

TABLE I. Comparison of Pretreatment Factors Between Model Building and Validation Patients

	Model (n = 304)	Validation (n = 201)	P-value
Age (years)	55.6 (9.4)	56.0 (12.2)	0.80
Male (%)	53 (%)	55 (%)	0.13
Body mass index (kg/m ²)	23.1 (3.1)	23.1 (4.0)	0.99
Albumin (g/dl)	4.0 (0.3)	4.0 (0.3)	0.47
Creatinine (mg/dl)	0.72 (0.15)	0.72 (0.14)	0.62
AST (IU/L)	63.3 (45.6)	58.9 (46.4)	0.91
ALT (IU/L)	78.7 (58.6)	74.5 (67.5)	0.68
GGT (IU/L)	53.2 (49.1)	57.4 (63.5)	0.43
Total cholesterol (mg/dl)	170.9 (32.6)	169.4 (34.1)	0.33
Triglyceride (mg/dl)	107.0 (44.7)	105.7 (48.0)	0.90
LDL-C (mg/dl)	95.5 (28.0)	96.4 (28.8)	0.34
White blood cell count (/μl)	4,902 (1,489)	4,906 (1,319)	0.86
Hemoglobin (g/dl)	14.1 (1.3)	14.3 (1.4)	0.09
Platelets (10 ⁹ /L)	164 (56)	172 (55)	0.68
HCVRNA (10 ³ IU/ml)	1,859 (1,468)	2,021 (1,393)	0.09
ISDR mutations: ≥2 (%)	15 (%)	20 (%)	0.11
Core70: mutant (%)	36 (%)	29 (%)	0.22
Core91: mutant (%)	40 (%)	36 (%)	0.20
Fibrosis: F2–4 (%)	49 (%)	48 (%)	0.36
Activity: A2–3 (%)	42 (%)	34 (%)	0.10

AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma-glutamyltransferase; LDL-C, low-density-lipoprotein-cholesterol; ISDR, interferon sensitivity-determining region. Data expressed as mean (SD).

0–3: A0 (no activity), A1 (mild activity), A2 (moderate activity), and A3 (severe activity). Sustained virological response was defined as undetectable HCV RNA by qualitative PCR with a lower detection limit of 50 IU/ml (Amplicor, Roche Diagnostic Systems) at week 24 after the completion of therapy.

Statistical Analysis

A database of pretreatment variables included hematological tests (hemoglobin level, white blood cell count, and platelet count), blood chemistry tests (serum levels of creatinine, albumin, aspartate aminotransferase, alanine aminotransferase (ALT), gamma-glutamyltransferase (GGT), total cholesterol, triglyceride, and low-density lipoprotein cholesterol (LDL-C)), viral factors (HCV RNA titer, number of substitutions in ISDR, substitutions in the amino acid positions 70 and 91 of the core region), histological findings (stage of fibrosis and grade of activity) and patient characteristics (age, sex, and body mass index). Based on this database, decision-tree analysis was used to define a predictive model for sustained virological response.

Student's *t*-test was used for the univariable comparison of quantitative variables and Fisher's exact test was used for the comparison of qualitative variables. For the multivariable analysis for factors associated with sustained virological response, logistic regression models with backward selection were used to identify independent predictors of sustained virological response. Variables that showed significant association with sustained virological response by univariable analysis were included in the multivariable analysis. IBM-SPSS software v.15.0 (SPSS, Inc., Chicago, IL) was used for these analyses. For the decision-tree analysis [Segal and

Bloch, 1989], the data mining software IBM SPSS Modeler 13 (IBM SPSS, Inc.) was used, as reported previously [Kurosaki et al., 2010a,b]. In brief, the software searched for the optimal split variables to build a decision-tree structure. The entire study population was first evaluated to determine the variables and cut-off points for the most significant division into two subgroups having different probabilities of sustained virological response. Thereafter, analysis was repeated on all subgroups in the same way until either no additional significant variable was detected or the sample size was below 20.

RESULTS

Generation of the Decision-Tree Model

The decision-tree analysis selected five predictive variables to produce six subgroups of patients (Fig. 1). The number of substitutions in ISDR was selected as the best predictor of sustained virological response. The possibility of achieving sustained virological response was 83% for patients with two or more substitutions in ISDR compared with 44% for patients with a single or no substitution. Among patients with a single or no substitution in ISDR, age, with an optimal cut-off of 60 years, was selected as the variable of second split. Patients younger than 60 had the higher probability of sustained virological response (55%) compared with those older than 60 years (31%). Among younger patients, amino acid substitution at Core70 was selected as the third variable of split—wild-type sequence being the predictor of favorable response compared with the mutant type (65% vs. 36%). Among patients with wild-type Core70, the level of serum LDL-C was selected as the fourth variable of split, with an optimal cutoff of

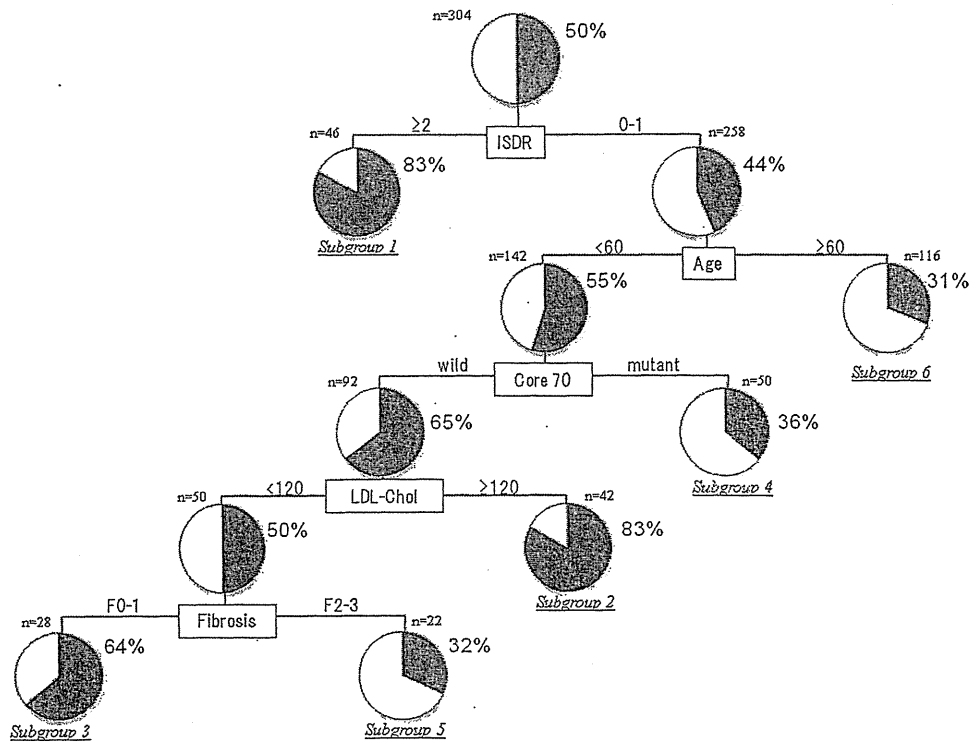


Fig. 1. Decision-tree model. Boxes indicate the factors used for splitting and the cutoff value for the split. Pie charts indicate the rate of sustained virological response for each group of patients after splitting. Terminal subgroups of patients discriminated by the analysis are numbered from 1 to 7. The rate of sustained virological response was >80% in subgroups 1 and 2, 64% in subgroup 3, and 31–36% in subgroups 4, 5, and 6. LDL-C represents low-density lipoprotein cholesterol and Core70 represents amino acid substitution at position 70 of the core region.

120 mg/dl. Patients with higher LDL-C level had the higher probability of sustained virological response (83% vs. 50%). The stage of fibrosis was selected as the final variable of split, with significant fibrosis (F2–4) being the predictor of lower sustained virological response probability (64% vs. 32%).

Among the six subgroups derived by this decision tree, the subgroup of patients with two or more substitutions in ISDR (subgroup 1) or with a single or no substitution in ISDR but younger than 60 years of age, having the wild-type Core70 and high serum level of LDL-C (≥ 120 mg/dl) (subgroup 2) showed the highest probability of sustained virological response (83%).

Validation of the Decision-Tree Model

The decision-tree model was validated using a validation dataset of 201 cases that were not included the model-building dataset. Each patient in the validation set was allocated to subgroups 1–6 using the flowchart form of the decision tree. The rates of sustained virological response were 75% for subgroup 1, 73% for subgroup 2, 65% for subgroup 3, 41% for subgroup 4, 46% for subgroup 5, and 33% for subgroup 6. The rates of sustained virological response for each subgroup of patients were correlated closely between the model building dataset and the validation dataset ($r^2 = 0.94$) (Fig. 2).

J. Med. Virol. DOI 10.1002/jmv

The six subgroups were reconstructed into three groups according to their rate of sustained virological response: the high-probability group consisted of subgroups 1 and 2, the intermediate-probability group consisted of subgroup 3, and the low-probability group consisted of subgroups 4, 5, and 6. The rate of sustained virological response in the high-probability group was high on a consistent basis: 83% for model-building patients and 74% for validation patients. The rate of sustained virological response in the intermediate-probability group was 64% for model building patients and 65% for internal validation patients. The rate of sustained virological response in the low-probability group was low on a consistent basis: 32% for model-building patients and 36% for internal validation patients (Fig. 3). Thirty percent of the patients were classified into the high-probability group and 10% of the patients were classified into intermediate-probability group, which means that about 40% of patients with higher than average probability of achieving sustained virological response were identified.

Effect of Dose Reductions of PEG-IFN and RBV

The possible effect of drug reductions was analyzed in the three groups of patients divided by decision tree (low-, intermediate-, and high-probability groups)

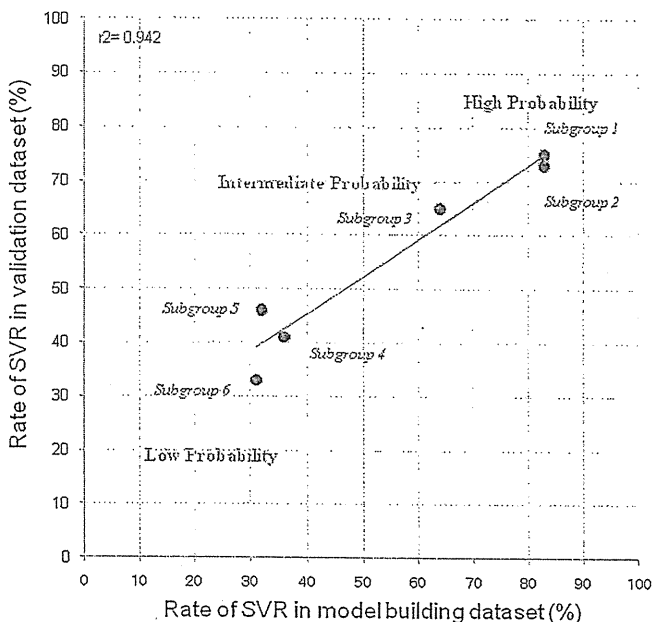


Fig. 2. Validation of the decision-tree analysis: Subgroup-stratified comparison of the rate of sustained virological response. Each patient in the validation set was allocated to subgroups 1–6 by following the flowchart form of the decision tree, and the rates of sustained virological response were then calculated and plotted for each subgroup. The x-axis represents the rate of sustained virological response in the model-building datasets and the y-axis represents the rate of sustained virological response in the validation datasets. The rates of achieving sustained virological response in each subgroup of patients correlated closely between the model-building dataset and the validation dataset (correlation coefficient: $r^2 = 0.94$).

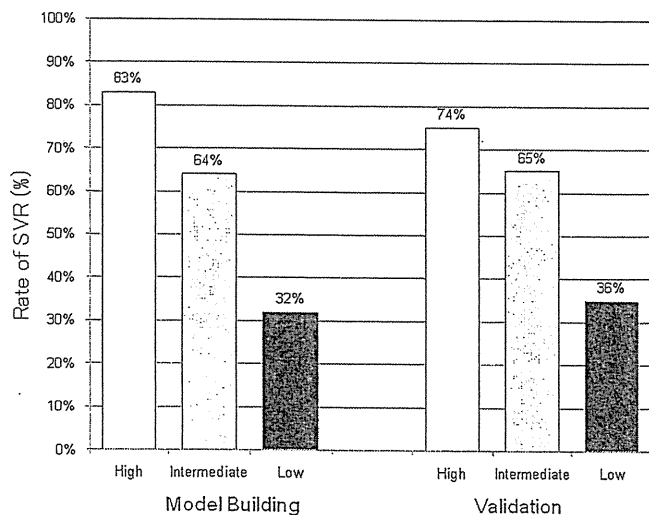


Fig. 3. Comparison of sustained virological response rates between groups divided by the decision tree. The rate of sustained virological response was compared between three groups of patients as divided by the decision-tree analysis. Black, gray, and white boxes indicate the low-probability group (subgroup 4, 5, and 6), intermediate-probability group (subgroup 3), and high-probability group (subgroup 1 and 2), respectively. The rate of sustained virological response showed significant difference between the three groups.

