using GENETYX Version 11.0 (GENETYX Corporation, Tokyo, Japan) Additional sequences were retrieved from the DNA Data Bank of Japan, EMBL Nucleotide Sequence Submissions and GenBank nucleic acid sequence databases for phylogenetic analysis. Phylogenetic relationships between sequences were determined using the neighbor-joining method using MEGA 4 software (14).

### Quantification of serum hepatitis B virus DNA

Hepatitis B virus DNA was quantified using real-time detection PCR as previously described (15), with modifications as previously described (16). The detection limit of this assay was 100 copies/mL.

### Statistical analysis

Statistical differences were evaluated by Fisher's exact test and the $X^2$ test, with Yates' correction for continuity where appropriate. Differences were considered significant for $P$-values < 0.05. All statistical analyses were performed using version 8.0 of the Stata Software package (StrataCorp LP, College Station, TX, USA).

## RESULTS

### Clinical characteristics of chronic hepatitis B patients in Turkey

Sera were collected from a wide area of Turkey. Figure 1 and Table 1 illustrate the locations of the four cities and the number of collected samples, respectively. The clinical characteristics of the 198 HBsAg-positive patients are summarized in Table 2. The LC/HCC patients were significantly older than the CH patients ($P < 0.0001$). Most (89.5%, 34 of 38 patients) were male. The concentrations of ALT, AST, and HBV DNA were not significantly different in LC/HCC patients compared to CH patients.

### Hepatitis B virus genotypes

Hepatitis B virus genotypes were successfully determined in 185/198 HBsAg-positive patients by the EIA genotyping method. Genotyping was not possible for the remaining 13 patients because no HBV PCR products were detected in their samples.

### Phylogenetic analysis of hepatitis B virus isolates based on complete genome sequencing

Whole HBV genomes were obtained from 36 of the HBsAg samples from four different geographical regions of Turkey, including 9 strains from Ankara, 15 from Samsun, 4 from Istanbul and 8 from Gaziantep (Fig. 2). All but one of these strains clustered with database reference strains representing genotype D, subgenotype D1. One strain isolated from a patient in Samsun clustered with subgenotype D3 references.

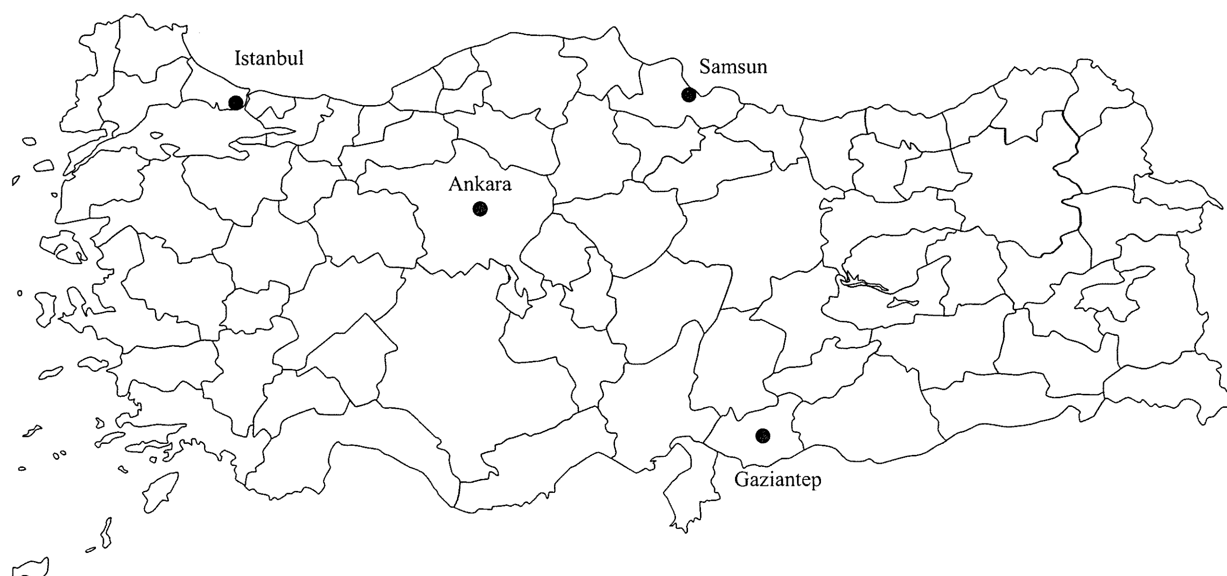



**Fig. 1.** Geographical locations from which samples were collected. All samples were collected at four clinical centers in geographically distinct parts of Turkey: Samsun (north), Ankara (center), Gaziantep (south), and Istanbul (west).



© 2012 The Societies and Wiley Publishing Asia Pty Ltd

**Table 2.** Clinical characteristics of chronic viral hepatitis patients from Turkey with and without liver cirrhosis and/or hepatocellular carcinoma

| Characteristic | Total ($n$ = 198) | CH ($n$ = 160) | LC/HCC ($n$ = 8) | $P$-values |
|---|---|---|---|---|
| Age (years)[†] | 41.4 ± 15 | 37.1 ± 13 | 58.9 ± 9 | < 0.001 |
| Male[‡] | 132 (66.7) | 98 (61.2) | 34 (89.5) | < 0.001 |
| Anti-HCV[‡] | 1 (0.5) | 1 (6%) | 0 | NS |
| Genotyped (HBV/D)[‡] | 185 (93.4) | 151 (94.4) | 34 (89.5) | NS |
| HBeAg[‡] | 68 (34.3) | 59 (36.9) | 9 (23.7) | NS |
| TP (g/dL)[†] | 7.4 ± 0.8 | 7.6 ± 0.6 | 6.5 ± 1 | < 0.001 |
| Alb (g/dL)[†] | 3.9 ± 0.7 | 4.1 ± 0.5 | 3.0 ± 0.7 | 0.001 |
| Glob (g/dL)[†] | 3.5 ± 0.6 | 3.5 ± 0.6 | 3.4 ± 0.6 | NS |
| PLT count ($\times 10^3$ μL)[†] | 150 ± 107 | 164 ± 112 | 94 ± 65 | < 0.001 |
| PT-INR[†] | 1.1 ± 0.3 | 1.0 ± 0.2 | 1.5 ± 0.2 | 0.01 |
| AST (IU/mL)[†] | 93 ± 194 | 90 ± 208 | 107 ± 119 | NS |
| ALT (IU/mL)[†] | 123 ± 226 | 133 ± 247 | 80 ± 75 | NS |
|  | 55 ± 47 | 52 ± 48 | 69 ± 37 | NS |
| ALP (IU/L)[†] | 133.1 ± 83.4 | 134 ± 84 | 130 ± 82 | NS |
| TB (mg/dL)[†] | 1.9 ± 5.5 | 1.1 ± 2.4 | 5.4 ± 10.9 | < 0.001 |
| DB (mg/dL)[†] | 1.2 ± 4.7 | 0.5 ± 1.5 | 4.1 ± 9.9 | < 0.001 |
| HBV DNA ($\log_{10}$ copies/mL)[†] | 5.4 ± 24 | 5.9 ± 25 | 0.2 ± 0.8 | NS |

NS, not significant.
[†]mean ± SD [‡]number (%) of patients (percentage).

### Basal core promoter and core region sequence analysis

To investigate genetic differences between LC/HCC and CH patients, the BCP/CP regions of HBV were successfully sequenced in samples from 22 LC/HCC patients and 52 age-, sex- and HBeAg-status-matched non-LC/HCC patients. Matching control subjects for these characteristics is important because HBV mutation rates are dependent on them. A summary of mutations observed in the BCP/CP region is presented in Table 3. There was a tendency toward a difference in the prevalence of the T1764G1766 double mutation ($P = 0.065$) and a statistically significant difference in prevalence of the A1896 mutation ($P = 0.03$) between LC/HCC and non-LC/HCC patients, a higher prevalence being found in LC/HCC patients. There was also a significant difference in the prevalence of the C1773 mutation, which was more frequently present in CH patients than in controls ($P = 0.05$).

Further, viral and host characteristics of the HBeAg-positive and HBeAg-negative patients were compared by using the samples from which the BCP/CP sequence were obtained (Table 4). Significant differences between the groups were observed in terms of the prevalence of V1753, A1757, and A1896 mutations ($P$ = 0.011, 0.024, and 0.0001, respectively).

The T1773 mutation is associated with HBeAg-negative patients and is less often found in patients with advancing liver disease and infection with HBV genotypes B and C (17). However, in the present study, an excess of the T1773 mutation was not observed in patients with mild liver damage (Table 3) or HBeAg-negative patients (Table 4). On the other hand, specific mutation patterns were observed in HBeAg-negative patients infected with HBV/D. All the HBeAg-negative samples shown in Table 4 were allocated to two groups based on their 1773 mutation patterns (T1773 or C1773) and analyzed to determine any correlations with other mutations in the core promoter region. As shown in Table 5, the T1773 mutation coupled with the double mutation, T1764/G1766. In addition, the prevalence of A1757/T1764/G1766 mutations in the T1773 mutation group was statistically significant. The T1773 group had a larger viral load than did the C1773 group without the T1764/G1766 double mutation.

## DISCUSSION

Viral hepatitis is one of the most prevalent and serious infectious diseases in the world and presents a serious public health problem. HBV infection follows different routes of inter- and intra-community transmission, various geographical, social and cultural factors playing important roles. The epidemiology of HBV genotypes provides useful information about population-specific behaviors, which may have direct or indirect roles in HBV transmission (18).

In this study, we investigated the genetic characteristics of HBV in a cohort of patients with CH with and without LC/HCC in Turkey. Phylogenetic analysis of complete genomes was carried out on HBV isolates from patients in different regions of Turkey to determine the distribution and transmission of different HBV genotypes within different areas of the country, and between Turkey and other parts of the world. In the present study in Turkey, all HBV genotypes (subgenotypes) were D1 type except for one isolate. There were no specific phylogenetic groupings of HBV isolates according to geography within Turkey. Previous studies have reported a high prevalence of genotype D1 (approximately 89%) in Mongolia (19, 20). They showed that HCV and Delta virus co-infections with HBV infection confer a high risk of HCC. These studies differ from the present study in that they investigated cases of co-infection with HBV and HCV or with HBV and Delta virus to assess association with HCC whereas we assessed cases of monoinfection. However, in Turkey genotype D1 infection without HCV co-infection is characterized by early HBeAg seroconversion, a small viral load upon seroconversion

