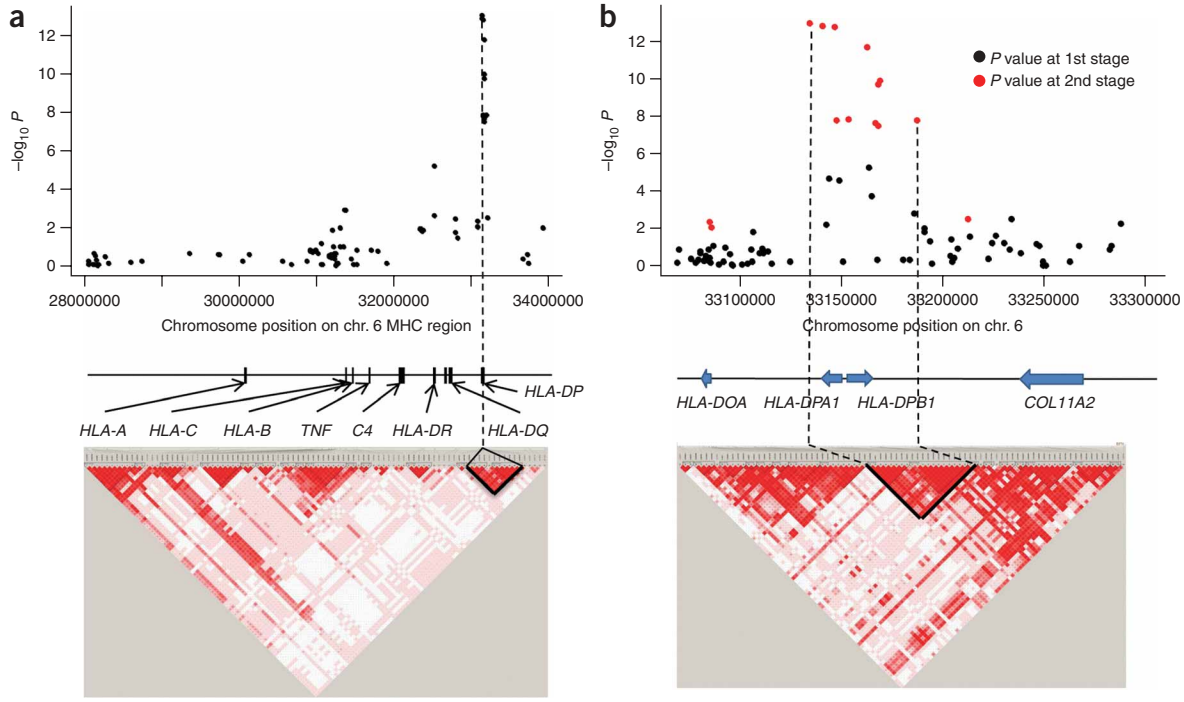

Figure 2 Case–control association results and linkage disequilibrium map of the MHC region. (**a**) *P*-value plot, genomic structure and LD map of the second stage within the extended MHC region of chromosome 6. The LD map based on *D′* was drawn using the genotype data of the cases and the controls in the second stage. (**b**) *P*-value plot, genomic structure and LD map around the *HLA-DPA1* and *HLA-DPB1* region. Black dots and red dots represent *P* values in the first and the second stage, respectively. The LD map based on *D′* was drawn using the genotype data of the cases and the controls in the first stage.

online). Case–control analyses revealed four associated haplotypes: *DPA1**0103-*DPB1**0402 and *DPA1**0103-*DPB1**0401 showed protective effects ($P = 6.00 \times 10^{-8}$, OR = 0.52, 95% CI = 0.35–0.75 and $P = 0.002$, OR = 0.57, 95% CI = 0.33–0.96, respectively), whereas *DPA1**0202-*DPB1**0501 and *DPA1**0202-*DPB1**0301 were associated with susceptibility to chronic hepatitis B ($P = 5.79 \times 10^{-6}$, OR = 1.45, 95% CI = 1.16–1.81 and $P = 0.002$, OR = 2.31, 95% CI = 1.39–3.84, respectively; **Table 2**). We also found various sets of SNPs (tagging SNPs) that could predict *HLA-DP* alleles (**Supplementary Table 4** online). Taken together, our findings strongly implicate an association of genetic variants in the *HLA-DPA1* and *HLA-DPB1* genes with chronic hepatitis B.

