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.

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

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

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.

Research Article

EASL JOURNAL OF HEPATOLOGY

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

Masayuki Kurosaki[1], Yasuhito Tanaka[2], Nao Nishida[3], Naoya Sakamoto[4], Nobuyuki Enomoto[5], Masao Honda[6], Masaya Sugiyama[2], Kentaro Matsuura[2], Fuminaka Sugauchi[2], Yasuhiro Asahina[1], Mina Nakagawa[4], Mamoru Watanabe[4], Minoru Sakamoto[5], Shinya Maekawa[5], Akito Sakai[6], Shuichi Kaneko[6], Kiyoaki Ito[7], Naohiko Masaki[7], Katsushi Tokunaga[3], Namiki Izumi[1,*], Masashi Mizokami[2,7]

[1]*Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan;* [2]*Department of Virology, Liver Unit, Nagoya City University, Graduate School of Medical Sciences, Nagoya, Japan;* [3]*Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan;* [4]*Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Tokyo, Japan;* [5]*First Department of Internal Medicine, University of Yamanashi, Yamanashi, Japan;* [6]*Department of Gastroenterology, Kanazawa University, Graduate School of Medicine, Kanazawa, Japan;* [7]*Research 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.

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).

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

ELSEVIER

Journal of Hepatology **2011** vol. 54 | 439–448

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%)
IL28B: 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

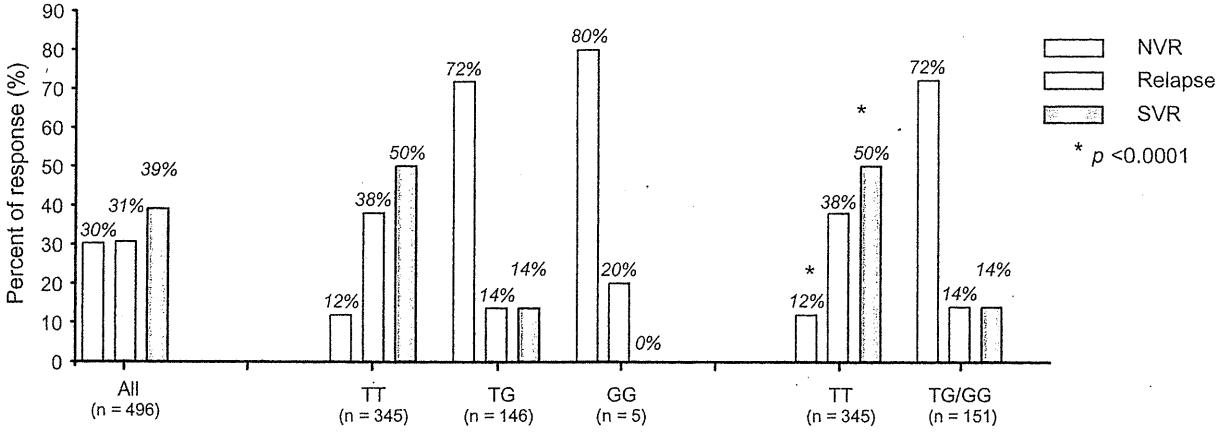

Fig. 1. Association between the IL28B genotype (rs8099917) and treatment response. The rates of response to treatment are shown for each rs8099917 genotype. The rate of null virological response (NVR), relapse, and sustained virological response (SVR) is shown. The *p* values are from Fisher's exact 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

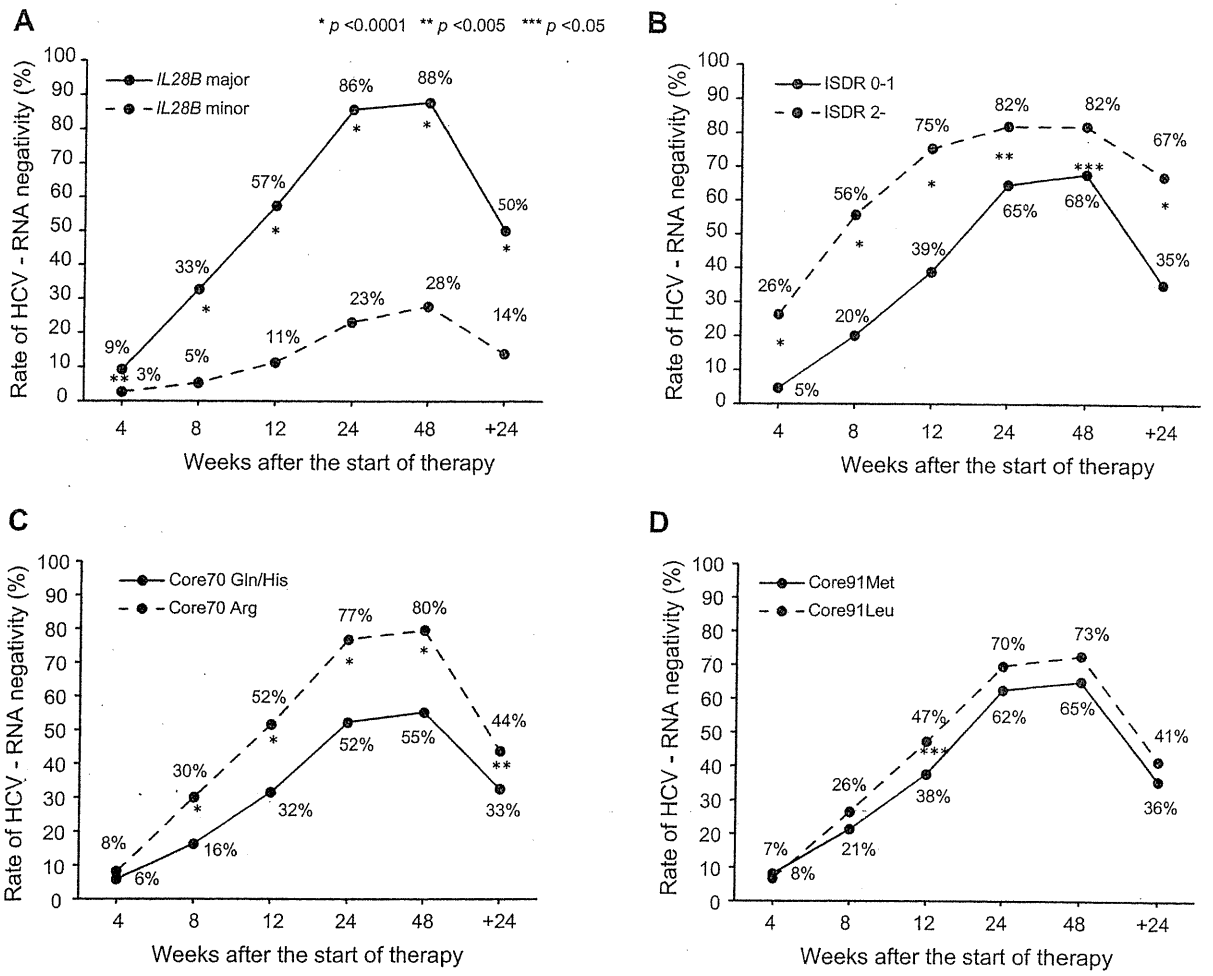

Fig. 2. Effect of *IL28B* mutations in the ISDR, Core70, and Core91 of HCV on time-dependent clearance of HCV. The rate of undetectable HCV-RNA was plotted for serial time points after the start of therapy (4, 8, 12, 24, and 48 weeks) and for 24 weeks after the completion of therapy. Patients were stratified 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.

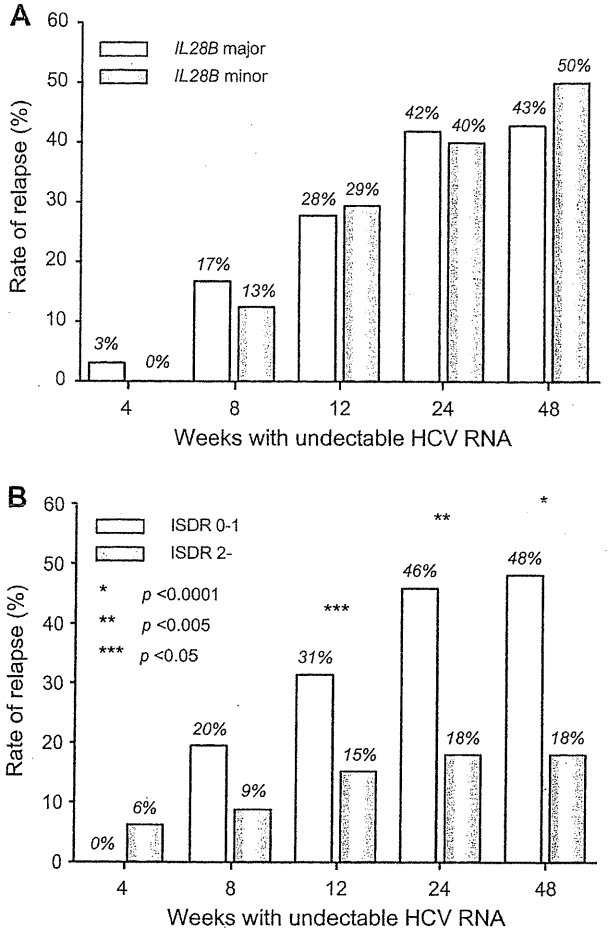

Fig. 3. Association between relapse and the *IL28B* allele or mutations in the ISDR. The rate of relapse was calculated for patients who had undetectable HCV-RNA at serial time points after the start of therapy (4, 8, 12, 24, and 48 weeks). Patients were stratified according to (A) the *IL28B* allele (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.

