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^{d}_{het}
					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]	$P_{adjusted\ for\ rs3077}$	OR (95% CI)	$P_{adjusted\ for\ rs9277535}$	OR (95% CI)	$P_{adjusted\ for\ rs2856718}$	OR (95% CI)	$P_{adjusted\ for\ rs7453920}$	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 ($P_{interaction} > 0.10$). 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

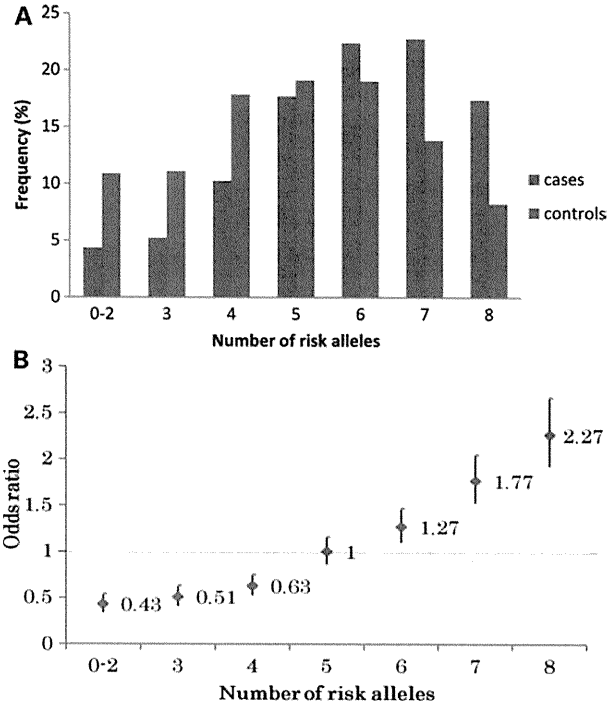

Figure 4. Cumulative effects of CHB risk alleles. (**A**) Distribution of risk alleles in CHB cases (red bars) and controls (blue bars). (**B**) Plot of the increasing OR for CHB with increasing number of risk alleles. The OR are relative to the median number of four risk alleles (rs3077, rs9277535, 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

≥ 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 ≤ 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.

Associations of *HLA-DP* Variants with Hepatitis B Virus Infection in Southern and Northern Han Chinese Populations: A Multicenter Case-Control Study

Jin Li[1], Daguo Yang[2], Yongwen He[3], Mengyi Wang[4], Zirong Wen[5], Lifeng Liu[1], Jinjian Yao[1], Koichi Matsuda[6], Yusuke Nakamura[6], Jinling Yu[1], Xiaorui Jiang[1], Shuzhen Sun[1], Qing Liu[1], Xiang Jiang[1], Qilong Song[1], Man Chen[1], Hong Yang[1], Feng Tang[1], Xiaowen Hu[1], Jing Wang[1], Ying Chang[1], Xingxing He[1], Yuan Chen[4]*, Jusheng Lin[1]*

1 Institute of Liver Disease, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei Province, China, **2** The Third People's Hospital of Shenzhen, Shenzhen, Guangdong Province, China, **3** Department of Infection Disease, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei Province, China, **4** Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei Province, China, **5** Qingdao Infectious Disease Hospital, Qingdao, Shandong Province, China, **6** Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Abstract

Background: Human leukocyte antigen DP (*HLA-DP*) locus has been reported to be associated with hepatitis B virus (HBV) infection in populations of Japan and Thailand. We aimed to examine whether the association can be replicated in Han Chinese populations.

Methodology/Principal Findings: Two *HLA-DP* variants rs2395309 and rs9277535 (the most strongly associated SNPs from each *HLA-DP* locus) were genotyped in three independent Han cohorts consisting of 2 805 cases and 1 796 controls. By using logistic regression analysis, these two SNPs in the *HLA-DPA1* and *HLA-DPB1* genes were significantly associated with HBV infection in Han Chinese populations ($P=0.021\sim3.36\times10^{-8}$ at rs2395309; $P=8.37\times10^{-3}\sim2.68\times10^{-10}$ at rs9277535). In addition, the genotype distributions of both sites (rs2395309 and rs9277535) were clearly different between southern and northern Chinese population ($P=8.95\times10^{-5}$ at rs2395309; $P=1.64\times10^{-9}$ at rs9277535). By using asymptomatic HBV carrier as control group, our study showed that there were no associations of two *HLA-DP* variants with HBV progression ($P=0.305\sim0.822$ and $0.163\sim0.881$ in southern Chinese population, respectively; $P=0.097\sim0.697$ and $0.198\sim0.615$ in northern Chinese population, respectively).

Conclusions: Our results confirmed that two SNPs (rs2395309 and rs9277535) in the *HLA-DP* loci were strongly associated with HBV infection in southern and northern Han Chinese populations, but not with HBV progression.

Citation: Li J, Yang D, He Y, Wang M, Wen Z, et al. (2011) Associations of *HLA-DP* Variants with Hepatitis B Virus Infection in Southern and Northern Han Chinese Populations: A Multicenter Case-Control Study. PLoS ONE 6(8): e24221. doi:10.1371/journal.pone.0024221

Editor: Sunil K. Ahuja, South Texas Veterans Health Care System, United States of America

Received January 27, 2011; **Accepted** August 5, 2011; **Published** August 31, 2011

Funding: This study was financially supported by the National Basic Research Program of China (973 Program, No. 2007CB512903) and the National Natural Science Foundation of China (No. 30872237). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jslin2010@Hotmail.com (J.Lin); chenyuan008@163.com (Y.Chen)

Introduction

More than 2 billion people have been infected with the hepatitis B virus (HBV) worldwide, of which 350 million are chronic carriers and about 600 000 die annually of HBV-related acute or chronic liver disease [1]. Although many individuals eventually achieve a state of nonreplicative infection, the prolonged immunologic response to infection leads to the development of cirrhosis, liver failure, or hepatocellular carcinoma (HCC) in up to 40% of patients [2]. In China, where HBV infection is endemic, there are estimated 93 million HBV carriers, and among them 30 million are patients with chronic hepatitis B [3]. Multiple causes influence the risk of chronic HBV infection in china, for example, age, gender, viral genotype, ethnicity, variation in genes of the immune system and so on [4].

Several polymorphisms of the *HLA* loci have been reported for hepatitis B virus infection [5,6]. A study in Gambian found that the allele DRB1*1302 was associated with the clearance of the virus [7]. Hepatitis B virus persistence and disease chronicity were associated with *HLA-DQA*1*0501 and *HLA-DQB*1*0301 in Chinese [8] and with *HLA-DR*9 in Koreans [9]. Although the association between common diseases and these *HLA* (or non-*HLA*) genes has become increasingly evident [10], their results are conflicting among the studies, and have not been confirmed by other investigators [11].

A recent study found that the *HLA-DP* locus was associated with chronic hepatitis B in Japanese and Thais [12]. As the frequencies of these *HLA-DP* alleles in Chinese populations were similar to those in Japanese populations, it would be necessary to confirm whether there was the association between the *HLA-DP* genetic

variation and HBV infection in Chinese populations. To this end, we selected the most strongly associated SNPs (the previous GWAS results) from each *HLA-DP* locus (rs9277535 at the *HLA-DPB1* and rs2395309 at the *HLA-DPA1*, respectively) and genotyped these two polymorphisms in a population-based case-control study of Chinese Hans, including 2 805 cases and 1 796 controls from Hubei province (Central China), Shandong province (North China) and Guangdong province (South China).

Materials and Methods

Ethic statement

The study was approved by the local research ethics committee (REC) at the Tongji Hospital of Huazhong University of Science and Technology in accordance with the principle of the Helsinki Declaration II. All written informed consent documents from each participant were obtained during the enrollment phase.

Study subjects

A total of 4 601 unrelated Han Chinese were recruited in this study between September 2007 and June 2011. All subjects were divided into six groups: a) HBV clearance group(Clear); b) Healthy control group(Health); c) Persistent asymptomatic HBV carriers group(AsC); d) Chronic active hepatitis B group(CHB); e) HBV-related liver cirrhosis group(LC); and f) HBV-related heptocellular carcinoma group(HCC). The diagnostic criteria for study inclusion were listed in Table S1, which had been described in the previous publication [13,14]. All individuals were gathered from three Han Chinese cohorts. First, we recruited 2 280 subjects from Tongji hostital and Union hospital in Wuhan, Hubei province. Second, we gathered additional 1 304 subjects from The Affiliated Hospital of Binzhou Medical College and Qingdao Infectious Disease Hospital in Shandong province, 1 017 subjects from Shenzhen Third People's Hospital, Shenzhen Fourth People's Hospital and Shenzhen Sixth People's Hospital in Guangdong province.

A uniform questionnaire was used at three enrollment sites and recorded self-report of risk factors for HBV transmission, family history of HBV infection, past and current smoking, alcohol ingestion, etc. The demographic information included gender, birth-date, birthplace, and past and current residency.

DNA Isolation and Genotyping

Genomic DNA was isolated from peripheral whole blood using TIANamp blood DNA kit (Tiangen Biotech [Beijing] Co., Ltd., China). The concentration and purity of the DNA were determined with a NanoDrop spectrophotometer and diluted to a final concentration of 8 ng/μL. The genotyping of genetic polymorphisms was performed via the TaqMan method according to the protocol of TaqMan® SNP Genotyping Assays (Applied Biosystems, California, USA). Allelic category was measured automatically using the Sequence Detection System 2.3 software (Applied Biosystems) according to the intensity of VIC and FAM dye. To detect these SNPs (rs2395309 and rs9277535), we customized the TaqMan® MGB Probe as well as the primers for PCR amplification (Table S2.).

Statistical analysis

Statistical analysis was conducted by using haploview 4.2, Arlequin 3.5, Stata10.0 and SPSS 17.0 softwares. Linkage disequilibrium was assessed by the haploview 4.2 softwares using frequencies obtained from the Health group. The (Bayesian) ELB algorithm was used to infer haplotypes by using Arlequin 3.5. The Hard-Weinberg equilibrium of alleles and population pairwise comparisons were also evaluated by using Arlequin 3.5 [15]. A meta-analysis of all studies was performed for each SNP associated with chronic hepatitis B by using Stata10.0 softwares. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated on the basis of the binary logistic regression analysis (adjustment for gender and age). The strength of association between the genotypes or alleles and HBV infection was estimated by using SPSS 17.0 softwares. A best-fit model was constructed by means of comparisons with other models. Values of $P<0.05$ were considered statistically significant.

Results

Hardy-Weinberg equilibrium test

Hardy-Weinberg equilibrium was estimated by Fisher's exact test using Arlequin 3.5 software. There was no significant difference between observed and expected frequencies of each genotype in these involved populations ($P>0.05$). This result indicated that these populations had a relatively stable genetic background and were suitable for further genetic statistical analysis.