(Fig. 4). Patients were stratified according to the cumulative drug exposure with PEG-IFN and RBV: the good adherence group consisted of patients who took $\geq 80\%$ planned doses of both PEG-IFN and RBV; the poor adherence group consisted of patients who took $< 80\%$ of planned doses of both PEG-IFN and RBV. Even after adjustment for drug adherence, the three groups of patients divided by decision-tree analysis still had low, intermediate, and high probability of achieving sustained virological response, respectively, indicating that this model predicts sustained virological response independent of drug exposure.

Multivariable Logistic Regression Analysis

Age, sex, serum levels of creatinine, ALT, GGT, LDL-C, hemoglobin, platelet count, HCV RNA titer, ISDR substitution, substitution at Core70, substitution at Core91, histological stage of fibrosis, and grade of activity were found to be associated with sustained virological response by standard univariable analysis. Multivariable analysis including these factors showed that age, sex, LDL-C levels, GGT levels, platelet count, ISDR substitution, and substitution at Core70 showed independent associations with sustained virological response (Table II). Substitution in ISDR had the highest odds ratio, at 9.92. Fibrosis, which was selected as a significant predictor of response in the decision-tree analysis, was not found to be an independent predictor of response in standard multivariable analysis, indicating that the decision-tree analysis could identify significant predictors that would apply specifically to selected patients.

DISCUSSION

The present study revealed that viral factors such as substitutions in ISDR and Core70 are significant and independent predictors of sustained virological response to PEG-IFN plus RBV in chronic hepatitis C. In a decision-tree model for the pretreatment prediction of sustained virological response, the number of substitutions in ISDR was the best predictor of sustained virological response, followed by younger age, wild-type sequence at Core70, higher level of LDL-C, and absent fibrosis. This decision-tree model could identify patients with high probability of sustained virological response (83%) among difficult-to-treat genotype 1b chronic hepatitis C patients. Using this model, rapid estimates of the response before treatment can be made by allocating patients to specific subgroups with a defined rate of response simply by following the flowchart form. Because more potent therapy, such as a combination of protease inhibitor, PEG-IFN, and RBV, is under clinical trial and may become available in the near future [Hezode et al., 2009; McHutchison et al., 2009], pretreatment prediction of the likelihood of sustained virological response may be useful for both patients and physicians to support clinical decisions whether to start current standard therapy or to wait for emerging new therapies.

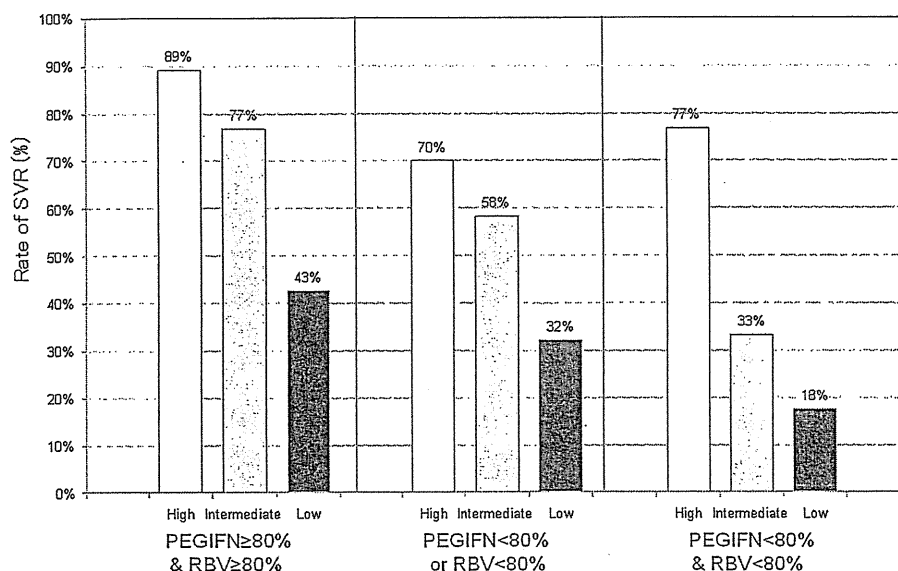


Fig. 4. Comparison of the rate of sustained virological response between the decision-tree groups stratified by drug adherence. The three groups of patients divided by the decision tree (black, gray, and white boxes indicating the low-, intermediate-, and high-probability groups, respectively) were further stratified according to cumulative drug exposure to PEG-IFN and RBV.

Two or more substitutions in ISDR had a strong impact on sustained virological response, because this factor was selected as a top variable in decision-tree analysis and had the highest odds ratio in multivariable analysis. Moreover, even among patients with unfavorable ISDR (0 or 1 mutation), younger patients (<60 years) with the wild-type sequence at Core70 and high level of LDL-C (≥ 120 mg/dl) had a high rate of sustained virological response. The sustained virological response rate of these two subgroups of patients was 83% in the model-building patients and 75% in the validation patients. Thus, patients with high possibility of sustained virological response could be extracted by the combined analysis of ISDR and Core70. These patients may be the best-suited candidates for treatment with the current combination therapy. Conversely, the following patients with 0–1 mutation in ISDR had a low probability of sustained virological response (32–35%): (1) older (>60 years); or (2) younger (<60 years) patients but having mutant-type sequence at Core70; or (3) younger (<60 years) patients having a wild-type sequence at Core70, but having a low level of LDL-C (<120 mg/dl) and advanced fibrosis. These patients may

be advised to wait for a more effective therapy. Decision may be made on a case-by-case basis, taking into account the potential risk of disease progression while waiting.

In a previous decision-tree model using simple and noninvasive standard tests that are available readily worldwide [Kurosaki et al., 2010b], the rate of sustained virological response was at most 65–76% among those in the high-probability group. That model focused on use by general physicians in routine general practice, especially where specialized resources, such as liver biopsy or determination of viral sequences, are not available. In that model, younger age, male sex, higher platelet counts, lower alpha-fetoprotein (AFP) levels, and lower GGT levels were identified as favorable predictive parameters. Higher AFP levels and lower platelet counts that are hallmarks of advanced fibrosis [Shiratori and Omata, 2000; Akuta et al., 2007b] were associated with low probability of sustained virological response in that model. On the other hand, the present analysis aimed to clarify the significance of viral factors for pretreatment prediction of sustained virological response, and to build an advanced model that may be used by specialist physicians engaged in the

TABLE II. Multivariable Logistic Regression Analysis for Factors Associated With SVR

Parameter		Odds	95% CI	P-value
Age (years)	<60 vs. ≥ 60	2.28	1.31–3.94	0.003
Sex	Male vs. female	3.36	1.87–5.99	<0.0001
GGT (IU/L)	<40 vs. ≥ 40	2.65	1.45–4.85	0.002
LDL-C (mg/dl)	≥ 120 vs. <120	1.79	0.91–3.53	0.094
Platelets (10 ⁹ /L)	≥ 120 vs. <120	2.69	1.22–5.90	0.014
ISDR mutations	≥ 2 vs. 0–1	9.92	3.71–26.54	<0.0001
Core70	Wild vs. mutant	1.92	1.07–3.47	0.030

GGT, gamma-glutamyltransferase; LDL-C, low-density-lipoprotein-cholesterol; ISDR, interferon sensitivity-determining region.

treatment of hepatitis. In the present model, stage of fibrosis was selected as a predictive factor, but at lower level of significance than HCV mutations. The predicted rate of sustained virological response in the high-probability group of the present model is higher than that in the previous model (75–83% vs. 65–76%). These results indicate that substitutions in ISDR and Core70 were important pretreatment predictors of sustained virological response. Determination of these viral factors is not available readily in clinical practice, but is of value for improving the accuracy of pretreatment prediction of sustained virological response.

Substitutions in ISDR and Core70 have been reported previously to be associated with efficacy of IFN therapy. The association between the number of substitutions in ISDR and response to therapy was demonstrated originally in patients treated with IFN mono-therapy [Enomoto et al., 1995, 1996; Kurosaki et al., 1997], but recent studies have reported a positive correlation with PEG-IFN and RBV combination therapy as well [Munoz de Rueda et al., 2008; Shirakawa et al., 2008; Ikeda et al., 2009]. Another important viral factor relevant to treatment response is amino acid substitution in Core70. The sequence of this amino acid was reported originally to be associated with nonresponse to therapy [Akuta et al., 2005], but subsequent studies confirmed the positive correlation of a wild-type Core70 with sustained virological response [Akuta et al., 2009]. The multiple logistic regression analysis showed that ISDR and Core70 were independent factors associated with sustained virological response along with host factors. How these important viral factors and other host factors can be combined to predict response to PEG-IFN plus RBV is an important clinical question. Decision-tree modeling can make the response probability apparent by combining all these factors. Some factors that may be associated with treatment outcome, such as levels of ferritin or homocysteine, were not included. This may be a potential limitation of the present study.

It is of interest that a recent study by Li et al. [2010] has shown that a high serum level of LDL-C is linked to the *IL28B* major allele (CC in rs12979860). In that study, a high serum level of LDL-C was associated with sustained virological response, but it was no longer significant when analyzed together with the *IL28B* genotype in multivariate analysis. Thus, the association between treatment response and LDL cholesterol levels in the present study may reflect the underlining link of LDL cholesterol levels to the *IL28B* genotype. Recent reports indicate that the *IL28B* genotype and HCV substitutions are correlated closely [Akuta et al., 2010; Kurosaki et al., 2010c]. Still, Core70 [Akuta et al., 2010] or ISDR [Kurosaki et al., 2010c] were predictors of response to therapy independent of *IL28B* genotype. Future study is needed to elucidate the possible mechanisms underlying the association between HCV sequences and host genetic factors, and also the role of host and viral factors for the prediction of treatment response.

In conclusion, a data mining analysis emphasized the impact of substitutions in ISDR and Core70 on pretreatment prediction of sustained virological response to PEG-IFN plus RBV therapy. A decision-tree model that includes substitutions in ISDR and Core70 of HCV could identify patients with high probability of sustained virological response, and could thereby improve the predictive accuracy over predictions that are based on standard tests.

