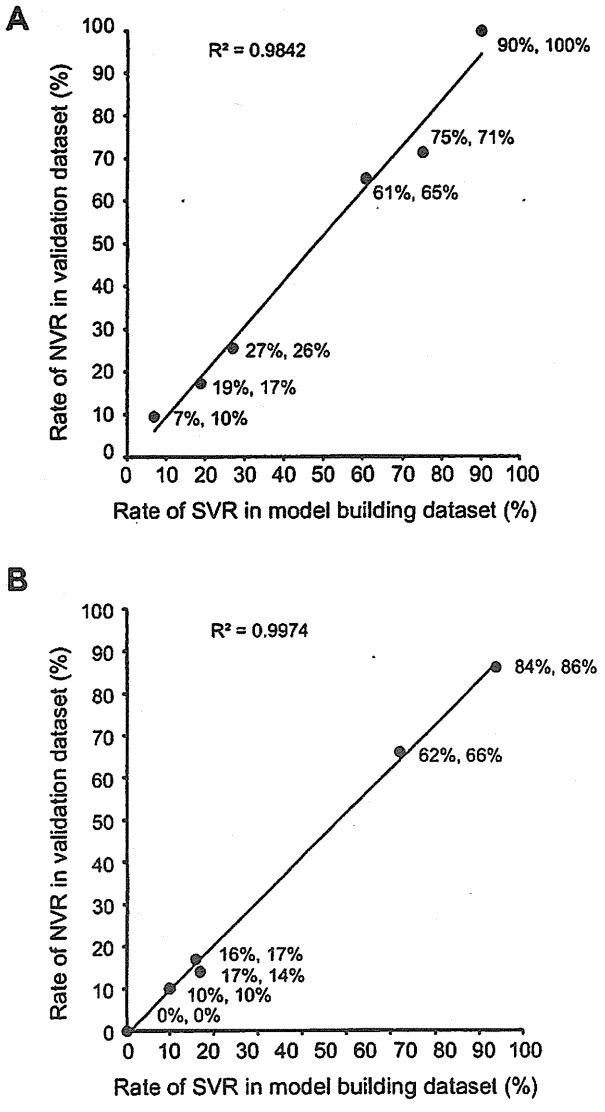

Fig. 6. Validation of the CART analysis. Each patient in the validation group was allocated to one of the six subgroups by following the flow-chart form of the decision tree. The rate of (A) sustained virological response (SVR) and (B) null virological response (NVR) in each subgroup was calculated and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

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

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

Conflict of interest

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

Financial support

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

References

[1] Ray Kim W. Global epidemiology and burden of hepatitis C. Microbes Infect 2002;4 (12):1219–1225.
[2] Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales Jr FL, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med 2002;347 (13):975–982.
[3] Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet 2001;358 (9286):958–965.
[4] Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, Someya T, et al. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. Intervirology 2005;48 (6):372–380.
[5] Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. Hepatology 2003;38 (3):645–652.
[6] Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 2009;10:1105–1109.
[7] Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate ML, et al. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. Nat Genet 2009;10:1100–1104.
[8] Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 2009;461 (7262):399–401.
[9] Rauch A, Kutalik Z, Descombes P, Cai T, Di Iulio J, Mueller T, et al. Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. Gastroenterology 2010;138 (4):1338–1345.
[10] Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. J Clin Invest 1995;96 (1):224–230.
[11] Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N Engl J Med 1996;334 (2):77–81.
[12] Kurosaki M, Enomoto N, Murakami T, Sakuma I, Asahina Y, Yamamoto C, et al. Analysis of genotypes and amino acid residues 2209 to 2248 of the NS5A region of hepatitis C virus in relation to the response to interferon-beta therapy. Hepatology 1997;25 (3):750–753.
[13] Shirakawa H, Matsumoto A, Joshita S, Komatsu M, Tanaka N, Umemura T, et al. Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors. Hepatology 2008;48 (6):1753–1760.
[14] Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, et al. Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. J Hepatol 2007;46 (3):403–410.

