# Model Incorporating the *ITPA* Genotype Identifies Patients at High Risk of Anemia and Treatment Failure With Pegylated-Interferon Plus Ribavirin Therapy for Chronic Hepatitis C

Masayuki Kurosaki,[1] Yasuhito Tanaka,[2] Nao Nishida,[3] Naoya Sakamoto,[4] Nobuyuki Enomoto,[5] Kentaro Matsuura,[2] Yasuhiro Asahina,[6] Mina Nakagawa,[6] Mamoru Watanabe,[6] Minoru Sakamoto,[5] Shinya Maekawa,[5] Katsushi Tokunaga,[3] Masashi Mizokami,[7] and Namiki Izumi[1]*

[1]*Department of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan*
[2]*Department of Virology, Liver Unit, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan*
[3]*Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan*
[4]*Department of Gastroenterology and Hematology, Hokkaido University, Sapporo, Japan*
[5]*First Department of Internal Medicine, University of Yamanashi, Yamanashi, Japan*
[6]*Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Tokyo, Japan*
[7]*Research Center for Hepatitis and Immunology, International Medical Center of Japan Konodai Hospital, Ichikawa, Japan*

This study aimed to develop a model for predicting anemia using the inosine triphosphatase (*ITPA*) genotype and to evaluate its relationship with treatment outcome. Patients with genotype 1b chronic hepatitis C (n = 446) treated with peg-interferon alpha and ribavirin (RBV) for 48 weeks were genotyped for the *ITPA* (rs1127354) and *IL28B* (rs8099917) genes. Data mining analysis generated a predictive model for anemia (hemoglobin (Hb) concentration <10 g/dl); the CC genotype of *ITPA*, baseline Hb <14.0 g/dl, and low creatinine clearance (CLcr) were predictors of anemia. The incidence of anemia was highest in patients with Hb <14.0 g/dl and CLcr <90 ml/min (76%), followed by Hb <14.0 g/dl and *ITPA* CC (57%). Patients with Hb ≥14.0 g/dl and *ITPA* AA/CA had the lowest incidence of anemia (17%). Patients with two predictors (high-risk) had a higher incidence of anemia than the others (64% vs. 28%, *P* < 0.0001). At baseline, the *IL28B* genotype was a predictor of a sustained virological response [adjusted odds ratio 9.88 (95% confidence interval 5.01–19.48), *P* < 0.0001]. In patients who achieved an early virological response, the *IL28B* genotype was not associated with a sustained virological response, while a high risk of anemia was a significant negative predictor of a sustained virological response [0.47 (0.24–0.91), *P* = 0.026]. For high-risk patients with an early virological response, giving >80% of the planned RBV dose increased sustained virological responses by 24%. In conclusion, a predictive model incorporating the *ITPA* genotype could identify patients with a high risk of anemia and reduced probability of sustained virological response. *J. Med. Virol.* 85:449–458, 2013.
© 2013 Wiley Periodicals, Inc.

KEY WORDS: hemolytic anemia; ribavirin; creatinine clearance; antiviral therapy

## INTRODUCTION

Hepatitis C virus (HCV) infection is a leading cause of cirrhosis and hepatocellular carcinoma worldwide [Kim, 2002]. The rate of eradication of HCV by pegylated interferon (PEG-IFN) plus ribavirin (RBV), defined as a sustained virological response, is around 50% in patients with HCV genotype 1 [Manns et al., 2001; Fried et al., 2002]. Failure of treatment is attributable to the lack of a virological response or relapse after completion of therapy. Genome-wide association studies and subsequent cohort studies

have shown that single nucleotide polymorphisms (SNPs) located near the *IL28B* gene are the most important determinant of virological response to PEG-IFN/RBV therapy [Ge et al., 2009; Suppiah et al., 2009; Tanaka et al., 2009; Rauch et al., 2010]. On the other hand, among patients with a virological response, the probability of a sustained virological response decreases when the patients become intolerant to therapy because of RBV-induced hemolytic anemia and receive a reduced dose of RBV [McHutchison et al., 2002; Kurosaki et al., 2012]. Genome-wide association studies have shown that variants of the inosine triphosphatase (*ITPA*) gene protect against hemolytic anemia [Fellay et al., 2010; Tanaka et al., 2011]. These variants are associated with a reduced requirement for an anemia-related dose reduction of RBV [Sakamoto et al., 2010; Thompson et al., 2010a; Kurosaki et al., 2011d; Seto et al., 2011]. However, factors other than the *ITPA* gene also contribute to the risk of severe anemia or RBV dose reduction [Ochi et al., 2010; Kurosaki et al., 2011d] and the results of studies on the impact of the *ITPA* genotype on treatment outcome are inconsistent [Ochi et al., 2010; Sakamoto et al., 2010; Thompson et al., 2010a, 2011; Kurosaki et al., 2011d].

Data mining is a novel statistical method used to extract relevant factors from a plethora of factors and combine them to predict the incidence of the outcome of interest [Breiman et al., 1980]. Decision tree analysis, a primary component of data mining analysis, has found medical applications recently [Averbook et al., 2002; Miyaki et al., 2002; Baquerizo et al., 2003; Leiter et al., 2004; Garzotto et al., 2005; Zlobec et al., 2005; Valera et al., 2007] and has proven to be a useful tool for predicting therapeutic efficacy [Kurosaki et al., 2010, 2011a,b,c, 2012] and adverse events [Hiramatsu et al., 2011] in patients with chronic hepatitis C treated with PEG-IFN/RBV therapy. Because the results of data mining analysis are presented as a flowchart [LeBlanc and Crowley, 1995], they are easily understandable and usable by clinicians lacking a detailed knowledge of statistics.

For the general application of this genetic information in clinical practice, this study aimed to construct a predictive model of severe anemia using the *ITPA* genotype, together with other relevant factors. This study also aimed to analyze the impact of the risk of anemia on treatment outcome, after adjustment for the *IL28B* genotype. These analyses were carried out at baseline and during therapy, when the early virological response became evident.

## MATERIALS AND METHODS

### Patients

Data were collected from a total of 446 genotype 1b chronic hepatitis C patients who were treated with PEG-IFN alpha and RBV at five hospitals and universities throughout Japan. The inclusion criteria were: (1) infection by hepatitis C genotype 1b; (2) no

co-infection with hepatitis B virus or human immunodeficiency virus; (3) no other causes of liver disease such as autoimmune hepatitis and primary biliary cirrhosis; and (4) availability of DNA for the analysis of the genetic polymorphisms of *IL28B* and *ITPA*. Patients received PEG-IFN alpha-2a (180 µg) and 2b (1.5 µg/kg) subcutaneously every week and a daily weight-adjusted dose of RBV (600 mg for patients weighing <60 kg, 800 mg for patients weighing 60–80 kg, and 1,000 mg for patients weighing >80 kg) for 48 weeks. Dose reduction or discontinuation of PEG-IFN and RBV was primarily based on the recommendations on the package inserts and the discretion of the physicians at each university and hospital. The standard duration of therapy was set at 48 weeks. No patient received erythropoietin or other growth factors for the treatment of anemia. Written informed consent was obtained from each patient, and the study protocol conformed to the ethical guidelines of the Declaration of Helsinki and was approved by the institutional ethics review committees.