REFERENCES

- Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, Someya T, Kobayashi M, Saitoh S, Watahiki S, Sato J, Matsuda M, Arase Y, Ikeda K, Kumada H. 2005. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. *Intervirology* 48:372–380.
- Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Arase Y, Ikeda K, Kumada H. 2007a. Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: Amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. *J Hepatol* 46:403–410.
- Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Arase Y, Ikeda K, Kumada H. 2007b. Predictors of viral kinetics to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b. *J Med Virol* 79:1686–1695.
- Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Kumada H. 2009. A matched case-controlled study of 48 and 72 weeks of peginterferon plus ribavirin combination therapy in patients infected with HCV genotype 1b in Japan: Amino acid substitutions in HCV core region as predictor of sustained virological response. *J Med Virol* 81:452–458.
- Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Chayama K, Nakamura Y, Kumada H. 2010. Amino acid substitution in hepatitis C virus core region and genetic variation near the interleukin 28B gene predict viral response to telaprevir with peginterferon and ribavirin. *Hepatology* 52:421–429.
- 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.
- Baquerizo A, Anselmo D, Shackleton C, Chen TW, Cao C, Weaver M, Gornbein J, Geevarghese S, Nissen N, Farmer D, Demetriou A, Busuttill RW. 2003. Phosphorus as an early predictive factor in patients with acute liver failure. *Transplantation* 75:2007–2014.
- Bedossa P, Poynard T. 1996. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 24:289–293.
- Breiman L, Friedman RA, Olshen CJ, Stone CM. 1980. Classification and regression trees. CA: Wadsworth.
- 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.
- Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Izumi N, Marumo F, Sato C. 1995. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J Clin Invest* 96:224–230.
- Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Ogura Y, Izumi N, Marumo F, Sato C. 1996. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med* 334:77–81.
- Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncalves FL, 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, Beer TM, Hudson RG, Peters L, Hsieh YC, Barrera E, Klein T, Mori M. 2005. Improved detection of prostate cancer using classification and regression tree analysis. *J Clin Oncol* 23:4322–4329.
- 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.
- Hezode C, Forestier N, Dusheiko G, Ferenci P, Pol S, Goeser T, Bronowicki JP, Bourliere M, Gharakhanian S, Bengtsson L, McNair L, George S, Kieffer T, Kwong A, Kauffman RS, Alam J, Pawlotsky JM, Zeuzem S. 2009. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N Engl J Med* 360:1839–1850.
- Ikeda H, Suzuki M, Okuse C, Yamada N, Okamoto M, Kobayashi M, Nagase Y, Takahashi H, Matsunaga K, Matsumoto N, Itoh F, Yotsuyanagi H, Koitabashi Y, Yasuda K, Iino S. 2009. Short-term prolongation of pegylated interferon and ribavirin therapy for genotype 1b chronic hepatitis C patients with early viral response. *Hepatol Res* 39:753–759.
- Izumi N, Nishiguchi S, Hino K, Suzuki F, Kumada Y, Itoh Y, Asahina Y, Tamori A, Hiramatsu N, Hayashi N, Kudo M. 2010. Management of hepatitis C; Report of the consensus meeting at the 45th annual meeting of the Japan society of hepatology (2009). *Hepatol Res* 40:347–368.
- Jensen DM, Morgan TR, Marcellin P, Pockros PJ, Reddy KR, Hadziyannis SJ, Ferenci P, Ackrill AM, Willems B. 2006. Early identification of HCV genotype 1 patients responding to 24 weeks peginterferon alpha-2a (40 kd)/ribavirin therapy. *Hepatology* 43:954–960.
- Kurosaki M, Enomoto N, Murakami T, Sakuma I, Asahina Y, Yamamoto C, Ikeda T, Tozuka S, Izumi N, Marumo F, Sato C, Ogura Y. 1997. Analysis of genotypes and amino acid residues 2209 to 2248 of the NS5A region of hepatitis C virus in relation to the response to interferon-beta therapy. *Hepatology* 25:750–753.
- 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, Higak M, Enomoto N, Izumi N. 2010a. 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, Yatsuhashi H, Izumi N. 2010b. Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis. *J Gastroenterol* DOI: 10.1007/s00535-010-0322-5.
- 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. 2010c. Pretreatment prediction of response to pegylated-interferon plus ribavirin for chronic hepatitis C using genetic polymorphism in IL28B and viral factors. *J Hepatol* DOI: 10.1016/j.jhep.2010.07.037.
- LeBlanc M, Crowley J. 1995. A review of tree-based prognostic models. *Cancer Treat Res* 75:113–124.
- Lee SS, Ferenci P. 2008. Optimizing outcomes in patients with hepatitis C virus genotype 1 or 4. *Antivir Ther* 13:9–16.
- 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.
- Li JH, Lao XQ, Tillmann HL, Rowell J, Patel K, Thompson A, Suchindran S, Muir AJ, Guyton JR, Gardner SD, McHutchison JG, McCarthy JJ. 2010. Interferon-lambda genotype and low serum low-density lipoprotein cholesterol levels in patients with chronic hepatitis C infection. *Hepatology* 51:1904–1911.
- 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, 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.
- Munoz de Rueda P, Casado J, Paton R, Quintero D, Palacios A, Gila A, Quiles R, Leon J, Ruiz-Extremera A, Salmeron J. 2008. Mutations in E2-PePHD, NS5A-PKRBD, NS5A-ISDR, and NS5A-V3 of hepatitis C virus genotype 1 and their relationships to pegylated interferon-ribavirin treatment responses. *J Virol* 82:6644–6653.
- Segal MR, Bloch DA. 1989. A comparison of estimated proportional hazards models and regression trees. *Stat Med* 8:539–550.
- Shirakawa H, Matsumoto A, Joshita S, Komatsu M, Tanaka N, Umemura T, Ichijo T, Yoshizawa K, Kiyosawa K, Tanaka E. 2008. Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors. *Hepatology* 48:1753–1760.
- Shiratori Y, Omata M. 2000. Predictors of the efficacy of interferon therapy for patients with chronic hepatitis C before and during therapy: How does this modify the treatment course? *J Gastroenterol Hepatol* 15:E141–E151.
- Supiah 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.
- 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.
- Yu ML, Dai CY, Huang JF, Chiu CF, Yang YH, Hou NJ, Lee LP, Hsieh MY, Lin ZY, Chen SC, Wang LY, Chang WY, Chuang WL. 2008. Rapid virological response and treatment duration for chronic hepatitis C genotype 1 patients: A randomized trial. *Hepatology* 47:1884–1893.
- 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.

Review Article

Predictors of Virological Response to a Combination Therapy with Pegylated Interferon Plus Ribavirin Including Virus and Host Factors

Namiki Izumi, Yasuhiro Asahina, and Masayuki Kurosaki

Department of Gastroenterology and Hepatology, Musashino Red-Cross Hospital, Kyonancho 1-26-1, Musashinoshi, 180-8610 Tokyo, Japan

Correspondence should be addressed to Namiki Izumi, nizumi@musashino.jrc.or.jp

Received 25 April 2010; Revised 29 June 2010; Accepted 19 July 2010

Academic Editor: Ming-Lung Yu

Copyright © 2010 Namiki Izumi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A combination therapy with pegylated interferon (PEG-IFN) plus ribavirin (RBV) has made it possible to achieve a sustained virological response (SVR) of 50% in refractory cases with genotype 1b and high levels of plasma HCVRNA. Several factors including virus mutation and host factors such as age, gender, fibrosis of the liver, lipid metabolism, innate immunity, and single nucleotide polymorphism (SNPs) are reported to be correlated to therapeutic effects. However, it is difficult to determine which factor is the most important predictor for an individual patient. Data mining analysis is useful for combining all these together to predict the therapeutic effects. It is important to analyze blood tests and to predict therapeutic effects prior to initiating treatment. Since new anti-HCV agents are under development, it will be necessary in the future to select the patients who have a high possibility of achieving SVR if treatment is performed with standard regimen.

1. Progress in Virological Response in the Difficult-to-Treat Patients with Genotype 1 Hepatitis C Virus (HCV) Infection and Factors Correlated to the Efficacy

Recently, the average age of the patients with chronic hepatitis C has been increasing in Japan. Incidence of hepatocellular carcinoma (HCC) in the elderly patients with chronic hepatitis C (65 years or older) has demonstrated to be higher than younger ones when adjusted by the stage of hepatic fibrosis [1]. In Japan, refractory cases with genotype 1b and high HCVRNA levels are seen in as high as 70 percent of chronic hepatitis C patients. The outcome of conventional IFN monotherapy has been an SVR response of 3%–5% after 6 months of treatment in genotype 1b and high HCVRNA patients [2, 3], and virus mutation such as interferon sensitivity-determining region (ISDR) is shown to be correlated with the virological response [2]. The association of ISDR mutations and virological response to IFN monotherapy was denied in an Italian study [4];

however, it was confirmed by a Chinese study [5] and an international meta-analysis [6].

However, pegylated IFN (PEG-IFN) extends the duration of therapy and reduces adverse effects, and for this reason, PEG-IFN has become the cornerstone of therapy. Furthermore, by the combination therapy with PEG-IFN and ribavirin (RBV), the rate of SVR has dramatically improved. Even in the patients with genotype 1b and high HCVRNA level, SVR rate reaches as high as 40%–50%, thereby improving the therapeutic effects both in Western countries [7, 8] and in Japan [9, 10].

It is important to predict the rate of achieving SVR in the individual patient, before initiating treatment. Both virus- and host-related elements have been reported as factors correlated to therapeutic effects of combination therapy [11–13]. A particular focus has been placed on virus mutations, age, gender, fibrosis of the liver, lipid metabolism, and degree of fatty metamorphosis of the liver.

Among these factors related to PEG-IFN and RBV, innate immunity has been shown to be correlated in virological

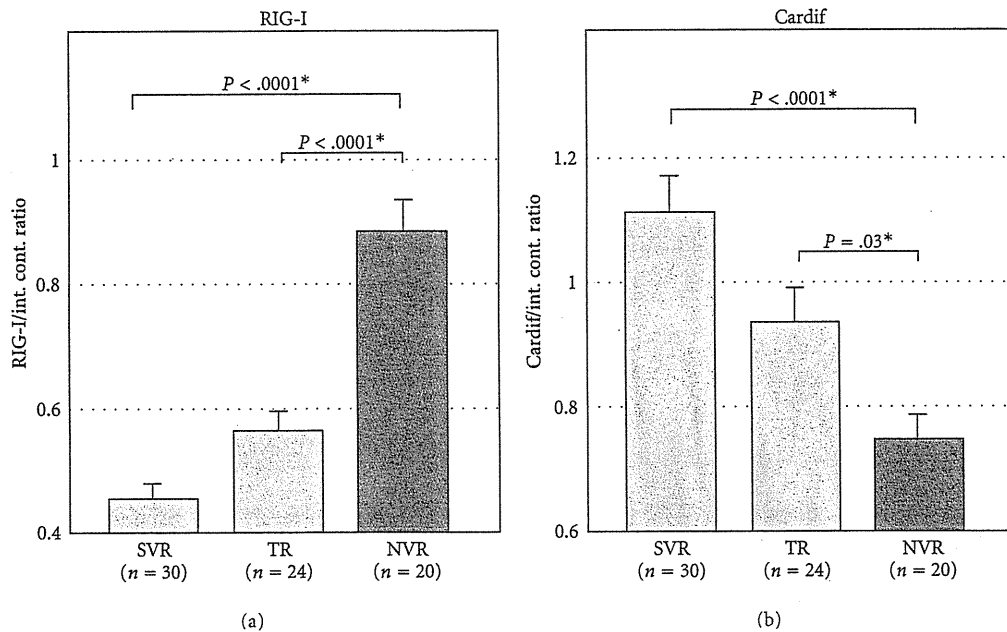


FIGURE 1: Expression of genes correlated to the intrahepatic innate immunity and virological response to PEG-IFN alpha-2b and RBV combination therapy. Open column indicates SVR ($n = 30$), gray column indicates TR ($n = 24$), and closed column indicates NVR ($n = 20$). Error bars indicate the standard error. The P values were analyzed by the Kruskal-Wallis test. Expression of Rig-I was significantly higher in NVR than in SVR patients, and Cardif expression was higher in SVR than in NVR. The figure was cited from [8].

response. Asahina et al. reported that liver biopsies were performed before the PEG-IFN and RBV combination therapy to examine the correlation between the gene expression involved in innate immunity and the therapeutic effects, and in the patients in whom RIG-I expression is high and the expression of Cardif, an adaptor gene, is low, it was found that there are many nonresponders (NVRs) in which HCVRNA does not become negative during the course of treatment [13]. It was therefore revealed that there are many NVRs among the patients in whom the ratio of RIG-I to Cardif in liver tissue is high and that this ratio is low in the SVR patients. Based on these findings, it has become clear that innate immunity is correlated to therapeutic effects (Figure 1).

Furthermore, it was recently discovered that a single nucleotide polymorphism (SNP) of the host gene IL28B is significantly involved in the therapeutic response to the PEG-IFN and RBV combination therapy [14, 15]. The possibility of becoming an NVR is high in cases of the minor allele carriers of IL28B. However, it is not possible to routinely measure an SNP of IL28B in the clinical setting. In this paper, factors which can actually be used in real clinical practice are discussed for the prediction of the efficacy of PEG-IFN and RBV combination therapy.

2. Amount of HCVRNA

In the patients with chronic hepatitis C, it is not possible to directly measure the amount of virus, and the

amount of HCVRNA is measured instead. Currently, a real-time PCR method which has an advantage of wide range and high sensitivity is utilized, and measurements can be taken from a single blood sample of a very small amount, that is, 1.2 log copies/ml, to a very large amount, that is, 8 log copies/ml. This method has a higher level of sensitivity than the conventional Amplicor monitor test and can therefore detect HCVRNA even if only a very small amount exists in the plasma. If the amount of HCVRNA in plasma is less than the range of sensitivity of the real-time PCR method, it is recorded as undetectable level. If the indication is "less than 1.2 log copies/ml of HCVRNA", it means that a very small amount of HCVRNA exists in the plasma. Since the indication of the real-time PCR method is based on log counts, a decrease of 1.0 in the numerical value means that the amount of HCVRNA has decreased to 1/10. With the application of this real-time PCR method, it has become possible to measure amounts of HCVRNA up to 8 log copies/ml, and it has also become possible to predict the efficacy before treatment and to monitor appropriately the reactivity during the course of treatment. However, in the patients in whom a PEG-IFN and RBV combination therapy is performed, SVR can be acquired even when the amount of virus prior to the treatment is quite large. It is therefore difficult to predict the virological response solely from the amount of HCVRNA before starting the treatment. Once treatment has commenced, at what level HCVRNA becomes negative is important for the prediction of therapeutic effects, and this serves as a parameter for deciding the duration of treatment [16].

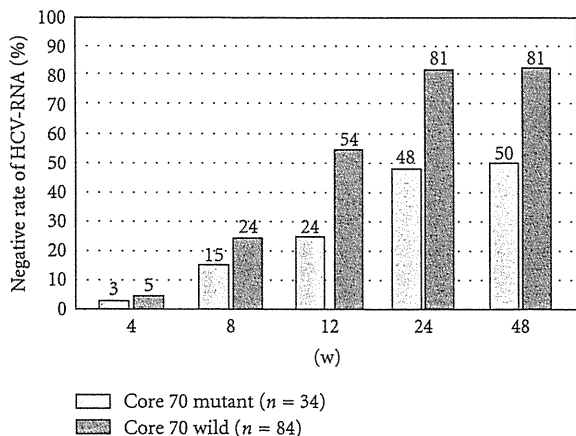


FIGURE 2: Comparison of aa70 mutations in the HCV core region and the rate of HCVRNA becoming negative during the course of treatment. Compared with the wild type, among the patients of aa70 mutations, there were fewer patients in whom HCVRNA had become negative during the course of treatment.