© 2012 The Societies and Wiley Publishing Asia Pty Ltd

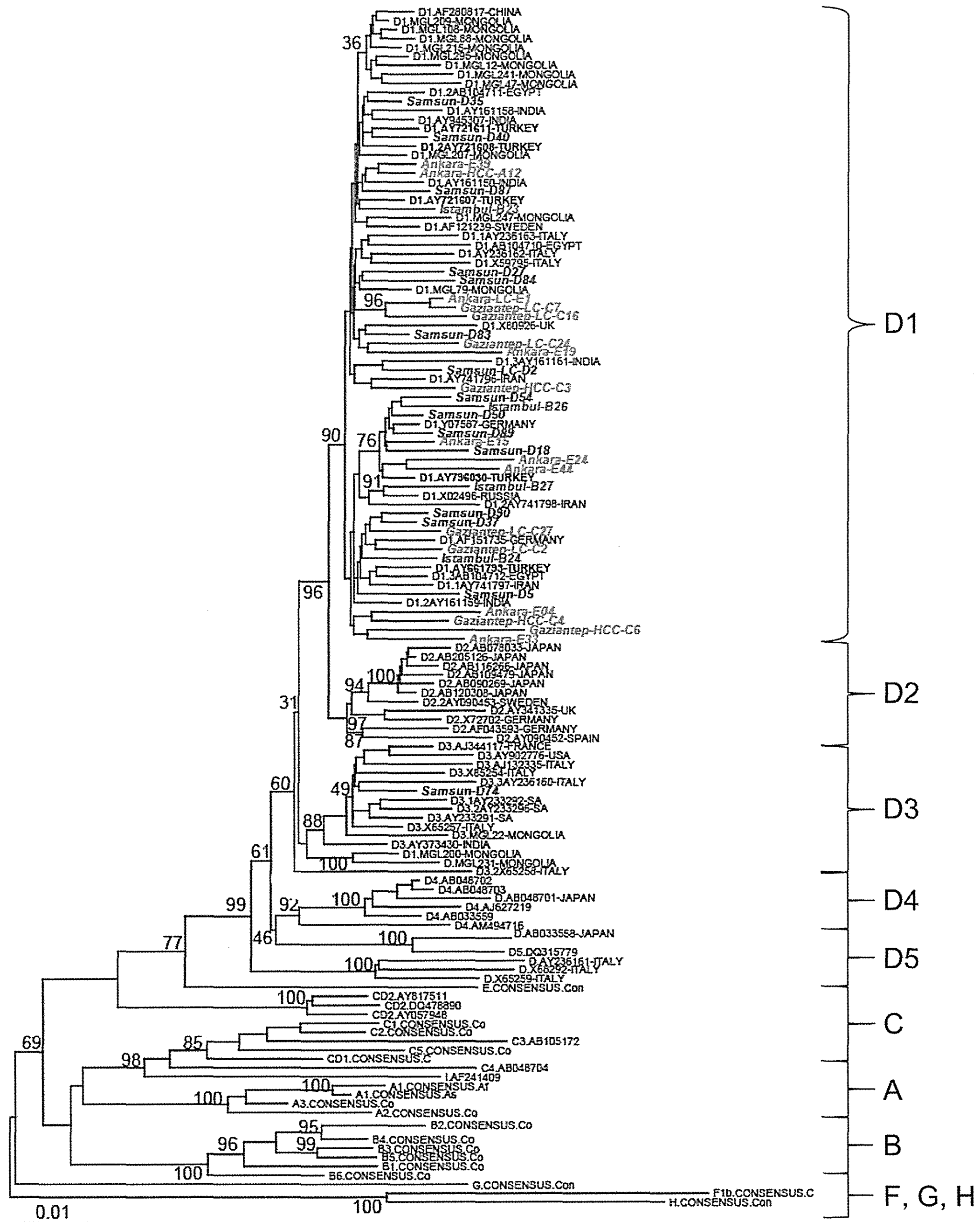



**Fig. 2.** Neighbor-joining phylogram based on complete hepatitis B virus genomes from Turkey and other countries. Sequences from the current study are color-coded according to the region of Turkey from which they were isolated (blue, Samsun; green, Ankara; orange, Gaziantep; red, Istanbul), and other Turkish sequences are labeled in bold. The study isolates were subjected to bootstrap re-sampling with all available complete genome sequences obtained from the EMBL, DDBJ, and GenBank nucleic acid sequence databases. Sequences used for the phylogenetic tree are indicated under the corresponding accession numbers from sequence databases and country of origin.

and a relatively low incidence of LC/HCC in those infected, which might indicate that HCV and Delta virus co-infection change the pathogenesis of HBV genotype D1.

Two previous studies on Iranian and Mongolian genotype D isolates reported a genotype-specific pattern of the functionally important BCP/CP region, with A1757 and T1764/G1766 (21, 22). Our *in vitro*


© 2012 The Societies and Wiley Publishing Asia Pty Ltd

**Table 3.** A comparison of BCP/PC mutations between age-, sex-, and HBeAg-status-matched chronic viral hepatitis patients from Turkey with and without LC and/or HCC

| | CH ($n$ = 52) | LC/HCC ($n$ = 22) | $P$-values |
|---|---|---|---|
| Age (years, mean ± SD) | 46.6 ± 11 | 57.1 ± 10.1 | NS |
| Male | 38 (73.1) | 20 (90.9) | NS |
| HBeAg | 21 (40.4) | 7 (31.8) | NS |
| T1653 | 3 (5.8) | 3 (13.6) | NS |
| C1727 | 12 (23.1) | 3 (13.6) | NS |
| C1752 | 12 (23.1) | 3 (13.6) | NS |
| V1753 | 14 (26.9) | 5 (22.7) | NS |
| G1757 | 13 (25) | 3 (13.6) | NS |
| T1762/A1764 | 19 (36.5) | 7 (31.8) | NS |
| T1764/G1766 | 7 (13.5) | 7 (31.8) | NS (0.065) |
| C1773 | 27 (51.9) | 6 (27.3) | NS (0.053) |
| Kozak | 8 (15.4) | 2 (9.1) | NS |
| H1862 | 5 (9.6) | 5 (22.7) | NS |
| A1896 | 22 (42.3) | 16 (72.7) | 0.017 |

NS, not significant.
Numbers in brackets represent percentages (%).
V base contains A, C or G bases. H base contains A, C or T bases. Kozak, Kozak sequence used in translation initiation site of eukaryotic mRNA.

**Table 4.** A comparison of BCP/PC mutations between age-matched patients with chronic viral hepatitis from Turkey with positive or negative HBeAg status

| | HBeAg-positive ($n$ = 47) | HBeAg-negative ($n$ = 50) | $P$-values |
|---|---|---|---|
| Age (years, mean ± SD) | 37.6 ± 14.4 | 37.6 ± 14.4 | Matched |
| LC/HCC | 7 (14.9) | 14 (28) | NS |
| Male | 30 (63.8) | 37 (74) | NS |
| T1653 | 2 (4.3) | 4 (8) | NS |
| C1727 | 7 (14.9) | 10 (20) | NS |
| C1752 | 9 (19.1) | 9 (18) | NS |
| V1753 | 5 (10.6) | 16 (32) | 0.011 |
| A1757 | 43 (91.5) | 37 (74) | 0.024 |
| T1762/A1764 | 10 (21.3) | 17 (34) | NS |
| T1764/G1766 | 9 (19.1) | 11 (22) | NS |
| C1773 | 16 (34) | 24 (48) | NS |
| Kozak | 2 (4.3) | 8 (16) | NS (0.057) |
| T1862 | 7 (14.9) | 7 (14) | NS |
| A1896 | 6 (12.8) | 38 (76) | <0.001 |

NS, not significant. Numbers in bracket represent percentages (%).
V base contains A, C or G bases. Kozak, Kozak sequence used in translation initiation site of eukaryotic mRNA.

experiments indicated that the A1757 and T1764/G1766 mutations are associated with the levels of viral. In the present study, we observed an association between T1773 and T1764/G1766 and a higher viral load in Turkish patients, but identified no clear correlations between

**Table 5.** Association between T1773 and T1764/G1766 double mutation

| | T1773 ($n$ = 26) | C1773 ($n$ = 24) | $P$-values |
|---|---|---|---|
| LC/HCC | 9 (34.6) | 5 (20.8) | NS |
| Male | 22 (84.6) | 15 (62.5) | NS |
| Age (years, mean ± SD) | 36.9 ± 12.5 | 38.3 ± 14.3 | NS |
| T1653 | 0 | 4 (16.7) | 0.030 |
| V1753 | 7 (25.9) | 9 (37.5) | NS |
| A1757 | 20 (76.9) | 17 (70.8) | NS |
| T1762/A1764 | 8 (29.6) | 9 (37.5) | NS |
| T1764/G1766 | 11 (40.7) | 0 | < 0.001 |
| Kozak | 2 (7.4) | 6 (25) | NS (0.095) |
| T1862 | 6 (22.2) | 1 (4.2) | NS (0.054) |
| A1896 | 19 (70.4) | 19 (79.2) | NS |
| A1757 + T1762/A1764 | 3 (11.1) | 4 (16.7) | NS |
| A1757 + T1764/G1766 | 11 (40.7) | 0 | 0.001 |
| Al 757 + wild 1762/1764/1766 | 6 (23.1) | 13 (54.2) | 0.024 |
| G1757 + T1762/A1764 | 5 (18.5) | 5 (20.8) | NS |
| Gl 757 + wild 762/1764/1766 | 1 (3.7) | 2 (8.3) | NS |
| HBV DNA $\log_{10}$ copies/mL | 5.4 ± 1.8 | 4.6 ± 1.3 | 0.009 |

NS, not significant. Numbers in brackets represent percentages (%).
V base contains A, C or G bases. Kozak, Kozak sequence used in translation initiation site of eukaryotic mRNA.

mutations in the BCP, PC, and/or core region and disease prognosis. This may have been a result of the uneven group sizes of the samples or a specific pattern of viral mutation that is dependent on geographical area. Further *in vitro* and clinical studies are needed to clarify the role of the 1773 mutation.

In this study, we observed an accumulation of T1773 mutations in CH patients and no statistically significant difference between HBeAg positive and HBeAg negative patients, in contrast to a previous paper on Taiwanese subjects (17). These discrepancies might be related to the different study populations because HBV mutation patterns are dependent on genotype and race. Turkey is a high prevalence area for HBV/D according to nationwide collection of samples, whereas Taiwan area is known to have a high prevalence of HBV/B and C. As previous studies have reported (21, 22), HBV/D1 has a unique mutation pattern in the BCP/CP region. The T1762/A1764 double mutation frequently occurs in HBV/B and C, whereas the T1764/G1766 double mutation tends to occur in HBV/D1. The amount of HBV-DNA in the A1757/T1764/G1766/T1773 mutation group was higher than that in the non-A1757/T1764/G1766/T1773 group; these findings are in concordance with those of Sendi *et al.* (22). Therefore, the specific mutation pattern of HBV/D1 might provide advantages in viral replication. Detection of coordinated mutations such as A1757/T1764/G1766/T1773 suggests the possibility that a

© 2012 The Societies and Wiley Publishing Asia Pty Ltd

mechanism such as secondary structure or a distinct transcriptional factor binding in the BCP/CP region of HBV/D1 is having an effect. Computer simulation shows binding of hepatocyte nuclear factor 3 on A1757/T1764/G1766/T1773.

Hepatitis B virus has a compact and constrained genome (23), and correlations between particular mutations in *cis*-acting elements of the virus and different phenotypic features of the virus have been shown clinically (17, 24–26), *in vitro*, and *in vivo* (27–29). In addition to viral factors, environmental factors such as exposure to aflatoxin (30) and the prevalence of co-infections (19, 31), may play important roles in causing regional differences in the clinical manifestation of HBV infections. Recent progress associated with the human genome project indicates the importance of host genetic factors in the outcome of HBV infections (32). There is still much to discover about HBV genotype D infection. We recommend that future work focus on characterizing the disease at a sub-genomic level in different parts of Asia in which genotype D is endemic, and broadening studies to include host factors.

## ACKNOWLEDGMENTS

We would like to extend our thanks to the following who helped us with the serum samples: Professor Necati Ormeci, Medical School of Ankara University; Professor Abdurrahman Kadayifci, Medical School of Gaziantep University; Professor Resat Ozaras, Cerrahpasa Medical School of Istanbul University; Associate Professor Rahmet Caylan, Ankara Ataturk State Hospital; Professor Ahmet Bektas, Medical School of Ondokuz Mayis University (Unit of Gastroenterology); and Associate Professor Hurrem Bodur, Ankara Numune State Hospital.

## DISCLOSURE

All authors have no conflicts of interest.

