

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 ($\geq 80\%$ 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 ($\geq 80\%$ 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 ($\geq 400,000$ 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 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 (μ l)	4,600 (2,200–10,900)	4,425 (1,800–10,810)	0.479
Neutrophils (μ l)	2,507 (667–7,870)	2,423 (900–7,281)	0.321
Red blood cell count (μ l)	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 $\geq 80\%$ 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, Goncalves 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 Invest*. 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 L, 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. Averbuch 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 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. Bedossa P, Poinard 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, Rekecawicz C, et al. Serum alpha-fetoprotein 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-glutamyl-transferase 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-glutamyl-transpeptidase 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.

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

Masayuki Kurosaki,¹ Naoya Sakamoto,² Manabu Iwasaki,³ Minoru Sakamoto,⁴ Yoshiyuki Suzuki,⁵ Naoki Hiramatsu,⁶ Fuminaka Sugauchi,⁷ Akihiro Tamori,⁸ Mina Nakagawa,² and Namiki Izumi, MD^{1*}

¹Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan

²Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Tokyo, Japan

³Department of Computer and Information Science, Seikei University, Tokyo, Japan

⁴First Department of Internal Medicine, University of Yamanashi, Yamanashi, Japan

⁵Department of Hepatology, Toranomon Hospital, Tokyo, Japan

⁶Department of Gastroenterology and Hepatology, Osaka University Graduate School of Medicine, Osaka, Japan

⁷Department of Gastroenterology and Metabolism, Nagoya City University Graduate School of Medical Sciences, Aichi, Japan

⁸Department of Hepatology, Osaka City University Medical School, Osaka, Japan

The aim of the present study was to clarify the significance of viral factors for pretreatment prediction of sustained virological response to pegylated-interferon (PEG-IFN) plus ribavirin (RBV) therapy for chronic hepatitis C using data mining analysis. Substitutions in the IFN sensitivity-determining region (ISDR) and at position 70 of the HCV core region (Core70) were determined in 505 patients with genotype 1b chronic hepatitis C treated with PEG-IFN plus RBV. Data mining analysis was used to build a predictive model of sustained virological response in patients selected randomly ($n = 304$). The reproducibility of the model was validated in the remaining 201 patients. Substitutions in ISDR (odds ratio = 9.92, $P < 0.0001$) and Core70 (odds ratio = 1.92, $P = 0.01$) predicted sustained virological response independent of other covariates. The decision-tree model revealed that the rate of sustained virological response was highest (83%) in patients with two or more substitutions in ISDR. The overall rate of sustained virological response was 44% in patients with a low number of substitutions in ISDR (0–1) but was 83% in selected subgroups of younger patients (<60 years), wild-type sequence at Core70, and higher level of low-density lipoprotein cholesterol (LDL-C) (≥ 120 mg/dl). Reproducibility of the model was validated ($r^2 = 0.94$, $P < 0.001$). In conclusion, substitutions in ISDR and Core70 of

HCV are significant predictors of response to PEG-IFN plus RBV therapy. A decision-tree model that includes these viral factors as predictors could identify patients with a high probability of sustained virological response. **J. Med. Virol.** 83:445–452, 2011.

© 2011 Wiley-Liss, Inc.

KEY WORDS: data mining; decision-tree model; ISDR; core region; PEG-interferon

INTRODUCTION

The combination of pegylated-interferon (PEG-IFN) plus ribavirin (RBV) is currently the most effective therapy for chronic hepatitis C, but the rate of sustained virological response after 48 weeks of therapy is about 50% in patients with HCV genotype 1b and a high HCV

Grant sponsor: Ministry of Health, Labor and Welfare, Japan. The authors report no conflicts of interest.

*Correspondence to: Namiki Izumi, MD, Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, 1-26-1 Kyonan-cho, Musashino-shi, Tokyo 180-8610, Japan. E-mail: nizumi@musashino.jrc.or.jp

Accepted 26 October 2010

DOI 10.1002/jmv.22005

Published online in Wiley Online Library (wileyonlinelibrary.com).

RNA titer [Manns et al., 2001; Fried et al., 2002]. The most reliable means to predict sustained virological response is to monitor the viral response during the early weeks of treatment. The early virological response, defined as undetectable HCV RNA at week 12, is associated with a high rate of sustained virological response [Davis et al., 2003; Lee and Ferenci, 2008]. The rapid virological response, defined as undetectable HCV RNA at week 4 of therapy, is even more predictive of sustained virological response than the early virological response [Jensen et al., 2006; Yu et al., 2008; Izumi et al., 2010]. However, there is no established means that predicts the virological response before commencing treatment. Recent reports have revealed that single nucleotide polymorphisms located near the *IL28B* gene show a strong association with the response to PEG-IFN plus RBV therapy [Ge et al., 2009; Suppiah et al., 2009; Tanaka et al., 2009; Kurosaki et al., 2010c]. These findings indicate that the host factor is an important determinant of the treatment response. On the other hand, the present study's authors have reported that a stretch of 40 amino acids in the NS5A region of HCV, designated as the interferon sensitivity-determining region (ISDR), has a close association with the virological response to interferon mono-therapy [Enomoto et al., 1995, 1996; Kurosaki et al., 1997]. More recently, amino acid substitutions at positions 70 and 91 of the core region have been reported to be associated with response to PEG-IFN plus RBV combination therapy [Akuta et al., 2005, 2007a]. The impact of these HCV substitutions on treatment response is yet to be validated.

Decision-tree analysis is a core component of data mining analysis that can be used to build predictive models [Breiman et al., 1980]. This method has been used to define prognostic factors in various diseases such as prostate cancer [Garzotto et al., 2005], diabetes [Miyaki et al., 2002], melanoma [Averbook et al., 2002; Leiter et al., 2004], colorectal carcinoma [Zlobec et al., 2005; Valera et al., 2007], and liver failure [Baquerizo et al., 2003]. The major advantage of decision-tree analysis over logistic regression analysis is that the results of analysis are easy to understand. The simple allocation of patients into subgroups by following the flowchart form could define the predicted possibility of outcome [LeBlanc and Crowley, 1995].

Decision-tree analysis was used for the prediction of early virological response (undetectable HCV RNA within 12 weeks of therapy) to PEG-IFN and RBV combination therapy in chronic hepatitis C [Kurosaki et al., 2010a], and more recently for the pretreatment prediction of sustained virological response [Kurosaki et al., 2010b]. In the latter model, simple and noninvasive standard tests were used as parameters; specialized tests such as viral mutations and host genetics, or invasive tests such as liver histology, were not included because the aim of that model was for use in general medical practice, especially in some countries or areas where resources are limited. Thus, the impact of viral mutations or liver histology was not considered in that model.

The present study examined whether including viral substitutions in ISDR and the core region of HCV in the decision-tree model could improve its predictive accuracy over the previous model to identify chronic hepatitis C patients who are likely to respond to PEG-IFN plus RBV therapy.

MATERIALS AND METHODS

Patients

This multicenter retrospective cohort study included 505 chronic hepatitis C patients who were treated 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, Osaka City University, and their related hospitals. The inclusion criteria were: (1) genotype 1b, (2) HCV RNA titer higher than 100 kIU/ml by quantitative PCR (Cobas Amplicor HCV Monitor v 2.0, Roche Diagnostic Systems, Pleasanton, CA), (3) no co-infection with hepatitis B virus or human immunodeficiency virus, (4) no other causes of liver disease, (5) patients having undergone liver biopsy prior to IFN treatment, (6) number of substitutions in ISDR having been determined, (7) substitutions in the amino acid positions 70 and 91 of the core region having been determined, and (8) completion of at least 12 weeks of therapy. Patients were treated with PEG-IFN alpha-2b (1.5 µg/kg) weekly plus RBV. The daily dose of RBV was adjusted by weight: 600 mg for <60 kg, 800 mg for 60–80 kg, and 1,000 mg for >80 kg. For the analysis, patients were assigned randomly to either the model building (304 patients) or validation (201 patients) groups. There were no significant differences in the clinical backgrounds between these two groups (Table I). Informed 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 committees of all concerned hospitals.

