A thymine-adenine dinucleotide repeat polymorphism near IL28B is associated with spontaneous clearance of hepatitis C virus

平峯, 智

https://doi.org/10.15017/1654717

出版情報：九州大学, 2015, 博士（医学）, 課程博士
バージョン：

Title:

A Thymine-Adenine Dinucleotide Repeat Polymorphism Near IL28B Is Associated with Spontaneous Clearance of Hepatitis C Virus

Authors:

Satoshi Hiramine(1,2), Masaya Sugiyama(1), Norihiro Furusyo(2), Akio Ido(3), Hirohito Tsubouchi(3), Hisayoshi Watanabe(4), Yoshiyuki Ueno(4), Masaaki Korenaga(1), Kazumoto Murata(1), Naohiko Masaki(1), Jun Hayashi(2), David L Thomas(5), and Masashi Mizokami(1)

Author affiliations:

1. Department of Hepatic Diseases, The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, 252-8516, Japan; 2. General Internal Medicine Department, Kyushu University Hospital, Fukuoka, 812-8582, Japan; 3. Digestive and Lifestyle Diseases, Department of Human and Environmental Sciences, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, 890-8544, Japan; 4. Department of Gastroenterology, Faculty of Medicine, Yamagata University, Yamagata, 990-9585, Japan; 5. Division of Infectious Diseases, Johns Hopkins School of Medicine, Baltimore, Maryland, 21205, USA.

Correspondence-

Masashi Mizokami M.D., Ph.D.

1-7-1 Kohonodi, Ichikawa 272-8516 Chiba, Japan

Tel: +81-47-372-3501 Fax: +81-47-375-4746

E-mail: mmizokami@hospk.ncgm.go.jp

Short title:

TA repeat and Spontaneous HCV Clearance

ABSTRACT:

BACKGROUND. Genome-wide-association studies have revealed several single nucleotide polymorphisms (SNPs) around interleukin(IL)28B that are strongly associated with Hepatitis C virus (HCV) clearance. However, their predictive value is not perfect, which suggests that other genetic factors may also be involved in HCV clearance. We previously reported a wide variation in the length of a thymine-adenine (TA) dinucleotide repeat in the promoter region of IL28B and that the transcriptional activity of the promoter increased gradually in a TA repeat length-dependent manner.

METHODS. We determined the length of the TA repeats of 1,060 Japanese and 201 African-American samples to investigate the relation to spontaneous HCV clearance.

RESULTS. The distribution of the TA repeats greatly differed between the two ethnicities. The variation ranged from 10 to 18 repeats, and the most frequent allele, 12, accounted for over 80% for Japanese. The African-American data showed a gently sloping distribution, and the allele with 6 repeats was detected only in the African-American sample. The TA repeats 11 or greater were correlated with spontaneous clearance. Multiple logistic regression analysis extracted the genotype of the TA repeats as an independent factor in both the Japanese (P=0.0004, odds ratio [OR]=13.02 95% confidence interval [CI]=2.59-237.0) and African-American (P=0.027, OR=3.70 95% CI=1.16-11.8) populations.

CONCLUSIONS. A long TA repeat in the promoter region of IL28B was associated with spontaneous HCV clearance. Although its efficacy may be limited in Japanese population because of its allele distribution, this novel genetic factor will be useful for predicting HCV clearance especially for the African-Americans.

Keywords: Genetic marker, microsatellite, Interferon lambda 3.

Abbreviations:

interferon, IFN; interleukin-28B, IL28B; hepatitis C virus, HCV; negative predictive value, NPV; pegylated, PEG; positive predictive value, PPV; ribavirin, RBV; single nucleotide polymorphisms, SNP; uridine diphosphoglucuronosyl transferase, UGT

Introduction

The World Health Organization estimated in 1999 that 170 million hepatitis C virus (HCV) carriers were present worldwide, with 3–4 million new cases appearing each year [1]. Although approximately 70% of the carriers develop chronic hepatitis, with a strong risk for cirrhosis or hepatocellular carcinoma, the remainder spontaneously clear infection and rarely have hepatic failure [2].

Although Interferon (IFN)-based treatment has improved, such as in combination with ribavirin (RBV) and pegylated (PEG) IFN, about half of the patients with HCV genotype 1 do not achieve HCV clearance in the U.S., Europe [3, 4], and Japan [5]. To avoid serious side-effects and unproductive expenditures, viral factors, such as genotype, viral load, and amino acid substitutions, are used to predict which patients are unlikely to respond, but with limited success.

Recent genome-wide association studies have revealed several single nucleotide polymorphisms (SNPs) around the *interleukin 28B* (*IL28B*) gene in chromosome 19 that are strongly associated with the response of chronic hepatitis C patients to IFN therapy [6-8]. *IL28B* is also known as IFN lambda 3, a class Ⅱ cytokine that induces antiviral activity and suppresses HCV replication [9, 10]. A high level of *IL28B* mRNA expression has been observed in persons with the advantageous *IL28B* SNPs genotype [7, 8]. Likewise, these SNPs were correlated to spontaneous HCV clearance [11, 12].

The *IL28B* SNPs genotype can clearly explain the heterogeneity of the clinical outcome of patients with HCV infection, however, approximately 20% of the patients with the advantageous genotype do not clear the infection and approximately 20% of patients with the disadvantageous genotype respond to the therapy [6, 8], which suggests that other factors may also be involved in HCV clearance.

Our recent study revealed a genetic polymorphism in the promoter region of *IL28*. This insertion/deletion polymorphism consists of a thymine-adenine (TA)

dinucleotide repeat, rs59702201 (rs72258881 has been integrated into rs59702201). We previously reported a wide variation in the length of the TA repeat, from 10 to 18, and that the transcriptional activity of the promoter increased gradually in a TA repeat length-dependent manner [13].

We therefore hypothesized that persons with longer TA repeats would have more success in clearing HCV. In this study, we determined the length of the TA repeat in the genomic samples of 1,060 Japanese and investigated the relation between the number of TA repeats and spontaneous HCV clearance. We then tested genomic samples of 201 African-Americans in order to validate the findings of the Japanese samples.

Materials and Methods

Genome samples

We acquired 1,060 samples for the genome testing of three independent Japanese HCV cohorts [14-16]. These samples were collected at free public health examinations. The examinations contain a questionnaire and a blood test analysis that included anti-HCV and HCV RNA. Subjects with a history of antiviral therapy using IFN or who were seropositive for HBsAg were excluded. Participants positive for anti-HCV were followed and tested again from three months to a year later. If the HCV RNA at the two time points were both negative, the patients (224) were assigned to a spontaneous clearance group. One third of the subjects were selected at random from all who were positive for HCV RNA and assigned to a chronic infection group (326). Half of the subjects of one of our cohorts [14] were selected at random from persons negative for anti-HCV and assigned to a healthy control group (510). To validate the findings of the Japanese samples, 201 samples were obtained from African-Americans enrolled in the ALIVE study, an ongoing study of injection drug users done in Baltimore, Maryland since 1988, as described elsewhere [17]. One hundred and one samples of patients with spontaneous clearance and 100 with chronic infection were selected randomly. Age and

sex were matched. Genomic DNA was extracted from whole blood samples by standard methods.