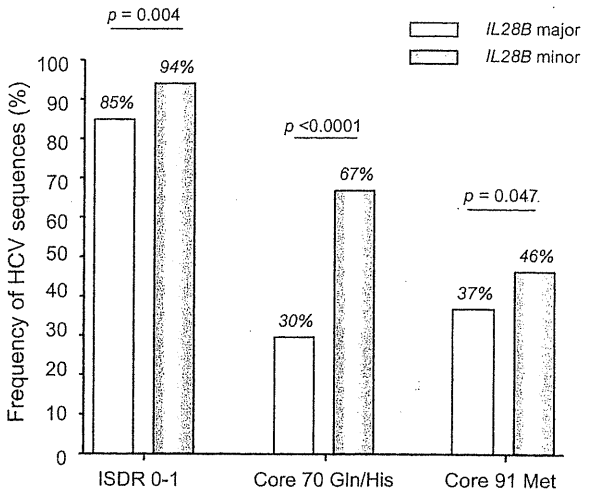

Fig. 4. Associations between the *IL28B* allele and HCV sequences. The prevalence of HCV sequences predicting a resistant phenotype to IFN was higher in patients with the minor *IL28B* allele than those with major allele. (A) 0 or 1 mutation in the ISDR of NS5A, (B) Gln or His at Core70, and (C) Met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

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 proportion 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.

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: ≥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	1.01-116	<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≤	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

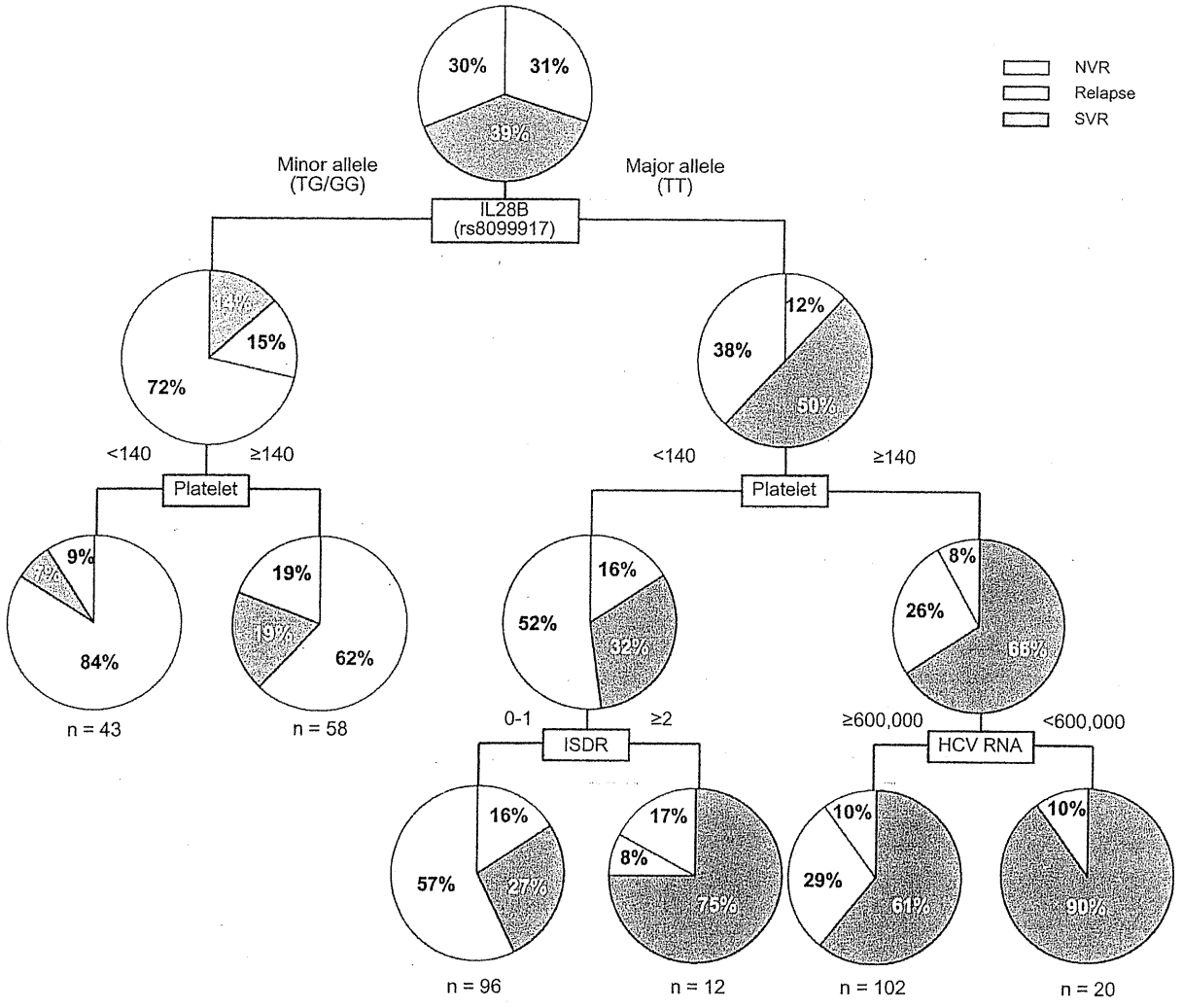

Fig. 5. Decision tree for the prediction of response to therapy. The boxes indicate the factors used for splitting. Pie charts indicate the rate of response for each group of patients after splitting. The rate of null virological response, relapse, and sustained virological response is shown.

estimate of the response before treatment, by simply allocating patients to subgroups by following the flow-chart form, which may facilitate clinical decision making. This is in contrast to the calculating formula, which was constructed by the traditional logistic regression model. This was not widely used in clinical practice as it is abstruse and inconvenient. These results support the evidence based approach of selecting the optimum treatment strategy for individual patients, such as treating patients with a low probability of NVR with current PEG-IFN/RBV combination therapy or advising those with a high probability of NVR to wait for more effective future therapies. Patients with a high probability of relapse may be treated for a longer duration to avoid a relapse. Decisions may be based on the possibility of a response against a potential risk of adverse events and the cost of the therapy, or disease progression while waiting for future therapy.

We have previously reported the predictive model of early virological response to PEG-IFN and RBV in chronic hepatitis C [26]. The top factor selected as significant was the grade of steatosis, followed by serum level of LDL cholesterol, age, GGT, and blood sugar. The mechanism of association between these factors and treatment response was not clear at that time. To our interest, a recent study by Li et al. [47] has shown that high serum level of LDL cholesterol was linked to the *IL28B* major allele (CC in rs12979860). High serum level of LDL cholesterol was associated with SVR 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 may reflect the underlining link of LDL cholesterol levels to *IL28B* genotype. Steatosis is reported to be correlated with low lipid levels [48] which suggest that *IL28B* genotypes may be also associated with steatosis. In fact, there were significant correlations between the *IL28B* genotype and the presence of steatosis in the present study (Table 4). In addition, the serum level of GGT, another predictive factor in our previous study, was signif-