Laboratory Tests

Hematological tests, blood chemistry, and HCV RNA titer were analyzed before therapy and at least once every month during therapy. 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. At position 70 of the core region (Core70), arginine was defined as the wild type, and glutamine or histidine was defined as the mutant type. At position 91 of the core region, leucine was defined as the wild type and methionine was defined as the mutant type, as described previously [Akuta et al., 2005]. Fibrosis and activity were scored according to the METAVIR scoring system [Bedossa and Poynard, 1996]. Fibrosis was staged on a scale of 0–4: F0 (no fibrosis), F1 (mild fibrosis), F2 (moderate fibrosis), F3 (severe fibrosis), and F4 (cirrhosis). Activity of necroinflammation was graded on a scale of

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

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

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

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

Statistical Analysis

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

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

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

RESULTS

Generation of the Decision-Tree Model

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

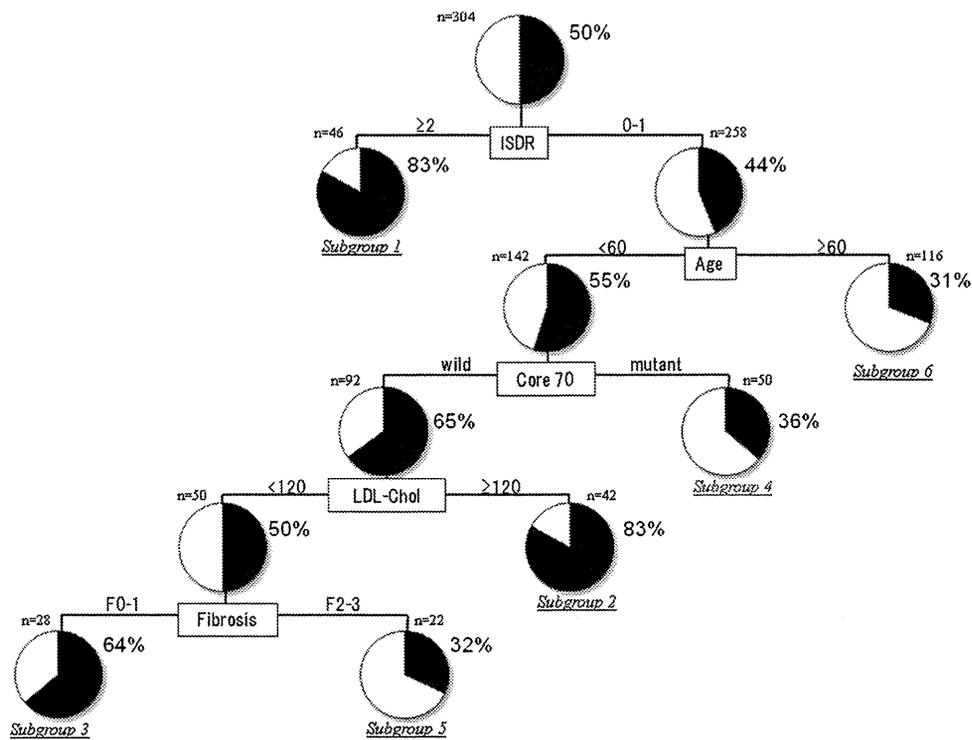


Fig. 1. Decision-tree model. Boxes indicate the factors used for splitting and the cutoff value for the split. Pie charts indicate the rate of sustained virological response for each group of patients after splitting. Terminal subgroups of patients discriminated by the analysis are numbered from 1 to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

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

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

Validation of the Decision-Tree Model

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

J. Med. Virol. DOI 10.1002/jmv

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

Effect of Dose Reductions of PEG-IFN and RBV

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

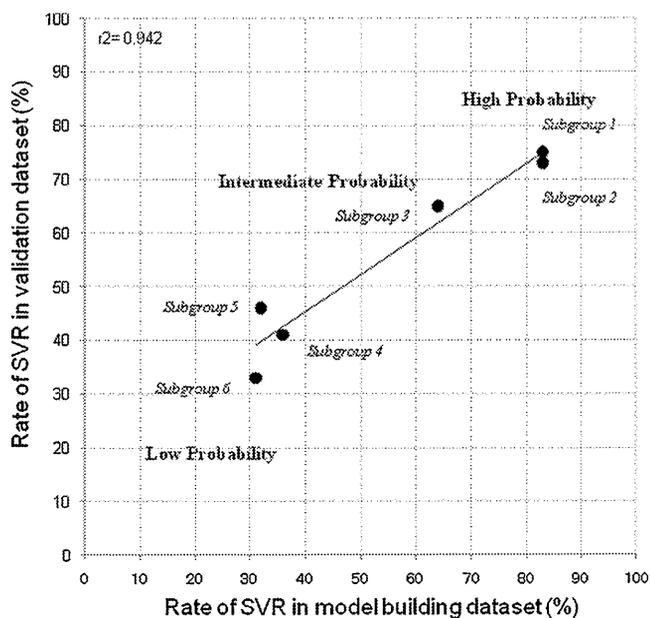


Fig. 2. Validation of the decision-tree analysis: Subgroup-stratified comparison of the rate of sustained virological response. Each patient in the validation set was allocated to subgroups 1–6 by following the flowchart form of the decision tree, and the rates of sustained virological response were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

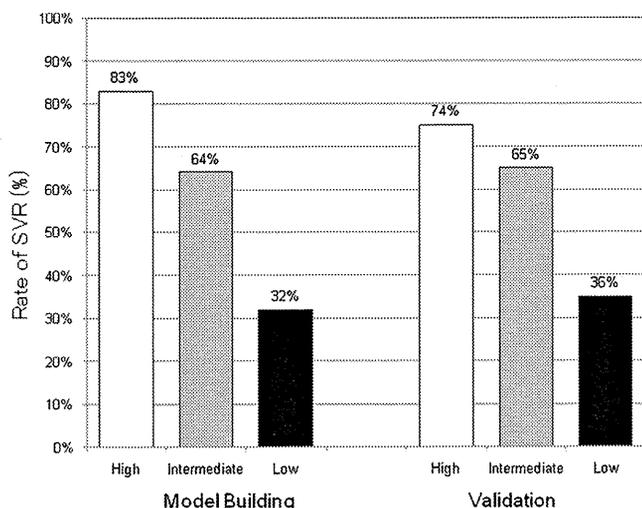


Fig. 3. Comparison of sustained virological response rates between groups divided by the decision tree. The rate of sustained virological response was compared between three groups of patients as divided by the decision-tree analysis. Black, gray, and white boxes indicate the low-probability group (subgroup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