Informed consent was obtained from each participant included in the study before the examination. The study protocol conforms to the ethical guidelines of the Declaration of Helsinki and each cohort study was approved by each institution's human research committee, including permission for human genome analysis.

***IL28B* SNPs polymorphism analysis.**

The rs8099917 and rs12979860 polymorphisms were determined using the Invader-Plus assay [18], which combines PCR and the Invader reaction [19, 20], on a LightCycler 480 (Roche Diagnostics, Basel, Switzerland). The Invader-Plus assay reagent kit, purchased from Third Wave Technology (Madison, USA), consists of a probe mix, a buffer mix, and an enzyme mix. The reagents were premixed according to the manufacturer’s instructions. Then, 10ng of genomic DNA was added to the master mix. The primer and probe sets are described in Table S1.The cycle conditions were 18 cycles of 15 s at 95°C and 60 s at 70°C. At the end of the PCR, the Taq polymerase was inactivated at 99°C for 10 min and the reaction temperature was lowered to 63°C for 15 to 30 min to permit the hybridization of the probe oligonucleotide and the formation of the overlap flap structure. Data were analyzed by endpoint genotyping software (Roche Diagnostics). Both rs8099917 and rs12979860 were determined from the African-American samples; however, we tested the Japanese samples for only rs8099917 because it was previously reported that rs8099917 and rs12979860 represented 98.6% of the Japanese population [18].

TA repeat genotyping.

To determine the genotype of the TA repeat polymorphism, we developed a new method based on GeneScan analysis (Applied Biosystems, Foster City, CA) that detects the fragment size of a fluorescent-labeled PCR amplicon. This method requires the use of nested PCR to prevent the amplification of the *IL29* region, which has a high

level of structural similarity to the *IL28A/B* region. The 1st PCR reaction was performed in a volume of 50 μl that contained 10 ng of genomic DNA, 10 pmol of each primer (5'-TAGCTGGGAATGGTGGCACA-3' and 5'-CAAACTCCTGGGCTCAAGCCATCCTCCTCACCCAG-3'), 5×PrimeSTAR GXL Buffer, 2.5 mM each deoxynucleotide triphosphates, and 1.25 units of PrimeStar GXL DNA polymerase (TAKARA Bio Inc, Tokyo, Japan). The cycle conditions were 35 cycles of 10 s at 98°C, 15 s at 65°C, and 60 s at 68°C, in addition to initial denaturation at 98°C for 5 min and a final extension at 68°C for 7 min. The 2nd PCR reaction was performed in a volume of 50 μl containing 1μl of the 1st PCR product and 10 pmol of each primer. The primers were 5'- TGAACCCAGGAGGCGGAGGTTGCAGTTAGC-3' and 5'- GTGCTGAGATTACAGGCCTGAGCCACCAC-3'. The former was labeled with FAM. The buffer, enzyme, and cycle conditions were the same as for the 1st PCR. One μL of the 2nd PCR product diluted 200-fold was mixed with 10 μl of formamide and 0.5 μl of 600 LIZ size standard (Applied Biosystems). The products were denatured at 95 °C for 2 min, immediately placed on ice for 10 min, and then subjected to GeneScan analysis. GeneScan analysis was done using the ABI 3130xl Genetic Analyzer (Applied Biosystems) with a G5 filter. Calibration of the G5 filter was performed using a DS-33 Matrix Standard Kit (Applied Biosystems). The GeneScan data were subsequently analyzed with the GeneMapper software (Applied Biosystems). The TA repeat genotype was determined automatically by GeneMapper software along with an *in house* standard marker. The standard marker consists of amplicons containing the TA repeat region from 9 to 31 repeats. (See Supplementary Figure. S1 & S2) For samples in which GeneMapper software could not automatically call the genotype, The repeat number was validated by capillary sequencing, as we reported previously [13].

Definition of positive predictive value and negative predictive value

To evaluate the precision rate of the *IL28B* SNPs and TA repeat for the prediction of spontaneous clearance, we calculated the positive (PPV) and negative

predictive values (NPV). The PPV was defined as the rate (%) of spontaneous clearance among subjects with an advantageous genotype(s) of *IL28B* SNPs and/or TA repeat. In contrast, the NPV was defined as the rate (%) of chronic infection among subjects with a disadvantageous genotype(s) of *IL28B* SNPs and/or TA repeat.

Statistical analysis.

Associations between spontaneous HCV clearance and the candidate variables were analyzed by univariate and multiple logistic regression analysis. Student's t-test and the Wilcoxon-Mann-Whitney U-test were used to compare continuous variables between groups, and the Kruskal-Wallis and Bonferroni post hoc test were used for multiple group comparisons. Chi-square test was used to compare categorical variables. A p value <0.05 was considered statistically significant. To identify independent factors for predicting spontaneous HCV clearance, variables that reached the p <0.1 level in univariate analysis were used as candidate factors for multiple logistic regression analysis. These statistical analyses were performed using the SAS system, version 9.1.3 (SAS Institute, Cary, NC).

Results

TA repeat distribution and HCV status in the Japanese population.

Table 1 shows the allele distribution of the TA repeats classified by HCV status. The distribution was similar among the three HCV groups. The percentage of the allele with 12 repeats was approximately 80%, with the percentage gradually decreasing with the increased length of the TA repeat. No allele was found with 11 repeats. Interestingly, the allele with 10 was significantly more frequent in the group with chronic infection than in the spontaneous clearance group (3.5% vs 0.2%, P<0.001) and the healthy controls (3.5% vs 0.5%, P<0.001).

Clinical characteristics and HCV status.

The relation between the genetic variations and the age, sex, and clinical outcome of the participants are shown in Table 2. Female and favorable *IL28B* SNP were significantly correlated with spontaneous clearance compared with the chronic infection group (70.1% vs 53.7%, P<0.001; 90.6% vs 64.1%, P<0.001; respectively). The *IL28B* SNP was also correlated with spontaneous clearance compared with the healthy control group (90.6% vs 79.8%, P<0.001). The spontaneous clearance group was significantly older than the chronic infection and healthy control groups. The percentage of participants with a TA repeat of 10 was significantly higher in the chronic infection group in the univariate analysis (p<0.0001, 7.1%, 0.4%, 1.0% in the chronic infection, spontaneous clearance, and healthy control groups, respectively). Although persons with a TA repeat of 10 were found more frequently in the healthy control (1.0%) than in the spontaneous clearance group (0.4%), the difference did not reach statistical significance. The association between the genotypes of *IL28B* SNP and the TA repeat was analyzed for each group, with no significant correlation found (p=0.694, P=0.094, P=0.900 in the chronic infection, spontaneous clearance, and healthy control groups, respectively).