[15] Okanoue T, Itoh Y, Hashimoto H, Yasui K, Minami M, Takehara T, et al. Predictive values of amino acid sequences of the core and NS5A regions in antiviral therapy for hepatitis C: a Japanese multi-center study. J Gastroenterol 2009;44 (9):952–963.
[16] Segal MR, Bloch DA. A comparison of estimated proportional hazards models and regression trees. Stat Med 1989;8 (5):539–550.
[17] LeBlanc M, Crowley J. A review of tree-based prognostic models. Cancer Treat Res 1995;75:113–124.
[18] Garzotto M, Beer TM, Hudson RG, Peters L, Hsieh YC, Barrera E, et al. Improved detection of prostate cancer using classification and regression tree analysis. J Clin Oncol 2005;23 (19):4322–4329.
[19] Averbook BJ, Fu P, Rao JS, Mansour EG. A long-term analysis of 1018 patients with melanoma by classic Cox regression and tree-structured survival analysis at a major referral center: implications on the future of cancer staging. Surgery 2002;132 (4):589–602.
[20] Leiter U, Buettner PG, Eigentler TK, Garbe C. Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the german dermatological society. J Clin Oncol 2004;22 (18):3660–3667.
[21] Valera VA, Walter BA, Yokoyama N, Koyama Y, Iiai T, Okamoto H, et al. Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. Ann Surg Oncol 2007;14 (1):34–40.
[22] Zlobec I, Steele R, Nigam N, Compton CC. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. Clin Cancer Res 2005;11 (15):5440–5443.
[23] Thabane M, Simunovic M, Akhtar-Danesh N, Marshall JK. Development and validation of a risk score for post-infectious irritable bowel syndrome. Am J Gastroenterol 2009;104 (9):2267–2274.
[24] Wu BU, Johannes RS, Sun X, Tabak Y, Conwell DL, Banks PA. The early prediction of mortality in acute pancreatitis: a large population-based study. Gut 2008;57 (12):1698–1703.
[25] Fonarow GC, Adams Jr KF, Abraham WT, Yancy CW, Boscardin WJ. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. Jama 2005;293 (5):572–580.
[26] Kurosaki M, Matsunaga K, Hirayama I, Tanaka T, Sato M, Yasui Y, et al. A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. Hepatol Res 2010;40 (3):251–260.
[27] Nishida N, Tanabe T, Takasu M, Suyama A, Tokunaga K. Further development of multiplex single nucleotide polymorphism typing method, the DigiTag2 assay. Anal Biochem 2007;364 (1):78–85.
[28] Hezode C, Forestier N, Dusheiko G, Ferenci P, Pol S, Goeser T, et al. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. N Engl J Med 2009;360 (18):1839–1850.
[29] McHutchison JG, Everson GT, Gordon SC, Jacobson IM, Sulkowski M, Kauffman R, et al. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. N Engl J Med 2009;360 (18):1827–1838.
[30] Rossignol JF, Elfert A, El-Gohary Y, Keeffe EB. Improved virologic response in chronic hepatitis C genotype 4 treated with nitazoxanide, peginterferon, and ribavirin. Gastroenterology 2009;136 (3):856–862.
[31] Marcello T, Grakoui A, Barba-Spaeth G, Machlin ES, Kotenko SV, MacDonald MR, et al. Interferons alpha and lambda inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics. Gastroenterology 2006;131 (6):1887–1898.
[32] Robek MD, Boyd BS, Chisari FV. Lambda interferon inhibits hepatitis B and C virus replication. J Virol 2005;79 (6):3851–3854.
[33] McCarthy JJ, Li JH, Thompson A, Suchindran S, Lao XQ, Patel K, et al. Replicated association between an IL28B Gene Variant and a Sustained Response to Pegylated Interferon and Ribavirin. Gastroenterology 2010;138:2307–2314.
[34] Tanaka Y, Nishida N, Sugiyama M, Tokunaga K, Mizokami M. A-interferons and the single nucleotide polymorphisms: a milestone to tailor-made therapy for chronic hepatitis C. Hepatol Res 2010;40:449–460.
[35] Backus LI, Boothroyd DB, Phillips BR, Mole LA. Predictors of response of US veterans to treatment for the hepatitis C virus. Hepatology 2007;46 (1):37–47.
[36] Mori N, Imamura M, Kawakami Y, Saneto H, Kawaoka T, Takaki S, et al. Randomized trial of high-dose interferon-alpha-2b combined with ribavirin in patients with chronic hepatitis C: correlation between amino acid substitutions in the core/NS5A region and virological response to interferon therapy. J Med Virol 2009;81 (4):640–649.
[37] Hung CH, Lee CM, Lu SN, Lee JF, Wang JH, Tung HD, et al. Mutations in the NS5A and E2-PePHD region of hepatitis C virus type 1b and correlation with the response to combination therapy with interferon and ribavirin. J Viral Hepat 2003;10 (2):87–94.
[38] Yen YH, Hung CH, Hu TH, Chen CH, Wu CM, Wang JH, et al. Mutations in the interferon sensitivity-determining region (nonstructural 5A amino acid 2209–2248) in patients with hepatitis C-1b infection and correlating response to combined therapy of pegylated interferon and ribavirin. Aliment Pharmacol Ther 2008;27 (1):72–79.
[39] Zeuzem S, Lee JH, Roth WK. Mutations in the nonstructural 5A gene of European hepatitis C virus isolates and response to interferon alfa. Hepatology 1997;25 (3):740–744.
[40] Squadrito G, Leone F, Sartori M, Nalpas B, Berthelot P, Raimondo G, et al. Mutations in the nonstructural 5A region of hepatitis C virus and response of chronic hepatitis C to interferon alfa. Gastroenterology 1997;113 (2):567–572.
[41] Sarrazin C, Berg T, Lee JH, Teuber G, Dietrich CF, Roth WK, et al. Improved correlation between multiple mutations within the NS5A region and virological response in European patients chronically infected with hepatitis C virus type 1b undergoing combination therapy. J Hepatol 1999;30 (6):1004–1013.
[42] Murphy MD, Rosen HR, Marousek GI, Chou S. Analysis of sequence configurations of the ISDR, PKR-binding domain, and V3 region as predictors of response to induction interferon-alpha and ribavirin therapy in chronic hepatitis C infection. Dig Dis Sci 2002;47 (6):1195–1205.
[43] Pascu M, Martus P, Hohne M, Wiedenmann B, Hopf U, Schreier E, et al. Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences. Gut 2004;53 (9):1345–1351.
[44] Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 2009;461 (7265):798–801.
[45] Kurosaki M, Enomoto N, Marumo F, Sato C. Evolution and selection of hepatitis C virus variants in patients with chronic hepatitis C. Virology 1994;205 (1):161–169.
[46] Enomoto N, Kurosaki M, Tanaka Y, Marumo F, Sato C. Fluctuation of hepatitis C virus quasispecies in persistent infection and interferon treatment revealed by single-strand conformation polymorphism analysis. J Gen Virol 1994;75 (Pt 6):1361–1369.
[47] Li JH, Lao XQ, Tillmann HL, Rowell J, Patel K, Thompson A, et al. Interferon-lambda genotype and low serum low-density lipoprotein cholesterol levels in patients with chronic hepatitis C infection. Hepatology 1904;51 (6):1904–1911.
[48] Serfaty L, Andreani T, Giral P, Carbonell N, Chazouilleres O, Poupon R. Hepatitis C virus induced hypobetalipoproteinemia: a possible mechanism for steatosis in chronic hepatitis C. J Hepatol 2001;34 (3):428–434.