The clinic and demographic characteristics

The clinical and demographic characteristics of the case-control study were summarized in Table 1, including gender, age, drinkers, serum total bilirubin level (T-Bil), HBV-DNA load, alanine transaminase (ALT) and serum markers of hepatitis B virus. There was no significant difference in the percentage of hepatitis B e antigen (HBeAg) positive ($P=0.10$) between asymptomatic HBV carriers (17.1%) and the patients of chronic hepatitis B group (20.2%). In addition, there was more alcohol consumption in patients ($P<0.05$) with HBV-related liver cirrhosis group (24.3%) and HBV-related heptocellular carcinoma group (30.5%) than those in HBV clearance group (10.2%) and healthy control group (8.9%). The difference in the alcohol consumption status was due to few drinkers in Chinese female population. Although an effort was made to obtain a good match on age and sex, there were more men in four case groups (averaged 73.5%) than those in HBV clearance group (51.6%, $P<0.05$) and healthy control group (47.5%, $P<0.05$).

Population pairwise comparisons and grouping of subjects

To explore whether differences in susceptibility loci were caused by the disease or by genetic background between populations, we first needed to determine which populations should be compared with each other, and whether there were populations that could be lumped together to simplify statistical analysis. To this end, we performed population pairwise comparisons $\mathbf{F_{ST}}$ testing between each population using Arlequin 3.5 software. The principle of population pairwise comparisons states that: if there is no difference in heredity between two populations, the data permuting of genotypes or haplotypes between two populations should not cause a significant difference, which can be evaluated by $\mathbf{F_{ST}}$ P value ($P>0.05$). According to the results shown in Table 2, we could infer that Shandong population had greater difference than Hubei population or Guangdong population in genetic background ($P<0.0001$). As the pairwise comparisons for Hubei population and Guangdong population were not significantly different ($P=0.191$), and both of them were the same geographic position (southern of china) [16], we had determined to merge Hubei population and Guangdong population into southern Chinese population. Meanwhile, Shandong population was taken as northern Chinese population. Furthermore, in order

Table 1. Clinical characteristics of study subjects.

Characteristics	Health(n = 962)	Clear(n = 834)	AsC(n = 910)	CHB(n = 964)	LC(n = 544)	HCC(n = 387)
Gender, no. (%)						
Male	454(47.5)	421(51.6)	519(58.4)	765(81.0)	435(80.3)	342(88.8)
Female	501(52.5)	395(48.4)	370(41.6)	180(19.0)	107(19.7)	43(11.2)
Age (years), mean(sd)	51.70±10.96	51.48±11.04	45.64±10.93	38.45±10.73	48.30±11.02	49.82±10.01
Drinkers, no. (%)	86(8.9)	85(10.2)	102(11.2)	179(18.6)	132(24.3)	118(30.5)
HBsAg	All−	All−	All+	All+	All+	All+
Anti-HBs IgG	All−	All+	All−	All−	All−	All−
HBeAg-positive,no. (%)	All−	All−	156(17.1)	195(20.2)	64(11.8)	39(10.1)
Anti-HBc IgG, no. (%)	All−	All+	All+	All+	All+	All+
Family history, no. (%)	No	No	36(4.0)	142(14.7)	69(12.7)	77(19.9)
ALT(U/L)	No	No	19.74±15.09	447.53±402.71	101.57±119.90	83.64±86.17
TBil (μmol/L)	No	No	11.71±9.01	161.55±171.06	124.19±125.86	69.24±65.74
HBV-DNA (copy/ml)	No	No	No	3.23E7±4.76E7	1.38E6±5.16E6	3.72E6±1.13E7

'No' means non-detected.
'Drinkers' was defined as alcohol consumption of >40 g/week, which included occasional drinkers and daily drinkers.
Abbreviations: Clear, HBV clearance group; Health, Healthy control group; AsC, Asymptomatic HBV carriers group; CHB, Chronic active hepatitis B group; LC,HBV-related liver cirrhosis group; HCC, HBV-related heptocellular carcinoma group;
doi:10.1371/journal.pone.0024221.t001

to identify whether the two polymorphisms were associated with HBV infection or clearance, we combined all the types of HBV infection populations into one group by using the healthy group or clearance group as the reference.

Logistic regression analysis of the *HLA-DP* loci polymorphisms

Then, to investigate which genotypic models were significantly associated with the various outcomes, we conducted comparisons of four models (Multiplicative model, Additive model, Dominant model and Recessive model) in southern and northern Chinese populations respectively (data not show). For the four models, the best-fit genotypic effect of these two SNPs (rs2395309 and rs9277535) was observed in the dominant model which was the protective genotype AA and AG (see Table 3). After compared with the Healthy control group, both single nucleotide polymorphism (SNP) sites (rs2395309 and rs9277535) showed associations with HBV infection in southern Chinese population (Odds ratio [OR] = 0.57; 95% Confidence intervals [CI] :0.47, 0.70; $P = 3.36\times10^{-8}$ at rs2395309; OR = 0.52; 95% CI :0.43, 0.64; $P = 2.68\times10^{-10}$ at rs9277535), as well as in northern Chinese population (OR = 0.50; 95% CI :0.35, 0.71; $P = 1.23\times10^{-4}$ at rs2395309; OR = 0.50; 95% CI :0.36, 0.68; $P = 1.74\times10^{-5}$ at rs9277535). And, interestingly, *HLA-DP* rs2395309 and rs9277535 sites also showed a strong protective effect for HBV clearance not only in southern Chinese population (OR = 1.31; 95% CI :1.17, 1.45; $P = 9.63\times10^{-7}$ at rs2395309; OR = 1.33; 95% CI :1.20, 1.49; $P = 1.67\times10^{-7}$ at rs9277535) but also in northern Chinese population (OR = 1.20; 95% CI :1.03, 1.40; $P = 0.021$ at rs2395309; OR = 1.26; 95% CI :1.06, 1.49; $P = 8.37\times10^{-3}$ at rs9277535). As shown in Table 3, notably, the genotype distributions of both sites (rs2395309 and rs9277535) were clearly different between southern and northern Healthy populations (P values = 8.95×10^{-5} and 1.64×10^{-9}, respectively. P values of Pearson's x^2 test for allele model). The two minor-allele frequencies (MAF) in both Healthy populations (southern and northern Han Chinese) were 30.1% vs 38.8% at rs2395309, 38.1% vs 52.2% at rs9277535. In addition, to decrease the bias of sex and age in population sampling, we further conducted the stratified analysis for sex and age. As presented in Table S3, male and female patients showed different associations with HBV diseases in these two SNPs (rs2395309 and rs9277535). Specially, in the northern Chinese population, this difference was notable between male patients and female patients. Furthermore, in the stratified analysis of age, most cases were no significant differences in genotype distributions of two SNPs sites between patients with age≤45 years and patients with age>45 years (Table S4.).

Table 2. Matrix of significant F_{ST} P values among populations.

Populations	Hubei	Guangdong	Shandong
Hubei	*		
Guangdong	0.191	*	
Shandong	**<0.0001**	**<0.0001**	*

Population pairwise comparisons F_{ST} tests were performed between pairs of groups using Arlequin 3.5software. Statistically significant values are in shown in bold.
Abbreviations: Hubei, Hubei populations; Guangdong, Guangdong populations; Shandong, Shandong populations.
doi:10.1371/journal.pone.0024221.t002

Associations of the *HLA-DP* loci polymorphisms with HBV progression

Considering the function of *HLA-DP* molecules, we were interested in the possible association between the polymorphisms in *HLA-DP* gene and the disease progression of chronic hepatitis B. To test our prediction, we further analysed the difference in two SNPs genotype distributions by using asymptomatic HBV carrier as control group. Unfortunately, there were not associations in chronic active hepatitis B group (OR = 1.03; 95% CI : 0.79, 1.34; $P = 0.822$ at rs2395309; OR = 0.92; 95% CI : 0.71, 1.18; $P = 0.501$ at rs9277535, in southern Chinese population; OR = 0.92; 95% CI : 0.62, 1.38; $P = 0.697$ at rs2395309; OR = 1.33; 95% CI : 0.86, 2.06; $P = 0.198$ at rs9277535, in northern Chinese population),

Table 3. Associations of two SNPs (rs2395309, rs9277535) with HBV infection and clearance in Han Chinese populations.

	South of china		North of china	
	Control group	Case group	Control group	Case group
HLA-DPB1 (rs2395309)- dominant model (AA+AGvsGG)				
AA/AG/GG	57/234/288†	112/709/1367‡	52/193/138†	63/249/302‡
P value OR (95%CI)	Reference	3.36×10^{-8} 0.57 (0.47,0.70)	Reference	1.23×10^{-4} 0.50 (0.35,0.71)
AA/AG/GG	112/709/1367‡	35/235/257‖	63/249/302‡	56/130/121‖
P value OR (95%CI)	Reference	9.63×10^{-7} 1.31 (1.17,1.45)	Reference	0.021 1.20 (1.03,1.40)
HLA-DPA1 (rs9277535)- dominant model (AA+AGvsGG)				
AA/AG/GG	80/277/216†	177/830/1195‡	97/203/80†	118/287/206‡
P value OR (95%CI)	Reference	2.68×10^{-10} 0.52 (0.43,0.64)	Reference	1.74×10^{-5} 0.50 (0.36,0.68)
AA/AG/GG	177/830/1195‡	67/251/208‖	118/287/206‡	67/165/75‖
P value OR (95%CI)	Reference	1.67×10^{-7} 1.33 (1.20,1.49)	Reference	8.37×10^{-3} 1.26 (1.06,1.49

†Healthy control group.
‖HBV clearance group.
‡HBV infection groups, including Asymptomatic HBV carriers, Chronic active hepatitis B group, HBV-related liver cirrhosis group, HBV-related heptocellular carcinoma group.
The *P* values, odds ratios (OR), and 95% confidence intervals (CI) were calculated on the basis of the binary logistic regression analysis, adjusted for sex and age.
doi:10.1371/journal.pone.0024221.t003

HBV-related liver cirrhosis group (OR = 1.11; 95% CI : 0.82, 1.52; $P=0.499$ at rs2395309; OR = 1.24; 95% CI : 0.92, 1.67; $P=0.163$ at rs9277535, in southern Chinese population; OR = 0.74; 95% CI : 0.48, 1.16; $P=0.189$ at rs2395309; OR = 1.29; 95% CI : 0.81, 2.06; $P=0.286$ at rs9277535, in northern Chinese population) and HBV-related heptocellular carcinoma group(OR = 0.85; 95% CI : 0.63, 1.16; $P=0.305$ at rs2395309; OR = 0.98; 95% CI : 0.73, 1.31; $P=0.881$ at

Table 4. Results of the association test for two SNPs(rs2395309,rs9277535) haplotypes in Han Chinese populations.