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

Multivariable Logistic Regression Analysis

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

DISCUSSION

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

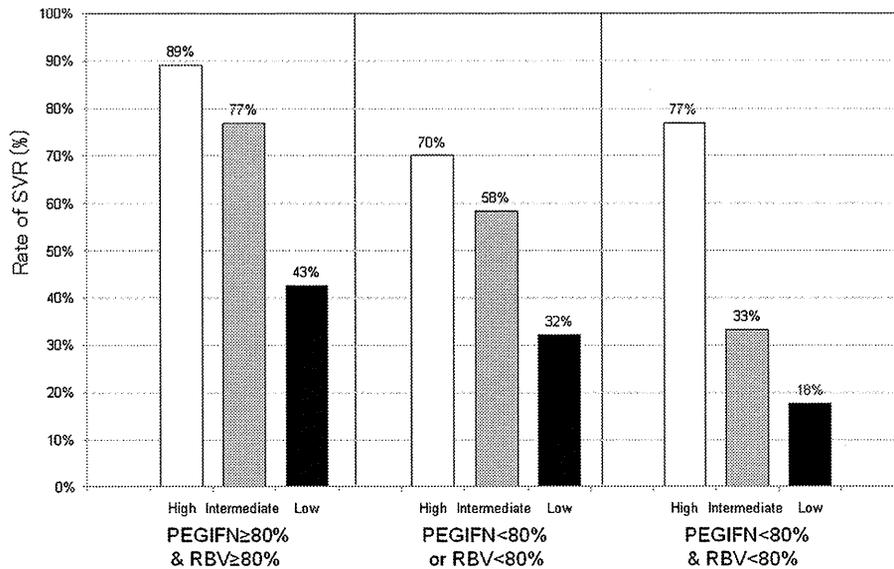


Fig. 4. Comparison of the rate of sustained virological response between the decision-tree groups stratified by drug adherence. The three groups of patients divided by the decision tree (black, gray, and white boxes indicating the low-, intermediate-, and high-probability groups, respectively) were further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

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

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

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

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

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

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

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

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

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

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

REFERENCES

- Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, Someya T, Kobayashi M, Saitoh S, Watahiki S, Sato J, Matsuda M, Arase Y, Ikeda K, Kumada H. 2005. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. *Intervirology* 48:372–380.
- Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Arase Y, Ikeda K, Kumada H. 2007a. Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: Amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. *J Hepatol* 46:403–410.
- Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Arase Y, Ikeda K, Kumada H. 2007b. Predictors of viral kinetics to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b. *J Med Virol* 79:1686–1695.
- Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Kumada H. 2009. A matched case-controlled study of 48 and 72 weeks of peginterferon plus ribavirin combination therapy in patients infected with HCV genotype 1b in Japan: Amino acid substitutions in HCV core region as predictor of sustained virological response. *J Med Virol* 81:452–458.
- Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Chayama K, Nakamura Y, Kumada H. 2010. Amino acid substitution in hepatitis C virus core region and genetic variation near the interleukin 28B gene predict viral response to telaprevir with peginterferon and ribavirin. *Hepatology* 52:421–429.
- Averbook BJ, Fu P, Rao JS, Mansour EG. 2002. A long-term analysis of 1018 patients with melanoma by classic Cox regression and tree-structured survival analysis at a major referral center: Implications on the future of cancer staging. *Surgery* 132:589–602.
- Baquerizo A, Anselmo D, Shackleton C, Chen TW, Cao C, Weaver M, Gornbein J, Geevarghese S, Nissen N, Farmer D, Demetriou A, Busuttil RW. 2003. Phosphorus and an early predictive factor in patients with acute liver failure. *Transplantation* 75:2007–2014.
- Bedossa P, Poinard T. 1996. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 24:289–293.
- Breiman L, Friedman RA, Olshen CJ, Stone CM. 1980. Classification and regression trees. CA: Wadsworth.
- Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. 2003. Early virologic response to treatment with peginterferon alpha-2b plus ribavirin in patients with chronic hepatitis C. *Hepatology* 38:645–652.
- Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Izumi N, Marumo F, Sato C. 1995. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J Clin Invest* 96:224–230.
- Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Ogura Y, Izumi N, Marumo F, Sato C. 1996. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med* 334:77–81.
- Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncalves FL, Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J. 2002. Peginterferon alpha-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 347:975–982.

- Garzotto M, Beer TM, Hudson RG, Peters L, Hsieh YC, Barrera E, Klein T, Mori M. 2005. Improved detection of prostate cancer using classification and regression tree analysis. *J Clin Oncol* 23:4322–4329.
- Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, Heinzen EL, Qiu P, Bertelsen AH, Muir AJ, Sulkowski M, McHutchison JG, Goldstein DB. 2009. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 461:399–401.
- Hezode C, Forestier N, Dusheiko G, Ferenci P, Pol S, Goeser T, Bronowicki JP, Bourliere M, Gharakhanian S, Bengtsson L, McNair L, George S, Kieffer T, Kwong A, Kauffman RS, Alam J, Pawlotsky JM, Zeuzem S. 2009. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N Engl J Med* 360:1839–1850.
- Ikeda H, Suzuki M, Okuse C, Yamada N, Okamoto M, Kobayashi M, Nagase Y, Takahashi H, Matsunaga K, Matsumoto N, Itoh F, Yotsuyanagi H, Koitabashi Y, Yasuda K, Iino S. 2009. Short-term prolongation of pegylated interferon and ribavirin therapy for genotype 1b chronic hepatitis C patients with early viral response. *Hepatol Res* 39:753–759.
- Izumi N, Nishiguchi S, Hino K, Suzuki F, Kumada Y, Itoh Y, Asahina Y, Tamori A, Hiramatsu N, Hayashi N, Kudo M. 2010. Management of hepatitis C; Report of the consensus meeting at the 45th annual meeting of the Japan society of hepatology (2009). *Hepatol Res* 40:347–368.
- Jensen DM, Morgan TR, Marcellin P, Pockros PJ, Reddy KR, Hadziyannis SJ, Ferenci P, Akrill 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, Yatsushashi 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.

Data mining model using simple and readily available factors could identify patients at high risk for hepatocellular carcinoma in chronic hepatitis C

Masayuki Kurosaki¹, Naoki Hiramatsu², Minoru Sakamoto³, Yoshiyuki Suzuki⁴, Manabu Iwasaki⁵, Akihiro Tamori⁶, Kentaro Matsuura⁷, Sei Kakinuma⁸, Fuminaka Sugauchi⁹, Naoya Sakamoto⁸, Mina Nakagawa⁸, Namiki Izumi^{1,*}

¹Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan; ²Department of Gastroenterology and Hepatology, Osaka University Graduate School of Medicine, Osaka, Japan; ³First Department of Internal Medicine, University of Yamanashi, Yamanashi, Japan; ⁴Department of Hepatology, Toranomon Hospital, Tokyo, Japan; ⁵Department of Computer and Information Science, Seikei University, Tokyo, Japan; ⁶Department of Hepatology, Osaka City University Medical School, Osaka, Japan; ⁷Department of Gastroenterology and Metabolism, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; ⁸Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Tokyo, Japan; ⁹Department of Gastroenterology, Nagoya Koseiin Medical Welfare Center, Nagoya, Japan

Background & Aims: Assessment of the risk of hepatocellular carcinoma (HCC) development is essential for formulating personalized surveillance or antiviral treatment plan for chronic hepatitis C. We aimed to build a simple model for the identification of patients at high risk of developing HCC.

Methods: Chronic hepatitis C patients followed for at least 5 years (n = 1003) were analyzed by data mining to build a predictive model for HCC development. The model was externally validated using a cohort of 1072 patients (472 with sustained virological response (SVR) and 600 with nonSVR to PEG-interferon plus ribavirin therapy).

Results: On the basis of factors such as age, platelet, albumin, and aspartate aminotransferase, the HCC risk prediction model identified subgroups with high-, intermediate-, and low-risk of HCC with a 5-year HCC development rate of 20.9%, 6.3–7.3%, and 0–1.5%, respectively. The reproducibility of the model was confirmed through external validation ($r^2 = 0.981$). The 10-year HCC development rate was also significantly higher in the high- and intermediate-risk group than in the low-risk group (24.5% vs. 4.8%; $p < 0.0001$). In the high- and intermediate-risk group, the incidence of HCC development was significantly reduced in patients with SVR compared to those with nonSVR (5-year rate, 9.5% vs. 4.5%; $p = 0.040$).