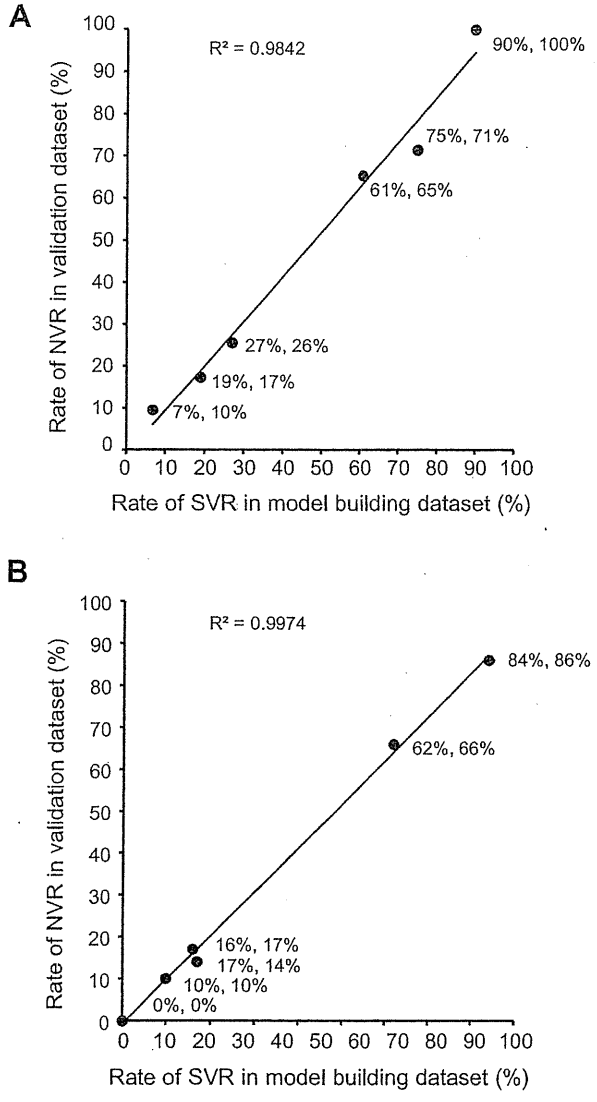

Fig. 6. Validation of the CART analysis. Each patient in the validation group was allocated to one of the six subgroups by following the flow-chart form of the decision tree. The rate of (A) sustained virological response (SVR) and (B) null virological response (NVR) in each subgroup was calculated and plotted. The *X*-axis represents the rate of SVR or NVR in the model building patients and the *Y*-axis represents those in the validation patients. The rate of SVR and NVR in each subgroup of patients is closely correlated between the model building and the validation patients (correlation coefficient: $r^2 = 0.98–0.99$).

icantly associated with *IL28B* genotype in the present study (Table 4). The serum level of GGT was significantly associated with NVR when examined independently but was no longer significant when analyzed together with the *IL28B* genotype. These observations indicate that some of the factors that we have previously identified may be associated with virological response to therapy through the underlining link to the *IL28B* genotype.

In conclusion, the present study highlighted the impact of the *IL28B* polymorphism and mutation in the ISDR on the pre-treatment prediction of response to PEG-IFN/RBV therapy. A decision model including these host and viral factors has the potential to support selection of the optimum treatment strategy for individual patients, which may enable personalized treatment.

Conflict of interest

The authors who have taken part in this study declare that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

Financial support

This study was supported by a grant-in-aid from the Ministry of Health, Labor and Welfare, Japan, (H19-kannen-013), (H20-kannen-006).

References

[1] Ray Kim W. Global epidemiology and burden of hepatitis C. Microbes Infect 2002;4 (12):1219–1225.
[2] Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales Jr FL, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med 2002;347 (13):975–982.
[3] Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet 2001;358 (9286):958–965.
[4] Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, Someya T, et al. 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 2005;48 (6):372–380.
[5] Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. Hepatology 2003;38 (3):645–652.
[6] Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 2009;10:1105–1109.
[7] Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate ML, et al. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. Nat Genet 2009;10:1100–1104.
[8] Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 2009;461 (7262):399–401.
[9] Rauch A, Kutalik Z, Descombes P, Cai T, Di Iulio J, Mueller T, et al. Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. Gastroenterology 2010;138 (4):1338–1345.
[10] Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. 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 1995;96 (1):224–230.
[11] Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N Engl J Med 1996;334 (2):77–81.
[12] Kurosaki M, Enomoto N, Murakami T, Sakuma I, Asahina Y, Yamamoto C, et al. 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 1997;25 (3):750–753.
[13] Shirakawa H, Matsumoto A, Joshita S, Komatsu M, Tanaka N, Umemura T, et al. Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors. Hepatology 2008;48 (6):1753–1760.
[14] Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, 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. J Hepatol 2007;46 (3):403–410.

[15] Okanoue T, Itoh Y, Hashimoto H, Yasui K, Minami M, Takehara T, 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. J Gastroenterol 2009;44 (9):952–963.
[16] Segal MR, Bloch DA. A comparison of estimated proportional hazards models and regression trees. Stat Med 1989;8 (5):539–550.
[17] LeBlanc M, Crowley J. A review of tree-based prognostic models. Cancer Treat Res 1995;75:113–124.
[18] Garzotto M, Beer TM, Hudson RG, Peters L, Hsieh YC, Barrera E, et al. Improved detection of prostate cancer using classification and regression tree analysis. J Clin Oncol 2005;23 (19):4322–4329.
[19] Averbook BJ, Fu P, Rao JS, Mansour EG. 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 2002;132 (4):589–602.
[20] Leiter U, Buettner PG, Eigentler TK, Garbe C. Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the german dermatological society. J Clin Oncol 2004;22 (18):3660–3667.
[21] Valera VA, Walter BA, Yokoyama N, Koyama Y, Iiai T, Okamoto H, et al. Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. Ann Surg Oncol 2007;14 (1):34–40.
[22] Zlobec I, Steele R, Nigam N, Compton CC. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. Clin Cancer Res 2005;11 (15):5440–5443.
[23] Thabane M, Simunovic M, Akhtar-Danesh N, Marshall JK. Development and validation of a risk score for post-infectious irritable bowel syndrome. Am J Gastroenterol 2009;104 (9):2267–2274.
[24] Wu BU, Johannes RS, Sun X, Tabak Y, Conwell DL, Banks PA. The early prediction of mortality in acute pancreatitis: a large population-based study. Gut 2008;57 (12):1698–1703.
[25] Fonarow GC, Adams Jr KF, Abraham WT, Yancy CW, Boscardin WJ. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. Jama 2005;293 (5):572–580.
[26] Kurosaki M, Matsunaga K, Hirayama I, Tanaka T, Sato M, Yasui Y, et al. A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. Hepatol Res 2010;40 (3):251–260.
[27] Nishida N, Tanabe T, Takasu M, Suyama A, Tokunaga K. Further development of multiplex single nucleotide polymorphism typing method, the DigiTag2 assay. Anal Biochem 2007;364 (1):78–85.
[28] Hezode C, Forestier N, Dusheiko G, Ferenci P, Pol S, Goeser T, et al. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. N Engl J Med 2009;360 (18):1839–1850.
[29] McHutchison JG, Everson GT, Gordon SC, Jacobson IM, Sulkowski M, Kauffman R, et al. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. N Engl J Med 2009;360 (18):1827–1838.
[30] Rossignol JF, Elfert A, El-Gohary Y, Keeffe EB. Improved virologic response in chronic hepatitis C genotype 4 treated with nitazoxanide, peginterferon, and ribavirin. Gastroenterology 2009;136 (3):856–862.
[31] Marcello T, Grakoui A, Barba-Spaeth G, Machlin ES, Kotenko SV, MacDonald MR, et al. Interferons alpha and lambda inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics. Gastroenterology 2006;131 (6):1887–1898.
[32] Robek MD, Boyd BS, Chisari FV. Lambda interferon inhibits hepatitis B and C virus replication. J Virol 2005;79 (6):3851–3854.
[33] McCarthy JJ, Li JH, Thompson A, Suchindran S, Lao XQ, Patel K, et al. Replicated association between an IL28B Gene Variant and a Sustained Response to Pegylated Interferon and Ribavirin. Gastroenterology 2010;138:2307–2314.
[34] Tanaka Y, Nishida N, Sugiyama M, Tokunaga K, Mizokami M. A-interferons and the single nucleotide polymorphisms: a milestone to tailor-made therapy for chronic hepatitis C. Hepatol Res 2010;40:449–460.
[35] Backus LI, Boothroyd DB, Phillips BR, Mole LA. Predictors of response of US veterans to treatment for the hepatitis C virus. Hepatology 2007;46 (1):37–47.
[36] Mori N, Imamura M, Kawakami Y, Saneto H, Kawaoka T, Takaki S, et al. Randomized trial of high-dose interferon-alpha-2b combined with ribavirin in patients with chronic hepatitis C: correlation between amino acid substitutions in the core/NS5A region and virological response to interferon therapy. J Med Virol 2009;81 (4):640–649.
[37] Hung CH, Lee CM, Lu SN, Lee JF, Wang JH, Tung HD, et al. Mutations in the NS5A and E2-PePHD region of hepatitis C virus type 1b and correlation with the response to combination therapy with interferon and ribavirin. J Viral Hepat 2003;10 (2):87–94.
[38] Yen YH, Hung CH, Hu TH, Chen CH, Wu CM, Wang JH, et al. Mutations in the interferon sensitivity-determining region (nonstructural 5A amino acid 2209–2248) in patients with hepatitis C-1b infection and correlating response to combined therapy of pegylated interferon and ribavirin. Aliment Pharmacol Ther 2008;27 (1):72–79.
[39] Zeuzem S, Lee JH, Roth WK. Mutations in the nonstructural 5A gene of European hepatitis C virus isolates and response to interferon alfa. Hepatology 1997;25 (3):740–744.
[40] Squadrito G, Leone F, Sartori M, Nalpas B, Berthelot P, Raimondo G, et al. Mutations in the nonstructural 5A region of hepatitis C virus and response of chronic hepatitis C to interferon alfa. Gastroenterology 1997;113 (2):567–572.
[41] Sarrazin C, Berg T, Lee JH, Teuber G, Dietrich CF, Roth WK, et al. Improved correlation between multiple mutations within the NS5A region and virological response in European patients chronically infected with hepatitis C virus type 1b undergoing combination therapy. J Hepatol 1999;30 (6):1004–1013.
[42] Murphy MD, Rosen HR, Marousek GI, Chou S. Analysis of sequence configurations of the ISDR, PKR-binding domain, and V3 region as predictors of response to induction interferon-alpha and ribavirin therapy in chronic hepatitis C infection. Dig Dis Sci 2002;47 (6):1195–1205.
[43] Pascu M, Martus P, Hohne M, Wiedenmann B, Hopf U, Schreier E, 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 2004;53 (9):1345–1351.
[44] Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 2009;461 (7265):798–801.
[45] Kurosaki M, Enomoto N, Marumo F, Sato C. Evolution and selection of hepatitis C virus variants in patients with chronic hepatitis C. Virology 1994;205 (1):161–169.
[46] Enomoto N, Kurosaki M, Tanaka Y, Marumo F, Sato C. Fluctuation of hepatitis C virus quasispecies in persistent infection and interferon treatment revealed by single-strand conformation polymorphism analysis. J Gen Virol 1994;75 (Pt 6):1361–1369.
[47] Li JH, Lao XQ, Tillmann HL, Rowell J, Patel K, Thompson A, et al. Interferon-lambda genotype and low serum low-density lipoprotein cholesterol levels in patients with chronic hepatitis C infection. Hepatology 1904;51 (6):1904–1911.
[48] Serfaty L, Andreani T, Giral P, Carbonell N, Chazouilleres O, Poupon R. Hepatitis C virus induced hypobetalipoproteinemia: a possible mechanism for steatosis in chronic hepatitis C. J Hepatol 2001;34 (3):428–434.