South of china						
Haplotype	**Health(2n = 1106)**	**Clear(2n = 1048)**	**AsC(2n = 1342)**	**CHB(2n = 1486)**	**LC(2n = 754)**	**HCC(2n = 632)**
A-A	247(22.3)	205(17.0)	227(16.8)	238(16.0)	122(16.2)	92(14.5)
A-G	88(8.0)	97(11.8)	62(4.6)	95(6.4)	36(4.8)	23(3.7)
G-A	173(15.6)	177(19.5)	140(10.4)	143(9.6)	95(12.6)	78(12.4)
G-G	598(54.1)	569(51.7)	913(68.1)	1010(68.0)	501(66.4)	439(69.4)
P value‖	Reference		1.47×10^{-6}	6.47×10^{-8}	2.53×10^{-5}	6.07×10^{-7}
OR (95%CI)			0.60 (0.49,0.74)	0.57 (0.47,0.70)	0.59 (0.46,0.76)	0.51 (0.39,0.66)
P value¶		Reference	7.35×10^{-4}	8.92×10^{-5}	2.45×10^{-3}	1.06×10^{-4}
OR (95%CI)			1.45 (1.17,1.80)	1.53 (1.24,1.89)	1.48 (1.15,1.91)	1.72 (1.31,2.27)
North of china						
Haplotype	**Health(2n = 734)**	**Clear(2n = 608)**	**AsC (2n = 422)**	**CHB (2n = 378)**	**LC(2n = 300)**	**HCC(2n = 100)**
A-A	226(30.8)	200(31.5)	118(28.0)	103(27.3)	71(23.6)	21(20.5)
A-G	55(7.5)	41(8.1)	27(6.4)	18(4.8)	16(5.4)	4(3.5)
G-A	157(21.4)	98(17.5)	66(15.6)	60(15.9)	62(20.7)	15(15.4)
G-G	296(40.3)	269(42.9)	211(50.0)	197(52.1)	151(50.3)	60(60.5)
P value‖	Reference		0.032	0.012	0.004	0.003
OR (95%CI)			0.73 (0.55,0.97)	0.68 (0.51,0.92)	0.62 (0.44,0.86)	0.46 (0.27,0.78)
P value¶		Reference	0.054	0.021	0.007	0.005
OR (95%CI)			1.33 (0.99,1.78)	1.42 (1.05,1.92)	1.58 (1.13,2.21)	2.12 (1.25,3.61)

‖Two SNPs haplotypes G-G, A-A in Health group compared with those in HBV infection groups.
¶Two SNPs haplotypes G-G, A-A in HBV infection groups compared with those in Clearance group.
The *P* values, odds ratios (OR), and 95% confidence intervals (CI) were calculated by Pearson Chi-Square test.
Abbreviations: Clear, HBV clearance group; Health, Healthy control group; AsC, Asymptomatic HBV carriers group; CHB, Chronic active hepatitis B group; LC,HBV-related liver cirrhosis group; HCC, HBV-related heptocellular carcinoma group; OR, odds ratio; CI, confidence interval.
doi:10.1371/journal.pone.0024221.t004

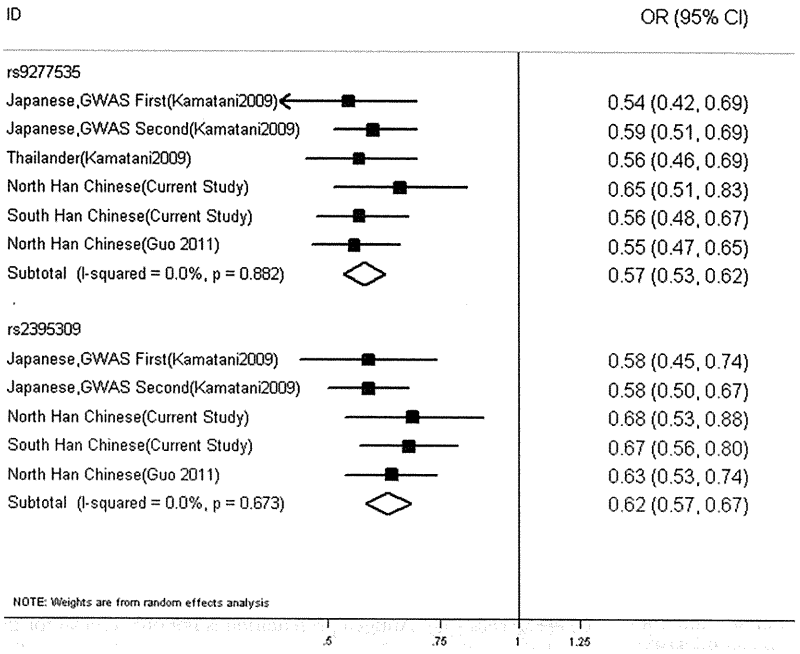

Figure 1. Meta-analasis of the rs9277535 and rs2395309. The meta-analysis combined with the results of previous studies, including more than 2,243 cases and 4,137 controls. Each effect size is shown with its confidence interval. Abbreviations: p, *P* heterogeneity value; OR, odds ratios; 95%CI, 95% confidence interval.
doi:10.1371/journal.pone.0024221.g001

rs9277535, in southern Chinese population; OR = 0.56; 95% CI : 0.28, 1.11; $P = 0.097$ at rs2395309; OR = 0.84; 95% CI : 0.42, 1.68; $P = 0.615$ at rs9277535, in northern Chinese population), compared with asymptomatic HBV group(Table S5.).

Associations of the *HLA-DP* loci polymorphisms with clinical factors

In order to analyze the associations between two SNPs and clinical factors (HBV-DNA load, ALT and TB), we used the independent-sample Kolmogorov-Smirnov t test in CHB group, LC group and HCC group. Although the GG patients have a higher mean on the HBV-DNA load, no significant difference was found between patients of different genotypes (see Fig. S1). In the analysis of ALT, the associations between two SNPs and the ALT level only be found in HBV-related liver cirrhosis group ($P = 0.002$ at rs2395309; $P = 0.009$ at rs9277535), rather than in other groups. Meanwhile, for the associations of the TB level, there was no difference between GG patients and AG+AA patients ($P > 0.05$ in each group).

Results of the Haplotype analysis and Meta-analysis

To further understand the contributions of these loci to HBV susceptibility, two-locus haplotypes were constructed for two SNPs rs2395309 and rs9277535 (Table 4.). Pairwise linkage disequilibrium (LD) analyses performed using all individuals from the health group showed that rs2395309 and rs9277535 SNPs were in LD with each other ($D' = 0.57$, $r^2 = 0.23$ in southern Chinese population; $D' = 0.58$, $r^2 = 0.20$ in northern Chinese population). In trying to derive HBV infection-specific haplotypes, the haplotype frequencies of two SNPs (rs2395309 and rs9277535) were evaluated in both Chinese populations. Four haplotypes were observed, and among them three haplotypes had frequencies more than 5% (Table 4.). Compared with protective A-A haplotype homozygotes, only G-G haplotype homozygotes had a significant increased risk for HBV infection (*P* value and odds ratios were shown in Table 4). Then, we summarized a meta-analysis combined with the results of related studies [12,17], including more than 2,243 cases and 4,137 controls. As shown in Figure 1 and Table S6, these odds ratios were quite similar among the three ethnic groups (Japanese, Thais and Chinese) and no heterogeneity was observed (P het = 0.673 at rs2395309; P het = 0.882 at rs 9577535).

Discussion

In this analysis, we confirmed that two SNPs sites (rs2395309 and rs9277535) in the *HLA-DPA1* and *HLA-DPB1* genes were significantly associated with HBV infection in southern and northern Han Chinese populations. Again, our haplotype analysis showed the frequency of G-G haplotype had a significant increase in the HBV infected populations, as compared with the healthy control group or HBV clearance group. As a result, we inferred that these persons with G-G haplotype have a higher risk of HBV infection than those persons with A-A haplotype. Meanwhile, the A-A haplotype could be strongly predictive for HBV clearance in HBV infection populations. Although our manuscript suggested that the genotype distributions of both sites (rs2395309 and rs9277535) were different between southern and northern Chinese population, the frequencies of two protective alleles A in Chinese populations were also similar to those in Asian populations, compared with European and Central American populations (data from public databases, HapMap). The results of the genetic association in our study were consistent with the previous study [12]. Hence, we could confirm that the polymorphisms of *HLA-DPA1* and *HLA-DPB1* gene play a very important role in chronic hepatitis B virus infection in southern and northern Han Chinese populations.

It has been well documented that men are more likely than women to be infected with HBV and develop liver cirrhosis and hepatocellular carcinoma [18,19]. The reasons for the gender distinction between HBV populations and health populations are complex, including occupation, alcohol drinking, tobacco smoking, family history of HBV infection and so on. Some previous reports suggested that sex hormones might interact with HBV in the infection process and lead to a dominant sex disparity in HBV populations. Naugler et al. [20] found that estrogen-mediated inhibition of interleukin-6 production by Kupffer cells reduced the risk of liver cancer in females. Wang et al. [21] study demonstrated that the androgen pathway could increase the transcription of HBV through direct binding to the androgen-responsive element sites in viral enhancer. Consequently, to decrease the bias of sex in population sampling, we further conducted the stratified analysis for sex. Although we found that male and female northern Chinese showed a different susceptibility to HBV infection, it only had 25% and 21% statistical power to detect these ORs of 0.73 and 0.74, which may lead to the false-negative results of rs2395309 and rs9277535 in northern female Chinese. The small sample for female HBV patients in this study might be the major reason for the non-significant associations in female Chinese. Hence, we only concluded that the genetic variants of *HLA-DPA1* and *HLA-DPB1* loci differ slightly between male and female Chinese, and the reasons why there is different between male and female for HBV infection need to be further studied.

And indeed, by consulting previous studies [22,23], we found that there are different distributions in some *HLA* alleles among Han Chinese populations. For instance, *HLA-DRB1**0301 [8], a risk-allele with respect to chronic HBV infection in Han Chinese, markedly has higher frequency in southern Han Chinese population than those in northern Han Chinese population. Since the frequency distribution of *HLA-DP* alleles were barely reported in China, it could be inferred only indirectly that there were also different distributions at *HLA-DP* alleles between two Han Chinese populations. And, it was the difference that led to the distinct distributions of both SNPs (rs9277535 and rs2395309) between southern and northern Han Chinese population. Nevertheless, this explanations why the distributions of the *HLA* alleles (or SNPs) differed between Han Chinese populations were complicated, such as evolution and migration history of the Chinese population [24,25,26], MHC-based mate choice [27], pathogen-driven selection at *HLA* alleles [28,29] and so on. Taking into account the different distributions of HBV genotypes [30] and HBV carrier rate [31] in China, as well as recent studies [12,17] and our results, we deduced that the mechanism of pathogen-driven selection (HBV and/or other pathogens) might be the leading cause of the different distributions at *HLA-DP* alleles between two Han Chinese populations.