Conclusions: The HCC risk prediction model uses simple and readily available factors and identifies patients at a high risk of HCC development. The model allows physicians to identify patients requiring HCC surveillance and those who benefit from IFN therapy to prevent HCC.

Keywords: Decision tree; Prediction; Pegylated interferon; Ribavirin; Risk.
Received 27 May 2011; received in revised form 8 August 2011; accepted 4 September 2011

* 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).



ELSEVIER

Journal of Hepatology 2011 vol. xxx | xxx–xxx

© 2011 European Association for the Study of the Liver. Published by Elsevier B.V. All rights reserved.

Introduction

Hepatocellular carcinoma (HCC) is the sixth most common cancer worldwide [1] and its incidence is increasing in many countries [2]. Chronic viral hepatitis is responsible for 80% of all HCC cases [2]. The need to conduct HCC surveillance should be determined according to the risk of HCC development because this surveillance is cost-effective only in populations with an annualized cancer development rate of $\geq 1.5\%$ [3]. The annualized rate of developing HCC from type C liver cirrhosis is 2–8% [4–6], indicating that this population with type C liver cirrhosis needs surveillance. However, the annualized rate of HCC development is $< 1.5\%$ in patients with chronic hepatitis C but without cirrhosis and the benefit of surveillance for all patients with chronic hepatitis has not yet been established [3]. HCC surveillance may be needed for patients with advanced fibrosis because the risk of HCC development increases in parallel with the progression of liver fibrosis [7,8]. Liver biopsy is the most accurate means of diagnosing fibrosis, but a single liver biopsy cannot indicate long-term prognosis because liver fibrosis progresses over time. Serial liver biopsies are not feasible because of the procedure's invasiveness. Moreover, factors other than fibrosis, such as advanced age, obesity, sex, lower albumin, and low platelet counts, also contribute to the development of HCC from chronic hepatitis C [8–11]. Therefore, these factors must be considered while assessing the risk of HCC development.

A meta-analysis of controlled trials [12] has shown that interferon (IFN) therapy reduced the rate of HCC development in patients with type C liver cirrhosis. However, there was a marked heterogeneity in the magnitude of the prevention effect

Research Article

of IFN on HCC development among the studies, probably due to the large differences in the baseline rate of HCC development among the different trials [12]. Whether the incidence of HCC development could be reduced in all patients with chronic hepatitis C, especially in those without liver cirrhosis, remains to be elucidated.

Data mining analysis, unlike conventional statistical analysis, is performed in an exploratory manner without considering a predefined hypothesis. Decision tree analysis, the major component of data mining analysis, is used to extract relevant factors from among various factors. These relevant factors are then combined in an orderly sequence to identify rules for predicting the incidence of the target outcome [13]. Data mining analysis has been used to define prognostic factors in various diseases [14–20]. In the field of hepatic diseases, data mining analysis has proven to be a useful tool for predicting early response [21], sustained virological response (SVR) [22–25], relapse [26], and adverse events [27] in patients with chronic hepatitis C treated with pegylated interferon (PEG-IFN) plus ribavirin (RBV). The findings of data mining analysis are expressed as flowcharts and are therefore easily understood [28] and readily available for clinical use, even by physicians without a detailed understanding of statistics.

In the present study, data mining analysis was used to identify risk factors for HCC development in a cohort of patients with chronic hepatitis C who had been followed for at least 5 years. An HCC risk prediction model was constructed on the basis of simple and generally available tests because the goal was to make the model easy to use in the clinic. The suitability, reproducibility, and generalizability of the results were validated using the data of an external cohort that was independent of the model derivation cohort.

Materials and methods

Patients

The model derivation cohort consisted of 1003 chronic hepatitis C patients without cirrhosis who had a non-sustained virological response (nonSVR) to previous IFN administered at the Musashino Red Cross Hospital and were followed for at least 5 years. Patients who had SVR or those who were followed for less than 5 years were not included. An analytical database on age, body mass index, albumin, aspartate aminotransferase (AST) levels, alanine aminotransferase (ALT) levels, γ -glutamyltransferase (GGT) levels, total bilirubin levels, total cholesterol levels, hemoglobin levels, and platelet count at the start of the observation was created. Histological data such as fibrosis stage, activity grade, or degree of steatosis was not included in the database because the goal of the present study was to make the model on the basis of simple and generally available tests. The patients who developed HCC more than 5 years after the start of the observation were considered not to have developed HCC by the 5-year point because the model was intended to predict HCC development within 5 years. The 1072 chronic hepatitis C patients included in the external validation cohort were treated with PEG-IFN and RBV at the University of Yamanashi, Tokyo Medical and Dental University, Osaka University, Osaka City University, Nagoya City University, or Toranomon Hospital and followed for at least 5 years. Among them, 600 had nonSVR and 472 had SVR. Data from nonSVR patients in this external cohort were used for external validation of the HCC prediction model. To assess the preventive effect of PEG-IFN plus RBV therapy on HCC development, the cumulative HCC development rate was compared between SVR and nonSVR patients in the external validation cohort after stratification by the risk of HCC development as determined by data mining analysis. Informed 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 committees of all concerned hospitals.

HCC surveillance and diagnosis

HCC surveillance was conducted by performing abdominal ultrasonography every 4–6 months. Contrast-enhanced computer tomography, magnetic resonance imaging, or angiography were performed when abdominal ultrasonography suggested a new lesion suspicious for HCC. Classical HCC was diagnosed for tumors showing vascular enhancement with washout on at least two types of diagnostic imaging. Tumor biopsy was used to diagnose tumors with non-classical imaging findings.

Statistical analysis

The IBM-SPSS Modeler 13 (IBM SPSS Inc., Chicago, IL, USA) was used for decision tree analysis. The statistical methods used have been described previously [21,22,24–27]. In brief, the software searched the analytical database for the factor that most effectively predicted HCC development and for its cutoff value. The patients were divided into two groups according to that predictor. Each divided group was repeatedly assessed and divided according to this 2-choice branching method. Branching was stopped when the number of patients decreased to ≤ 20 to avoid over fitting. Finally, an HCC risk prediction model was created through this analysis. The model classified patients into subgroups with different HCC development rates in a flowchart form. For model validation, nonSVR patients from an external cohort were individually fitted into the model and classified into the subgroups and the HCC development rates of those subgroups were then calculated. The suitability and reproducibility of the model were validated by comparing the subgroup HCC development rates of the model derivation group to those of the validation group.

On univariate analysis, Student's *t*-test was used for continuous variables and Fisher's exact test was used for categorical data. Logistic regression was used for multivariate analysis. A log-rank test for Kaplan–Meier analysis was used to statistically test HCC development rates over time. *p*-Values of < 0.05 were considered significant. SPSS Statistics 18 (IBM SPSS Inc.) was used for these analyses.

Results

Univariate and multivariate analysis of factors associated with HCC development

The baseline characteristics of patients are shown in Table 1. The 5-year HCC development rate in the model derivation group was 6.2%, which did not differ significantly from the rate of 6.0% in the nonSVR group of the external cohort, but the rate of 2.0% in the SVR group of the external cohort was significantly lower than that in the model derivation group ($p = 0.0003$) and the nonSVR group of the external cohort ($p = 0.0012$). On univariate analysis, the factors found to be associated with HCC development in the model derivation cohort were age, AST levels, albumin levels, total cholesterol levels, and platelet count. On multivariate analysis, age (odds ratio 1.086), albumin levels (odds ratio 0.248), and platelet count (odds ratio 0.842) were significant predictors of HCC development (Table 2).