HLA-DR13 was reported to have a protective effect against persistent HBV infection in different populations[9,13,14]. Comparison of genotypes of *HLA-DRB1**1301 and *1302 alleles (both corresponding to HLA-DR13) and Illumina HumanHap550 SNPs in 333 of the first-stage control samples revealed that the A allele of rs11752643 was in strong LD with HLA-DR13 ($r^2 = 0.83$, $D' = 1$). However, the association between rs11752643 and chronic hepatitis B was not significant in our second stage GWAS, with an uncorrected *P* value of 1.04×10^{-4} (**Supplementary Table 5** online). In addition, the association of chronic hepatitis B with rs3077 and rs9277535 remained highly significant ($P = 2.11 \times 10^{-10}$ and 1.73×10^{-9}, respectively) after adjustment for rs11752643 using a logistic

Table 1 Results of replication studies and meta-analysis

SNP	Nearest gene	Allele (1/2)	Stage	Cases			Controls			OR (95%CI)[a]	P^b	P_{het}^c
				11	12	22	11	12	22			
rs3077	*HLA-DPA1*	A/G	GWAS second stage	42	240	324	197	598	472	0.57 (0.49–0.66)	1.26E–13	
			First replication	25	95	152	50	122	102	0.53 (0.41–0.69)	1.73E–06	
			Second replication	64	237	410	197	596	485	0.55 (0.47–0.63)	1.06E–16	
			Third replication	28	109	163	85	250	210	0.61 (0.49–0.75)	6.53E–06	
			Meta-analysis[d]							0.56 (0.51–0.61)	2.31E–38	0.84
rs9277535	*HLA-DPB1*	A/G	GWAS second stage	58	254	294	230	619	418	0.59 (0.51–0.69)	1.78E–12	
			First replication	26	102	144	49	132	91	0.54 (0.42–0.69)	1.96E–06	
			Second replication	68	264	376	227	604	445	0.56 (0.48–0.64)	1.81E–16	
			Third replication	29	136	139	107	273	155	0.56 (0.46–0.69)	6.52E–08	
			Meta-analysis[d]							0.57 (0.52–0.62)	6.34E–39	0.85

Odds ratio and *P* values for independence test were calculated by the Mantel-Haenszel method.
[a]Odds ratio of minor allele from two-by-two allele frequency table. [b]*P* values of Pearson's χ^2 test for allele model. [c]Result of Breslow-Day test. [d]Meta-analysis of all four studies.

Table 2 Haplotype analysis

No.	Haplotype[a]	Frequency (cases)	Frequency (controls)	P^b	OR[b] (95% CI)
1	GG-*DPA1*0202*-TCG-*DPB1*0501*-GAGATT	0.428	0.347	5.79E–06	1.45 (1.16–1.81)
2	AA-*DPA1*0103*-CCA-*DPB1*0201*-AGTGCC	0.165	0.192	0.052	Reference
3	GG-*DPA1*0201*-TCG-*DPB1*0901*-GGGGTC	0.129	0.124	0.642	1.21 (0.91–1.61)
4	AA-*DPA1*0103*-CTA-*DPB1*0402*-AGTGCC	0.042	0.096	6.00E–08	0.52 (0.35–0.75)
5	AA-*DPA1*0103*-CCA-*DPB1*0401*-AGTGCC	0.018	0.038	0.002	0.57 (0.33–0.96)
6	GG-*DPA1*0202*-TCG-*DPB1*0301*-GGGGTC	0.036	0.018	0.002	2.31 (1.39–3.84)
7	GG-*DPA1*0202*-TCG-*DPB1*0202*-AGTGCC	0.020	0.027	0.257	0.88 (0.51–1.52)
8	GG-*DPA1*0202*-TCG-*DPB1*0201*-AGTGCC	0.022	0.024	0.662	0.97 (0.57–1.65)
9	GG-*DPA1*0201*-TCG-*DPB1*0501*-GAGATT	0.029	0.018	0.057	1.81 (1.06–3.08)
10	GG-*DPA1*0201*-TCA-*DPB1*1301*-GGTGCC	0.022	0.016	0.172	1.69 (0.95–3.03)
11	AA-*DPA1*0103*-CTG-*DPB1*0301*-GGGGTC	0.011	0.016	0.246	0.74 (0.36–1.53)
12	GG-*DPA1*0201*-TCG-*DPB1*1401*-GGGGTC	0.012	0.012	0.877	1.25 (0.61–2.53)