Independent factors contributing to spontaneous clearance.

Multiple logistic regression analysis of the aforementioned four variables was done for the spontaneous clearance and chronic infection groups, and all four retained their association as independent factors contributing to the spontaneous clearance of HCV: Sex (odds ratio [OR], 1.79; 95% confidence interval [CI], 1.22-2.65), age (OR, 1.04; 95% CI, 1.02-1.06), the genotype of rs8099917 (OR, 5.14; 95% CI, 3.12-8.82), and the TA repeats (OR, 13.02; 95% CI, 2.59-237.0) (Table 3).

Distribution of the TA repeat among African-Americans.

To validate the findings of our Japanese samples, we subsequently tested African-American samples with spontaneous clearance or chronic HCV infection. The allele frequencies as compared to our Japanese population for the spontaneous clearance

and chronic infection groups are shown in Fig. 1. Although 12 repeats was common to all four groups, the percentages differed greatly; approximately 30% for both African-American groups and 80% for both Japanese groups. The African-American data showed a gently sloping distribution of the TA repeat, in contrast to a sharp drop for the Japanese. Furthermore, the allele with 6 repeats, which was not detected in the Japanese sample, accounted for 16.3% of the spontaneous clearance group and 30.0% of the chronic infection group for the African-Americans. Interestingly, none of the samples in this study had alleles with 7, 8, or 9 repeats.

Association between spontaneous HCV clearance and the number of TA repeats in the African-American samples.

The clinical data of the African-American samples was analyzed in the same manner as the Japanese data. We classified the African-American samples into three groups according to a TA repeat of 10, the meaningful cut-off value for the Japanese samples; persons in whom the TA repeats of both alleles are 10 or shorter, persons in whom one of the two alleles is 10 or shorter, and persons with no allele of 10 or shorter. The rate of spontaneous clearance was significantly lower for persons in whom the TA repeats of both alleles were 10 or shorter compared to those with no allele of 10 or shorter (20.0% vs 59.4%, p=0.001) and compared to those with at least one allele 11 or longer. (20.0% vs 53.6%, p=0.004). Although the power did not reach statistical significance, the rate of spontaneous clearance decreased as the number of alleles with 10 or shorter TA repeats increased (59.4%, 46.3% and 20.0%) (Fig. 2). This suggests that an allele with TA repeats of 10 or shorter is a risk factor for HCV persistence.

Multivariate Analysis of the African-American samples.

The rs8099917 and rs12979860 genotypes and the TA repeats were analyzed by multiple logistic regression analysis. Comparing the persons in whom both alleles were 10 or shorter to the others, rs12979860 (OR, 3.24; 95% CI, 1.55-6.76) and the TA repeat length (OR, 3.70; 95% CI, 1.16-11.8) were extracted as independent variables

associated with spontaneous HCV clearance (Table 4). In the other models, such as comparing persons with no allele of 10 or shorter to the others or comparing the three groups independently, those with longer TA repeats tended to spontaneously clear the infection; however, the association did not reach statistical significance (data not shown).

Probability of the prediction of spontaneous HCV clearance for combinations of the *IL28B* SNPs and TA repeat.

For our Japanese samples, the PPV for spontaneous clearance based on the *IL28B* SNP alone was 49.3% (203 / 412). Among 412 persons with the favorable genotype of the *IL28B* SNP, 13 had an unfavorable TA repeat of 10, all of whom belonged to chronic infection group. The addition of the TA repeat raised the PPV to 50.9% (203 / 399). The NPV for spontaneous clearance based on the *IL28B* SNP alone was 84.8% (117 / 138). Among 138 persons with an unfavorable genotype of *IL28B* SNP, 11 had an unfavorable TA repeat, only one of whom was in the spontaneous clearance group. The NPV increased to 90.9% (10 / 11) with the addition of the TA repeat. (See Supplementary Figure. S3.) For the African-American sample, the PPV for *IL28B* SNP alone did not differ from the PPV for the *IL28B* SNP and the TA repeat (75.0%: 39 / 52) because none of the subjects with a favorable *IL28B* SNP had an unfavorable TA repeat. On the other hand, the NPV increased from 58.4% (87 / 149) to 80.0% (16 / 20) with the TA repeat. (See Supplementary Figure. S4.)

DISCUSSION

This study shows that the length of the TA repeat is an independent factor associated with spontaneous HCV clearance in Japanese and African-American populations. We previously reported that long TA repeats are associated with viral response to PEG-IFN/RBV therapy in a study of 48 patients with chronic hepatitis C

[13]. These investigations suggest that the length of the TA repeat plays an important role in the elimination of HCV infection.

Although the favorable *IL28B* SNPs genotype has been reported to be associated with the clinical outcome of IFN therapy [6-8] and spontaneous clearance [11], it alone cannot explain all HCV clearance. In this study, a combination of the *IL28B* SNPs and the TA repeat was shown to have improved the prediction of spontaneous HCV clearance. Although the impact of the TA repeat may be limited in the clinical setting of the Japanese population due to the extremely low prevalence of the unfavorable genotype, among a population such as African -Americans in which the favorable alleles of *IL28B* and the TA repeat are not predominant, the TA repeat will be a more useful marker. The TA repeat is a novel and helpful genetic marker for use with the *IL28B* SNPs. The prediction of the course of acute HCV infection is important, especially so for determining which patients are likely to transition to the chronic phase and for which patient IFN-based therapy should be considered. IFN-based therapy has serious side effects and is costly. By identifying patients who have a high probability of spontaneous clearance, unnecessary treatment can be avoided and the cost to both the patients and the medical system reduced. In addition, if an HCV vaccine becomes available in the future, the identification of patients who are likely to develop persistent HCV infection would be useful for determining who should receive preference for vaccination.

We investigated the distribution of the length of the TA repeat in a large Japanese population, 1,060 samples that included 510 of healthy volunteers, and found that the allele with 12 repeats accounts for approximately 80%. Although our samples of African-Americans were from a selected population, it is likely that the distribution of the whole African-American population is quite different from the Japanese population, as shown by the allele containing six repeats only being detected in this population. Because most persons infected with HCV are asymptomatic and do not have

a medical examination during the period of acute hepatitis, it is difficult to clarify the precise rate of spontaneous clearance of HCV. However, it has been reported to be from 14 to 46 % and different by race and ethnicity [2]. Studies of Japanese cohorts have reported rates from 22 to 30 %.[14, 21, 22] HCV persistence has been reported to be more likely among black people [23, 24] and the favorable *IL28B* SNPs allele has been recently found to be less frequent among persons of African descent [6, 11]. Although the *IL28B* SNPs partly explains the racial and ethnic discrepancy in the frequency of spontaneous HCV clearance, the addition of the TA repeat adds useful information to the genetic factors that contribute to racial differences. We have shown that the distribution of the TA repeat is markedly different between African-American and Japanese populations. TA repeats of 10 or shorter, which were shown to be disadvantageous for spontaneous HCV clearance, were more frequent in the African-American population than in the Japanese population tested. Further studies are needed to validate the distribution and predictive power of the TA repeat among other races and ethnicities.