HCC risk prediction model by data mining analysis

The results of decision tree analysis are presented in Fig. 1. Age was selected as the first predictor. The 5-year HCC development rate was 3.4% in younger patients (< 60 years) and 8.6% in older patients (≥ 60 years). The second predictor for younger patients (< 60 years) was platelet count. The HCC development rate was 6.9% in patients with a lower platelet count ($< 150 \times 10^9/L$) and 0.8% in patients with a higher count ($\geq 150 \times 10^9/L$). The second predictor for older patients (≥ 60 years) was also platelet count. The HCC development rate was 13.1% in patients with a lower platelet count ($< 150 \times 10^9/L$) and 1.8% in patients with a higher count ($\geq 150 \times 10^9/L$). The third predictor was albumin levels,

Table 1. Baseline characteristics of patients for model derivation and external validation.

	Model derivation (n = 1003)	External cohort, non-SVR (n = 600)	External cohort, SVR (n = 472)
Sex: Male/Female*	463 (46%)/540 (54%)	306 (51%)/294 (49%)	299 (63%)/173 (37%)
Age (yr)	57.3 (11.1)	55.9 (9.6)	51.4 (10.6)
Body mass index (kg/m ²)	23.5 (3.2)	23.4 (3.3)	23.3 (3.1)
Albumin (g/dl)	4.1 (0.3)	4.0 (0.4)	4.0 (0.3)
AST (IU/L)	64.2 (36.5)	67.3 (43.8)	62.5 (48.3)
ALT (IU/L)	80.6 (55.1)	81.2 (62.3)	88.6 (82.1)
GGT (IU/L)	59.3 (50.5)	67.6 (65.1)	55.7 (71.2)
Total cholesterol (mg/dl)	172.1 (31.5)	168.2 (31.0)	174.3 (33.7)
Platelet (10 ⁹ /L)	154.0 (53.0)	153.7 (53.2)	176.6 (49.7)
Hemoglobin (g/dl)	13.3 (1.5)	14.2 (1.5)	14.4 (1.4)
HCC development within 5 years: n (%)*	62 (6.2%)	36 (6.0%)	10 (2.0%)

Data expressed as mean (standard deviation) unless otherwise indicated.

AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma-glutamyltransferase; HCC, hepatocellular carcinoma; SVR, sustained virological response.

*Data expressed as number of patients (percentage).

whose cutoff value was 3.75 g/dl in patients with a higher platelet count ($\geq 150 \times 10^9/L$). The HCC development rate was 6.3% when albumin levels were lower (< 3.75 g/dl) and 1.5% when levels were higher (≥ 3.75 g/dl). The cutoff value for albumin levels was 4.0 g/dl in patients with a lower platelet count ($< 150 \times 10^9/L$). The HCC development rate was 20.9% when albumin levels were lower (< 4.0 g/dl) and 6.4% when levels were higher (≥ 4.0 g/dl). The fourth and final predictor was AST levels. The HCC development rate was 7.3% when AST levels were at least 40 IU/L and 0% when the levels were < 40 IU/L. On the basis of this analysis, seven subgroups with a 5-year HCC development rate of 0–20.9% were identified. The area under the receiver operating characteristic curve according to the HCC risk prediction model was 0.817.

External validation of the HCC risk prediction model with an independent external cohort

Six hundred nonSVR patients from an external cohort were fitted into the HCC risk prediction model and classified into the seven subgroups. The 5-year HCC development rate of these subgroups was 0–17.9%. The HCC development rate in the individual subgroups of the model derivation group was closely correlated to that in the corresponding subgroups of the external validation group (Fig. 2; correlation coefficient $r^2 = 0.981$). The HCC development rate in the subgroup of patients with the highest risk of HCC development (high-risk group) according to the model older age (≥ 60 years) with a lower platelet count ($< 150 \times 10^9/L$) and lower albumin levels (< 4.0 g/dl) was 20.9% in the model derivation

Table 2. Multivariable analysis of factors associated with subsequent development of HCC within 5 years.

	Odds ratio	95% CI	p value
Age	1.086	1.029-1.146	0.003
Albumin	0.248	0.100-0.613	0.003
Platelet	0.842	0.769-0.921	< 0.0001

CI, confidence interval.

group and 17.9% in the external validation group. The intermediate-risk group or the patients with an HCC development rate of at least 5% consisted of the following three subgroups: (1) older age (≥ 60 years), lower platelet count ($< 150 \times 10^9/L$), higher albumin levels (≥ 4.0 g/dl), and higher AST levels (≥ 40 IU/L); (2) older age (≥ 60 years), higher platelet count ($\geq 150 \times 10^9/L$), and lower albumin levels (< 3.75 g/dl); and (3) younger age (< 60 years) and lower platelet count ($< 150 \times 10^9/L$). In these intermediate-risk groups, the 5-year HCC development rate was 6.3–7.3% in the model derivation group and 5.3–7.9% in the external validation group. The low-risk group consisted of the following three subgroups: (1) younger age (< 60 years) and higher platelet count ($\geq 150 \times 10^9/L$); (2) older age (≥ 60 years), lower platelet count ($< 150 \times 10^9/L$), higher albumin levels (≥ 4.0 g/dl), and lower AST levels (< 40 IU/L); and (3) older age (≥ 60 years), higher platelet count ($\geq 150 \times 10^9/L$), and higher albumin levels (≥ 3.75 g/dl). In these low-risk groups, the 5-year HCC development rate was 0–1.5% in the model derivation group and 0–2.9% in the external validation group.

Predictability of the HCC risk prediction model on HCC development rate beyond 5 years

Cumulative HCC development rates in the high-, intermediate-, and low-risk groups were compared over time using the Kaplan–Meier method. The 10-year rates were 28.9% in the high-risk group, 22.9% in the intermediate-risk group, and 4.8% in the low-risk group (Fig. 3A). The high and intermediate-risk group created by pooling data from the high- and intermediate-risk groups had a significantly higher cumulative HCC development rate than the low-risk group beyond 5 years (Fig. 3B; 5-year rate, 11.6% vs. 1.0%; 10-year rate, 24.5% vs. 4.8%; $p < 0.0001$).

Effect of response to PEG-IFN plus RBV therapy in the reduction of HCC development: analysis stratified by the HCC risk prediction model

The 600 nonSVR patients and 472 SVR patients in the external cohort were fitted into the HCC risk prediction model and