Controls of the first stage and cases of the second stage were analyzed.
[a]Haplotypes consisting of rs2595309, rs3077, *HLA-DPA1*, rs2301220, rs9277341, rs3135021, *HLA-DPB1*, rs9277535, rs10484569, rs3128917, rs2281388, rs3117222 and rs9380343 are shown. [b]*P* values, odds ratios and its 95% confidence intervals of each haplotype were calculated as described in the Methods.

regression model. Thus, our findings clearly indicate that hepatitis B is associated with variants in the *HLA-DP* loci.

A number of reports have described association of several *HLA* and non-*HLA* genes with persistent HBV infection[12,15], but their results were not consistent among the studies, and none of them indicated a possible involvement of the *HLA-DP* locus. This study is the first GWAS to investigate host genetic factors associated with chronic hepatitis B. One genome-wide linkage analysis using 318 microsatellite markers in the Gambian population suggested that the chromosome 21q22 region contains a susceptibility locus for persistent HBV infection[16]. However, our GWAS analysis failed to support this result, possibly owing to ancestry differences or different modes of viral transmission (the vertical transmission in Japan versus the horizontal transmission in Gambia).

To investigate the correlation between the incidence of hepatitis B infection and these polymorphisms, we evaluated the frequencies of rs3077 and rs9277535 in 11 different HapMap3 populations (**Supplementary Table 6** online). Our association analysis indicated that A alleles at both rs3077 and rs9277535 were associated with protective effects for chronic hepatitis B. Notably, the frequencies of these two alleles were lower in Asian and African populations, especially in the Chinese population, compared with European and Central American populations. Although disease prevalence is not determined solely by genetic factors, the findings presented in our manuscript suggest that genetic factors might exert substantial influence on the prevalence of infectious disease.

Antigen presentations on HLA class II molecules to CD4-positive helper T cells and on class-I molecules to CD8-positive cytotoxic T cells are considered to be critical for the immune response against exposure to HBV. Although cytotoxic T cells are suspected to have major roles in viral clearance, helper T cells are also essential in the immune response to acute infections[17]. HLA-DPs have a structure similar to other classical HLA class II molecules, but their roles in the immune response have not been well characterized, except the association with berylliosis[18]. The 11 SNPs we found showing strong association with chronic HBV infection were in very strong LD with *HLA-DP* alleles. Because the subsequent haplotype analyses identified significant association of chronic hepatitis B with haplotypes containing the *HLA-DPA1* and *HLA-DPB1* genes, we suspected that variations in HLA-DP molecules would affect the ability for antigen presentation of HLA class II molecules on immune cells and result in weak (or no) immune response and persistent HBV infection. A previous report that implicated *HLA-DPA1**0103 and *DPB1**0402 to be candidate predictive factors for antibody production after HBV vaccination[19] supports this hypothesis. It should be noted that the lack of information regarding exposure to HBV for each control might underestimate the effect size obtained in this study but does not inflate the type 1 error rate.

In summary, we have demonstrated that genetic variants in the *HLA-DP* genes are strongly associated with chronic hepatitis B in the Asian population. Considering the function of HLA-DP molecules, our findings suggest that antigen presentation on HLA-DP molecules might be critical for virus elimination and have an important role in the pathogenesis of chronic hepatitis B. An understanding of the molecular mechanism by which *HLA-DP* variants confer risk of chronic hepatitis B should shed light on its pathogenesis and facilitate development of new therapies for treatment of the disease and prevention of disease progression.

