Table 2. Results of replication study for protective effects against CHB.

	Position			MAF[a]	Allele	Stage	HBV carriers			Healthy controls				OR[b]		
dbSNP rsID	Chr	Buld 36.3	Nearest Gene	(allele)	(1/2)	(population)	11	12	22	11	12	22	HWEp	95% CI	*P*-value[c]	P_{het}[d]
rs3077	6	33141000	HLA-DPA1	0.44	T/C	GWAS	13	51	117	28	88	67	0.919	0.42	1.14×10^{-7}	
				(T)		(Japanese)	(7.2)	(28.2)	(64.6)	(15.3)	(48.1)	(36.6)		(0.30–0.58)		
						Replication-1	26	95	134	46	125	65	0.309	0.48	2.70×10^{-8}	
						(Japanese)	(10.2)	(37.3)	(52.5)	(19.5)	(53.0)	(27.5)		(0.37–0.62)		
						Replication-2	23	81	111	31	74	40	0.767	0.47	2.08×10^{-6}	
						(Korean)	(10.7)	(37.7)	(51.6)	(21.4)	(51.0)	(27.6)		(0.35–0.65)		
						Meta-analysis[e]								0.46	4.40×10^{-19}	0.80
														(0.39–0.54)		
rs9277542	6	33163225	HLA-DPB1	0.45	T/C	GWAS	18	53	110	29	102	52	0.073	0.42	5.32×10^{-8}	
				(T)		(Japanese)	(9.9)	(29.3)	(60.8)	(15.8)	(55.7)	(28.4)		(0.31–0.58)		
						Replication-1	30	106	118	54	114	67	0.681	0.54	3.33×10^{-6}	
						(Japanese)	(11.8)	(41.7)	(46.5)	(23.0)	(48.5)	(28.5)		(0.42–0.70)		
						Replication-2	30	87	94	35	72	36	0.933	0.54	8.29×10^{-5}	
						(Korean)	(14.2)	(41.2)	(44.5)	(24.5)	(50.3)	(25.2)		(0.40–0.74)		
						Meta-analysis[e]								0.50	1.28×10^{-15}	0.40
														(0.43–0.60)		

[a]Minor allele frequency and minor allele in 198 healthy Japanese (ref#19).
[b]Odds ratio of minor allele from two-by-two allele frequency table.
[c]P value of Pearson's chi-square test for allelic model.
[d]Heterogeneity was tested using general variance-based method.
[e]Meta-analysis was tested using the random effects model.
doi:10.1371/journal.pone.0039175.t002

comparison of HBV patients with HBV resolved individuals (OR = 0.56 for rs3077, 0.66 for rs9277535 and 0.55 for rs9277542) than in the one with healthy controls (OR = 0.55 for rs3077, 0.61 for rs9277535 and 0.51 for rs9277542), which was the opposite result as we expected (Table S2 and Table S4). These results may suggest that other unknown immune system(s) exist to eliminate the HBV in the HBV resolved individuals.

Among the HLA class II loci (*HLA-DPA1*, *HLA-DPB1* and *HLA-DQB2*), which were associated with CHB and HBV clearance, a weak linkage disequilibrium ($r^2<0.1$) was observed between *HLA-DQB2* locus and *HLA-DPA1/−DPB1* loci in Japanese and Korean populations (Figure S2). We also found that similar linkage disequilibrium blocks (r^2) were observed among three subgroups (HBV carriers, HBV resolved individuals and Healthy controls). Moreover, logistic regression analysis of *HLA-DP* (rs3077 and rs92775542) with use of *HLA-DQ* (rs9276431 and rs768538) as covariates showed that the same level of significant associations of *HLA-DP* with CHB and HBV clearance as shown in the single-point association analysis, while no associations of *HLA-DQ* with $P_{log}>0.05$ were detected both in Japanese and in Korean (Table S5). These results show that *HLA-DP* is the main genetic factor for susceptibility to CHB and HBV clearance, and the associations of *HLA-DQB2* would result from linkage disequilibrium of *HLA-DP*A1/−DPB1.

In this study, we confirmed the significant associations between *HLA-DPA1* and *HLA-DPB1*, and protective effects against CHB and HBV clearance in Japanese and Korean individuals. These results suggest that the associations between the *HLA-DP* locus, CHB and HBV clearance are widely replicated in East Asian populations, including Chinese, Thai, Japanese and Korean individuals; however, there have been no similar GWAS performed in Caucasian and African populations. Moreover, there were no significant SNPs associated with HCC development in this study, thus suggesting that it is necessary to increase the sample size. To clarify the pathogenesis of CHB or the mechanisms of HBV clearance, further studies are necessary, including a functional study of the *HLA-DP* molecule, identification of novel host genetic factors other than *HLA-DP*, and variation analysis of HBV.

Materials and Methods

Ethics Statement

All study protocols conform to the relevant ethical guidelines, as reflected in the *a priori* approval by the ethics committees of all participating universities and hospitals. The written informed consent was obtained from each patient who participated in this study and all samples were anonymized.

Genomic DNA Samples and Clinical Data

All of the 1,793 Japanese and Korean samples, including individuals with CHB, healthy controls and HBV-resolved individuals (HBsAg-negative and anti-HBc-positive), were collected at 20 multi-center hospitals (liver units with hepatologists) throughout Japan and Korea. The 19 hospitals in Japan were grouped into the following 8 areas: Hokkaido area (Hokkaido University Hospital, Teine Keijinkai Hospital), Tohoku area (Iwate Medical University Hospital), Kanto area (Musashino Red Cross Hospital, Saitama Medical University, Kitasato University Hospital, University of Tokyo), Koshin area (Shinshu University Hospital, Kanazawa University Hospital), Tokai area (Nagoya City University Hospital, Nagoya Daini Red Cross Hospital), Kinki area (Kyoto Prefectural University of Medicine Hospital, National Hospital Organization Osaka National Hospital, Osaka

Table 3. Results of replication study for clearance of hepatitis B virus.

	Position			MAF[a]	Allele	Stage	HBV carriers			Resolved individuals			OR[b]		
dbSNP rsID	Chr	Buld 36.3	Nearest Gene	(allele)	(1/2)	(population)	11	12	22	11	12	22	95% CI	*P*-value[c]	P_{het}[d]
rs3077	6	33141000	HLA-DPA1	0.44	T/C	GWAS	13	51	117	29	82	74	0.44	9.24×10^{-7}	
				(T)		(Japanese)	(7.2)	(28.2)	(64.6)	(15.7)	(44.3)	(40.0)	(0.32–0.61)		
						Replication-1	26	95	134	20	64	60	0.72	3.32×10^{-2}	
						(Japanese)	(10.2)	(37.3)	(52.5)	(13.9)	(44.4)	(41.7)	(0.53–0.97)		
						Replication-2	23	81	111	29	48	28	0.41	2.35×10^{-7}	
						(Korean)	(10.7)	(37.7)	(51.6)	(27.6)	(45.7)	(26.7)	(0.29–0.58)		
						Meta-analysis[e]							0.51	1.56×10^{-4}	0.03
													(0.36–0.72)		
						Meta-analysis[e]							0.43	1.89×10^{-12}	0.75
						(GWAS+replication-2)							(0.34–0.54)		
rs9277542	6	33163225	HLA-DPB1	0.45	T/C	GWAS	18	53	110	28	88	69	0.51	3.15×10^{-5}	
				(T)		(Japanese)	(9.9)	(29.3)	(60.8)	(15.1)	(47.6)	(37.3)	(0.37–0.70)		
						Replication-1	30	106	118	28	62	52	0.68	1.25×10^{-2}	
						(Japanese)	(11.8)	(41.7)	(46.5)	(19.7)	(43.7)	(36.6)	(0.51–0.92)		
						Replication-2	30	87	94	30	53	22	0.46	4.97×10^{-6}	
						(Korean)	(14.2)	(41.2)	(44.5)	(28.6)	(50.5)	(21.0)	(0.33–0.64)		
						Meta-analysis[e]							0.55	5.91×10^{-7}	0.19
													(0.43–0.69)		
						Meta-analysis[e]							0.49	9.69×10^{-10}	0.65
						(GWAS+replication-2)							(0.39–0.61)		

[a]Minor allele frequency and minor allele in 198 healthy Japanese (ref#19).
[b]Odds ratio of minor allele from two-by-two allele frequency table.
[c]P value of Pearson's chi-square test for allelic model.
[d]Heterogeneity was tested using general variance-based method.
[e]Meta-analysis was tested using the random effects model.
doi:10.1371/journal.pone.0039175.t003

City University), Chugoku/Shikoku area (Tottori University Hospital, Ehime University Hospital, Yamaguchi University Hospital, Kawasaki Medical College Hospital) and Kyushu area (Kurume University Hospital). Korean samples were collected at Yonsei University College of Medicine.

HBV status was measured based on serological results for HBsAg and anti-HBc with a fully automated chemiluminescent enzyme immunoassay system (Abbott ARCHITECT; Abbott Japan, Tokyo, Japan, or LUMIPULSE f or G1200; Fujirebio, Inc., Tokyo, Japan). For clinical staging, inactive carrier (IC) state was defined by the presence of HBsAg with normal ALT levels over 1 year (examined at least four times at 3-month intervals) and without evidence of portal hypertension. Chronic hepatitis (CH) was defined by elevated ALT levels (>1.5 times the upper limit of normal [35 IU/L]) persisting over 6 months (at least by 3 bimonthly tests). Liver cirrhosis (LC) was diagnosed principally by ultrasonography (coarse liver architecture, nodular liver surface, blunt liver edges and hypersplenism), platelet counts <100,000/cm^3, or a combination thereof. Histological confirmation by fine-needle biopsy of the liver was performed as required. Hepatocellular carcinoma (HCC) was diagnosed by ultrasonography, computerized tomography, magnetic resonance imaging, angiography, tumor biopsy or a combination thereof.

The Japanese control samples from HBV-resolved subjects (HBsAg-negative and anti-HBc-positive) at Nagoya City University-affiliated healthcare center were used by comprehensive agreement (anonymization in an unlinkable manner) in this study. Some of the unrelated Japanese healthy controls were obtained from the Japan Health Science Research Resources Bank (Osaka, Japan). One microgram of purified genomic DNA was dissolved in 100 µl of TE buffer (pH 8.0) (Wako, Osaka, Japan), followed by storage at −20°C until use.

SNP Genotyping and Data Cleaning

For GWAS, we genotyped a total of 550 individuals, including 181 Japanese HBV carriers, 184 Japanese healthy controls and 185 spontaneously HBV-resolved Japanese individuals (HBsAg-negative and anti-HBc-positive), using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Inc., Santa Clara, CA), in accordance with the manufacturer's instructions. The average QC call rate for 550 samples reached 98.47% (95.00–99.92%), which had an average sample call rate of 98.91% (93.55–99.74%) by determining the genotype calls of over 900 K SNPs using the Genotyping Console v4.1 software (with Birdseed v1 algorithm) provided by the manufacturer [19]. We then applied the following thresholds for SNP quality control in data cleaning: SNP call rate ≥95% and MAF ≥1% for three groups (HBV carriers, healthy controls and HBV-resolved individuals), and HWE *P*-value ≥0.001 for healthy controls [20]. Here, SNP call rate is defined for each SNP as the number of successfully genotyped samples divided by the number of total samples genotyped. A total of 597,789 SNPs and 590,278 SNPs on autosomal chromosomes

passed the quality control filters in the genome-wide association analysis using HBV carriers and healthy controls, and using HBV carriers and HBV-resolved individuals, respectively (Figure 1). All cluster plots for the SNPs showing $P<0.0001$ on association analyses in the allele frequency model were confirmed by visual inspection, and SNPs with ambiguous cluster plots were excluded.

In the following replication stage, we selected a set of 32 SNPs with $P<0.0001$ in the GWAS using HBV carriers and HBV-resolved individuals. SNP genotyping in two independent sets of 256 Japanese HBV carriers, 236 Japanese healthy controls and 150 Japanese HBV-resolved individuals (Table 1, replication-1), and 344 Korean HBV carriers, 151 Korean healthy controls and 106 Korean HBV-resolved individuals (Table 1, replication-2) was completed for the selected 32 SNPs using the DigiTag2 assay [21,22] and custom TaqMan SNP Genotyping Assays (Applied Biosystems, Foster City, CA) on the LightCycler 480 Real-Time PCR System (Roche, Mannheim, Germany).

Statistical Analysis

The observed associations between SNPs and the protective effects on chronic hepatitis B or clearance of hepatitis virus B were assessed by chi-squared test with a two-by-two contingency table in allele frequency model. SNPs on chromosome X were removed because gender was not matched among HBV carriers, healthy controls and HBV-resolved individuals. A total of 597,789 SNPs and 590,278 SNPs passed the quality control filters in the GWAS stage; therefore, significance levels after Bonferroni correction for multiple testing were $P=8.36\times10^{-8}$ (0.05/597,789) and $P=8.47\times10^{-8}$ (0.05/590,278), respectively. For the replication study, 29 of 32 SNPs were successfully genotyped; therefore, we applied $P=0.0017$ (0.05/29) as a significance level, and none of the 29 markers genotyped in the replication stage showed deviations from the Hardy-Weinberg equilibrium in healthy controls ($P>0.01$).