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

ORIGINAL ARTICLE—LIVER, PANCREAS, AND BILIARY TRACT

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

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

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

Abstract

Background This study aimed to develop a model to predict the development of severe anemia during pegylated interferon alpha-2b plus ribavirin combination therapy.

Methods Data were collected from 1081 genotype 1b chronic hepatitis C patients who were treated at 6 hospitals in Japan. These patients were randomly assigned to a model-building group ($n = 691$) or an internal validation group ($n = 390$). Factors predictive of severe anemia (hemoglobin, Hb < 8.5 g/dl) were explored using data-mining analysis.

Results Hb values at baseline, creatinine clearance (Ccr), and an Hb concentration decline by 2 g/dl at week 2 were used to build a decision-tree model, in which the patients were divided into 5 subgroups based on variable rates of severe anemia ranging from 0.4 to 11.8%. The reproducibility of the model was confirmed by the internal validation group ($r^2 = 0.96$). The probability of severe anemia was high in patients whose Hb value was <14 g/dl before treatment (6.5%), especially (a) in those whose Ccr was <80 ml/min (11.8%) and (b) those whose Ccr was ≥80 ml/min but whose Hb concentration decline at week 2 was ≥2 g/dl (11.5%). The probability of severe anemia was low in the other patients (0.4–2.5%).

Conclusions The decision-tree model that included Hb values at baseline, Ccr, and an Hb concentration decline by