## REFERENCES

1. Chen D.S. (1993) From hepatitis to hepatoma: lessons from type B viral hepatitis. *Science* **262**: 369–70.
2. Usuda S., Okamoto H., Iwanari H., Baba K., Tsuda F., Miyakawa Y., Mayumi M. (1999) Serological detection of hepatitis B virus genotypes by ELISA with monoclonal antibodies to type-specific epitopes in the preS2-region product. *J Virol Methods* **80**: 97–112.
3. Kao J.H., Chen D.S. (2003) Clinical relevance of hepatitis B virus genotypes Ba and Bj in Taiwan. *Gastroenterology* **125**: 1916–7 (author reply 17-8).
4. Miyakawa Y., Mizokami M. (2003) Classifying hepatitis B virus genotypes. *Intervirology* **46**: 329–38.
5. Kao J.H., Chen P.J., Lai M.Y., Chen D.S. (2003) Basal core promoter mutations of hepatitis B virus increase the risk of hepatocellular carcinoma in hepatitis B carriers. *Gastroenterology* **124**: 327–34.
6. Takahashi K., Ohta Y., Kanai K., Akahane Y., Iwasa Y., Hino K., Ohno N., Yoshizawa H., Mishiro S. (1999) Clinical implications of mutations C-to-T1653 and T-to-C/A/G1753 of hepatitis B virus genotype C genome in chronic liver disease. *Arch Virol* **144**: 1299–308.
7. Kurbanov F., Tanaka Y., Mizokami M. (2010) Geographical and genetic diversity of the human hepatitis B virus. *Hepatol Res* **40**: 14–30.
8. Thakur V., Guptan R.C., Kazim S.N., Malhotra V., Sarin S.K. (2002) Profile, spectrum and significance of HBV genotypes in chronic liver disease patients in the Indian subcontinent. *J Gastroenterol Hepatol* **17**: 165–70.
9. Gandhe S.S., Chadha M.S., Arankalle V.A. (2003) Hepatitis B virus genotypes and serotypes in western India: lack of clinical significance. *J Med Virol* **69**: 324–30.
10. Yalcin K., Degertekin H., Alp M.N., Tekes S., Satici O., Budak T. (2003) Determination of serum hepatitis B virus DNA in chronic HBsAg carriers: clinical significance and correlation with serological markers. *Turk J Gastroenterol* **14**: 157–63.
11. Dienstag J.L. (2008) Hepatitis B virus infection. *N Engl J Med* **359**: 1486–500.
12. Ganem D., Prince A.M. (2004) Hepatitis B virus infection—natural history and clinical consequences. *N Engl J Med* **350**: 1118–29.
13. Sugauchi F., Mizokami M., Orito E., Ohno T., Kato H., Suzuki S., Kimura Y., Ueda R., Butterworth L.A., Cooksley W.G. (2001) A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. *J Gen Virol* **82**: 883–92.
14. Tamura K., Dudley J., Nei M., Kumar S. (2007) MEG A4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–9.
15. Abe A., Inoue K., Tanaka T., Kato J., Kajiyama N., Kawaguchi R., Tanaka S., Yoshiba M., Kohara M. (1999) Quantitation of hepatitis B virus genomic DNA by real-time detection PCR. *J Clin Microbiol* **37**: 2899–903.
16. Tanaka Y., Hasegawa I., Kato T., Orito E., Hirashima N., Acharya S.K., Gish R.G., Kramvis A., Kew M.C., Yoshihara N., Shrestha S. M., Khan M., Miyakawa Y., Mizokami M. (2004) A case-control study for differences among hepatitis B virus infections of genotypes A (subtypes Aa and Ae) and D. *Hepatology* **40**: 747–55.
17. Chen C.H., Lee C.M., Lu S.N., Changchien C.S., Eng H.L., Huang C.M., Wang J.H., Hung C.H., Hu T.H. (2005) Clinical significance of hepatitis B virus (HBV) genotypes and precore and core promoter mutations affecting HBV e antigen expression in Taiwan. *J Clin Microbiol* **43**: 6000–6.
18. Khan A., Tanaka Y., Saito H., Ebinuma H., Sekiguchi H., Iwama H., Wakabayashi G., Kamiya T., Kurbanov F., Elkady A., Mizokami M. (2008) Transmission of hepatitis B virus (HBV) genotypes among Japanese immigrants and natives in Bolivia. *Virus Res* **132**: 174–80.
19. Oyunsuren T., Kurbanov F., Tanaka Y., Elkady A., Sanduijav R., Khajidsuren O., Dagvadorj B., Mizokami M. (2006) High frequency of hepatocellular carcinoma in Mongolia; association with mono-, or co-infection with hepatitis C, B, and delta viruses. *J Med Virol* **78**: 1688–95.
20. Tsatsralt-Od B., Takahashi M., Nishizawa T., Endo K., Inoue J., Okamoto H. (2005) High prevalence of dual or triple infection



© 2012 The Societies and Wiley Publishing Asia Pty Ltd

of hepatitis B, C, and delta viruses among patients with chronic liver disease in Mongolia. *J Med Virol* **77**: 491–9.
21. Elkady A., Tanaka Y., Kurbanov F., Oynsuren T., Mizokami M. (2008) Virological and clinical implication of core promoter C1752/V1753 and T1764/G1766 mutations in hepatitis B virus genotype D infection in Mongolia. *J Gastroenterol Hepatol* **23**: 474–81.
22. Sendi H., Mehrab-Mohseni M., Zali M.R., Norder H., Magnius L.O. (2005) T1764G1766 core promoter double mutants are restricted to Hepatitis B virus strains with an A1757 and are common in genotype D. *J Gen Virol* **86**: 2451–8.
23. Mizokami M., Orito E., Ohba K., Ikeo K., Lau J.Y., Gojobori T. (1997) Constrained evolution with respect to gene overlap of hepatitis B virus. *J Mol Evol* **44** (Suppl 1): S83–90.
24. Aye T.T., Uchida T., Becker S.O., Hirashima M., Shikata T., Komine F., Moriyama M., Arakawa Y., Mima S., Mizokami M., Lau J.Y.N. (1994) Variations of hepatitis B virus precore/core gene sequence in acute and fulminant hepatitis B. *Dig Dis Sci* **39**: 1281–7.
25. Kramvis A., Arakawa K., Yu M.C., Nogueira R., Stram D.O., Kew M.C. (2008) Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. *J Med Virol* **80**: 27–46.
26. Sugiyama M., Tanaka Y., Kato T., Orito E., Ito K., Acharya S.K., Gish R.G., Kramvis A., Shimada T., Izumi N., Kaito M., Miyakawa Y., Mizokami M. (2006) Influence of hepatitis B virus genotypes on the intra- and extracellular expression of viral DNA and antigens. *Hepatology* **44**: 915–24.
27. Ozasa A., Tanaka Y., Orito E., Sugiyama M., Kang J.H., Hige S., Kuramitsu T., Suzuki K., Tanaka E., Okada S., Tokita H., Asahina Y., Inoue K., Kakumu S., Okanoue T., Murawaki Y., Hino K., Onji M., Yatsuhashi H., Sakugawa H., Miyakawa Y., Ueda R., Mizokami M. (2006) Influence of genotypes and precore mutations on fulminant or chronic outcome of acute hepatitis B virus infection. *Hepatology* **44**: 326–34.
28. Sugiyama M., Tanaka Y., Kurbanov F., Maruyama I., Shimada T., Takahashi S., Shirai T., Hino K., Sakaida I., Mizokami M. (2009) Direct cytopathic effects of particular hepatitis B virus genotypes in severe combined immunodeficiency transgenic with urokinase-type plasminogen activator mouse with human hepatocytes. *Gastroenterology* **136**: 652–62 e3.
29. Sugiyama M., Tanaka Y., Kurbanov F., Nakayama N., Mochida S., Mizokami M. (2007) Influences on hepatitis B virus replication by a naturally occurring mutation in the core gene. *Virology* **365**: 285–91.
30. Tanaka Y., Kurbanov F., Mano S., Orito E., Vargas V., Esteban J.I., Yuen M.F., Lai C.L., Kramvis A., Kew M.C., Smuts H.E., Netesov S.V., Alter H.J., Mizokami M. (2006) Molecular tracing of the global hepatitis C virus epidemic predicts regional patterns of hepatocellular carcinoma mortality. *Gastroenterology* **130**: 703–14.
31. Ding X., Park Y.N., Taltavull T.C., Thung S.N., Jin X., Jin Y., Trung N.S., Edamoto Y., Sata T., Abe K. (2003) Geographic characterization of hepatitis virus infections, genotyping of hepatitis B virus, and p53 mutation in hepatocellular carcinoma analyzed by in situ detection of viral genomes from carcinoma tissues: comparison among six different countries. *Jpn J Infect Dis* **56**: 12–8.
32. Zhang H., Zhai Y., Hu Z., Wu C., Qian J., Jia W., Ma F., Huang W., Yu L., Yue W., Wang Z., Li P., Zhang Y., Liang R., Wei Z., Cui Y., Xie W., Cai M., Yu X., Yuan Y., Xia X., Zhang X., Yang H., Qiu W., Yang J., Gong F., Chen M., Shen H., Lin D., Zeng Y. X., He F., Zhou G. (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**: 755–8.

© 2012 The Societies and Wiley Publishing Asia Pty Ltd

# Genome-Wide Association Study Confirming Association of HLA-DP with Protection against Chronic Hepatitis B and Viral Clearance in Japanese and Korean

**Nao Nishida[1,2]*, Hiromi Sawai[2], Kentaro Matsuura[3], Masaya Sugiyama[1], Sang Hoon Ahn[4], Jun Yong Park[4], Shuhei Hige[5], Jong-Hon Kang[6], Kazuyuki Suzuki[7], Masayuki Kurosaki[8], Yasuhiro Asahina[8], Satoshi Mochida[9], Masaaki Watanabe[10], Eiji Tanaka[11], Masao Honda[12], Shuichi Kaneko[12], Etsuro Orito[13], Yoshito Itoh[14], Eiji Mita[15], Akihiro Tamori[16], Yoshikazu Murawaki[17], Yoichi Hiasa[18], Isao Sakaida[19], Masaaki Korenaga[20], Keisuke Hino[20], Tatsuya Ide[21], Minae Kawashima[2], Yoriko Mawatari[1,2], Megumi Sageshima[2], Yuko Ogasawara[2], Asako Koike[22], Namiki Izumi[8], Kwang-Hyub Han[4], Yasuhito Tanaka[3], Katsushi Tokunaga[2], Masashi Mizokami[1]**

**1** Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Chiba, Japan, **2** Department of Human Genetics, The University of Tokyo, Bunkyo-ku, Tokyo, Japan, **3** Department of Virology and Liver Unit, Nagoya City University Graduate School of Medical Sciences, Nagoya, Aichi, Japan, **4** Department of Internal Medicine, Yonsei University College of Medicine, Seoul, South Korea, **5** Department of Internal Medicine, Hokkaido University Graduate School of Medicine, Sapporo, Japan, **6** Department of Internal Medicine, Teine Keijinkai Hospital, Sapporo, Japan, **7** Department of Gastroenterology and Hepatology, Iwate Medical University, Morioka, Japan, **8** Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan, **9** Division of Gastroenterology and Hepatology, Saitama Medical University, Saitama, Japan, **10** Department of Gastroenterology, Kitasato University School of Medicine, Sagamihara, Kanagawa, Japan, **11** Department of Medicine, Shinshu University School of Medicine, Matsumoto, Japan, **12** Department of Gastroenterology, Kanazawa University Graduate School of Medicine, Kanazawa, Japan, **13** Department of Gastroenterology, Nagoya Daini Red Cross Hospital, Nagoya, Japan, **14** Molecular Gastroenterology and Hepatology, Kyoto Prefectural University of Medicine, Kyoto, Japan, **15** Department of Gastroenterology and Hepatology, National Hospital Organization Osaka National Hospital, Osaka, Japan, **16** Department of Hepatology, Osaka City University Graduate School of Medicine, Osaka, Japan, **17** Second Department of Internal Medicine, Faculty of Medicine, Tottori University, Yonago, Japan, **18** Department of Gastroenterology and Metabology, Ehime University Graduate School of Medicine, Ehime, Japan, **19** Gastroenterology and Hepatology, Yamaguchi University Graduate School of Medicine, Yamaguchi, Japan, **20** Division of Hepatology and Pancreatology, Kawasaki Medical College, Kurashiki, Japan, **21** Division of Gastroenterology, Department of Medicine, Kurume University School of Medicine, Fukuoka, Japan, **22** Central Research Laboratory, Hitachi Ltd., Kokubunji, Tokyo, Japan