Research Article

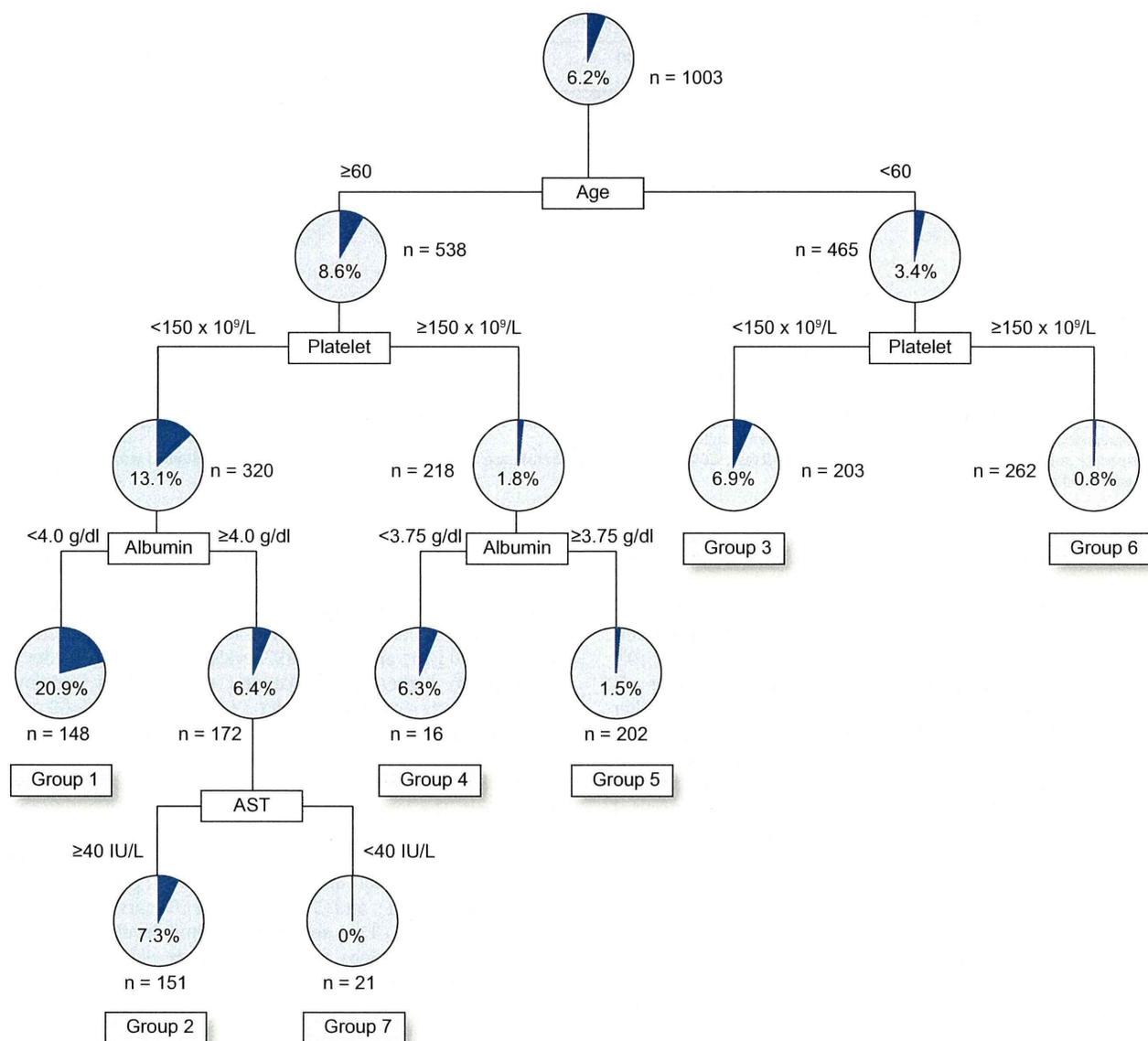


Fig. 1. The decision tree model of HCC development within 5 years. Boxes indicate the factors used to differentiate patients and the cutoff values for those different groups. Pie charts indicate the HCC development rate within 5 years for each group of patients after differentiation. Terminal groups of patients differentiated by analysis are numbered from 1 to 7.

classified into the high- and intermediate-risk group or the low-risk group, as defined above. The HCC development rate was significantly lower in SVR patients than in nonSVR patients in the high- and intermediate-risk group (5-year HCC rate, 9.5% vs. 4.5%; $p = 0.040$, log-rank test). In the low-risk group, the 5-year rate was 1.8% in nonSVR patients and 0.9% in SVR patients. Both rates were low and not significantly different ($p = 0.331$, log-rank test) (Fig. 4).

Discussion

An awareness of the risk of HCC development in the context of routine care for chronic hepatitis C is essential for formulating

an HCC surveillance plan personalized for individual patients. The risk of developing HCC from chronic hepatitis is lower than that from cirrhosis [7]; therefore, across-the-board surveillance for chronic hepatitis C is not recommended [3]. A method to easily determine this risk, without performing serial liver biopsies, would be extremely significant clinically. In the present study, an HCC risk prediction model that included the factors such as age, platelet count, albumin levels, and AST levels was constructed. The model was found to have excellent reproducibility when validated with an external cohort. This model could identify subgroups of chronic hepatitis C patients at high risk of HCC development; the 5-year HCC development rate for the high- and intermediate-risk groups was 11.6%, yielding an annual incidence of 2.3%. This HCC risk prediction model requires only

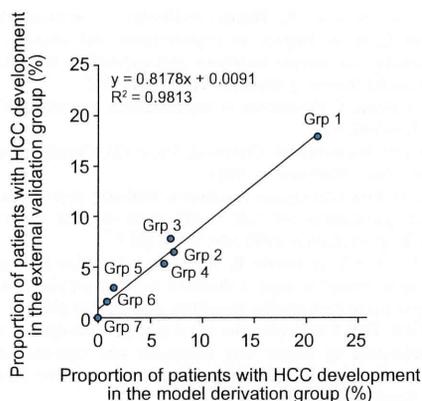


Fig. 2. External validation of the decision tree model with an independent cohort. Each patient in the external validation group was allocated to groups 1–7 following the flowchart of the decision tree. The HCC development rates were then calculated for each group and the graph plotted. The x-axis represents the HCC development rate in the model derivation group, and the y-axis represents the HCC development rate in the external validation group. The HCC development rates in each subgroup of patients are closely correlated between the model derivation group and the external validation group (correlation coefficient: $R^2 = 0.981$).

simple test values that are readily obtained in routine care and can therefore be easily used at the patient bedside. The model can be used to identify patients with a high risk of HCC development and therefore requiring surveillance, thereby allowing the formulation of surveillance plans personalized for individual patients.

Advanced fibrosis has been reported as independent risk factors for HCC development [7,8]. Platelet counts and albumin levels, which were factors selected for discrimination of the risk of HCC development, are closely related to the stage of fibrosis. Their correlation with the HCC risk has been repeatedly demonstrated [9–11,29–31]. The present study confirmed the impact of old age and advanced fibrosis, as reflected by low platelet counts and albumin levels. These results are consistent with our previous report [32]. What is unique to the present study was the study design to build a simple and reliable model for

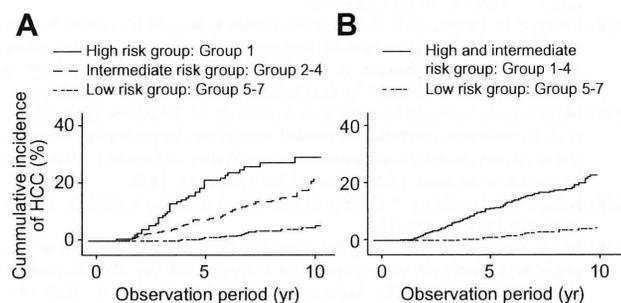


Fig. 3. Cumulative incidence of HCC development beyond 5 years in subgroups of patients defined by the decision tree model. Cumulative incidences of HCC in the groups classified by the decision tree model are compared. (A) The cumulative HCC development rate beyond 5 years is higher in the high- (group 1) and intermediate-risk (groups 2–4) groups compared to the low-risk group (groups 5–7). (B) The high and intermediate-risk group created by pooling data from the high- and intermediate-risk groups has a significantly higher cumulative HCC development rate than the low-risk group (5-year rate, 11.6% vs. 1.0%; 10-year rate, 24.5% vs. 4.8%; $p < 0.0001$).