The genetic inflation factor λ was estimated by applying the Cochrane-Armitage test on all SNPs and was found to be 1.056 and 1.030 in the GWAS using HBV carriers and healthy controls, and using HBV carriers and HBV-resolved individuals, respectively (Figure S3). These results suggest that the population substructure should not have any substantial effect on statistical analysis. In addition, the principal component analysis in a total of 550 individuals in the GWAS stage together with the HapMap samples also revealed that the effect of population stratification was negligible (Figure S4).

Based on the genotype data of a total of 1,793 samples including 1,192 Japanese samples and 601 Korean samples in both GWAS and replication stages, haplotype blocks were estimated using the Gabriel's algorithm using the Haploview software (v4.2) (Figure S2). In the logistic regression analysis, two SNPs (rs9276431 and rs7768538) within the HLA-DQ locus were individually involved as a covariate (Table S5). Statistical analyses were performed using the SNP & Variation Suite 7 software (Golden Helix, MT, USA).

Supporting Information

Figure S1 GWAS using samples from HBV carriers with LC or HCC, and HBV carriers without LC and HCC. *P* values were calculated using chi-squared test for allele frequencies.
(PPTX)

Figure S2 Estimation of linkage disequilibrium blocks in HBV patients, HBV resolved individuals and healthy controls in Japanese and Korean. The LD blocks (r^2) were analyzed using the Gabriel's algorithm.
(PPTX)

Figure S3 Quantile-quantile plot for test statistics (allele-based chi-squared tests) for GWAS results. Dots represent P values of each SNP that passed the quality control filters. Inflation factor λ was estimated to be: a) 1.056 in the analysis with HBV carriers and healthy controls; and b) 1.030 with HBV carriers and HBV-resolved individuals.
(PPTX)

Figure S4 Principal component analysis on a total of 550 individuals in GWAS, together with HapMap samples (CEU, YRI and JPT).
(PPTX)

Table S1 Results for 29 SNPs selected in replication study using samples of HBV carriers and healthy controls. [a]*P* values by chi-squared test for allelic model. [b]Odds ratio of minor allele from two-by-two allele frequency table. [c]Meta-analysis was tested using additive, two-tailed CMH fixed-effects model.
(XLSX)

Table S2 Results of meta-analysis for protective effects against persistent HB infection across 6 independent studies, including this study. [a]Minor allele frequency and minor allele in 198 healthy Japanese (ref#19). [b]Odds ratio of minor allele from two-by-two allele frequency table. [c]*P* value of Pearson's chi-squared test for allele model. [d]Heterogeneity was tested using general variance-based method. [e]Meta-analysis was tested using the random effects model.
(XLSX)

Table S3 Results for 29 SNPs selected in replication study using samples from HBV carriers and HBV-resolved individuals. [a]*P* values by chi-squared test for allelic model. [b]Odds ratio of minor allele from two-by-two allele frequency table. [c]Meta-analysis was tested using additive, two-tailed CMH fixed-effects model.
(XLSX)

Table S4 Results of meta-analysis for clearance of HBV across 6 independent studies, including this study. [a]Minor allele frequency and minor allele in 198 healthy Japanese (ref#19). [b]Odds ratio of minor allele from two-by-two allele frequency table. [c]*P* value of Pearson's chi-squared test for allele model. [d]Heterogeneity was tested using general variance-based method. [e]Meta-analysis was tested using the random effects model.
(XLSX)

Table S5 Logistic regression analysis of *HLA-DP* (rs3077 and rs9277542) and *HLA-DQ* (rs9276431 and rs7768538) with susceptibility to CHB and HBV clearance using the *HLA-DQ* genotypes individually as a covariate.
(XLSX)

Acknowledgments

We thank all the patients and families who contributed to the study and Ms. Yasuka Uehara-Shibata and Ms. Yoshimi Ishibashi for technical assistance.

Author Contributions

Conceived and designed the experiments: NN HS YT. Performed the experiments: HS Y. Mawatari M. Sageshima YO. Analyzed the data: NN MK AK. Contributed reagents/materials/analysis tools: KM M. Sugiyama SHA JYP SH JHK KS M. Kurosaki YA SM MW ET MH SK EO YI EM AT Y. Murawaki YH IS M. Korenaga KH TI NI KHH YT MM. Wrote the paper: NN M. Kawashima YT KT MM.

References

1. Arauz-Ruiz P, Norder H, Robertson BH, Magnius LO (2002) Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. J Gen Virol 83: 2059–2073.
2. Hoofnagle JH, Doo E, Liang TJ, Fleischer R, Lok ASF (2007) Management of hepatitis B: Summary of a clinical research workshop. Hepatology 45: 1056–1075.
3. Yokosuka O, Kurosaki M, Imazeki F, Arase Y, Tanaka Y, et al. (2011) Management of hepatitis B: Consensus of the Japan society of Hepatology 2009. Hepatol Res 41: 1–21.
4. Tada H, Uga N, Fuse Y, Shimizu M, Nemoto Y, et al. (1992) Prevention of perinatal transmission of hepatitis B virus carrier state. Acta Paediatr Jpn 34: 656–659.
5. Stevens CE, Toy PT, Taylor PE, Lee T, Yip HY (1992) Prospects for control of hepatitis B virus infection: implications of childhood vaccination and long-term protection. Pediatrics 90: 170–173.
6. Szmuness W (1979) Large-scale efficacy trials of hepatitis B vaccines in the USA: baseline data and protocols. J Med Virol 4: 327–340.
7. Kwon H, Lok AS (2011) Hepatitis B therapy. Nat Rev Gastroenterol Hepatol 8: 275–284.
8. Kobayashi M, Ikeda K, Arase Y, Suzuki F, Akuta N, et al. (2008) Change of Hepatitis B virus genotypes in acute and chronic infections in Japan. J Med Virol 80: 1880–1884.
9. Yano K, Tamada Y, Yatsuhashi H, Komori A, Abiru S, et al. (2010) Dynamic epidemiology of acute viral hepatitis in Japan. Intervirology 53: 70–75.
10. 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. Nat Genet 41: 591–595.
11. Mbarek H, Ochi H, Urabe Y, Kumar V, Kubo M, et al. (2011) A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population. Hum Mol Genet 20: 3884–3892.
12. 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: e24221.
13. 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. Hepatol 53: 422–428.
14. Thursz MR, Kwiatkowski D, Allsopp CEM, 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.
15. 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.
16. An P, Winkler C, Guan L, O'Brien SJ, Zeng Z, et al. (2011) A common HLA-DPA1 variant is a major determinant of hepatitis B virus clearance in Han Chinese. J Infect Dis 203: 943–947.
17. Wang L, Wu X-P, Zhang W, Zhu D-H, Wang Y, et al. (2011) Evaluation of genetic susceptibility loci for chronic hepatitis B in Chinese: two independent case-control study.
18. Hu L, Zhai X, Liu J, Chu M, Pan S, et al. (2011) Genetic variants in HLA-DP/DQ influence both hepatitis B virus clearance and Hepatocellular carcinoma development. Hepatology (in press).
19. Nishida N, Koike A, Tajima A, Ogasawara Y, Ishibashi Y, et al. (2008) Evaluating the performance of Affymetrix SNP Array 6.0 platform. BMC Genomics 9: 431.
20. Miyagawa T, Nishida N, Ohashi J, Kimura R, Fujimoto A, et al. (2008) Appropriate data cleaning methods for genome-wide association study. J Hum Genet 53: 886–893.
21. Nishida N, Tanabe T, Takasu M, Suyama A, Tokunaga K (2007) Further development of multiplex single nucleotide polymorphism typing method, the DigiTag2 assay. Anal Biochem 364: 78–85.
22. Nishida N, Mawatari Y, Sageshima M, Tokunaga K (2012) Highly parallel and short-acting amplification with locus-specific primers to detect single nucleotide polymorphisms by the DigiTag2 assay. PLoS ONE 7: e29967.

OPEN ACCESS Freely available online

PLOS ONE

Soluble MICA and a *MICA* Variation as Possible Prognostic Biomarkers for HBV-Induced Hepatocellular Carcinoma

Vinod Kumar[1,2]*, Paulisally Hau Yi Lo[1], Hiromi Sawai[3], Naoya Kato[4], Atsushi Takahashi[2], Zhenzhong Deng[1], Yuji Urabe[1], Hamdi Mbarek[1], Katsushi Tokunaga[3], Yasuhito Tanaka[5], Masaya Sugiyama[6], Masashi Mizokami[6], Ryosuke Muroyama[4], Ryosuke Tateishi[7], Masao Omata[7], Kazuhiko Koike[7], Chizu Tanikawa[1], Naoyuki Kamatani[2], Michiaki Kubo[2], Yusuke Nakamura[1], Koichi Matsuda[1]

1 Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, **2** Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan, **3** Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, **4** Unit of Disease Control Genome Medicine, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, **5** Department of Clinical Molecular Informative Medicine, Nagoya City University Graduate School of Medical Sciences, Aichi, Japan, **6** The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Chiba, Japan, **7** Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Abstract

MHC class I polypeptide-related chain A (MICA) molecule is induced in response to viral infection and various types of stress. We recently reported that a single nucleotide polymorphism (SNP) rs2596542 located in the *MICA* promoter region was significantly associated with the risk for hepatitis C virus (HCV)-induced hepatocellular carcinoma (HCC) and also with serum levels of soluble MICA (sMICA). In this study, we focused on the possible involvement of MICA in liver carcinogenesis related to hepatitis B virus (HBV) infection and examined correlation between the *MICA* polymorphism and the serum sMICA levels in HBV-induced HCC patients. The genetic association analysis revealed a nominal association with an SNP rs2596542; a G allele was considered to increase the risk of HBV-induced HCC ($P = 0.029$ with odds ratio of 1.19). We also found a significant elevation of sMICA in HBV-induced HCC cases. Moreover, a G allele of SNP rs2596542 was significantly associated with increased sMICA levels ($P = 0.009$). Interestingly, HCC patients with the high serum level of sMICA (>5 pg/ml) exhibited poorer prognosis than those with the low serum level of sMICA (≤5 pg/ml) ($P = 0.008$). Thus, our results highlight the importance of *MICA* genetic variations and the significance of sMICA as a predictive biomarker for HBV-induced HCC.

Citation: Kumar V, Yi Lo PH, Sawai H, Kato N, Takahashi A, et al. (2012) Soluble MICA and a *MICA* Variation as Possible Prognostic Biomarkers for HBV-Induced Hepatocellular Carcinoma. PLoS ONE 7(9): e44743. doi:10.1371/journal.pone.0044743

Editor: Erica Villa, University of Modena & Reggio Emilia, Italy

Received May 3, 2012; **Accepted** August 7, 2012; **Published** September 14, 2012

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. 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: koichima@ims.u-tokyo.ac.jp

Introduction

Hepatocellular carcinoma (HCC) reveals a very high mortality rate that is ranked the third among all cancers in the world [1]. HCC is known to develop in a multistep process which has been related to various risk factors such as genetic factors, environment toxins, alcohol and drug abuse, autoimmune disorders, elevated hepatic iron levels, obesity, and hepatotropic viral infections [2]. Among them, chronic infection with hepatitis B virus (HBV) is one of the major etiological factors for developing HCC with considerable regional variations ranging from 20% of HCC cases in Japan to 65% in China [3].

Interestingly, clinical outcome after the exposure to HBV considerably varies between individuals. The great majority of individuals infected with HBV spontaneously eliminate the viruses, but a subset of patients show the persistent chronic hepatitis B infection (CHB), and then progresses to liver cirrhosis and HCC through a complex interplay between multiple genetic and environmental factors [4]. In this regard, genome wide association studies (GWAS) using single nucleotide polymorphisms (SNPs) have highlighted the importance of genetic factors in the pathogenesis of various diseases including CHB as well as HBV-induced HCC [5,6,7,8,9,10,11,12,13]. Recently, we identified a genetic variant located at 4.7 kb upstream of the *MHC class I polypeptide-related chain A* (*MICA*) gene to be strongly associated with hepatitis C virus (HCV) -induced HCC development [14].

MICA is highly expressed on viral-infected cells or cancer cells, and acts as ligand for NKG2D to activate antitumor effects of Natural killer (NK) cells and $CD8^+$ T cells [15,16]. Our previous results indicated that a G allele of SNP rs2596542 was significantly associated with the lower cancer risk and the higher level of soluble MICA (sMICA) in the serum of HCV-induced HCC patients, demonstrating the possible role of MICA as a tumor suppressor. However, elevation of serum sMICA was shown to be associated with poor prognosis in various cancer patients [17,18,19,20].

Matrix metalloproteinases (MMPs) can cleave MICA at a transmembrane domain [21] and release sMICA proteins from cells. Since sMICA was shown to inhibit the antitumor effects of NK cells and $CD8^+$ T cells by reduction of their affinity to binding to target cells [22,23], the effect of MICA in cancer cells would be modulated by the expression of MMPs. To elucidate the role of MICA in HBV-induced hepatocellular carcinogenesis, we here report analysis of the *MICA* polymorphism and serum sMICA level in HBV-induced HCC cases.