### Laboratory Tests

Blood samples obtained before therapy were analyzed for hematologic data, blood chemistry, and HCV RNA. Genetic polymorphisms in SNPs of the *ITPA* gene (rs1127354) and the *IL28B* gene (rs8099917) were determined using ABI TaqMan Probes (Applied Biosystems, Carlsbad, CA) and the DigiTag2 assay, respectively. Baseline creatinine clearance (CLcr) levels were calculated using the formula of Cockcroft and Gault [1976]: for males, CLcr = [(140 − age in years) × body weight in kg] ÷ (72 × serum creatinine in mg/dl) and for females, CLcr = 0.85 × [(140 − age in years) × body weight in kg] ÷ (72 × serum creatinine in mg/dl). The stage of liver fibrosis was scored according to the METAVIR scoring system: F0 (no fibrosis), F1 (mild fibrosis: portal fibrosis without septa), F2 (moderate fibrosis: few septa), F3 (severe fibrosis: numerous septa without cirrhosis), and F4 (cirrhosis). A rapid virological response was defined as undetectable HCV RNA by qualitative PCR with a lower detection limit of 50 IU/ml (Amplicor, Roche Diagnostic Systems, Pleasanton, CA) at week 4 of therapy and a complete early virological response was defined as undetectable HCV RNA at week 12. A sustained virological response was defined as undetectable HCV RNA at 24 weeks after completion of therapy. Severe anemia was defined as hemoglobin (Hb) <10 g/dl.

### Statistical Analysis

Database for analysis included the following variables: age, sex, body mass index, serum aspartate aminotransferase (AST) levels, alanine aminotransferase (ALT) levels, gamma-glutamyltransferase (GGT) levels, creatinine levels, CLcr, Hb, platelet count, serum levels of HCV RNA, and the stage of liver fibrosis

TABLE I. Patients' Baseline Characteristics

| | | |
|---|---|---|
| Age (years) | 58.6 | (9.6) |
| Gender: male (n, %) | 185 | (42%) |
| Body mass index (kg/m$^2$) | 23.1 | (3.7) |
| AST (IU/L) | 59.9 | (53.8) |
| ALT (IU/L) | 69.8 | (53.8) |
| GGT (IU/L) | 48.5 | (41.6) |
| Creatinine (mg/dl) | 0.7 | (0.2) |
| Creatinine clearance (ml/min) | 89.5 | (23.0) |
| Hemoglobin (g/dl) | 14 | (1.4) |
| Platelet count (10$^9$/L) | 154.5 | (52.1) |
| HCV RNA > 600,000 IU/ml (n, %) | 354 | (79%) |
| Liver fibrosis: F3-4 (n, %) | 108 | (24%) |
| Initial ribavirin dose (n, %) | | |
|    600 mg/day | 300 | (67%) |
|    800 mg/day | 138 | (31%) |
|    1,000 mg/day | 9 | (2%) |
| Pegylated interferon (n, %) | | |
|    alpha2a 180 mcg | 58 | (13%) |
|    alpha2b 1.5 mcg/kg | 388 | (87%) |
| ITPA rs1127354: CC (n, %) | 317 | (71%) |
| IL28B rs809917: TT (n, %) | 311 | (70%) |

AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma-glutamyltransferase.
Data expressed as mean (standard deviation) unless otherwise mentioned.

(Table I). Based on these data set, a model for predicting the risk of developing severe anemia was constructed by data mining analysis using the IBM-SPSS Modeler 13 as described previously [Kurosaki et al., 2010, 2011a,b,c; Hiramatsu et al., 2011]. Briefly, the software was used to explore the database automatically to search for optimal predictors that discriminated most efficiently patients with severe anemia from those without. The software also determined the optimal cutoff values of each predictor. Patients were divided into two groups according to the predictor and each of the two groups was repeatedly divided in the same way until no significant factor remained or 20 or fewer patients were in a group.

The incidence of severe anemia, the total dose of RBV, and treatment outcome were compared between groups with high and low risks of anemia. On univariate analysis, Student's t-test was used for continuous variables, and Fisher's exact test was used for categorical data. Logistic regression was used for multivariate analysis. P values of <0.05 were considered significant. SPSS Statistics 18 was used for these analyses.

## RESULTS

### Predictive Model of Severe Anemia

The incidence of severe anemia in the whole cohort was 49% (Fig. 1). The best predictor of severe anemia was the baseline Hb concentration. Patients with a low baseline Hb concentration (<14 g/dl) were more likely to develop severe anemia (67%) than those with a higher Hb (>14 g/dl) (34%). The second best predictor for those patients with a baseline Hb <14.0 g/dl was CLcr. Patients with a CLcr below 90 ml/min had

the highest incidence of severe anemia (76%). In those with a CLcr above >90 ml/min the incidence of severe anemia was 57% in patients with the CC allele of the ITPA gene while it was 37% in patients with the CA or AA allele. On the other hand, the second best predictor for those patients with a baseline Hb concentration above 14 g/dl was the ITPA genotype. Patients with the AA or AC allele had the lowest incidence of anemia (17%). For those with the ITPA CC allele, CLcr was the third best predictor; the optimal cutoff value was 85 ml/min for this group. The incidence of severe anemia was 49% in patients with a CLcr below 85 ml/min while it was 32% in those with a CLcr above 85 ml/min.

Following this analysis, the patients were divided into six groups, with the incidence of severe anemia ranging from 17% to 76%. Three groups with two predictors, having an incidence of anemia >40%, were defined as the high-risk group and the remainder were defined as the low-risk group. The incidence of severe anemia was higher in the high-risk group than the low-risk group (65% vs. 28%, P = 0.029) (Fig. 2). Comparison of the ITPA genotype and the predictive model showed that the sensitivity for the prediction of severe anemia was similar (75.9% vs. 76.4%) but the specificity of the predictive model was greater (33.6% vs. 59.3%).