Other studies of TA dinucleotide repeat polymorphism and disease reported an association between the uridine diphosphoglucuronosyl transferase 1A1 (UGT1A1) gene and Gilbert syndrome, a benign form of unconjugated hyperbilirubinemia. There is a TA repeat polymorphism in the promoter of UGT1A1, and elongated TA repeats have been shown to cause Gilbert's syndrome [25]. Although the number of TA repeats in UGT1A1 is 6 or 7 in Caucasian [25, 26] and Asian [27] populations, some Africans have 5 and 8 [27, 28], similar to our present study. On the other hand, the transcriptional activity of the promoter increases gradually with an increase in the number of TA repeats in *IL28B* [13] while the length of the TA repeat in UGT1A1 has an inverse relationship to the transcriptional activity [27]. These similarities and differences between these two TA repeats may be hints to the mechanism of how microsatellite regions contribute to gene expression.

In our present study, the genotype of the TA repeat that was significantly correlated with spontaneous clearance was different between the Japanese and African-American cohorts. African-Americans who had at least one allele with a TA repeat of 11 or longer were significantly more advantageous compared to persons in whom both alleles had TA repeats of 10 or shorter. This is slightly different from the results for Japanese, in which persons who had no allele with a TA repeat of 10 were advantageous compared to the others. However, it is possible that the statistically significant genotype would change if the allele frequency were different among populations. For example, in addition to the above-mentioned TA repeat near UGT, two functional SNPs were correlated with the expression of UGT, and it was reported that a combination of these polymorphisms was an effective predictor of severe side-effects induced by irinotecan in a Caucasian population [29]. However, the allele frequencies of these polymorphisms were markedly different and the Caucasian criterion was not applicable to the African population [28]. Similarly, the allele frequencies of our TA repeats were shown to be markedly different between African-Americans and Japanese; the African-Americans had a wider range and more a gently sloping distribution of the TA repeat than the Japanese. It is not surprising that different races and ethnicities would have different criteria.

In the African-American cohort, the highest percentage of spontaneous clearance was found for persons in whom the TA repeats of both alleles were 11 or longer, the lowest for persons in whom the TA repeats of both alleles were 10 or shorter, and those with one long and one short allele were intermediate. This result strongly supports the hypothesis that a TA repeat of 10 or shorter is a risk factor for HCV, however, the power did not reach statistical significance. In this study, only two racial groups were investigated and the sample size of African-Americans was smaller than that of Japanese. If a larger population of many racial groups and ethnicities were to be studied, we would be able to develop a classification for the TA repeats and their

contribution to spontaneous HCV clearance. The advantageous length of the TA repeat may be different among races and ethnicities. Further study is needed to clarify these possibilities.

In this study, we mainly analyzed inherent factors such as sex and genotype. Some laboratory values, such as platelet count, serum albumin, and alanine aminotransferase, are well documented as being associated to the clinical course of chronic hepatitis. We excluded these parameters from our analyses because they were missing for many of the participants. However, it is thought that they were normal before the infection for almost all the persons in the spontaneous clearance and chronic infection groups and thus would have little affect on the findings.

Although the correlation of sex [12, 30] and *IL28B* SNPs [11, 12] to spontaneous HCV clearance was previously reported, older age, which was an independent factor in our study of the Japanese population, is contradictory to the common wisdom [23]. Because this is an epidemiological study that used samples collected at the time of a general medical check, we could not determine the age at which the participants had been infected or cleared the infection. Therefore, it is possible that older age was erroneously extracted as an independent factor that contributes to spontaneous clearance.

A limitation of this study is that our African-Americans samples were selected from persons who were positive for anti-HCV, and we could not investigate the distribution of the genotype of the TA repeat in general African-American population. Although the genotypes of *IL28B* and the TA repeat were not associated each other in the general Japanese population, we could not validate the association for the African-Americans in this study because of the small sample size and because only anti-HCV positive patients were included. The fact that the linkage disequilibrium around the *IL28B* gene is in much weaker in African than in Japanese or Caucasian people [31] suggests that the IL28B SNP and TA repeat had no or little relation in

African-American general population. However, further study of the general population is needed.

In conclusion, a long TA repeat in the promoter region of *IL28B* was associated with spontaneous HCV clearance. The findings indicate that this novel genetic factor may be useful for improving the prediction of HCV clearance, along with the *IL28B* SNPs. However, further study is needed to validate the utility of the TA repeat among people of various races and ethnicities.

Acknowledgments

This work was supported by grants from the National Center for Global Health and Medicine of Japan (22-108, 25-202), from the Japanese Ministry of Health, Labor and Welfare (H23-007), and from the Japanese Ministry of Education, Culture, Sports, and Science (No. 22590750). We are also grateful to all the technical staffs of Department of Hepatic Diseases, The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine for their excellent lab works on genotyping the SNPs of IL28B and TA repeat.

References

1. Ray Kim W. Global epidemiology and burden of hepatitis C. Microbes Infect. 2002;4:1219-25.

2. Seeff LB. Natural history of chronic hepatitis C. Hepatology. 2002;36:S35-46.

3. Fried MW, Shiffman ML, Reddy KR, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med. 2002;347:975-82.

4. Manns MP, McHutchison JG, Gordon SC, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet. 2001;358:958-65.

5. Furusyo N, Kajiwara E, Takahashi K, et al. Association between the treatment length and cumulative dose of pegylated interferon alpha-2b plus ribavirin and their effectiveness as a combination treatment for Japanese chronic hepatitis C patients: project

of the Kyushu University Liver Disease Study Group. J Gastroenterol Hepatol. 2008;23:1094-104.

6. Ge D, Fellay J, Thompson AJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature. 2009;461:399-401.

7. Suppiah V, Moldovan M, Ahlenstiel G, et al. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. Nat Genet. 2009;41:1100-4.

8. Tanaka Y, Nishida N, Sugiyama M, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet. 2009;41:1105-9.

9. Kotenko SV, Gallagher G, Baurin VV, et al. IFN-lambdas mediate antiviral protection through a distinct class II cytokine receptor complex. Nat Immunol. 2003;4:69-77.

10. Marcello T, Grakoui A, Barba-Spaeth G, et al. Interferons alpha and lambda inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics. Gastroenterology. 2006;131:1887-98.