Materials and Methods

Study participants

The demographic details of study participants are summarized in Table 1. A total of 181 HCC cases, 597 CHB patients, and 4,549 non-HBV controls were obtained from BioBank Japan that was initiated in 2003 with the funding from the Ministry of Education, Culture, Sports, Science and Technology, Japan [24]. In the Biobank Japan Project, DNA and serum of patients with 47 diseases were collected through collaborating network of 66 hospitals throughout Japan. List of participating hospitals is shown in the following website (http://biobankjp.org/plan/member_hospital.html). A total of 226 HCC cases, 102 CHB patients, and 174 healthy controls were additionally obtained from the University of Tokyo. The diagnosis of chronic hepatitis B was conducted on the basis of HBsAg-seropositivity and elevated serum aminotransferase levels for more than six months according to the guideline for diagnosis and treatment of chronic hepatitis (The Japan Society of Hepatology, http://www.jsh.or.jp/medical/gudelines/index.html). Control Japanese DNA samples (n = 934) were obtained from Osaka-Midosuji Rotary Club, Osaka, Japan. All HCC patients were histopathologically diagnosed. Overall survival was defined as the time from blood sampling for sMICA test to the date of death due to HCC. Patients who were alive on the date of last follow-up were censored on that date. All participants provided written informed consent. This research project was approved by the ethics committee of the University of Tokyo and the ethics committee of RIKEN. All clinical assessments and specimen collections were conducted according to Declaration of Helsinki principles.

SNP genotyping

Genotyping platforms used in this study were shown in Table 1. We genotyped 181 HCC cases and 5,483 non-HBV control samples using either Illumina Human Hap610-Quad or Human Hap550v3. The other samples were genotyped at SNP rs2596542 by the Invader assay system (Third Wave Technologies, Madison, WI).

MICA variable number tandem repeat (VNTR) locus genotyping

Genotyping of the *MICA* VNTR locus in 176 HBV-induced HCC samples was performed using the primers reported previously by the method recommended by Applied Biosystems (Foster City, CA) [14]. Briefly, the 5′ end of forward primer was labeled with 6-FAM, and reverse primer was modified with GTGTCTT non-random sequence at the 5′ end to promote Plus A addition. The PCR products were mixed with Hi-Di Formamide and GeneScan-600 LIZ size standard, and separated by GeneScan system on a 3730×l DNA analyzer (Applied Biosystems, Foster City, CA). GeneMapper software (Applied Biosystems, Foster City, CA) was employed to assign the repeat fragment size (Figure S1).

Quantification of soluble MICA

We obtained serum samples of 111 HBV-positive HCC samples, 129 HCV-positive HCC samples, and 60 non-HBV controls from Biobank Japan. Soluble MICA levels were measured by sandwich enzyme-linked immunosorbent assay, as described in the manufacturer's instructions (R&D Systems, Minneapolis, MN).

Statistical analysis

The association between an SNP rs2596542 and HBV-induced HCC was tested by Cochran-Armitage trend test. The Odds ratios were calculated by considering a major allele as a reference. Statistical comparisons between genotypes and sMICA levels were performed by Kruskal-Wallis test (if more than two classes for comparison) or Wilcoxon rank test using R. Overall survival rate of the patients was analyzed by Kaplan-Meier method in combination with log-rank test with SPSS 20 software. The period for the survival analysis was calculated from the date of blood sampling to the recorded date of death or the last follow-up date. Differences with a P value of <0.05 were considered statistically significant.

Results

Association of SNP rs2596542 with HBV-induced HCC

In order to examine the effect of rs2596542 genotypes on the susceptibility to HBV-induced HCC, a total of 407 HCC cases and 5,657 healthy controls were genotyped. The Cochran Armitage trend test of the data revealed a nominal association

Table 1. Demographic details of subjects analyzed.

Subjects	Source	Genotyping platform	Number of Sample	Female (%)	Age (mean+/−sd)
Liver Cancer	BioBank Japan	Illumina Human Hap610-Quad	181	17.9	62.94±9.42
	University of Tokyo	Invader assay	226		
Control	BioBank Japan	Illumina Human Hap550v3	4549	47.95	55.19±12.5
	Osaka**	Illumina Human Hap550v3	934		
	University of Tokyo	Invader assay	174		
Chronic hepatitis B*	BioBank Japan	Invader assay	597	45.66	61.31±12.6
	University of Tokyo	Invader assay	102		

*Chronic hepatitis B patients without liver cirrhosis and liver cancer during enrollment.
**Healthy volunteers from Osaka Midosuji Rotary Club, Osaka, Japan.
doi:10.1371/journal.pone.0044743.t001

between HBV-induced HCC and rs2596542 in which a risk allele G was more frequent among HBV-induced HCC cases than an A allele (P = 0.029, OR = 1.19, 95% CI: 1.02–1.4; Table 2). To further investigate the effect of rs2596542 on the progression from CHB to HBV-induced HCC, we genotyped a total of 699 CHB cases without HCC. Although the progression risk from CHB to HBV-induced HCC was not statistically significant with rs2596542 (P = 0.197 by the Cochran Armitage trend test with an allelic OR = 1.3 (0.94–1.36); Table 2), we found a similar trend of association in which the frequency of a risk-allele G was higher among HBV-induced HCC patients than that of CHB subjects. Since we previously revealed that an A allele was associated with a higher risk of HCV-induced HCC with OR of 1.36 [14], the s2596542 alleles that increased the risk of HCC were opposite in HBV-induced HCC and HCV-induced HCC.

Soluble MICA levels are associated with SNP rs2596542

We subsequently performed measurement of soluble MICA (sMICA) in serum samples using the ELISA method in 176 HBV-positive HCC cases and 60 non-HBV controls. Nearly 30% of the HBV-induced HCC cases revealed the serum sMICA level of >5 pg/ml (defined as high) while the all control individuals except one showed that of ≤5 pg/ml (defined as low) ($P = 4.5 \times 10^{-6}$; Figure 1A). Then, we examined correlation between SNP rs2596542 genotypes and serum sMICA levels in HBV-positive HCC cases. Interestingly, rs2596542 genotypes were significantly associated with serum sMICA levels (P = 0.009; Figure 1B); 39% of individuals with the GG genotype and 20% of those with the AG genotype were classified as high for serum sMICA, but only 11% of those with the AA genotype were classified as high (AA+AG vs GG; P = 0.003) (Figure 1B). These findings were similar with our previous reports in which a G allele was associated with higher serum sMICA levels in HCV-induced HCC patients [14].

Negative association of variable number of tandem repeat (VNTR) with sMICA level

The *MICA* gene harbors a VNTR locus in exon 5 that consists of 4, 5, 6, or 9 repeats of GCT as well as a G nucleotide insertion into a five-repeat allele (referred as A4, A5, A6, A9, and A5.1, respectively). The insertion of G (A5.1) causes a premature translation termination and results in loss of a transmembrane domain, which may produce the shorter form of the MICA protein that is likely be secreted into serum [25]. However, the association of this VNTR locus with serum sMICA level was controversial among studies [14,26,27,28]. Therefore, we examined the association between the VNTR locus and sMICA level in HBV-induced HCC patients, and found no significant association (Figure S1 and S2), concordant with our previous report for HCV-induced HCC patients [14].

Soluble MICA levels are associated with survival of HCC patients

In order to evaluate the prognostic significance of serum sMICA levels in HCC patients, we performed survival analysis of HCC patients. A total of 111 HBV-infected HCC patients and 129 HCV-infected HCC patients were included in this analysis. The mean survival period for HBV- and HCV-infected patients with less than 5 pg/ml of serum sMICA were 67.1 months (95% CI: 61.1–73.1, n = 83), and 58.2 months (95% CI: 51.4–65.0, n = 85), respectively. On the other hand, for patients with more than 5 pg/ml of serum sMICA, the mean survival periods were 47.8 months (95% CI: 34.8–30.9, n = 28) for HBV-induced HCC patients and 59.5 months (95% CI: 51.9–67.1, n = 44) for HCV-induced HCC patients. The Kaplan-Maier analysis and log-rank test indicated that among HBV-induced HCC subjects, the patients in the high serum sMICA group showed a significantly shorter survival than those in the low serum sMICA (P = 0.008; Figure 2). In addition, we performed multi-variate analysis to test whether sMICA is an independent prognostic factor by including age and gender as covariates. The results revealed significant association of sMICA levels with overall survival (P = 0.017) but not with age and gender (Table S1).However, we found no association between the serum sMICA level and the overall survival in the HCV-induced HCC subjects (P = 0.414; Figure S3). Taken together, our findings imply the distinct roles of the *MICA* variation and sMICA between HBV- and HCV-induced hepatocellular carcinogenesis.

Vascular invasion in HBV-related HCC patients is associated with soluble MICA levels

Since sMICA levels were associated with the overall survival of HBV-related HCC patients, we tested whether sMICA levels affect survival through modulating invasive properties of tumors or size of the tumors. We tested the association between sMICA levels and vascular invasion in 35 HBV-related HCC cases, among whom 7 cases were positive and 21 cases were negative for vascular invasion. We found significant association between sMICA levels and vascular invasion (Figure 3; P = 0.014) in which 7 cases with positive vascular invasion showed high levels of sMICA (mean = 54 pg/ml) than 21 cases without vascular invasion (mean = 7.51 pg/ml). However, we found no association between tumor size and sMICA levels (P = 0.56; data not shown). These results suggest that sMICA may reduce the survival of HBV-related HCC patients by affecting the invasive properties of tumors.

Discussion

Several mechanisms such as HBV-genome integration into host chromosomal DNA [29] and effects of viral proteins including HBx [30] are shown to contribute to development and progression of HCC, while the immune cells such as NK and T cells function as key antiviral and antitumor effectors. MICA protein has been

Table 2. Association between HCC and rs2596542.

SNP	Comparison	Chr	Locus	Case MAF	Control MAF	*P**	OR*	95% CI
rs2596542	HCC vs. Healthy control	6	*MICA*	0.294	0.332	0.029	1.19	1.02–1.4
rs2596542	HCC vs. CHB	6	*MICA*	0.294	0.320	0.197	1.13	0.94–1.36

Note: 407 HCC cases, 699 CHB subjects and 5,657 non-HBV controls were used in the analysis.
Chr.,chromosome; MAF, minor allele frequency; OR, odds ratio for minor allele; CI, confidence interval.
*Obtained by Armitage trend test.
doi:10.1371/journal.pone.0044743.t002

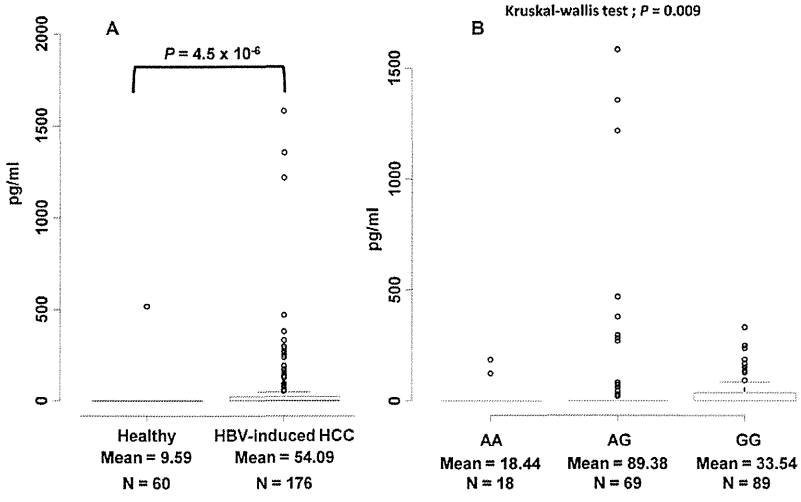

Figure 1. Soluble MICA levels are associated with HBV-related HCC. (A) Correlation between soluble MICA levels and HBV-induced HCC subjects. The y-axis displays the concentration of soluble MICA in pg/ml. The number of independent samples tested in each group is shown in the x-axis. Each group is shown as a box plot and the mean values are shown in the x-axis. The difference between two groups is tested by Wilcoxon rank test. The box plots are plotted using default settings in R. (B) Correlation between soluble MICA levels and rs2596542 genotype in HBV-positive HCC subjects. The x-axis shows the genotypes at rs2596542 and y-axis display the concentration of soluble MICA in pg/ml. Each group is shown as a box plot. $P = 0.027$ and 0.013 for AA vs. GG and AA vs. AG, respectively. The association between genotypes and sMICA levels was tested by Kruskal-wallis test, whereas the difference in the sMICA levels between AA and GG is tested by Wilcoxon rank test. The box plots are plotted using default settings in R.
doi:10.1371/journal.pone.0044743.g001

considered as a stress marker of gastrointestinal epithelial cells because of its induced expression by several external stimuli such as heat, DNA damage, and viral infections [31,32,33,34]. Here, we examined the association of rs2596542 and serum sMICA levels with HBV-induced HCC. Like in HCV-induced HCC [14], our results from ELISA revealed a significantly higher proportion

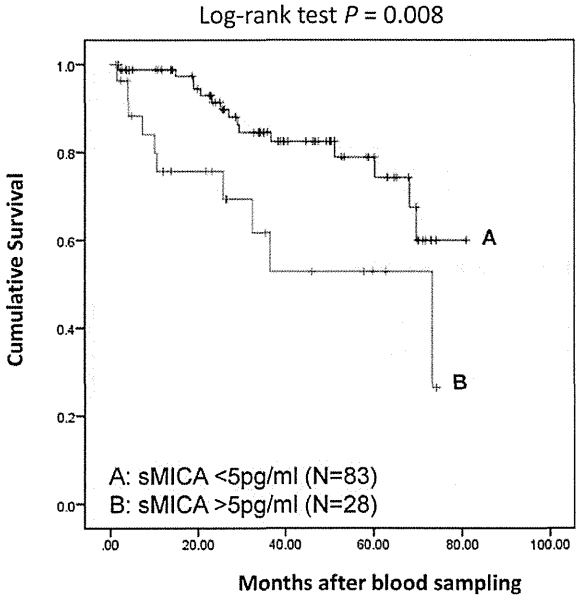

Figure 2. Kaplan-Meier curves of the patients with HBV-induced HCC. The patients were divided into two groups according to their sMICA concentration (high: >5 pg/ml and low: =<5 pg/ml). Statistical difference was analyzed by log-rank test. The y-axis shows the cumulative survival probability and x-axis display the months of the patients survival after blood sampling.
doi:10.1371/journal.pone.0044743.g002