### The Risk of Anemia Impacts on Sustained Virological Responses by Patients Who Achieved an Early Virological Response

The impact of IL28B genotype, ITPA genotype, and risk group of anemia on the rate of sustained virological response was studied at baseline and week 12. At baseline, patients with the TT allele of the IL28B gene had a significantly higher rate of sustained virological response than those with the TG or GG allele (43% vs. 10%, P < 0.0001), the high-risk group for anemia had a significantly lower rate of sustained virological response than the low-risk group (28% vs. 40%, P = 0.011), and the ITPA genotype was not associated with a sustained virological response (Fig. 3A–C). At week 4, patients with rapid virological response had a high rate of sustained virological response, irrespective of the IL28B genotype (TT vs. TG/GG; 97% vs. 100%, P = 1.000), the ITPA genotype (CC vs. CA/AA; 95% vs. 100%, P = 1.000), and the risk of anemia (high vs. low; 95% vs. 100%, P = 1.000). Among the patients who did not achieve a rapid virological response, those with the IL28B TT allele had a significantly higher rate of sustained virological response than those with the TG or GG allele (38% vs. 8%, P < 0.0001), and the high-risk group for anemia had a significantly lower rate of sustained virological response than the low-risk group (24% vs. 35%, P = 0.015). At week 12, in patients who achieved a complete early virological response, the IL28B genotype was not associated with a sustained virological response, while the high-risk group for anemia had a
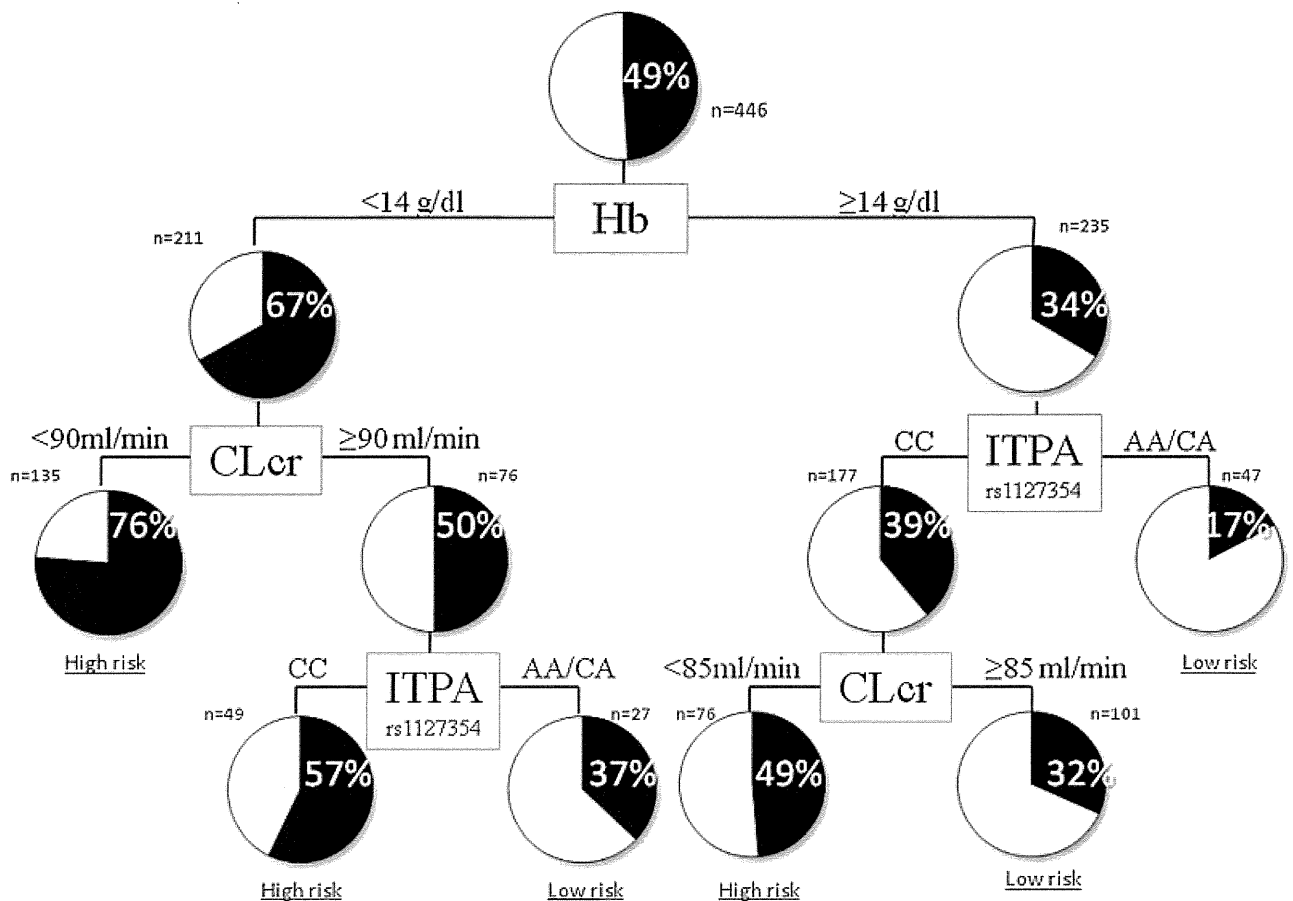
Fig. 1. The predictive model for severe anemia. The boxes indicate the factors used to differentiate patients and the cutoff values for the different groups. The pie charts indicate the rate of severe anemia (Hb <10.0 g/dl) for each group of patients, after differentiation. Terminal groups of patients differentiated by analysis are classified as at high risk if the rate is >40% and low risk if the rate is <40%. ITPA, inosine triphosphatase; CLcr, creatinine clearance; Hb, hemoglobin.

significantly lower rate of sustained virological response than the low-risk group (59% vs. 76%, $P = 0.013$) (Fig. 3D–F). In patients who did not achieve a complete early virological response, the *IL28B* genotype was a significant predictor of a sustained virological response (TT vs. TG/GG; 14% vs. 2%, $P < 0.0001$) but a high risk for anemia was not (high vs. low; 10% vs. 6%, $P = 0.361$).

From multivariate analysis (Table II), the *IL28B* genotype was the most important predictor of a sustained virological response at baseline [adjusted odds ratio 9.88 (95% confidence interval 5.01–19.48), $P < 0.0001$], along with female sex [0.42 (0.26–0.68), $P < 0.0001$], platelet count [1.09 (1.04–1.15), $P < 0.0001$], advanced fibrosis [0.49 (0.27–0.91), $P = 0.024$], and baseline HCV RNA load [4.14 (2.27–7.55), $P < 0.0001$]. At week 4, in patients without a rapid virological response, the *IL28B* genotype remained the most important predictor of a sustained virological response [7.16 (3.60–14.25), $P < 0.0001$], along with female sex and platelet count. At week 12, in patients with a complete early virological response, the risk of anemia was an independent and significant

predictor of a sustained virological response [0.47 (0.24–0.91), $P = 0.026$], together with the platelet count and HCV RNA load, but the *IL28B* genotype was not associated with a sustained virological response. In patients without a complete early virological response, the *IL28B* genotype was a predictor of a sustained virological response [9.13 (2.02–41.3), $P = 0.004$] along with the platelet count. Thus, *IL28B* was a significant predictor of a sustained virological response at baseline and among virological non-responders at weeks 4 and 12. On the other hand, once a complete early virological response was achieved, the *IL28B* genotype was no longer associated with a sustained virological response but the risk of anemia was an independent predictor of a sustained virological response.

### The Risk of Anemia, RBV Dose, and Treatment Outcome in Patients With a Complete Early Virological Response

Patients who achieved a complete early virological response were stratified according to adherence to
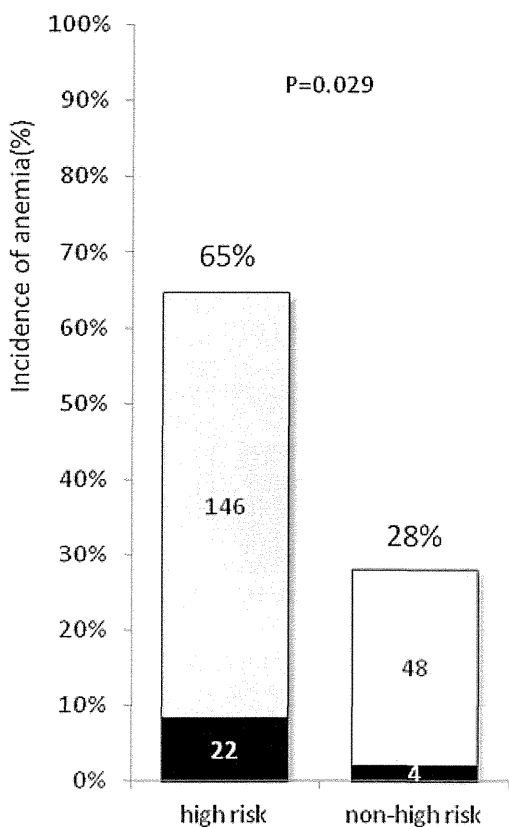
Fig. 2. The incidence of severe anemia stratified by risk of anemia. The incidence of anemia during therapy is shown for each group of patients at high and low risk of anemia. The black and white bars represent the percentages of patients with Hb concentrations below 8.5 g/dl and above 10 g/dl, respectively.

RBV ($\leq$40%, 41–60%, 61–80%, and >80%), which showed that patients with a high risk of anemia were predominantly in subgroups with a lower adherence to RBV ($\leq$40%, 41–60%, and 61–80%), whereas patients with a low risk of anemia were predominantly in subgroups with a higher adherence to RBV (>80%) (Fig. 4, upper panel). The percentage of patients who received >80% of the planned dose of RBV was significantly higher in the low-risk group for anemia than in the high-risk group (74% vs. 55%, $P < 0.0001$).

Within the groups with high and low risks of anemia, there was a stepwise increase in the rate of sustained virological response according to the increase in adherence to RBV (Fig. 4, lower panel). The rate of sustained virological response was higher in patients who received >80% of the planned dose of RBV than those who received less, for both high-risk patients (71% vs. 47%, $P = 0.016$) and low-risk patients (81% vs. 60%, $P = 0.072$). Within the same subgroup of RBV adherence, however, the rate of sustained virological response did not differ between patients with a high risk and a low risk of anemia. Taken together, these results suggest that patients with a high risk of anemia have a disadvantage because they are likely

to be intolerant to RBV, leading to reduced adherence to RBV throughout the 48 weeks of therapy and a reduced rate of sustained virological response. However, if >80% adherence to RBV could be obtained, the rate of sustained virological response would increase by 24%.