Moreover, after infection with hepatitis B virus (HBV), the host's inflammatory immune response induces hepatocellular damage and is followed by the pathogenesis of liver cirrhosis and cancer [32]. Liver cancer arises most frequently in the setting of chronic liver inflammation [33]. Considering the function of *HLA-DP* molecules, HBV antigen presentation on *HLA-DP* molecules may be critical for virus elimination and has an important role in the progression of hepatitis B [34]. Therefore, we further analysed the possible association between the polymorphisms in *HLA-DP* gene and the disease progression of chronic hepatitis B. Unfortunately, compared with asymptomatic HBV carrier, there were no associations in chronic active hepatitis B group, HBV-related liver cirrhosis group and HBV-related hepatocellular carcinoma group. Although chronic HBV infection is the most important cause of HCC worldwide and contributes to at least 70% of cases of HCC in Asian-Africa [35], only a tiny fraction of chronic HBV carriers develop HCC in their lifetime [36]. It is suggested that the risk of HCC is caused by a complex interplay between multiple genetic and environmental factors. Recently, Zhang et al. have conducted the first liver GWAS for HCC in Chinese ancestry and identified a single susceptibility locus in the UBE4B-KIF1B-PGD region on 1p36.22 [37]. Since the region involve in these aspects of vesicles transport, cell apoptosis, DNA repair, and other intracellular pathways, it seems likely that different genes play disparate roles in HBV infection and HBV progression. For example, immune pathway (*HLA-DP* or other genes) is the primary cause of HBV infection, but intracellular pathway (Ubiquitin or other pathways) is the major reason of HBV progression. Thus, by combining our results with the aforementioned discussion, we inferred that the polymorphisms in *HLA-DPA1* and *HLA-DPB1* gene influence the infection of HBV in Chinese populations, rather than the progression of HBV disease.

Since the early 1970s [38], classical human leukocyte antigen loci have stood out as the leading candidates for infectious disease susceptibility. The classical *HLA* loci are the class I (*HLA-A*, *-B*, *-C*, *-E*, *-F*, and *-G*) and class II (*HLA-DR*, *-DQ*, *-DM*, and *-DP*) molecules. *HLA* class II molecules are the central part in the immune system by presenting peptides to the antigen receptor of CD4+ T cells [39]. Antigen presentation is not only crucial for the regulation of protective immune responses against invading pathogens, but also necessary for the maintenance of self-tolerance. It is therefore perhaps not surprising to find that the human MHC class II gene region holds the largest number, and some of the longest recognised, associations with a autoimmune, inflammatory and infectious diseases [40,41]. Although *HLA-DPs* have a structure similar to other classical *HLA* class II molecules, *HLA-DP* molecule roles in the immune response have not been well characterized until now. In a previous study, Hirayama et al. [42] indicated that the *HLA* class II genes for the *HLA-DR-DQ* alleles were associated with protection against early changes in liver fibrosis, whereas *HLA-DP* alleles were associated with protection from the late phase of schistosomal hepatic fibrosis. Owing to lack of replication of the previously report, more studies are essential to provide conclusive genetic and functional evidence to support a role for *HLA-DP* in HBV disease susceptibility.

In summary, in this multicenter case-control study, we have confirmed that the G alleles of two SNPs sties in the *HLA-DPA1* and *HLA-DPB1* were significantly associated with hepatitis B virus (HBV) infection in Han Chinese populations, and both A alleles (rs2395309 and rs9277535) also showed a strong protective effect for HBV clearance. Furthermore, we found that the genotype distributions of both sites (rs2395309 and rs9277535) were clearly different between southern and northern Han Chinese population. By using asymptomatic HBV carrier as control group, our study showed that there were no associations of *HLA-DP* variants (rs2395309 and rs9277535) with HBV progression. Although HBV disease is not determined solely by genetic factors, the experimental results offer the foundation for further study of genetic variations in the *HLA-DPA1* and *HLA-DPB1* for the prevention and therapy of chronic HBV infection.

Supporting Information

Figure S1 Associations of these two SNPs (rs2395309, rs9277535) genotypes with HBV DNA levels. *P* values of independent-sample Kolmogorov-Smirnov t test for dominant model (AA+AG vs GG). Abbreviations:SNPs, single nucleotide polymorphisms.
(TIF)

Table S1 Diagnosis criteria for Healthy control group (Health), HBV clearance group (Clear), Asymptomatic chronic HBV carriers group (AsC), Chronic active hepatitis B group (CHB), HBV-related liver cirrhosis group (LC) and HBV-related heptocellular carcinoma group (HCC).
(DOC)

Table S2 TaqMan probes and Primers for two SNPs (rs2395309 and rs9277535).
(DOC)

Table S3 The stratified analysis of gender between two SNPs (rs2395309, rs9277535) genotypes and different populations. Male and female patients showed different genotype distributions in these two SNPs (rs2395309 and rs9277535), specially in the northern Chinese population. The *P* values, odds ratios (OR), and 95% confidence intervals (CI) were calculated on the basis of the binary logistic regression analysis, adjusted for age.
(DOC)

Table S4 The stratified analysis of age between two SNPs (rs2395309, rs9277535) genotypes in south Chinese population and north Chinese population. Most cases were no significant difference in genotype distributions of two SNPs sites between patients with age≤45 years and patients with age>45 years. The *P* values, odds ratios (OR), and 95% confidence intervals (CI) were calculated on the basis of the binary logistic regression analysis, adjusted for sex.
(DOC)

Table S5 Associations of two SNPs (rs2395309, rs9277535) with HBV progression in Han Chinese populations. Compared with asymptomatic HBV group, those two sites (rs2395309 and rs9277535) in HLA-DPA1 or HLA-DPB1 gene had no associations with the chronic active hepatitis B, the HBV-related liver cirrhosis, and the HBV-related heptocellular carcinoma in southern and northern Chinese population.
(DOC)

Table S6 A Meta-analysis for previous study and current study (more than 2,243 cases and 4,137 controls). Genotype distributions of rs9277535 and rs2395309 in three ethnic groups (Japanese, Thais, Chinese) between healthy control group and chronic active hepatitis B group. *P* values of Pearson's x^2 test for allele model. Odds ratios (OR) and 95% confidence intervals (CI) of minor allele from two-by-two allele frequency table.
(DOC)

Acknowledgments

We thank all subjects for their ongoing participation in this study. We also thank Dr. Xinguo Peng (The Affiliated Hospital of Binzhou Medical College), Xinjuan Kong (Qingdao Infectious Disease Hospital), Xiaoliang Li, Bin Wen, Zhiyu Li (The Third People's Hospital of Shenzhen), Lu Yuan(The Fourth People's Hospital of Shenzhen) and Yan Gao (The Sixth People's Hospital of Shenzhen) for the help in recruitment of patients; Mr. Yuncheng Zhang, Zhixian Liu and Mrs. Hong Yuan, Ying Zhang, Lizhen Yang, Juzhen Zhong (The Third People's Hospital of Shenzhen) for the laboratory assistance.

Author Contributions

Conceived and designed the experiments: J.Lin Y.Chen J.Li. Performed the experiments: J.Li MW LL J.Yao QS MC HY. Analyzed the data: J.Li Y.Chang X.He. Contributed reagents/materials/analysis tools: J.Yu J.Li Xiaorui Jiang SS QL Xiang Jiang. Wrote the paper: J.Li. Provided the unpublished data: KM YN. Subject recruitment, biological sample collection and medical records in Guangdong province: DY J.Li J.Yao QS. Subject recruitment, biological sample collection and medical records in Shandong province: ZW LL MC HY. Subject recruitment, biological sample collection and medical records in Hubei province: YH FT X.Hu JW J.Yu QS MC HY.

References

1. World Health Organization. Hepatitis B. World Health Organization Fact Sheet No. 204. Available: http://who.int/mediacentre/factsheets/fs204/en. (Revised August 2008).
2. Ganem D, Prince A (2004) Hepatitis B virus infection natural history and clinical consequences. New England Journal of Medicine 350: 1118.
3. Lu FM, Zhuang H (2009) Management of hepatitis B in China. Chin Med J (Engl) 122: 3–4.
4. Wright TL (2006) Introduction to Chronic Hepatitis B Infection. The American Journal of Gastroenterology 101: S1–S6.
5. Blackwell JM, Jamieson SE, Burgner D (2009) HLA and Infectious Diseases. Clinical Microbiology Reviews 22: 370–385.
6. Godkin A, Davenport M, Hill AVS (2005) Molecular analysis of HLA class II associations with hepatitis B virus clearance and vaccine nonresponsiveness. Hepatology 41: 1383–1390.
7. Thursz MR, Kwiatkowski D, Allsopp CE, Greenwood BM, Thomas HC, et al. (1995) Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. N Engl J Med 332: 1065–1069.
8. Singh R, Kaul R, Kaul A, Khan K (2007) A comparative review of HLA associations with hepatitis B and C viral infections across global populations. WORLD JOURNAL OF GASTROENTEROLOGY 13: 1770.
9. Ahn SH, Han KH, Park JY, Lee CK, Kang SW, et al. (2000) Association between hepatitis B virus infection and HLA-DR type in Korea. Hepatology 31: 1371–1373.
10. Consortium TWTCC (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.
11. Ober C, Thompson E (2005) Rethinking genetic models of asthma: the role of environmental modifiers. Current Opinion in Immunology 17: 670–678.
12. Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A, et al. (2009) A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nature Genetics 41: 591–595.
13. Zeng Z, Guan L, An P, Sun S, O'Brien SJ, et al. (2008) A population-based study to investigate host genetic factors associated with hepatitis B infection and pathogenesis in the Chinese population. BMC Infectious Diseases 8: 1.
14. He X, Chang Y, Jiang H, Tang F, Meng F, et al. (2010) Persistent Effect of IFNAR-1 Genetic Polymorphism on the Long-Term Pathogenesis of Chronic HBV Infection. Viral Immunology 23: 251–257.
15. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10: 564–567.
16. Xu S, Yin X, Li S, Jin W, Lou H, et al. (2009) Genomic Dissection of Population Substructure of Han Chinese and Its Implication in Association Studies. The American Journal of Human Genetics 85: 762–774.
17. Guo X, Zhang Y, Li J, Ma J, Wei Z, et al. (2011) Strong influence of human leukocyte antigen (HLA)-DP gene variants on development of persistent chronic hepatitis B virus carriers in the Han Chinese population. Hepatology 53: 422–428.
18. Prieto J (2008) Inflammation, HCC and sex: IL-6 in the centre of the triangle. Journal of Hepatology 48: 380–381.
19. Chen S, Yeh J, Chang M, Liao Y, Hsiao L, et al. (2010) Gender Difference of Alanine Aminotransferase Elevation May Be Associated with Higher Hemoglobin Levels among Male Adolescents. PLoS ONE 5: e13269.
20. Naugler W, Sakurai T, Kim S, Maeda S, Kim K, et al. (2007) Gender disparity in liver cancer due to sex differences in MyD88-dependent IL-6 production. Science 317: 121.
21. Wang S-H, Yeh S-H, Lin W-H, Wang H-Y, Chen D-S, et al. (2009) Identification of androgen response elements in the enhancer I of hepatitis B virus: A mechanism for sex disparity in chronic hepatitis B. Hepatology 50: 1392–1402.
22. Trachtenberg E, Vinson M, Hayes E, Hsu YM, Houtchens K, et al. (2007) HLA class I (A, B, C) and class II (DRB1, DQA1, DQB1, DPB1) alleles and haplotypes in the Han from southern China. Tissue antigens 70: 455–463.
23. Yang G, Deng YJ, Hu SN, Wu DY, Li SB, et al. (2006) HLA-A, -B, and -DRB1 polymorphism defined by sequence-based typing of the Han population in Northern China. Tissue antigens 67: 146–152.
24. Wen B, Li H, Lu D, Song X, Zhang F, et al. (2004) Genetic evidence supports demic diffusion of Han culture. Nature 431: 302–305.
25. Zhang F, Su B, Zhang YP, Jin L (2007) Genetic studies of human diversity in East Asia. Philos Trans R Soc Lond B Biol Sci 362: 987–995.
26. Zhao TM, Lee TD (1989) Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. Hum Genet 83: 101–110.
27. Chaix R, Cao C, Donnelly P (2008) Is mate choice in humans MHC-dependent? PLoS Genet 4: e1000184.