## Abstract

Hepatitis B virus (HBV) infection can lead to serious liver diseases, including liver cirrhosis (LC) and hepatocellular carcinoma (HCC); however, about 85–90% of infected individuals become inactive carriers with sustained biochemical remission and very low risk of LC or HCC. To identify host genetic factors contributing to HBV clearance, we conducted genome-wide association studies (GWAS) and replication analysis using samples from HBV carriers and spontaneously HBV-resolved Japanese and Korean individuals. Association analysis in the Japanese and Korean data identified the *HLA-DPA1* and *HLA-DPB1* genes with $P_{meta} = 1.89\times10^{-12}$ for rs3077 and $P_{meta} = 9.69\times10^{-10}$ for rs9277542. We also found that the *HLA-DPA1* and *HLA-DPB1* genes were significantly associated with protective effects against chronic hepatitis B (CHB) in Japanese, Korean and other Asian populations, including Chinese and Thai individuals ($P_{meta} = 4.40\times10^{-19}$ for rs3077 and $P_{meta} = 1.28\times10^{-15}$ for rs9277542). These results suggest that the associations between the *HLA-DP* locus and the protective effects against persistent HBV infection and with clearance of HBV were replicated widely in East Asian populations; however, there are no reports of GWAS in Caucasian or African populations. Based on the GWAS in this study, there were no significant SNPs associated with HCC development. To clarify the pathogenesis of CHB and the mechanisms of HBV clearance, further studies are necessary, including functional analyses of the HLA-DP molecule.

**Citation:** Nishida N, Sawai H, Matsuura K, Sugiyama M, Ahn SH, et al. (2012) Genome-Wide Association Study Confirming Association of HLA-DP with Protection against Chronic Hepatitis B and Viral Clearance in Japanese and Korean. PLoS ONE 7(6): e39175. doi:10.1371/journal.pone.0039175

**Editor:** Anand S. Mehta, Drexel University College of Medicine, United States of America

**Received** February 1, 2012; **Accepted** May 16, 2012; **Published** June 21, 2012

**Copyright:** © 2012 Nishida et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by Grants-in-Aid from the Ministry of Health, Labour, and Welfare of Japan (H22-kanen-005, H23-kanen-005), the Japan Science and Technology Agency (09038024), and the Miyakawa Memorial Research Foundation. Partial support by Grant-in-Aid for Young Scientists (B) (22710191) from the Ministry of Education, Culture, Sports, Science, and Technology is also acknowledged. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** AK is an employee of the Central Research Laboratory, Hitachi Ltd. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials, as detailed online in the guide for authors.

* E-mail: nishida-75@umin.ac.jp

## Introduction

Overall, one-third of the world's population (2.2 billion) is infected with hepatitis B virus (HBV), and about 15% of these are chronic carriers. About 75% of the chronic carriers live in the east-south Asia and east pacific area, and there are 1.3–1.5 million chronic carriers living in Japan [1]. Of chronic carriers, 10–15% develop liver cirrhosis (LC), liver failure and hepatocellular carcinoma (HCC), and the remaining individuals eventually achieve a state of nonreplicative infection, resulting in hepatitis B surface antigen (HBsAg) negative and hepatitis B core antibody (anti-HBc) positive, i.e. HBV-resolved individuals [2–3]. In Japan, although the major route of HBV transmission was perinatal transmission and horizontal transmission in early childhood, infant HBV carriers have successfully been reduced since 1986 through a selective vaccination policy by the Japanese government [4–7]. However, the prevalence of HBV genotype A in acute HBV (AHB) infection has increased markedly since 2000, reaching approximately 52% in 2008 due to the lack of a universal HB vaccination, and around 10% of AHB cases could be persistent infection [8–9]. Viral factors, as well as host factors, are thought to be associated with persistent HB infection.

In 2009, significant associations between chronic hepatitis B (CHB) and a region including *HLA-DPA1* and *HLA-DPB1* were identified using 786 Japanese individuals having CHB and 2,201 control individuals through a two-stage genome-wide association study (GWAS) [10]. The same group was also subjected to a second GWAS using a total of 2,667 Japanese persistent HBV infection cases and 6,496 controls, which confirmed significant associations between the *HLA-DP* locus and CHB, in addition to associations with another two SNPs located in the genetic region including the *HLA-DQ* gene [11]. The associations between *HLA-DP* variants with HBV infection were replicated in other Asian populations, including Thai and Han Chinese individuals [10,12–13]. With regard to HBV clearance, the association between the human leukocyte antigen (HLA) class II allele and clearance of HBV was confirmed by the candidate gene approach in African, Caucasian and Asian populations [14–18]. However, in a previous GWAS using samples of Japanese CHB and control individuals, the clinical data on HBV exposure in the control individuals were unknown, and this may have led to bias. Moreover, there have been no reports of GWAS using samples from HBV carriers and HBV-resolved individuals to identify host genetic factors associated with HBV clearance other than HLA class II molecules.

Here, we performed a GWAS using samples from Japanese HBV carriers, healthy controls and spontaneously HBV-resolved individuals in order to confirm or identify the host genetic factors related to CHB and viral clearance. In the subsequent replication analysis, we validated the associated SNPs in the GWAS using two independent sets of Japanese and Korean individuals. In our study, healthy controls were randomly selected with clinically no evidence of HBV exposure, therefore, HBV-resolved individuals were prepared to clearly identify the host genetic factors related with CHB or HBV clearance.

## Results

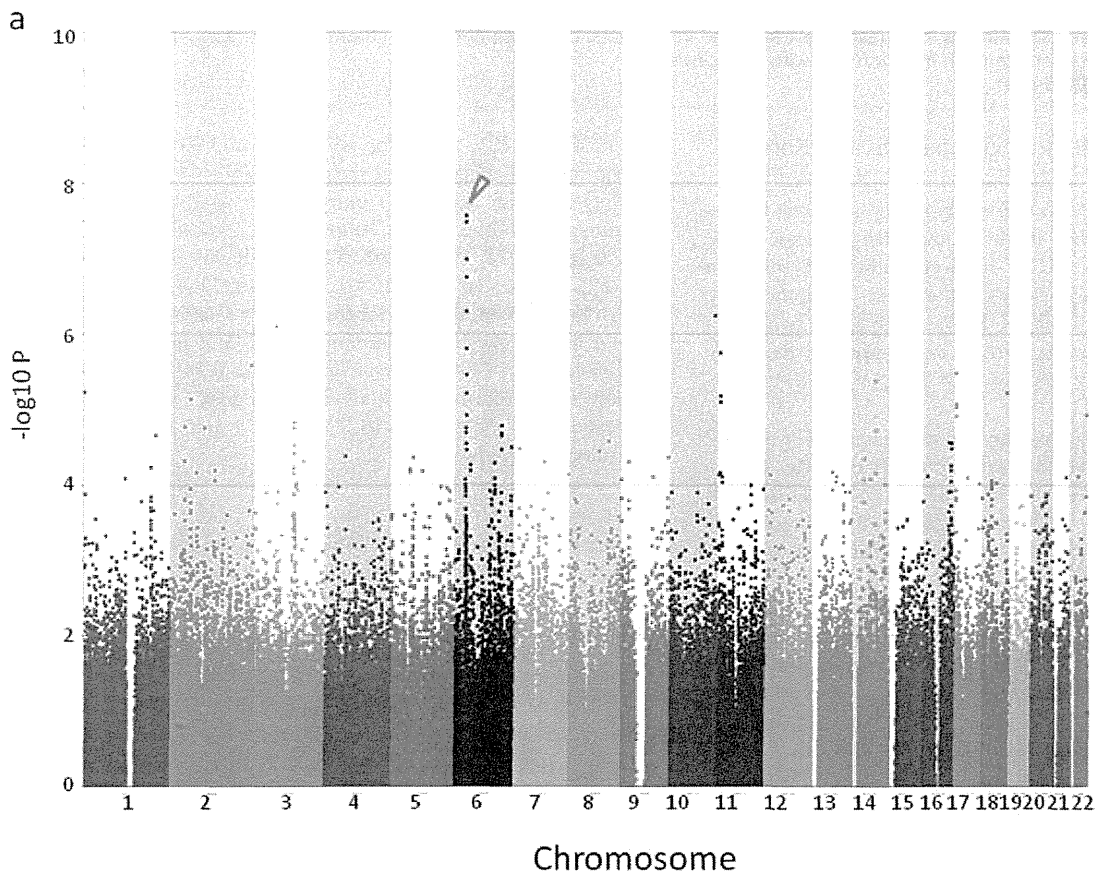
### Protective Effects Against Chronic Hepatitis B in Japanese and Korean Individuals

In this study, we conducted a GWAS using samples from 181 Japanese HBV carriers (including asymptomatic carriers (ASC), CHB cases, LC cases and HCC cases, based on the criteria described in Materials and Methods) and 184 healthy controls in order to identify the host genetic factors related to progression of CHB. All samples were genotyped using a genome-wide SNP typing array (Affymetrix Genome-Wide Human SNP Array 6.0 for 900 K SNPs). Figure 1a shows a genome-wide view of the single point association data based on allele frequencies using the SNPs that met the following filtering criteria: (i) SNP call rate ≥95%; (ii) minor allele frequency (MAF) ≥1% for HBV carriers and healthy controls; and (iii) no deviation from Hardy-Weinberg equilibrium (HWE) $P \geq 0.001$ in healthy controls. We identified significant associations of protective effects against CHB with two SNPs (rs3077 and rs9277542) using the allele frequency model, both of which are located in the 3′ UTR of *HLA-DPA1* and in the sixth exon of *HLA-DPB1*, respectively (rs3077, $P = 1.14\times10^{-7}$, and rs9277542, $P = 5.32\times10^{-8}$, respectively). The association for rs9277542 reached a genome-wide level of significance in the GWAS panel (Bonferroni criterion $P < 8.36\times10^{-8}$ (0.05/597,789)).

