Received October 19, 2010. Accepted March 18, 2011.

Reprint requests

Address requests for reprints to: Shuchi Kaneko, MD, PhD, Department of Gastroenterology, Graduate School of Medicine, Kanazawa University, Takara-Machi 13-1, Kanazawa 920-8641, Japan. e-mail: skaneko@m-kanazawa.jp; e-mail: fax: (81) 76-234-4250.

Acknowledgments

Participating investigators from the Hokuriku Liver Study Group are listed in Appendix A.

The authors thank Mina Nishiyama and Yuki Hatayama for excellent technical assistance.

Conflicts of interest

The authors disclose no conflicts.

Supplementary Materials and Methods

Plasma Amino Acid Analysis

Plasma sample amino acid concentrations were measured by high-performance liquid chromatography-electrospray ionization-mass spectrometry followed by derivatization.[1] An MSQ Plus LC/MS system (Thermo Fischer Scientific, Waltham, MA) equipped with an electrospray ionization source was used in positive ionization mode for selected ion monitoring. Xcalibur version 1.4 SR1 software (Thermo Fischer Scientific, Yokohama, Japan) was used for data collection and processing. The high-performance liquid chromatography separation system consisted of an L-2100 pump, L-2200 autosampler, and L-2300 column oven (Hitachi High-Technologies Corporation, Tokyo, Japan). A Wakosil-II 3C8-100HG column (100, 2.1, 3 mm; Wako Pure Chemical Industries, Osaka, Japan) was used for the separation, and the mobile phase consisted of eluent A (25-mmol/L ammonium formate in water, pH 6.0) and eluent B (water:acetonitrile = 40:60).

Western Blotting

The expression of HCV core protein, Socs3, Foxo3a, phospho-Foxo3a (Ser253) (pFoxo3a), STAT1, pSTAT1 (Tyr701), S6K, pS6K, p-mTOR (Ser2448), Raptor, and β-actin were evaluated with mouse anti-core (Affinity BioReagents, Golden, CO), mouse anti-Socs3 (Santa Cruz Biotechnology, Santa Cruz, CA), rabbit anti-Foxo3a, rabbit anti-β-actin (Sigma-Aldrich, St Louis, MO), rabbit anti-phospho-Foxo3a (Ser253), rabbit anti-STAT1, rabbit anti-p-STAT1 (Tyr701), rabbit anti-p70 S6K, rabbit anti-pS6K, rabbit anti-p-mTOR (Ser2448), and rabbit anti-Raptor (Cell Signaling Technology, Beverly, MA), respectively. Densitometric analysis was conducted directly on the blotted membrane using a charge coupled device camera system (LAS-3000 Mini; Fujifilm, Tokyo, Japan) and Scion Image software (Frederick, MD).

Primer Sequences for PCR and siRNA

Primer sequences for PCR and siRNA were as follows: 2′5′OAS: forward 5′- CTC AGA AAT ACC CCA GCC AAA TC-3′, reverse 5′-GTG GTG AGA GGA CTG AGG AA-3′; Socs3: forward 5′-TAC CAC CTG AGT CTC CAG CTT CTC-3′, reverse 5′-CCT GGC AGT TCT CAT TAG TTC AGC ATT C-3′; Foxo3a: forward 5′-TGC TGT ATG CAA GAA CTT TCC AGT AGC AG-3′, reverse 5′-ACT CTA GCC CCC ATG CTA CTA GTG-3′; glyceraldehyde-3-phosphate dehydrogenase: forward 5′-GAA GGT GAA GGT CGG AGT-3′, reverse 5′-GAA GAT GGT GAT GGG ATT TC-3′, siFoxo3a (SASI_Hs01_00119127; Sigma) sense: 5′-GAA UGA UGG GCU GAC UGA AdTdT-3′, antisense: 5′-UUC AGU CAG CCC AUC AUU CdTdT-3′. Small interfering Raptor was purchased as part of KIAA1303 siGENOME SMART pool siRNA reagents from Dharmacon, Inc (Lafayette, CO).

Construction of ISRE-Luc Reporter and FBEmut-luc Reporter Plasmids

Oligonucleotides containing the ISRE tandem repeat sequence (sense 5′-TCG AGA ACT GAA ACT GAA ACT GAA ACT GAA ACT GAA ACT GAA ACT GAA ACT GAA A-3′, antisense 5′-AGC TTT TCA GTT TCA GTT TCA GTT TCA GTT TCA GTT TCA GTT TCA GTT TCA GTT C-3′, consensus 5′-GAA Ann GAA ACT-3′) were annealed, and integrated into Xho I and Hind III sites of the pGL4.23 luciferase vector (Promega). The human Socs3 promoter region (-109/+217) was amplified by genomic PCR using specific primers (forward, 5′-TGC TGC GAG TAG TGA CTA AAC ATT ACA AG-3′ and reverse, 5′-CCG TGA AGT CCA CAA AGG AGC CTT C-3′) and cloned into the EcoR V site of the pGL4.10-luc2 reporter vector (Promega). The Socs3 FBE mutant reporter vector was created by substituting 2 adenines in the putative FBE with guanines (wild-type sequence 5′-CTAAACA-3′, mutated sequence 5′-CTGAGCA-3′).

ChIP Assay

For the ChIP assay using the anti-ISGF3γ antibody, 1×10^6 Huh-7 cells were treated with IFN-alfa (0 or 100 U/mL) and BCAA (2 mmol/L) in low-amino-acid medium for 6 hours. For ChIP using the anti-Foxo3a antibody, 1×10^6 Huh-7 cells were cultured in low-amino-acid medium for 24 hours.

Cells were cross-linked with 1% formaldehyde in PBS for 10 minutes at 37°C, and the reaction was stopped with 250 mmol/L glycine for 10 minutes. Cells were suspended in sodium dodecyl sulfate-lysis buffer (1% sodium dodecyl sulfate, 10 mmol/L ethylenediaminetetraacetic acid [EDTA], 50 mmol/L Tris-HCl [pH 8.1]), complete protease inhibitor cocktail (Roche Applied Science), and incubated for 30 minutes at 10°C. Cell lysate was sonicated with Bioruptor (Cosmo Bio, Tokyo, Japan) to obtain chromatin fragments and diluted 10-fold in ChIP dilution buffer (0.01% sodium dodecyl sulfate, 1.1% Triton-X 100, 1.2 mmol/L EDTA, 16.7 mmol/L Tris-HCl [pH 8.1], 150 mmol/L NaCl, complete protease inhibitor cocktail). Chromatin fragments were incubated with 2 μg ISGF3γ antibody (Santa Cruz Biotechnology), 2 μg Foxo3a antibody (H-144; Santa Cruz Biotechnology), or normal rabbit immunoglobulin G for 18 hours at 4°C. Dynabeads (30 μL) protein G (Invitrogen) was added and incubated for 1 hour at 4°C. The beads were washed with low-salt-wash buffer (0.1% sodium dodecyl sulfate, 1% Triton-X 100, 2.0 mmol/L EDTA, 20 mmol/L Tris-HCl [pH 8.1], 150 mmol/L NaCl), high-salt-wash buffer (0.1% sodium dodecyl sulfate, 1% Triton-X 100, 2.0 mmol/L EDTA, 20 mmol/L Tris-HCl [pH 8.1], 500 mmol/L NaCl), LiCl wash buffer (250 mmol/L LiCl, 1% NP-40, 1% de-

oxycholate, 1.0 mmol/L EDTA, 1.0 mmol/L Tris-HCl [pH 8.1]) and Tris-EDTA buffer. Immunoprecipitated chromatin fragments were eluted with elution buffer (1% sodium dodecyl sulfate, 100 mmol/L NaHCO3, 10 mmol/L dithiothreitol), and reverse cross-linked by incubating for 6 hours at 65°C in elution buffer containing 200 mmol/L NaCl. DNA fragments were purified and quantified by real-time detection PCR with primers for putative ISRE in the 2′5′OAS promoter region (forward, 5′-AAA TGC ATT TCC AGA GCA GAG TTC AGA G-3′, reverse, 5′-GGG TAT TTC TGA GAT CCA TCA TTG ACA GG-3′) or putative FBE in the Socs3 promoter region (forward, 5′-TGC TGC GAG TAG TGA CTA AAC ATT ACA AG -3′, reverse, 5′-AGC GGA GCA GGG AGT CCA AGT C -3′). Values were normalized by the measurement of input DNA.