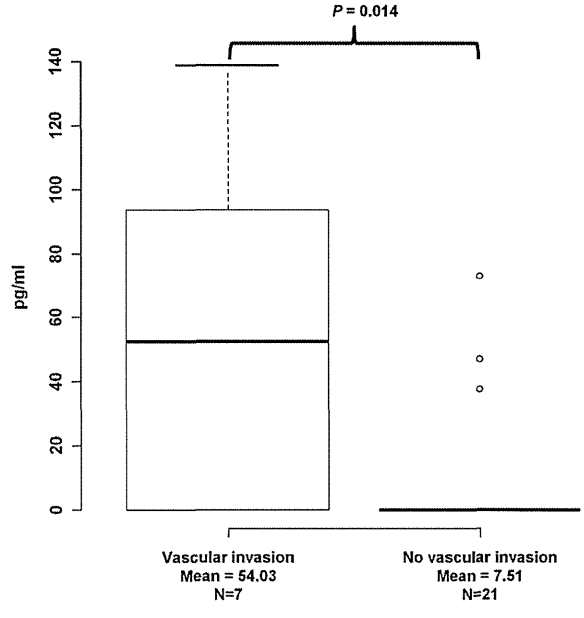

Figure 3. Correlation between soluble MICA levels and vascular invasion in HBV-induced HCC subjects. The y-axis displays the concentration of soluble MICA in pg/ml. The number of independent samples tested in each group is shown in the x-axis. Each group is shown as a box plot and the mean values are shown in the x-axis. The difference between two groups is tested by Wilcoxon rank test. The box plots are plotted using default settings in R.
doi:10.1371/journal.pone.0044743.g003

of high serum sMICA cases (nearly 30%) in the HBV-induced HCC group, compared to non-HBV individuals (1.7%). Moreover, the serum sMICA level was significantly associated with rs2596542, but not with the copy number differences of the VNTR locus, as concordant with our previous report [14].

Several studies have already indicated the roles of sMICA as prognostic markers for different types of malignant diseases [17,18,19,20]. Therefore, it is of medical importance to test whether serum sMICA levels can be used as a prognostic marker for patients with HCC. To our best knowledge, this is the first study to demonstrate the prognostic potential of sMICA for HBV-positive HCC patients; we found 19.3 months of improvement in survival among patients carrying less than 5 pg/ml of serum sMICA, compared to those having more than 5 pg/ml.

On the contrary, we found no significant correlation between sMICA levels and the prognosis of HCV-induced HCC cases. These opposite effects of *MICA* variation could be explained by the following mechanism. The individuals who carry the G allele would express high levels of membrane-bound MICA upon HCV infection and thus lead to the activation of immune cells against virus infected cells. On one hand, HBV infection results in increased expression of membrane-bound MICA as well as MMPs through viral protein HBx [35], which would result in the elevation of sMICA and the reduction of membrane-bound MICA. Since sMICA could block CD8+T cells, NK-CTL, and NK cells, higher sMICA would cause the inactivation of immune surveillance system against HBV infected cells. In other words, HBV may use this strategy to evade immune response and hence, higher levels of sMICA could be associated with lower survival rate among HBV-associated HCC. On the other hand, since HCV is not known to induce the cleavage of membrane bound MICA, individuals with low level membrane bound MICA expression (carriers of rs2596542-allele A) could be inherently susceptible for HCV-induced HCC. Thus, HBx-medicated induction of MMPs could partially explain the intriguing contradictory effect of MICA between HBV-induced HCC and HCV-induced HCC. Since we observed significant correlation of sMICA levels with vascular invasion, it may be the case that high levels of sMICA cause poor prognosis of HBV-related HCC cases by making tumors more aggressive and invasive. However it is important in future to determine the ratio of membrane-bound MICA to sMICA in case of HCV- and HBV-related HCC.

Interestingly, the immune therapy against melanoma patients induced the production of auto-antibodies against MICA [36]. Anti-MICA antibodies would exert antitumor effects through antibody-dependent cellular cytotoxicity against cells expressing membrane-bound MICA and/or activation of NK cells by inhibiting the sMICA-NKG2D interaction. However, further studies are necessary, using well-defined HBV-related HCC cohort, to investigate whether sMICA levels could be included as an additional factor to predict the survival rate among HBV-related HCC subjects. Taken together, our results indicate the potential of *MICA* variant and sMICA as prognostic biomarkers. Thus, MICA could be a useful therapeutic target for HBV-induced HCC.

Supporting Information

Figure S1 MICA repeat genotyping using capillary-based method. The alleles are annotated using GeneMapper software based on the size of the PCR product (185 bp = A4 allele, 188 bp = A5, 189 bp = A5.1, 191 bp = A6 and 200 bp = A9). The inset at the base of each peak shows the size of the PCR product with corresponding allele call by the software. The figure display all observed heterozygotes at A5.1 allele.
(TIF)

Figure S2 MICA VNTR alleles are not associated with soluble MICA levels. Each group is shown as a box plot. The difference in the sMICA values among each group is tested by Wilcoxon rank test. The box plots are plotted using default settings in R.
(TIF)

Figure S3 Kaplan-Meier curves of the patients with HCV-induced HCC. The patients were divided into two groups according to their sMICA concentration (<5 pg/ml or >5 pg/ml). Statistical difference was analyzed by log-rank test. The y-axis shows the cumulative survival probability and x-axis display the months of the patients survival after blood sampling.
(TIF)

Table S1 Clinical parameters of HBV-related HCC patients available for prognostic analyses.
(XLS)

Acknowledgments

We would like to thank all the patients 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 Ayako Matsui and Hiroe Tagaya (the University of Tokyo), and the technical staff of the Laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN for their technical support.

Author Contributions

Conceived and designed the experiments: VK KM YN. Performed the experiments: VK PHL YU HM ZD. Analyzed the data: VK PHL CT RM. Contributed reagents/materials/analysis tools: YN NK AT MK HS KT YT MS MM RT MO KK NK. Wrote the paper: VK PHL KM YN.

References

1. Kew MC (2010) Epidemiology of chronic hepatitis B virus infection, hepatocellular carcinoma, and hepatitis B virus-induced hepatocellular carcinoma. Pathol Biol (Paris) 58: 273–277.
2. Sherman M (2010) Hepatocellular carcinoma: epidemiology, surveillance, and diagnosis. Semin Liver Dis 30: 3–16.
3. Perz J, Armstrong G, Farrington L, Hutin Y, Bell B (2006) The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. J Hepatol 45: 529–538.
4. Chen CJ, Chen DS (2002) Interaction of hepatitis B virus, chemical carcinogen, and genetic susceptibility: multistage hepatocarcinogenesis with multifactorial etiology. Hepatology 36: 1046–1049.
5. Cui R, Okada Y, Jang SG, Ku JL, Park JG, et al. (2011) Common variant in 6q26–q27 is associated with distal colon cancer in an Asian population. Gut 60: 799–805.
6. Kumar V, Matsuo K, Takahashi A, Hosono N, Tsunoda T, et al. (2011) Common variants on 14q32 and 13q12 are associated with DLBCL susceptibility. J Hum Genet England. pp. 436–439.
7. Cui R, Kamatani Y, Takahashi A, Usami M, Hosono N, et al. (2009) Functional variants in ADH1B and ALDH2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. Gastroenterology 137: 1768–1775.
8. Urabe Y, Tanikawa C, Takahashi A, Okada Y, Morizono T, et al. (2012) A genome-wide association study of nephrolithiasis in the Japanese population identifies novel susceptible loci at 5q35.3, 7p14.3 and 13q14.1. PLOS Genet 8(3): e1002541.
9. Tanikawa C, Urabe Y, Matsuo K, Kubo M, Takahashi A, et al. (2012) A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population. Nat Genet 44(4): 430–434.

10. Hata J, Matsuda K, Ninomiya T, Yonemoto K, Matsushita T, et al. (2007) Functional SNP in an Sp1-binding site of AGTRL1 gene is associated with susceptibility to brain infarction. Hum Mol Genet 16: 630–639.
11. 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. Nat Genet 41: 591–595.
12. Mbarek H, Ochi H, Urabe Y, Kumar V, Kubo M, et al. (2011) A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population. Human Molecular Genetics 20: 3884–3892.
13. 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.
14. Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, et al. (2011) Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nature genetics 43: 455–458.
15. Jinushi M, Takehara T, Tatsumi T, Kanto T, Groh V, et al. (2003) Expression and role of MICA and MICB in human hepatocellular carcinomas and their regulation by retinoic acid. Int J Cancer 104: 354–361.
16. Bauer S, Groh V, Wu J, Steinle A, Phillips JH, et al. (1999) Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. Science 285: 727–729.
17. Holdenrieder S, Stieber P, Peterfi A, Nagel D, Steinle A, et al. (2006) Soluble MICA in malignant diseases. Int J Cancer 118: 684–687.
18. Nückel H, Switala M, Sellmann L, Horn PA, Dürig J, et al. (2010) The prognostic significance of soluble NKG2D ligands in B-cell chronic lymphocytic leukemia. Leukemia 24: 1152–1159.
19. Tamaki S, Sanefuzi N, Kawakami M, Aoki K, Imai Y, et al. (2008) Association between soluble MICA levels and disease stage IV oral squamous cell carcinoma in Japanese patients. Hum Immunol 69: 88–93.
20. Li K, Mandai M, Hamanishi J, Matsumura N, Suzuki A, et al. (2009) Clinical significance of the NKG2D ligands, MICA/B and ULBP2 in ovarian cancer: high expression of ULBP2 is an indicator of poor prognosis. Cancer Immunol Immunother 58: 641–652.
21. Salih H, Rammensee H, Steinle A (2002) Cutting edge: down-regulation of MICA on human tumors by proteolytic shedding. J Immunol 169: 4098–4102.
22. Groh V, Wu J, Yee C, Spies T (2002) Tumour-derived soluble MIC ligands impair expression of NKG2D and T-cell activation. Nature 419: 734–738.
23. Jinushi M, Takehara T, Tatsumi T, Hiramatsu N, Sakamori R, et al. (2005) 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.
24. Nakamura Y (2007) The BioBank Japan Project. Clin Adv Hematol Oncol 5: 696–697.
25. Ota M, Katsuyama Y, Mizuki N, Ando H, Furihata K, et al. (1997) 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.
26. Tamaki S, Sanefuzi N, Ohgi K, Imai Y, Kawakami M, et al. (2007) An association between the MICA-A5.1 allele and an increased susceptibility to oral squamous cell carcinoma in Japanese patients. J Oral Pathol Med 36: 351–356.
27. Tamaki S, Kawakami M, Yamanaka Y, Shimomura H, Imai Y, et al. (2009) Relationship between soluble MICA and the MICA A5.1 homozygous genotype in patients with oral squamous cell carcinoma. Clin Immunol 130: 331–337.
28. Lü M, Xia B, Ge L, Li Y, Zhao J, et al. (2009) Role of major histocompatibility complex class I-related molecules A*A5.1 allele in ulcerative colitis in Chinese patients. Immunology 128: e230–236.
29. Bonilla Guerrero R, Roberts LR (2005) The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. J Hepatol 42: 760–777.
30. Bouchard MJ, Schneider RJ (2004) The enigmatic X gene of hepatitis B virus. J Virol 78: 12725–12734.
31. Groh V, Bahram S, Bauer S, Herman A, Beauchamp M, et al. (1996) Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. Proc Natl Acad Sci U S A 93: 12445–12450.
32. Groh V, Steinle A, Bauer S, Spies T (1998) Recognition of stress-induced MHC molecules by intestinal epithelial gammadelta T cells. Science 279: 1737–1740.
33. Groh V, Rhinehart R, Randolph-Habecker J, Topp M, Riddell S, et al. (2001) Costimulation of CD8alphabeta T cells by NKG2D via engagement by MIC induced on virus-infected cells. Nat Immunol 2: 255–260.
34. Gasser S, Orsulic S, Brown EJ, Raulet DH (2005) The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. Nature 436: 1186–1190.
35. Lara-Pezzi E, Gomez-Gaviro MV, Galvez BG, Mira E, Iniguez MA, et al. (2002) The hepatitis B virus X protein promotes tumor cell invasion by inducing membrane-type matrix metalloproteinase-1 and cyclooxygenase-2 expression. J Clin Invest 110: 1831–1838.
36. Jinushi M, Hodi F, Dranoff G (2006) Therapy-induced antibodies to MHC class I chain-related protein A antagonize immune suppression and stimulate antitumor cytotoxicity. Proc Natl Acad Sci U S A 103: 9190–9195.