11. Thomas DL, Thio CL, Martin MP, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature. 2009;461:798-801.

12. Rao HY, Sun DG, Jiang D, et al. IL28B genetic variants and gender are associated with spontaneous clearance of hepatitis C virus infection. J Viral Hepat. 2012;19:173-81.

13. Sugiyama M, Tanaka Y, Wakita T, et al. Genetic variation of the IL-28B promoter affecting gene expression. PLoS One. 2011;6:e26620.

14. Kishihara Y, Furusyo N, Kashiwagi K, et al. Human T lymphotropic virus type 1 infection influences hepatitis C virus clearance. J Infect Dis. 2001;184:1114-9.

15. Uto H, Stuver SO, Hayashi K, et al. Increased rate of death related to presence of viremia among hepatitis C virus antibody-positive subjects in a community-based cohort study. Hepatology. 2009;50:393-9.

16. Watanabe H, Saito T, Shinzawa H, et al. Spontaneous elimination of serum hepatitis C virus (HCV) RNA in chronic HCV carriers: a population-based cohort study. J Med Virol. 2003;71:56-61.

17. Vlahov D, Anthony JC, Munoz A, et al. The ALIVE study, a longitudinal study of HIV-1 infection in intravenous drug users: description of methods and characteristics of participants. NIDA Res Monogr. 1991;109:75-100.

18. Ito K, Higami K, Masaki N, et al. The rs8099917 polymorphism, when determined by a suitable genotyping method, is a better predictor for response to pegylated alpha interferon/ribavirin therapy in Japanese patients than other single nucleotide polymorphisms associated with interleukin-28B. J Clin Microbiol. 2011;49:1853-60.

19. Lyamichev V, Mast AL, Hall JG, et al. Polymorphism identification and

quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. Nat Biotechnol. 1999;17:292-6.

20. Lyamichev VI, Kaiser MW, Lyamicheva NE, et al. Experimental and theoretical analysis of the invasive signal amplification reaction. Biochemistry. 2000;39:9523-32.

21. Ishibashi M, Shinzawa H, Kuboki M, et al. Prevalence of inhabitants with anti-hepatitis C virus antibody in an area following an acute hepatitis C epidemic: age-and area-related features. J Epidemiol. 1996;6:1-7.

22. Uto H, Hayashi K, Kusumoto K, et al. Spontaneous elimination of hepatitis C virus RNA in individuals with persistent infection in a hyperendemic area of Japan. Hepatol Res. 2006;34:28-34.

23. Thomas DL, Astemborski J, Rai RM, et al. The natural history of hepatitis C virus infection: host, viral, and environmental factors. JAMA. 2000;284:450-6.

24. Seeff LB, Miller RN, Rabkin CS, et al. 45-year follow-up of hepatitis C virus infection in healthy young adults. Ann Intern Med. 2000;132:105-11.

25. Raijmakers MT, Jansen PL, Steegers EA, et al. Association of human liver bilirubin UDP-glucuronyltransferase activity with a polymorphism in the promoter region of the UGT1A1 gene. J Hepatol. 2000;33:348-51.

26. Borlak J, Thum T, Landt O, et al. Molecular diagnosis of a familial nonhemolytic hyperbilirubinemia (Gilbert's syndrome) in healthy subjects. Hepatology. 2000;32:792-5.

27. Beutler E, Gelbart T, Demina A. Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? Proc Natl Acad Sci U S A. 1998;95:8170-4.

28. Horsfall LJ, Zeitlyn D, Tarekegn A, et al. Prevalence of clinically relevant UGT1A alleles and haplotypes in African populations. Ann Hum Genet. 2011;75:236-46.

29. Lankisch TO, Schulz C, Zwingers T, et al. Gilbert's Syndrome and irinotecan toxicity: combination with UDP-glucuronosyltransferase 1A7 variants increases risk. Cancer Epidemiol Biomarkers Prev. 2008;17:695-701.

30. Kenny-Walsh E. Clinical outcomes after hepatitis C infection from contaminated anti-D immune globulin. Irish Hepatology Research Group. N Engl J Med. 1999;340:1228-33.

31. Tanaka Y, Nishida N, Sugiyama M, et al. lambda-Interferons and the single nucleotide polymorphisms: A milestone to tailor-made therapy for chronic hepatitis C. Hepatol Res. 2010;40:449-60.

FIGURE

Fig.1.

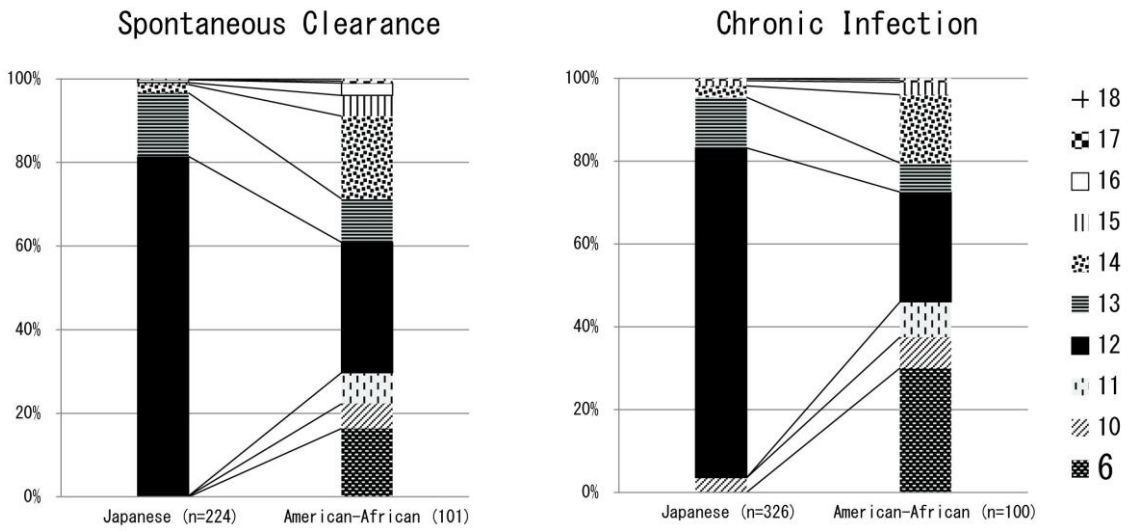

Fig.1. Length of the TA repeat of African-Americans in comparison with the Japanese population

Contrary to the Japanese distribution that has a very high peak at 12, the African-American data showed a gently sloping distribution. In addition, 6 repeats was only detected in the African-American cohort, accounting for 16.3% in the spontaneous clearance group and 30.0% in the chronic infection group.