pH77S.3/GLuc2A

pH77S.2 is a modification of pH77S[2] containing an additional mutation within the E2 protein (N476D in the polyprotein) that promotes infectious virus yields from RNA-transfected cells (Yi et al, unpublished data). To monitor replication, the GLuc sequence, fused at its C terminus to the foot-and-mouth disease virus 2A autoprotease, was inserted between p7 and NS2 of pH77S.2 (Supplementary Figure 4). To insert the GLuc-coding sequence between p7 and NS2 in pH77S.2, followed by the foot-and-mouth disease virus 2A protein-coding sequence, Mlu I, EcoR V, and Spe I restriction sites were created between the p7 and NS2 coding sequences by site-directed mutagenesis. DNA coding for GLuc was subcloned into the Mlu I and EcoR V sites of the modified plasmid after PCR amplification using the primers: 5′- ATA ATA TTA CGC GTA *TGG GAG TCA AAG TTC TGT TTG CC*-3′ (sequence corresponding to the N-terminal GLuc is italicized and that corresponding to Mlu I is underlined) and 5′-ATA AAT AGAT ATC *GTC ACC ACC GGC CCC CTT GAT CTT*-3′ (C terminal GLuc is italicized and EcoR V is underlined). A DNA fragment encoding the 17 amino acids of the foot-and-mouth disease virus 2A protein was generated by annealing the following complementary oligonucleotides: 5′- ATA TGA TAT CAA CTT TGA CCT TCT CAA GTT GGC CGG CGA CGT CGA GTC CAA CCC AGG GCC CAC TAG CAT AT-3′ and 5′-ATA TGC TAG TGG GCC CTG GGT TGG ACT CGA CGT CGC CGG CCA ACT TGA GAA GGT CAA AGT TGA TAT CAT AT-3′ (underlined sequences indicate EcoR V and Spe I sites). The annealed oligonucleotides were digested by both restriction enzymes and the product inserted into the corresponding sites of pH77S.2 containing GLuc to generate pH77S.2/GLuc2A. Q41R is a cell-culture adaptive mutation within the NS3 protease domain of pH77S. Because it is not essential for production of infectious virus (Yi et al, unpublished data), pH77S.2 and pH77S.2/GLuc2A constructs underwent this mutation by site-directed mutagenesis of a PCR fragment spanning the Afe I and BsrG I sites to replace Gln_{41} with wild-type Arg. The resulting plasmids (pH77S.2/R41Q and pH77S.2/GLuc2A/R41Q) were redesignated pH77S.3 and pH77S.3/GLuc2A, respectively.[3,4] GLuc has several advantages over other luciferase reporter enzymes in that it is smaller and allows more sensitive detection than either firefly or Renilla luciferase.[3,4] In addition, a signal sequence directs its secretion into cell-culture media, allowing real-time dynamic measurements of GLuc expression without the need for cell lysis. H77S.3/GLuc2A RNA produces infectious virus, although with lower efficiency than H77S.3 RNA (10-fold less).

References

1. Shimbo K, Kubo S, Harada Y, et al. Automated precolumn derivatization system for analyzing physiological amino acids by liquid chromatography/mass spectrometry. Biomed Chromatogr 2009; 24:683–691.
2. Yi M, Villanueva RA, Thomas DL, et al. Production of infectious genotype 1a hepatitis C virus (Hutchinson strain) in cultured human hepatoma cells. Proc Natl Acad Sci U S A 2006;103:2310–2315.
3. Shetty S, Kim S, Shimakami T, et al. Hepatitis C virus genomic RNA dimerization is mediated via a kissing complex intermediate. RNA 2010;16:913–925.
4. Shimakami T, Welsch C, Yamane D, et al. Protease inhibitor-resistant hepatitis C virus mutants with reduced fitness from impaired production of infectious virus. Gastroenterology 2011; 140:667–675.

J Gastroenterol
DOI 10.1007/s00535-010-0322-5

ORIGINAL ARTICLE—LIVER, PANCREAS, AND BILIARY TRACT

Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis

Masayuki Kurosaki · Naoya Sakamoto · Manabu Iwasaki · Minoru Sakamoto · Yoshiyuki Suzuki · Naoki Hiramatsu · Fuminaka Sugauchi · Hiroshi Yatsuhashi · Namiki Izumi

Received: 22 June 2010 / Accepted: 21 August 2010

Abstract
Background This study aimed to develop a model for the pre-treatment prediction of sustained virological response (SVR) to peg-interferon plus ribavirin therapy in chronic hepatitis C.
Methods Data from 800 genotype 1b chronic hepatitis C patients with high viral load (>100,000 IU/ml) treated by peg-interferon plus ribavirin at 6 hospitals in Japan were randomly assigned to a model building ($n = 506$) or an internal validation ($n = 294$). Data from 524 patients treated at 29 hospitals in Japan were used for an external validation. Factors predictive of SVR were explored using data mining analysis.
Results Age (<50 years), alpha-fetoprotein (AFP) (<8 ng/mL), platelet count ($\geq 120 \times 10^9$/l), gamma-glutamyl-transferase (GGT) (<40 IU/l), and male gender were used to build the decision tree model, which divided patients into 7 subgroups with variable rates of SVR ranging from 22 to 77%. The reproducibility of the model was confirmed by the internal and external validation ($r^2 = 0.92$ and 0.93, respectively). When reconstructed into 3 groups, the rate of SVR was 75% for the high probability group, 44% for the intermediate probability group and 23% for the low probability group. Poor adherence to drugs lowered the rate of SVR in the low probability group, but not in the high probability group.

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

M. Kurosaki
e-mail: kurosaki@musashino.jrc.or.jp

N. Sakamoto
Department of Gastroenterology and Hepatology,
Tokyo Medical and Dental University, Tokyo, Japan
e-mail: nsakamoto.gast@tmd.ac.jp

M. Iwasaki
Department of Computer and Information Science,
Seikei University, Tokyo, Japan
e-mail: iwasaki@st.seikei.ac.jp

M. Sakamoto
First Department of Internal Medicine, University of Yamanashi,
Yamanashi, Japan
e-mail: msakamoto@yamanashi.ac.jp

Y. Suzuki
Department of Hepatology, Toranomon Hospital, Tokyo, Japan
e-mail: suzunari@interlink.or.jp

N. Hiramatsu
Department of Gastroenterology and Hepatology,
Osaka University Graduate School of Medicine,
Osaka, Japan
e-mail: hiramatsu@gh.med.osaka-u.ac.jp

F. Sugauchi
Department of Gastroenterology and Metabolism,
Nagoya City University Graduate School of Medical Sciences,
Nagoya, Japan
e-mail: fsugauch@med.nagoya-cu.ac.jp

H. Yatsuhashi
Clinical Research Center, National Nagasaki Medical Center,
Nagasaki, Japan
e-mail: yatsuhashi@nmc.hosp.go.jp

Published online: 10 September 2010

Conclusions A decision tree model that includes age, gender, AFP, platelet counts, and GGT is useful for predicting the probability of response to therapy with peg-interferon plus ribavirin and has the potential to support clinical decisions regarding the selection of patients for therapy.

Keywords Data mining · Decision tree · Alpha-fetoprotein · HCV · Peg-interferon

Introduction

The current standard therapy for genotype 1 chronic hepatitis C is 48 weeks of pegylated interferon (PEG-IFN) plus ribavirin (RBV) [1]. Sustained virological response (SVR), defined as undetectable HCVRNA post-treatment is regarded as a cure of chronic hepatitis C. However, the rate of SVR to this regimen is only 50% in patients with HCV genotype 1b and a high HCVRNA titer [2, 3]. Since PEG-IFN and RBV combination therapy is costly and accompanied by potential adverse effects, the ability to predict the possibility of SVR before therapy may significantly influence the selection of patients for therapy. A recent report revealed that single nucleotide polymorphisms located in the *IL28B* are strongly associated with a response to PEG-IFN plus RBV therapy [4–6]. Besides, the amino acid substitutions in the NS5A [7–9] or core region of HCV were also associated with response to therapy [10, 11]. Unfortunately, these host genetic and viral factors are not yet readily available for general application in actual clinical practice. Fibrosis of the liver is also an important predictor of response, but resources may be limited in some countries. Clinical and non-invasive parameters may be better suited for general practice, but there is no established means by which the likelihood of a response can be predicted prior to therapy.