METHODS

Samples. Characteristics of each cohort group are shown in **Supplementary Table 1**. Case and control samples used in this study for the Japanese population were obtained from the BioBank Japan at the Institute of Medical Science, the University of Tokyo[20], except case samples of the second replication and control samples of the first stage of the GWAS. From the registered samples in BioBank Japan, we selected individuals that were clinically diagnosed as having chronic hepatitis B. The diagnosis of chronic hepatitis B was conducted based on HBsAg-seropositivity and elevated serum aminotransferase levels for more than six months according to the guideline for diagnosis and treatment of chronic hepatitis (see URLs section below). The control groups consisted of 2,821 individuals that were registered in BioBank Japan as subjects with diseases other than chronic hepatitis B. Subjects who were positive for HBsAg were excluded from the controls. We obtained 934 Japanese control DNAs in the first stage from volunteers in the Osaka-Midosuji Rotary Club, Osaka, Japan. Case samples for the second replication cohort ($n = 718$, RIKEN) were collected at Toranomon Hospital as well as at hospitals participating in the Hiroshima Liver Study Group (for a list of doctors participating in this study group, see URLs section below). Cases and controls for the Thai replication study ($n = 308$ and 546, respectively) were collected at Ramathibodi Hospital, Mahidol University, Thailand. The diagnosis of chronic hepatitis B was based on HBsAg-seropositivity and elevated serum aminotransferase levels. All participants provided written informed consent. This research project was approved by the ethical committees at the Institute of Medical Science, the University of Tokyo, the Center for Genomic Medicine (formerly SNP Research Center), RIKEN and Ramathibodi Hospital, Mahidol University.

SNP genotyping. We applied the two-stage approach as described previously[21]. For the first stage, we genotyped 188 individuals with chronic hepatitis B and 934 controls using the Illumina HumanHap550v3 Genotyping BeadChip. After excluding nine cases with call rate of <0.98, we applied SNP quality control (call rate of ≥ 0.99 in both cases and controls and P value of Hardy-Weinberg equilibrium test of $\geq 1.0 \times 10^{-6}$ in controls): 499,544 SNPs on autosomal chromosomes passed the quality control filters and were further analyzed. Among the SNPs analyzed in the first stage, we selected the top 12,000 SNPs showing the smallest P values for the second stage. SNPs with minor allele frequency (MAF) of ≤ 0.1 in both case and control samples were excluded from the further analysis. In the second stage, we genotyped an additional panel of 616 cases using an

Affymetrix GeneChip Custom 10K array. After excluding nine cases with call rate of <0.95, all cluster plots were checked by visual inspection by trained staff, and SNPs with ambiguous calls were excluded. Ninety-four randomly selected case samples in the first stage were re-genotyped in the second stage, and SNPs with concordance rates of $<98\%$ between two assays (Illumina and Affymetrix) were excluded from the further analysis. We used genome-wide screening data of other diseases (uterine cervical cancer, esophageal cancer, hematological cancer, pulmonary tuberculosis, ovarian cancer, uterine body cancer and keroid) as controls for the second stage. All the samples were genotyped using the Illumina HumanHap550v3 Genotyping BeadChip, and the same quality-control filters as the first screening were applied. As a result, we analyzed 9,875 SNPs in 607 cases and 1,267 controls in the second stage and found 11 SNPs ($P < 5.06 \times 10^{-6}$) to be significantly associated with chronic hepatitis B after Bonferroni correction. These first and second stages are defined as the discovery phase of the research, and the following replication studies are defined as the replication phase. In the replication analyses, we used TaqMan genotyping system (Applied Biosystems) or the multiplex PCR-based Invader assay (Third Wave Technologies).