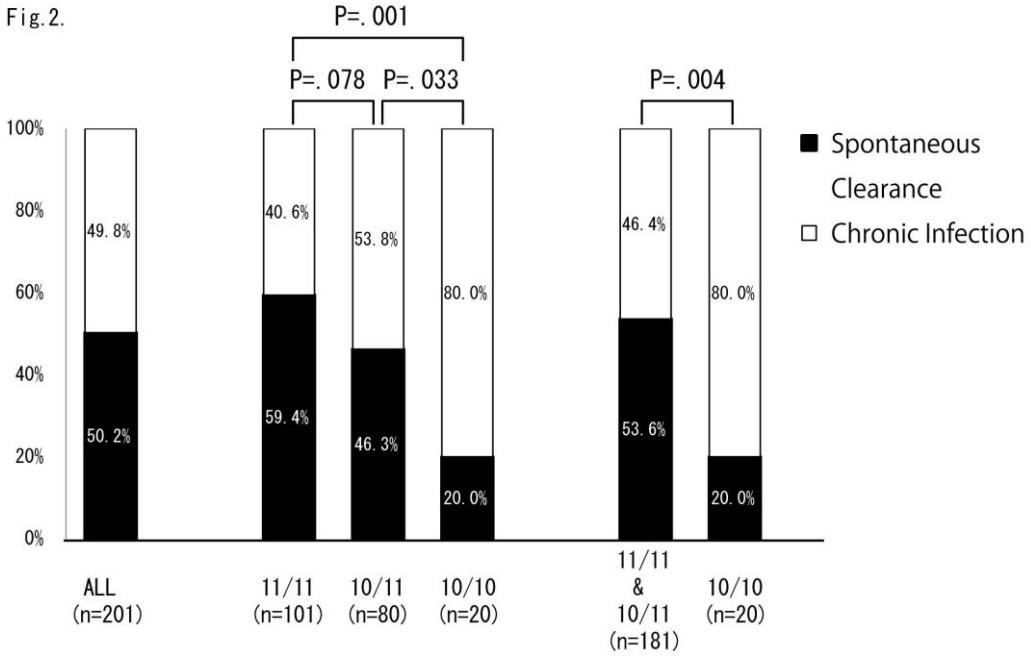

Fig.2. Association between spontaneous HCV clearance and the TA repeat length of the African-American cohort

11/11, persons in whom both alleles of the TA repeats are 11 or greater; 10/11, persons in whom one of two alleles of the TA repeat is under 11 and the other is 11 or greater; 10/10, persons in whom both alleles of the TA repeat are under 11: Chi-square test was used to investigate the association between the TA repeat and spontaneous clearance of HCV. The rate of spontaneous clearance was significantly lower for 10/10 (20.0%) than for 11/11 (59.4%) or the combined group of 11/11 and 10/11 (53.6%).

Supplementary figure legends

Supplementary Fig. S1

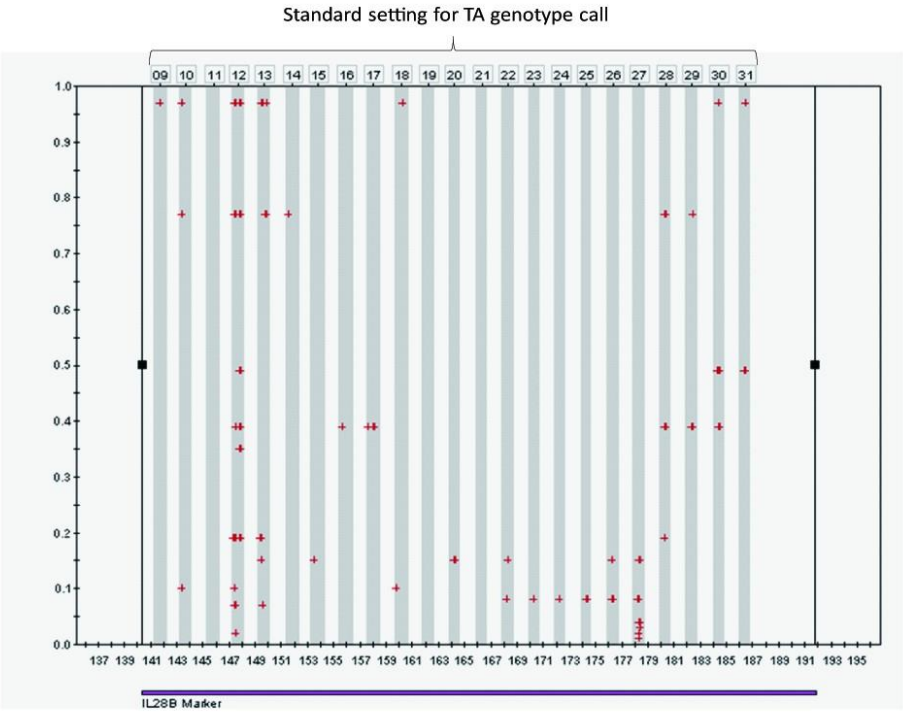

Fig. S1. GeneScan Analysis of TA repeats near *IL28B*: Genetic variations of the TA dinucleotide repeat near the *IL28B* gene were genotyped by GeneScan analysis. Human genome samples were cloned into plasmids, and the TA genotypes were determined by capillary sequencing. These standard plasmids were amplified using specific primers, and the DNA fragment consisting of 9 to 31 TA repeats were sized by GeneScan software. Each peak was set as the standard TA genotype (gray bar). The TA genotype of the samples was automatically determined using the size standard.

Supplementary Figure. S2

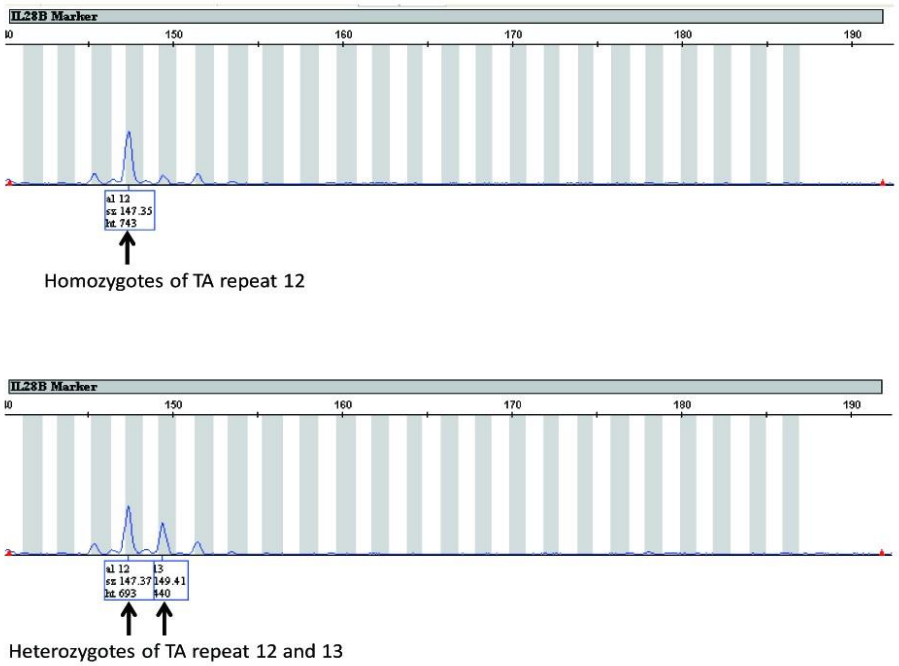

Fig. S2. GeneScan Analysis of TA repeats near *IL28B*: The software automatically calculated the genotype of the TA repeats along with an *in house* standard marker obtained from clinical samples. The upper data shows TA repeats of the 12/12 genotype and the lower displays the 12/13 genotype.