Data mining is a method of predictive analysis that explores data, without setting the hypothesis, 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 fields of business. Recent publications on decision tree analysis 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 the analysis are presented as a tree structure, which is intuitive and facilitates the allocation of patients into subgroups by following the flow chart form [20]. We have recently reported the usefulness of decision tree analysis for the prediction of early virological response (undetectable HCVRNA within 12 weeks of therapy) to PEG-IFN and RBV combination therapy in chronic hepatitis C [21].

In the present study, we used decision tree analysis to explore baseline predictors of response to PEG-IFN/RBV therapy so that a pre-treatment algorithm could be created to discriminate chronic hepatitis C patients who are likely to respond to PEG-IFN/RBV therapy from those who are not. For the purpose of use in general practice, only clinical and non-invasive parameters were included in the analysis.

Materials and methods

Patients

This was a multicenter retrospective cohort study supported by the Japanese Ministry of Health, Labor and Welfare. Data were collected from a total of 800 chronic hepatitis C patients who received therapy for 48 weeks with PEG-IFN alpha-2b and RBV at Musashino Red Cross Hospital, Toranomon Hospital, Tokyo Medical and Dental University, Osaka University, Nagoya City University Graduate School of Medical Sciences, Yamanashi University, and their related hospitals. The inclusion criteria to be enrolled in this study were as follows (1) infection by genotype 1b, (2) HCVRNA higher than 100,000 IU/ml by quantitative PCR (Cobas Amplicor HCV Monitor v 2.0, Roche Diagnostic systems, CA), which is typically used for the definition of high viral load in Japan, (3) lack of co-infection with hepatitis B virus or human immunodeficiency virus, (4) lack of other causes of liver disease 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 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 1,000 mg for >80 kg), which is the recommended dosage in Japan. Patients who were treated for more than 49 weeks were not included in the study. For the analysis, patients were randomly assigned to either the model building ($n = 506$) or the internal validation ($n = 295$) 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. The overall rate of SVR was 47% in the model building set and 49% in the validation set. There were no significant differences in the clinical backgrounds between these 2 groups.

For external validation of the model, we collaborated with another study group supported by the Japanese Ministry of Health, Labor and Welfare. This multicenter study group consisted of 29 medical centers and hospitals

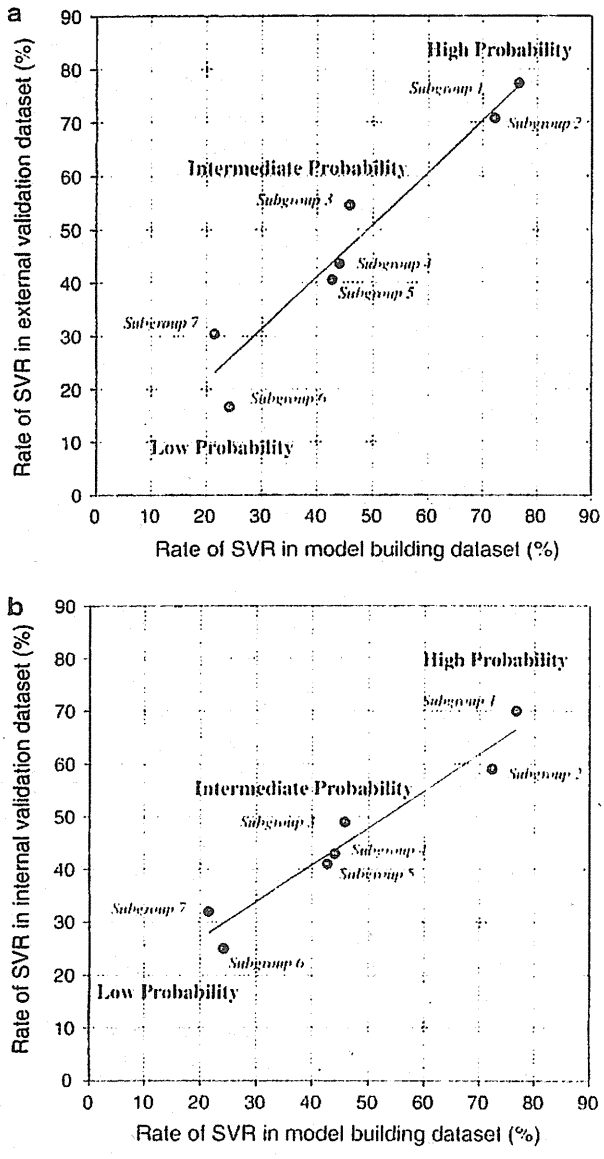

Fig. 2 Validation of the decision tree analysis by an internal and external validation dataset: subgroup-stratified comparison of the SVR rate. The rate of SVR in each subgroup was plotted. The *X* axis represents the model building, and the *Y* axis represents the validation datasets. a Internal validation and b external validation. There was a close correlation between the model building and the internal validation dataset (correlation coefficient $r^2 = 0.925$) and between the model building and the external validation dataset (correlation coefficient $r^2 = 0.936$)

original dataset used for model building. Each patient in the external validation set was allocated to subgroups 1–7 using the flow-chart form of the tree. The rates of SVR were 70% for subgroup 1, 59% for subgroup 2, 49% for subgroup 3, 43% for subgroup 4, 41% for subgroup 5, 25% for subgroup 6, and 32% for subgroup 7. The rates of SVR for each subgroup of patients were closely correlated

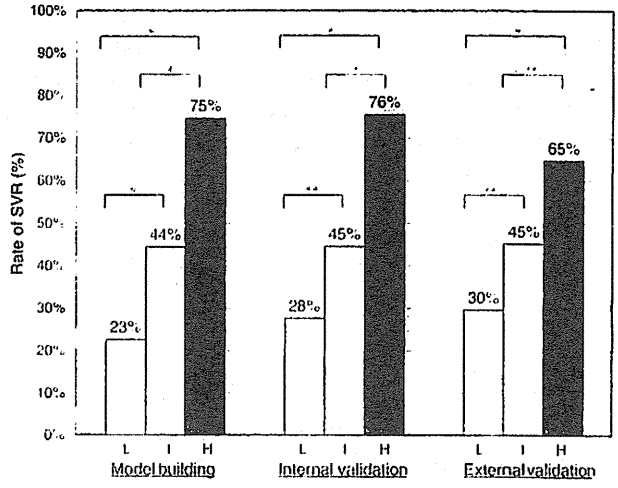

Fig. 3 Comparison of SVR rates between groups divided by the decision tree. The rate of SVR was compared among the 3 groups of patients divided by the decision tree analysis (*white*, *gray* and *black boxes*, indicating a low (*L*), intermediate (*I*) and high (*H*) probability group, respectively). The rate of SVR was significantly different among the 3 groups. $*p < 0.0001$, $**p < 0.001$

between the model-building dataset and the validation dataset ($r^2 = 0.936$) (Fig. 2b).

Construction of 3 groups according to the probability of SVR

Seven subgroups were reconstructed into 3 groups according to their predicted rates of SVR: the high probability group consisted of subgroups 1 and 2, the intermediate probability group consisted of subgroups 3, 4 and 5, and the low probability group consisted of subgroups 6 and 7. The rate of SVR was significantly different among the 3 groups (Fig. 3). The rate of SVR in the high probability group was consistently high: 75% for model building patients, 76% for internal validation patients and 65% for external validation patients. Conversely, the rate of SVR in the low probability group was consistently low: 23% for model building patients, 28% for internal validation patients and 30% for external validation patients. The rate of SVR in the intermediate probability group was 44% for model building patients, 45% for internal validation patients and 45% for external validation patients. Since 28–32% of patients were classified as high probability and 30–32% were classified as low probability, roughly 60% of patients were classified as having either a high or low probability of achieving SVR.