In order to validate the results of GWAS, a total of 32 SNPs, including the associated two SNPs (rs3077 and rs9277542), were selected for replication in two independent sets of HBV carriers and healthy controls (replication-1:256 Japanese HBV carriers and 236 Japanese healthy controls; and replication-2:344 Korean HBV carriers and 151 Korean healthy controls; Table 1). The associations for the original significant SNP (rs9277542) and marginal SNP (rs3077) on GWAS were replicated in both replication sets [replication-1 (Japanese); rs3077, $P = 2.70\times10^{-8}$, OR = 0.48 and rs9277542, $P = 3.33\times10^{-6}$, OR = 0.54; replication-2 (Korean); rs3077, $P = 2.08\times10^{-6}$, OR = 0.47 and rs9277542, $P = 8.29\times10^{-5}$, OR = 0.54, Table 2]. We conducted meta-analysis to combine these studies using the DerSimonian Laird method (random effects model) to incorporate variation among studies. As shown in Table 2, the odds ratios were quite similar across the three studies (GWAS and two replication studies) and no heterogeneity was observed ($P_{het} = 0.80$ for rs3077 and 0.40 for rs9277542). $P_{meta}$ values were $4.40\times10^{-19}$ for rs3077 (OR = 0.46, 95% confidence interval (CI) = 0.39–0.54), and $1.28\times10^{-15}$ for rs9277542 (OR = 0.50, 95% CI = 0.43–0.60). Among the remaining 30 SNPs in the replication study, 27 SNPs were successfully genotyped by the DigiTag2 assay with SNP call rate ≥ 95% and HWE *p*-value ≥ 0.01. Two SNPs (rs9276431 and rs7768538), located in the genetic region including the *HLA-DQ* gene, were marginally replicated in the two sets of HBV carriers and healthy controls with Mantel-Haenszel *P* values of $2.80\times10^{-7}$ (OR = 0.56, 95% CI = 0.45–0.70) and $1.09\times10^{-7}$ (OR = 0.53, 95% CI = 0.42–0.67), respectively, when using additive, two-tailed Cochran Mantel-Haenszel (CMH) fixed-effects model with no evidence of heterogeneity ($P_{het} = 0.67$ for rs9276431 and 0.70 for rs7768538) (Table S1).

Meta-analysis using the random effects model across 6 independent studies, including 5 additional published data, showed $P_{meta} = 3.94\times10^{-45}$, OR = 0.55 for rs3077, $P_{meta} = 1.74\times10^{-21}$, OR = 0.61 for rs9277535 and $P_{meta} = 1.69\times10^{-15}$, OR = 0.51 for rs9277542, with the SNP rs9277535 being located about 4-kb upstream from rs9277542 and showing strong linkage disequilibrium of $r^2 = 0.955$ on the HapMap JPT (Table S2). As shown in Table S2, the odds ratio was very similar among the 6 studies, and heterogeneity was negligible with $P_{het} > 0.01$.

Moreover, based on GWAS using samples from 94 chronic HBV carriers with LC or HCC and 87 chronic HBV carriers without LC and HCC, we found no significant SNPs associated with CHB progression (Figure S1).


a
-log10 P
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Chromosome

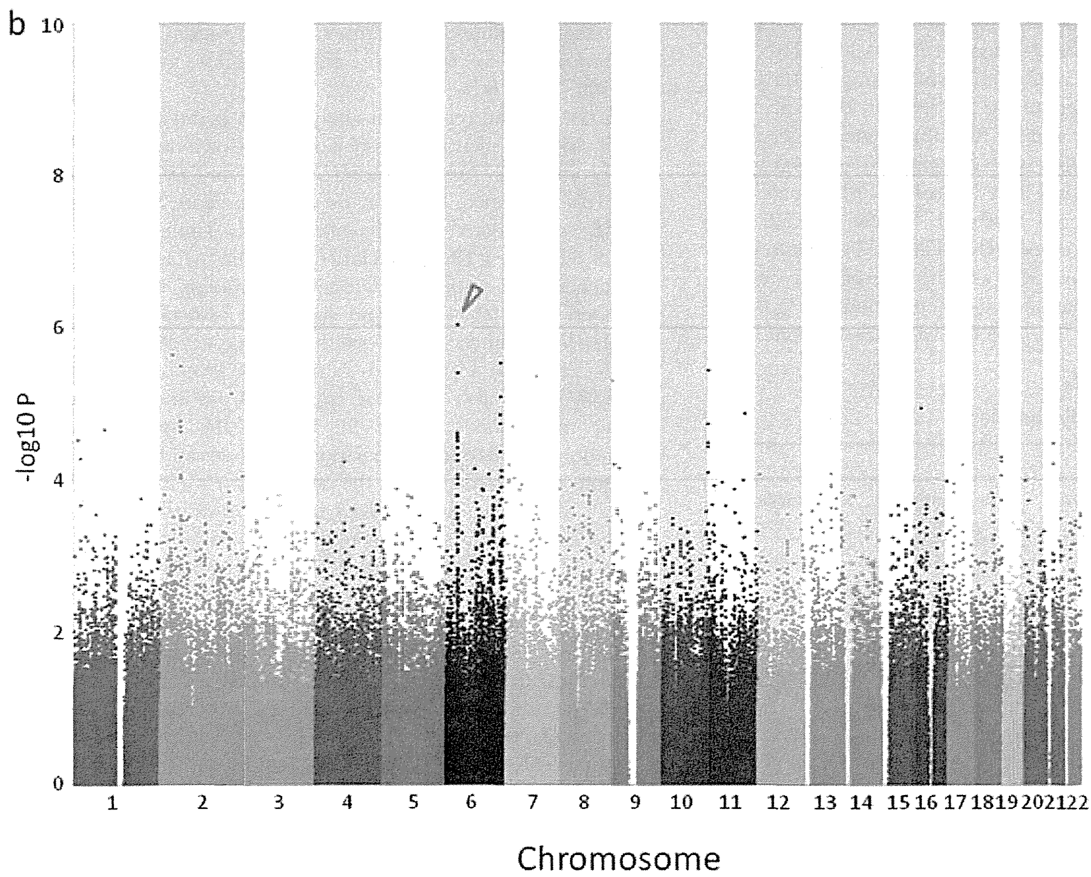

b
-log10 P
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Chromosome

**Figure 1. Results of genome-wide association studies.** a) HBV carriers and healthy controls, and b) HBV carriers and HBV-resolved individuals were compared. *P* values were calculated by chi-squared test for allele frequencies. Dots with arrows on chromosome 6 show strong associations with protective effects against persistent HB infection and with HBV clearance.
doi:10.1371/journal.pone.0039175.g001

## Clearance of Hepatitis B virus in Japanese and Korean Individuals

We also conducted a GWAS to identify the host genetic factors related to clearance of HBV in the above 181 Japanese HBV carriers and 185 Japanese HBV-resolved individuals using a genome-wide SNP typing array (Affymetrix Genome-Wide Human SNP Array 6.0 for 900 K SNPs). The same two SNPs (rs3077 and rs9277542) showed strong associations in the allele frequency model ($P = 9.24\times10^{-7}$ and $P = 3.15\times10^{-5}$) with clearance of HBV (Figure 1b).

The above 32 SNPs, including the two associated SNPs (rs3077 and rs9277542), were selected for a replication study in two independent sets of HBV carriers and HBV resolved individuals (replication-1:256 Japanese HBV carriers and 150 Japanese HBV resolved individuals; and replication-2:344 Korean HBV carriers and 106 Korean HBV resolved individuals; Table 1). All 32 SNPs were genotyped using the DigiTag2 assay and 29 of 32 SNPs were successfully genotyped (Table S3). The associations of the original SNPs were replicated in both replication sets [replication-1 (Japanese): rs3077, $P = 3.32\times10^{-2}$, OR = 0.72 and rs9277542, $P = 1.25\times10^{-2}$, OR = 0.68; replication-2 (Korean): rs3077, $P = 2.35\times10^{-7}$, OR = 0.41 and rs9277542, $P = 4.97\times10^{-6}$, OR = 0.46; Table 3]. Meta-analysis using random effects model showed $P_{meta} = 1.56\times10^{-4}$ for rs3077 (OR = 0.51, 95% CI = 0.36–0.72), and $5.91\times10^{-7}$ for rs9277542 (OR = 0.55, 95% CI = 0.43–0.69). While there was evidence of heterogeneity between these studies for rs3077 ($P_{het} = 0.03$) and no evidence for rs9277542 ($P_{het} = 0.19$), significant associations with HBV clearance were observed with Mantel-Haenszel $P_{meta} = 3.28\times10^{-12}$ for rs3077 and $1.42\times10^{-10}$ for rs9277542, when using CMH fixed-effects model. Among the remaining 27 SNPs in the replication study, two SNPs (rs9276431 and rs7768538), located in a genetic region including *HLA-DQ* gene, were marginally replicated in the two sets of HBV carriers and HBV resolved individuals with Mantel-Haenszel *P* values of $2.10\times10^{-5}$ (OR = 0.59) and $1.10\times10^{-5}$ (OR = 0.56), respectively (Table S3), when using CMH fixed-effect model. Due to the existing heterogeneity among three groups (GWAS, Replication-1 and Replication-2) ($P_{het} = 0.03$ for rs9276431 and 0.04 for rs7768538), weak associations were observed with $P_{meta} = 0.03$ for rs9276431 and 0.02 for rs7768538 by the random effects model meta-analysis.

**Table 1.** Number of study samples.

| | | GWAS | Replication-1 | Replication-2 |
|---|---|---|---|---|
| **population** | | **Japanese** | **Japanese** | **Korean** |
| HBV carriers | Total | **181** | **256** | **344** |
| | IC | 20 | 94 | – |
| | CH | 67 | 101 | 177 |
| | LC | 3 | 10 | – |
| | HCC | 91 | 51 | 167 |
| Healthy controls | | **184** | **236** | **151** |
| Resolved individuals | | **185** | **150** | **106** |

Abbreviation: IC, Inactive Carrier; CH, Chronic Hepatitis; LC, Liver Cirrhosis; HCC, Hepatocellular Carcinoma.
doi:10.1371/journal.pone.0039175.t001

Meta-analysis across 6 independent studies, including 5 additional published data, showed $P_{meta} = 1.48\times10^{-9}$, OR = 0.60 for rs3077, $P_{meta} = 1.08\times10^{-17}$, OR = 0.66 for rs9277535 and $P_{meta} = 5.14\times10^{-5}$, OR = 0.55 for rs9277542 (Table S4). As shown in Table S4, the OR for the rs9277535 and rs9277542 were similar among the 6 independent studies, and heterogeneity was negligible ($P_{het} = 0.03$ for rs9277535 and 0.14 for rs9277542). However, significant level of heterogeneity for rs3077 was observed with $P_{het} = 9.57\times10^{-6}$ across 5 independent studies, including our study.

## URLs

The results of the present GWAS are registered at a public database: https://gwas.lifesciencedb.jp/cgi-bin/gwasdb/gwas_top.cgi.

## Discussion

The recent genome-wide association study showed that the SNPs located in a genetic region including *HLA-DPA1* and *HLA-DPB1* genes were associated with chronic HBV infection in the Japanese and Thai population [10,11]. In this study, we confirmed a significant association between SNPs (rs3077 and rs9277542) located in the same genetic region as *HLA-DPA1* and *HLA-DPB1* and protective effects against CHB in Korean and Japanese individuals. Mata-analysis using the random effects model across 6 independent studies including our study suggested that, widely in East Asian populations, variants in antigen binding sites of *HLA-DP* contribute to protective effects against persistent HBV infection (Table S2).