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

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

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

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

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

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

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

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

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

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

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

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

Abbreviations

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

Introduction

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

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

Materials and methods

Patients

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

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

Laboratory tests

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

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

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

Data are expressed as median ± standard deviation unless otherwise indicated

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

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

Database of variables and decision-tree analysis

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

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

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

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

Results

Decision-tree analysis

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

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

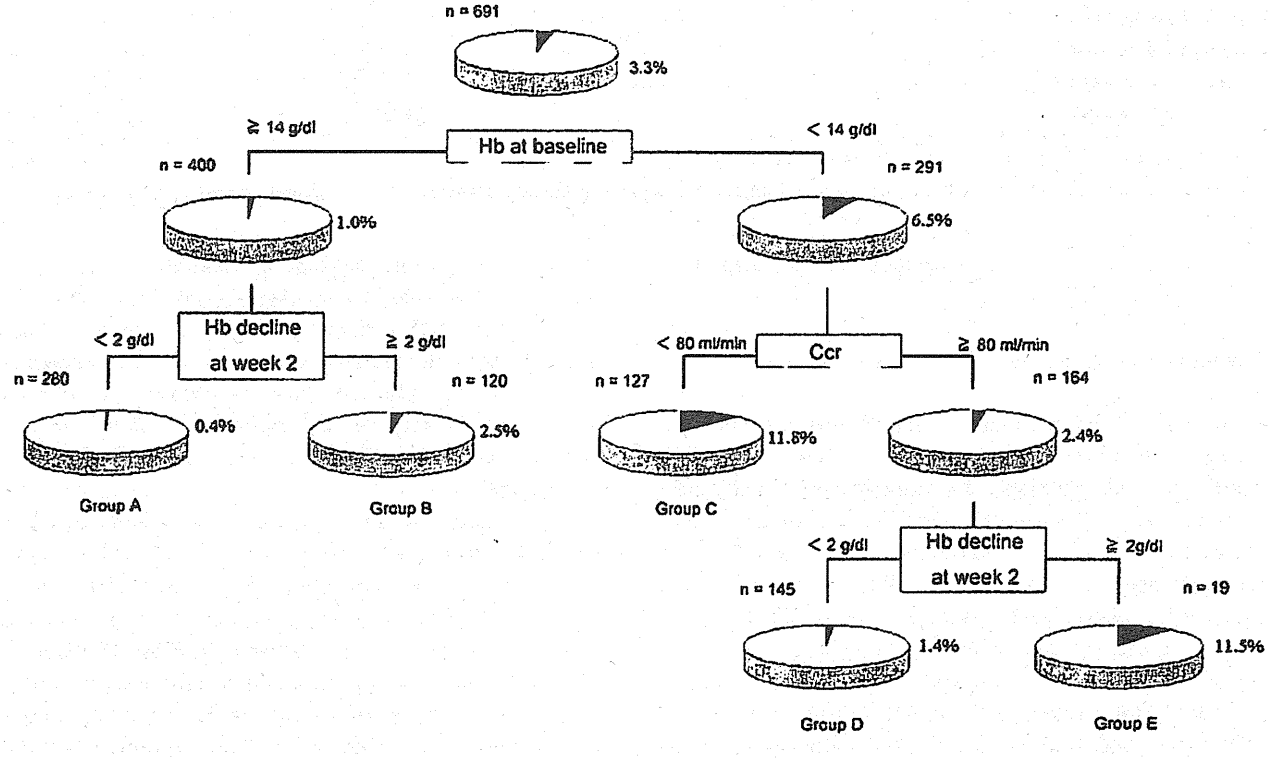

Fig. 1 Decision-tree analysis. *Boxes* indicate the splitting factors and the cutoff value for the split. *Pie charts* indicate the rate of severe anemia (Hb < 8.5 g/dl) for each group. Terminal groups classified by the analysis were labeled from A to E. *Hb* hemoglobin, *Ccr* creatine clearance

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

Validation of the decision tree

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

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

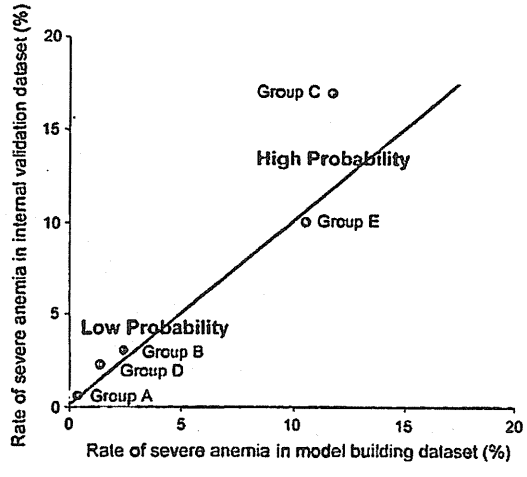

Fig. 2 Validation of decision-tree analysis with the internal validation dataset: subgroup-stratified comparison of the rate of severe anemia. The rate of severe anemia in each subgroup was plotted. The *X*-axis represents the model-building dataset and the *Y*-axis represents the internal validation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

Factors associated with severe anemia determined by multivariate logistic regression analysis

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

Discussion

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

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

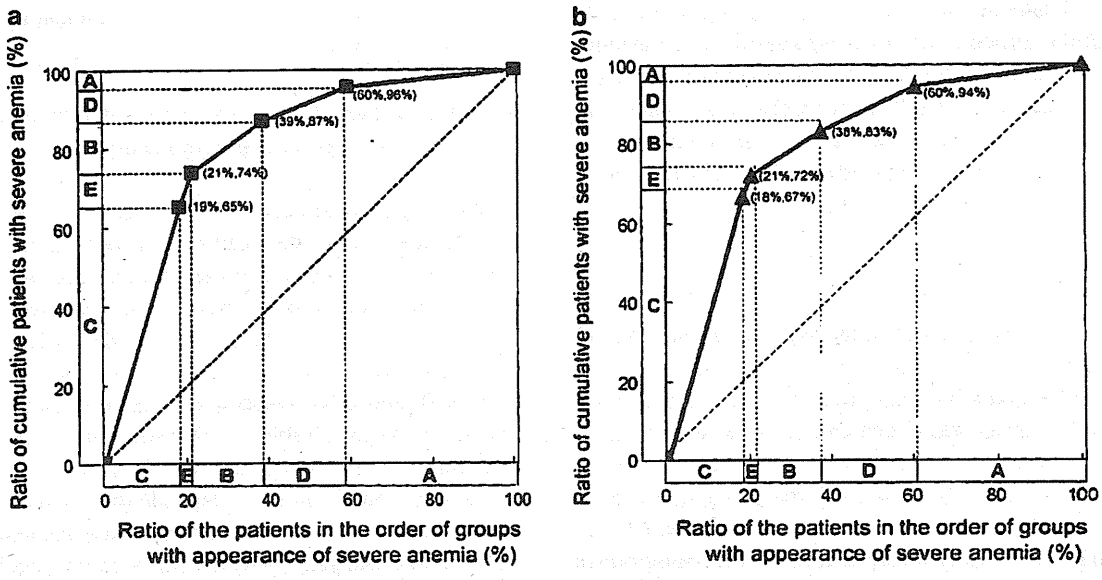

Fig. 3 Validation of the efficiency and stability by the discrimination efficiency curve. a Model-building group and b internal validation group. The groups were sorted in the order of incidence rate of severe anemia and validated using the correlation between cumulative cases (%) and the cumulative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