## DISCUSSION

This study confirmed previous reports that the *IL28B* genotype is the most significant predictor of a sustained virological response to PEG-IFN plus RBV therapy in chronic hepatitis C patients at baseline [Ge et al., 2009; Suppiah et al., 2009; Tanaka et al., 2009; Rauch et al., 2010; Kurosaki et al., 2011c] and at week 4 [Thompson et al., 2010b], but it had no impact on the rate of sustained virological response among those patients who achieved a complete early virological response [Thompson et al., 2010b; Kurosaki et al., 2011c]. In contrast, the risk of anemia, assessed by the combination of the *ITPA* genotype, baseline Hb concentration, and baseline CLcr, was found to be associated with a sustained virological response in patients who achieved a complete early virological response. Generally, a complete early virological response is the hallmark of a high probability of a sustained virological response, but the rate of sustained virological responses in patients who achieved a complete early virological response and had a high risk of anemia was as low as 59%. This reduced rate of sustained virological response in these patients was attributable to poor adherence to RBV throughout the 48 weeks of therapy. Because administration of >80% of the planned RBV dose increased the rate of sustained virological response by 24%, it may be postulated that personalizing the treatment schedule to achieve a sufficient dose of RBV, such as extension of treatment duration, may improve sustained virological response rates in these patients. Clearly, this postulate needs to be confirmed in future study. Thus, the findings presented here may have the potential to support selection of the optimum, personalized treatment strategy for an individual patient, based on risk of anemia.

The degree of hemolytic anemia caused by RBV varies among individuals. A reduction of the Hb concentration early during therapy predicts the likely development of severe anemia [Hiramatsu et al., 2008, 2011] but there are no reliable predictors at baseline. A breakthrough came from the results of a genome-wide association study that revealed that variants of the *ITPA* gene are protective against hemolytic anemia [Fellay et al., 2010]. The *ITPA* genotype has been shown repeatedly to be associated with the degree of hemolytic anemia and dose reduction of RBV [Fellay et al., 2010; Sakamoto et al., 2010; Thompson et al., 2010a; Seto et al., 2011; Tanaka et al., 2011; Kurosaki et al., 2011d]. However, factors other than the *ITPA* gene, such as baseline Hb concentrations [Ochi et al., 2010; Kurosaki et al., 2011d], platelet counts [Ochi
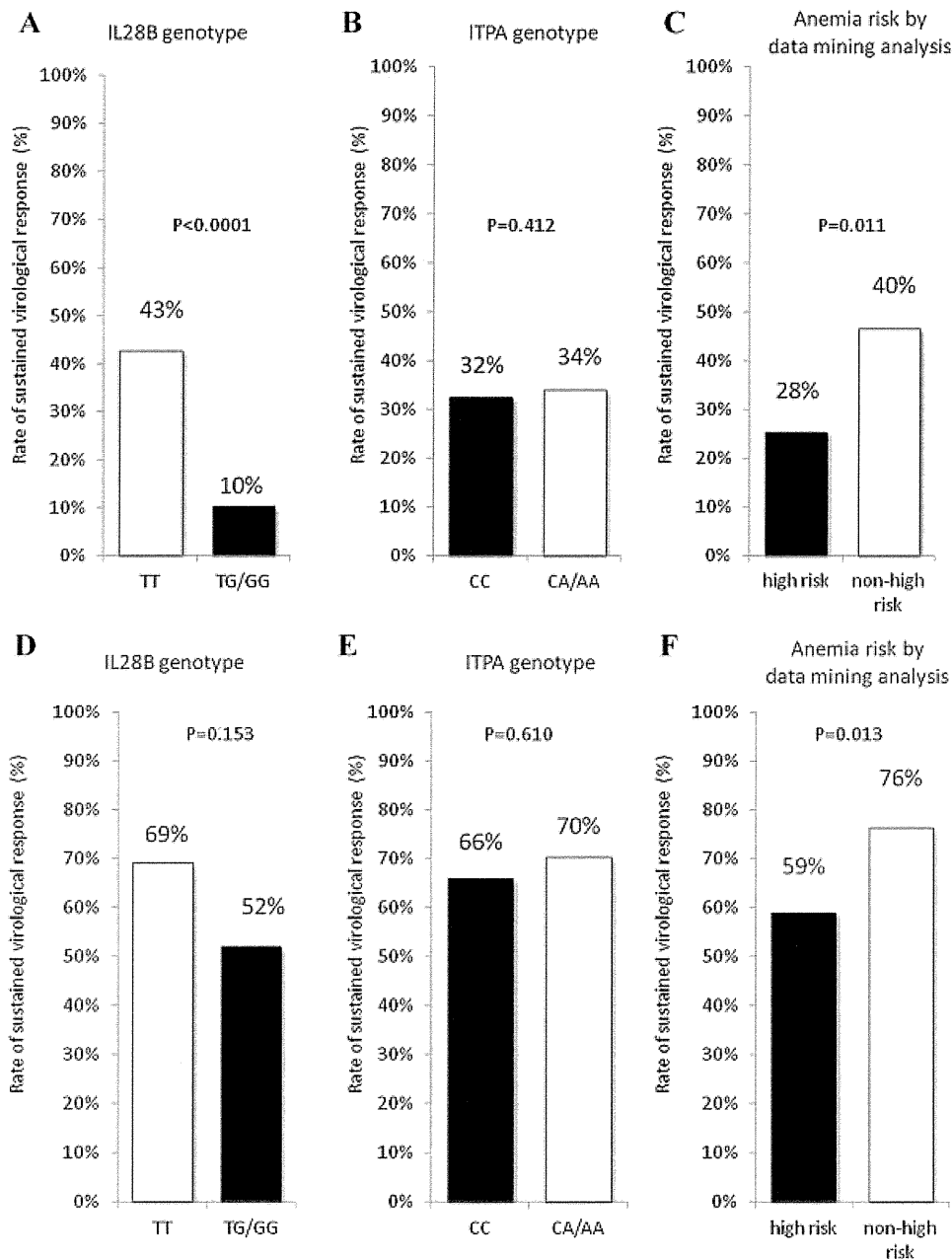
Fig. 3. Rates of sustained virological responses at baseline and among those with a virological response at week 12. The impacts of *IL28B* genotype, *ITPA* genotype, and risk group of anemia on the rate of sustained virological response were studied at baseline (**A–C**) and among those with complete early virological responses (defined as undetectable HCV RNA at week 12) (**D–F**). At baseline, those with the TT allele of the *IL28B* gene had a significantly higher rate of sustained virological response than those with the TG or GG allele and the group at high-risk of anemia had a significantly lower rate of sustained virological response than the low-risk group. Among patients with complete early virological responses, the *IL28B* genotype was not associated with a sustained virological response, while the group at high-risk of anemia had a significantly lower rate of sustained virological response than the low-risk group.

et al., 2010], and CLcr [Kurosaki et al., 2011d], also contribute to the risk of severe anemia or RBV dose reduction. In the present study, the predictive model of anemia based on the data mining analysis selected the *ITPA* genotype, baseline Hb concentration, and

baseline CLcr as predictive factors and identified six subgroups of patients with a variable rate of severe anemia, ranging from 17% to 76%. The specificity of the prediction of severe anemia was improved by 25.7% in the predictive model, compared to *ITPA*