On GWAS and replication analysis with Japanese and Korean individuals, we identified associations between the same SNPs (rs3077 and rs9277542) in the *HLA-DPA1* and *HLA-DPB1* genes and HBV clearance; however, no new candidate SNPs from the GWAS were detected on replication analysis (Table S3). When the data of reference#18 was excluded from the meta-analysis across 6 independent studies, heterogeneity among 4 studies was estimated to be $P_{het} = 0.15$ and significant association of rs3077 with HBV clearance was observed with $P_{meta} = 5.88\times10^{-24}$, OR = 0.56 (Table S4). In our study, a negligible level of heterogeneity for rs3077 was also observed ($P_{het} = 0.03$) on meta-analysis by adding replication-1 (Table 3). Despite the heterogeneity in replication-1, a marginal association was observed for rs3077 with the same downward trend in the odds ratio ($P = 3.32\times10^{-2}$, OR = 0.72). Moreover, meta-analysis using GWAS and replication-2 showed significant association of $P_{meta} = 1.89\times10^{-12}$, OR = 0.43 for rs3077 with no evidence of heterogeneity ($P_{het} = 0.75$). Although the reason why heterogeneity was observed in replication-1 is unclear, one possible reason is the clinical heterogeneity due to different kits being used for antibody testing. The associations of *HLA-DPA1/−DPB1* with CHB and HBV clearance showed the same level of significance in the comparison of HBV patients with HBV resolved individuals (OR = 0.43 for rs3077 and 0.49 for rs9277542) as the one with healthy controls (OR = 0.46 for rs3077 and 0.50 for rs9277542), when the replication-1 was excluded in the analysis (Table 2 and Table 3). The results of meta-analysis across 6 independent studies including our study also showed the same or slightly weaker associations in the

**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 | $1.14\times10^{-7}$ | |
| | | | | (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 | $2.70\times10^{-8}$ | |
| | | | | | | (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 | $2.08\times10^{-6}$ | |
| | | | | | | (Korean) | (10.7) | (37.7) | (51.6) | (21.4) | (51.0) | (27.6) | | (0.35–0.65) | | |
| | | | | | | Meta-analysis[e] | | | | | | | | 0.46 | $4.40\times10^{-19}$ | 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 | $5.32\times10^{-8}$ | |
| | | | | (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 | $3.33\times10^{-6}$ | |
| | | | | | | (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 | $8.29\times10^{-5}$ | |
| | | | | | | (Korean) | (14.2) | (41.2) | (44.5) | (24.5) | (50.3) | (25.2) | | (0.40–0.74) | | |
| | | | | | | Meta-analysis[e] | | | | | | | | 0.50 | $1.28\times10^{-15}$ | 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 | $9.24\times10^{-7}$ | |
| | | | | (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 | $3.32\times10^{-2}$ | |
| | | | | | | (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 | $2.35\times10^{-7}$ | |
| | | | | | | (Korean) | (10.7) | (37.7) | (51.6) | (27.6) | (45.7) | (26.7) | (0.29–0.58) | | |
| | | | | | | Meta-analysis[e] | | | | | | | 0.51 | $1.56\times10^{-4}$ | 0.03 |
| | | | | | | | | | | | | | (0.36–0.72) | | |
| | | | | | | Meta-analysis[e] | | | | | | | 0.43 | $1.89\times10^{-12}$ | 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 | $3.15\times10^{-5}$ | |
| | | | | (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 | $1.25\times10^{-2}$ | |
| | | | | | | (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 | $4.97\times10^{-6}$ | |
| | | | | | | (Korean) | (14.2) | (41.2) | (44.5) | (28.6) | (50.5) | (21.0) | (0.33–0.64) | | |
| | | | | | | Meta-analysis[e] | | | | | | | 0.55 | $5.91\times10^{-7}$ | 0.19 |
| | | | | | | | | | | | | | (0.43–0.69) | | |
| | | | | | | Meta-analysis[e] | | | | | | | 0.49 | $9.69\times10^{-10}$ | 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 $\lambda$ 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 $\lambda$ 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.

Sawai *et al. BMC Medical Genetics* 2012, **13**:47
http://www.biomedcentral.com/1471-2350/13/1/47

BMC Medical Genetics

RESEARCH ARTICLE
Open Access

# No association for Chinese HBV-related hepatocellular carcinoma susceptibility SNP in other East Asian populations

Hiromi Sawai[1*], Nao Nishida[1,2], Hamdi Mbarek[3], Koichi Matsuda[3], Yoriko Mawatari[2], Megumi Yamaoka[1], Shuhei Hige[4], Jong-Hon Kang[5], Koichi Abe[6], Satoshi Mochida[7], Masaaki Watanabe[8], Masayuki Kurosaki[9], Yasuhiro Asahina[9], Namiki Izumi[9], Masao Honda[10], Shuichi Kaneko[10], Eiji Tanaka[11], Kentaro Matsuura[12], Yoshito Itoh[13], Eiji Mita[14], Masaaki Korenaga[15], Keisuke Hino[15], Yoshikazu Murawaki[16], Yoichi Hiasa[17], Tatsuya Ide[18], Kiyoaki Ito[2], Masaya Sugiyama[2], Sang Hoon Ahn[19], Kwang-Hyub Han[19], Jun Yong Park[19], Man-Fung Yuen[20], Yusuke Nakamura[3], Yasuhito Tanaka[12], Masashi Mizokami[2] and Katsushi Tokunaga[1]

**Abstract**

**Background:** A recent genome-wide association study (GWAS) using chronic HBV (hepatitis B virus) carriers with and without hepatocellular carcinoma (HCC) in five independent Chinese populations found that one SNP (rs17401966) in *KIF1B* was associated with susceptibility to HCC. In the present study, a total of 580 HBV-derived HCC cases and 1351 individuals with chronic hepatitis B (CHB) or asymptomatic carrier (ASC) were used for replication studies in order to evaluate the reported association with HBV-derived HCC in other East Asian populations.

**Results:** We did not detect any associations between rs17401966 and HCC in the Japanese cohorts (replication 1: OR = 1.09, 95 % CI = 0.82-1.43; replication 2: OR = 0.79, 95 % CI = 0.54-1.15), in the Korean cohort (replication 3: OR = 0.95, 95 % CI = 0.66-1.36), or in the Hong Kong Chinese cohort (replication 4: OR = 1.17, 95 % CI = 0.79-1.75). Meta-analysis using these cohorts also did not show any associations with P = 0.97.

**Conclusions:** None of the replication cohorts showed associations between rs17401966 and HBV-derived HCC. This may be due to differences in the genetic diversity among the Japanese, Korean and Chinese populations. Other reasons could be the high complexity of multivariate interactions between the genomic information and the phenotype that is manifesting. A much wider range of investigations is needed in order to elucidate the differences in HCC susceptibility among these Asian populations.

**Keywords:** Hepatitis B, hepatocellular carcinoma, candidate SNP, replication study, genome-wide association study

## Background

Hepatitis B (HB) is a potentially life-threatening liver infection caused by the hepatitis B virus (HBV), and approximately 360 million people worldwide are thought to be chronically infected with HBV. The clinical course of HBV infection is variable, including acute self-limiting infection, fulminant hepatic failure, inactive carrier state and chronic hepatitis with progression to cirrhosis and hepatocellular carcinoma (HCC). Although some HBV carriers spontaneously eliminate the virus, 2-10 % of individuals with chronic HB (CHB) develop liver cirrhosis every year, and a subset of these individuals suffer from liver failure or HCC. Around 600,000 new HCC cases are diagnosed annually worldwide, with HCC being relatively common in Asia-Pacific countries and sub-Saharan Africa; more than 70 % of HCC patients are diagnosed in Asia (with 55 % in China) [1]. However, HCC is relatively uncommon in the USA, Europe and Australia [1,2]. The majority of HCC develops in patients with cirrhosis, which is most often attributable

* Correspondence: sawai@m.u-tokyo.ac.jp
[1]Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
Full list of author information is available at the end of the article

BioMed Central
© 2012 Sawai et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

to chronic HBV infection followed by chronic HCV in the Asia-Pacific region [3].

A recent genome-wide association study (GWAS) using Japanese CHB cases and controls confirmed that 11 SNPs in a region including *HLA-DPA1* and *-DPB1* were associated with CHB [4]. Moreover, a GWAS using chronic HBV carriers with and without HCC in five independent Chinese populations reported that one SNP (rs17401966) in *KIF1B* was associated with HCC susceptibility [5]. In the present study, we performed replication studies using Japanese, Korean and Hong Kong Chinese cases and controls in order to evaluate the reported association with HBV-derived HCC in other East Asian populations.

## Results

We performed SNP genotyping of rs17401966 located in the *KIF1B* gene for the purpose of replication analysis of the previous GWAS report [5]. Four distinct cohorts were used for these replication analyses (Table 1). We first examined two independent Japanese case–control samples including 179 cases and 769 controls from Biobank Japan (replication 1), and 142 cases and 251 controls from various hospitals (replication 2). We did not detect any associations between rs17401966 and HCC in the Japanese cohorts (replication 1: OR = 1.09; 95 % CI = 0.82-1.43, replication 2: OR = 0.79; 95 % CI = 0.54-1.15). We further examined Korean case–control samples comprising 164 cases and 144 controls (replication 3) and Hongkongese 94 HCC cases and 187 CHB controls (replication 4), but again did not detect any association (replication 3: OR = 0.95; 95 % CI = 0.66-1.36, replication 4: OR = 1.17; 95 % CI = 0.79-1.75). Logistic regression analysis adjusted for age and gender also did not show any association ($P_{log}$ = 0.65, 0.27, 0.11, 0.56 for each replication panel). Moreover, we conducted meta-analysis to combine these studies, also not detect any association ($P_{meta}$ = 0.97).

## Discussion and conclusions

Zhang et al. [5] reported that SNP rs17401966 was significantly associated with HBV-related HCC (joint OR = 0.61). They conducted a GWAS using 348 cases and 359 controls in a population in Guangxi in southern China, and selected 45 SNPs for the replication study based on the results ($P < 10^{-4}$). In the first replication study, they used 276 cases and 266 controls from Beijing in northern China, and 5 SNPs showed the same direction of association as in the GWAS ($P < 0.05$). They performed a further replication study (of 507 cases and 215 controls) in Jiangsu in eastern China and only one SNP showed the same trend ($P = 3.9 \times 10^{-5}$). Guangdong and Shanghai samples from southern and eastern China were used for further replication studies. The association yielded a p-value of $1.7 \times 10^{-18}$ on meta-analysis.

We performed four replication analyses using Japanese, Korean and Hong Kong Chinese samples (Table 1). Although sample size of each cohort is smaller than that of the previous GWAS, we conducted mete-analysis of all our study. The result did not show any association between rs17401966 and HBV-derived HCC ($P_{meta}$ = 0.97).

This may be due to differences in genetic diversity among Japanese, Korean and Chinese populations. A maximum-likelihood tree of 126 populations based on 19,934 SNPs showed that Japanese and Korean populations form a monophyletic clade with a 100 % bootstrap value [6]. However, Chinese populations form a paraphyletic clade with two other populations. This indicates that Japanese and Korean populations are genetically closer to one another than the Chinese population.

**Table 1 Association between rs17401966 and HBV-derived HCC**

| cohort | sample size | cases | | | controls | | | | OR | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | (cases/controls) | GG | AG | AA | GG | AG | AA | HWE p | (95 % CI) | $P$ [a] | $P_{het}$ [b] |
| replication 1 | 179/769 | 13 | 61 | 105 | 45 | 261 | 463 | 0.599 | 1.09 | 0.578 | |
| (Japan 1) | | (7.2) | (34.1) | (58.7) | (5.9) | (33.9) | (60.2) | | (0.82-1.43) | | |
| replication 2 | 142/251 | 5 | 46 | 91 | 14 | 91 | 146 | 1 | 0.79 | 0.212 | |
| (Japan 2) | | (3.5) | (32.4) | (64.1) | (5.6) | (36.2) | (58.2) | | (0.54-1.15) | | |
| replication 3 | 164/144 | 17 | 59 | 88 | 15 | 55 | 74 | 0.616 | 0.95 | 0.790 | |
| (Korea) | | (10.4) | (36.0) | (53.6) | (10.4) | (38.2) | (51.4) | | (0.66-1.36) | | |
| replication 4 | 94/187 | 10 | 39 | 44 | 13 | 80 | 94 | 0.767 | 1.17 | 0.432 | |
| (Hong Kong) | | (10.6) | (41.5) | (46.8) | (6.9) | (42.8) | (50.3) | | (0.79-1.75) | | |
| Meta-analysis [c] | | | | | | | | | 0.996 | 0.965 | 0.423 |
| | | | | | | | | | (0.84-1.18) | | |

[a]P value of fisher's exact test for allele model.
[b]Result of Breslow-Day test.
[c]Results of meta-analysis were calculated by the Mantel-Haenzel method.