Supplementary Fig. S3.

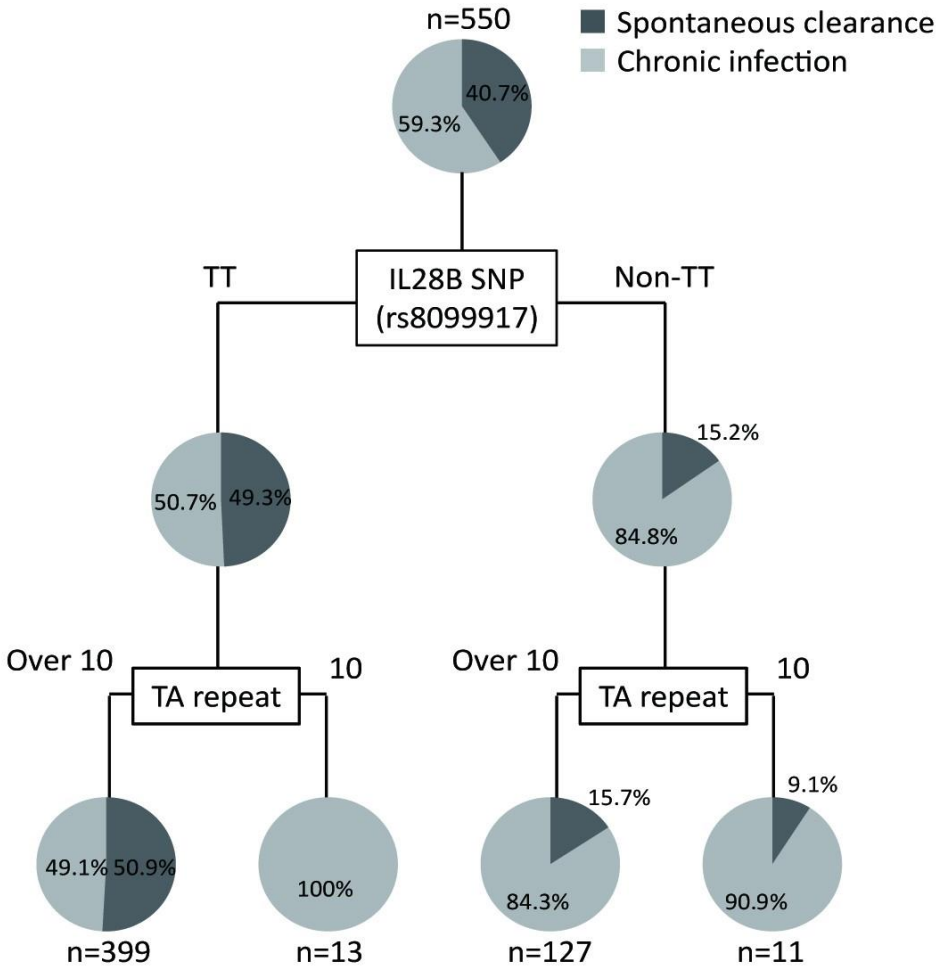

Fig. S3. Tree Diagram for Predicting Spontaneous HCV Clearance among the Japanese population: Among 550 persons, 412 had TT (favorable) genotype rs8099917. Among them, 203 (49.3%) were in the spontaneous clearance group. Among the 412 with TT genotype rs8099917, 399 had the favorable genotype of the TA repeat. Among them, 203 (50.9%) were in the spontaneous clearance group. Among the 138 with non-TT genotype rs8099917, 117 (84.8%) were in the chronic infection group and 11 had the unfavorable TA repeat, 10 (90.9%) of whom were in the chronic infection group. Although the most efficient factor was *IL28B* SNPs, none of the 13 persons with the favorable *IL28B* SNPs genotype and short TA repeat spontaneously cleared the infection. The percentage of spontaneous clearance reached over 50% among paticipants with the favorable IL28B SNPs genotype and a long TA repeat.

Supplementary Fig. S4.