28. Barreiro LB, Quintana-Murci L (2010) From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet 11: 17–30.
29. Prugnolle F, Manica A, Charpentier M, Guegan JF, Guernier V, et al. (2005) Pathogen-driven selection and worldwide HLA class I diversity. Curr Biol 15: 1022–1027.
30. Zeng G, Wang Z, Wen S, Jiang J, Wang L, et al. (2005) Geographic distribution, virologic and clinical characteristics of hepatitis B virus genotypes in China. J Viral Hepat 12: 609–617.
31. Huang ZJ, Zhou MG, Wang LJ (2007) Study on the geographic distribution of liver cancer mortality and HBsAg carrier rate in China. Disease Surveillance (Chinese) 22: 242–245.
32. Giovanna F, Bortolotti F, Francesco D (2008) Natural history of chronic hepatitis B: Special emphasis on disease progression and prognostic factors. Journal of Hepatology 48: 335–352.
33. Yang HI, Lu SN, Liaw YF, You SL, Sun CA, et al. (2002) Hepatitis B e antigen and the risk of hepatocellular carcinoma. N Engl J Med 347: 168–174.
34. Handunnetthi L, Ramagopalan SV, Ebers GC, Knight JC (2009) Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. Genes and Immunity 11: 99–112.
35. Llovet JM, Burroughs A, Bruix J (2003) Hepatocellular carcinoma. Lancet 362: 1907–1917.
36. Chen JD, Yang HI, Iloeje UH, You SL, Lu SN, et al. (2010) Carriers of inactive hepatitis B virus are still at risk for hepatocellular carcinoma and liver-related death. Gastroenterology 138: 1747–1754. e1741.
37. Zhang H, Zhai Y, Hu Z, Wu C, Qian J, 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.
38. Thorsby E (1974) The human major histocompatibility system. Transplant Rev 18: 51–129.
39. Pieters J (2000) MHC class II-restricted antigen processing and presentation. Adv Immunol 75: 159–208.
40. Jones EY, Fugger L, Strominger JL, Siebold C (2006) MHC class II proteins and disease: a structural perspective. Nat Rev Immunol 6: 271–282.
41. Grivennikov SI, Greten FR, Karin M (2010) Immunity, Inflammation, and Cancer. Cell 140: 883–899.
42. Hirayama K, Chen H, Kikuchi M, Yin T, Gu X, et al. (1999) HLA-DR-DQ alleles and HLA-DP alleles are independently associated with susceptibility to different stages of post-schistosomal hepatic fibrosis in the Chinese population. Tissue antigens 53: 269–274.

nature genetics

Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma

Vinod Kumar[1,2], Naoya Kato[3], Yuji Urabe[1], Atsushi Takahashi[2], Ryosuke Muroyama[3], Naoya Hosono[2], Motoyuki Otsuka[4], Ryosuke Tateishi[4], Masao Omata[4], Hidewaki Nakagawa[2], Kazuhiko Koike[4], Naoyuki Kamatani[2], Michiaki Kubo[2], Yusuke Nakamura[1,2] & Koichi Matsuda[1]

To identify the genetic susceptibility factor(s) for hepatitis C virus–induced hepatocellular carcinoma (HCV-induced HCC), we conducted a genome-wide association study using 432,703 autosomal SNPs in 721 individuals with HCV-induced HCC (cases) and 2,890 HCV-negative controls of Japanese origin. Eight SNPs that showed possible association ($P < 1 \times 10^{-5}$) in the genome-wide association study were further genotyped in 673 cases and 2,596 controls. We found a previously unidentified locus in the 5′ flanking region of *MICA* on 6p21.33 (rs2596542, $P_{combined} = 4.21 \times 10^{-13}$, odds ratio = 1.39) to be strongly associated with HCV-induced HCC. Subsequent analyses using individuals with chronic hepatitis C (CHC) indicated that this SNP is not associated with CHC susceptibility ($P = 0.61$) but is significantly associated with progression from CHC to HCC ($P = 3.13 \times 10^{-8}$). We also found that the risk allele of rs2596542 was associated with lower soluble MICA protein levels in individuals with HCV-induced HCC ($P = 1.38 \times 10^{-13}$).

It is estimated that more than 170 million people are infected with HCV worldwide[1]. Persistent HCV infection causes CHC and, subsequently, fatal liver diseases such as liver cirrhosis and HCC. Therefore, the treatment of HCV carriers is an issue of global importance. HCC is the third most common cause of cancer-related deaths[2], and HCV infection accounts for 30–70% of the individuals with HCC[3,4]. HCV-induced HCC is a multistep and progressive liver disease in which disease progression may be influenced by both environmental and genetic risk factors. The impact of host genetic variation on progression to CHC after HCV exposure is well documented by recent genome-wide association studies (GWAS)[5–7]. However, no comprehensive analyses have been performed to explore the genetic basis of HCV-induced HCC. Therefore, we conducted a GWAS for HCV-induced HCC.

We genotyped the DNA of 721 individuals with HCV-induced HCC and 2,890 HCV-negative controls (**Supplementary Table 1**) from BioBank Japan[8]. After the initial standard SNP quality filters, we obtained genotyping results for 432,703 SNPs for association analysis. Because progression from CHC to liver cancer is strongly affected by age and gender[3], we performed a logistic regression analysis by including age and gender as covariates at all tested loci in our analyses. The genetic inflation factor (λ) was 1.03, indicating that there is no or little population stratification (**Supplementary Fig. 1**). Although no SNPs cleared the GWAS significance threshold ($P < 5 \times 10^{-8}$) at this stage, we identified eight independent loci showing possible association ($P < 1 \times 10^{-5}$; **Supplementary Fig. 2**).

In the replication stage, 673 cases from an independent HCC cohort from the University of Tokyo and 2,596 HCV-negative controls from BioBank Japan were genotyped at these eight SNPs. We observed a significant replication of association at rs2596542 on chromosome 6p21.33 ($P = 8.62 \times 10^{-9}$, odds ratio (OR) = 1.44, 95% confidence interval (CI) 1.27–1.63; **Table 1**), whereas the remaining seven SNPs failed to replicate the association (**Supplementary Table 2**). Furthermore, the combination analysis of the GWAS and replication study data at rs2596542 revealed a highly significant association in which the frequency of the risk allele A is higher in cases ($P = 4.21 \times 10^{-13}$, OR = 1.39; **Fig. 1** and **Table 1**) after the age and gender adjustment, without any heterogeneity ($P = 0.24$) between the two stages. To further investigate the impact of rs2596542 on the complex nature of the HCV-induced HCC phenotype, we genotyped 1,730 individuals with CHC who had not developed liver cirrhosis or HCC during their recruitment. As a result, rs2596542 was found to have no association with chronic hepatitis C susceptibility ($P = 0.61$) but was significantly associated with progression from CHC to HCC ($P = 3.13 \times 10^{-8}$, OR = 1.36; **Table 2**).

Because heavy alcohol consumption (>50 g per day) as well as poor response to interferon (IFN) treatment were shown to be the major risk factors for HCC among individuals with CHC[9], we evaluated the effect of alcohol consumption as a confounding factor and found that rs2596542 remained highly significant even after adjustment for this factor (non-HCV versus HCC, OR = 1.39, $P = 1.22 \times 10^{-11}$; CHC versus HCC, OR = 1.25, $P = 2.31 \times 10^{-4}$; **Supplementary Table 3**). The major genotypes of HCV can be determined by a serotyping

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [2]Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan. [3]Unit of Disease Control Genome Medicine, The Institute of Medical Science, University of Tokyo, Tokyo, Japan. [4]Department of Gastroenterology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. Correspondence should be addressed to K.M. (koichima@ims.u-tokyo.ac.jp).