Effect of dose reductions of PEG-IFN and RBV on SVR

The cumulative dose of PEG-IFN and RBV was not included as a variable of analysis since the present study

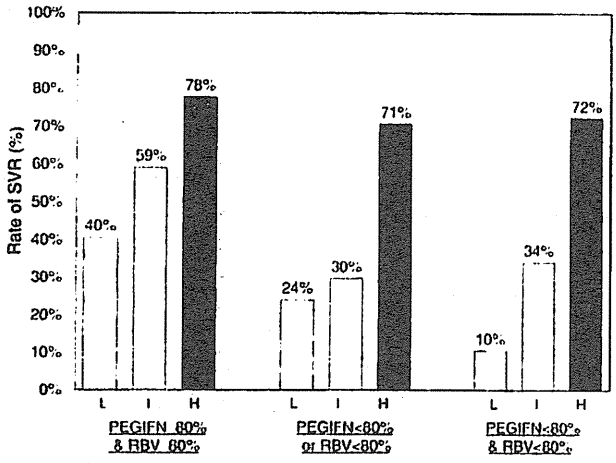

Fig. 4 Comparison of SVR rates among groups stratified by drug adherence. The 3 groups of patients divided by the decision tree analysis (*white*, *gray* and *black boxes* indicating a low (*L*), intermediate (*I*) and high (*H*) probability group, respectively) were further stratified according to the cumulative drug exposure of PEG-IFN and RBV. The good adherence group (≥80% planned dose of both PEG-IFN and RBV) had a higher rate of SVR compared with the poor adherence group (<80% planned dose of both PEG-IFN and RBV) in the low ($p = 0.0003$) and intermediate ($p = 0.007$) but not in the high probability group ($p = 0.53$)

aimed to develop a pre-treatment model for the prediction of response. To analyze the possible effect of drug reductions on the result of the decision tree analysis, 3 groups of patients divided by the decision tree analysis (low, intermediate and high probability group) were further stratified according to the cumulative drug exposure of PEG-IFN and RBV (Fig. 4). Even after adjustment for adherence, 3 groups of patients still had low, intermediate and high probability of achieving SVR, respectively. Of note, the good adherence group (≥80% planned dose of both PEG-IFN and RBV) had higher rates of SVR compared with the poor adherence group (<80% planned dose of both PEG-IFN and RBV) in the low ($p = 0.0003$) and intermediate ($p = 0.007$) probability group, but not in the high probability group ($p = 0.53$).

Factors associated with SVR by multivariate logistic regression analysis

We also explored the factors associated with SVR using a standard statistical analysis. By univariate analysis, age, gender, serum albumin, creatinine, alanine aminotransferase, GGT, red blood cell count, hemoglobin, hematocrit, platelet count and AFP were found to be associated with SVR (Table 2). HCVRNA load was not associated with SVR. By multivariate analysis, age, gender, GGT and platelet count were found to be independently associated with SVR (Table 3). Of note, AFP, which was selected as a significant predictor of response in the decision tree analysis, was not found to be an independent response predictor in the standard multivariate analysis. This indicates a unique feature of the decision tree analysis; i.e., it could identify significant predictors that specifically apply to selected patients, in this case patients younger than 50 years old.

Relationships between decision tree model and stage of fibrosis or HCV RNA load

Liver biopsy was performed in 664 patients. The distribution of fibrosis in three probability groups differed significantly. Advanced fibrosis (F3 or F4) was higher in the low probability group (39%) compared to the intermediate probability group (13%) ($p < 0.0001$) and to the high probability group (6%) ($p < 0.0001$). Advanced fibrosis was also higher in the intermediate group compared to the high probability group ($p = 0.01$). AFP was significantly associated with liver fibrosis stage: medians of AFP levels were 4.9, 5.9, 13.0 and 18.6 for F1, F2, F3 and F4, respectively ($p < 0.0001$, Spearman's rank correlations). Lower platelet counts correlated with advanced fibrosis stages (data not shown). The SVR rate was higher in the high probability group compared to the intermediate or low probability group after stratification by HCV RNA load. Among patients with low HCVRNA load (<400,000 IU/ml), the rate of SVR was 93, 59 and 50% for the high, intermediate and low probability group, respectively ($p = 0.002$ for high vs. intermediate and $p < 0.001$ for high vs. low probability groups). Among patients with a high HCVRNA load (≥400,000 IU/ml), the rate of SVR was 73, 42 and 21% for the high, intermediate and low probability group, respectively ($p < 0.001$ for high vs. low, high vs. intermediate and intermediate vs. low probability groups).

Discussion

Currently, the combination of PEG-IFN and RBV is the recommended therapy for chronic HCV infection. The rate of SVR with 48 weeks of therapy is around 50% in patients with HCV genotype 1b and a high HCV RNA titer [2, 3]. To date, the virological response during therapy is the most reliable means for predicting the likelihood of SVR [2, 24, 25]. More potent therapy, such as a triple combination of protease inhibitor, PEG-IFN and RBV, is being evaluated in clinical trials but is not readily available [26, 27]. Under the circumstances, pre-treatment prediction of the likelihood of SVR may be useful for both patients and physicians to support clinical decisions as to whether to start PEG-IFN/RBV therapy or delay treatment until a new more effective therapy becomes available.

Table 1 Comparison of pre-treatment factors between model building and internal validation patients

	Model (n = 506)	Validation (n = 295)
Age (years)	56 (14–75)	55 (18–74)
Male gender[a]	261/506 (52%)	160/295 (54%)
Body mass index (kg/m^2)	22.9 (14.3–34.0)	23.2 (16.1–33.8)
Albumin (g/dl)	4 (2.7–5.0)	4 (2.8–4.9)
Creatinine (mg/dl)	0.7 (0.4–1.5)	0.7 (0.4–1.1)
AST (IU/l)	60 (11–370)	62 (11–240)
ALT (IU/l)	73 (11–413)	73 (14–390)
GGT (IU/l)	56 (10–328)	55 (7–409)
Total cholesterol (mg/dl)	173 (73–297)	171 (29–273)
Triglyceride (mg/dl)	105 (33–474)	109 (32–372)
White blood cell count (/μl)	4,745 (1,800–10,900)	4,823 (1,200–9,700)
Neutrophil count (/μl)	2,563 (667–7,870)	2,484 (508–7,579)
Red blood cell count (/μl)	448 (313–577)	451 (313–574)
Hemoglobin (g/dl)	14.1 (9.4–18.3)	14.1 (10.0–18.0)
Hematocrit (%)	41.7 (13.3–53.7)	41.9 (15.5–52.7)
Platelets (10^9/l)	164 (52–380)	158 (43–312)
AFP (ng/ml)	14.7 (0.9–680)	13 (0.8–323)
HCVRNA (10^3 IU/ml)	1,852 (100–5,100)	1,870 (100–5,100)
Fibrosis stage: F3–4	73/417 (18%)	48/247 (19%)

Data expressed as median (range) unless otherwise indicated

AST aspartate aminotransferase, *ALT* alanine aminotransferase, *GGT* gamma-glutamyltransferase, *AFP* alpha-fetoprotein

[a] Data expressed as number/available data (percentage)

belonging to the National Hospital Organization. A dataset collected from 524 patients who were treated with PEG-IFN alpha-2b/RBV was used as an external validation dataset, i.e., completely independent from the dataset that was used for model building.

Laboratory tests

Blood samples were obtained before therapy and at least once every month during therapy, and were used for hematologic tests, blood chemistry analysis and determination of HCV RNA. Pretreatment levels of HCV RNA were quantified by Cobas Amplicor (Roche Diagnostic Systems, Pleasanton, CA). SVR was defined as undetectable HCV RNA at week 24 after completion of therapy, as determined by qualitative PCR with a lower end detection limit of 50 IU/ml (Amplicor, Roche Diagnostic Systems). Liver biopsy was available in 664 patients. Fibrosis and activity were scored according to the METAVIR scoring system [22]. Fibrosis was staged on a scale of 0–4: 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). Activity of necroinflammation was graded on a scale of 0–3: A0 (no activity), A1 (mild activity), A2 (moderate activity) and A3 (severe activity).