***HLA-DPA1* and *HLA-DPB1* genotyping.** We analyzed *HLA-DP* genotypes using 607 cases (in the second stage of GWAS) and 934 controls (in the first stage of GWAS). Exon 2 of the *HLA-DPA1* and *HLA-DPB1* genes were amplified and directly sequenced according to the protocol of International Histocompatibility Workshop Group[22]. *HLA-DPA1* and *DPB1* alleles were determined based on the alignment database of dbMHC.

Statistical analysis. In the first stage of the GWAS, Fisher's exact test was applied to a two-by-two contingency table in three genetic models: an allele frequency model, a dominant-effect model and a recessive-effect model. At the second stage of GWAS and replication analyses, statistical significance of the association with each SNP was assessed using a 1-degree-of-freedom Cochrane-Armitage trend test. Significance levels after Bonferroni correction for multiple testing were $P = 5.06 \times 10^{-6}$ (0.05/9,875) in the second stage and $P = 0.025$ (0.05/2) in replication analyses. Age- and sex-adjusted odds ratios were obtained by logistic regression analysis. Odds ratios and confidence intervals were calculated using the major allele as a reference. The meta-analysis was conducted using the Mantel-Haenszel method. Heterogeneity among studies was examined by using the Breslow-Day test. To assess the association of each *HLA* allele, we used Fisher's exact tests 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 odds ratios and 95% confidence intervals were calculated from coefficients of GLM model. Odds ratios of each haplotype were calculated relative to the second major haplotype in **Table 2**, because the most common haplotype was the disease-associated haplotype. All of these statistical values were calculated by function haplo.cc. We used Haploview software to analyze linkage disequilibrium values between *HLA-DR13* and SNPs.

Software. For general statistical analysis, we used R statistical environment version 2.6.1 or PLINK1.03 (ref. 23). To draw the LD map, we used Haploview software[24]. Estimation of haplotype frequencies and analysis of haplotype association were performed by R package haplo.stats[25]. Sequence variants in exon2 of *HLA-DPA1* and *HLA-DPB1* were analyzed by Polyphred.

URLs. The Japan Society of Hepatology, http://www.jsh.or.jp/medical/gudelines/index.html; Hiroshima Liver Study Group, http://home.hiroshima-u.ac.jp/naika1/hepatology/english/study.html; PLINK1.03, http://pngu.mgh.harvard.edu/~purcell/plink/; R package haplo.stats, http://mayoresearch.mayo.edu/mayo/research/schaid_lab/software.cfm; Polyphred, http://droog.gs.washington.edu/polyphred/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank K. Tokunaga for useful advice of on HLA-DP genotyping and interpretation, and technical staff of Laboratory for Genotyping Development at RIKEN for SNP genotyping at the first and second stages of the GWAS. We are also grateful to members of Hiroshima Liver Study Group and The Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan for supporting our study. This work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Sciences and Technology of the Japanese government.