Received 20 July 2010; accepted 23 March 2011; published online 17 April 2011; doi:10.1038/ng.809

Table 1 Association results of rs2596542 in the GWAS, replication stage and combined analysis

SNP	Chr. (locus)	Stage	Case RAF	Control RAF	*P*	OR (95% CI)
rs2596542 (A/G)	6 (*MICA*)	GWAS[a]	0.388	0.331	4.50×10^{-6}	1.34 (1.16–1.53)
		Replication[a]	0.413	0.331	8.62×10^{-9}	1.44 (1.27–1.63)
		Combined[a]	0.400	0.331	4.21×10^{-13}	1.39 (1.27–1.52)
				MH test	7.76×10^{-12}	1.35 (1.24–1.47)

We analyzed 1,394 cases with HCC (721 in the GWAS and 673 in the replication) and 5,486 controls (2,890 in GWAS and 2,596 in replication). Chr., chromosome; RAF, risk allele frequency (allele A); OR, odds ratio for the minor allele calculated by considering the major allele as a reference; MH, Mantel-Haenszel.
[a]*P* values and ORs are adjusted for age and gender by logistic regression analysis under an additive model.

assay that is based on the type-specific antibodies produced by the infected host[10]. A subgroup analysis for HCV serotypes or history of IFN therapy indicated that this variation is associated with HCC susceptibility independently of HCV genotypes or treatment response (**Supplementary Fig. 3**). Consistent with this result, rs1051796, which had $r^2 = 0.7$ and $D' = 0.95$ with rs2596542, was not associated with IFN response ($P = 0.89$) according to previously published data in the Japanese population[11].

rs2596542 is located within the class I major histocompatibility complex (MHC) region. The human MHC region encompasses the complex and extended linkage disequilibrium (LD) structure[12,13]. Several HLA alleles and genes within MHC region have been implicated in HCV infection or clearance or in response to treatment[14–16]. Therefore, we searched the whole 7.5-Mb extended MHC region using GWAS data to test the possibility of other associated loci. We found a moderate association peak at rs9275572 ($P = 4.99 \times 10^{-5}$), which is located between *HLA-DQA* and *HLA-DQB* loci (**Supplementary Fig. 4**). Subsequent replication and combination analyses at rs9275572 indicated a significant association with HCV-induced HCC ($P = 9.38 \times 10^{-9}$, OR = 1.30; **Supplementary Table 4**). The multiple logistic regression analysis to control for alcohol consumption along with age and gender also indicated a significant association at rs9275572 ($P = 3.21 \times 10^{-8}$, OR = 1.29; **Supplementary Table 5**). However, rs2596542 was not in high LD with rs9275572 ($D' = 0.41$, $r^2 = 0.16$), and both SNPs remained associated with HCC even after conditional analysis on each other and had small reductions in their ORs upon conditioned analysis (OR = 1.23, $P = 4.43 \times 10^{-6}$ and OR = 1.17, $P = 0.00059$, respectively; **Supplementary Table 6**). A haplotype analysis between these two markers showed four possible haplotypes, with haplotype AA showing higher risk (with OR = 1.44) compared to the major haplotype GG (**Supplementary Table 7**). However, the OR for the risk haplotype was 1.32 with $P = 2.31 \times 10^{-10}$ after comparing against all observed haplotypes in the population (**Supplementary Table 7**), which is weaker than that of rs2596542 alone (OR = 1.39, $P = 4.21 \times 10^{-13}$). Hence, the impact of rs2596542 is much stronger than the haplotype of two SNPs, suggesting that rs2596542 is a principal genetic factor in this region. We also found that rs9275572 has a moderate association with CHC susceptibility as well as progression from CHC to HCC ($P = 0.03$ and $P = 2.58 \times 10^{-5}$, OR = 1.09 and OR = 1.29, respectively; **Supplementary Table 8**). Because *HLA-DQ* and *HLA-DR* alleles were shown to be associated with viral persistence and early liver disease among Japanese individuals[16], further study will be needed to confirm whether the association at rs9275572 is because of its LD with *HLA-DQ* or *DR* alleles.

In this regard, it is interesting to note that rs9275572 had a very strong expression quantitative trait locus effect on *HLA-DQB1* ($\log_{10}$ odds (LOD) ≥ 19.48) and *HLA-DRB4* alleles (LOD ≥ 26.88)[17]. Thus, it will be important to test the functional effect of the common haplotype (AA; **Supplementary Table 7**), which tags the risk alleles at these two SNPs.

Two SNPs, rs12979860 and rs8099917, at the *IL28B* locus were reported to be associated with spontaneous clearance of HCV virus[18] and response to pegylated IFN-α and ribavirin therapy[11], respectively. However, we found no association at rs12979860 and rs8099917 in our dataset (**Supplementary Table 9**). Because we used non-HCV control subjects rather than subjects who had cleared HCV infection spontaneously, and because only about 20% of the cases with HCC had been treated with IFN, our study may not be suitable to detect associations at the *IL28B* locus. In addition, the protective C allele at rs12979860 is nearly fixed throughout east Asia, with a frequency of more than 91% in the Japanese population as compared to 67% in European Americans[6], indicating a role for other factors in spontaneous clearance.

The top associated SNP, rs2596542, is located 4.7 kb upstream of *MICA*, the MHC class I polypeptide-related sequence A gene, and 41.7 kb downstream of the *HLA-B* gene (**Supplementary Fig. 5**). The regional association plot at the rs2596542 locus, made using genotype data from the GWAS (**Fig. 1**) and imputation analysis (**Supplementary Fig. 6**), revealed that all of the modestly associated SNPs are tightly

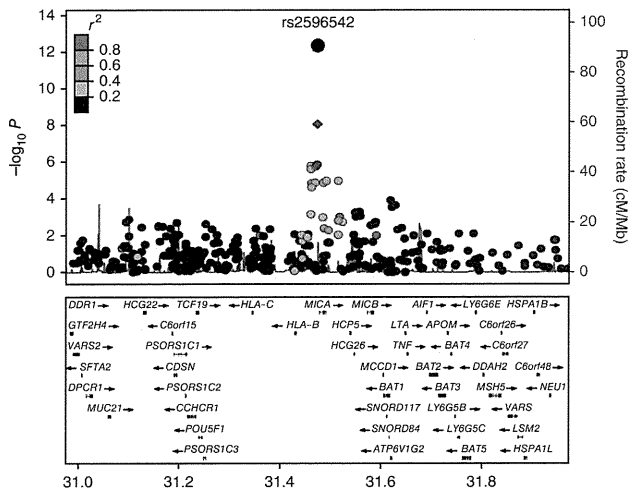

Figure 1 Regional association plot at rs2596542. Above, the *P* values of genotyped SNPs are plotted (as $-\log_{10}$ values) against their physical position on chromosome 6 (NCBI Build 36). The *P* value for rs2596542 at the GWAS stage, replication stage and combination analysis is represented by a purple diamond, circle and diamond, respectively. Estimated recombination rates from the HapMap JPT population show the local LD structure. Inset, the SNP's colors indicate LD with rs2596542 according to a scale from $r^2 = 0$ to $r^2 = 1$ based on pairwise r^2 values from HapMap JPT. Below, gene annotations from the UCSC genome browser.

Table 2 rs2596542 (A/G) is associated with progression from CHC to HCC

Subjects	RAF	(Comparison) *P*[a]	OR[a]	95% CI
Healthy	0.331			
CHC	0.333	(Healthy vs. CHC) 0.61	1.02	0.94–1.10
HCC	0.398	(CHC vs. HCC) 3.13×10^{-8}	1.36	1.22–1.51

We analyzed 5,486 controls, 1,730 CHC cases and 1,394 HCC cases. RAF, risk allele frequency (allele A); OR, odds ratio for the minor allele by considering the major allele as a reference.
[a]Calculated by logistic regression analysis, by PLINK upon age and gender adjustment under additive model.

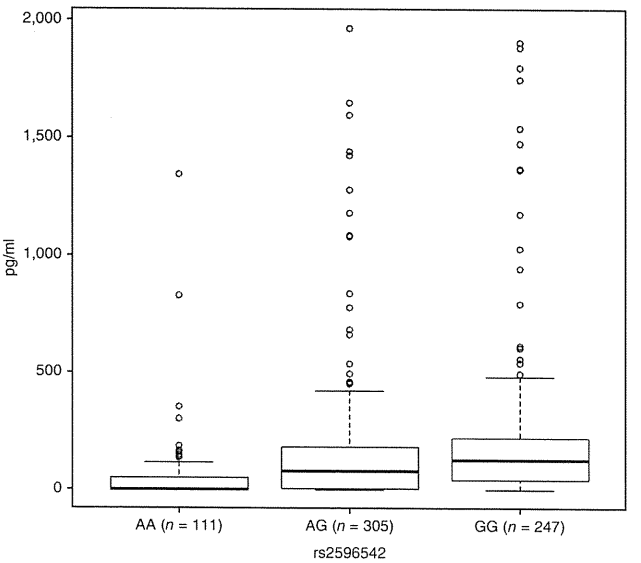

Figure 2 Correlation between soluble MICA levels and rs2596542 genotype. The *x* axis shows the genotypes at rs2596542, and the *y* axis shows the concentration of soluble MICA in pg/ml. The number of independent samples tested in each group is shown in parentheses. Each group is shown as a box plot, and the median values are shown as thick dark horizontal lines (median values of AA = 0, AG = 43.6 and GG = 77.74). The box covers the twenty-fifth to seventy-fifth percentiles, and the whiskers outside the box extend to the highest and lowest value within 1.5 times the interquartile range. Points outside the whiskers are outliers. We tested the difference in the median values among genotypes using the Kruskal-Wallis test ($P = 1.6 \times 10^{-13}$). We plotted the box plots using default settings in R (see URLs).

linked to rs2596542 ($r^2 > 0.4$) and are confined to the *MICA* gene locus. On the other hand, the imputation analysis of *HLA*-tagging SNPs did not show any evidence of linkage with rs2596542 (Online Methods and **Supplementary Table 10**), suggesting that *MICA* is a disease-associated candidate gene at this locus.

MICA is a membrane protein that acts as a ligand for NKG2D to activate anti-tumor effects through natural killer cells and $CD8^+$ T cells[19]. On the other hand, MICA is secreted into the serum by cleavage at the transmembrane domain with matrix metalloproteinases[20,21] and inhibits the anti-tumor effect of natural killer cells and $CD8^+$ T cells by blocking their action[22–24]. Elevated expression of both the membrane-bound and soluble forms of MICA (sMICA) have been reported in several cancers, including HCC[25–27]. Exon 5 of *MICA* encodes the transmembrane domain and contains a variable number of tandem repeats (VNTR) consisting of 4, 5, 6 or 9 repeats of GCT or one additional G nucleotide insertion into the 5-GCT-repeat allele (referred as A4, A5, A6, A9 and A5.1, respectively). The insertion of G (A5.1) causes a premature stop codon and subsequent loss of the transmembrane domain, leading to altered subcellular localization[28]. Therefore, we tested whether rs2596542 is in linkage with functional *MICA* VNTR alleles.

We further genotyped 673 cases with HCV-induced HCC and 890 non-HCV controls for the *MICA* VNTR locus with capillary-based electrophoresis (**Supplementary Fig. 7**). A case-control analysis revealed that the *MICA* VNTR is associated with HCV-induced HCC (global $P = 4.55 \times 10^{-7}$; **Supplementary Table 11**). Particularly, alleles A9 and A6 were associated with conferring a higher risk of HCC (OR = 1.73 and OR = 1.34, respectively), whereas the A5 and A5.1 alleles had a protective effect. Comparison of the genotypes at rs2596542 and the VNTR locus revealed that the A risk allele at rs2596542 is in LD with the A9 and A4 alleles, and the non-risk G allele is in LD with the A5 and A5.1 alleles, whereas we observed no linkage between an A6 allele and rs2596542 (**Supplementary Table 12**). We also genotyped 124 individuals with CHC; however, we observed no significant association between individuals with CHC and controls or individuals with CHC and HCC (**Supplementary Tables 13,14**).