We did not find any association with Hong Kong Chinese cohort (P = 0.43). Moreover, a study using 357 HCC cases and 354 HBV-positive non-HCC controls in Hong Kong Chinese did not show any significant difference (P = 0.91) [7]. Previous population studies have revealed that various Han Chinese populations show varying degrees of admixture between a northern Altaic cluster and a southern cluster of Sino-Tibetan/Tai-Kadai populations in southern China and northern Thailand [6]. Although Hong Kong is located closed to the Guangdong (cohort 3 of Zhang et al study), there is great heterogeneity for rs17401966 between Hong Kong cohorts (our study and Chan's study [7]) and Guangdong cohort (our study versus Zhang's study: $P_{het} = 0.0066$; Chan's study versus Zhang's study: $P_{het} = 0.035$). This result suggests the existence of other confounding factors, which can differentiate the previous study in China and this study.

One of the possible reasons could be the high complexity of multivariate interactions between the genomic information and the phenotype that is manifesting. HCC development is a multiple process which links to causative factors such as age, gender, environmental toxins, alcohol and drug abuse, higher HBV DNA levels, and HBV genotype variations [8]. The eight HBV genotypes display distinct geographical and ethnic distributions. Genotypes B and C are prevalent in Asia. Specific variations in HBV have been associated with cirrhosis and HCC. These variations include in particular mutations in pre-core region (Pre-C), in basal core promoter (BCP) and in ORF encoding Pre-S1/Pre-S2/S and Pre-C/C. Because there is an overlap between Pre-C or BCP mutations and genotypes, these mutations appear to be more common in genotype C as compared to other genotypes [9].

Aflatoxins are a group of 20 related metabolites and Aflatoxin B1 is the most potent naturally occurring chemical liver carcinogen known. Aflatoxin exposures multiplicatively increase the risk of HCC in people chronically infected with HBV, which illustrates the deleterious impact that even low toxin levels in the diet can have on human health [10–12]. Liu and Wu estimated population risk for aflatoxin-induced HCC around the world [13]. Most cases occur in sub-Saharan Africa, Southeast Asia and China, where populations suffer from both high HBV prevalence and largely uncontrolled exposure to aflatoxin in food. But we could not obtain the information of these confounding factors from both of the previous GWAS study and this study. A much wider range of investigations is thus needed in order to elucidate the differences in HCC susceptibility among these Asian populations.

## Methods

### Samples

Case and control samples used in this study were collected from Japan, Korea and Hong Kong listed in supplementary Additional file 1: Table S1. A total of 179 cases and 769 control subjects were analyzed in the first replication study. DNA samples from both CHB controls and HBV-related HCC cases used in this study were obtained from the BioBank Japan at the Institute of Medical Science, the University of Tokyo [14]. Among the BioBank Japan samples, we selected HBsAg-seropositive CHB patients with elevated serum aminotransferase levels for more than six months, according to the guidelines for diagnosis and treatment of chronic hepatisis from The Japan Society of Hepatology (http://www.jsh.or.jp/medical/gudelines/index.html). The mean (and standard deviation; SD) age was 62.0 (9.4) years for the cases and 54.7 (13.5) years for the controls. The second Japanese replication sample sets for the cases (n = 142) and controls (n = 251) study were obtained from 16 hospitals. The case samples for the second replication included 142 HCC patients and the controls included 135 CHB patients and 116 asymptomatic carriers (ASC). The mean (SD) age was 61.3 (10.2) years for the cases and 56.2 (10.9) years for the controls. The Korean replication samples were collected from Yonsei University College of Medicine. The third replication set was composed of 165 HCC patients and 144 CHB patients. The mean (SD) age was 52.2 (8.9) and 37.3 (11.3) years for the cases and controls, respectively. The samples in Hong Kong were collected from the University of Hong Kong, Queen Mary Hospital. The fourth replication set was composed of 94 HCC patients and 187 CHB patients. The mean (SD) age was 58.0 (10.5) and 56.9 (8.3) years for the cases and controls, respectively. All participants provided written informed consent. This research project was approved by the Research Ethics Committees at the Institute of Medical Science and the Graduate School of Medicine, the University of Tokyo, Yonsei University College of Medicine, the University of Hong Kong, Nationa Center for Global Health and Medicine, Hokkaido University Graduate School of Medicine, Teine Keijinkai Hospital, Iwate Medical University, Saitama Medical University, Kitasato University School of Medicine, Musashino Red Cross Hospital, Kanazawa University Graduate School of Medicine, Shinshu University School of Medicine, Nagoya City University Graduate School of Medical Sciences, Kyoto Prefectural University of Medicine, National Hospital Organization Osaka National Hospital, Kawasaki Medical College, Tottori University, Ehime University Graduate School of Medicine, and Kurume University School of Medicine.

### SNP Genotyping

For the first replication samples, we genotyped rs17401966 using PCR-based Invader assay (Third Wave Technologies, Madison, WI) [15], and for the second, third and fourth replication samples, we used TaqMan genotyping assay (Applied Biosystems, Carlsbad, CA). In the TaqMan SNP

genotyping assay, PCR amplification was performed in a 5-μl reaction mixture containing 1 μl of genomic DNA, 2.5 μl of KAPA PROBE FAST qPCR Master Mix (Kapa Biosystems, Woburn, MA), and 40 x TaqMan SNP Genotyping Assay probe (ABI) for this SNP. QPCR thermal cycling was performed as follows: 95°C for 3 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. The SNP call rate of each replication panel was 100 %, 100 %, 99.7 % and 99.6 %.

### Statistical analysis

We performed Hardy-Weinberg equilibrium test for the case and control samples in each replication study. Fisher's exact test was applied to two-by-two contingency tables for three different genetic models; allele frequency, dominant and recessive model. Odds ratios and confidence intervals were calculated using the major alleles as references. Meta-analysis was conducted using the Mantel-Haenszel method. Heterogeneity among studies was examined by using the Breslow-Day test. Genotype-phenotype association for the SNP rs17401966 was assessed using logistic regression analysis adjusted for age and gender in plink 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml).

## Additional file

**Additional file 1: Table S1.** Samples used in this study.

#### Abbreviations

HB: Hepatitis b; HBV: Hepatitis b virus; HCC: Hepatocellular carcinoma; CHB: Chronic hepatitis b; HCV: Hepatitis c virus; GWAS: Genome-wide association study; ASC: Asymptomatic carrier.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

This study was supported by a grant-in-aid from the Ministry of Health, Labour, and Welfare of Japan (H23-kanen-005), and Japan Science and Technology Agency (09038024). We thank Dr. Minae Kawashima to giving us technical advices.

#### Author details

[1]Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [2]The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan. [3]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [4]Department of Internal Medicine, Hokkaido University Graduate School of Medicine, Sapporo, Japan. [5]Department of Internal Medicine, Teine Keijinkai Hospital, Sapporo, Japan. [6]First Department of Internal Medicine, Iwate Medical University, Iwate, Japan. [7]Division of Gastroenterology and Hepatology, Internal Medicine, Saitama Medical University, Saitama, Japan. [8]Department of Gastroenterology, Kitasato University School of Medicine, Sagamihara, Kanagawa, Japan. [9]Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan. [10]Department of Gastroenterology, Kanazawa University Graduate School of Medicine, Kanazawa, Japan. [11]Department of Medicine, Shinshu University School of Medicine, Matsumoto, Japan. [12]Department of Clinical Molecular Informative Medicine, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan. [13]Molecular Gastroenterology and Hepatology, Kyoto Prefectural University of Medicine, Kyoto, Japan. [14]National Hospital Organization Osaka National Hospital, Osaka, Japan. [15]Division of Hepatology and Pancreatology, Kawasaki Medical College, Kurashiki, Japan. [16]Second department of Internal Medicine, Faculty of Medicine, Tottori University, Yonago, Japan. [17]Department of Gastroenterology and Metabology, Ehime University Graduate School of Medicine, Ehime, Japan. [18]Division of Gastroenterology, Department of Medicine, Kurume University School of Medicine, Fukuoka, Japan. [19]Department of International Medicine, Yonsei University College of Medicine, Seoul, Korea. [20]Department of Medicine, the University of Hong Kong, Queen Mary Hospital, Hong Kong, China.

#### Author contributions

Study design and discussion: H.S., N.N., Y.T., Ko.M., M.M., K.T.; sample collection: Y.T., Ko.M., Y.N., S.H.A., K.H.H., J.Y.P., M.F.Y., S.H., J.H.K., K.A., S.M., M.W., M.Ku., Y.A., N. I., M.H., S.K., E.T., Ke.M., Y.I., E.M., M.Ko., K.H., Y.Mu., Y.H., T.I., K.I., M.S., M.M.; genotyping: H.S., Y.M., M.Y., H.M.; statistical analysis: H.S.; manuscript writing: H.S., N.N., Y.T., M.M., K.T. All authors read and approved the final manuscript.

Received: 2 March 2012 Accepted: 19 June 2012
Published: 19 June 2012

#### References

1. Parkin DM, Bray F, Ferlay J, Pisani P: **Global cancer statistics, 2002.** *CA: a cancer journal for clinicians* 2005, **55**(2):74–108.
2. Parkin DM: **Global cancer statistics in the year 2000.** *The lancet oncology* 2001, **2**(9):533–543.
3. Marrero CR, Marrero JA: **Viral hepatitis and hepatocellular carcinoma.** *Archives of medical research* 2007, **38**(6):612–620.
4. Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A, Hosono N, Kubo M, Tsunoda T, Kamatani N, Kumada H, *et al*: **A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians.** *Nature genetics* 2009, **41**(5):591–595.
5. Zhang H, Zhai Y, Hu Z, Wu C, Qian J, Jia W, Ma F, Huang W, Yu L, Yue W, *et al*: **Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers.** *Nature genetics* 2010, **42**(9):755–758.
6. Abdulla MA, Ahmed I, Assawamakin A, Bhak J, Brahmachari SK, Calacal GC, Chaurasia A, Chen CH, Chen J, Chen YT *et al*: **Mapping human genetic diversity in Asia.** *Science (New York, NY* 2009, **326**(5959):1541–1545.
7. Chan KY, Wong CM, Kwan JS, Lee JM, Cheung KW, Yuen MF, Lai CL, Poon RT, Sham PC, Ng IO: **Genome-wide association study of hepatocellular carcinoma in Southern Chinese patients with chronic hepatitis B virus infection.** *PLoS One* 2011, **6**(12):e28798.
8. Sherman M: **Hepatocellular carcinoma: epidemiology, surveillance, and diagnosis.** *Semin Liver Dis* 2010, **30**(1):3–16.
9. Yang HI, Yeh SH, Chen PJ, Iloeje UH, Jen CL, Su J, Wang LY, Lu SN, You SL, Chen DS, *et al*: **Associations between hepatitis B virus genotype and mutants and the risk of hepatocellular carcinoma.** *J Natl Cancer Inst* 2008, **100**(16):1134–1143.
10. Qian GS, Ross RK, Yu MC, Yuan JM, Gao YT, Henderson BE, Wogan GN, Groopman JD: **A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China.** *Cancer Epidemiol Biomarkers Prev* 1994, **3**(1):3–10.
11. Ross RK, Yuan JM, Yu MC, Wogan GN, Qian GS, Tu JT, Groopman JD, Gao YT, Henderson BE: **Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma.** *Lancet* 1992, **339**(8799):943–946.
12. Wang LY, Hatch M, Chen CJ, Levin B, You SL, Lu SN, Wu MH, Wu WP, Wang LW, Wang Q, *et al*: **Aflatoxin exposure and risk of hepatocellular carcinoma in Taiwan.** *International journal of cancer* 1996, **67**(5):620–625.
13. Liu Y, Wu F: **Global burden of aflatoxin-induced hepatocellular carcinoma: a risk assessment.** *Environmental health perspectives* 2010, **118**(6):818–824.
14. Nakamura Y: **The BioBank Japan Project.** *Clin Adv Hematol Oncol* 2007, **5** (9):696–697.
15. Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y: **A high-throughput SNP typing system for genome-wide association studies.** *Journal of human genetics* 2001, **46**(8):471–477.

doi:10.1186/1471-2350-13-47
**Cite this article as:** Sawai *et al.*: **No association for Chinese HBV-related hepatocellular carcinoma susceptibility SNP in other East Asian populations.** *BMC Medical Genetics* 2012 **13**:47.