J Gastroenterol (2011) 46:1111–1119
DOI 10.1007/s00535-011-0412-z

ORIGINAL ARTICLE—LIVER, PANCREAS, AND BILIARY TRACT

Pretreatment prediction of anemia progression by pegylated interferon alpha-2b plus ribavirin combination therapy in chronic hepatitis C infection: decision-tree analysis

Naoki Hiramatsu · Masayuki Kurosaki · Naoya Sakamoto · Manabu Iwasaki · Minoru Sakamoto · Yoshiyuki Suzuki · Fuminaka Sugauchi · Akihiro Tamori · Sei Kakinnuma · Kentaro Matsuura · Namiki Izumi

Received: 16 February 2011 / Accepted: 2 April 2011 / Published online: 17 June 2011

Abstract
Background This study aimed to develop a model to predict the development of severe anemia during pegylated interferon alpha-2b plus ribavirin combination therapy.
Methods Data were collected from 1081 genotype 1b chronic hepatitis C patients who were treated at 6 hospitals in Japan. These patients were randomly assigned to a model-building group ($n = 691$) or an internal validation group ($n = 390$). Factors predictive of severe anemia (hemoglobin, Hb < 8.5 g/dl) were explored using data-mining analysis.
Results Hb values at baseline, creatinine clearance (Ccr), and an Hb concentration decline by 2 g/dl at week 2 were used to build a decision-tree model, in which the patients were divided into 5 subgroups based on variable rates of severe anemia ranging from 0.4 to 11.8%. The reproducibility of the model was confirmed by the internal validation group ($r^2 = 0.96$). The probability of severe anemia was high in patients whose Hb value was <14 g/dl before treatment (6.5%), especially (a) in those whose Ccr was <80 ml/min (11.8%) and (b) those whose Ccr was ≥80 ml/min but whose Hb concentration decline at week 2 was ≥2 g/dl (11.5%). The probability of severe anemia was low in the other patients (0.4–2.5%).
Conclusions The decision-tree model that included Hb values at baseline, Ccr, and an Hb concentration decline by

N. Hiramatsu and M. Kurosaki contributed equally to this work.

N. Hiramatsu
Department of Gastroenterology and Hepatology,
Osaka University Graduate School of Medicine,
Osaka, Japan

M. Kurosaki · N. Izumi (✉)
Division of Gastroenterology and Hepatology,
Musashino Red Cross Hospital,
1-26-1 Kyonan-cho, Musashino,
Tokyo 180-8610, Japan
e-mail: nizumi@musashino.jrc.or.jp

N. Sakamoto · S. Kakinnuma
Department of Gastroenterology and Hepatology,
Tokyo Medical and Dental University, Tokyo, Japan

M. Iwasaki
Department of Computer and Information Science,
Seikei University, Tokyo, Japan

M. Sakamoto
First Department of Internal Medicine,
University of Yamanashi, Yamanashi, Japan

Y. Suzuki
Department of Hepatology, Toranomon Hospital,
Tokyo, Japan

F. Sugauchi
Department of Gastroenterology, Nagoya Koseiin
Medical Welfare Center, Nagoya, Japan

A. Tamori
Department of Hepatology,
Osaka City University Medical School, Osaka, Japan

K. Matsuura
Department of Gastroenterology and Metabolism,
Nagoya City University Graduate School of Medical Sciences,
Nagoya, Japan

2 g/dl at week 2 was useful for predicting the probability of severe anemia, and has the potential to support clinical decisions regarding early dose reduction of ribavirin.

Keywords Data mining · Decision tree · Severe anemia · Chronic hepatitis C · Pegylated interferon · Ribavirin

Abbreviations

Hb	Hemoglobin
PEG-IFN	Pegylated interferon
Ccr	Creatinine clearance
SVR	Sustained virological response
AFP	Alpha-fetoprotein

Introduction

The current standard therapy for chronic hepatitis C is 48 weeks of pegylated interferon (PEG-IFN) plus ribavirin [1]. Sustained virological response (SVR), defined as negative hepatitis C virus (HCV) RNA for 24 weeks after cessation of therapy, can be achieved by the current treatment regimen, but this outcome can be attained in only less than 50% of patients infected with genotype 1 HCV [2, 3]. Hemolytic anemia is a common side effect of ribavirin and is the major reason for dose reduction. Age, gender, baseline platelet level, baseline hemoglobin (Hb) level [4, 5], haptoglobin phenotype [6], drug dose [7], plasma concentration of ribavirin [8], apparent clearance of ribavirin (CL/F) [9], and an early decline in Hb concentration [10, 11] have been reported to contribute to ribavirin-induced anemia. Predicting the possibility of severe anemia before therapy or at the early phase of therapy can help modify ribavirin dosage, decrease the discontinuance rate for ribavirin, and raise the SVR rate.

Data mining is a method of predictive analysis that explores data to discover hidden patterns and relationships in highly complex datasets and enables the development of predictive models. Decision-tree analysis is a core component of data mining and predictive modeling [12], and it is utilized by decision makers in various business fields. Recent publications concerning decision-tree analysis in the medical field indicate its usefulness for defining prognostic factors in various diseases such as prostate cancer [13], diabetes [14], melanoma [15, 16], colorectal carcinoma [17, 18], and liver failure [19]. The results of decision-tree analysis are presented in the form of a flow chart, which is easy to use in clinical practice [20]. This analysis was also used to predict early virological response (undetectable HCV RNA within 12 weeks of therapy) and SVR to PEG-IFN plus ribavirin combination therapy in chronic hepatitis C [21–24]. In the present study, we used decision-tree analysis to explore before- and during-treatment predictors of severe anemia during PEG-IFN alpha-2b/ribavirin combination therapy and used a prediction algorithm to try to identify chronic hepatitis C patients who are likely to develop severe anemia.