TABLE II. Logistic Regression Analysis for Factors Associated With Sustained Virological Response at Baseline, Week 4 and Week 12

| | Multi-variable | | |
| --- | --- | --- | --- |
| | Odds | 95% CI | P-value |
| Pre-treatment | | | |
| Sex: female | 0.42 | 0.26–0.68 | <0.0001 |
| Platelet ($10^9$/L) | 1.09 | 1.04–1.15 | <0.0001 |
| Fibrosis: F3-4 | 0.49 | 0.27–0.91 | 0.024 |
| HCV RNA: <600,000 IU/L | 4.14 | 2.27–7.55 | <0.0001 |
| IL28B rs8099917: TT | 9.88 | 5.01–19.48 | <0.0001 |
| At week 4 | | | |
| Non-RVR patients | | | |
| Sex: female | 0.45 | 0.28–0.72 | 0.001 |
| Platelet ($10^9$/L) | 1.10 | 1.05–1.16 | 0.000 |
| IL28B rs8099917: TT | 7.16 | 3.60–14.25 | <0.0001 |
| At week 12 | | | |
| cEVR patients | | | |
| Platelet ($10^9$/L) | 1.09 | 1.02–1.17 | 0.015 |
| HCV RNA: <600,000 IU/L | 3.21 | 1.39–7.55 | 0.007 |
| High-risk of anemia[a] | 0.47 | 0.24–0.91 | 0.026 |
| At week 12 | | | |
| Non-cEVR patients | | | |
| Platelet ($10^9$/L) | 1.11 | 1.02–1.21 | 0.017 |
| IL28B rs8099917: TT | 9.13 | 2.02–41.3 | 0.004 |

RVR: rapid virological response, defined as undetectable HCV RNA at week 4.
cEVR: complete early virological response, defined as undetectable HCV RNA at week 12.
[a]High-risk of anemia defined by decision tree analysis includes the following groups: (1) baseline hemoglobin <14.0 g/dl and creatinine clearance <90 ml/min, (2) baseline hemoglobin <14.0 g/dl, creatinine clearance ≥90 ml/min and ITPA rs1127354 genotype CC, and (3) baseline hemoglobin ≥14.0 g/dl, ITPA rs1127354 genotype CC, and creatinine clearance <85 ml/min.

genotyping alone. Because hemolytic anemia induced by RBV is one of the major adverse events leading to premature termination of therapy [Fried et al., 2002], a method to predict the risk of severe anemia before treatment is important clinically. A predictive model of anemia may have the potential to support individualized treatment strategies; patients at high risk of anemia may be tested intensively for anemia or may be candidates for erythropoietin therapy, whereas those with a low risk of anemia may be treated with a higher dose of RBV. Prediction of anemia will remain important in the era of direct antiviral agents for chronic hepatitis C, because these newer therapies still require RBV and PEG-IFN in combination, and the degree of anemia complicating these therapies may be even greater than with the current combination therapy [McHutchison et al., 2009; Kwo et al., 2010].

Studies of the impact of the ITPA genotype on treatment outcome have produced conflicting results. Previous studies of American [Thompson et al., 2010a] and Italian [Thompson et al., 2011] cohorts did not find any association between the ITPA genotype and treatment outcome, whereas a marginal difference was observed in a report from Japan [Ochi et al., 2010]. Moreover, with a subgroup analysis of Japanese patients, the variant of the ITPA gene was

associated with a sustained virological response in patients with the IL28B major genotype [Kurosaki et al., 2011d], in patients infected with HCV other than genotype 1[Sakamoto et al., 2010], and in patients with pre-treatment Hb concentrations between 13.5 and 15 g/dl [Azakami et al., 2011]. These inconsistent results may be because the impact of anemia may be greater on a cohort of aged patients, such as in Japan. Another reason may be that the ITPA genotype is not the sole determinant of anemia; the ITPA genotype alone was not associated with treatment outcome in the present study but a high-risk of anemia, defined by the combination of the ITPA genotype, baseline Hb concentration, and baseline CLcr, was associated with sustained virological responses by patients with complete early virological responses, even after adjustment for the IL28B genotype and other relevant factors. This is in contrast to the finding that the IL28B genotype is an independent and significant predictor at baseline of a sustained virological response by patients without a rapid virological response and those without a complete early virological response, but not those with a complete early virological response. These results indicate that the IL28B genotype could be used to predict a sustained virological response at baseline or during therapy in patients in whom HCV RNA has not yet become undetectable, but it has no predictive value in patients in whom HCV RNA has become undetectable. The risk of anemia may be used to predict sustained virological responses in a selected subgroup of patients who achieve a complete early virological response.

Patients who received more than 80% of the planned dose of PEG-IFN or RBV had a higher rate of sustained virological responses than those who received a lower cumulative dose [McHutchison et al., 2002; Davis et al., 2003]. Patients who achieve a complete early virological response usually have a good chance of a sustained virological response and the treatment duration is not extended beyond 48 weeks. However, reduced adherence to drugs in these patients was related to relapse after the completion of 48 weeks of therapy [Hiramatsu et al., 2009; Kurosaki et al., 2012]. In the present study, the rate of sustained virological response was 59% in patients who achieved a complete early virological response but had a high risk of anemia, 17% lower than in patients with a low risk of anemia. However, there was a stepwise increase in the rate of sustained virological response according to the increase in adherence to RBV, and the rate of sustained virological response was higher in high-risk patients who received >80% of the planned dose of RBV (71% vs. 47%). This 24% increase in sustained virological response was observed among the patients in the present study who received 48 weeks of treatment. These findings suggest that receiving a sufficient RBV dose is essential for patients with a complete early virological response to attain a sustained virological response and that the treatment strategy should be personalized for patients with a
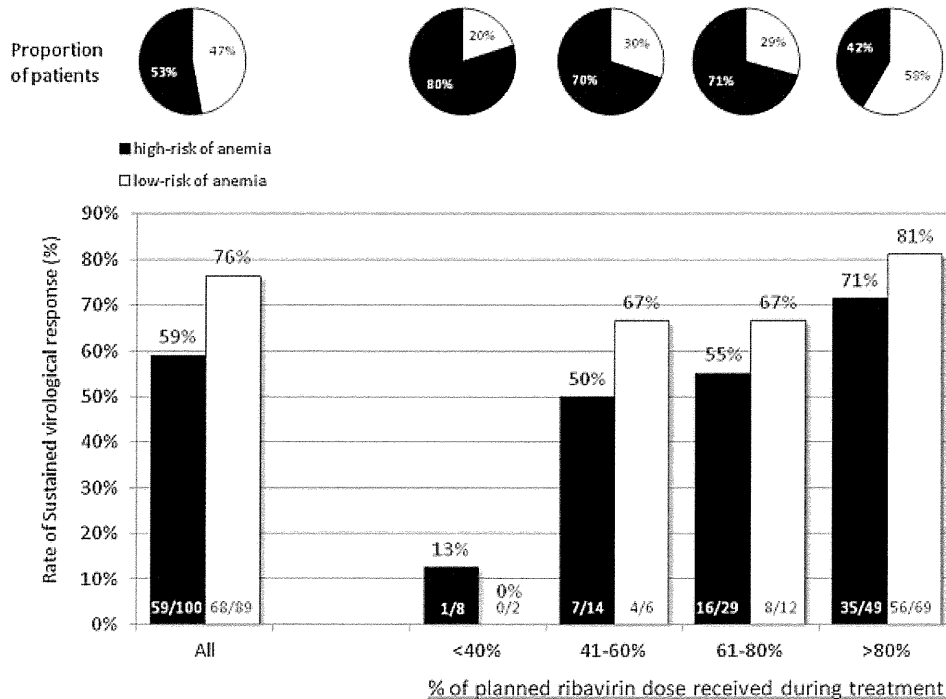
Fig. 4. The impact of risk of anemia and RBV dose on treatment outcome after a complete early virological response. Patients with complete early virological responses were divided into subgroups according to their adherence to RBV: ≤40%, 41–60%, 61–80%, and >80%. For each subgroup, the proportion of patients with a high risk and a low risk of anemia is shown in the upper panel by pie charts, and the rates of sustained virological responses, stratified by high risk and low risk of anemia, are shown in the lower panel by bar graphs. The black and white bars or charts represent patients with high and low risks of anemia, respectively.

high risk of anemia to extend the duration of treatment, even those patients with a complete early virological response, to obtain >80% adherence to RBV.