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

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

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

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

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

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

Data are expressed as median ± standard deviation unless otherwise indicated

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

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

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

Hb hemoglobin, *Ccr* creatinine clearance

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

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

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

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

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

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

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

References

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

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

Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY

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

Masayuki Kurosaki[1], Naoki Hiramatsu[2], Minoru Sakamoto[3], Yoshiyuki Suzuki[4], Manabu Iwasaki[5], Akihiro Tamori[6], Kentaro Matsuura[7], Sei Kakinuma[8], Fuminaka Sugauchi[9], Naoya Sakamoto[8], Mina Nakagawa[8], Namiki Izumi[1,*]

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

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 ⩾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

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 ($\geqslant 150 \times 10^9$/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 ($\geqslant 150 \times 10^9$/L). The third predictor was albumin levels,

Table 1. Baseline characteristics of patients for model deviation 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/m²)	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 ($\geq$3.75 g/dl). The cutoff value for albumin levels was 4.0 g/dl in patients with a lower platelet count ($<150 \times 10^9$/L). The HCC development rate was 20.9% when albumin levels were lower (<4.0 g/dl) and 6.4% when levels were higher ($\geq$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 ($\geq$60 years) with a lower platelet count ($<150 \times 10^9$/L) and lower albumin levels (<4.0 g/dl) was 20.9% in the model derivation group and 17.9% in the external validation group. The intermediate-risk group or the patients with an HCC development rate of at least 5% consisted of the following three subgroups: (1) older age ($\geq$60 years), lower platelet count ($<150 \times 10^9$/L), higher albumin levels ($\geq$4.0 g/dl), and higher AST levels ($\geq$40 IU/L); (2) older age ($\geq$60 years), higher platelet count ($\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$/L); (2) older age ($\geq$60 years), lower platelet count ($<150 \times 10^9$/L), higher albumin levels ($\geq$4.0 g/dl), and lower AST levels (<40 IU/L); and (3) older age ($\geq$60 years), higher platelet count ($\geq 150 \times 10^9$/L), and higher albumin levels ($\geq$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.

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.