Measuring the rate of viral clearance from serum is helpful for predicting the likelihood of a response to PEGIFN and RBV and useful for determining the optimal duration of therapy if the patients start to receive the treatment [17]. In the AASLD practice guideline, response-guided therapy is highly recommended [18]. In two nationwide registration trials conducted in Japan [9, 10], the SVR rate was high, from 76% to 100%, in patients whose HCVRNA was cleared rapidly from serum by week 4 (rapid virological response; RVR), and 71% to 73% in patients who achieved undetectable HCVRNA from week 5 to week 12 (early virological response; EVR). In contrast, the SVR rate in patients with late clearance of HCVRNA from week 13 to week 24 was low at 29% to 36%. No patients without clearance of HCVRNA by week 24 achieved SVR.

The strategy of extending therapy in patients with delayed virological responses, defined as clearance of HCV-RNA between weeks 12 and 24, was evaluated in five studies [19–23]. These results cannot be compared directly with each other because of the heterogeneous study populations, differences in the baseline characteristics, and the different regimens utilized amongst them. Nevertheless, the results showed a trend toward a higher SVR rate by extending therapy from 48 to 72 weeks in patients with delayed virological response.

3. Viral Mutations in Core and NS5A Region

In the patients with genotype 1b HCV infection, the mutations in aa70 and aa91 in the core region have been shown to correlate with early virological response (EVR) during PEG-IFN and RBV combination treatment [11]. If aa70 in the core region is mutated to anything other than arginine and aa91 to anything other than leucine, it is difficult to achieve EVR, and it is thus highly possible that such cases

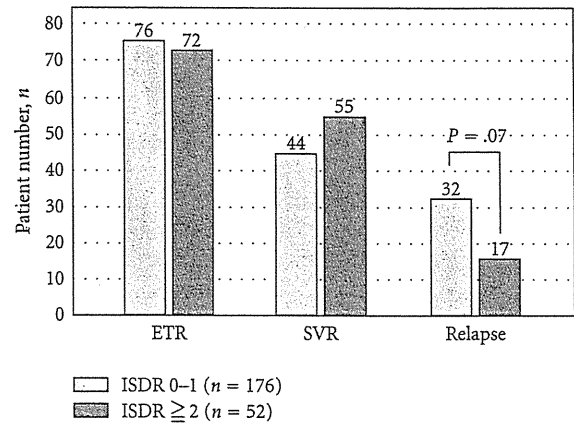


FIGURE 3: Number of ISDR substitutions and the comparison of virological response, SVR, and relapse at the end of the treatment. Compared with the patients with 2 or more sites of substitutions, the rate of SVR was lower and the rate of relapse was higher in the patients in whom there were fewer substitutions in ISDR, that is, 0 or 1 sites.

will become nonresponders. The examination results at our institution including 292 patients with genotype 1b infection demonstrated that, in the cases with mutations in aa70 in the core region, the rate of HCVRNA becoming negative during the course of combination treatment was low compared to the wild type of aa70 (Figure 2). However, core aa70 mutation is shown to have quasispecies detectable by cloning, and 70Q clone was positively selected during combination treatment with PEGIFN and RBV [24].

Furthermore, Enomoto et al. reported that the patients with 4 or more amino acid mutations were observed in interferon sensitivity-determining region (ISDR) within NS5A region [2]; SVR rate is higher than 90% by IFN monotherapy, and SVR is less than 10% in the patients with no mutation in ISDR. It has also been reported that, in PEG-IFN and RBV combination therapy, the number of ISDR mutations is involved in the SVR [12].

We analyzed the relationship between virological response and ISDR mutations in the patients with genotype 1b infection treated by PEG-IFN alpha-2b and RBV combination therapy. In the patients with 0 or 1 ISDR mutation, even if the rate of HCVRNA becoming negative at the end of treatment was the same, the rate of SVR would be lower compared with the patients having 2 or more mutations (Figure 3). This demonstrates that there is a higher incidence of relapse after the end of treatment in the patients with 0 or 1 ISDR mutation.

Enomoto and Maekawa reported that mutations both in NS5A-ISDR (interferon sensitivity-determining region) and core 70Q substitution are associated with no early viral response during PEGIFN and RBV combination therapy [25]. Association of core aa70 substitution and mutations in NS5A region is confirmed to be associated with viral response by PEGIFN and RBV combination therapy in a Japanese multicenter cooperative study [26]. The number of

mutations in the interferon sensitivity-determining region was shown to be associated with the viral response to PEGIFN and RBV combination treatment not only in Japan [27], but also in Tunisia [28].

Recently, a consensus has been established that mutations in aa70 in the core region are important for the prediction of HCVRNA becoming negative during the early course of treatment, and the number of ISDR mutations is important for the prediction of relapse after the end of treatment.

4. Adherence

It has been confirmed that it is important to ensure 80% or more of the scheduled dose of both PEG-IFN and RBV in order to improve the rate of SVR, and together with the duration of treatment, the 80 · 80 · 80 rule has been established. However, Schiffman et al. recently reported that the dose of PEG-IFN in the initial stage of administration is important and that, if a sufficient dose of PEG-IFN is administered, then 60% or more of the RBV dose would be enough [29]. It is therefore of primary importance to ensure the dose of PEG-IFN.

In Japan, the average age of patients with chronic hepatitis C is increasing, and achieving good adherence is difficult in many patients. Consequently, the rate of SVR is low in elderly patients. How to improve the rate of SVR in elderly patients is an important issue. With regard to the dose of RBV, reducing the RBV dose based on the calculation of the total body clearance (CL/F) has been proposed to be useful for decreasing the discontinuation and improving the rate of SVR. Although there is no consensus on an appropriate dose of PEG-IFN in elderly patients, if the initial dose is set lower than the usual dose, discontinuation would be reduced. Thus, it is necessary to investigate whether such an initial dose would improve the rate of SVR.

Recently, the risk of hemolytic anemia was clearly demonstrated to correlate with ITAP gene SNP [30]. The predictive implication should be analyzed prospectively in clinical practice.

5. Host Factors

Zeuzem et al. described the factors related to the less response to interferon-based therapy, and he showed that several host factors such as older age, race, and obesity are responsible factors for the poor response to IFN [31]. Recently, insulin resistance which was examined by homeostasis model assessment index (HOMA-IR) was shown to be associated with a lower rate of SVR, and body mass index (BMI) was not identified as a significant factor for the poor response to PEGIFN and ribavirin combination therapy [32]. Insulin resistance was confirmed as a related factor to the nonresponse to interferon-based treatment [33]. However, Charlton et al. reported that obesity itself is an associated factor for decreased efficacy of interferon-based therapies, and they discussed the possible mechanism [34], and obesity was shown to be associated with the increased enhancement

of suppressor of cytokine signaling (SOCS) family in the hepatocytes [35].

6. Data Mining Analysis

Both virus- and host-related factors are correlated to therapeutic effects of PEG-IFN and RBV. One important question is which of these factors should be focused on in order to predict the therapeutic effects in an individual patient. In addition, in each individual patient, the host and virological factors are different. It is therefore difficult to predict the virological response in each case before treatment. Furthermore, although it is important to predict the relapse rate when HCVRNA becomes within an undetectable level in an individual patient, prediction of the rate of SVR including virological and host factors and adherence to the treatment has never been carried out in an individual patient.

A data mining analysis is the process of analyzing a large amount of data by a computer in order to develop an algorithm. Conventional statistics have been used to examine certain hypothesis. Data mining is superior in that it can set an algorithm, using a computer, based on a large amount of data without a hypothesis.

We therefore conducted at our institute a data mining analysis of the patients with genotype 1b infection having high levels of HCVRNA to whom a PEG-IFN alpha-2b and RBV combination therapy was administered to investigate whether by the 12th week after the commencement of treatment HCVRNA became negative (EVR) (Figure 4) [36]. The most important factor for the prediction of EVR was the steatosis of the liver: when steatosis was observed in 30% or more of hepatocytes, EVR was found to be difficult to achieve. In the patients in whom steatosis was not severe, the second most important factor was the serum LDL cholesterol value. While the rate of EVR was 57% in the patients in whom this value was 100 mg/ml or above, the rate of EVR was 32% in the patients in whom the LDL cholesterol was less than 100 mg/dl.

The higher the LDL cholesterol value, the earlier the HCVRNA became negative. Among the patients with low LDL cholesterol values, while the rate of EVR was 15% in patients 60 years of age or older, the rate was high in the patients under the age of 60 years old, that is, 49%. Among patients under the age of 60, the rate of EVR was low, that is, 31%, in patients with a blood glucose level of 120 mg/dl or above whereas EVR was achieved in 71% of the patients with a blood glucose level of less than 120 mg/dl (Figure 4).

On the other hand, in the patients with high LDL cholesterol values, the next most important factor was age. While the rate of EVR was 50% in patients 50 years of age or older, EVR was achieved in 77% of the patients under the age of 50. Among patients of 50 years of age or older, the next most important factor was the gamma GTP value. While the rate of EVR was 35% in the patients in whom gamma GTP was 40 IU/L or above, EVR was achieved in 60% of the patients where the value was less than 40 IU/L.

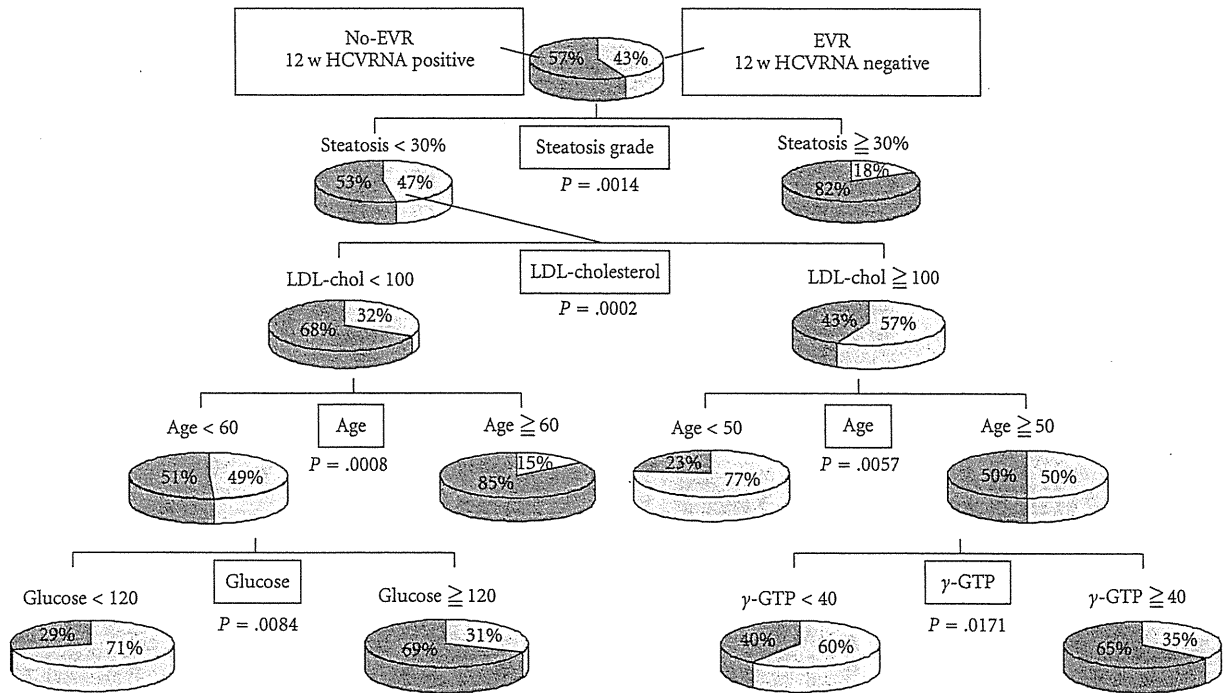


FIGURE 4: HCV RNA negative (EVR) algorithm at 12th week from data mining analysis of PEG-IFN alpha-2b plus RBV combination in the genotype 1b and high levels of HCV RNA. Both virological and host factors were evaluated by data mining analysis software from SPSS. This figure was cited from [36].