In conclusion, the combination of the *ITPA* genotype, baseline Hb concentration, and baseline CLcr could be used as a pre-treatment predictor of anemia. The risk of anemia thus identified is associated with adherence to RBV and impacts on the treatment outcome of patients who achieve a complete early virological response. This is in contrast to the major role of the *IL28B* genotype in the prediction of sustained virological responses at baseline and among non-responders at weeks 4 and 12. Patients who achieve a complete early virological response generally have a high probability of a sustained virological response but those who have a high risk of anemia have a high rate of relapse because of reduced adherence to RBV. To improve the rate of sustained virological responses in these patients, it may be postulated that the treatment schedule may be personalized to obtain >80% adherence to RBV. Clearly, this postulate needs to be confirmed in a future study.

## REFERENCES

Averbook BJ, Fu P, Rao JS, Mansour EG. 2002. A long-term analysis of 1018 patients with melanoma by classic Cox regression and tree-structured survival analysis at a major referral center: Implications on the future of cancer staging. Surgery 132:589–602.

Azakami T, Hayes CN, Sezaki H, Kobayashi M, Akuta N, Suzuki F, Kumada H, Abe H, Miki D, Tsuge M, Imamura M, Kawakami Y, Takahashi S, Ochi H, Nakamura Y, Kamatani N, Chayama K. 2011. Common genetic polymorphism of ITPA gene affects ribavirin-induced anemia and effect of peg-interferon plus ribavirin therapy. J Med Virol 83:1048–1057.

Baquerizo A, Anselmo D, Shackleton C, Chen TW, Cao C, Weaver M, Gornbein J, Geevarghese S, Nissen N, Farmer D, Demetriou A, Busuttil RW. 2003. Phosphorus ans an early predictive factor in patients with acute liver failure. Transplantation 75:2007–2014.

Breiman LJH, Friedman RA, Olshen CJ, Stone CM. 1980. Classification and regression trees. Calif: Wadsworth.

Cockcroft DW, Gault MH. 1976. Prediction of creatinine clearance from serum creatinine. Nephron 16:31–41.

Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. 2003. Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. Hepatology 38:645–652.

Fellay J, Thompson AJ, Ge D, Gumbs CE, Urban TJ, Shianna KV, Little LD, Qiu P, Bertelsen AH, Watson M, Warner A, Muir AJ, Brass C, Albrecht J, Sulkowski M, McHutchison JG, Goldstein DB. 2010. ITPA gene variants protect against anaemia in patients treated for chronic hepatitis C. Nature 464: 405–408.

Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL, Jr., Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J. 2002. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med 347:975–982.

Garzotto M, Park Y, Mongoue-Tchokote S, Bledsoe J, Peters L, Blank BH, Austin D, Beer TM, Mori M. 2005. Recursive partitioning for

risk stratification in men undergoing repeat prostate biopsies. Cancer 104:1911–1917.

Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, Heinzen EL, Qiu P, Bertelsen AH, Muir AJ, Sulkowski M, McHutchison JG, Goldstein DB. 2009. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 461:399–401.

Hiramatsu N, Kurashige N, Oze T, Takehara T, Tamura S, Kasahara A, Oshita M, Katayama K, Yoshihara H, Imai Y, Kato M, Kawata S, Tsubouchi H, Kumada H, Okanoue T, Kakumu S, Hayashi N. 2008. Early decline of hemoglobin can predict progression of hemolytic anemia during pegylated interferon and ribavirin combination therapy in patients with chronic hepatitis C. Hepatol Res 38:52–59.

Hiramatsu N, Oze T, Yakushijin T, Inoue Y, Igura T, Mochizuki K, Imanaka K, Kaneko A, Oshita M, Hagiwara H, Mita E, Nagase T, Ito T, Inui Y, Hijioka T, Katayama K, Tamura S, Yoshihara H, Imai Y, Kato M, Yoshida Y, Tatsumi T, Ohkawa K, Kiso S, Kanto T, Kasahara A, Takehara T, Hayashi N. 2009. Ribavirin dose reduction raises relapse rate dose-dependently in genotype 1 patients with hepatitis C responding to pegylated interferon alpha-2b plus ribavirin. J Viral Hepat 16:586–594.

Hiramatsu N, Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, Sugauchi F, Tamori A, Kakinuma S, Matsuura K, Izumi N. 2011. Pretreatment prediction of anemia progression by pegylated interferon alpha-2b plus ribavirin combination therapy in chronic hepatitis C infection: Decision-tree analysis. J Gastroenterol 46:1111–1119.

Kim RW. 2002. Global epidemiology and burden of hepatitis C. Microbes Infect 4:1219–1225.

Kurosaki M, Matsunaga K, Hirayama I, Tanaka T, Sato M, Yasui Y, Tamaki N, Hosokawa T, Ueda K, Tsuchiya K, Nakanishi H, Ikeda H, Itakura J, Takahashi Y, Asahina Y, Higaki M, Enomoto N, Izumi N. 2010. A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. Hepatol Res 40:251–260.

Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, Hiramatsu N, Sugauchi F, Tamori A, Nakagawa M, Izumi N. 2011a. Sequences in the interferon sensitivity-determining region and core region of hepatitis C virus impact pretreatment prediction of response to PEG-interferon plus ribavirin: Data mining analysis. J Med Virol 83:445–452.

Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, Hiramatsu N, Sugauchi F, Yatsuhashi H, Izumi N. 2011b. Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis. J Gastroenterol 46:401–409.

Kurosaki M, Tanaka Y, Nishida N, Sakamoto N, Enomoto N, Honda M, Sugiyama M, Matsuura K, Sugauchi F, Asahina Y, Nakagawa M, Watanabe M, Sakamoto M, Maekawa S, Sakai A, Kaneko S, Ito K, Masaki N, Tokunaga K, Izumi N, Mizokami M. 2011c. Pretreatment prediction of response to pegylated-interferon plus ribavirin for chronic hepatitis C using genetic polymorphism in IL28B and viral factors. J Hepatol 54:439–448.

Kurosaki M, Tanaka Y, Tanaka K, Suzuki Y, Hoshioka Y, Tamaki N, Kato T, Yasui Y, Hosokawa T, Ueda K, Tsuchiya K, Kuzuya T, Nakanishi H, Itakura J, Takahashi Y, Asahina Y, Matsuura K, Sugauchi F, Enomoto N, Nishida N, Tokunaga K, Mizokami M, Izumi N. 2011d. Relationship between genetic polymorphisms of the inosine triphosphatase gene and anemia or outcome after treatment with pegylated-interferon and ribavirin. Antivir Ther 16:685–694.

Kurosaki M, Hiramatsu N, Sakamoto M, Suzuki Y, Iwasaki M, Tamori A, Matsuura K, Kakinuma S, Sugauchi F, Sakamoto N, Nakagawa M, Yatsuhashi H, Izumi N. 2012. Age and total ribavirin dose is an independent predictor of relapse among early virological responders to peg-interferon plus ribavirin therapy in chronic hepatitis C revealed by data mining analysis. Antivir Ther 17:35–43.

Kwo PY, Lawitz EJ, McCone J, Schiff ER, Vierling JM, Pound D, Davis MN, Galati JS, Gordon SC, Ravendhran N, Rossaro L, Anderson FH, Jacobson IM, Rubin R, Koury K, Pedicone LD, Brass CA, Chaudhri E, Albrecht JK. 2010. Efficacy of boceprevir, an NS3 protease inhibitor, in combination with peginterferon alfa-2b and ribavirin in treatment-naive patients with genotype 1 hepatitis C infection (SPRINT-1): An open-label, randomised, multicentre phase 2 trial. Lancet 376:705–716.