AUTHOR CONTRIBUTIONS

Y.N. conceived the study; Y.N., Y.K., Y.D., M.K. and K.M. designed the study; Y.K., S.W., H.O. and N.H. performed genotyping; Y.K., T.T., M.K., N.K., Y.N. and K.M. wrote the manuscript; T.K., A.T., T.T. and N.K. performed data analysis at the genome-wide phase; Y.N., K.M. and M.K. managed DNA samples belong to BioBankJapan; K.C. and H.K. managed second replication samples; W.C., A.P. and T.S. managed third replication samples in Thailand; Y.K. summarized the whole results; Y.N. obtained funding for the study.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Pungpapong, S., Kim, W. & Poterucha, J. Natural history of hepatitis B virus infection: an update for clinicians. *Mayo Clin. Proc.* **82**, 967–975 (2007).
2. Custer, B. *et al.* Global epidemiology of hepatitis B virus. *J. Clin. Gastroenterol.* **38**, S158–S168 (2004).
3. Lai, C.L., Ratziu, V., Yuen, M.F. & Poynard, T. Viral hepatitis B. *Lancet* **362**, 2089–2094 (2003).
4. Kobayashi, M. *et al.* Viral genotypes and response to interferon in patients with acute prolonged hepatitis B virus infection of adulthood in Japan. *J. Med. Virol.* **68**, 522–528 (2002).
5. Laskus, T. *et al.* Prevalence of markers of hepatitis viruses in out-patient alcoholics. *J. Hepatol.* **15**, 174–178 (1992).
6. Sheen, I., Liaw, Y., Lin, D. & Chu, C. Role of hepatitis C and delta viruses in the termination of chronic hepatitis B surface antigen carrier state: a multivariate analysis in a longitudinal follow-up study. *J. Infect. Dis.* **170**, 358–361 (1994).
7. Lin, T. *et al.* Hepatitis B virus markers in Chinese twins. *Anticancer Res.* **9**, 737–741 (1989).
8. Ben-Ari, Z. *et al.* Cytokine gene polymorphisms in patients infected with hepatitis B virus. *Am. J. Gastroenterol.* **98**, 144–150 (2003).
9. Thursz, M.R. *et al.* Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. *N. Engl. J. Med.* **332**, 1065–1069 (1995).
10. Bellamy, R. *et al.* Tuberculosis and chronic hepatitis B virus infection in Africans and variation in the vitamin D receptor gene. *J. Infect. Dis.* **179**, 721–724 (1999).
11. Deng, G. *et al.* Association of estrogen receptor polymorphisms with susceptibility to chronic hepatitis B virus infection. *Hepatology* **40**, 318–326 (2004).
12. Singh, R., Kaul, R., Kaul, A. & Khan, K. A comparative review of HLA associations with hepatitis B and C viral infections across global populations. *World J. Gastroenterol.* **13**, 1770–1787 (2007).
13. Hohler, T. *et al.* HLA-DRB1* 1301 and* 1302 protect against chronic hepatitis B. *J. Hepatol.* **26**, 503–507 (1997).
14. Ahn, S.H. *et al.* Association between hepatitis B virus infection and HLA-DR type in Korea. *Hepatology* **31**, 1371–1373 (2000).
15. Frodsham, A. Host genetics and the outcome of hepatitis B viral infection. *Transpl. Immunol.* **14**, 183–186 (2005).
16. Frodsham, A. *et al.* Class II cytokine receptor gene cluster is a major locus for hepatitis B persistence. *Proc. Natl. Acad. Sci. USA* **103**, 9148–9153 (2006).
17. Jung, M. *et al.* Activation of a heterogeneous hepatitis B (HB) core and e antigen-specific CD4+ T-cell population during seroconversion to anti-HBe and anti-HBs in hepatitis B virus infection. *J. Virol.* **69**, 3358–3368 (1995).
18. Fontenot, A., Torres, M., Marshall, W., Newman, L. & Kotzin, B. Beryllium presentation to CD4+ T cells underlies disease-susceptibility HLA-DP alleles in chronic beryllium disease. *Proc. Natl. Acad. Sci. USA* **97**, 12717–12722 (2000).
19. Mineta, M. *et al.* Contribution of HLA class I and class II alleles to the regulation of antibody production to hepatitis B surface antigen in humans. *Int. Immunol.* **8**, 525–531 (1996).
20. Nakamura, Y. The BioBank Japan project. *Clin. Adv. Hematol. Oncol.* **5**, 696–697 (2007).
21. Saito, A. & Kamatani, N. Strategies for genome-wide association studies: optimization of study designs by the stepwise focusing method. *J. Hum. Genet.* **47**, 360–365 (2002).
22. van der Zwan, A., Griffith, B., Rozemuller, E., Williams, T. & Tilanus, M.G.J. *IHWG Technical Manual Genomic Analysis of the Human MHC: DNA-Based Typing for HLA Alleles and Linked Polymorphisms* (ed. Tilanus, M.G.J.) (Seattle, Washington, International Histocompatibility Working Group, 2002).
23. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
24. Barrett, J., Fry, B., Maller, J. & Daly, M. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
25. Schaid, D.J., Rowland, C.M., Tines, D.E., Jacobson, R.M. & Poland, G.A. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am. J. Hum. Genet.* **70**, 425–434 (2002).