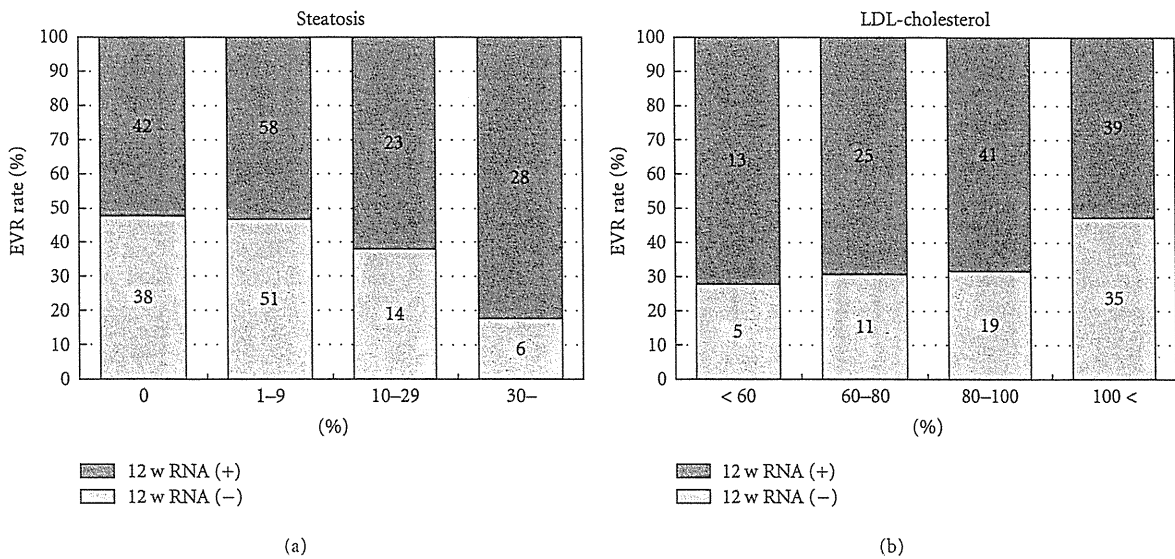


FIGURE 5: The rate of EVR in the patients with genotype 1b and high levels of HCV RNA, based on fatty deposition in the liver (a), and the LDL cholesterol value (b), respectively. EVR was highly achieved in the patients with less steatosis in the liver, and in those with high serum LDL-cholesterol levels. This is univariate analysis, and cited from [36].

We therefore compared these factors based on the original data. A univariate comparison of the fatty infiltration of the liver and the rate of EVR demonstrated that the rate of EVR was higher when the steatosis of the liver was less severe (Figure 5(a)). In addition, a comparison of the LDL cholesterol value and the rate of EVR demonstrated a significant correlation, confirming that the higher the LDL cholesterol value, the higher the rate of EVR (Figure 5(b)). Therefore, it was also proposed by the results of univariate analysis of each factor extracted from the data mining analysis that these factors were correlated to the rate of EVR.

From these observations, it is likely to improve the viral response to PEGIFN and ribavirin by reducing steatosis of the liver through daily walking or abstaining alcohol intake or by refraining from high-fat diet.

By utilizing data mining, it is therefore possible to assess virus- and host-related factors together and to predict the virological response in each patient, and thereby clinically useful information can be obtained. The algorithm should be validated including a large number of the patients.

Acknowledgment

This paper was supported by a grant from the Japanese Ministry of Health, Welfare and Labor.

References

- [1] Y. Asahina, K. Tsuchiya, I. Hirayama, et al., "Effect of aging on risk for hepatocellular carcinoma in chronic hepatitis C virus infection," *Hepatology*, vol. 52, no. 2, pp. 518–527, 2010.
- [2] N. Enomoto, I. Sakuma, Y. Asahina et al., "Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection," *New England Journal of Medicine*, vol. 334, no. 2, pp. 77–81, 1996.
- [3] S. Iino, F. Ichida, A. Sakuma et al., "A randomized clinical trial with natural interferon- α monotherapy for 24 or 48 weeks on patients with chronic hepatitis C having genotype 1b infection in high viral titers," *Hepatology Research*, vol. 24, no. 4, pp. 338–345, 2002.
- [4] G. Squadrito, M. E. Orlando, I. Cacciola et al., "Long-term response to interferon alpha is unrelated to "interferon sensitivity determining region" variability in patients with chronic hepatitis C virus-1b infection," *Journal of Hepatology*, vol. 30, no. 6, pp. 1023–1027, 1999.
- [5] C. Shen, T. Hu, L. Shen, L. Gao, W. Xie, and J. Zhang, "Mutations in ISDR of NS5A gene influence interferon efficacy in Chinese patients with chronic hepatitis C virus genotype 1b infection," *Journal of Gastroenterology and Hepatology*, vol. 22, no. 11, pp. 1898–1903, 2007.
- [6] M. Pascu, P. Martus, M. Höhne et al., "Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences," *Gut*, vol. 53, no. 9, pp. 1345–1351, 2004.
- [7] S. J. Hadziyannis, H. Sette Jr., T. R. Morgan et al., "Peginterferon- α 2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose," *Annals of Internal Medicine*, vol. 140, no. 5, pp. 346–355, 2004.
- [8] M. P. Manns, J. G. McHutchison, S. C. Gordon et al., "Peginterferon alfa-2b plus ribavirin compared with interferon- α 2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial," *Lancet*, vol. 358, no. 9286, pp. 958–965, 2001.
- [9] S. Iino, E. Tomita, H. Kumada et al., "Prediction of treatment outcome with daily high-dose IFN α -2b plus ribavirin in patients with chronic hepatitis C with genotype 1b and high HCV RNA levels: relationship of baseline viral levels and viral dynamics during and after therapy," *Hepatology Research*, vol. 30, no. 2, pp. 63–70, 2004.
- [10] G. Yamada, S. Iino, T. Okuno et al., "Virological response in patients with hepatitis C virus genotype 1b and a high viral load: impact of peginterferon- α -2a plus ribavirin dose reductions and host-related factors," *Clinical Drug Investigation*, vol. 28, no. 1, pp. 9–16, 2008.
- [11] N. Akuta, F. Suzuki, Y. Kawamura et al., "Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: amino acid substitutions in the core region and low-density lipoprotein cholesterol levels," *Journal of Hepatology*, vol. 46, no. 3, pp. 403–410, 2007.
- [12] H. Shirakawa, A. Matsumoto, S. Joshita et al., "Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors," *Hepatology*, vol. 48, no. 6, pp. 1753–1760, 2008.
- [13] Y. Asahina, N. Izumi, I. Hirayama et al., "Potential relevance of cytoplasmic viral sensors and related regulators involving innate immunity in antiviral response," *Gastroenterology*, vol. 134, no. 5, pp. 1396–1405, 2008.
- [14] D. Ge, J. Fellay, A. J. Thompson et al., "Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance," *Nature*, vol. 461, no. 7262, pp. 399–401, 2009.
- [15] Y. Tanaka, N. Nishida, M. Sugiyama et al., "Genome-wide association of IL28B with response to pegylated interferon- α and ribavirin therapy for chronic hepatitis C," *Nature Genetics*, vol. 41, no. 10, pp. 1105–1109, 2009.
- [16] P. Ferenci, H. Laferl, T. Scherzer et al., "Peginterferon alfa-2a and ribavirin for 24 weeks in hepatitis C type 1 and 4 patients with rapid virological response," *Gastroenterology*, vol. 135, no. 2, pp. 451–458, 2008.
- [17] N. Izumi, S. Nishiguchi, K. Hino et al., "Management of hepatitis C; Report of the Consensus Meeting at the 45th Annual Meeting of the Japan Society of Hepatology (2009)," *Hepatology Research*, vol. 40, no. 4, pp. 347–368, 2010.
- [18] M. G. Ghany, D. B. Strader, D. L. Thomas, and L. B. Seeff, "Diagnosis, management, and treatment of hepatitis C: an update," *Hepatology*, vol. 49, no. 4, pp. 1335–1374, 2009.
- [19] T. Berg, M. von Wagner, S. Nasser et al., "Extended treatment duration for hepatitis C virus type 1: comparing 48 versus 72 weeks of peginterferon- α -2a plus ribavirin," *Gastroenterology*, vol. 130, no. 4, pp. 1086–1097, 2006.
- [20] J. M. Sánchez-Tapias, M. Diago, P. Escartín et al., "Peginterferon- α 2a plus ribavirin for 48 versus 72 weeks in patients with detectable hepatitis C virus RNA at week 4 of treatment," *Gastroenterology*, vol. 131, no. 2, pp. 451–460, 2006.
- [21] P. Ferenci, H. Laferl, T. M. Scherzer, et al., "Customizing treatment with peginterferon alfa-2a (40kD)(PEGASYS[®]) plus ribavirin (COPEGUS[®]) in patient with HCV genotype 1 or 4 infection: interim results of a prospective randomized trial," *Hepatology*, vol. 44, no. 336a, 2006.
- [22] B. L. Pearlman, C. Ehleben, and S. Saifee, "Treatment extension to 72 weeks of peginterferon and ribavirin in hepatitis C

- genotype 1-infected slow responders," *Hepatology*, vol. 46, no. 6, pp. 1688–1694, 2007.
- [23] M. Buti, Y. Lurie, N. G. Zakharova, et al., "Extended treatment duration in chronic hepatitis C genotype 1-infected slow responders: final results of the SUCCESS study," *Journal of Hepatology*, vol. 50, supplement 1, p. S58, abstract 141, 2009.
- [24] F. Kurbanov, Y. Tanaka, K. Matsuura et al., "Positive selection of core 70Q variant genotype 1b hepatitis C virus strains induced by pegylated interferon and ribavirin," *Journal of Infectious Diseases*, vol. 201, no. 11, pp. 1663–1671, 2010.
- [25] N. Enomoto and S. Maekawa, "HCV genetic elements determining the early response to peginterferon and ribavirin therapy," *Intervirology*, vol. 53, no. 1, pp. 66–69, 2010.
- [26] T. Okanoue, Y. Itoh, H. Hashimoto et al., "Predictive values of amino acid sequences of the core and NS5A regions in antiviral therapy for hepatitis C: a Japanese multi-center study," *Journal of Gastroenterology*, vol. 44, no. 9, pp. 952–963, 2009.
- [27] M. Nakagawa, N. Sakamoto, M. Ueyama et al., "Mutations in the interferon sensitivity determining region and virological response to combination therapy with pegylated-interferon alpha 2b plus ribavirin in patients with chronic hepatitis C-1b infection," *Journal of Gastroenterology*, vol. 45, no. 6, pp. 656–665, 2010.
- [28] N. Bouzgarrou, E. Hassen, W. Mahfoudh et al., "NS5AISDR-V3 region genetic variability of Tunisian HCV-1b strains: correlation with the response to the combined interferon/ribavirin therapy," *Journal of Medical Virology*, vol. 81, no. 12, pp. 2021–2028, 2009.
- [29] M. L. Shiffman, M. G. Ghany, T. R. Morgan et al., "Impact of reducing peginterferon alfa-2a and ribavirin dose during retreatment in patients with chronic hepatitis C," *Gastroenterology*, vol. 132, no. 1, pp. 103–112, 2007.
- [30] J. Fellay, A. J. Thompson, D. Ge et al., "TPA gene variants protect against anaemia in patients treated for chronic hepatitis C," *Nature*, vol. 464, no. 7287, pp. 405–408, 2010.
- [31] S. Zeuzem, "Heterogeneous virologic response rates to interferon-based therapy in patients with chronic hepatitis C: who responds less well?" *Annals of Internal Medicine*, vol. 140, no. 5, pp. 370–381, 2004.
- [32] H. S. Conjeevaram, D. E. Kleiner, J. E. Everhart et al., "Race, insulin resistance and hepatic steatosis in chronic hepatitis C," *Hepatology*, vol. 45, no. 1, pp. 80–87, 2007.
- [33] G. Tarantino, P. Conca, P. Sorrentino, and M. Ariello, "Metabolic factors involved in the therapeutic response of patients with hepatitis C virus-related chronic hepatitis," *Journal of Gastroenterology and Hepatology*, vol. 21, no. 8, pp. 1266–1268, 2006.
- [34] M. R. Charlton, P. J. Pockros, and S. A. Harrison, "Impact of obesity on treatment of chronic hepatitis C," *Hepatology*, vol. 43, no. 6, pp. 1177–1186, 2006.
- [35] M. J. Walsh, J. R. Jonsson, M. M. Richardson et al., "Non-response to antiviral therapy is associated with obesity and increased hepatic expression of suppressor of cytokine signaling 3 (SOCS-3) in patients with chronic hepatitis C, viral genotype 1," *Gut*, vol. 55, no. 4, pp. 529–535, 2006.
- [36] M. Kurosaki, K. Matsunaga, I. Hirayama, et al., "A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis," *Hepatology Research*, vol. 40, no. 3, pp. 251–260, 2010.

Pre-treatment prediction of response to pegylated-interferon plus ribavirin for chronic hepatitis C using genetic polymorphism in *IL28B* and viral factors

Masayuki Kurosaki¹, Yasuhito Tanaka², Nao Nishida³, Naoya Sakamoto⁴, Nobuyuki Enomoto⁵, Masao Honda⁶, Masaya Sugiyama², Kentaro Matsuura², Fuminaka Sugauchi², Yasuhiro Asahina¹, Mina Nakagawa⁴, Mamoru Watanabe⁴, Minoru Sakamoto⁵, Shinya Maekawa⁵, Akito Sakai⁶, Shuichi Kaneko⁶, Kiyooki Ito⁷, Naohiko Masaki⁷, Katsushi Tokunaga³, Namiki Izumi^{1,*}, Masashi Mizokami^{2,7}

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

Background & Aims: Pegylated interferon and ribavirin (PEG-IFN/RBV) therapy for chronic hepatitis C virus (HCV) genotype 1 infection is effective in 50% of patients. Recent studies revealed an association between the *IL28B* genotype and treatment response. We aimed to develop a model for the pre-treatment prediction of response using host and viral factors.