OPEN ACCESS Freely available online

PLOS ONE

Evolutionary Analysis of Classical *HLA* Class I and II Genes Suggests That Recent Positive Selection Acted on *DPB1*04:01* in Japanese Population

Minae Kawashima[1]*, Jun Ohashi[2]*, Nao Nishida[1,3], Katsushi Tokunaga[1]

1 Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, **2** Molecular and Genetic Epidemiology, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan, **3** The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Chiba, Japan

Abstract

The human leukocyte antigen (*HLA*) genes exhibit the highest degree of polymorphism in the human genome. This high degree of variation at classical *HLA* class I and class II loci has been maintained by balancing selection for a long evolutionary time. However, little is known about recent positive selection acting on specific *HLA* alleles in a local population. To detect the signature of recent positive selection, we genotyped six *HLA* loci, *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DQB1*, and *HLA-DPB1* in 418 Japanese subjects, and then assessed the haplotype homozygosity (*HH*) of each *HLA* allele. There were 120 *HLA* alleles across the six loci. Among the 80 *HLA* alleles with frequencies of more than 1%, *DPB1*04:01*, which had a frequency of 6.1%, showed exceptionally high *HH* (0.53). This finding raises the possibility that recent positive selection has acted on *DPB1*04:01*. The *DPB1*04:01* allele, which was present in the most common 6-locus *HLA* haplotype (4.4%), *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01*, seems to have flowed from the Korean peninsula to the Japanese archipelago in the Yayoi period. A stochastic simulation approach indicated that the strong linkage disequilibrium between *DQB1*06:04* and *DPB1*04:01* observed in Japanese cannot be explained without positive selection favoring *DPB1*04:01*. The selection coefficient of *DPB1*04:01* was estimated as 0.041 (95% credible interval 0.021–0.077). Our results suggest that *DPB1*04:01* has recently undergone strong positive selection in Japanese population.

Citation: Kawashima M, Ohashi J, Nishida N, Tokunaga K (2012) Evolutionary Analysis of Classical *HLA* Class I and II Genes Suggests That Recent Positive Selection Acted on *DPB1*04:01* in Japanese Population. PLoS ONE 7(10): e46806. doi:10.1371/journal.pone.0046806

Editor: Henry Harpending, University of Utah, United States of America

Received May 2, 2012; **Accepted** September 10, 2012; **Published** October 3, 2012

Funding: This work was supported by National Bioscience Database Center (NBDC) of Japan Science and Technology Agency (JST), by KAKENHI (22133008) Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and by KAKENHI (23133502) Grant-in-Aid for Scientific Research on Innovative Areas. 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: mina-e@umin.ac.jp (MK); juno-tky@umin.ac.jp(JO)

Introduction

The crucial immunological function of human leukocyte antigen (HLA) molecules is to present pathogen-derived antigenic peptides to T lymphocytes [1]. The HLA proteins are encoded by genes in the major histocompatibility complex region, which spans approximately 4 megabases (Mb) on the short arm of chromosome 6 (6p21.3) and includes the most polymorphic loci in the human genome [2]. A remarkable feature of the classical *HLA* class I and class II genes is the high degree of polymorphism. More than 1,750 *HLA-A*, 2,330 *HLA-B*, 1,300 *HLA-C*, 1,060 *HLA-DRB1*, 160 *HLA-DQB1*, and 150 *HLA-DPB1* alleles have been reported (IMGT/HLA database; http://www.ebi.ac.uk/imgt/hla/).

Positive selection has been shown as a driving force for the high degree of polymorphism at *HLA* loci [3,4]. The *HLA* genes show three remarkable signatures of positive selection: (1) the rate of nonsynonymous (amino acid altering) nucleotide substitution is substantially higher than that of synonymous substitution at antigen-recognition sites [5,6], (2) there are trans-species polymorphisms (i.e., similar alleles are present in multiple species) [7], and (3) there is a significant excess of heterozygosity [8,9]. Balancing selection, including overdominant selection and frequency-dependent selection, can easily account for these observations [3,4].

A number of studies have reported common long-range *HLA* haplotypes [10–16]. The extended length of common haplotype is a key feature of recent positive selection [17,18]. The *HLA* alleles on long-range haplotypes may have been subject to recent positive selection. In this study, to identify the signature of recent positive selection that has acted on specific *HLA* alleles in a local (i.e., geographically restricted) population, we investigated the allele frequencies and haplotype frequencies at *HLA-A*, *HLA-C*, *HLA-B*, *HLA-DRB1*, *HLA-DQB1*, and *HLA-DPB1* in 418 Japanese individuals. Our theoretical and computer simulation analyses suggested that *DPB1*04:01* has recently undergone strong positive selection in Japanese population.

Results

HLA Class I and Class II Alleles in Japanese

The genotypes of six *HLA* genes (three class I and three class II genes) were determined for each of 418 Japanese individuals. The frequencies of the 67 alleles found at the three *HLA* class I genes are listed in Table 1. Of the 17 *HLA-A* alleles, two–*A*02:01* and *A*24:02*–had frequencies higher than 10% (10.2 and 37.7 percent,

respectively). Of the 17 *HLA-C* alleles, four–*C*01:02*, *C*03:03*, *C*03:04*, and *C*07:02*–had frequencies higher than 10%: 16.5, 13.5, 12.6, and 14.5 percent, respectively. There were 33 *HLA-B* alleles, and not one had an allele frequency greater than 10%. The allele with the highest frequency (9.6%) was *B*52:01*; this allele was followed by *B*15:01* (8.5%), *B*51:01* (8.5%), *B*4403* (8.1%), and *B*35:01* (8.0%).

The frequencies of 53 alleles at three *HLA* class II genes are listed in Table 2. Of the 27 alleles at the *HLA-DRB1* locus, two–*DRB1*09:01* and *DRB1*04:05*–had frequencies of more than 10% (15.2% and 14.6%, respectively), and five–*DRB1*15:02* (8.4%), *DRB1*15:01* (8.0%), *DRB1*13:02* (7.8%), *DRB1*08:03* (7.5%), and *DRB1*01:01* (6.8%)–were also common. Of the 14 alleles at *HLA-DQB1*, four–*DQB1*03:03*, *DQB1*06:01*, *DQB1*04:01*, and *DQB1* 03:01*–were observed at frequencies of greater than 10% (15.9%, 15.9%, 14.6%, and 11.8%, respectively). There were four other common alleles at *HLA-DQB1–DQB1*03:02* (9.2%), *DQB1*06:02* (7.8%), *DQB1*05:01*, and *DQB1*06:04* (7.5%). Of the six *HLA* loci genotyped, *HLA-DPB1* had the fewest alleles with just 12. The *DPB1*05:01* (38.5%) and *DPB1*02:01* (25.1%) alleles were the most frequent alleles at this locus.

Of the six *HLA* loci examined, the *HLA-B* locus showed the highest heterozygosity (0.937), and *HLA-DPB1* showed the lowest (0.765) (Tables 1 and 2). None of the *HLA* class I or II loci

Table 1. Frequencies of *HLA* class I alleles.

HLA-A						HLA-C						HLA-B					
Allele	Count	Freq.	H[a]	HWE[b] P-val	EW[c] P-val	Allele	Count	Freq.	H[a]	HWE[b] P-val	EW[c] P-val	Allele	Count	Freq.	H[a]	HWE[b] P-val	EW[c] P-val
*A*01:01*	10	0.012	0.810	0.667	0.294	*C*01:02*	138	0.165	0.891	0.919	0.003	*B*07:02*	57	0.068	0.937	0.286	0.002
*A*02:01*	85	0.102				*C*01:03*	4	0.005				*B*13:01*	13	0.016			
*A*02:06*	61	0.073				*C*03:02*	3	0.004				*B*15:01*	71	0.085			
*A*02:07*	23	0.028				*C*03:03*	113	0.135				*B*15:07*	5	0.006			
*A*02:10*	2	0.002				*C*03:04*	105	0.126				*B*15:11*	5	0.006			
*A*03:01*	4	0.005				*C*04:01*	42	0.050				*B*15:18*	13	0.016			
*A*03:02*	1	0.001				*C*05:01*	5	0.006				*B*15:27*	1	0.001			
*A*11:01*	80	0.096				*C*06:02*	7	0.008				*B*15:28*	1	0.001			
*A*24:02*	315	0.377				*C*07:02*	121	0.145				*B*27:04*	2	0.002			
*A*24:08*	1	0.001				*C*07:04*	7	0.008				*B*35:01*	67	0.080			
*A*24:20*	10	0.012				*C*08:01*	47	0.056				*B*37:01*	7	0.008			
*A*26:01*	67	0.080				*C*08:03*	12	0.014				*B*39:01*	34	0.041			
*A*26:02*	12	0.014				*C*12:02*	81	0.097				*B*39:04*	5	0.006			
*A*26:03*	22	0.026				*C*12:03*	1	0.001				*B*40:01*	46	0.055			
*A*26:05*	1	0.001				*C*14:02*	50	0.060				*B*40:02*	57	0.068			
*A*31:01*	66	0.079				*C*14:03*	69	0.083				*B*40:03*	7	0.008			
*A*33:03*	76	0.091				*C*15:02*	31	0.037				*B*40:06*	34	0.041			
												*B*40:52*	1	0.001			
												*B*44:02*	5	0.006			
												*B*44:03*	68	0.081			
												*B*46:01*	38	0.045			
												*B*48:01*	22	0.026			
												*B*51:01*	71	0.085			
												*B*51:02*	4	0.005			
												*B*52:01*	80	0.096			
												*B*54:01*	64	0.077			
												*B*55:02*	20	0.024			
												*B*55:04*	1	0.001			
												*B*56:01*	5	0.006			
												*B*56:03*	2	0.002			
												*B*58:01*	3	0.004			
												*B*59:01*	16	0.019			
												*B*67:01*	11	0.013			

[a]Heterozygosity.
[b]Hardy-Weinberg equilibrium test.
[c]Ewens-Watterson test.
doi:10.1371/journal.pone.0046806.t001

Table 2. Frequencies of *HLA* class II alleles.

HLA-DRB1						*HLA-DQB1*						*HLA-DPB1*					
Allele	**Count**	**Freq.**	**H[a]**	**HWE[b]**	**EW[c]**	**Allele**	**Count**	**Freq.**	**H[a]**	**HWE[b]**	**EW[c]**	**Allele**	**Count**	**Freq.**	**H[a]**	**HWE[b]**	**EW[c]**
				P-val	P-val					P-val	P-val					P-val	P-val
*DRB1*01:01*	57	0.068	0.918	0.247	0.013	*DQB1*02:01*	1	0.001	0.885	0.222	0.001	*DPB1*02:01*	210	0.251	0.765	0.398	0.225
*DRB1*03:01*	1	0.001				*DQB1*03:01*	99	0.118				*DPB1*02:02*	35	0.042			
*DRB1*04:01*	10	0.012				*DQB1*03:02*	77	0.092				*DPB1*03:01*	36	0.043			
*DRB1*04:03*	24	0.029				*DQB1*03:03*	133	0.159				*DPB1*04:01*	51	0.061			
*DRB1*04:04*	2	0.002				*DQB1*04:01*	122	0.146				*DPB1*04:02*	83	0.099			
*DRB1*04:05*	122	0.146				*DQB1*04:02*	26	0.031				*DPB1*05:01*	322	0.385			
*DRB1*04:06*	28	0.033				*DQB1*05:01*	63	0.075				*DPB1*06:01*	5	0.006			
*DRB1*04:07*	1	0.001				*DQB1*05:02*	17	0.020				*DPB1*09:01*	65	0.078			
*DRB1*04:10*	12	0.014				*DQB1*05:03*	30	0.036				*DPB1*13:01*	12	0.014			
*DRB1*08:02*	32	0.038				*DQB1*06:01*	133	0.159				*DPB1*14:01*	10	0.012			
*DRB1*08:03*	63	0.075				*DQB1*06:02*	65	0.078				*DPB1*19:01*	5	0.006			
*DRB1*09:01*	127	0.152				*DQB1*06:03*	5	0.006				*DPB1*41:01*	2	0.002			
*DRB1*10:01*	6	0.007				*DQB1*06:04*	63	0.075									
*DRB1*11:01*	23	0.028				*DQB1*06:09*	2	0.002									
*DRB1*12:01*	30	0.036															
*DRB1*12:02*	18	0.022															
*DRB1*13:01*	5	0.006															
*DRB1*13:02*	65	0.078															
*DRB1*14:02*	1	0.001															
*DRB1*14:03*	11	0.013															
*DRB1*14:05*	17	0.020															
*DRB1*14:06*	13	0.016															
*DRB1*14:07*	3	0.004															
*DRB1*14:54*	26	0.031															
*DRB1*15:01*	67	0.080															
*DRB1*15:02*	70	0.084															
*DRB1*16:02*	2	0.002															

[a]Heterozygosity.
[b]Hardy-Weinberg equilibrium test.
[c]Ewens-Watterson test.
doi:10.1371/journal.pone.0046806.t002

exhibited significant deviation from HWE. Results of a Ewens-Watterson neutrality test [19,20] of *HLA* allele frequencies in this study population revealed that the observed distributions of allele frequencies at *HLA-C* ($P = 0.003$), *HLA-B* ($P = 0.002$), *HLA-DRB1* ($P = 0.013$), and *HLA-DQB1* ($P = 0.001$) differed significantly (i.e., there was excess heterozygosity) from the distributions expected based on the assumption of neutrality, whereas there was no significant difference between the expected and observed distributions of allele frequencies at *HLA-A* or *HLA-DPB1* (Tables 1 and 2).