LeBlanc M, Crowley J. 1995. A review of tree-based prognostic models. Cancer Treat Res 75:113–124.

Leiter U, Buettner PG, Eigentler TK, Garbe C. 2004. Prognostic factors of thin cutaneous melanoma: An analysis of the Central Malignant Melanoma Registry of the German Dermatological Society. J Clin Oncol 22:3660–3667.

Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, Goodman ZD, Koury K, Ling M, Albrecht JK. 2001. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: A randomised trial. Lancet 358:958–965.

McHutchison JG, Manns M, Patel K, Poynard T, Lindsay KL, Trepo C, Dienstag J, Lee WM, Mak C, Garaud JJ, Albrecht JK. 2002. Adherence to combination therapy enhances sustained response in genotype-1-infected patients with chronic hepatitis C. Gastroenterology 123:1061–1069.

McHutchison JG, Everson GT, Gordon SC, Jacobson IM, Sulkowski M, Kauffman R, McNair L, Alam J, Muir AJ. 2009. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. N Engl J Med 360:1827–1838.

Miyaki K, Takei I, Watanabe K, Nakashima H, Omae K. 2002. Novel statistical classification model of type 2 diabetes mellitus patients for tailor-made prevention using data mining algorithm. J Epidemiol 12:243–248.

Ochi H, Maekawa T, Abe H, Hayashida Y, Nakano R, Kubo M, Tsunoda T, Hayes CN, Kumada H, Nakamura Y, Chayama K. 2010. Inosine triphosphate pyrophosphatase polymorphism affects ribavirin-induced anemia and outcomes of therapy—A genome-wide study of Japanese hepatitis C virus patients. Gastroenterology 139:1190–1197.

Rauch A, Kutalik Z, Descombes P, Cai T, Di Iulio J, Mueller T, Bochud M, Battegay M, Bernasconi E, Borovicka J, Colombo S, Cerny A, Dufour JF, Furrer H, Gunthard HF, Heim M, Hirschel B, Malinverni R, Moradpour D, Mullhaupt B, Witteck A, Beckmann JS, Berg T, Bergmann S, Negro F, Telenti A, Bochud PY. 2010. Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: A genome-wide association study. Gastroenterology 138:1338–1345.

Sakamoto N, Tanaka Y, Nakagawa M, Yatsuhashi H, Nishiguchi S, Enomoto N, Azuma S, Nishimura-Sakurai Y, Kakinuma S, Nishida N, Tokunaga K, Honda M, Ito K, Mizokami M, Watanabe M. 2010. ITPA gene variant protects against anemia induced by pegylated interferon-alpha and ribavirin therapy for Japanese patients with chronic hepatitis C. Hepatol Res 40:1063–1071.

Seto WK, Tanaka Y, Liu K, Lai CL, Yuen MF. 2011. The effects of IL-28B and ITPA polymorphisms on treatment of hepatitis C virus genotype 6. Am J Gastroenterol 106:1007–1008.

Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate ML, Bassendine M, Spengler U, Dore GJ, Powell E, Riordan S, Sheridan D, Smedile A, Fragomeli V, Muller T, Bahlo M, Stewart GJ, Booth DR, George J. 2009. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. Nat Genet 41:1100–1104.

Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, Nakagawa M, Korenaga M, Hino K, Hige S, Ito Y, Mita E, Tanaka E, Mochida S, Murawaki Y, Honda M, Sakai A, Hiasa Y, Nishiguchi S, Koike A, Sakaida I, Imamura M, Ito K, Yano K, Masaki N, Sugauchi F, Izumi N, Tokunaga K, Mizokami M. 2009. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 41:1105–1109.

Tanaka Y, Kurosaki M, Nishida N, Sugiyama M, Matsuura K, Sakamoto N, Enomoto N, Yatsuhashi H, Nishiguchi S, Hino K, Hige S, Itoh Y, Tanaka E, Mochida S, Honda M, Hiasa Y, Koike A, Sugauchi F, Kaneko S, Izumi N, Tokunaga K, Mizokami M. 2011. Genome-wide association study identified ITPA/DDRGK1 variants reflecting thrombocytopenia in pegylated interferon and ribavirin therapy for chronic hepatitis C. Hum Mol Genet 20:3507–3516.

Thompson AJ, Fellay J, Patel K, Tillmann HL, Naggie S, Ge D, Urban TJ, Shianna KV, Muir AJ, Fried MW, Afdhal NH, Goldstein DB, McHutchison JG. 2010a. Variants in the ITPA gene protect against ribavirin-induced hemolytic anemia and decrease the need for ribavirin dose reduction. Gastroenterology 139:1181–1189.

Thompson AJ, Muir AJ, Sulkowski MS, Ge D, Fellay J, Shianna KV, Urban T, Afdhal NH, Jacobson IM, Esteban R, Poordad F,

Lawitz EJ, McCone J, Shiffman ML, Galler GW, Lee WM, Reindollar R, King JW, Kwo PY, Ghalib RH, Freilich B, Nyberg LM, Zeuzem S, Poynard T, Vock DM, Pieper KS, Patel K, Tillmann HL, Noviello S, Koury K, Pedicone LD, Brass CA, Albrecht JK, Goldstein DB, McHutchison JG. 2010b. Interleukin-28B polymorphism improves viral kinetics and is the strongest pretreatment predictor of sustained virologic response in genotype 1 hepatitis C virus. Gastroenterology 139:120–129 e118.

Thompson AJ, Santoro R, Piazzolla V, Clark PJ, Naggie S, Tillmann HL, Patel K, Muir AJ, Shianna KV, Mottola L, Petruzzellis D, Romano M, Sogari F, Facciorusso D, Goldstein DB, McHutchison JG, Mangia A. 2011. Inosine triphosphatase genetic variants are protective against anemia during antiviral therapy for HCV2/3 but do not decrease dose reductions of RBV or increase SVR. Hepatology 53:389–395.

Valera VA, Walter BA, Yokoyama N, Koyama Y, Iiai T, Okamoto H, Hatakeyama K. 2007. Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. Ann Surg Oncol 14:34–40.

Zlobec I, Steele R, Nigam N, Compton CC. 2005. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. Clin Cancer Res 11:5440–5443.

PLoS one

# 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 $P_{meta} = 1.89 \times 10^{-12}$ for rs3077 and $P_{meta} = 9.69 \times 10^{-10}$ 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 ($P_{meta} = 4.40 \times 10^{-19}$ for rs3077 and $P_{meta} = 1.28 \times 10^{-15}$ 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.