Methods: Data were collected from 496 patients with HCV genotype 1 treated with PEG-IFN/RBV at five hospitals and universities in Japan. *IL28B* genotype and mutations in the core and IFN sensitivity determining region (ISDR) of HCV were analyzed to predict response to therapy. The decision model was generated by data mining analysis.

Results: The *IL28B* polymorphism correlated with early virological response and predicted null virological response (NVR) (odds ratio = 20.83, $p < 0.0001$) and sustained virological response (SVR) (odds ratio = 7.41, $p < 0.0001$) independent of other covariates. Mutations in the ISDR predicted relapse and SVR independent of *IL28B*. The decision model revealed that patients with the minor *IL28B* allele and low platelet counts had the highest NVR (84%) and lowest SVR (7%), whereas those with the major *IL28B* allele and mutations in the ISDR or high platelet counts had the lowest NVR (0–17%) and highest SVR (61–90%). The model had high reproducibility and predicted SVR with 78% specificity and 70% sensitivity.

Conclusions: The *IL28B* polymorphism and mutations in the ISDR of HCV were significant pre-treatment predictors of response to PEG-IFN/RBV. The decision model, including these host and viral factors may support selection of optimum treatment strategy for individual patients.

© 2010 European Association for the Study of the Liver. Published by Elsevier B.V. All rights reserved.

Introduction

Hepatitis C virus (HCV) infection is the leading cause of cirrhosis and hepatocellular carcinoma worldwide [1]. The successful eradication of HCV, defined as a sustained virological response (SVR), is associated with a reduced risk of developing hepatocellular carcinoma. Currently, pegylated interferon (PEG-IFN) plus ribavirin (RBV) is the most effective standard of care for chronic hepatitis C but the rate of SVR is around 50% in patients with HCV genotype 1 [2,3], the most common genotype in Japan, Europe, the United States, and many other countries. Moreover, 20–30% of patients with HCV genotype 1 have a null virological response (NVR) to PEG-IFN/RBV therapy [4]. The most reliable method for predicting the response is to monitor the early decline of serum HCV-RNA levels during treatment [5] but there is no established method for prediction before treatment. Because PEG-IFN/RBV therapy is costly and often accompanied by adverse effects such as flu-like symptoms, depression and hematological abnormalities, pre-treatment predictions of those patients who are unlikely to benefit from this regimen enables ineffective treatment to be avoided.

Recently, it has been reported through a genome-wide association study (GWAS) of patients with genotype 1 HCV that single nucleotide polymorphisms (SNPs) located near the *IL28B* gene are strongly associated with a response to PEG-IFN/RBV therapy in

Keywords: *IL28B*; ISDR; Peg-interferon; Ribavirin; Data mining; Decision tree.
Received 14 March 2010; received in revised form 22 June 2010; accepted 7 July 2010;
available online 19 September 2010

* Corresponding author. Address: Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, 1-26-1 Kyonan-cho, Musashino-shi, Tokyo 180-8610, Japan. Tel.: +81 422 32 3111; fax: +81 422 32 9551.

E-mail address: nizumi@musashino.jrc.or.jp (N. Izumi).



Research Article

Table 1. Baseline characteristics of all patients, and patients assigned to the model building or validation groups.

	All patients n = 496	Model group n = 331	Validation group n = 165
Gender: male	250 (50%)	170 (51%)	80 (48%)
Age (years)	57.1 ± 9.9	56.8 ± 9.7	57.5 ± 10.2
ALT (IU/L)	78.6 ± 60.8	78.1 ± 61.4	79.7 ± 59.6
GGT (IU/L)	59.3 ± 63.6	58.9 ± 62.0	60.2 ± 66.9
Platelets (10 ⁹ /L)	154 ± 53	153 ± 52	154 ± 56
Fibrosis: F3-4	121 (24%)	80 (24%)	41 (25%)
HCV-RNA: >600,000 IU/ml	409 (82%)	273 (82%)	136 (82%)
ISDR mutation: ≤1	220 (88%)	290 (88%)	145 (88%)
Core 70 (Arg/Gln or His)	293 (59%)/203 (41%)	197 (60%)/134 (40%)	96 (58%)/69 (42%)
Core 91 (Leu/Met)	299 (60%)/197 (40%)	200 (60%)/131 (40%)	99 (60%)/66 (40%)
<i>IL28B</i> : Minor allele	151 (30%)	101 (31%)	50 (30%)
SVR	194 (39%)	129 (39%)	65 (39%)
Relapse	152 (31%)	103 (31%)	49 (30%)
NVR	150 (30%)	99 (30%)	51 (31%)

ALT, alanine aminotransferase; GGT, gamma-glutamyltransferase; ISDR, interferon sensitivity determining region; Arg, arginine; Gln, glutamine; His, histidine; Leu, leucine; Met, methionine; Minor, heterozygote or homozygote of minor allele; SVR, sustained virological response; NVR, null virological response.

Japanese [6], European [7], and a multi-ethnic population [8,9]. The last three studies focused on the association of SNPs in the *IL28B* region with SVR [7–9] but we found a stronger association with NVR [6]. In addition to these host genetic factors, we have reported that mutations within a stretch of 40 amino acids in the NS5A region of HCV, designated as the IFN sensitivity determining region (ISDR), are closely associated with the virological response to IFN therapy: a lower number of mutations is associated with treatment failure [10–13]. Amino acid substitutions at positions 70 and 91 of the HCV core region (Core70, Core91) also have been reported to be associated with response to PEG-IFN/RBV therapy: glutamine (Gln) or histidine (His) at Core70 and methionine (Met) at Core91 are associated with treatment resistance [4,14]. The importance of substitutions in the HCV core and ISDR was confirmed recently by a Japanese multicenter study [15]. How these viral factors contribute to response to therapy is yet to be determined. For general application in clinical practice, host genetic factors and viral factors should be considered together.

Data mining analysis is a family of non-parametric regression methods for predictive modeling. Software is used to automatically explore the data to search for optimal split variables and to build a decision tree structure [16]. The major advantage of decision tree analysis over logistic regression analysis is that the results of the analysis are presented in the form of flow chart, which can be interpreted intuitively and readily made available for use in clinical practice [17]. The decision tree analysis has been utilized to define prognostic factors in various diseases [18–25]. We have reported recently its usefulness for the prediction of an early virological response (undetectable HCV-RNA within 12 weeks of therapy) to PEG-IFN/RBV therapy in chronic hepatitis C [26].

This study aimed to define the pre-treatment prediction of response to PEG-IFN/RBV therapy through the integrated analysis of host factors, such as the *IL28B* genetic polymorphism and various clinical covariates, as well as viral factors, such as mutations in the HCV core and ISDR and serum HCV-RNA load. In addition,

for the general application of these results in clinical practice, decision models for the pre-treatment prediction of response were determined by data mining analysis.

Materials and methods

Patients

This was a multicentre retrospective study supported by the Japanese Ministry of Health, Labor and Welfare. Data were collected from a total of 496 chronic hepatitis C patients who were treated with PEG-IFN alpha and RBV at five hospitals and universities throughout Japan. Of these, 98 patients also were included in the original GWAS analysis [6]. The inclusion criteria in this study were as follows (1) infection by genotype 1b, (2) lack of co-infection with hepatitis B virus or human immunodeficiency virus, (3) lack of other causes of liver disease, such as autoimmune hepatitis, and primary biliary cirrhosis, (4) completion of at least 24 weeks of therapy, (5) adherence of more than 80% to the planned dose of PEG-IFN and RBV for the NVR patients, (6) availability of DNA for the analysis of the genetic polymorphism of *IL28B*, and (7) availability of serum for the determination of mutations in the ISDR and substitutions of Core70 and Core91 of HCV. Patients received PEG-IFN alpha-2a (180 µg) or 2b (1.5 µg/kg) subcutaneously every week and were administered a weight adjusted dose of RBV (600 mg for <60 kg, 800 mg for 60–80 kg, and 1000 mg for >80 kg daily) which is the recommended dosage in Japan. 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 committee. The baseline characteristics are listed in Table 1. For the data mining analysis, 67% of the patients (331 patients) were assigned randomly to the model building group and 33% (165 patients) to the validation group. There were no significant differences in the clinical backgrounds between these two groups.

Laboratory and histological tests

Blood samples were obtained before therapy and were analyzed for hematologic tests and for blood chemistry and HCV-RNA. Sequences of ISDR and the core region of HCV were determined by direct sequencing after amplification by reverse-transcription and polymerase chain reaction as reported previously [4,11]. Genetic polymorphism in one tagging SNP located near the *IL28B* gene (rs8099917) was determined by the GWAS or DigiTag2 assay [27]. Homozygosity (GG) or heterozygosity (TG) of the minor sequence was defined as having the *IL28B* minor allele, whereas homozygosity for the major sequence (TT) was

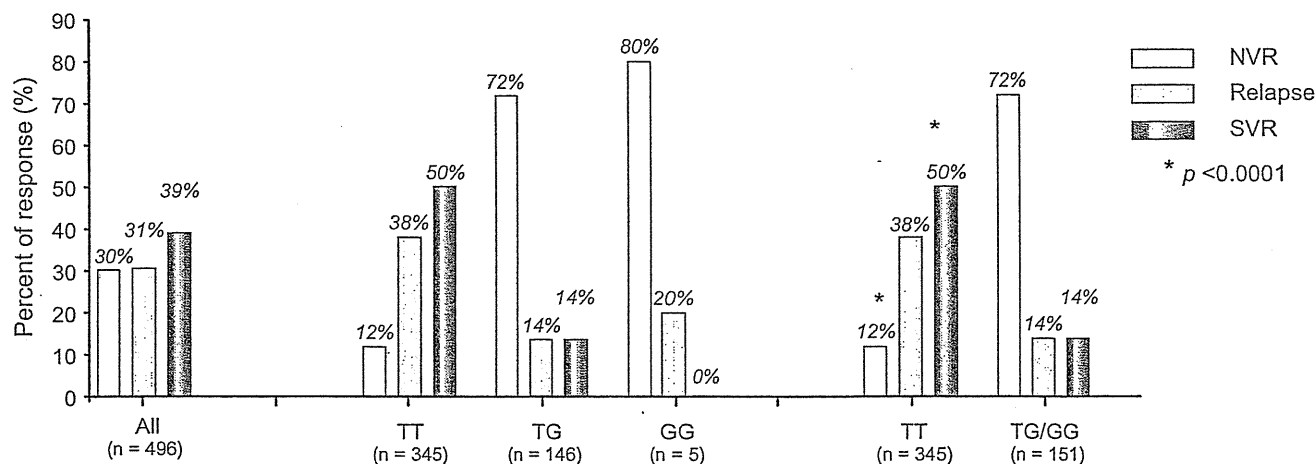


Fig. 1. Association between the IL28B genotype (rs8099917) and treatment response. The rates of response to treatment are shown for each rs8099917 genotype. The rate of null virological response (NVR), relapse, and sustained virological response (SVR) is shown. The *p* values are from Fisher's exact test. The rate of NVR was significantly higher ($p < 0.0001$) and the rate of SVR was significantly lower ($p < 0.0001$) in patients with the IL28B minor allele compared to those with the major allele.

defined as having the IL28B major allele. In this study, NVR was defined as a less than 2 log reduction of HCV-RNA at week 12 and detectable HCV-RNA by qualitative PCR with a lower detection limit of 50 IU/ml (Amplicor, Roche Diagnostic systems, CA) at week 24 during therapy. RVR (rapid virological response) and complete early virological response (cEVR) were defined as undetectable HCV-RNA at 4 weeks and 12 weeks during therapy and SVR was defined as undetectable HCV-RNA 24 weeks after the completion of therapy. Relapse was defined as reappearance of HCV-RNA after the completion of therapy. 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). Percentage of steatosis was quantified in 111 patients by determining the average proportion of hepatocytes affected by steatosis.

Statistical analysis

Associations between pre-treatment variables and treatment response were analyzed by univariate and multivariate logistic regression analysis. Associations between the IL28B polymorphism and sequences of HCV were analyzed by Fisher's exact test. SPSS software v.15.0 (SPSS Inc., Chicago, IL) was used for these analyses. For the data mining analysis, IBM-SPSS Modeler version 13.0 (IBM-SPSS Inc., Chicago, IL) software was utilized as reported previously [26]. The patients used for model building were divided into two groups at each step of the analysis based on split variables. Each value of each variable was considered as a potential split. The optimum variables and cut-off values were determined by a statistical search algorithm to generate the most significant division into two prognostic subgroups that were as homogeneous as possible for the probability of SVR. Thereafter, each subgroup was evaluated again and divided further into subgroups. This procedure was repeated until no additional significant variable was detected or the sample size was below 15. To avoid over-fitting, 10-fold cross validation was used in the tree building process. The reproducibility of the resulting model was tested with the data from the validation patients.