Materials and methods

Patients

This multicenter retrospective cohort study was supported by the Japanese Ministry of Health, Labour and Welfare. Data were collected from 1081 chronic hepatitis C patients who were treated with PEG-IFN alpha-2b plus ribavirin at Osaka University, Musashino Red Cross Hospital, Toranomon Hospital, Tokyo Medical and Dental University, Nagoya City University, Yamanashi University, and their related hospitals. The inclusion criteria applied in the present study were as follows: (1) infection by genotype 1b, (2) HCV RNA $\geq$ 100 KIU/ml by quantitative PCR (Cobas Amplicor HCV Monitor v 2.0, Roche Diagnostic Systems, CA, USA), (3) lack of co-infection with hepatitis B virus or human immunodeficiency virus, (4) lack of other causes of liver diseases such as autoimmune hepatitis and primary biliary cirrhosis, and (5) completion of at least 12 weeks of therapy. Patients received PEG-IFN alpha-2b (1.5 g/kg) subcutaneously every week and were administered a weight-adjusted dose of ribavirin (600 mg for <60 kg, 800 mg for 60–80 kg, and 1000 mg for >80 kg). The dosage of ribavirin was reduced from 1000 to 600 mg, 800 to 600 mg, or 600 to 400 mg when the Hb concentration decreased to less than 10 g/dl, and was discontinued when the Hb concentration decreased to less than 8.5 g/dl, based on the recommendations in the package inserts. No patient received erythropoietin or blood transfusion for the treatment of anemia. Anemia with Hb $<$ 8.5 g/dl was defined as severe anemia in this study.

For the analysis, patients were randomly assigned to either the model-building ($n = 691$) group or the internal validation ($n = 390$) group. Consent was obtained from each patient. The study protocol conformed to the ethical guidelines of the Declaration of Helsinki and was approved by the institutional review committee. The baseline characteristics and representative laboratory test results are listed in Table 1. There were no significant differences between the clinical backgrounds of the two groups.

Laboratory tests

Blood samples were obtained before therapy and at least once every month during therapy, and were used for hematological tests, blood chemistry analyses, and determination of HCV RNA. Pretreatment levels of HCV RNA

Table 1 Comparison of clinical parameters of model-building and internal validation groups

	All patients ($N = 1081$)	Model building ($N = 691$)	Internal validation ($N = 390$)
Age (years)	55.6 ± 10.5	55.6 ± 10.8	55.6 ± 10.4
Gender (male/female)	612/469	393/298	219/171
Body mass index (kg/m^2)	23.2 ± 3.3	23.4 ± 3.8	23.1 ± 3.0
Creatinine (mg/dl)	0.73 ± 0.16	0.74 ± 0.17	0.73 ± 0.16
AST (IU/l)	62.0 ± 44.8	63.2 ± 48.6	61.4 ± 42.5
ALT(IU/l)	74.6 ± 56.1	75.4 ± 60.5	74.2 ± 53.5
GGT (IU/l)	58.6 ± 57.0	59.5 ± 58.5	58.0 ± 56.2
Albumin	4.0 ± 0.3	4.0 ± 0.3	4.0 ± 0.4
Total cholesterol	171.8 ± 31.7	171.5 ± 32.3	172.2 ± 30.8
HDL cholesterol	50.9 ± 14.5	51.1 ± 14.3	50.5 ± 15.0
LDL cholesterol	95.5 ± 27.7	96.1 ± 27.9	94.1 ± 27.2
Triglyceride	108.8 ± 55.4	107.8 ± 57.3	110.9 ± 51.7
Glucose	111.2 ± 39.0	111.7 ± 39.8	110.3 ± 37.6
Alpha-fetoprotein	14.5 ± 43.9	13.3 ± 37.7	16.8 ± 54.1
White blood cell count (/μl)	4946 ± 1427	4851 ± 1355	4999 ± 1464
Hemoglobin (g/dl)	14.2 ± 1.4	14.2 ± 1.4	14.2 ± 1.4
Platelets (10^9/mm^3)	166.1 ± 51.4	165.6 ± 51.7	166.4 ± 51.2
Ccr (ml/min)	95.1 ± 26.5	94.8 ± 25.9	95.4 ± 26.9
HCV RNA (KIU/ml)	1978 ± 1442	1937 ± 1382	2001 ± 1476
Fibrosis stage (F0–2/F3–4/ND)	695/148/238	454/85/152	241/63/86
Activity (A0–1/2–3/ND)	457/383/241	295/241/154	162/141/87
PEG-IFN alpha-2b dosage (μg/kg/body weight)	1.48 ± 0.13	1.49 ± 0.13	1.48 ± 0.13
Ribavirin dosage (600/800/1000 mg)	581/457/43	370/298/23	211/159/20
Decline of Hb at week 1	−0.2 ± 0.8	−0.2 ± 0.8	−0.2 ± 0.8
Decline of Hb at week 2	−1.2 ± 1.2	−1.2 ± 1.3	−1.2 ± 1.1
Decline of Hb at week 4	−2.3 ± 1.5	−2.2 ± 1.4	−2.4 ± 1.5
Decline of Hb at week 8	−2.8 ± 1.4	−2.8 ± 1.4	−2.7 ± 1.4

Data are expressed as median ± standard deviation unless otherwise indicated

AST aspartate aminotransferase, *ALT* alanine aminotransferase, *GGT* gamma-glutamyltransferase, *Ccr* creatinine clearance, *Hb* hemoglobin

were quantified by Cobas Amplicor (Roche Diagnostic Systems, CA, USA).

Database of variables and decision-tree analysis

A database of pretreatment variables was created containing 3 variables from hematological tests (Hb, white blood cells, and platelets), 11 variables from blood biochemical tests (creatinine, albumin, aspartate aminotransferase, alanine aminotransferase, gamma-glutamyl transpeptidase, total cholesterol, HDL cholesterol, LDL cholesterol, triglyceride, fasting blood glucose, and alpha-fetoprotein), creatinine clearance (Ccr), serum level of HCV RNA, liver histology (activity, fibrosis), 3 variables from patient characteristics (age, gender, and body mass index), 2 variables from therapeutic factors (PEG-IFN alpha-2b dosage, ribavirin dosage), and the level of decline of Hb concentration (at the end of 1, 2, 4, and 8 weeks from the start of treatment). Ccr levels were calculated using the Cockcroft–Gault formula [25]. Variables with data deficiency of greater than 15% were not included in the decision-tree analysis. Data deficiency was 21% in liver histology (activity, fibrosis) and 16% in the level of decline of Hb concentration at the end of 1 week. Accordingly, these variables were excluded from the database.

On the basis of this database, we implemented the recursive partitioning analysis algorithm referred to as the decision-tree analysis algorithm [26] to define subgroups of patients with respect to the possibility of severe anemia. The data-mining software used was IBM SPSS Modeler 13 (IBM SPSS Inc, Chicago, IL, USA), as reported previously [21–24]. In brief, the software searched the patient population for the most significant variables and cutoffs to be used for dividing the total population into 2 subgroups, having different probabilities of severe anemia. Thereafter,

the analysis was repeated on all subgroups in the same manner until either no additional significant variable was detected or the sample size was less than 20.

For other statistical analyses, including multivariable analysis, IBM SPSS Statistics software v.15.0 (IBM SPSS Inc, Chicago, IL, USA) was used. Differences in proportions were tested by the chi-squared test. Differences in continuous variables were compared by Student's t test. For univariate and multivariate analyses, logistic regression analysis was used to predict ribavirin-induced severe anemia. A value of $P < 0.05$ (two-tailed) was considered to indicate significance.

Results

Decision-tree analysis

Decision-tree analysis was carried out on the data of the model-building group using 27 variables, as described above. The analysis automatically selected 3 predictive variables to produce a total of 5 patient subgroups to build the decision tree (Fig. 1). Baseline Hb was selected as the first splitting variable, with an optimal cutoff of 14 g/dl. The possibility of severe anemia was 6.5% for patients with Hb levels <14 g/dl compared to 1.0% for patients with Hb levels ≥14 g/dl. Among patients with Hb ≥ 14 g/dl, the level of decline of Hb at the end of 2 weeks from the start of treatment, with an optimal cutoff of 2 g/dl, was selected as the second splitting variable. Patients with lower decline levels had a lower probability of developing severe anemia [<2 g/dl (group A) 0.4% vs. ≥2 g/dl (group B) 2.5%]. Among patients whose Hb was less than 14 g/dl, Ccr was selected as the second splitting variable, with an optimal cutoff of 80 ml/min. Patients with higher Ccr levels had a lower probability of developing severe anemia [≥80 ml/min, 2.4% vs. <80 ml/min (group C) 11.8%]. Among patients with a Ccr ≥ 80 ml/min, the level of decline of Hb at the end of 2 weeks from the start of the treatment was selected as the third splitting variable, with an optimal cutoff of 2 g/dl. Patients with lower decline levels had a lower probability of developing severe anemia [<2 g/dl (group D) 1.4% vs. ≥2 g/dl (group E) 11.5%].