We then tested whether the VNTR alleles, rs2596542 alleles, or VNTR-rs2596542 haplotypes had any association with MICA expression in individuals with HCV-induced HCC. We determined sMICA levels by ELISA using a total of 665 HCC serum samples (**Supplementary Table 15**). Notably, rs2596542 was significantly correlated with sMICA levels, and specifically, the risk genotype AA was associated with low levels of sMICA ($P = 1.38 \times 10^{-13}$; **Fig. 2**), whereas VNTR alleles (**Supplementary Fig. 8**) and VNTR- rs2596542 haplotypes (**Supplementary Table 16**) showed no strong association. The absence of any correlation between *MICA* VNTR alleles and sMICA suggests that sMICA levels are not regulated by post-translational processing or a premature stop codon caused by A5.1 alleles in individuals with HCC. We also examined the sMICA level in different stages of HCV-induced liver disease (in non-HCV subjects and those with CHC and HCV-induced liver cirrhosis) and found that sMICA level was elevated at the early stage of disease and was not correlated with disease progression (**Fig. 3**). Additionally, the risk allele A was also correlated with low sMICA levels in subjects with CHC (**Supplementary Fig. 9**). These findings suggest that MICA expression was induced by factors caused by chronic HCV infection,

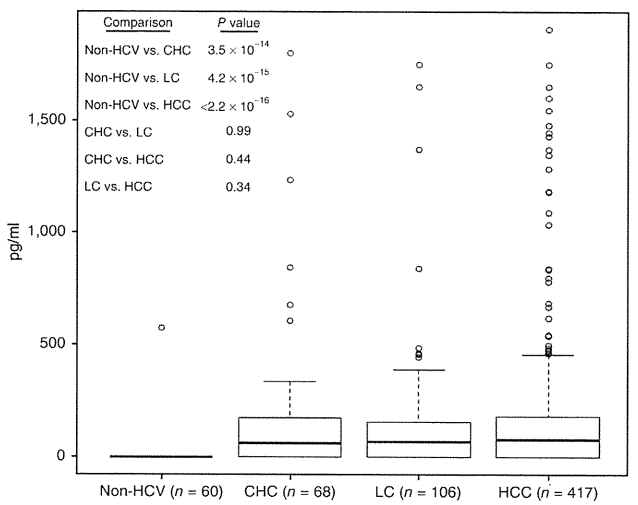

Figure 3 Correlation between soluble MICA and HCV-related diseases. The *x* axis shows the disease stages after HCV infection, and the *y* axis shows the concentration of soluble MICA in pg/ml. The number of independent samples tested in each group is shown in parentheses. Each group is shown as a box plot, and the median values are shown as thick dark horizontal lines (median values of non-HCV = 0, CHC = 64.55, LC = 72.11 and HCC = 77.98). The box covers the twenty-fifth to seventy-fifth percentiles, and the whiskers outside the box extend to the highest and lowest value within 1.5 times the interquartile range. Points outside the whiskers are outliers. We tested the difference in the median values among the disease groups using the Wilcoxon rank test. The box plots were plotted using default settings in R. Non-HCV, individuals not exposed to HCV infection; CHC, individuals with chronic hepatitis C; LC, individuals with liver cirrhosis; HCC, individuals with hepatocellular carcinoma.

similar to various types of stresses such as viral infection, inflammation and heat shock[29,30]. The levels of sMICA were shown to be directly proportional to the level of membrane-bound MICA[25], and membrane bound MICA is essential for activating natural killer cells and $CD8^+$ T cells to eliminate virus-infected cells[19]. Considering the association of the risk allele A with low levels of sMICA, our findings suggest that the individuals who carry the rs2596542 A allele would express low levels of membrane-bound MICA in response to HCV infection, which thus leads to poor or no activation of natural killer cells and $CD8^+$ T cells against virus-infected cells. Eventually, these individuals are likely to progress from CHC to HCC. Notably, several SNPs that are in absolute linkage with rs2596542 are located within the promoter or enhancer region of *MICA* and may alter the binding of stress-inducible transcriptions factors such as heat shock proteins (**Supplementary Table 17**). In this regard, it is important to analyze the factors that regulate *MICA* expression, particularly in the context of CHC. Although, the molecular mechanism whereby *MICA* polymorphisms confer the risk of disease progression should be characterized in the future, our findings reveal a crucial role of genetic variations in the host innate immune system in the development of HCV-induced HCC.

URLs. R, http://cran.r-project.org/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; Primer3 v0.3.0, http://frodo.wi.mit.edu/primer3/; LocusZoom, http://csg.sph.umich.edu/locuszoom/; FastSNP, http://fastsnp.ibms.sinica.edu.tw/pages/input_CandidateGeneSearch.jsp.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

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

ACKNOWLEDGMENTS

We would like to thank all the subjects and the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan who donated their DNA for this work. We also thank the technical staff of the Laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN, and the Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, the University of Tokyo. 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.

AUTHOR CONTRIBUTIONS

K.M. and Y.N. conceived of the study; Y.N., V.K., M.K. and K.M. designed the study; V.K., Y.U., R.M. and N.H. performed genotyping; V.K., Y.N. and K.M. wrote the manuscript; A.T. and N. Kamatani performed quality control at the genome-wide phase; Y.N., K.M., H.N. and M.K. managed DNA and serum samples belonging to BioBank Japan; N. Kato, R.T., M. Otsuka, M. Omata and K.K. managed replication DNA and serum samples; V.K. analyzed the data, performed VNTR genotyping, ELISA and summarized the whole results; Y.N. obtained funding for the study.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Global Burden of Hepatitis C Working Group. Global burden of disease (GBD) for hepatitis C. *J. Clin. Pharmacol.* **44**, 20–29 (2004).
2. Parkin, D.M., Bray, F., Ferlay, J. & Pisani, P. Global cancer statistics, 2002. *CA Cancer J. Clin.* **55**, 74–108 (2005).
3. Umemura, T., Ichijo, T., Yoshizawa, K., Tanaka, E. & Kiyosawa, K. Epidemiology of hepatocellular carcinoma in Japan. *J. Gastroenterol.* **44** (Suppl 19), 102–107 (2009).
4. Vong, S. & Bell, B.P. Chronic liver disease mortality in the United States, 1990–1998. *Hepatology* **39**, 476–483 (2004).
5. Ge, D. *et al.* Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).
6. Thomas, D.L. *et al.* Genetic variation in *IL28B* and spontaneous clearance of hepatitis C virus. *Nature* **461**, 798–801 (2009).
7. Rauch, A. *et al.* Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. *Gastroenterology* **138**, 1338–1345, 1345.e1–7 (2010).
8. Kamatani, Y. *et al.* A genome-wide association study identifies variants in the *HLA-DP* locus associated with chronic hepatitis B in Asians. *Nat. Genet.* **41**, 591–595 (2009).
9. Schütte, K., Bornschein, J. & Malfertheiner, P. Hepatocellular carcinoma–epidemiological trends and risk factors. *Dig. Dis.* **27**, 80–92 (2009).
10. Tanaka, T. *et al.* Significance of specific antibody assay for genotyping of hepatitis C virus. *Hepatology* **19**, 1347–1353 (1994).
11. Tanaka, Y. *et al.* Genome-wide association of *IL28B* with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat. Genet.* **41**, 1105–1109 (2009).
12. Anonymous. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* **401**, 921–923 (1999).
13. de Bakker, P.I. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* **38**, 1166–1172 (2006).
14. Kuniholm, M.H. *et al.* Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia. *Hepatology* **51**, 1514–1522 (2010).
15. Wang, J.H. *et al.* Ethnic and geographical differences in HLA associations with the outcome of hepatitis C virus infection. *Virol. J.* **6**, 46 (2009).
16. 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).
17. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
18. Ge, D. *et al.* Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).
19. Bauer, S. *et al.* Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* **285**, 727–729 (1999).
20. Salih, H.R., Rammensee, H. & Steinle, A. Cutting edge: down-regulation of MICA on human tumors by proteolytic shedding. *J. Immunol.* **169**, 4098–4102 (2002).
21. Waldhauer, I. *et al.* Tumor-associated MICA is shed by ADAM proteases. *Cancer Res.* **68**, 6368–6376 (2008).
22. Jinushi, M. *et al.* Impairment of natural killer cell and dendritic cell functions by the soluble form of MHC class I-related chain A in advanced human hepatocellular carcinomas. *J. Hepatol.* **43**, 1013–1020 (2005).
23. Groh, V., Wu, J., Yee, C. & Spies, T. Tumour-derived soluble MIC ligands impair expression of NKG2D and T-cell activation. *Nature* **419**, 734–738 (2002).
24. Doubrovina, E.S. *et al.* Evasion from NK cell immunity by MHC class I chain-related molecules expressing colon adenocarcinoma. *J. Immunol.* **171**, 6891–6899 (2003).
25. Kohga, K. *et al.* Serum levels of soluble major histocompatibility complex (MHC) class I-related chain A in patients with chronic liver diseases and changes during transcatheter arterial embolization for hepatocellular carcinoma. *Cancer Sci.* **99**, 1643–1649 (2008).
26. Jinushi, M. *et al.* Expression and role of MICA and MICB in human hepatocellular carcinomas and their regulation by retinoic acid. *Int. J. Cancer* **104**, 354–361 (2003).
27. Groh, V. *et al.* Broad tumor-associated expression and recognition by tumor-derived gamma delta T cells of MICA and MICB. *Proc. Natl. Acad. Sci. USA* **96**, 6879–6884 (1999).
28. Ota, M. *et al.* Trinucleotide repeat polymorphism within exon 5 of the MICA gene (MHC class I chain-related gene A): allele frequency data in the nine population groups Japanese, Northern Han, Hui, Uygur, Kazakhstan, Iranian, Saudi Arabian, Greek and Italian. *Tissue Antigens* **49**, 448–454 (1997).
29. Groh, V. *et al.* Costimulation of CD8αβ T cells by NKG2D via engagement by MIC induced on virus-infected cells. *Nat. Immunol.* **2**, 255–260 (2001).
30. Groh, V. *et al.* Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proc. Natl. Acad. Sci. USA* **93**, 12445–12450 (1996).

ONLINE METHODS

Sample collections. We obtained DNA from 721 HCV-related HCC cases, 1,730 CHC cases and 5,486 HCV-negative controls from the BioBank Japan project[31]. For replication analysis, DNA from 673 HCV-induced HCC cases was obtained from a prospective HCC study cohort of the University of Tokyo. A diagnosis of CHC, liver cirrhosis or HCC were based on histological, clinical and laboratory findings obtained by trained physicians. Case samples with HBV co-infection were excluded from the analysis. Interferon was administrated to 20.4% of HCC cases and 70.1% of cases were not treated. The remaining 9.5% of the cases lacked information about interferon treatment. The non-HCV controls obtained from BioBank Japan contained case-mixed individuals after excluding all individuals with cancer, chronic hepatitis B, diabetes or tuberculosis. All subjects were of Japanese origin and provided written informed consent. The clinical and demographic details of the samples are summarized in **Supplementary Table 1**. We also obtained serum samples from BioBank Japan and the University of Tokyo (**Supplementary Table 12**). This research project was approved by the ethical committees of the University of Tokyo and RIKEN.