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

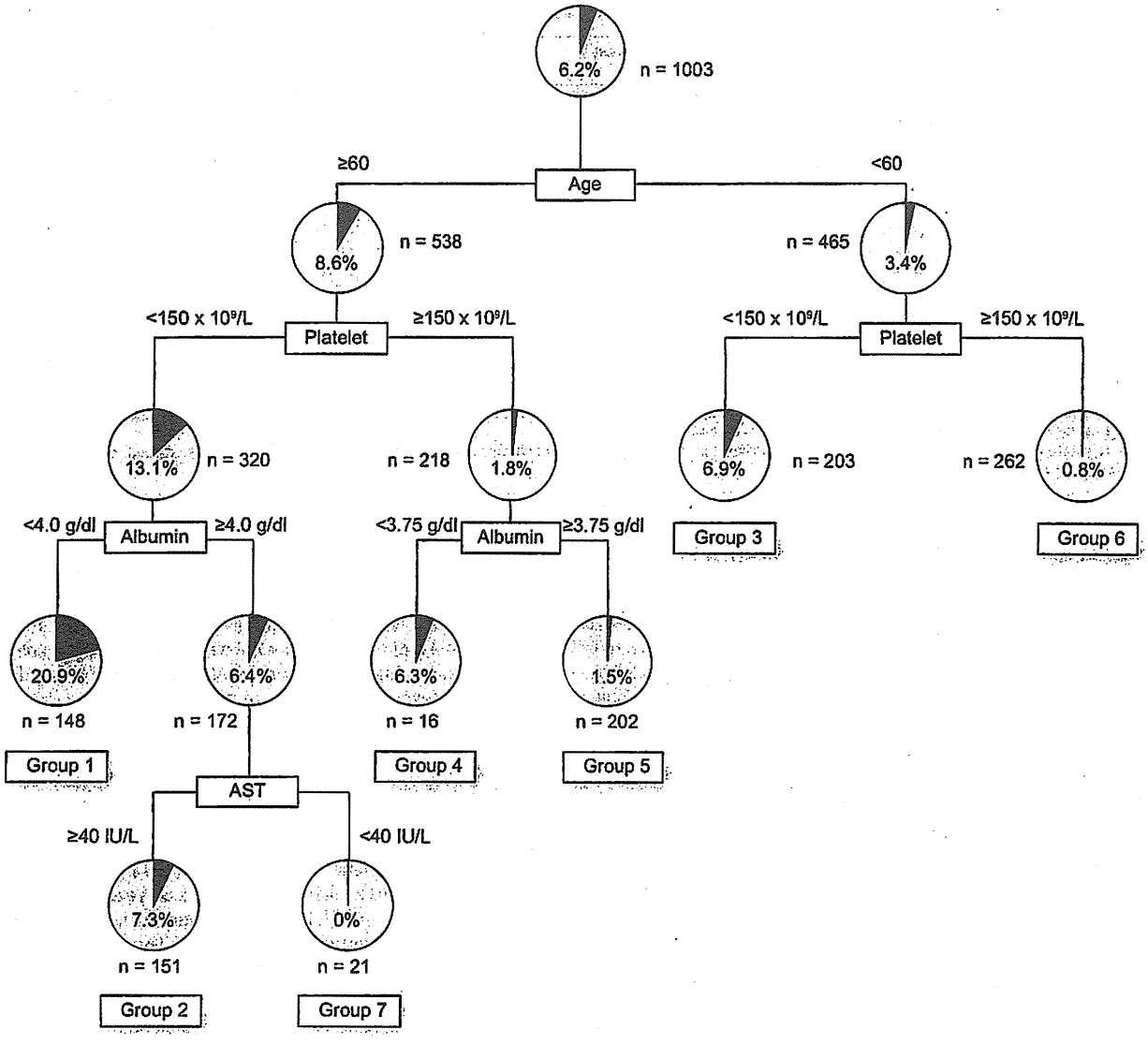

Fig. 1. The decision tree model of HCC development within 5 years. Boxes indicate the factors used to differentiate patients and the cutoff values for those different groups. Pie charts indicate the HCC development rate within 5 years for each group of patients after differentiation. Terminal groups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

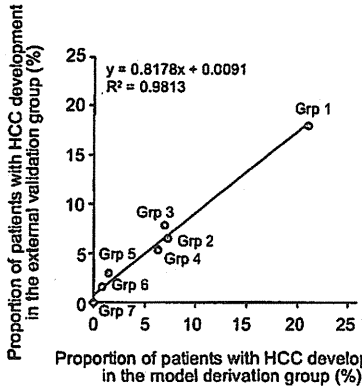

Fig. 2. External validation of the decision tree model with an independent cohort. Each patient in the external validation group was allocated to groups 1–7 following the flowchart of the decision tree. The HCC development rates were then calculated for each group and the graph plotted. The *x*-axis 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

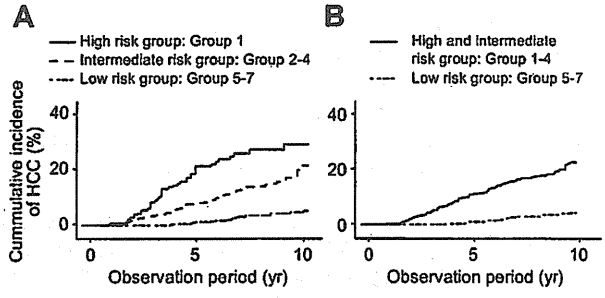

Fig. 3. Cumulative incidence of HCC development beyond 5 years in subgroups of patients defined by the decision tree model. Cumulative incidences of HCC in the groups classified by the decision tree model are compared. (A) The cumulative HCC development rate beyond 5 years is higher in the high- (group 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%; $p < 0.0001$).

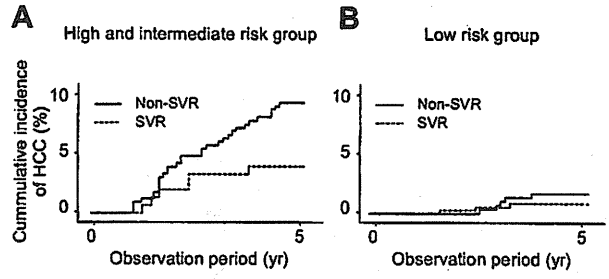

Fig. 4. Sustained virological response to PEG-IFN plus RBV therapy reduces the incidence of HCC development after stratification by the HCC risk. The 600 nonSVR patients and the 472 SVR patients in the external cohort were fitted into the HCC risk prediction model and classified into the high and intermediate-risk 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

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. Hepatol Res 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, Moucari R, Figueiredo-Mendes C, Ripault MP, Giuily 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 LJH, 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, 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.
[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. Hepatol Res 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, Benvegnu 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.

Analysis of the Complete Open Reading Frame of Genotype 2b Hepatitis C Virus in Association with the Response to Peginterferon and Ribavirin Therapy

Makoto Kadokura[1☯], Shinya Maekawa[1*☯], Ryota Sueki[1], Mika Miura[1], Kazuki Komase[1], Hiroko Shindo[1], Fumitake Amemiya[1], Tomoyoshi Uetake[1], Taisuke Inoue[1], Minoru Sakamoto[1], Mina Nakagawa[2], Naoya Sakamoto[2], Mamoru Watanabe[2], Nobuyuki Enomoto[1]

1 First Department of Internal Medicine, Faculty of Medicine, University of Yamanashi, Chuo, Yamanashi, Japan, 2 Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Bunkyo, Tokyo, Japan

Abstract

Background and Aims: Patients infected with genotype 2b hepatitis C virus (HCV) generally can achieve favorable responses to pegylated-interferon plus ribavirin therapy (PEG-IFN/RBV). However, a proportion of patients show poorer responses and the correlation between viral sequence variation and treatment outcome remains unclear.