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 hepato-cellular 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 $\geq$95%; (ii) minor allele frequency (MAF) $\geq$1% for HBV carriers and healthy controls; and (iii) no deviation from Hardy-Weinberg equilibrium (HWE) $P\geq0.001$ 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 $\geq$ 95% and HWE $p$-value $\geq$ 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).
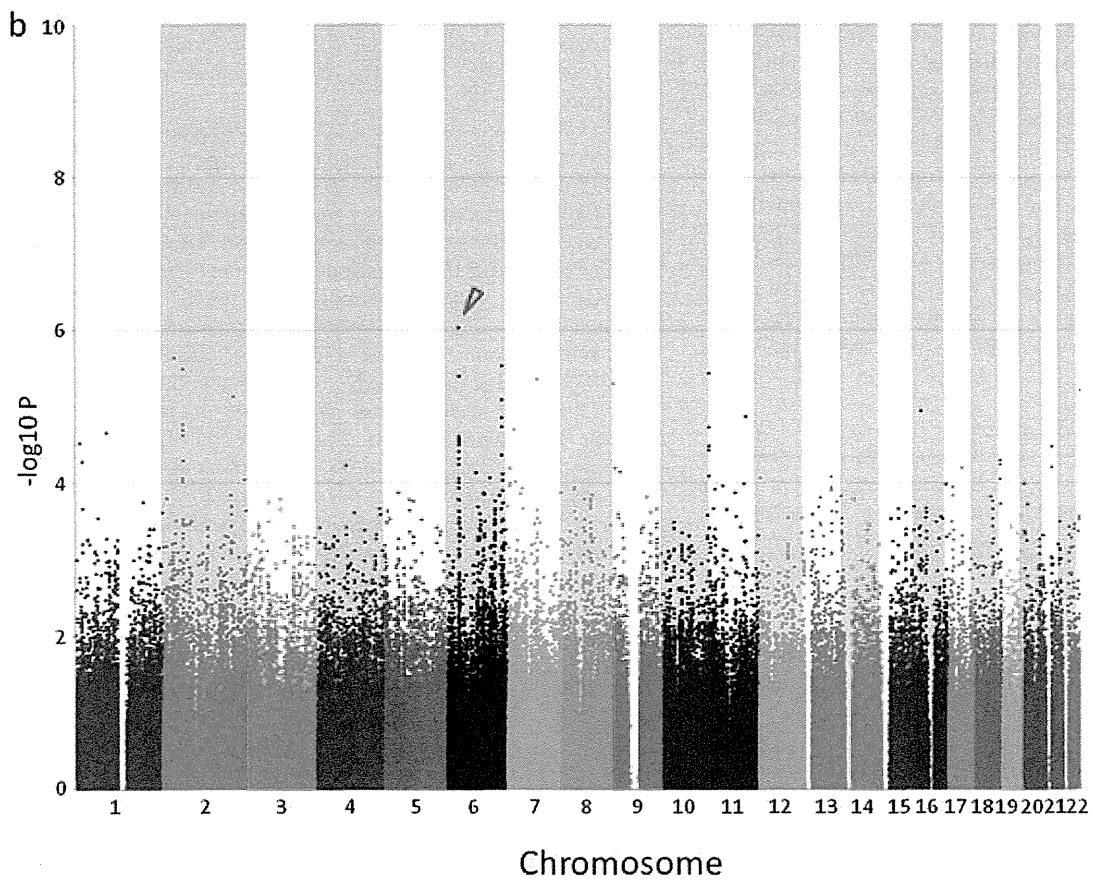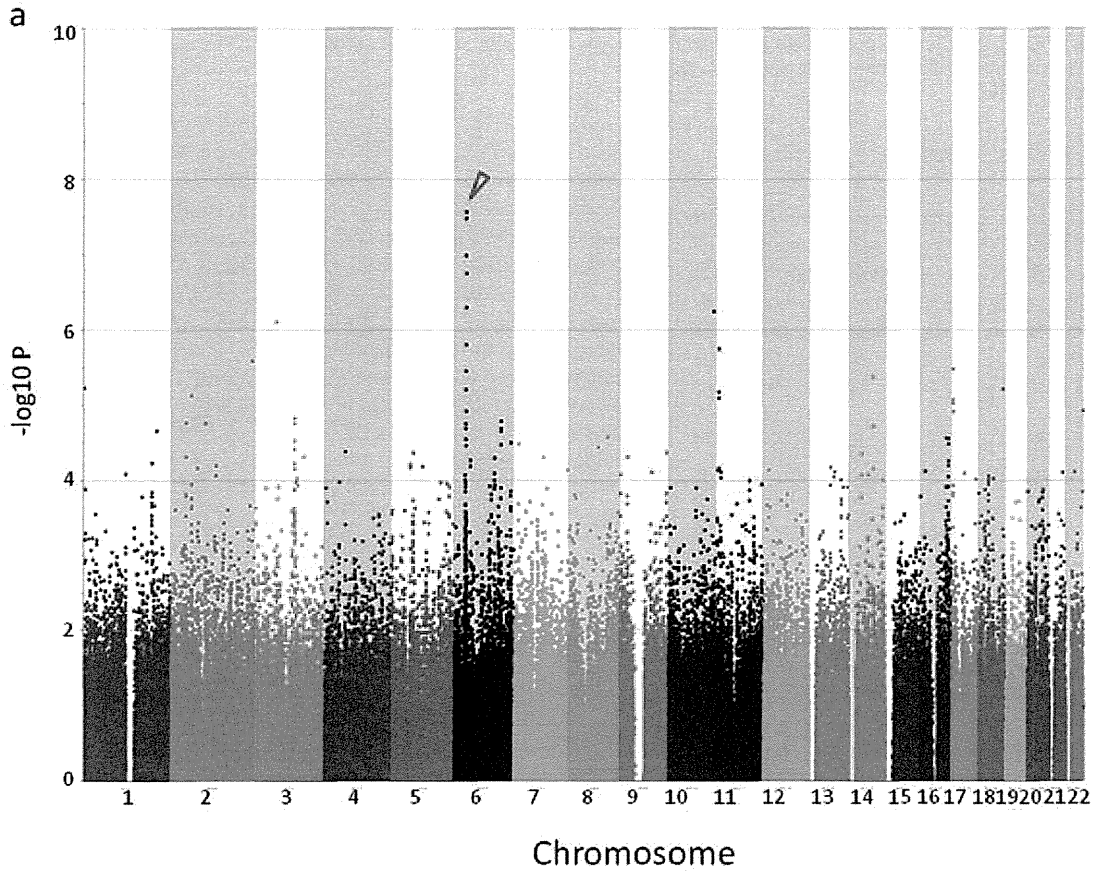
- 628 -

**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 ($P = 9.24 \times 10^{-7}$ and $P = 3.15 \times 10^{-5}$) 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, $P = 3.32 \times 10^{-2}$, OR $= 0.72$ and rs9277542, $P = 1.25 \times 10^{-2}$, OR $= 0.68$; replication-2 (Korean): rs3077, $P = 2.35 \times 10^{-7}$, OR $= 0.41$ and rs9277542, $P = 4.97 \times 10^{-6}$, OR $= 0.46$; Table 3]. Meta-analysis using random effects model showed $P_{meta} = 1.56 \times 10^{-4}$ for rs3077 (OR $= 0.51$, 95% CI $= 0.36$–$0.72$), and $5.91 \times 10^{-7}$ 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 $P_{meta} = 3.28 \times 10^{-12}$ for rs3077 and $1.42 \times 10^{-10}$ 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 $2.10 \times 10^{-5}$ (OR $= 0.59$) and $1.10 \times 10^{-5}$ (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.

Meta-analysis across 6 independent studies, including 5 additional published data, showed $P_{meta} = 1.48 \times 10^{-9}$, OR $= 0.60$ for rs3077, $P_{meta} = 1.08 \times 10^{-17}$, OR $= 0.66$ for rs9277535 and $P_{meta} = 5.14 \times 10^{-5}$, 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 $P_{het} = 9.57 \times 10^{-6}$ 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 $P_{meta} = 5.88 \times 10^{-24}$, 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 ($P = 3.32 \times 10^{-2}$, OR $= 0.72$). Moreover, meta-analysis using GWAS and replication-2 showed significant association of $P_{meta} = 1.89 \times 10^{-12}$, 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 1.** Number of study samples.

| population | | GWAS<br>Japanese | Replication-1<br>Japanese | Replication-2<br>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

- 630 -

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

| dbSNP rsID | Chr | Buld 36.3 | Nearest Gene | MAF[a] (allele) | Allele (1/2) | Stage (population) | HBV carriers 11 | 12 | 22 | Healthy controls 11 | 12 | 22 | HWEp | OR[b] 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-DPA1/−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 agree-ment (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

- 632 -

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)

## 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.

- 634 -

BMC
Medical Genetics

# 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

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.