Statistical analysis

A database of pretreatment variables was created containing 6 variables from hematological tests (red blood cells, hemoglobin, hematocrit, white blood cells, neutrocytes and platelets), 8 variables from the blood chemistry test [creatinine, albumin, aspartate aminotransferase, alanine aminotransferase, gamma-glutamyltransferase (GGT), total cholesterol, triglyceride and alpha-fetoprotein (AFP)], serum level of HCV RNA and 3 variables for patient characteristics (age, gender and body mass index). Based on this database, the recursive partitioning analysis algorithm referred to as decision tree analysis was implemented to define meaningful subgroups of patients with respect to the possibility of achieving SVR.

Decision tree analysis is a family of nonparametric regression methods. Software is used to automatically explore the data to search for optimal split variables and to build a decision tree structure [23]. For the analysis, the entire study population was evaluated to determine which variables and cutoff points yielded the most significant division into 2 prognostic subgroups that were as homogeneous as possible for the probability of SVR. Thereafter, the same analytic process was applied to all newly defined subgroups. A restriction was imposed on the tree construction such that the procedure stopped when either no additional significant variable was detected or when the sample size was below 20. For this analysis, the data mining software IBM SPSS Modeler 13 (IBM SPSS Inc., Chicago, IL) was utilized. SPSS software v.15.0 (SPSS Inc., Chicago, IL) was used for multivariate logistic regression analysis.

Results

Decision tree analysis

Decision tree analysis was carried out on the model building dataset from 506 patients using 18 variables. Figure 1 shows the results. The analysis automatically selected 5 predictive variables to produce a total of 7 subgroups of patients. Age was selected as the variable of initial split with an optimal cutoff of 50 years. The possibility of achieving SVR was 41% for patients older than 50 compared to 70% for patients

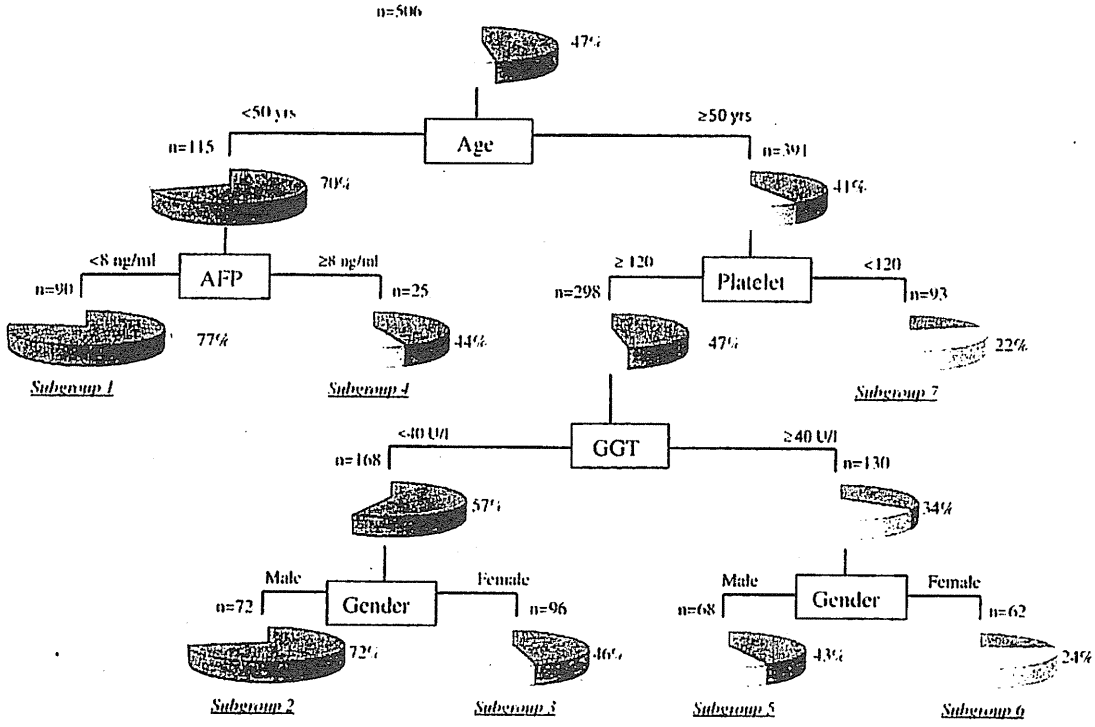

Fig. 1 Decision tree analysis. *Boxes* indicate the factors for splitting and the cutoff value for the split. *Pie charts* indicate the rate of SVR for each group. Terminal subgroups of patients discriminated by the analysis are numbered from 1 to 7. *AFP* alpha-fetoprotein, *GGT* gamma-glutamyltransferase

younger than 50. Among patients younger than 50, the level of serum AFP, with an optimal cutoff of 8 ng/ml, was selected as the variable of second split. Patients with lower AFP levels had a higher probability of SVR (77 vs. 44%). Among older patients, platelet count was selected as the second variable of split, with an optimal cutoff of 120×10^9/l. Patients with higher platelet counts had a higher probability of SVR (47 vs. 22%). Among patients with platelet counts higher than 120×10^9/l, GGT was selected as the third variable of split with an optimal cutoff of 40 IU/l. Patients with a lower GGT level had a higher probability of SVR (57 vs. 34%). Gender was selected as the fourth variable of split, with male gender being a predictor of a higher SVR probability (72 vs. 46% in patients with GGT levels <40 IU/l and 43 vs. 24% in those with GGT ≥40 IU/l). HCVRNA load was included in the analysis but was not selected as a significant variable.

The probabilities of SVR for the 7 subgroups derived by this process were highly variable. The subgroup of young patients (<50 years) with low serum AFP (<8 ng/ml) (subgroup 1) or the subgroup of older (≥50 years) male patients with high platelet counts ($\geq 120 \times 10^9$/l) and low serum GGT (<40 IU/l) (subgroup 2) showed the highest probability of SVR (72 and 77%), while the subgroup of older (≥50 years) patients with low platelet counts ($<120 \times 10^9$/l) (subgroup 7) and older (≥50 years) female patients with high serum GGT (subgroup 6) showed the lowest probability of SVR (22 and 24%).

Validation of the decision tree

The results of the decision tree analysis were validated with an internal validation dataset of 295 cases, which was independent of the model building dataset. Each patient in the validation set was allocated to subgroups 1–7 using the flow-chart form of the decision tree. The rates of SVR were 77% for subgroup 1, 71% for subgroup 2, 55% for subgroup 3, 44% for subgroup 4, 41% for subgroup 5, 17% for subgroup 6, and 30% for subgroup 7. The rates of SVR for each subgroup of patients were closely correlated between the model building dataset and the internal validation dataset ($r^2 = 0.925$) (Fig. 2a).