Pairwise LD between *HLA* Alleles

The pairwise linkage disequilibrium (LD) parameters, r^2 and $|D'|$ [21], for each possible pair of two *HLA* alleles were estimated (Figure 1 and Data S1). Most alleles at *HLA-A* were not in strong LD with any of the alleles at the other loci because the physical distance from *HLA-A* to each of the other loci is large. To evaluate the relative strength of LD between two *HLA* loci, 2-locus r^2 and 2-locus $|D'|$ (see Materials and Methods for details), were calculated based on the pairwise LD parameters for all the allelic pairs (Table S1). The values of 2-locus $|D'|$ for *HLA-C* and *HLA-B* ($|D'| = 0.91$) and for *HLA-DRB1* and *HLA-DQB1* ($|D'| = 0.80$) were high, whereas the lowest 2-locus $|D'|$ value was observed for *HLA-A* and *HLA-DPB1* ($|D'| = 0.25$). These values reflected the physical distances between the respective loci. The values of 2-locus $|D'|$ for *HLA-DRB1* and *HLA-DPB1* and for *HLA-DQB1* and *HLA-DPB1* were relatively low compared to the values for the other pairs (Figure 2). These low values probably result from the recombination hotspot in the *HLA* class II region [22–24].

Major 6-locus *HLA* Haplotypes in Japanese

Frequencies of multi-locus haplotypes were estimated using the PHASE program [25,26] (Table 3 and Tables S2, S3, S4, S5). In 418 Japanese subjects (i.e., 839 chromosomes), 489 different 6-locus *HLA* haplotypes were inferred. Based on the frequencies of 6-locus *HLA* haplotypes, the probability of selecting two identical 6-locus *HLA* haplotypes at random from the Japanese population was estimated as 0.0075. Six 6-locus *HLA* haplotypes had

frequencies higher than 1% (Table 3). Of these, *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* was the most common (4.4%).

The intensity of recombination in the *HLA* region has been estimated at 0.67 cM/Mb [27], which corresponds to a recombination fraction of approximately 2% between *HLA-A* and *HLA-DPB1*. Thus, association between the six *HLA* alleles in any 6-locus *HLA* haplotype is not generally strong due to the frequent recombination in the *HLA* region. The expected frequency of the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype is 2.5×10^{-7} under the assumption of linkage equilibrium, which is much smaller than the observed frequency of 0.044. The strong LD among *HLA* alleles on the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype may result from recent positive selection acting on one of *HLA* alleles on the haplotype, although other mechanisms such as neutral random genetic drift, recent admixture, recent migration, recent bottlenecks, and suppression of recombination can also cause the strong LD [10,12,13,15,16].

Haplotype Omozygosity

Strong positive selection leads to a rapid increase in the frequency of a selected (target) allele in a population. The number of recombination events between the target allele and the surrounding polymorphic sites is limited while the advantageous allele increases in frequency; therefore, the diversity of haplotypes carrying the advantageous allele becomes low. Accordingly, strong LD is expected in the genomic region bearing the selected allele. In this study, the degree of LD for each HLA allele was measured by haplotype homozygosity (*HH*); this term is defined as the probability that any two randomly chosen samples of haplotype bearing a focal *HLA* allele have the same 6-locus *HLA* haplotype. Like *EHH* [17], a high *HH* value can be regarded as a signature of recent positive selection acting on a focal *HLA* allele.

To detect *HLA* alleles that have been subject to recent positive selection, *HH* was calculated for each allele based on the estimated number of 6-locus haplotypes in 418 Japanese subjects. Of the 80

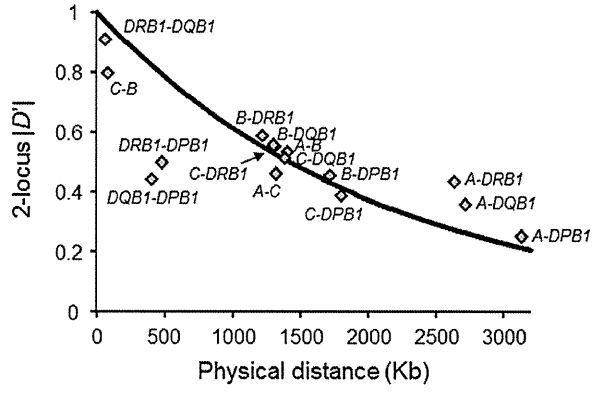

Figure 2. Relationship between two-locus |D'| and physical distance (Kb). A solid-line curve, $2-locus|D'|=(1-0.67\times10^{-5}\times x)^{75,13}$, was obtained using the least-squares method, where x represents the physical distance (Kb). The recombination rate in the *HLA* region was assumed to be 0.67 cM/Mb [27]. Spearman's rank correlation coefficient between 2-locus |D'| and the physical distance was −0.8607 ($P<0.0001$).
doi:10.1371/journal.pone.0046806.g002

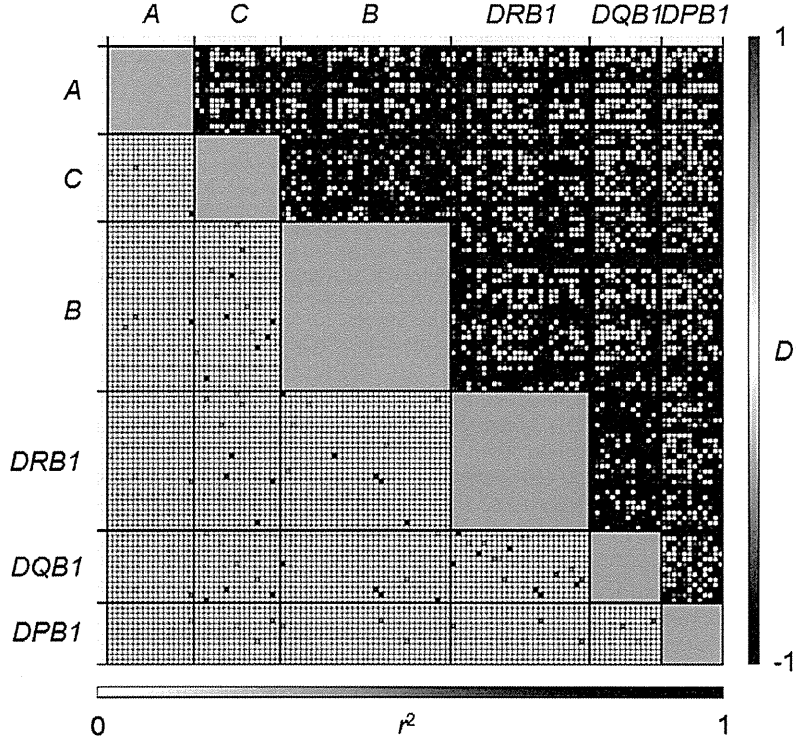

Figure 1. Pairwise estimates of LD parameters, |D'| (upper diagonal) and r^2 (lower diagonal) for every pair of *HLA* alleles. The name of each allele is presented in Data S1.
doi:10.1371/journal.pone.0046806.g001

Table 3. Estimated frequencies of 6-locus *HLA* haplotypes.

Association						# of haplotypes[a]	HF[b]
*A*33:03*	*C*14:03*	*B*44:03*	*DRB1*13:02*	*DQB1*06:04*	*DPB1*04:01*	37	0.044
*A*24:02*	*C*12:02*	*B*52:01*	*DRB1*15:02*	*DQB1*06:01*	*DPB1*09:01*	33	0.039
*A*24:02*	*C*07:02*	*B*07:02*	*DRB1*01:01*	*DQB1*05:01*	*DPB1*04:02*	29	0.035
*A*24:02*	*C*01:02*	*B*54:01*	*DRB1*04:05*	*DQB1*04:01*	*DPB1*05:01*	13	0.016
*A*24:02*	*C*12:02*	*B*52:01*	*DRB1*15:02*	*DQB1*06:01*	*DPB1*02:01*	12	0.014
*A*11:01*	*C*04:01*	*B*15:01*	*DRB1*04:06*	*DQB1*03:02*	*DPB1*02:01*	11	0.013

[a]Estimated by the PHASE program version 2.1.
[b]Haplotype frequency.
doi:10.1371/journal.pone.0046806.t003

HLA alleles that had frequencies of more than 1%, one allele at each class I locus (*A*33:03*, *C*14:03*, and *B*44:03*) had the highest *HH* for that locus; similarly, one allele at each class II locus (*DRB1*13:02*, *DQB1*06:04*, and *DPB1*04:01*) had the highest *HH* for that locus (Figure 3). These six *HLA* alleles made up the 6-locus haplotype, *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01*, with the highest frequency in this Japanese population (Table 3).

The *HH* values are generally reduced by loci with high heterozygosity. Therefore, it was relatively difficult for an allele at *HLA-DPB1* to show high *HH*, because heterozygosities at the other loci are high. Nevertheless, the *DPB1*04:01* allele, which had a population frequency of 6.1%, showed the highest *HH* value (0.53) of the 80 *HLA* alleles with frequencies higher than 1% (Figure 3). The values of *HH* of the remaining 79 *HLA* alleles were less than 0.33. This finding suggests that *DPB1*04:01* had undergone recent positive selection in Japan. The large *HH* values of the five other alleles (*A*33:03*, *C*14:03*, *B*44:03*, *DRB1*13:02*, and *DQB1*06:04*) in this 6-locus *HLA* haplotype (i.e., *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01*) appear to be due to the hitchhiking effect of *DPB1*04:01*.

To investigate the effect of recombination on the decay of the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype, the value of extended haplotype homozygosity (*EHH*) was calculated for *DPB1*04:01* (Figure 4). Although the *EHH* of *DPB1*04:01* was reduced at *HLA-DQB1*, the decrease in *EHH* was almost negligible at *HLA-DRB1*, *HLA-B*, and *HLA-C* loci; these findings indicate that, in this haplotype, recombination mainly has occurred between *DQB1*06:04* and *DPB1*04:01*.

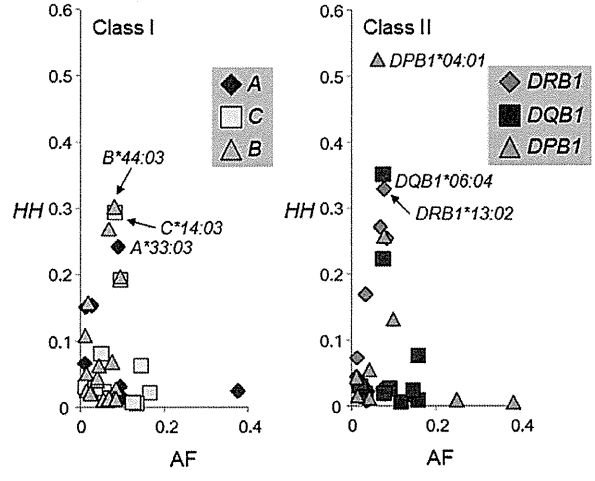

Figure 3. Haplotype homozygosity (*HH*) × allele frequency (AF) of each *HLA* allele. The left and right panels show *HH* values of *HLA* class I alleles and *HLA* class II alleles, respectively. The class I alleles were designated as follows: *HLA-A* (red diamond), *HLA-C* (yellow square), and *HLA-B* (green triangle); the class II alleles were designated as follows: *HLA-DRB1* (blue diamond), *HLA-DQB1* (purple square), and *HLA-DPB1* (pink triangle). In both panels, only *HH* values of alleles with frequencies of more than 0.01 are shown.
doi:10.1371/journal.pone.0046806.g003

Origin of *DPB1*04:01* in Japanese

*DPB1*04:01* is common (>30%) in European populations [9,28], whereas the frequency of *DPB1*04:01* is 6.1% in Japanese (Table 2). Given the worldwide distribution of *DPB1*04:01*, it is unlikely that *DPB1*04:01* originated in Japan. *DPB1*04:01* seems to have entered Japan. Archaeological studies of Japanese history have suggested that the Yayoi people came from the Korean peninsula circa 300 B.C., and mixed with the indigenous Jomon people. A recent large-scale survey of single nucleotide polymorphisms (SNPs) on autosomal chromosomes [29] revealed that most people presently inhabiting mainland Japan are genetically closer to Koreans than to Ryukuans. Ryukuans are considered to be more pure descendants of the Jomon people than are mainland Japanese. These observations indicate that a large population of Yayoi people migrated from the Korean peninsula. Although the frequency of the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype in Koreans has not been reported, *DPB1*04:01*, which was carried by *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01*, appears to have derived from the Korean population because the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04* and *DRB1*13:02-DQA1*01:02-DQB1*06:04-DPB1*04:01* haplotypes are observed at the frequencies of 4.2% and 4.7% in Korean populations [28,30,31]. These and similar haplotypes have not been reported in other Asian populations (http://www.allelefrequencies.net) [28].