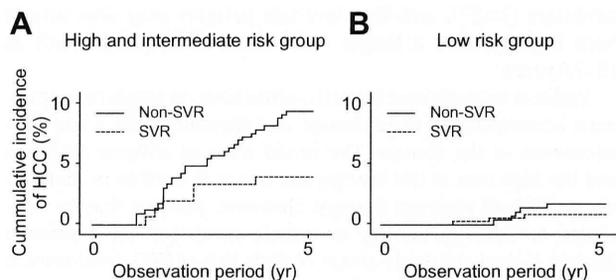


Fig. 4. Sustained virological response to PEG-IFN plus RBV therapy reduces the incidence of HCC development after stratification by the HCC risk. The 600 nonSVR patients and the 472 SVR patients in the external cohort were fitted into the HCC risk prediction model and classified into the high and intermediate-risk group or the low-risk group. The HCC development rate is significantly lower in SVR patients than in nonSVR patients in the high and intermediate-risk group (groups 1–4) (5-year HCC rate, 9.5% vs. 4.5%; $p = 0.040$). In the low-risk group (groups 5–7), the 5-year rate is 1.8% in nonSVR patients and 0.9% in SVR patients. Both rates are low and not significantly different ($p = 0.331$).

the prediction of HCC development that could be easily used in the clinic. For this purpose, a novel statistical method was used, histological factors were excluded in the analysis, the model derivation cohort was restricted to those who had nonSVR and had a long follow-up period duration (5 years), and the reproducibility of the model was independently validated by an external cohort. These are the major differences of the present study compared to our previous report. Many researchers have put a lot of efforts to formulate regression models for HCC prediction [9,10,33]. These prediction models are useful for identifying high-risk patients but are somewhat complicated to use at the bedside because they require calculations to be performed. Our prediction model is used simply by incorporating patients' data obtained through simple tests into the decision tree and following the flowchart. These prediction models based on factors easily accessible in routine clinical settings help physicians identify high-risk patients out of chronic hepatitis.

Viral eradication is the short-term goal of IFN therapy, but the ultimate goal is the prevention of HCC occurrence. Previous reports have shown that SVR to IFN therapy suppresses HCC occurrence in patients with type C liver cirrhosis and chronic hepatitis [7,12,30,34,35]. However, there is a marked heterogeneity in the magnitude of the treatment effect on the risk of HCC among studies, probably due to differences in the baseline risk of HCC among different trials [12]. Thus, the question remains whether the preventive effect of IFN therapy on HCC development could apply to all patients with chronic hepatitis C, especially those without liver cirrhosis. The result of the present study indicated that among high- and intermediate-risk patients, as assessed with our HCC risk prediction model, the cumulative HCC development rate was significantly reduced in SVR patients compared with nonSVR patients. This finding suggests that patients with chronic hepatitis, in whom disease has not yet progressed to hepatic cirrhosis but who are at a high risk of HCC development, benefit from antiviral treatment. The preventive effect of IFN on HCC development was not evident in low-risk patients within 5 years of observation. A longer observation term may be required to analyze the possible effect of antiviral therapy in these patients. Application of the present model on treatment decision may have limitations in that effect to prevent HCC development may differ in newer therapeutic agents such as protease

Research Article

inhibitors [36,37], and that low-risk patients may also benefit from therapy after a longer term observation period such as 15–20 years.

Patients with chronic hepatitis often have no subjective symptoms accompanying their disease and therefore have a low consciousness of the disease. The broad array of adverse reactions and the high cost of IFN therapy are frequent hurdles in motivating patients to undergo therapy. However, patients may be convinced to undergo therapy or remain motivated for continued therapy if they are made aware of their risk of HCC development and the preventive effect of IFN on HCC development.

In conclusion, a reproducible HCC risk prediction model, which includes the factors such as age, platelet count, albumin levels, and AST levels, was constructed to predict the 5-year HCC development rate in patients with chronic hepatitis C. The model requires only a combination of readily available test values and can therefore be easily used at the bedside. The information provided by the model allows the physician to identify patients requiring IFN therapy for the prevention of HCC and formulate plans for imaging HCC surveillance.

Conflict of interest

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

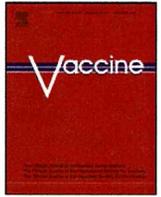
Financial support

This study was supported by a Grant-in-Aid from the Ministry of Health, Labor and Welfare, Japan (H20-kanen-006).

References

- [1] Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;55:74–108.
- [2] Bosch FX, Ribes J, Diaz M, Cleries R. Primary liver cancer: worldwide incidence and trends. *Gastroenterology* 2004;127:S5–S16.
- [3] Bruix J, Sherman M. Management of hepatocellular carcinoma. *Hepatology* 2005;42:1208–1236.
- [4] Fattovich G, Giustina G, Degos F, Tremolada F, Diodati G, Almasio P, et al. Morbidity and mortality in compensated cirrhosis type C: a retrospective follow-up study of 384 patients. *Gastroenterology* 1997;112:463–472.
- [5] Niederau C, Lange S, Heintges T, Erhardt A, Buschkamp M, Hurter D, et al. Prognosis of chronic hepatitis C: results of a large, prospective cohort study. *Hepatology* 1998;28:1687–1695.
- [6] Degos F, Christidis C, Ganne-Carrie N, Farmachidi JP, Degott C, Guettier C, et al. Hepatitis C virus related cirrhosis: time to occurrence of hepatocellular carcinoma and death. *Gut* 2000;47:131–136.
- [7] Yoshida H, Tateishi R, Arakawa Y, Sata M, Fujiyama S, Nishiguchi S, et al. Benefit of interferon therapy in hepatocellular carcinoma prevention for individual patients with chronic hepatitis C. *Gut* 2004;53:425–430.
- [8] Kurosaki M, Hosokawa T, Matsunaga K, Hirayama I, Tanaka T, Sato M, et al. Hepatic steatosis in chronic hepatitis C is a significant risk factor for developing hepatocellular carcinoma independent of age, sex, obesity, fibrosis stage and response to interferon therapy. *Hepatology* 2010;40:870–877.
- [9] Velazquez RF, Rodriguez M, Navascues CA, Linares A, Perez R, Sotorrios NG, et al. Prospective analysis of risk factors for hepatocellular carcinoma in patients with liver cirrhosis. *Hepatology* 2003;37:520–527.
- [10] Lok AS, Seeff LB, Morgan TR, di Bisceglie AM, Sterling RK, Curto TM, et al. Incidence of hepatocellular carcinoma and associated risk factors in hepatitis C-related advanced liver disease. *Gastroenterology* 2009;136:138–148.
- [11] Cardoso AC, Mouchari R, Figueiredo-Mendes C, Ripault MP, Giully N, Castelnau C, et al. Impact of peginterferon and ribavirin therapy on hepatocellular carcinoma: incidence and survival in hepatitis C patients with advanced fibrosis. *J Hepatol* 2010;52:652–657.
- [12] Craxi A, Camma C. Prevention of hepatocellular carcinoma. *Clin Liver Dis* 2005;9:329–346, viii.
- [13] Breiman IJH, Friedman RA, Olshen CJ, Stone CM. Classification and regression trees. Calif: Wadsworth; 1980.
- [14] Garzotto M, Park Y, Mongoue-Tchokote S, Bledsoe J, Peters L, Blank BH, et al. Recursive partitioning for risk stratification in men undergoing repeat prostate biopsies. *Cancer* 2005;104:1911–1917.
- [15] 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;12:243–248.
- [16] 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.
- [17] 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;22:3660–3667.
- [18] Valera VA, Walter BA, Yokoyama N, Koyama Y, Iai 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.
- [19] 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;11:5440–5443.
- [20] 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;75:2007–2014.
- [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. *Hepatology* 2010;40:251–260.
- [22] 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;83:445–452.
- [23] Izumi N, Asahina Y, Kurosaki M. Predictors of virological response to a combination therapy with pegylated interferon plus ribavirin including virus and host factors. *Hepat Res Treat* 2010;2010:703602.
- [24] 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;46:401–409.
- [25] 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;54:439–448.
- [26] Kurosaki M, Hiramatsu N, Sakamoto M, Suzuki Y, Iwasaki M, Tamori A, et al. Age and total ribavirin dose is an independent predictor of relapse among early virological responders to peg-interferon plus ribavirin therapy in chronic hepatitis C revealed by data mining analysis. *Antivir Ther*, in press.
- [27] Hiramatsu N, Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, et al. Pretreatment prediction of anemia progression by pegylated interferon alpha-2b plus ribavirin combination therapy in chronic hepatitis C infection: decision-tree analysis. *J Gastroenterol* 2011;46:1111–1119.
- [28] LeBlanc M, Crowley J. A review of tree-based prognostic models. *Cancer Treat Res* 1995;75:113–124.
- [29] Lu SN, Wang JH, Liu SL, Hung CH, Chen CH, Tung HD, et al. Thrombocytopenia as a surrogate for cirrhosis and a marker for the identification of patients at high-risk for hepatocellular carcinoma. *Cancer* 2006;107:2212–2222.
- [30] Bruno S, Stroffolini T, Colombo M, Bollani S, Benvegna L, Mazzella G, et al. Sustained virological response to interferon-alpha is associated with improved outcome in HCV-related cirrhosis: a retrospective study. *Hepatology* 2007;45:579–587.
- [31] Bonis PA, Tong MJ, Blatt LM, Conrad A, Griffith JL. A predictive model for the development of hepatocellular carcinoma, liver failure, or liver transplantation for patients presenting to clinic with chronic hepatitis C. *Am J Gastroenterol* 1999;94:1605–1612.