The probabilities of severe anemia for the 5 subgroups derived by this process were highly variable. The subgroup of patients with higher Hb levels (≥14 g/dl) (groups A and B) had a low probability of developing severe anemia (0.4–2.5%). Also, the subgroup of patients with lower Hb (<14 g/dl) but with a higher Ccr (≥80 ml/min) and lower Hb decline levels at the end of 2 weeks from the start of the treatment (<2 g/dl) (group D) showed a low probability of developing severe anemia (1.4%). On the other hand, the

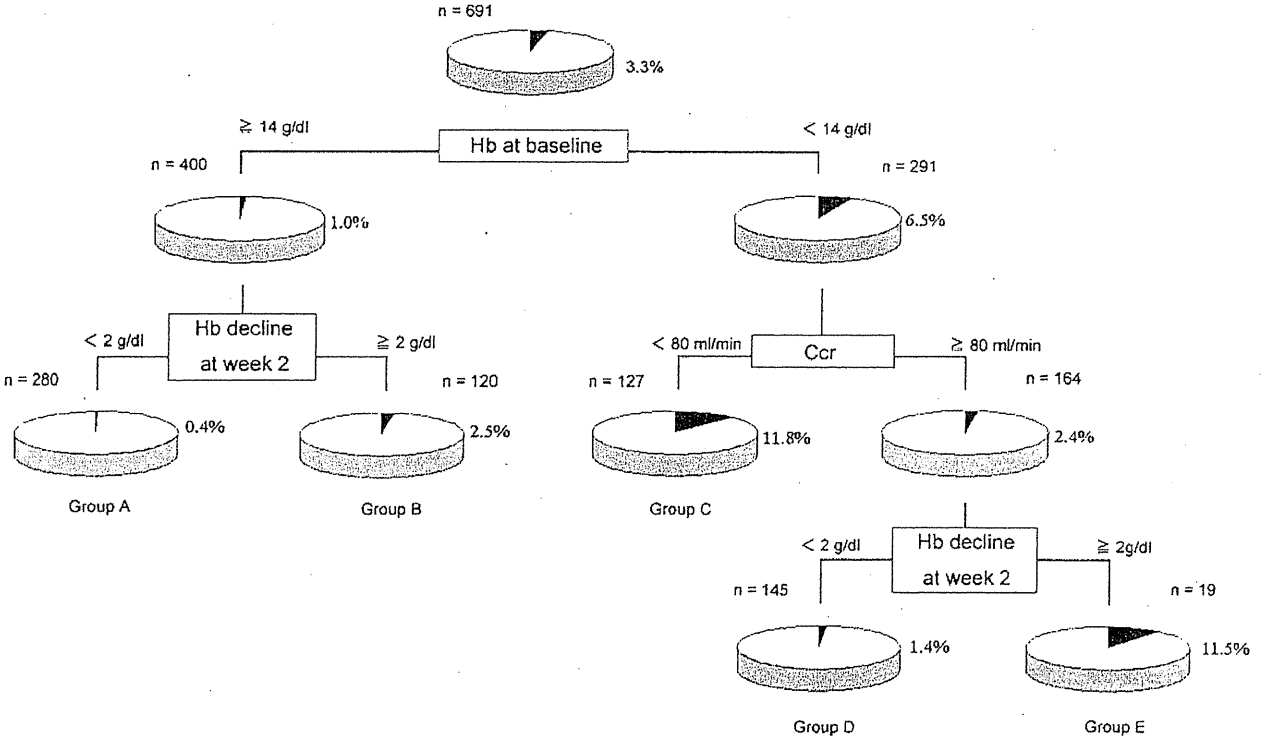

Fig. 1 Decision-tree analysis. *Boxes* indicate the splitting factors and the cutoff value for the split. *Pie charts* indicate the rate of severe anemia (Hb < 8.5 g/dl) for each group. Terminal groups classified by the analysis were labeled from A to E. *Hb* hemoglobin, *Ccr* creatine clearance

subgroup of patients with lower Hb (<14 g/dl) and lower Ccr (<80 ml/min) (group C) levels showed the highest probability of severe anemia (11.8%). Also, the subgroup of patients with lower Hb levels (<14 g/dl), higher Ccr (≥80 ml/min), and higher Hb decline levels at the end of 2 weeks from the start of treatment (≥2 g/dl) (group E) showed a high probability of developing severe anemia (11.5%).

Validation of the decision tree

The results of the decision tree were validated with the dataset of the internal validation group, which was independent of the model-building group dataset. Each patient in the validation group was allocated to groups A–E using the flow chart form of the decision tree. The rates of severe anemia (Hb < 8.5 g/dl) were 0.6% for group A, 3.0% for group B, 16.9% for group C, 2.3% for group D, and 11.0% for group E. The rates of severe anemia for each subgroup of patients were closely correlated between the model-building group and the internal validation group ($r^2 = 0.96$) (Fig. 2).

The efficiency and stability of the decision-tree model were validated using the discrimination efficiency curve (Fig. 3). The subgroups were sorted according to the order of incidence rate of severe anemia and validated using the correlation between the cumulative cases (%) and the cumulative incidence of severe anemia (%). The curve of the model-building group was located at the left upper part compared with the standard curve, indicating that the discrimination efficiency was high. Furthermore, the curve of the model-building group was extremely similar to the curve of the internal validation group, indicating that the stability was high.

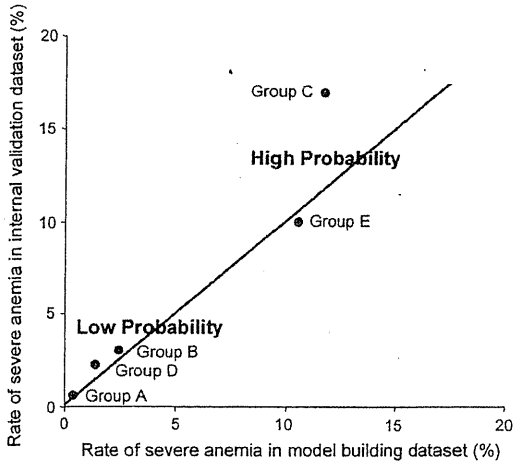

Fig. 2 Validation of decision-tree analysis with the internal validation dataset: subgroup-stratified comparison of the rate of severe anemia. The rate of severe anemia in each subgroup was plotted. The *X*-axis represents the model-building dataset and the *Y*-axis represents the internal validation dataset. There was a close correlation between the model-building and internal validation datasets ($r^2 = 0.96$)

Factors associated with severe anemia determined by multivariate logistic regression analysis

We also explored the factors associated with severe anemia using standard statistical analysis. By univariable analysis, age, creatinine, Hb, Ccr, fibrosis stage, and decline of Hb at 2, 4, and 8 weeks from the start of treatment were found to be associated with severe anemia (Table 2) and the odds ratio for these factors were 1.06, 9.61, 0.47, 0.95, 3.14, 0.76, 0.70, and 0.68, respectively, by univariable logistic regression analysis (Table 3). By multivariate analysis, Hb, Ccr, and decline of Hb at 2 weeks from the start of treatment were found to be independently associated with severe anemia (Table 3). Fibrosis was not included in the multivariable analysis because data were not available for 238 patients. Creatinine was not included in the multivariable analysis because creatinine and Ccr were confounding factors due to their close correlation. Decline of Hb at week 2, 4, and 8 were also closely correlated. We selected decline of Hb at week 2 in the multivariable analysis because we think that variables at earlier time points may be more useful in clinical use. As a result, decision-tree and multivariable logistic regression analyses identified the same factors for prediction of severe anemia.

Discussion

Hemolytic anemia, a major common side effect of ribavirin treatment, is one of the most important adverse effects of PEG-IFN and ribavirin combination treatment. Therefore, before- and during-treatment prediction of the likelihood of severe anemia can be very useful for physicians to support clinical decisions concerning the dose reduction of ribavirin. Reducing the dose of ribavirin has been shown to affect the HCV RNA negativity [27], and the discontinuation of ribavirin has been reported to lead to a marked decrease of SVR [9]. Therefore, averting ribavirin discontinuance, even if its dose must be reduced, can lead to an improvement in the SVR rate. It is important to identify patients prone to develop severe anemia leading to ribavirin dose reduction or discontinuance in the early phase of treatment.