Methods: The pretreatment complete open reading frame (ORF) sequences of genotype 2b HCV determined by direct sequencing were investigated for correlation with the final outcome in a total of 60 patients.

Results: In this study group, 87.5% (14/16) of non-sustained virological response (non-SVR) patients (n = 16) were relapsers. Compared to sustained virological response (SVR) patients (n = 44), non-SVR patients were older and could not achieve prompt viral clearance after the therapy induction. Comparing each viral protein between the two groups, viral sequences were more diverse in SVR patients and that diversity was found primarily in the E1, p7, and NS5A proteins. In searching for specific viral regions associated with the final outcome, several regions in E2, p7, NS2, NS5A, and NS5B were extracted. Among these regions, part of the interferon sensitivity determining region (ISDR) was included. In these regions, amino acid substitutions were associated with the final outcome in an incremental manner, depending upon the number of substitutions.

Conclusions: Viral sequences are more diverse in SVR patients than non-SVR patients receiving PEG-IFN/RBV therapy for genotype-2b HCV infection. Through systematic comparison of viral sequences, several specific regions, including part of the ISDR, were extracted as having significant correlation with the final outcome.

Citation: Kadokura M, Maekawa S, Sueki R, Miura M, Komase K, et al. (2011) Analysis of the Complete Open Reading Frame of Genotype 2b Hepatitis C Virus in Association with the Response to Peginterferon and Ribavirin Therapy. PLoS ONE 6(9): e24514. doi:10.1371/journal.pone.0024514

Editor: John E. Tavis, Saint Louis University, United States of America

Received March 10, 2011; **Accepted** August 12, 2011; **Published** September 15, 2011

Funding: This study was supported in part by a grant-in-aid scientific research fund of the Ministry of Education, Science, Sports and Culture number 20390206 and in part by a grant-in-aid from the Ministry of Health, Labour, and Welfare of Japan (H19-kanen-002). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding was received for this study.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: maekawa@yamanashi.ac.jp

☯ These authors contributed equally to this work.

Introduction

Worldwide, 180 million people are estimated to be infected with hepatitis C virus (HCV), a major cause of chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma (HCC) [1]. In HCV-infected patients with chronic hepatitis, treatment with interferon (IFN)-based therapy can result in viral clearance as well as biochemical and histological improvements [2]. In this IFN-based therapy, HCV genotype is the most significant factor affecting treatment responses [3,4].

In genotype 2b HCV infection, 80% of patients with high viral titers can achieve a sustained virological response (SVR) to the regimen of pegylated-interferon (PEG-IFN) plus ribavirin (RBV) for 24 weeks [5,6]. This response is high considering that much lower percentages of patients infected with other genotypes can achieve SVR, especially with genotype 1 [1]. However, in other words, 20% of patients infected with genotype 2b HCV still cannot clear the virus and remain at risk of developing HCC. On the other hand, although various studies have been undertaken to clarify the factors contributing to the response to IFN-based therapy in genotype 1 infection, it remains poorly understood which patients with genotype 2b HCV infection will show unfavorable responses. Recently, the significance of IL28B single nucleotide polymorphisms (SNPs) in determining the response to PEG-IFN/RBV therapy was demonstrated in genotype 1 HCV infection [7,8]. However, the significance of IL28B SNPs was rather weak in genotype 2 HCV infection [9].

In terms of the association between HCV sequence variation and treatment responses, previous studies have reported that

Table 1. Baseline Characteristics of Studied Patients.

Characteristic	SVR (n=44)	non-SVR (n=16)	*P* value
Gender (Male/Female)	26/18	9/7	NS†
Age (yrs)	56 (22-72)*	59 (30-80)	0.04‡
BMI	23.5 (16.6-30.3)	24.7 (18.5-31.7)	NS‡
ALT (IU/l)	51 (19-380)	41 (17-390)	NS‡
GGTP (IU/l)	36 (11-133)	40 (17-292)	NS‡
T.Chol (mg/dl)	169 (119-225)	178 (145-217)	NS‡
WBC (/μl)	4600 (2620-7200)	5080 (3270-8600)	NS‡
Hb (g/dl)	14.2 (11.5-17.3)	14.6 (11.8-16.4)	NS‡
Platelet ($\times 10^4/mm^3$)	19 (7.1-31.8)	17.8 (8-36.7)	NS‡
Fibrosis score (0-2/≥3)§	38/5	7/3	NS†
HCV RNA (KIU/ml)	2050 (100-16000)	1800 (140-6300)	NS‡
IFN dose (≥80%/60-80%)	36/8	13/3	NS†
Ribavirin dose (≥80%/60-80%)	32/12	10/6	NS†
RVR rate (%)	55.8	6.3	0.0008†
EVR rate (%)	97.7	68.8	0.004†
ETR rate (%)	100	87.5	NS†