If the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype has a single origin, the current genetic diversity of this haplotype must be low. To assess the genetic diversity of this haplotype, we performed a sliding window analysis of individual heterozygosity, defined as a proportion of heterozygous SNPs to all SNPs in the window (Figure 5). Reduced individual heterozygosity was only found in the *HLA* region on the short arm of chromosome 6 in all the three subjects that were

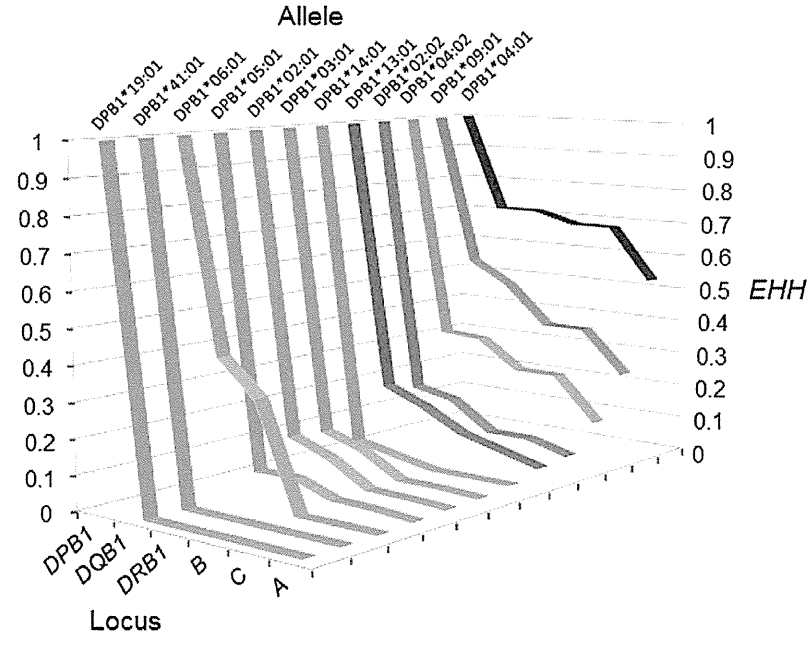

Figure 4. Extended *HH* (*EHH*) × relative locus position for12 *HLA-DPB1* alleles.
doi:10.1371/journal.pone.0046806.g004

homozygous for the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype (Figure 5A); in contrast, such a reduction was not observed in two subjects that were heterozygous for this haplotype (Figure 5B). Furthermore, three subjects that were homozygous for the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype shared the same SNP haplotype that spanned more than 4 Mb in the *HLA* region (Figure 5A). These observations suggest that the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype in Japanese has a single origin, and has not been generated repeatedly by recombination.

Computer Simulation

The analysis of *EHH* revealed that the reduction in *EHH* for *DPB1*04:01* resulted from recombination between *DQB1*06:04* and *DPB1*04:01* that inhabited the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype (Figure 4). Therefore, the relationship between *DQB1*06:04* and *DPB1*04:01* was focused in the following analyses. The high *HH* and *EHH* values of *DPB1*04:01* (Figures 3 and 4) may merely reflect that a neutral random genetic drift, rather than a recent positive selection, occurred after the Yayoi people reached the Japanese archipelago (300 B.C. or 2300 years ago). To assess this possibility, we conducted a computer simulation assuming a two-locus two-allele model in which changes in the frequency of four haplotypes carrying *DPB1*04:01* or non-*DPB1*04:01* alleles at the *HLA-DPB1* locus and *DQB1*06:04* or non-*DQB1:06:04* alleles at the *HLA-DQB1* locus were evaluated. In the simulation, the values of three parameters: selection intensity, s, recombination rate, c, and frequency of *DQB1*06:04-DPB1*04:01* haplotype, $f_1(0)$, in the beginning of the Yayoi period were drawn by a random number generator in every run. Haplotype frequencies were subject to change based on a stochastic model of positive selection, recombination, and random genetic drift. Dominant selection was assumed for *DPB1*04:01*, and, for the sake of simplicity, no selection (i.e., selectively neutral) was assumed for all alleles at the *DQB1* locus. The rejection method [18,32,33] was applied to accept only simulation runs that gave results similar to the observed values (see Materials and Methods for details). The uniform distribution was used for each parameter as a prior distribution (see Materials and Methods for detail). Figure 6A shows 2,500 parameter sets (i.e., posterior distributions) that were accepted in these simulations. The posterior distribution of the initial frequency of *DQB1*06:04-DPB1*04:01* haplotype was similar to the prior one, whereas the posterior distributions of selection intensity and recombination rate were different from the prior ones. In the posterior distribution, s ranged from 0.009 to 0.098, and the mean and 95% credible interval of s were 0.041 and 0.021−0.077, respectively (Figure 6B). It should be noted that neutral random genetic drift (i.e., $s \approx 0$) did not yield the results similar to the observed values. The findings from the simulations indicated that *DPB1*04:01* has been subject to relatively strong positive selection in Japanese since the Yayoi period.

Discussion

A number of *HLA* alleles have been shown to be associated with variations in immune responses to infectious diseases (e.g., human immunodeficiency virus [HIV]/AIDS, malaria, tuberculosis, hepatitis, leprosy, leishmaniasis, and schistosomiasis) caused by pathogenic microorganisms (see review by Blackwell et al. [34]). The most plausible explanation for positive selection favoring *DPB1*04:01* would be its function in resistance to infections. A recent genome-wide association study showed that the *DPA1*01:03-DPB1*04:01* haplotype confers protection against hepatitis B virus (HBV) infection (OR = 0.57, 95% CI = 0.33–0.96) [35]. Hepatitis B is a deadly infectious disease. Acute hepatitis B, which can cause fatal complications such as fulminant hepatitis, occurs in a percentage of the people infected with HBV. Although the estimated selection coefficient of s (0.0254–0.0550) for

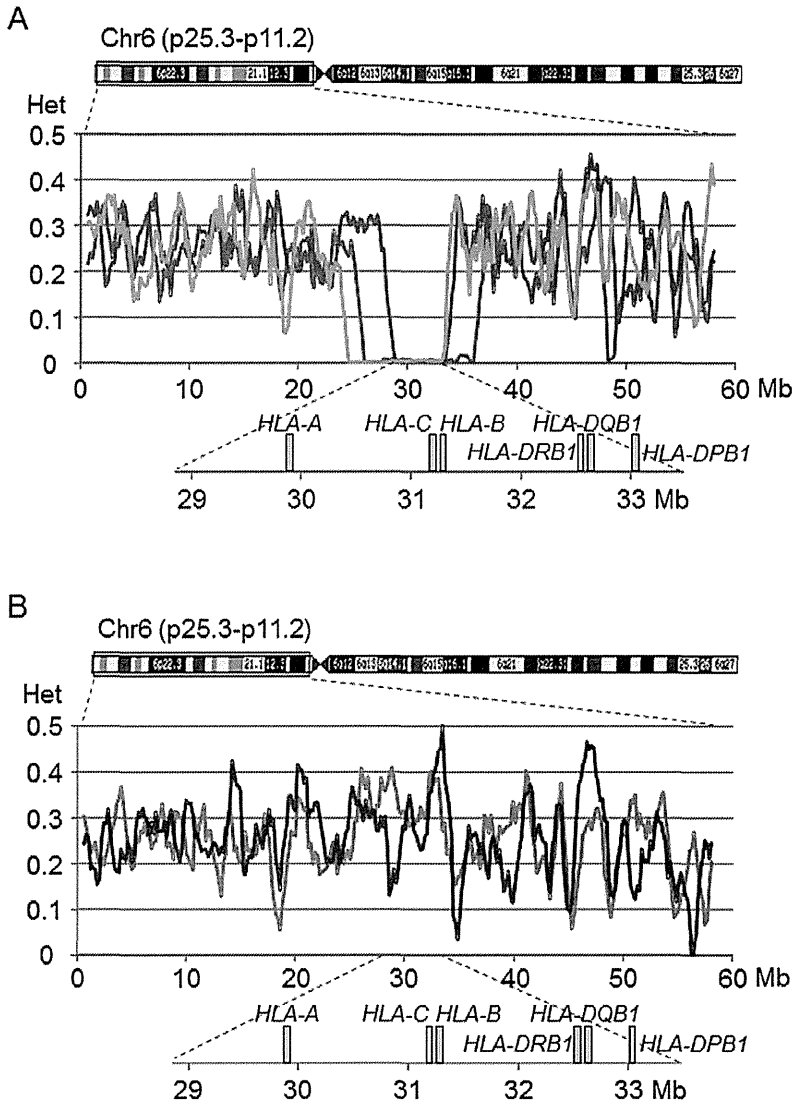

Figure 5. Individual heterozygosity of each subject with the most common 6-locus *HLA* haplotype. The individual heterozygosity in the genomic region on the short arm of chromosome 6 was assessed using the sliding window analysis; in this analysis, the window and step sizes were set to be 1 Mb and 200 kb, respectively. The individual heterozygosity was defined as a proportion of heterozygous SNPs to SNPs genotyped in a single subject. This analysis was performed for five Japanese subjects with the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype: (A) three of these five subjects were homozygous for this haplotype (blue, red, and green) and (B) two subjects had the heterozygous genotypes of the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype and the *A*24:02-C*07:02-B*07:02-DRB1*01:01–DQB1*05:01-DPB1*04:02* haplotype (orange) and of the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype and the *A*24:02-C*12:02-B*52:01-DRB1*15:02-DQB1*06:01-DPB1*09:01* haplotype (purple).
doi:10.1371/journal.pone.0046806.g005

*DPB1*04:01* does not seem to result solely from protection against infection with HBV, HBV infection may have been one of the key driving forces for the rapid increase in frequency of *DPB1*04:01* in the Japanese population.

Here, the analysis of *HH* was used to detect a signature of recent positive selection. The advantage of using *HH* in the analysis of *HLA* genes is that alleles with similar frequencies not only at the same *HLA* locus, but also at different loci, can be compared. This feature of analyses based on *HH* allows us to compare *HLA* alleles even within the same long-range haplotype. Since the same polymorphic markers are used for all *HLA* alleles in the calculation of *HH*, the effect of recombination on the value of *HH* can be well controlled. However, the *HH* analysis has a disadvantage in that the empirical distribution of *HH* value has to be obtained from only those alleles that are in the targeted region. Therefore, unlike conventional long-range haplotype tests based on *EHH* values [17,36], the statistical test based on *HH* values cannot be performed using genome-wide data. Nevertheless, *HH*-based test is thought to be suitable for analysis of *HLA* genes because each locus has a number of alleles to be examined and strong LD exists between alleles even at distant loci. The use of *HH* in the analysis of various human populations would help us to detect other *HLA*

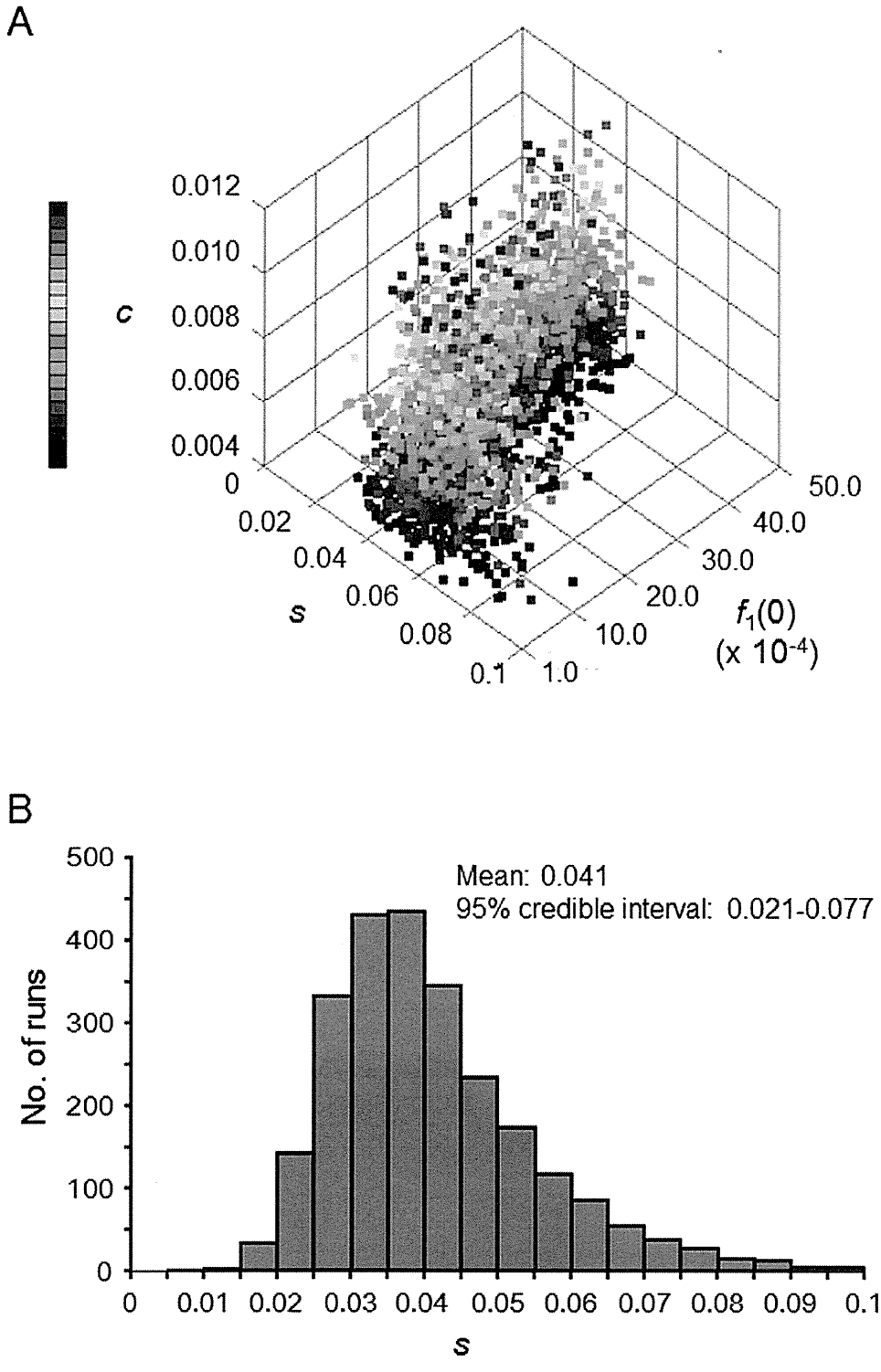

Figure 6. Estimation of model parameters for positive selection acting on *DPB1*04:01*. The recombination rate (*c*), initial haplotype frequency ($f_1(0)$), and selection coefficient (*s*), were estimated by comparing the four haplotype frequencies observed in our study population with the respective values predicted via simulation. (A) Posterior distributions of the three parameters that produced simulated data that resemble the observed data. (B) Frequency distribution of *s* accepted in simulation runs. The mean and 95% credible interval of *s* are 0.041 and 0.021−0.077.
doi:10.1371/journal.pone.0046806.g006

alleles that have been subject to geographically-restricted positive selection and to understand the role of *HLA* genes in the adaptation of human population to local environments over evolutionary time.