Using decision-tree analysis, we constructed a simple model for predicting the incidence of severe anemia during therapy. The analysis highlighted 3 variables relevant to virological response: Hb, Ccr, and the decline of Hb concentration by 2 g/dl at the end of the 2 weeks from the start of treatment. Classification based on these variables identified subgroups of genotype 1b chronic hepatitis C patients with high probabilities of developing severe anemia. The

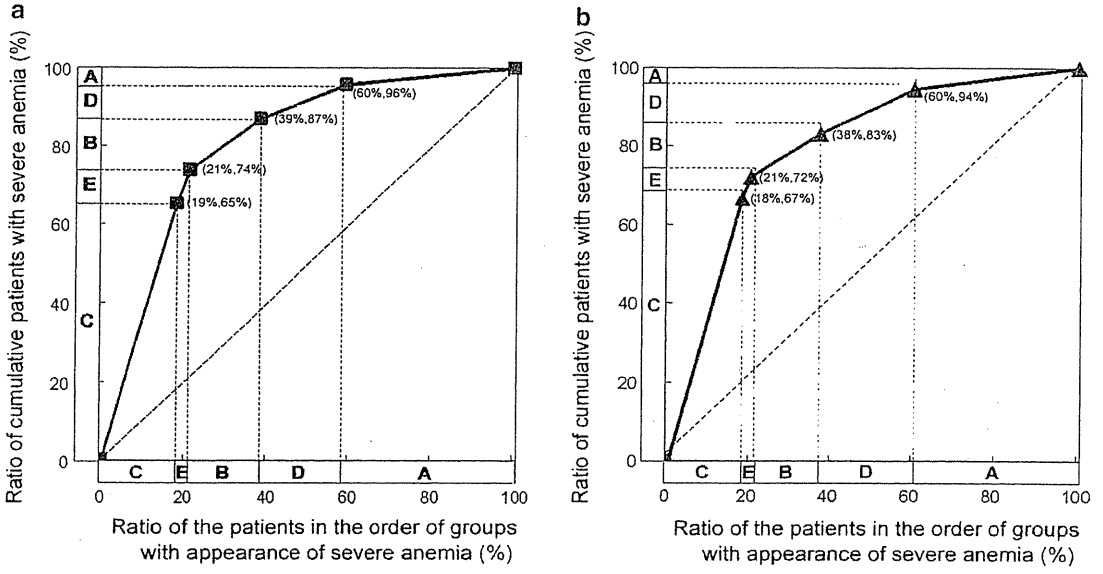

Fig. 3 Validation of the efficiency and stability by the discrimination efficiency curve. **a** Model-building group and **b** internal validation group. The groups were sorted in the order of incidence rate of severe anemia and validated using the correlation between cumulative cases (%) and the cumulative incidence of severe anemia (%). The *X*-axis represents the ratio of patients in the order of groups predicting the development of anemia and the *Y*-axis represents the cumulative patients suffering from severe anemia. The discrimination efficiency and stability of the curve of the model-building group were high

reproducibility of the model was confirmed with the internal validation dataset. An advantage of decision-tree analysis over traditional regression models is that the decision-tree model is user-intuitive and can be readily interpreted by medical professionals without the need for any specific knowledge of statistics. Patients can be allocated to specific subgroups based on a defined rate of severe anemia simply by following the flow chart format. Using this model, an estimate of the incidence of severe anemia can be obtained rapidly, which may facilitate clinical decision making for the reduction of ribavirin dosage. Thus, this model could be readily applicable for clinical practice.

According to the results of the decision tree, patients were categorized into 2 groups. The rates of severe anemia were 0.4–2.5% for the low probability group and 11.5–11.8% for the high probability group. For example, patients in the high probability group may be the most suitable candidates for dose reduction of ribavirin. Decision-tree analysis revealed that the high probability groups are patient groups with lower Hb (<14 g/dl) and lower Ccr (<80 ml/min) levels (group C) and patient groups with lower Hb (<14 g/dl), higher Ccr (≥80 ml/min), and higher Hb decline levels at 2 weeks from the start of treatment (≥2 g/dl) (group E). In particular, groups C and A were shown to be clinically significant in Fig. 3; group C includes the majority of patients suffering from severe anemia (65% in the model-building group and 67% in the internal validation group) and the very steep tilt angle of the group C slope means that group C patients have a very high probability of developing severe anemia. On the other hand, group A includes a large number of patients (40% in the model-building group and 40% in the internal validation group), and the very gentle tilt angle of the group A slope implies that group A patients have a very low probability of developing severe anemia.

Predicting the progression of anemia is necessary to decide whether medication can be continued while minimizing the disadvantages of anemia. The apparent clearance of ribavirin (CL/F), which reflects its plasma concentration at 4 weeks after the start of combination therapy, has been used as a predictive factor for developing ribavirin-induced hemolytic anemia before the start of treatment [9, 10]. However, the use of CL/F is not practical for general clinicians, because the calculation of CL/F is complicated. We revealed that a decline of Hb concentration by 2 g/dl at 2 weeks from the start of treatment ("2 by 2" standard) is both sensitive and convenient for identifying patients at high risk for severe anemia [10, 11]. The present study using decision-tree analysis revealed that Hb decline at week 2 was a significant and independent predictor of severe anemia. When considered along with other predictive factors, decision-tree analysis enables more exact identification of the patients prone to severe anemia.

Recently, a genome-wide association technique was used to show that ITPA polymorphism affects ribavirin-induced anemia. Polymorphisms (rs 1127354 and rs 7270101) that cause ITPase deficiency are strongly

Table 2 Comparison of clinical parameters of patients with and without severe anemia

	Anemia ($N = 41$)	No anemia ($N = 1040$)	P value
Age (years)	61.0 ± 7.6	55.4 ± 10.6	0.001
Gender (male/female)	18/23	594/446	0.109
Body mass index (kg/m^2)	22.4 ± 2.9	23.2 ± 3.3	0.119
Creatinine (mg/dl)	0.79 ± 0.24	0.7 ± 0.16	0.011
AST (IU/L)	74.2 ± 62.9	61.6 ± 43.9	0.075
ALT (IU/L)	79.6 ± 68.7	74.5 ± 55.6	0.565
GGT (IU/L)	40.7 ± 31.0	59.2 ± 57.6	0.071
Albumin	3.9 ± 0.3	4.0 ± 0.3	0.260
Total cholesterol	177.1 ± 23.1	171.6 ± 32.0	0.258
HDL cholesterol	50.8 ± 8.0	50.9 ± 14.7	0.986
LDL cholesterol	93.6 ± 22.0	95.5 ± 27.9	0.717
Triglyceride	109.1 ± 45.0	108.8 ± 55.8	0.974
Glucose	114.1 ± 32.9	111.1 ± 39.2	0.738
Alpha-fetoprotein	29.0 ± 71.4	13.9 ± 42.5	0.229
White blood cell count (/μl)	4632 ± 1828	4958 ± 1408	0.152
Hemoglobin (g/dl)	12.9 ± 1.3	14.2 ± 1.4	0.0001
Platelets (10^9/mm^3)	152.1 ± 51.7	166.6 ± 51.3	0.075
Ccr (ml/min)	75.4 ± 23.6	95.9 ± 26.4	0.0001
HCV RNA (KIU/ml)	1807 ± 1456	1985 ± 1442	0.438
Fibrosis stage (F0–2/F3–4/ND)	18/10/13	677/138/225	0.019
Activity (A0–1/2–3/ND)	14/14/13	443/369/228	0.701
PEG-IFN alpha-2b dosage	1.50 ± 0.13	1.48 ± 0.13	0.260
Ribavirin dosage (600/800/1000 mg)	28/13/0	553/444/43	0.105
Decline of hemoglobin at week 1	−0.2 ± 1.1	−0.2 ± 0.8	0.644
Decline of hemoglobin at week 2	−1.6 ± 1.9	−1.2 ± 1.2	0.022
Decline of hemoglobin at week 4	−3.0 ± 1.7	−2.2 ± 1.4	0.005
Decline of hemoglobin at week 8	−3.6 ± 1.6	−2.7 ± 1.4	0.003

Data are expressed as median ± standard deviation unless otherwise indicated

AST aspartate aminotransferase, *ALT* alanine aminotransferase, *GGT* gamma-glutamyltransferase, *Ccr* creatinine clearance, *Hb* hemoglobin

Table 3 Univariable and multivariable logistic regression analysis of factors associated with severe anemia