To further confirm the universality of the results, data collected from 524 patients by a collaborating study group were used for external validation. Thus, the dataset used for external validation was completely independent of the

Table 2 Comparison of pre-treatment factors between patients with and without sustained virological response (SVR) among the model building dataset ($n = 506$)

	SVR ($n = 240$)	Non-SVR ($n = 266$)	p
Age (years)	54 (25–75)	60 (36–73)	<0.0001
Male gender[a]	151/240 (63%)	171/266 (41%)	<0.0001
Body mass index (kg/m^2)	22.5 (16.8–32.0)	22.6 (15.5–33.3)	0.244
Albumin (g/dl)	4.1 (3.2–5.0)	4 (2.7–4.9)	0.004
Creatinine (mg/dl)	0.7 (0.44–1.14)	0.69 (0.39–1.47)	<0.0001
AST (IU/l)	59 (11–370)	61 (17–261)	0.457
ALT (IU/l)	58 (11–413)	53 (11–316)	0.031
GGT (IU/l)	31 (10–322)	43 (12–328)	0.005
Total cholesterol (mg/dl)	175 (87–297)	171 (73–274)	0.184
Triglyceride (mg/dl)	105 (36–474)	105 (33–294)	0.992
White blood cell count (/μl)	4,600 (2,200–10,900)	4,425 (1,800–10,810)	0.479
Neutrophils (/μl)	2,507 (667–7,870)	2,423 (900–7,281)	0.321
Red blood cell count (/μl)	455 (336–577)	441 (313–564)	0.001
Hemoglobin (g/dl)	14.3 (10.2–17.6)	13.9 (9.4–17.9)	0.004
Hematocrit (%)	42.1 (13.3–53.7)	41.2 (30.7–52.0)	0.031
Platelets (10^9/l)	178 (81–380)	142 (60–320)	<0.0001
AFP (ng/ml)	4.3 (0.9–680)	6.4 (1.9–468)	0.041
HCVRNA (10^3 IU/ml)	1,400 (100–5,100)	1,700 (100–5,100)	0.659
Fibrosis stage: F3–4[a]	21/198 (11%)	52/219 (24%)	<0.0001

Data expressed as median (range) unless otherwise indicated

AST aspartate aminotransferase, *ALT* alanine aminotransferase, *GGT* gamma-glutamyltransferase, *AFP* alpha-fetoprotein

[a] Data expressed as number/available data (percentage)

Table 3 Multivariate logistic regression analysis for factors associated with sustained virological response (SVR)

	Odds	95% CI	p value
Age (years)	0.96	0.94–0.98	0.001
Platelets (10^9/l)	1.09	1.04–1.14	<0.0001
ALT (IU/l)	1.01	1.00–1.01	0.001
GGT (IU/l)	0.99	0.98–0.99	<0.0001
Male gender	2.92	1.87–4.55	<0.0001

GGT gamma-glutamyltransferase

Using the data mining analysis, we constructed a simple decision tree model for the pre-treatment prediction of response to PEG-IFN/RBV. The analysis highlighted 5 variables relevant to response: age, gender, platelet count, AFP and GGT. Classification based on these variables identified subgroups of patients with high probabilities of achieving SVR among difficult to treat genotype 1b chronic hepatitis C patients. The reproducibility of the model was confirmed by the independent internal and external validation datasets. An advantage of the 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 any specific knowledge of statistics. Patients can be allocated to specific subgroups with a defined rate of response simply by following the flow-chart form. Using this model, an estimate of the response before treatment can be rapidly obtained, which may facilitate clinical decision making. Thus, this model could be readily applicable to clinical practice.

According to the results of the decision tree analysis, patients were categorized into 3 groups: the rate of SVR was 23–30% for the low probability group, 44–45% for the intermediate probability group and 65–76% for the high probability group. About 30% of patients were each categorized in the high and low probability group and the remaining 40% of patients in the intermediate probability group. These results support the evidence-based approach for selecting an optimum treatment strategy for individual patients. For example, patients in the high probability group may be the most suitable candidates for PEG-IFN/RBV therapy, while patients in the low probability group may be advised to wait for a future therapy, such as the combination of protease inhibitor, PEG-IFN and RBV. However, the estimation of low probability should not be used to preclude patients from therapy, and the final decision should be made on a case-by-case basis, taking into consideration the acceptance by the patient of a low likelihood of response and the potential risk of disease progression while waiting for a future therapy.

Another important finding was that poor adherence to drugs lowered the rate of SVR in the low and intermediate probability groups, which implies that effort should be made to maintain ≥80% of the planned dose of PEG-IFN and RBV in those patients. On the other hand, the rate of SVR was high irrespective of drug adherence in the high probability group. Whether shorter duration of therapy is sufficient in this group of patients should be confirmed in future study.

The variables used in the decision tree have been previously reported to associate with the efficacy of IFN therapy. Younger age and male gender are associated with a favorable response [28]. Lower platelet count is a hallmark of advanced fibrosis in chronic hepatitis C and is reported to be associated with poor response to IFN [29]. AFP is usually used for the screening or the diagnosis of hepatocellular carcinoma, but recent studies suggest an association between higher AFP levels and poor response to IFN therapy [30–33]. Previous report speculated that higher expression of AFP by hepatic progenitor cells may be associated with non-response to therapy [30]. Another report speculated that AFP levels predict poor response to therapy through the underlining link to advanced liver fibrosis [31]. Our data support the latter speculation since advanced fibrosis was associated with elevation of AFP levels. Fibrosis of the liver is an important predictor of response, but we did not include this factor in the decision tree analysis since liver biopsy may not always be available in general practice. As a result, two predictive factors that correlate with fibrosis stage (platelet counts and AFP) were selected in the model, and three probability groups reflected the different distribution of fibrosis stage. GGT is reported to be associated with insulin resistance and hepatic steatosis [34–37], a factor that confers resistance to IFN therapy [38–44]. What is unique to the present study is the visualization of response probability by combining these factors and its high reproducibility revealed by a high-quality validation of the model by internal and external validation datasets that were completely independent of the model building dataset. Since factors used in the model were clinical parameters that are readily available by the usual workup of patients, this model could be immediately applicable to clinical practice without imposing costs for additional examinations.

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

In conclusion, we built a pre-treatment model for the prediction of virological response to PEG-IFN/RBV. Because this decision tree model was made up of simple variables, it can be easily applied to clinical practice. This model may have the potential to support decisions about patient selection for PEG-IFN/RBV based on a possibility of response weighed against the potential risk of adverse events or costs.

Acknowledgments This study was supported by a grant-in-aid from the Ministry of Health, Labor and Welfare, Japan.

References

1. Strader DB, Wright T, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C. Hepatology. 2004;39:1147–71.
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;347:975–82.
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;358:958–65.
4. 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;41:1105–9.
5. 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;41:1100–4.
6. 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;461:399–401.
7. 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 Investig. 1995;96:224–30.
8. 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;334:77–81.
9. 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;25:750–3.
10. 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;48:372–80.
11. 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;44:952–63.
12. Breiman LJH, Friedman RA, Olshen CJ, Stone CM. Classification and regression trees. Calif: 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:

implications on the future of cancer staging. 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 ans 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. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR cooperative study group. Hepatology. 1996;24:289–93.
23. Segal MR, Bloch DA. A comparison of estimated proportional hazards models and regression trees. Stat Med. 1989;8:539–50.
24. 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;38:645–52.
25. Lee SS, Ferenci P. Optimizing outcomes in patients with hepatitis C virus genotype 1 or 4. Antivir Ther. 2008;13(Suppl 1):9–16.
26. 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;360:1839–50.
27. 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;360:1827–38.
28. Sezaki H, Suzuki F, Kawamura Y, Yatsuji H, Hosaka T, Akuta N, et al. Poor response to pegylated interferon and ribavirin in older women infected with hepatitis C virus of genotype 1b in high viral loads. Dig Dis Sci. 2009;54:1317–24.
29. Shiratori Y, Omata M. 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. 2000;15(Suppl):E141–51.
30. Gad RR, Males S, El Makhzangy H, Shouman S, Hasan A, Attala M, et al. Predictors of a sustained virological response in patients with genotype 4 chronic hepatitis C. Liver Int. 2008;28: 1112–9.
31. Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, et al. Predictors of viral kinetics to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b. J Med Virol. 2007;79:1686–95.
32. Males S, Gad RR, Esmat G, Abobakr H, Anwar M, Rekacewicz C, et al. Serum alpha-foetoprotein level predicts treatment outcome in chronic hepatitis C. Antivir Ther. 2007;12:797–803.
33. Bayati N, Silverman AL, Gordon SC. Serum alpha-fetoprotein levels and liver histology in patients with chronic hepatitis C. Am J Gastroenterol. 1998;93:2452–6.
34. Fraser A, Ebrahim S, Smith GD, Lawlor DA. A comparison of associations of alanine aminotransferase and gamma-glutamyltransferase with fasting glucose, fasting insulin, and glycated hemoglobin in women with and without diabetes. Hepatology. 2007;46:158–65.
35. Marchesini G, Avagnina S, Barantani EG, Ciccarone AM, Corica F, Dall'Aglio E, et al. Aminotransferase and gamma-glutamyltranspeptidase levels in obesity are associated with insulin resistance and the metabolic syndrome. J Endocrinol Investig. 2005;28:333–9.
36. Soresi M, Tripi S, Franco V, Giannitrapani L, Alessandri A, Rappa F, et al. Impact of liver steatosis on the antiviral response in the hepatitis C virus-associated chronic hepatitis. Liver Int. 2006;26:1119–25.
37. Yaginuma R, Ikejima K, Okumura K, Kon K, Suzuki S, Takei Y, et al. Hepatic steatosis is a predictor of poor response to interferon alpha-2b and ribavirin combination therapy in Japanese patients with chronic hepatitis C. Hepatol Res. 2006;35:19–25.
38. Adinolfi LE, Gambardella M, Andreana A, Tripodi MF, Utili R, Ruggiero G. Steatosis accelerates the progression of liver damage of chronic hepatitis C patients and correlates with specific HCV genotype and visceral obesity. Hepatology. 2001;33:1358–64.
39. 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;46:403–10.
40. Berg T, Sarrazin C, Herrmann E, Hinrichsen H, Gerlach T, Zachoval R, et al. Prediction of treatment outcome in patients with chronic hepatitis C: significance of baseline parameters and viral dynamics during therapy. Hepatology. 2003;37:600–9.
41. Mazzella G, Salzetta A, Casanova S, Morelli MC, Villanova N, Miniero R, et al. Treatment of chronic sporadic-type non-A, non-B hepatitis with lymphoblastoid interferon: gamma GT levels predictive for response. Dig Dis Sci. 1994;39:866–70.
42. Thomopoulos KC, Theocharis GJ, Tsamantas AC, Siagris D, Dimitropoulou D, Gogos CA, et al. Liver steatosis is an independent risk factor for treatment failure in patients with chronic hepatitis C. Eur J Gastroenterol Hepatol. 2005;17:149–53.
43. Villela-Nogueira CA, Perez RM, de Segadas Soares JA, Coelho HS. Gamma-glutamyl transferase (GGT) as an independent predictive factor of sustained virologic response in patients with hepatitis C treated with interferon-alpha and ribavirin. J Clin Gastroenterol. 2005;39:728–30.
44. Camps J, Crisostomo S, Garcia-Granero M, Riezu-Boj JI, Civeira MP, Prieto J. Prediction of the response of chronic hepatitis C to interferon alfa: a statistical analysis of pretreatment variables. Gut. 1993;34:1714–7.

Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | 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, Kohodai 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 (10^4/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

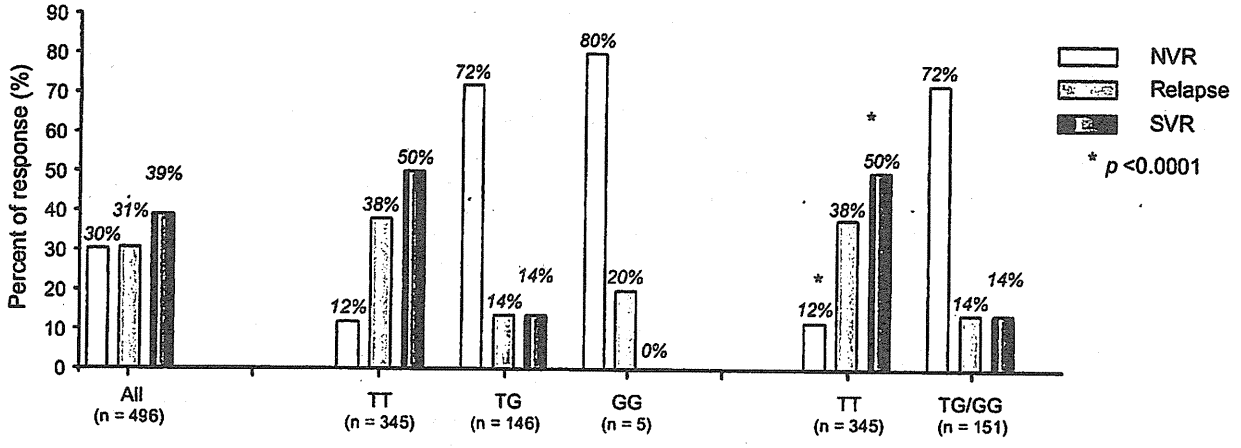

Fig. 1. Association between the IL28B genotype (rs8099917) and treatment response. The rates of response to treatment are shown for each rs8099917 genotype. The rate of null virological response (NVR), relapse, and sustained virological response (SVR) is shown. The *p* values are from Fisher's exact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

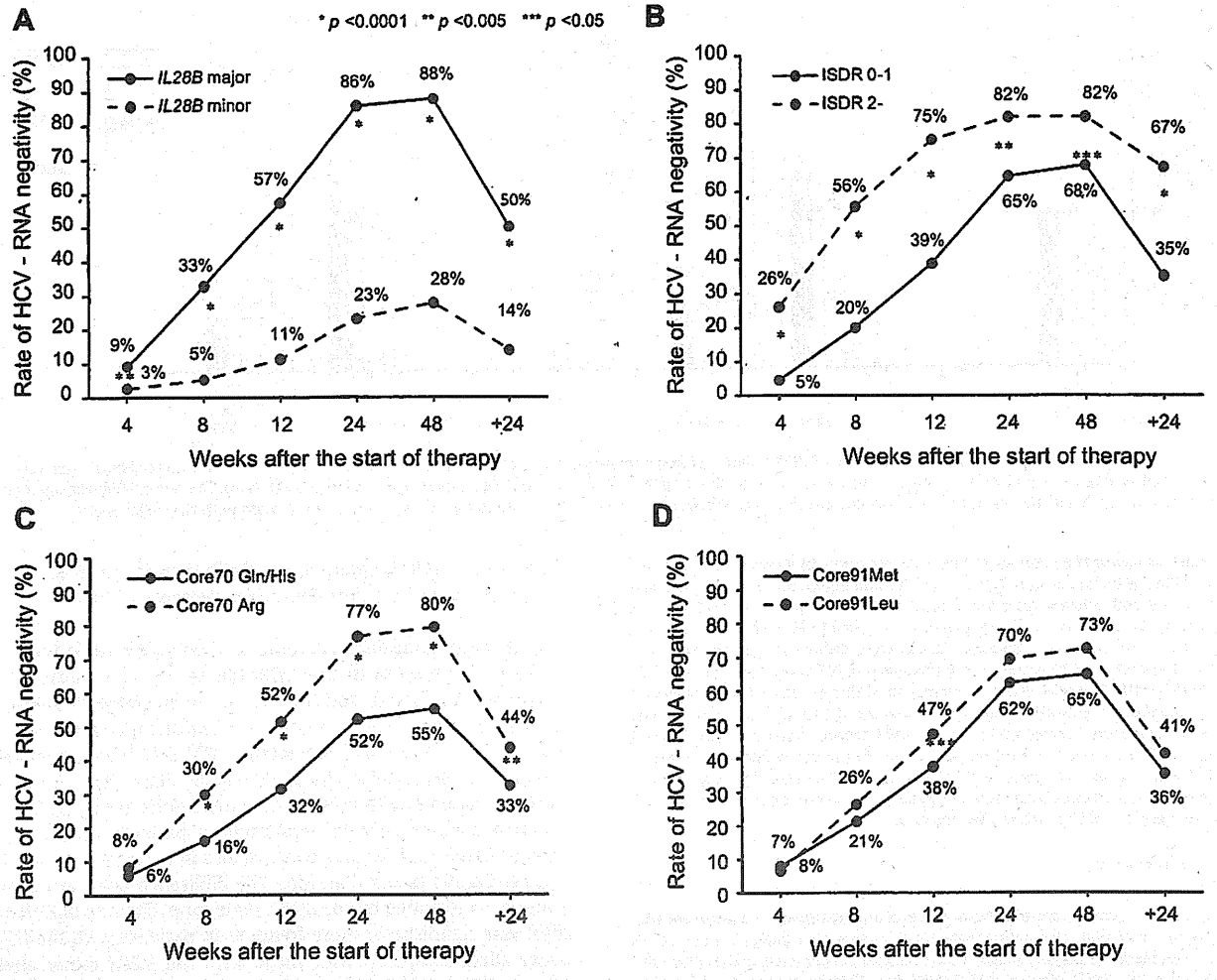

Fig. 2. Effect of *IL28B* mutations in the ISDR, Core70, and Core91 of HCV on time-dependent clearance of HCV. The rate of undetectable HCV-RNA was plotted for serial time points after the start of therapy (4, 8, 12, 24, and 48 weeks) and for 24 weeks after the completion of therapy. Patients were stratified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