SNP genotyping and quality control. In the GWAS, 721 individuals with HCV-related liver cancer and 2,890 controls were genotyped using Illumina HumanHap610-Quad and Illumina HumanHap550v3 Genotyping BeadChip, respectively. In the replication stage, 673 cases with HCV-related disease, 1,730 cases with CHC and 2,596 controls were genotyped by the multiplex PCR-based Invader assay (Third Wave Technologies) and the Illumina HumanHap610-Quad, respectively. The common SNPs between the Illumina HumanHap550v3 and the Illumina HumanHap610-Quad arrays from all autosomal chromosomes were included for the analysis. We applied standard SNP quality control filters to exclude SNPs with low call rate (<99%), a Hardy-Weinberg equilibrium $P < 1.0 \times 10^{-6}$ for controls and minor allele frequency of <0.01. In the end, we obtained 432,703 SNPs for the analysis. In the replication analysis, the allele discrimination plots were validated by two well-trained researchers (the plots are available on request). We excluded samples with low genotyping rate (<99%) and employed principal component analysis to avoid the population stratification issue, in which individuals belonging only to Hondo cluster were included in the analysis (**Supplementary Fig. 10**)[32].

Statistical analysis. The association of SNPs with the disease phenotype in the GWAS, replication stage and combination analyses was tested using multivariate logistic regression analysis after adjusting for age at recruitment (continuous) and gender by assuming an additive model and using PLINK[33]. In the GWAS, the genetic inflation factor (λ) was derived by applying logistic regressed P values for all the tested SNPs. The quantile-quantile plot was drawn using R. The ORs were calculated by considering the major allele as a reference, unless it was stated otherwise elsewhere. The combined analysis of the GWAS and replication stage was verified by conducting the Mantel-Haenszel method. We considered $P < 5 \times 10^{-8}$ as the genome-wide significance threshold, which is the Bonferroni-corrected threshold for the number of independent SNPs genotyped in HapMap Phase 2 (ref. 34). Heterogeneity across the two stages was examined by using the Breslow-Day test[35].

For multiple logistic regression analysis at rs2596542 using the R program, we considered age at recruitment (≤60 or >60 years)[3], gender (male or female) and alcohol consumption (non-drinkers, ≤ 50 g alcohol per day or >50 g alcohol per day) as covariates from both the GWAS and replication stage cases with HCC and non-HCV controls. Association at the *MICA* VNTR locus was analyzed by Fisher's exact test, and the global P value was calculated using a χ^2 test. Statistical comparisons between genotypes and sMICA levels were performed by Kruskal-Wallis test or Wilcoxon rank test using R. We employed the R package haplo.stats to infer haplotypes and to perform haplotype association analysis. P values for association between sMICA levels and haplotype distribution were obtained by score test under an additive model by using the haplo.score function. ORs and 95% confidence intervals were calculated from the coefficients of the GLM model by considering the major haplotype as a reference. We used the haplo.cc function to calculate these statistical values.

HCV serotype. HCV serotype data was available for 531 cases with HCC from the replication stage. HCV serotype was examined by serotyping assay (SRL Laboratory) according to previously reported methods[36]. According to the Simmonds classification[37], serotype 1 corresponded to disease types 1a and 1b, whereas serotype 2 corresponded to disease types 2a and 2b.

***MICA* VNTR locus genotyping.** We followed the method suggested by Applied Biosystems. Briefly, the 5′ end of the forward primer was labeled with 6-FAM, and the 5′ end of reverse primer was labeled with the GTGTCTT non-random sequence to promote addition of As. The primer sequences were previously reported[28]. The PCR products were mixed with Hi-Di Formamide and GeneScan-600 LIZ size standard and separated using a GeneScan system on a 3730xl DNA analyzer (Applied Biosystems). GeneMapper software (Applied Biosystems) was used to assign the repeat fragment size (**Supplementary Fig. 7**).

Quantification of soluble MICA. sMICA levels were measured by sandwich enzyme-linked immunosorbent assay, as described in the manufacturer's instructions (R&D Systems).

Imputation and association analysis at HLA allele tagging SNPs. We obtained a SNP or a combination of SNPs which can tag HLA alleles in the Japanese population from a previous study[13]. The untyped genotypes of these SNPs were imputed in the GWAS samples by using a hidden Markov model programmed in MACH[38] and haplotype information from HapMap JPT samples. We applied the same SNP quality criteria as in the GWAS for selecting SNPs for the analysis. The association was tested on all SNPs that passed the quality control criteria using logistic regression analysis conditioned on age and gender.

Initially, we obtained the pair-wise LD between HLA alleles tagging SNPs and rs2596542. We performed case-control association analysis in our GWAS dataset. As shown in **Supplementary Table 9**, none of the HLA-tagging SNPs showed evidence of linkage or association except rs2844521, and rs2844521 was in absolute linkage with rs2596542 ($r^2 = 1$, $D' = 1$) and thus showed similar association. We obtained actual genotype data at rs2596501, as this SNP is included on the 550K SNP platform, and inferred the haplotype between rs2844521 and rs2596501. However, the haplotype GT (the G allele of rs2844521 and the T allele of rs2596501), which is reported to tag the *HLA-B*3501* allele ($r^2 = 1$, $D' = 1$), was not associated with HCC in our GWAS dataset ($P = 0.39$). We also performed a conditional logistic regression analysis on rs2596501 (data not shown) and found no effect on the association between rs2596542 and HCV-induced HCC. This data suggested that rs2596542 association is independent of *HLA-B*3501*. Although we observed mild association between other *HLA-B* alleles (*HLA-B*5401*, $P = 0.004$; *HLA-B*6701*, $P = 0.012$) and HCV-induced HCC, the association at rs2596542 alone was the most significant. Taken together, we found no strong evidence for linkage of HLA alleles with rs2596542.

Software. For general statistical analysis, we used R statistical environment version 2.6.1 or plink version 1.06. The Haploview software version 4.2 (ref. 39) was used to calculate LD and to draw Manhattan plots. Primer3 v0.3.0 web tool was used to design primers. We used LocusZoom for plotting regional association plots. We used FastSNP[40] web tool for functional annotation of SNPs (see URLs for all software packages).

31. Nakamura, Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5**, 696–697 (2007).
32. Yamaguchi-Kabata, Y. *et al.* Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am. J. Hum. Genet.* **83**, 445–456 (2008).
33. 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).
34. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
35. Breslow, N. & Day, N. Statistical methods in cancer research. Volume II–The design and analysis of cohort studies. *IARC Sci. Publ.* 1–406 (1987).
36. Tsukiyama-Kohara, K. *et al.* A second group of hepatitis C viruses. *Virus Genes* **5**, 243–254 (1991).
37. Simmonds, P. *et al.* Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *J. Gen. Virol.* **75**, 1053–1061 (1994).
38. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
39. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
40. Yuan, H.Y. *et al.* FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res.* **34**, W635–W641 (2006).

Hepatol Int
DOI 10.1007/s12072-011-9306-7

ORIGINAL ARTICLE

Characterization of naturally occurring protease inhibitor-resistance mutations in genotype 1b hepatitis C virus patients

Hiroko Shindo · Shinya Maekawa · Kazuki Komase · Ryota Sueki · Mika Miura · Makoto Kadokura · Kuniaki Shindo · Fumitake Amemiya · Takatoshi Kitamura · Yasuhiro Nakayama · Taisuke Inoue · Minoru Sakamoto · Shun-ichi Okada · Yasuhiro Asahina · Namiki Izumi · Masao Honda · Shuichi Kaneko · Nobuyuki Enomoto

Received: 21 April 2011 / Accepted: 22 July 2011

Abstract

Background and aims Protease inhibitor (PI)-resistant hepatitis C virus (HCV) variants may be present in substantial numbers in PI-untreated patients according to recent reports. However, influence of these viruses in the clinical course of chronic hepatitis C has not been well characterized.

Methods The dominant HCV nonstructural 3 (NS3) amino acid sequences were determined in 261 HCV genotype 1b-infected Japanese patients before pegylated interferon plus ribavirin (PEG-IFN/RBV) therapy, and investigated the patients' clinical characteristics as well as treatment responses including sustained virological response (SVR) rate. HCV-NS3 sequences were also determined in 39 non-SVR patients after completion of the therapy.

Results Four single mutations (T54S, Q80K, I153V, and D168E) known to confer PI resistance were found in 35 of 261 patients (13.4%), and double mutations (I153V plus T54S/D168E) were found in 6 patients (2.3%). Responses to PEG-IFN/RBV therapy did not differ between patients with and without PI-resistance mutations (mutation group, SVR 48%; wild-type group, SVR 40%; $P = 0.38$). On the other hand, two mutations appeared in two non-SVR patients after PEG-IFN/RBV therapy (I153V and E168D, 5.1%).

Conclusions PI-resistance-associated NS3 mutations exist in a substantial proportion of untreated HCV-1b-infected patients. The impact of these mutations in the treatment of PIs is unclear, but clinicians should pay attention to avoid further development of PI resistance.

Keywords HCV · Protease inhibitor · Naturally occurring viral resistance mutations

Introduction

Hepatitis C virus (HCV) infects more than 170 million persons worldwide and thus represents a global health problem. At least 130 million infected individuals are chronic carriers of HCV and are at significant risk of developing liver cirrhosis and hepatocellular carcinoma [1]. The current standard treatment with pegylated interferon plus ribavirin (PEG-IFN/RBV) is complicated by frequent adverse reactions, and a sustained virologic response (SVR) can be achieved only in 50% of patients infected with the most prevalent genotype 1 [2]. In Japan, since 70% of patients are infected with intractable genotype 1b HCV, more effective treatments are urgently required.

A promising approach is the development of specifically targeted antiviral therapies for hepatitis C (STAT-C). HCV-specific protease inhibitors (PIs) target an essential step in HCV replication by blocking the nonstructural 3/4A (NS3/4A) protease-dependent cleavage of the HCV polyprotein

Electronic supplementary material The online version of this article (doi:10.1007/s12072-011-9306-7) contains supplementary material, which is available to authorized users.

H. Shindo · S. Maekawa (✉) · K. Komase · R. Sueki · M. Miura · M. Kadokura · K. Shindo · F. Amemiya · T. Kitamura · Y. Nakayama · T. Inoue · M. Sakamoto · S. Okada · N. Enomoto
First Department of Internal Medicine, University of Yamanashi, 1110, Shimokato, Chuo, Yamanashi 409-3898, Japan
e-mail: maekawa@yamanashi.ac.jp

Y. Asahina · N. Izumi
Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan

M. Honda · S. Kaneko
Department of Gastroenterology, Kanazawa University Graduate School of Medicine, Kanazawa, Japan

Published online: 18 August 2011