	Univariable analysis			Multivariable analysis		
	Odds	95% CI	P value	Odds	95% CI	P value
Age (years)	1.06	1.03–1.11	<0.0001	1.02	0.96–1.08	0.984
Creatinine (mg/dl)	9.61	1.91–48.4	0.006	–	–	–
Hb (g/dl)	0.47	0.37–0.59	<0.0001	0.40	0.29–0.55	<0.0001
Ccr (ml/min)	0.95	0.94–0.97	<0.0001	0.97	0.95–0.99	0.012
Fibrosis (F3–4)	3.14	1.49–6.60	0.003	–	–	–
Decline of Hb at week 2	0.76	0.61–0.95	0.017	0.54	0.39–0.74	0.0001
Decline of Hb at week 4	0.70	0.57–0.87	0.001	–	–	–
Decline of Hb at week 8	0.68	0.55–0.85	0.001	–	–	–

Hb hemoglobin, *Ccr* creatinine clearance

associated with protection from ribavirin-induced hemolytic anemia and with a lesser need for ribavirin dose reduction [28–30]. These polymorphisms are very valuable, but the indication for treatment is determined not by them but by viral genotypes or by *IL28B* variations. The present decision tree, which involves a factor attained after initiation of PEG-IFN plus ribavirin therapy, i.e., Hb decline at week 2, is useful for selecting the best regimen, and can be easily used by general clinicians.

What is unique to the present study is the visualization of the probability of severe anemia by combining factors and its high reproducibility, as revealed by high-quality validation of the internal validation dataset that was completely independent of the model-building dataset. The

factors used in the decision-tree model were clinical parameters that are readily available through the usual work-up of patients. This model can be immediately applied to clinical practice without imposing any cost for additional examinations.

A potential limitation of the present study is that data-mining analysis has an intrinsic risk of showing relationships that are relevant to the original dataset but are not reproducible across different populations. Although internal validation showed that our model had high reproducibility, we recognize that further validation using a larger external validation cohort, especially in populations other than Japanese, is necessary to verify the reliability of our model.

In conclusion, we built the decision-tree model for predicting severe anemia caused by PEG-IFN alpha-2b plus ribavirin combination therapy in chronic hepatitis C with genotype 1b and high viral load. Because this decision-tree model was composed of simple variables, it can be easily applied to clinical practice. This model may have the potential to support decisions concerning ribavirin dose reduction during PEG-IFN alpha-2b plus ribavirin combination therapy and contribute to increasing the rate of SVR.

Acknowledgments This study was supported by a grant-in-aid from the Ministry of Health, Labour and Welfare of Japan (H20-kannen-006).

Conflict of interest All authors have no financial relationship relevant to this study to disclose.

References

1. Strader DB, Wright T, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C. Hepatology. 2004;39(4):1147–71.
2. Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL Jr, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med. 2002;347(13):975–82.
3. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet. 2001;358(9286):958–65.
4. Nomura H, Tanimoto H, Kajiwara E, Shimono J, Maruyama T, Yamashita N, et al. Factors contributing to ribavirin-induced anemia. J Gastroenterol Hepatol. 2004;19:1312–7.
5. Takaki S, Tsubota A, Hosaka T, Akuta N, Someya T, Kobayashi M, et al. Factors contributing to ribavirin dose reduction due to anemia during interferon alfa2b and ribavirin combination therapy for chronic hepatitis C. J Gastroenterol. 2004;39:668–73.
6. Van Vlierbergh H, Delanghe JR, De Vos M, Leroux-Roel G, BASL Steering Committee. Factors influencing ribavirin-induced hemolysis. J Hepatol. 2001;34:911–6.
7. Jen JF, Glue P, Gupta S, Zambas D, Hajian G. Population pharmacokinetic and pharmacodynamic analysis of ribavirin in patients with chronic hepatitis C. Ther Drug Monit. 2000;22:555–65.
8. Lindhal K, Schvarcz R, Bruchfeld A, Ståhle L. Evidence that plasma concentration rather than dose per kilogram body weight predicts ribavirin-induced anemia. J Viral Hepat. 2004;11:84–7.
9. Kamar N, Chatelut E, Manolis E, Lafont T, Izopet J, Rostaing L. Ribavirin pharmacokinetics in renal and liver transplant patients: evidence that it depends on renal function. Am J Kidney Dis. 2004;43:140–6.
10. Hiramatsu N, Kurashige N, Oze T, Takehara T, Tamura S, Kasahara A, et al. Early decline of hemoglobin can predict progression of hemolytic anemia during pegylated interferon and ribavirin combination therapy in patients with chronic hepatitis C. Hepatol Res. 2008;38:52–9.
11. Oze T, Hiramatsu N, Kurashige N, Tsuda N, Yakushijin T, Kanto T, et al. Early decline of hemoglobin correlates with progression of ribavirin-induced hemolytic anemia during interferon plus ribavirin combination therapy in patients with chronic hepatitis C. J Gastroenterol. 2006;41:862–87.
12. Breiman LJH, Friedman RA, Olshen CJ, Stone CM. Classification and regression trees. Monterey: Wadsworth; 1980.
13. Garzotto M, Beer TM, Hudson RG, Peters L, Hsieh YC, Barrera E, et al. Improved detection of prostate cancer using classification and regression tree analysis. J Clin Oncol. 2005;23:4322–9.
14. Miyaki K, Takei I, Watanabe K, Nakashima H, Omae K. Novel statistical classification model of type 2 diabetes mellitus patients for tailor-made prevention using data mining algorithm. J Epidemiol. 2002;12:243–8.
15. Averbook BJ, Fu P, Rao JS, Mansour EG. A long-term analysis of 1018 patients with melanoma by classic Cox regression and tree-structured survival analysis at a major referral center. Surgery. 2002;132:589–602.
16. Leiter U, Buettner PG, Eigentler TK, Garbe C. Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the German Dermatological Society. J Clin Oncol. 2004;22:3660–7.
17. Valera VA, Walter BA, Yokoyama N, Koyama Y, Iiai T, Okamoto H, et al. Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. Ann Surg Oncol. 2007;14:34–40.
18. Zlobec I, Steele R, Nigam N, Compton CC. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. Clin Cancer Res. 2005;11:5440–3.
19. Baquerizo A, Anselmo D, Shackleton C, Chen TW, Cao C, Weaver M, et al. Phosphorus as an early predictive factor in patients with acute liver failure. Transplantation. 2003;75:2007–14.
20. LeBlanc M, Crowley J. A review of tree-based prognostic models. Cancer Treat Res. 1995;75:113–24.
21. Kurosaki M, Matsunaga K, Hirayama I, Tanaka T, Sato M, Yasui Y, et al. A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. Hepatol Res. 2010;40:251–60.
22. Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, Hiramatsu N, et al. Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis. J Gastroenterol. 2011;46:401–9.
23. Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, Hiramatsu N, et al. Sequences in the interferon sensitivity-determining region and core region of hepatitis C virus impact pretreatment prediction of response to peg-interferon plus ribavirin: data mining analysis. J Med Virol. 2011;83:445–52.
24. Kurosaki M, Tanaka Y, Nishida N, Sakamoto N, Enomoto N, Honda M, et al. Pre-treatment prediction of response to

pegylated-interferon plus ribavirin for chronic hepatitis C using genetic polymorphism in *IL28B* and viral factors. J Hepatol. 2011;54:439–48.
25. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. Nephron. 1976;16:31–41.
26. Segal MR, Bloch DA. A comparison of estimated proportional hazards models and regression trees. Stat Med. 1989;8:539–50.
27. Hiramatsu N, Oze T, Yakushijin T, Inoue Y, Igura T, Mochizuki K, et al. Ribavirin dose reduction raises relapse rate dose-dependently in genotype 1 patients with hepatitis C responding to pegylated interferon alfa-2b plus ribavirin. J Viral Hepat. 2009; 16:586–94.
28. Fellay J, Thompson AJ, Ge D, Gumbs CE, Urban TJ, Shianna KV, et al. ITPA gene variants protect against anaemia in patients treated for chronic hepatits C. Nature. 2010;464:405–8.
29. Thompson AJ, Fellay J, Patel K, Tillmann HL, Naggie S, Ge D, et al. Variants in the ITPA gene protect against ribavirin-induced hemolytic anemia and decrease the need for ribavirin dose reduction. Gastroenterology. 2010;139:1181–9.
30. Ochi H, Maekawa T, Abe H, Hayashida Y, Nakano R, Kubo M, et al. ITPA polymorphism affects ribavirin-induced anemia and outcomes of therapy-A genome-wide study of Japanese HCV virus patients. Gastroenterology. 2010;139:1190–7.