- [32] Asahina Y, Tsuchiya K, Tamaki N, Hirayama I, Tanaka T, Sato M, et al. Effect of aging on risk for hepatocellular carcinoma in chronic hepatitis C virus infection. *Hepatology* 2010;52:518–527.
- [33] Ganne-Carrie N, Chastang C, Chapel F, Munz C, Pateron D, Sibony M, et al. Predictive score for the development of hepatocellular carcinoma and additional value of liver large cell dysplasia in Western patients with cirrhosis. *Hepatology* 1996;23:1112–1118.
- [34] Nishiguchi S, Kuroki T, Nakatani S, Morimoto H, Takeda T, Nakajima S, et al. Randomised trial of effects of interferon-alpha on incidence of hepatocellular carcinoma in chronic active hepatitis C with cirrhosis. *Lancet* 1995;346:1051–1055.
- [35] Shiratori Y, Ito Y, Yokosuka O, Imazeki F, Nakata R, Tanaka N, et al. Antiviral therapy for cirrhotic hepatitis C: association with reduced hepatocellular carcinoma development and improved survival. *Ann Intern Med* 2005;142:105–114.
- [36] 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;360:1839–1850.
- [37] 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;360:1827–1838.



New neutralizing antibody epitopes in hepatitis C virus envelope glycoproteins are revealed by dissecting peptide recognition profiles[☆]

Alla Kachko^a, Galina Kochneva^b, Galina Sivolobova^b, Antonina Grazhdantseva^b, Tatyana Lupan^b, Iryna Zubkova^a, Frances Wells^a, Michael Merchlinsky^c, Ollie Williams^c, Hisayoshi Watanabe^d, Alla Ivanova^b, Aleksander Shvalov^b, Valeriy Loktev^b, Sergei Netesov^b, Marian E. Major^{a,*}

^a Laboratory of Hepatitis Viruses, Division of Viral Products, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, MD 20892, USA

^b Laboratory of Viral Hepatitis, State Research Center of Virology and Biotechnology "VECTOR", Koltsovo, Novosibirsk 630559, Russian Federation

^c Laboratory of DNA Viruses, Division of Viral Products, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, MD 20892, USA

^d Yamagata University School of Medicine, Department of Gastroenterology, 2-2-2 Iida-nishi, Yamagata 990-9585, Japan

ARTICLE INFO

Article history:

Received 15 February 2011
Received in revised form 22 August 2011
Accepted 18 October 2011
Available online 29 October 2011

Keywords:

Vaccine
Cross neutralization
Binding affinity
Antibody blocking
Antibody purification

ABSTRACT

One of the greatest challenges to HCV vaccine development is the induction of effective immune responses using recombinant proteins or vectors. In order to better understand which vaccine-induced antibodies contribute to neutralization of HCV the quality of polyclonal anti-E1E2 antibody responses in immunized mice and chimpanzees was assessed at the level of epitope recognition using peptide scanning and neutralization of chimeric 1a/2a, 1b/2a and 2a HCVcc after blocking or affinity elution of specific antibodies. Mice and chimpanzees were immunized with genotype 1a (H77) HCV gpE1E2; all samples contained cross-neutralizing antibody against HCVcc. By functionally dissecting the polyclonal immune responses we identified three new regions important for neutralization within E1 (aa264–318) and E2 (aa448–483 and aa496–515) of the HCV glycoproteins, the third of which (aa496–515) is highly conserved (85–95%) amongst genotypes. Antibodies to aa496–515 were isolated by affinity binding and elution from the serum of a vaccinated chimpanzee and found to specifically neutralize chimeric 1a/2a, 1b/2a and 2a HCVcc. IC50 titres (IgG ng/mL) for the aa496–515 eluate were calculated as 142.1, 239.37 and 487.62 against 1a/2a, 1b/2a and 2a HCVcc, respectively. Further analysis demonstrated that although antibody to this new, conserved neutralization epitope is efficiently induced with recombinant proteins in mice and chimpanzees; it is poorly induced during natural infection in patients and chimpanzees (7 out of 68 samples positive) suggesting the epitope is poorly presented to the immune system in the context of the viral particle. These findings have important implications for the development of HCV vaccines and strategies designed to protect against heterologous viruses. The data also suggest that recombinant or synthetic antigens may be more efficient at inducing neutralizing antibodies to certain epitopes and that screening virally infected patients may not be the best approach for finding new cross-reactive epitopes.

Published by Elsevier Ltd.

Abbreviations: HCVcc, cell cultured Hepatitis C virus; 1a/2a, chimeric HCVcc containing core-E1E2-NS2 fragment of 1a genotype; 1b/2a, chimeric HCVcc containing core-E1E2-NS2 fragment of 1b genotype; rE1E2, recombinant HCV E1E2 protein; VV-E1E2, rE1E2 protein purified from Vaccinia virus system; SIN-E1E2, rE1E2 protein purified from Sindbis virus system; VV-AB, antibody from mice immunized against recombinant protein VV-E1E2; SIN-AB, antibody from mice immunized against recombinant protein SIN-E1E2; ID50 titre, inhibitory dose 50 titre (the titre determined to neutralize 50% of the virus); NMS, normal mouse serum; EPI, epitope 1 HCV glycoproteins E1E2; EPII, epitope II HCV glycoproteins E1E2.

[☆] The findings and conclusions in this article have not been formally disseminated by the Food and Drug Administration and should not be construed to represent any Agency determination or policy.

* Corresponding author at: Laboratory of Hepatitis Viruses, Division of Viral Products, Bldg 29A/Rm 1D10/HFM 448, 8800 Rockville Pike, Bethesda, MD 20892, USA. Tel.: +1 301 827 1881; fax: +1 301 496 1810.

E-mail addresses: alla.kachko@fda.hhs.gov (A. Kachko), kochneva@vector.nsc.ru (G. Kochneva), sgf@vector.nsc.ru (G. Sivolobova), gaa@vector.nsc.ru (A. Grazhdantseva), Lupan@ngs.ru (T. Lupan), Iryna.zubkova@fda.hhs.gov (I. Zubkova), frances.wells@fda.hhs.gov (F. Wells), Michael.Merchlinsky@hhs.gov (M. Merchlinsky), ollie.williams@fda.hhs.gov (O. Williams), h-watanabe@med.id.yamagata-u.ac.jp (H. Watanabe), ali@vector.nsc.ru (A. Ivanova), shvalov@mail.ru (A. Shvalov), loktev@vector.nsc.ru (V. Loktev), nauka@nsu.ru (S. Netesov), marian.major@fda.hhs.gov (M.E. Major).