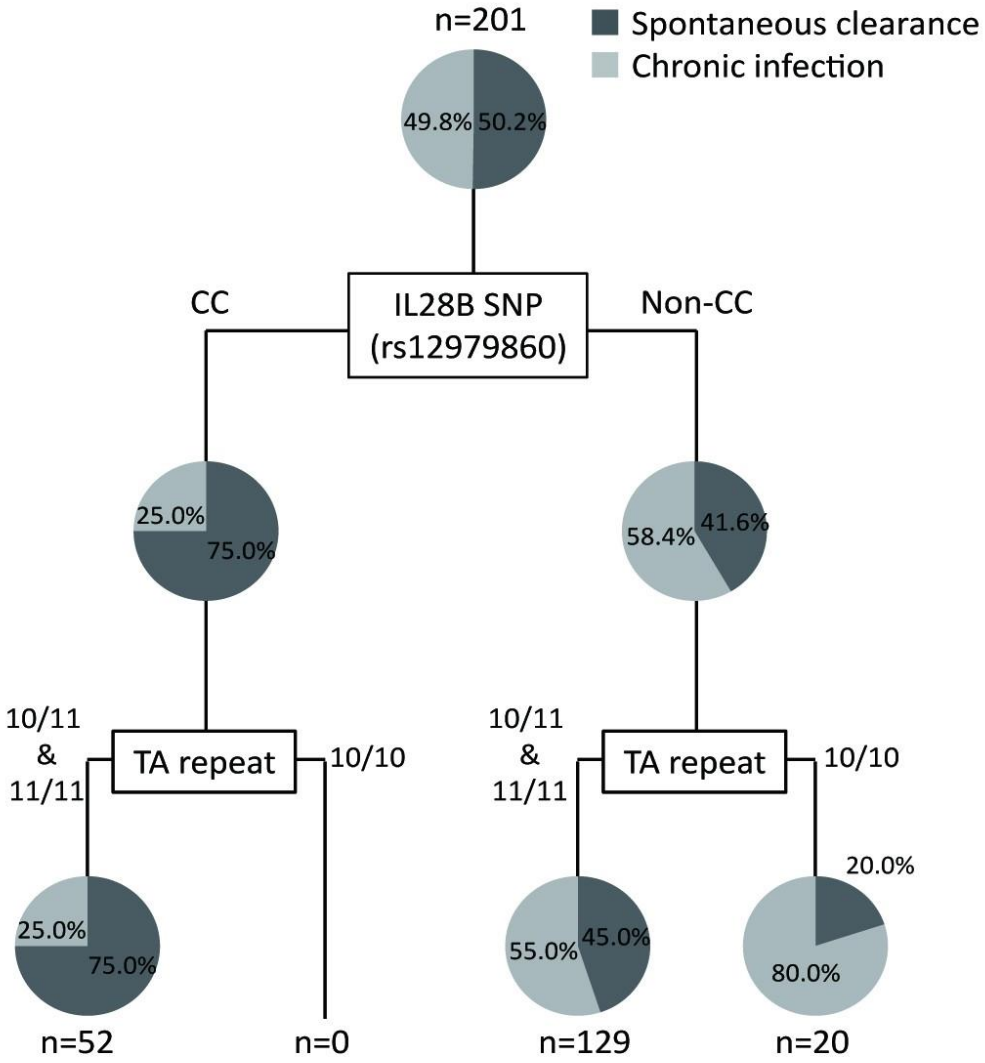

Fig. S4. Tree Diagram for Predicting Spontaneous HCV Clearance among the American-African population: Among 201 persons, 52 had CC (favorable) genotype rs12979860. None had the unfavorable TA repeat genotype, and 39 (75.0%) were in the spontaneous clearance group. Among the 149 with non-CC genotype rs12979860, 87 (58.4%) were in the chronic infection group and 20 had the unfavorable TA repeat, 16 (80.0%) of whom were in chronic infection group.

Table 1. Allele frequency of rs59702201 (TA repeat) among persons with spontaneous clearance, patients with chronic hepatitis C and a general Japanese population

	Spontaneous Clearance (n=224)	Chronic Infection (n=326)	Healthy Control (n=510)
Repeat number of rs59702201	n (%)	n (%)	n (%)
10	1 (0.2)	23 (3.5)	5 (0.5)
11	0 (0.0)	0 (0.0)	0 (0.0)
12	364 (81.3)	519 (79.6)	873 (85.6)
13	68 (15.2)	80 (12.3)	101 (9.9)
14	9 (2.0)	18 (2.8)	26 (2.6)
15	2 (0.5)	8 (1.2)	4 (0.4)

16	3 (0.8)	0 (0.0)	6 (0.6)
17	0 (0.0)	3 (0.5)	5 (0.5)
18	1(0.2)	1 (0.2)	0 (0.0)

1 Abbreviations: TA, Thymine-Adenine.

Table 2. Characteristics of three Japanese groups classified by HCV status

	Spontaneous Clearance (n=224)	Chronic Infection (n=326)	Healthy Control (n=510)	P value
Age, median (interquartile range)	70 (14.75) [a]	63 (15.25)	67 (16.0)	<0.0001
Female, n (%)	157 (70.1) [b]	175 (53.7) [c]	342 (67.1)	<0.0001
rs8099917 genotype: TT, n (%)	203 (90.6) [b, c]	209 (64.1) [c]	407 (79.8)	<0.0001
rs59702201 genotype: Over10, n (%)	223 (99.6) [b]	303 (92.9) [c]	505 (99.0)	<0.0001

P values were based on one-way ANOVA or Chi-square test for continuous or categorical variables, respectively. [a] P <0.0001 vs. Chronic Infection group and P =0.0006 vs. Healthy Control group. [b] P <0.001 vs. Chronic Infection group, by Bonferroni correction. [c] P <0.001 vs. Healthy Control group, by Bonferroni correction.

Abbreviations: Over10, persons who had no allele with a TA repeat of 10.

Table 3. Multivariate analysis of HCV clearance by the Japanese participants

	Spontaneous Clearance (n=224)	Chronic Infection (n=326)	Univariate Analysis	Multivariate Analysis	
			P value	OR(95% CI)	P value
Age, median (interquartile range)	70 (14.75)	63 (15.25)	<0.0001	1.04 (1.02-1.06)	<0.0001
Female, n (%)	157 (70.1)	175 (53.7)	<0.0001	1.79 (1.22-2.65)	0.0031
rs8099917 genotype: TT, n (%)	203 (90.6)	209 (64.1)	<0.0001	5.14 (3.12-8.82)	<0.0001
rs59702201 genotype: Over10, n (%)	223 (99.6)	303 (92.9)	<0.0001	13.02 (2.59-237.0)	0.0004

The multiple logistic regression analysis was done using variables identified in univariate analysis with P values of <0.1 to investigate the associations between spontaneous HCV clearance and the TA repeat among Japanese population. A P value <0.05 was considered statistically significant.

Abbreviations: OR, Odds Ratio; CI, Confidence Intervals; Over10, persons who had no allele with a TA repeat of 10.

Table 4. Multivariate Analysis of HCV clearance by the American-African participants

	Spontaneous Clearance (n=101)	Chronic Infection (n=100)	Univariate Analysis	Multivariate Analysis	
			P value	OR (95% CI)	P value
Female, n (%)	49 (48.5)	48 (48.0)	0.942		
rs8099917 genotype: TT, n (%)	93 (92.1)	83 (83.0)	0.051	1.97 (0.78-4.96)	0.151
rs12979860 genotype: CC, n (%)	39 (38.6)	13 (13.0)	<0.001	3.24 (1.55-6.76)	0.002
rs59702201 genotype: 10/11 & 11/11, n (%)	97 (96.0)	84 (84.0)	0.004	3.70 (1.16-11.8)	0.027

The multiple logistic regression analysis was done using variables identified in univariate analysis with P values of <0.1 to investigate the associations between spontaneous HCV clearance and the TA repeat among American-African population. A P value <0.05 was considered statistically significant.

Abbreviations: OR, Odds Ratio; CI, Confidence Intervals; 10/11, persons in whom one of two alleles of the TA repeats is under 11; 11/11, persons in whom both alleles of the TA repeats are 11 or greater.

Supplementary Table. The primer and probe sets for InvaderPlus assay

	Primer name	Sequence
rs8099917	Primer F	TCATCCCTCATCCCACTTCTGGAACA
	Primer R	CAGCAGGAAACAGATGGCCCG
	Invader probe	TTCCTTTCTGTGAGCAATKTCACCCAAATTGGAACCATGCTGTTATACAGTTTGGTAGC
rs12979860	Primer F	GGATGGGTACTGGCAGCGC
	Primer R	GCTGACATAGGAGAGGCGCCT
	Invader probe	CCAGGGAGCTCCCCGAAGGCGYGAACCAGGGTTGAATT