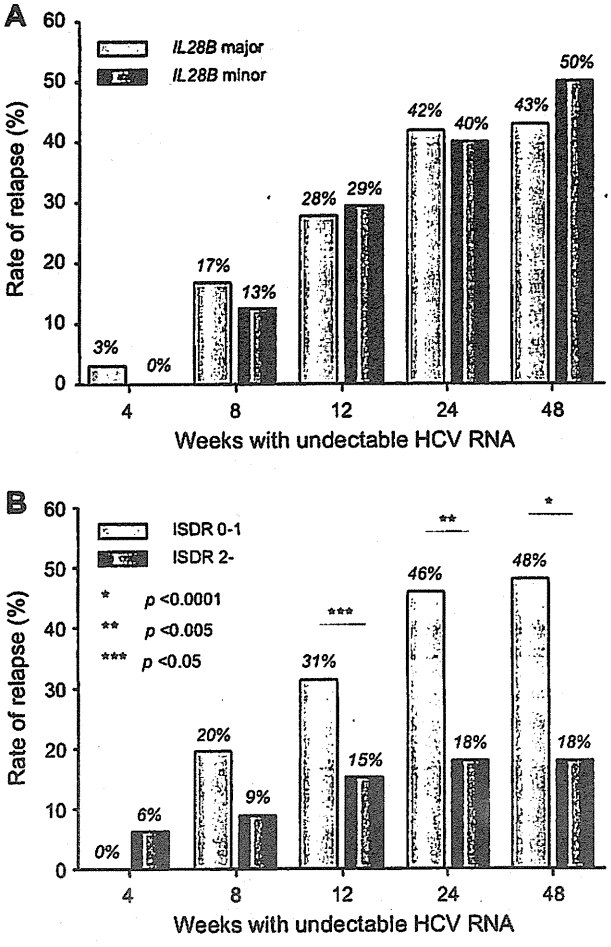

Fig. 3. Association between relapse and the *IL28B* allele or mutations in the ISDR. The rate of relapse was calculated for patients who had undetectable HCV-RNA at serial time points after the start of therapy (4, 8, 12, 24, and 48 weeks). Patients were stratified according to (A) the *IL28B* allele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

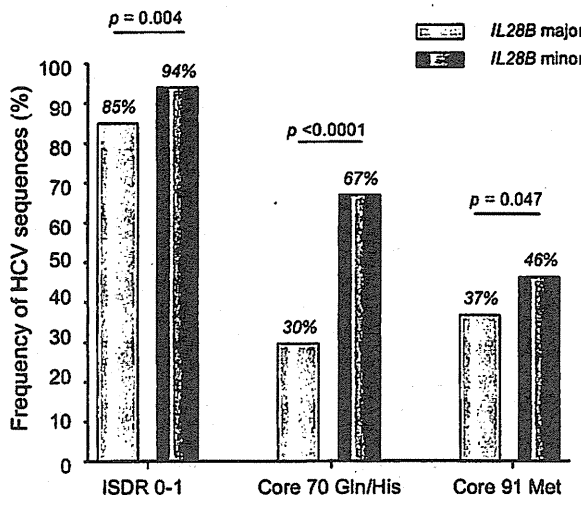

Fig. 4. Associations between the *IL28B* allele and HCV sequences. The prevalence of HCV sequences predicting a resistant phenotype to IFN was higher in patients with the minor *IL28B* allele than those with major allele. (A) 0 or 1 mutation in the ISDR of NS5A, (B) Gln or His at Core70, and (C) Met 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	1.01-1.16	<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≤	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	76 ± 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

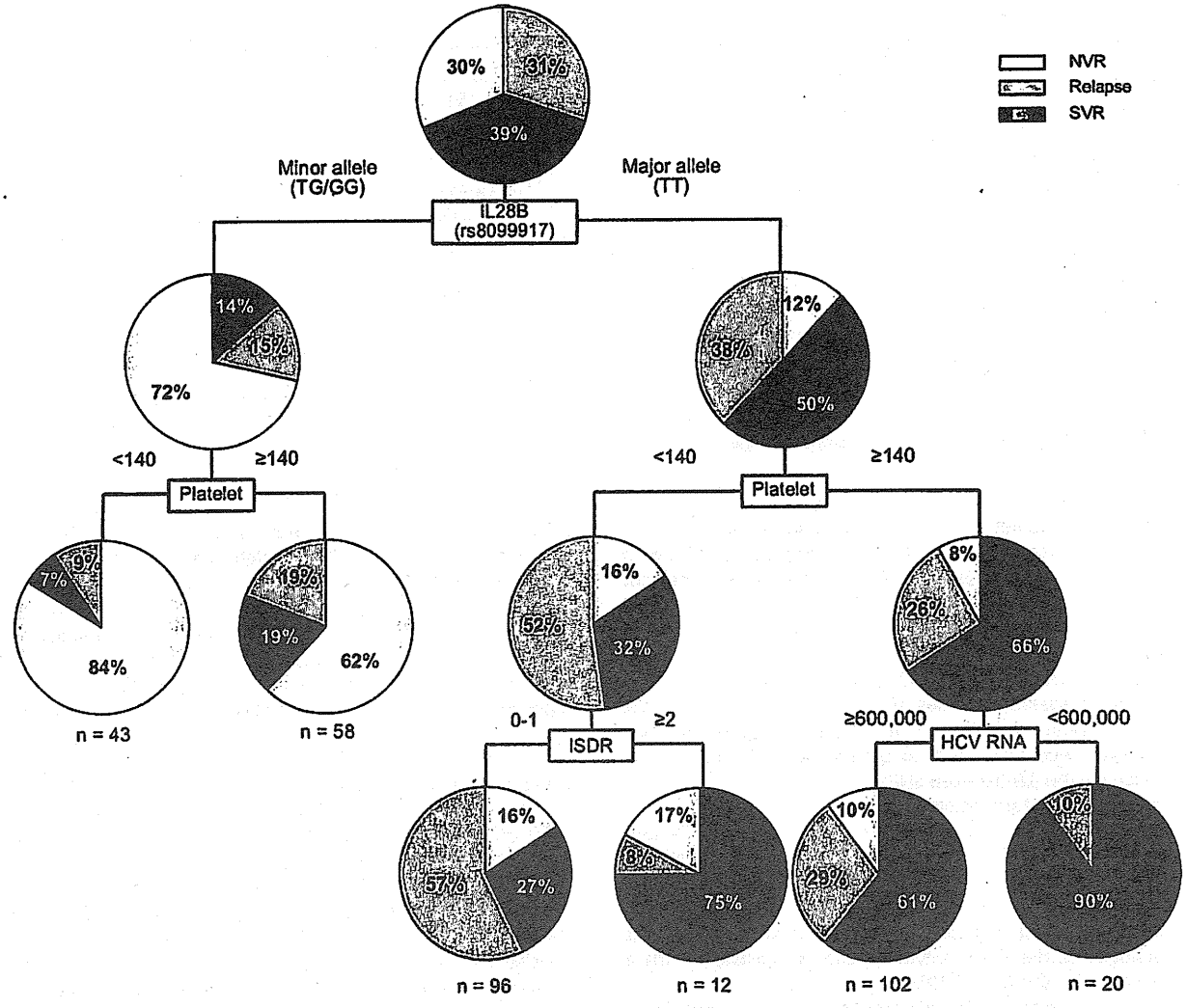

Fig. 5. Decision tree for the prediction of response to therapy. The boxes indicate the factors used for splitting. Pie charts indicate the rate of response for each group of patients after splitting. The rate of null virological response, relapse, and sustained virological response is shown.

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-