§: SVR : n=43, non-SVR : n=10.
*: median (range).
†: Fisher's exact probability test.
‡: Mann-Whitney's U test.
doi:10.1371/journal.pone.0024514.t001

amino acid variation in the NS5A-ISDR [10], NS5A-IRRDR [11], NS5B [12], PKR-eIF2 phosphorylation homology domain (PePHD) of E2 [13], and Core [14] correlate with the clinical outcome of IFN-based therapy, including PEG-IFN/RBV therapy for genotype 1b HCV infection. In the meantime, these viral sequence studies have been controversial regarding their true clinical importance, because the results of different studies were not always coincident [15,16,17]. On this background, recent studies trying to analyze the correlation of complete HCV open reading frame diversity, clinical characteristics, and the response to PEG-IFN/RBV therapy for genotype 1 HCV infection, in the most comprehensive approach yet attempted, have clarified that viral amino acid variation is associated with treatment responses, with consideration of racial background [18,19]. In genotype 2 infection, however, only a few studies have investigated the association of HCV sequence variation and treatment response [20,21] and the clinical significance has been yet established. We reported recently that variation of amino acid (aa) 110 in Core and amino acids (aa) 2258–2308 in NS5A were significantly associated with treatment outcome of the PEG-IFN/RBV therapy for genotype 2a HCV infection, through the analysis of the complete HCV ORFs in Japanese patients [22].

In this study, to assess comprehensively the influence of viral sequence variation on the response to the PEG-IFN/RBV therapy in genotype 2b HCV infection, we determined the complete pretreatment HCV ORFs from Japanese patients and investigated amino acid variation and its correlation with the response to combination therapy with PEG-IFN plus RBV.

Methods

Patients

A total of 77 adult Japanese patients infected with genotype 2b HCV, who received the combination therapy with PEG-IFN

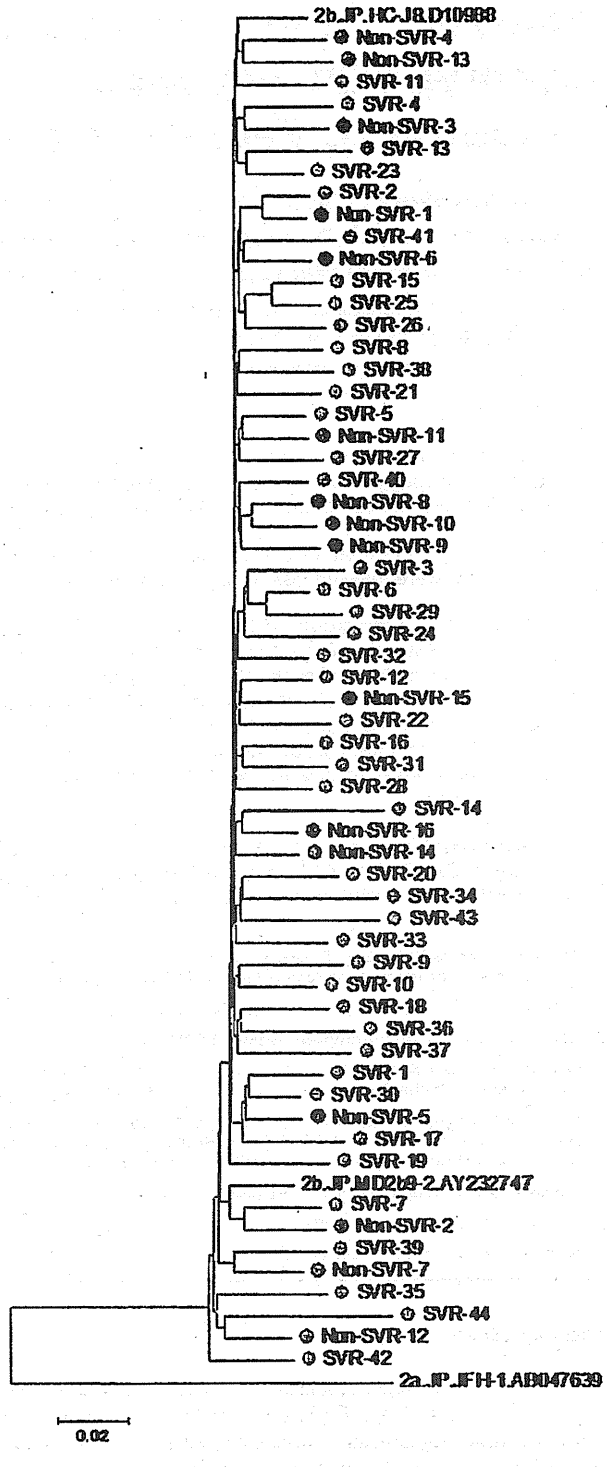

Figure 1. Phylogenetic analysis of the genotype-2b polyprotein sequences. In order to perform the phylogenetic analysis, we first aligned all 60 HCV complete ORF amino acid sequences obtained from the patients along with reference sequences (2b.HC-J8.D10988, 2.JP.MD2b9-2, and 2a.JP.JFH-1.AB047639), using the ClustalW program, and constructed the phylogenetic tree using the Neighbor-Joining method with MEGA version 4 software. Blue circles indicate SVR patients and red circles indicate non-SVR patients.
doi:10.1371/journal.pone.0024514.g001