To estimate the selection coefficient of *DPB1*04:01*, we used a simple two-locus two-allele genetic model that was based on two assumptions, directional selection at *DPB1* and selective neutrality at *HLA-DQB1*. The problem associated with the use of this model was that the Ewens-Watterson test revealed that the allele frequency distribution at *HLA-DQB1* in this study population deviated significantly from that expected under neutrality (Table 2); therefore, the assumption of selective neutrality at *HLA-DQB1* may not be valid. If balancing selection is operating at *HLA-DQB1*, the allele frequency of *DQB1*06:04* is maintained at a certain frequency, and the change in the allele frequency of *DPB1*04:01* must be influenced by this selection at *HLA-DQB1*, although the effect of balancing selection at *HLA-DQB1* on the estimation of s is considered to be much smaller than that of directional selection favoring *DPB1*04:01*.

In this study, six *HLA* loci were investigated in 418 Japanese subjects. Of *HLA* alleles with high population frequencies, *DPB1*04:01*, which was present in the most common 6-locus *HLA* haplotype spanning more than 4 Mb, showed exceptionally high *HH*. A computer simulation estimated the selection coefficient of *DPB1*04:01* as 0.041. Taken together with high *HH* value of *DPB1*04:01*, we conclude that *DPB1*04:01* has recently undergone strong positive selection in Japanese population.

Materials and Methods

Subjects

All 418 individuals investigated in this study were unrelated Japanese adults living in Tokyo or neighboring areas. The genomic DNAs were extracted from peripheral blood samples using a commercial kit (QIAamp Blood Kit [Qiagen, Hilden, Germany]). All blood and DNA samples were de-identified. Verbal informed consent was obtained from all the participants before 1990. In this study, written informed consent was not obtained because the blood sampling was conducted before the "Ethical Guidelines for Human Genome and Genetic Sequencing Research" were established in Japan. Under the condition that DNA sample is permanently de-linked from the individual, this study was approved by the Research Ethics Committee of the Faculty of Medicine, University of Tokyo.

HLA Typing

DNA typing of *HLA* alleles was performed by HLA LABORATORY (Kyoto, Japan) using a Luminex Multi-Analyte profiling system (xMAP; Luminex, Austin, TX, USA) [37].

SNP Typing

Five Japanese subjects who had at least one *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype were genotyped using the Axiom™ Genome-Wide ASI 1 Array Plate (Affymetrix Inc., Santa Clara, CA, USA). Of five subjects, three subjects were homozygous for the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype and two subjects had the heterozygous genotypes of the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype and the *A*24:02-C*07:02-B*07:02-DRB1*01:01-DQB1*05:01-DPB1*04:02* haplotype and of the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype and the *A*24:02-C*12:02-B*52:01-DRB1*15:02-DQB1*06:01-DPB1*09:01* haplotype.

Statistical Analysis

Deviation from HWE for each *HLA* locus was tested using an exact test available in a web-based software, Genepop 4.0.10 [38]. Using Arlequin version 3.5 [39], the Ewens-Watterson test [40], which is based on Ewens sampling theory of neutral alleles [19], was performed to assess whether the observed distribution of allele frequencies at each *HLA* locus was different from an expectation that was based on neutrality.

To evaluate the degree of LD between *HLA* alleles, values of r^2 and D' [21] for all pairwise combinations of *HLA* alleles were calculated based on the haplotype frequencies estimated using the expectation maximization algorithm [20]. Here, each *HLA* allele was regarded as a single nucleotide polymorphism (SNP). For example, the *A*01:01* allele and the other alleles at the *HLA-A* locus were designated as "A" and "G", respectively. Accordingly, the algorithm for the estimation of haplotype frequencies for two loci, each with two alleles, could be applied to the *HLA* loci with multiple alleles for the purposes of these pairwise comparisons.

The LD parameter, 2-locus $|D'|$, between any two *HLA* loci (locus 1 and locus 2) was calculated based on the pairwise LD parameter, D'_{ij}, between ith allele at locus 1 and jth allele at locus 2 as follows: 2-locus $|D'| = \sum_{i=1}^{m} \sum_{j=1}^{n} p_i q_j |D'_{ij}|$, where p_i and q_j represent the frequencies of ith allele at locus 1 with m different alleles and jth allele at locus 2 with n different alleles. Spearman's rank correlation coefficient between 2-locus $|D'|$ and the physical distance was calculated. Assuming a model: 2-locus $|D'| = \left(1 - 0.67 \times 10^{-5} \times x\right)^a$, the curve fitting model parameter, a, was estimated using the least squares method; this method minimizes the sum-of-squared residual between an observed value and a fitted value that was determine by a model. In the above equation, the physical distance (Kb) between two loci is denoted by x and the recombination intensity in the *HLA* region was set at 0.65 cM/Mb [27,41].

The phased haplotypes consisting of two or more *HLA* loci were estimated using the PHASE program version 2.1 [25,26]. The estimated 6-locus haplotypes were further used for the calculation of extended haplotype homozygosity (*EHH*) [17] and of haplotype homozygosity (*HH*). In this study, *HH* of each *HLA* allele was defined as the probability that any two randomly chosen samples of haplotype bearing the *HLA* allele have the same 6-locus *HLA* haplotype.

A sliding window analysis of individual heterozygosity, which was defined as the proportion of heterozygous SNPs to SNPs genotyped in a single subject, was conducted to examine whether the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype had a single origin in Japan. 19,949 SNPs located on 6p were genotyped, and the average SNP density was 0.34 SNP/kb. The window and step sizes were 1 Mb and 200 kb, respectively. This analysis was performed using the SNP data from the five subject included in the SNP typing: three subjects were homozygous for the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype and two subjects had the heterozygous genotypes of the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype and the *A*24:02-C*07:02-B*07:02-DRB1*01:01-DQB1*05:01-DPB1*04:02* haplotype and of the *A*33:03-C*14:03-B*44:03-DRB1*13:02-DQB1*06:04-DPB1*04:01* haplotype and the *A*24:02-C*12:02-B*52:01-DRB1*15:02-DQB1*06:01-DPB1*09:01* haplotype.

Computer Simulation

To estimate the intensity of recent positive selection acting on *DPB1*04:01*, a stochastic population genetic model (two-locus two-allele model) assuming both positive selection and random

genetic drift was built and assessed. The diploid population size, N, was set to be 10,000 (i.e., 20,000 chromosomes). Four haplotypes carrying *DPB1*04:01* or non-*DPB1*04:01* alleles (designated by *DPB1*X*) at the *HLA-DPB1* locus and *DQB1*06:04* or non-*DQB1:06:04* alleles (designated by *DQB1*X*) at the *HLA-DQB1* locus were used in this model. The frequencies of the *DQB1*06:04-DPB1*04:01*, *DQB1*X-DPB1*04:01*, *DQB1*06:04-DPB1*X*, and *DQB1*X-DPB1*X* haplotypes at generation t were denoted by $f_1(t)$, $f_2(t)$, $f_3(t)$, and $f_4(t)$, respectively. The current frequencies of the corresponding haplotypes in our study population were denoted by f_1, f_2, f_3, and f_4. A dominant selection was assumed for *DPB1*04:01* (i.e., relative finesses of *DPB1*04:01/DPB1*04:01*, *DPB1*04:01/DPB1*X*, and *DPB1*X/DPB1*X* are 1, 1, and 1 - s, respectively). The initial haplotype frequencies were set as $f_1(t) = z$, $f_2(t) = 0$, $f_3(t) = (1-z)f_3/(f_3+f_4)$, and $f_4(t) = (1-z)f_4/(f_3+f_4)$. The recombination between *HLA-DPB1* and *HLA-DQB1* loci was assumed to occur at a rate of c. Since the recombination rate between *HLA-DQB1* and *HLA-DPB1* has been estimated to be between 0.004 and 0.012 [41,42], a uniform recombination rate (c) within this range was used as a prior distribution. To estimate suitable parameter sets of z, s, and c, each value was drawn by a random number generator in every simulation run. The random numbers were between 0.0001 (i.e., $2/2N$) and 0.005 (i.e., $100/2N$) for z, between 0 and 0.1 for s, and between 0.004 and 0.012 for c.

Next, to evaluate the similarity between simulated and observed frequencies,

$$e = \sum_{i=1}^{4} \frac{(f_i(t) - f_i)^2}{f_i(t) + f_i}$$

was calculated. As the simulated haplotype frequencies, $f_1(t)$, $f_2(t)$, $f_3(t)$, and $f_4(t)$, approaches values close to the observed frequencies, f_1, f_2, f_3, and f_4, the value of e approaches 0. The rejection method [18,32,33] was used to accept only simulation runs that resulted in (i) e of less than 0.01, (ii) $f_1(t)$ of not less than $f_1 - 0.01$ nor more than $f_1 + 0.01$, and (iii) t of not less than 92 nor more than 115 generations. A total of 2,500 runs were accepted. The mean and 95% credible interval of s were obtained from the 2,500 accepted runs.

Supporting Information

Data S1 Pairwise LD measures for individual HLA allele pairs.
(XLSX)

Table S1 Linkage Disequilibrium between pairs of HLA loci.
(XLSX)

Table S2 Estimated frequencies of 2-locus HLA haplotypes.
(XLSX)

Table S3 Estimated frequencies of 3-locus HLA haplotypes.
(XLSX)

Table S4 Estimated frequencies of 4-locus HLA haplotypes.
(XLSX)

Table S5 Estimated frequencies of 5-locus HLA haplotypes.
(XLSX)

Acknowledgments

We deeply thank all the subjects for their participation in the study. We also thank Ms Yoriko Mawatari, Ms Megumi Sageshima, Ms Yuko Ogasawara, Ms Natsumi Baba, and Ms Rieko Hayashi (University of Tokyo) for technical assistance.

Author Contributions

Conceived and designed the experiments: MK JO. Performed the experiments: MK NN. Analyzed the data: MK JO. Contributed reagents/materials/analysis tools: JO NN KT. Wrote the paper: MK JO. Assembled the data: MK NN. Performed the computer simulation: JO.

References

1. Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, et al. (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329: 512–518.
2. Campbell RD, Trowsdale J (1993) Map of the human MHC. Immunology today 14: 349–352.
3. Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 124: 967–978.
4. Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc Natl Acad Sci U S A 87: 2419–2423.
5. Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335: 167–170.
6. Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci U S A 86: 958–962.
7. Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Hum Immunol 19: 155–162.
8. Hedrick PW, Thomson G (1983) Evidence for balancing selection at HLA. Genetics 104: 449–456.
9. Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, et al. (2008) Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. Hum Immunol 69: 443–464.
10. Aly TA, Eller E, Ide A, Gowan K, Babu SR, et al. (2006) Multi-SNP analysis of MHC region: remarkable conservation of HLA-A1-B8-DR3 haplotype. Diabetes 55: 1265–1269.
11. Horton R, Gibson R, Coggill P, Miretti M, Allcock RJ, et al. (2008) Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. Immunogenetics 60: 1–18.
12. Alper CA, Larsen CE, Dubey DP, Awdeh ZL, Fici DA, et al. (2006) The haplotype structure of the human major histocompatibility complex. Hum Immunol 67: 73–84.
13. Yunis EJ, Larsen CE, Fernandez-Vina M, Awdeh ZL, Romero T, et al. (2003) Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks. Tissue Antigens 62: 1–20.
14. Traherne JA, Horton R, Roberts AN, Miretti MM, Hurles ME, et al. (2006) Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. PLoS Genet 2: e9.
15. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, et al. (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nature genetics 38: 1166–1172.
16. Awdeh ZL, Raum D, Yunis EJ, Alper CA (1983) Extended HLA/complement allele haplotypes: evidence for T/t-like complex in man. Proc Natl Acad Sci U S A 80: 259–263.
17. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419: 832–837.
18. Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Brittenham G, et al. (2004) Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. Am J Hum Genet 74: 1198–1208.
19. Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Popul Biol 3: 87–112.
20. Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12: 921–927.
21. Lewontin RC (1964) The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. Genetics 49: 49–67.
22. Mitsunaga S, Kuwata S, Tokunaga K, Uchikawa C, Takahashi K, et al. (1992) Family study on HLA-DPB1 polymorphism: linkage analysis with HLA-DR/DQ and two "new" alleles. Human immunology 34: 203–211.