# Morphobase, an Encyclopedic Cell Morphology Database, and Its Use for Drug Target Identification

Yushi Futamura,[1,3] Makoto Kawatani,[1,3] Sayaka Kazami,[1] Kenichi Tanaka,[1] Makoto Muroi,[1] Takeshi Shimizu,[1] Koji Tomita,[1,2] Nobumoto Watanabe,[1] and Hiroyuki Osada[1,2,*]
[1]Chemical Biology Core Facility, Chemical Biology Department, RIKEN Advanced Science Institute, Wako-shi, Saitama 351-0198, Japan
[2]Graduate School of Biomedical Sciences, Tokyo Medical and Dental University, Tokyo 113-8510, Japan
[3]These authors contributed equally to this work
*Correspondence: hisyo@riken.jp
http://dx.doi.org/10.1016/j.chembiol.2012.10.014

## SUMMARY

**Visual observation is a powerful approach for screening bioactive compounds that can facilitate the discovery of attractive druggable targets following their chemicobiological validation. So far, many high-content approaches, using sophisticated imaging technology and bioinformatics, have been developed. In our study, we aimed to develop a simpler method that focuses on intact cell images because we found that dynamic changes in morphology are informative, often reflecting the mechanism of action of a drug. Here, we constructed a chemical-genetic phenotype profiling system, based on the high-content cell morphology database Morphobase. This database compiles the phenotypes of cancer cell lines that are induced by hundreds of reference compounds, wherein those of well-characterized anticancer drugs are classified by mode of action. Furthermore, we demonstrate the applicability of this system in identifying NPD6689, NPD8617, and NPD8969 as tubulin inhibitors.**

## INTRODUCTION

With the completion of the sequencing of the human genome and the demonstration of dramatic clinical efficacy for emerging molecular-target drugs, particularly imatinib, the trend in anticancer drug discovery has shifted to a molecular target-based approach. Whereas target-based screenings are believed to be the most rational and powerful, classical phenotype screenings, such as morphology-based assay, have been reconsidered and reintroduced as a complementary strategy for drug discovery, especially in the exploration of first-in-class therapeutics. For instance, certain benzoquinone ansamycins, such as herbimycin and geldanamycin, were found to revert tyrosine kinase-dependent oncogenic transformation (Uehara et al., 1986), the molecular target of which Whitesell et al. (1994) identified as HSP90. Similarly, lactacystin, discovered as an inducer of differentiation in Neuro2A cells, was demonstrated to target the 20S proteasome, indicating that proteasome inhibitors generally have selective anticancer activity (Fenteany et al., 1995; Omura et al., 1991).

Accordingly, phenotype-based screens are yet recognized as a new method of drug discovery, and newly modified phenotypic screens such as high-throughput, high-content imaging-driven, and omics-based screens have been developed as the initial step in the discovery of small molecule probes and drugs (Feng et al., 2009; Houle et al., 2010; Roti and Stegmaier, 2012). With this renewed interest in phenotype screens, however, the deduction of a molecular target of putative hits still remains the rate-limiting stage, despite the significant technological advances in chemical biology. The success rate of identifying a compound with a specific mechanism of action and its efficacy depends on our ability to compile and analyze the knowledge of the possible molecular targets related to phenotypes. In this regard, a multidimensional cell-based phenotype profiling approach that is supported by genetic modification, e.g., genome-wide gene depletion (genetics) or the application of well-validated chemotherapeutics (chemical genetics), can offer a promising strategy for discovering new drugs and defining their mechanisms of action. This strategy has allowed us and other groups to establish such profiling systems and report many important findings in recent years. For example, Perlman et al. (2004) presented a high-throughput cytological profiling method, in which the effects of 100 compounds were examined at 13 3-fold dilution of each drug on various cellular components in HeLa cells. To assess the similarity of compounds, they integrated a titration-invariant similarity score to succeed in grouping compounds with similarly reported targets and classifying a poorly characterized austocystin to transcription and translation inhibitors (Perlman et al., 2004). The CalMorph system, developed by Ohya and colleagues, which compiled comprehensive yeast cell morphological profiles of 4,718 nonessential gene deletion mutants, deduced the cellular functions of well-characterized drugs by comparing the morphological changes that were induced by gene deletion and versus the test compound (Ohnuki et al., 2010). Abassi et al. (2009) developed a cell-based kinetic profiling approach using impedance readout to monitor the effects of approximately 260 small molecules. This approach has been used to analyze the calcium-modulating activity of celecoxib, a Cox2 inhibitor, and identify an additional mechanism (calcium channel inhibitory activity) for the Eg5 inhibitor, monastrol (Abassi et al., 2009). Similarly,

 ©2012 Elsevier Ltd All rights reserved

Westwick group performed high-content protein-fragment complementation assays to measure changes in protein complexes in response to 107 drugs, characterizing the off-target effects and hidden phenotypes of certain drugs (MacDonald et al., 2006). Recently, we reported a proteome-based profiling system to predict the mechanism of action and molecular target of a compound of interest (Muroi et al., 2010) and demonstrated that BNS22 targets the catalytic domain of DNA topoisomerase II (TOP2). This new type of TOP2 inhibitor is an attractive chemotherapeutic agent as an alternative to topoisomerase poisons such as etoposide (Kawatani et al., 2011).

Even so, a primary screening is merely the starting point, and one must challenge how to exploit unique bioactive substances from a huge library rapidly and simply. To this end, we propose chemical-genetic cell morphological profiling as a promising method. The characteristic changes in morphology on exposure to an agent are often associated with its mechanism of action. Moreover, our data indicate that intact morphological changes alone—without excessive immunological or fluorescent staining—are sufficient to profile drug responses (Osada et al., 1988, 1997). In this study, using two cancer cell lines, we constructed an encyclopedia of cellular morphology, Morphobase, consisting of the cell shape changes induced by various compounds. Specifically, we developed a high-content image analysis method to examine the effects of approximately 200 routinely used chemicals on morphology, accompanied by statistical characterization of the obtained phenotypic data. The importance of this database was verified by its identification of the mechanism of action of three candidate drugs, comparing the similarity in morphological features by test compounds and those in the training data set. The Morphobase strategy not only reproduces "drug-target-phenotype" relationships for drugs with known targets, it also predicts unreported mechanisms of action and facilitates the discovery of novel drug candidates.

## RESULTS

### Tumor Cells Undergo Dynamic Changes Related to the Mode of Action of a Drug

To construct the cell morphology database, we examined the effects of roughly 30 well-characterized drugs on cell shape in two mammalian cell lines—the human cervix epidermoid carcinoma cell line HeLa and rat kidney cells that were infected with ts25, a T class mutant of Rous sarcoma virus Prague strain, *src*$^{ts}$-NRK cells—over time and by dose. As summarized in Figures 1 and S1A (available online), the cells that were exposed to an agent underwent typical morphological changes, which were similar between compounds of a pharmacological class. For instance, protein synthesis inhibitors, such as reveromycin A, cycloheximide, and anisomycin, induced the reversion of the spherical shape of transformed *src*$^{ts}$-NRK cells to a flattened shape, characteristic of normal cells (Takahashi et al., 1992). The morphological changes in *src*$^{ts}$-NRK cells were classified roughly into three categories—(1) flattened, (2) polygonal, and (3) rounded—and subgrouped by variation in size and the presence of unique granular structures, spikes, and vacuoles (Figure S1B). In contrast, the morphological changes in HeLa cells were less complicated, falling into three broad classes: (1) flattened, (2) round up, and (3) toxic/growth inhibition, which usually reflect stages in the cell cycle (G1/S phase, flattened with embossed nucleus; G2/M, round up). These preliminary observations indicated that the data on one cell line were not sufficient to characterize each therapeutic group of anticancer agents. However, the phenotypic response data of two or more distinct cell lines can discriminate features that are related to the mechanism of action of a particular compound.

### High-Content Image Analysis of Subtle and Complex Changes in Morphology

To eliminate unintentional human errors during visual inspection, we recorded morphological changes using an automatic system, the IN Cell Analyzer, to perform a high-content image analysis. This system can automatically recognize cytoskeletal morphology and subcellular components and systematically generate quantitative morphological data.

To manage the vast variety of phenotypes that are induced by mechanistically distinct compounds, a custom-made image analysis algorithm was designed using IN Cell Developer Toolbox (GE Healthcare). On the first attempt, bright-field cell images were collected by the IN Cell Analyzer, but it appeared to be difficult to define every single cell accurately. Thus, nuclear staining with Hoechst 33342 was introduced to define the boundary between the cells and background. The unique textures that were induced by a drug on the cell surface or inside the cell were defined by the descriptor termed "Granular" for the segmentation of these components (see Figure 2A and Experimental Procedures for detailed segmentation algorithm). Using this method, various types of morphological changes in *src*$^{ts}$-NRK and HeLa cells that were induced by small molecules—the sizes and shapes, either of single cell or even tightly packed cells—could be successfully segmented from one another, confirming that this general cellular segmentation algorithm can monitor complex phenotypic responses in relevant cancer cell types (Figure 2B). Figure 2C shows the flowchart of image processing and measurements of morphological parameters for the next data mining step. Once the bright-field and Hoechst images of *src*$^{ts}$-NRK or HeLa cells that were exposed to various drugs were acquired, the "Nuclear," "Cell," and "Granular" fragments were properly segmented, and 12 morphological parameters were output automatically.

### Statistical Analysis of Multiparametric Phenotype Data and Classification by Mode of Action

A test set of 207 authentic compounds, supplied by the RIKEN Natural Products Depository (NPDepo) (http://www.npd.riken.jp/npd) and SCADS inhibitor kits, was created (Table S1). Of these compounds, 54 were drugs with well-characterized mechanisms of action, 118 were commonly used experimental drugs, 26 were FDA-approved drugs, 7 were reported to have multiple biological targets, and 2 were antibiotics. To standardize the conditions for monitoring the resulting morphological changes, the following growth inhibitory concentrations were used: $IC_{85-95}$ values for HeLa cells and $IC_{50-60}$ for *src*$^{ts}$-NRK cells, calculated after a 48 hr exposure. The corresponding data are shown in Table S1. For compounds with low cytotoxicity, a maximum concentration of 25 $\mu$M was chosen for the exposure. Following incubation, cells were fixed and stained with