Results

Association between the IL28B (rs8099917) genotype and the PEG-IFN/RBV response

The rs8099917 allele frequency was 70% for TT ($n = 345$), 29% for TG ($n = 146$), and 1% for GG ($n = 5$). We defined the IL28B major allele as homozygous for the major sequence (TT) and the IL28B minor allele as homozygous (GG) or heterozygous (TG) for the minor sequence. The rate of NVR was significantly higher (72% vs. 12%, $p < 0.0001$) and the rate of SVR was significantly lower (14% vs. 50%, $p < 0.0001$) in patients with the IL28B minor allele compared to those with the major allele (Fig. 1).

Effect of the IL28B polymorphism, substitutions in the ISDR, Core70, and Core91 of HCV on time-dependent clearance of HCV

Patients were stratified according to their IL28B allele type, the number of mutations in the ISDR, the amino acid substitutions in Core70 and Core91, and the rate of undetectable HCV-RNA at 4, 8, 12, 24, and 48 weeks after the start of therapy were analyzed (Fig. 2A–D). The rate of undetectable HCV-RNA was significantly higher in patients with the IL28B major allele than the minor allele, in patients with two or more mutations in the ISDR compared to none or only one mutation, in patients with arginine (Arg) at Core70 rather than Gln/His, and in patients with leucine (Leu) at Core91 rather than Met. The difference was most significant when stratified by the IL28B allele type. The rate of RVR and cEVR was significantly more frequent in patients with the IL28B major allele compared with those with the IL28B minor allele: 9% vs. 3% for RVR ($p < 0.005$) and 57% vs. 11% for cEVR ($p < 0.0001$). These findings suggest that IL28B has the greatest impact on early virological response to therapy.

Association between substitutions in the ISDR and relapse after the completion of therapy

Patients were stratified according to the IL28B allele, number of mutations in the ISDR, and amino acid substitutions of Core70 and Core91, and the rate of relapse was analyzed (Fig. 3A and B). Among patients who achieved cEVR, the rate of relapse was significantly lower in patients with two or more mutations in the ISDR compared to those with only one or no mutations (15% vs. 31%, $p < 0.005$) (Fig. 3 B). On the other hand, the relapse rate was not different between the IL28B major and minor alleles within patients who achieved RVR (3% vs. 0%) or cEVR (28% vs. 29%) (Fig. 3A). Amino acid substitutions of Core70 and Core91 were not associated with the rate of relapse (data not shown).

Factors associated with response by multivariate logistic regression analysis

By univariate analysis, the minor allele of IL28B ($p < 0.0001$), one or no mutations in the ISDR ($p = 0.03$), high serum level of

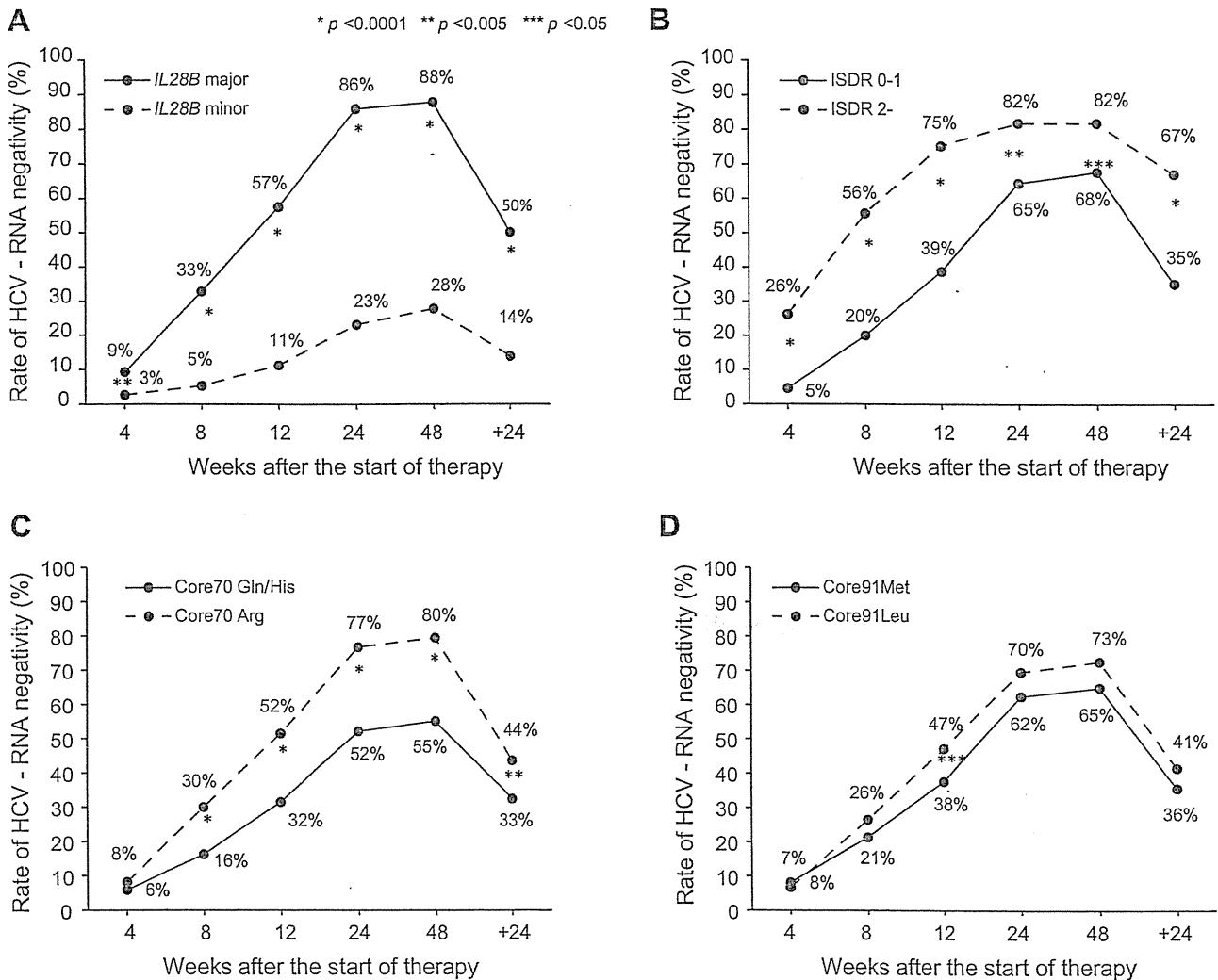


Fig. 2. Effect of *IL28B* mutations in the ISDR, Core70, and Core91 of HCV on time-dependent clearance of HCV. The rate of undetectable HCV-RNA was plotted for serial time points after the start of therapy (4, 8, 12, 24, and 48 weeks) and for 24 weeks after the completion of therapy. Patients were stratified according to (A) the *IL28B* allele (minor allele vs. major allele), (B) the number of mutations in the ISDR (0–1 mutation vs. 2 or more mutations), amino acid substitutions of (C) Core70 (Gln/His vs. Arg), and (D) Core91 (Met vs. Leu). The *p* values are from Fisher's exact test.

HCV-RNA ($p = 0.035$), Gln or His at Core70 ($p < 0.0001$), low platelet counts ($p = 0.009$), and advanced fibrosis ($p = 0.0002$) were associated with NVR. By multivariate analysis, the minor allele of *IL28B* (OR = 20.83, 95%CI = 11.63–37.04, $p < 0.0001$) was associated with NVR independent of other covariates (Table 2). Notably, mutations in the ISDR ($p = 0.707$) and at amino acid Core70 ($p = 0.207$) were not significant in multivariate analysis due to the positive correlation with the *IL28B* polymorphism ($p = 0.004$ for ISDR and $p < 0.0001$ for Core70, Fig. 4).

Genetic polymorphism of *IL28B* also was associated with SVR (OR = 7.41, 95% CI = 4.05–13.57, $p < 0.0001$) independent of other covariates, such as platelet counts, fibrosis, and serum levels of HCV-RNA. Mutation in the ISDR was an independent predictor of SVR (OR = 2.11, 95% CI = 1.06–4.18, $p = 0.033$) but the amino acid at Core70 was not (Table 3).

Factors associated with the *IL28B* polymorphism

Patients with the *IL28B* minor allele had significantly higher serum level of gamma-glutamyltransferase (GGT) and a higher

frequency of hepatic steatosis (Table 4). When the association between the *IL28B* polymorphism and HCV sequences was analyzed, Gln or His at Core70, that is linked to resistance to PEG-IFN and RBV therapy [4,14,15], was significantly more frequent in patients with the minor *IL28B* allele than in those with the major allele (67% vs. 30%, $p < 0.0001$) (Fig. 4). Other HCV sequences with an IFN resistant phenotype also were more prevalent in patients with the minor *IL28B* allele than those with the major allele: Met at Core91 (46% vs. 37%, $p = 0.047$) and one or no mutations in the ISDR (94% vs. 85%, $p = 0.004$) (Fig. 4).

Data mining analysis

Data mining analysis was performed to build a model for the prediction of SVR and the result is shown in Fig. 5. The analysis selected four predictive variables, resulting in six subgroups of patients. Genetic polymorphism of *IL28B* was selected as the best predictor of SVR. Patients with the minor *IL28B* allele had a lower probability of SVR and a higher probability of NVR than those with the major *IL28B* allele (SVR: 14% vs. 50%, NVR: 72% vs.

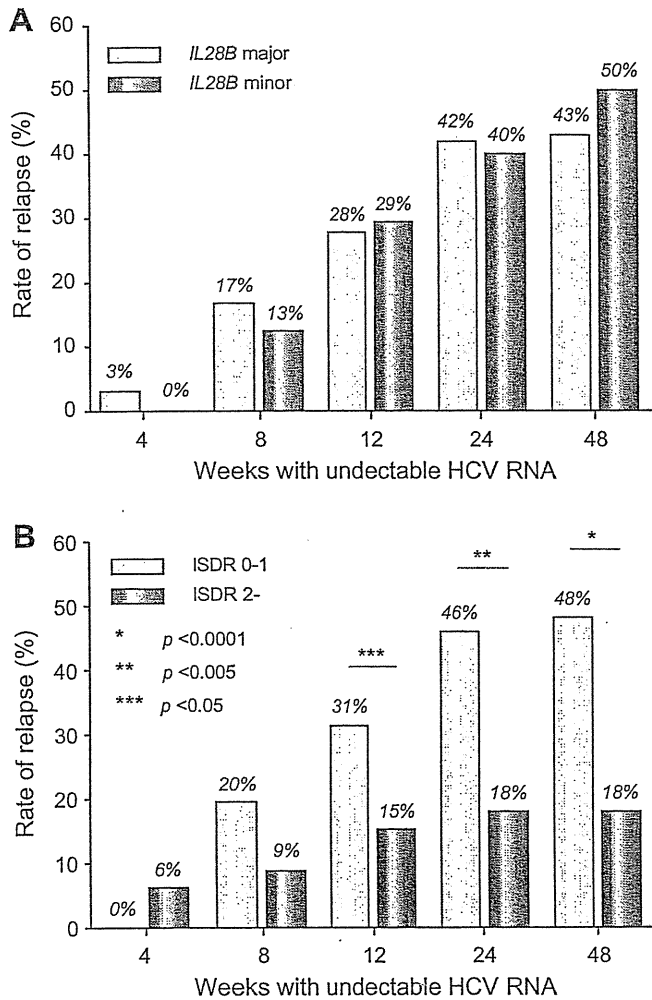


Fig. 3. Association between relapse and the *IL28B* allele or mutations in the ISDR. The rate of relapse was calculated for patients who had undetectable HCV-RNA at serial time points after the start of therapy (4, 8, 12, 24, and 48 weeks). Patients were stratified according to (A) the *IL28B* allele (minor allele vs. major allele) and (B) the number of mutations in the ISDR (0–1 mutation vs. 2 or more mutations). The *p* values are from Fisher's exact test.

12%). After stratification by the *IL28B* allele, patients with low platelet counts ($<140 \times 10^9/L$) had a lower probability of SVR and higher probability of NVR than those with high platelet counts ($\geq 140 \times 10^9/L$): for the minor *IL28B* allele, SVR was 7% vs. 19%, and NVR was 84% vs. 62%, and for the major *IL28B* allele, SVR was 32% vs. 66% and NVR was 16% vs. 8%. Among patients with the major *IL28B* allele and low platelet counts, those with two or more mutations in the ISDR had a higher probability of SVR and lower probability of relapse than those with one or no mutations in the ISDR (SVR: 75% vs. 27%, and relapse: 8% vs. 57%). Among patients with the major *IL28B* allele and high platelet counts, those with a low HCV-RNA titer ($<600,000$ IU/ml) had a higher probability of SVR and lower probability of NVR and relapse than those with a high HCV-RNA titer (SVR: 90% vs. 61%, NVR: 0% vs. 10%, and relapse: 10% vs. 29%). The sensitivity and specificity of the decision tree were 78% and 70%, respectively. The area under the receiver operating characteristic (ROC) curve of the model was 0.782 (data not shown). The pro-

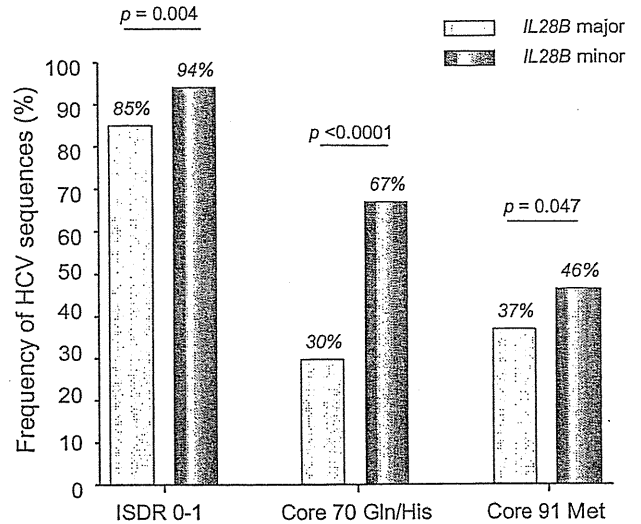


Fig. 4. Associations between the *IL28B* allele and HCV sequences. The prevalence of HCV sequences predicting a resistant phenotype to IFN was higher in patients with the minor *IL28B* allele than those with major allele. (A) 0 or 1 mutation in the ISDR of NS5A, (B) Gln or His at Core70, and (C) Met at Core91. *p* values are from Fisher's exact test.

portion of patients with advanced fibrosis (F3–4) was 39% (84/217) in patients with low platelet counts ($<140 \times 10^9/L$) compared to 13% (37/279) in those with high platelet counts ($\geq 140 \times 10^9/L$).

Validation of the data mining analysis

The results of the data mining analysis were validated with 165 patients who differed from those used for model building. Each patient was allocated to one of the six subgroups for the validation using the flow-chart form of the decision tree. The rate of SVR and NVR in each subgroup was calculated. The rates of SVR and NVR for each subgroup of patients were closely correlated between the model building and the validation patients ($r^2 = 0.99$ and 0.98) (Fig. 6).

Discussion

The rate of NVR after 48 weeks of PEG-IFN/RBV therapy among patients infected with HCV of genotype 1 is around 20–30%. Previously, there have been no reliable baseline predictors of NVR or SVR. Because more potent therapies, such as protease and polymerase inhibitor of HCV [28,29] and nitazoxanide [30], are in clinical trials and may become available in the near future, a pre-treatment prediction of the likelihood of response may be helpful for patients and physicians, to support clinical decisions about whether to begin the current standard of care or whether to wait for emerging therapies. This study revealed that the *IL28B* polymorphism was the overwhelming predictor of NVR and is independent of host factors and viral sequences reported previously. The *IL28B* encodes a protein also known as IFN-lambda 3, which is thought to suppress the replication of various viruses including HCV [31,32]. The results of the current study and the findings of the GWAS studies [6–9] may provide the rationale for developing diagnostic testing or an IFN-lambda based therapy for chronic hepatitis C in the future.

Research Article

Table 2. Factors associated with NVR analyzed by univariate and multivariate logistic regression analysis.

	Univariate			Multivariate		
	Odds ratio	95%CI	p value	Odds ratio	95%CI	p value
Gender: female	0.98	0.67-1.45	0.938	1.29	0.75-2.23	0.363
Age	1.01	0.97-1.01	0.223	0.99	0.97-1.02	0.679
ALT	1.00	1.00-1.00	0.867	1.00	0.99-1.00	0.580
GGT	1.004	1.00-1.01	0.029	1.00	1.00-1.00	0.715
Platelets	0.95	0.91-0.99	0.009	0.92	0.87-0.98	0.006
Fibrosis: F3-4	2.23	1.46-3.42	0.0002	1.97	1.09-3.57	0.025
HCV-RNA: $\geq 600,000$ IU/ml	1.83	1.05-3.19	0.035	2.49	1.17-5.29	0.018
ISDR mutation: ≤ 1	2.14	1.08-4.22	0.030	0.96	0.78-1.18	0.707
Core 70 (Gln/His)	3.23	2.16-4.78	<0.0001	1.41	0.83-2.42	0.207
Core 91 (Met)	1.39	0.95-2.06	0.093	1.21	0.72-2.04	0.462
IL28B: Minor allele	19.24	11.87-31.18	<0.0001	20.83	11.63-37.04	<0.0001

ALT, alanine aminotransferase; GGT, gamma-glutamyltransferase; ISDR, interferon sensitivity determining region; Gln, glutamine; His, histidine; Met, methionine; Minor allele, heterozygote or homozygote of minor allele.

Table 3. Factors associated with SVR analyzed by univariate and multivariate logistic regression analysis.

	Univariate			Multivariate		
	Odds ratio	95%CI	p value	Odds ratio	95%CI	p value
Gender: female	0.81	0.56-1.16	0.253	0.86	0.55-1.35	0.508
Age	0.97	0.95-0.99	0.0003	0.99	0.96-1.01	0.199
ALT	1.00	1.00-1.00	0.337	1.00	1.00-1.01	0.108
GGT	1.00	1.00-1.00	0.273	1.00	1.00-1.00	0.797
Platelets	1.12	1.01-1.16	<0.0001	1.13	1.08-1.19	<0.0001
Fibrosis: F0-2	2.64	1.65-4.22	<0.0001	1.87	1.07-3.28	0.029
HCV-RNA: <600,000 IU/ml	2.49	1.55-3.98	0.0001	2.75	1.55-4.90	0.001
ISDR mutation: $2\leq$	3.78	2.14-6.68	<0.0001	2.11	1.06-4.18	0.033
Core 70 (Arg)	1.61	1.11-2.28	0.012	0.84	0.52-1.35	0.470
Core 91 (Leu)	1.28	0.88-1.85	0.185	1.26	0.81-1.96	0.300
IL28B: Major allele	6.21	3.75-10.31	<0.0001	7.41	4.05-13.57	<0.0001

ALT, alanine aminotransferase; GGT, Gamma-glutamyltransferase; ISDR, interferon sensitivity determining region; Arg, arginine; Leu, leucine; Major allele, homozygote of major allele.

Among baseline factors, *IL28B* was the most significant predictor of NVR and SVR. Moreover, the *IL28B* allele type was also correlated with early virological response: the rate of RVR and cEVR was significantly high for the *IL28B* major allele compared to the *IL28B* minor allele: 9% vs. 3% for RVR and 57% vs. 11% for cEVR (Fig. 2). On the other hand, the relapse rate was not different between the *IL28B* genotypes within patients who achieved RVR or cEVR (Fig. 3). We believe that optimal therapy should be based on baseline features and a response-guided approach. Our findings suggest that the *IL28B* genotype is a useful baseline predictor of virological response which should be used for selecting the treatment regimen: whether to treat patients with PEG-IFN and RBV or to wait for more effective future therapy including direct acting antiviral drugs. On the other hand, baseline *IL28B* genotype might not be suitable for determining the treatment duration in patients who started PEG-IFN/RBV therapy

and whose virological response is determined because the *IL28B* genotype is not useful for the prediction of relapse. The duration of therapy should be personalized based on the virological response. Future studies need to explore whether the combination of baseline *IL28B* genotype and response-guided approach further improves the optimization of treatment duration.

The SVR rate in patients having the *IL28B* minor allele was 14% in the present study while it was 23% in Caucasians and 9% in African Americans in a study by McCarthy et al. [33]. On the other hand, the SVR rate in patients having the *IL28B* minor allele was 28% in genotypes 1/4 compared to 80% in genotypes 2/3 in a study by Rauch et al. [9]. These data imply that the impact of the *IL28B* polymorphism on response to therapy may be different in terms of race, geographical areas, or HCV genotypes, and that our data need to be validated in future studies including different populations and geographical areas before generalization.

Table 4. Factors associated with IL28B genotype.

	IL28B major allele n = 345	IL28B minor allele n = 151	p value
Gender: male	166 (48%)	84 (56%)	0.143
Age (years)	57 ± 10	57 ± 10	0.585
ALT (IU/L)	79 ± 60	78 ± 62	0.842
Platelets (10 ⁹ /L)	153 ± 54	155 ± 52	0.761
GGT (IU/L)	51 ± 45	78 ± 91	0.001
Fibrosis: F3-4	76 (22%)	45 (30%)	0.063
Steatosis:			
>10%	16/88 (18%)	13/23 (57%)	0.024
>30%	6/88 (7%)	6/23 (26%)	0.017
HCV-RNA: >600,000 IU/ml	284 (82%)	125 (83%)	1.000

ALT, alanine aminotransferase; GGT, gamma-glutamyltransferase.

Four GWAS studies have shown the association between a genetic polymorphism near the *IL28B* gene and response to PEG-IFN plus RBV therapy. The SNPs that showed significant association with response were rs12979860 [8] and rs8099917 [6,7,9]. There is a strong linkage-disequilibrium (LD) between these two SNPs as well as several other SNPs near the *IL28B* gene in Japanese patients [34] but the degree of LD was weaker in Caucasians and Hispanics [8]. Thus, the combination of SNPs is not useful for predicting response in Japanese patients but may improve the predictive value in patients other than Japanese who have weaker LD between SNPs.

Other significant predictors of response independent of *IL28B* genotype were platelet counts, stage of fibrosis, and HCV RVA load. A previous study reported that platelet count is a predictor of response to therapy [35], and the lower platelet count was related with advanced liver fibrosis in the present study. The association between response to therapy and advanced fibrosis independent of the *IL28B* polymorphism is consistent with a recent study by Rauch et al. [9].

There is agreement that the viral genotype is significantly associated with the treatment outcome. Moreover, viral factors such as substitutions in the ISDR of the NS5A region [10] or in the amino acid sequence of the HCV core [4] have been studied in relation to the response to IFN treatment. The amino acid Gln or His at Core70 and Met at Core91 are repeatedly reported to be associated with resistance to therapy [4,14,15] in Japanese patients but these data wait to be validated in different populations or other geographical areas. In this study, we confirmed that patients with two or more mutations in the ISDR had a higher rate of undetectable HCV-RNA at each time point during therapy. In addition, the rate of relapse among patients who achieved cEVR was significantly lower in patients with two or more mutations in ISDR compared to those with only one or no mutations (15% vs. 31%, *p* < 0.05). Thus, the ISDR sequence may be used to predict a relapse among patients who achieved virological response during therapy, while the *IL28B* polymorphism may be used to predict the virological response before therapy. A higher number of mutations in the ISDR are reported to have close association with SVR in Japanese [11–13,15,36] or Asian [37,38] populations but data from Western countries have been controversial [39–42]. A meta-analysis of 1230 patients including 525 patients from Europe has shown that there was a positive correlation

between the SVR and the number of mutations in the ISDR in Japanese as well as in European patients [43] but this correlation was more pronounced in Japanese patients. Thus, geographical factors may account for the different impact of ISDR on treatment response, which may be a potential limitation of our study.

To our surprise, these HCV sequences were associated with the *IL28B* genotype: HCV sequences with an IFN resistant phenotype were more prevalent in patients with the minor *IL28B* allele than those with the major allele. This was an unexpected finding, as we initially thought that host genetics and viral sequences were completely independent. A recent study reported that the *IL28B* polymorphism (rs12979860) was significantly associated with HCV genotype: the *IL28B* minor allele was more frequent in HCV genotype 1-infected patients compared to patients infected with HCV genotype 2 or 3 [33]. Again, patients with the *IL28B* minor allele (IFN resistant genotype) were infected with HCV sequences that are linked to an IFN resistant phenotype. The mechanism for this association is unclear, but may be related to an interaction between the *IL28B* genotype and HCV sequences in the development of chronic HCV infection as discussed by McCarthy et al., since the *IL28B* polymorphism was associated with the natural clearance of HCV [44]. Alternatively, the HCV sequence within the patient may be selected during the course of chronic infection [45,46]. These hypotheses should be explored through prospective studies of spontaneous HCV clearance or by testing the time-dependent changes in the HCV sequence during the course of chronic infection.

How these host and viral factors can be integrated to predict the response to therapy in future clinical practice is an important question. Because various host and viral factors interact in the same patient, predictive analysis should consider these factors in combination. Using the data mining analysis, we constructed a simple decision tree model for the pre-treatment prediction of SVR and NVR to PEG-IFN/RBV therapy. The classification of patients based on the genetic polymorphism of *IL28B*, mutation in the ISDR, serum levels of HCV-RNA, and platelet counts, identified subgroups of patients who have the lowest probabilities of NVR (0%) with the highest probabilities of SVR (90%) as well as those who have the highest probabilities of NVR (84%) with the lowest probability of SVR (7%). The reproducibility of the model was confirmed by the independent validation based on a second group of patients. Using this model, we can rapidly develop an