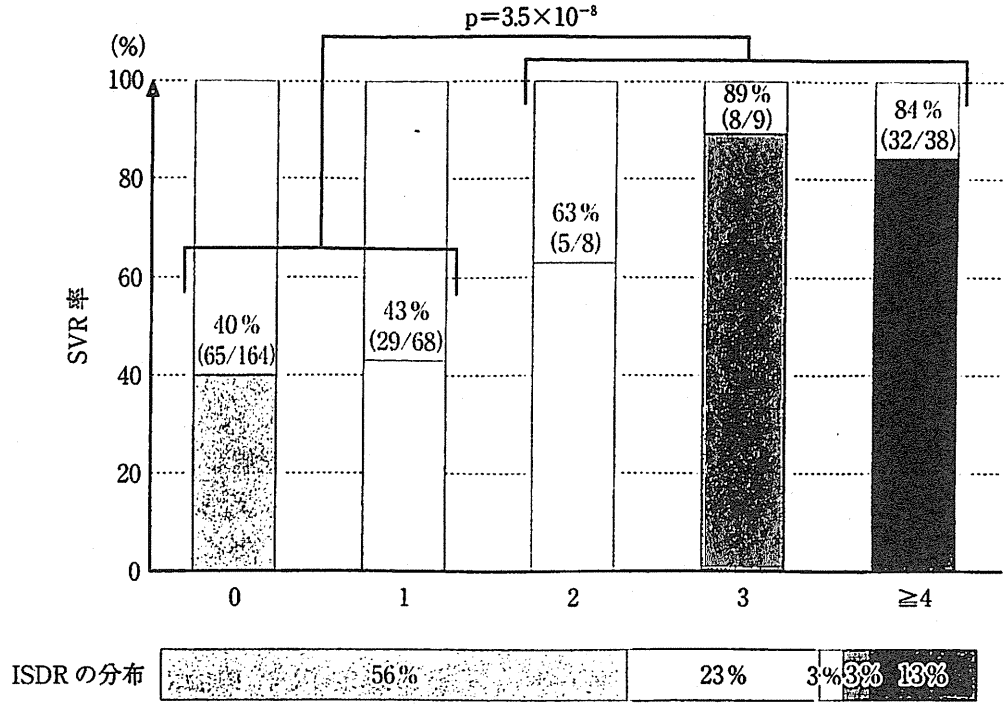


図4 PEG-IFN+RBV 療法(12カ月)の ISDR 変異数別 SVR 率(n=292)

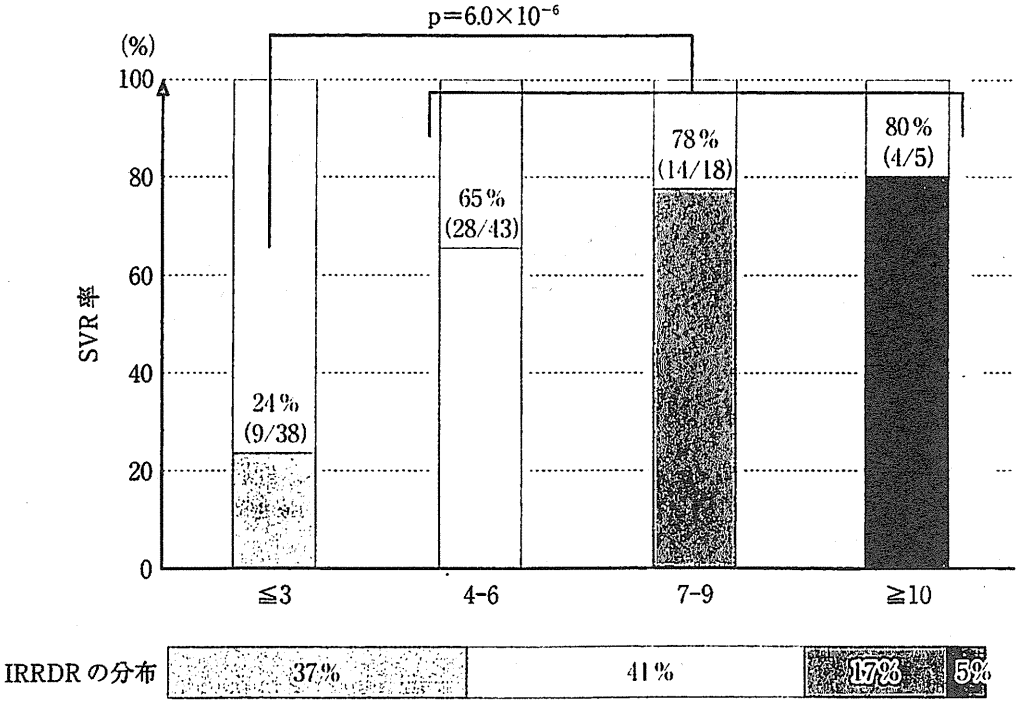


図5 PEG-IFN+RBV 療法(12カ月)の IRRDR 変異数別 SVR 率(n=104)

4. 宿主側因子(IL28B)と NS5A 変異との関連

他方，ペグインターフェロン＋リバビリン療法の治療効果に関連する宿主(ヒト)ゲノムワイド関連解析によって，19番染色体上のインターフェロンラムダ遺伝子近傍の IL28B 領域の1塩基多型(single nucleotide polymorphism: SNP)

表1 PEG-IFN+RBV療法(12カ月)のSVRに寄与する因子(n=292)

		odds比	95%CI	p
年齢	<60/≧60	0.097	0.021-0.453	0.0030
ISDR変異数	0/≧1	7.989	1.651-38.663	0.0098
IRRDR変異数	≦3/≧4	18.626	3.367-103.050	0.0008
IL28B(rs8099917)	GG+TG/TT	20.240	3.091-132.553	0.0017

多変量ロジステック回帰分析

が同定された．すなわちIL28B遺伝子座の代表的SNPであるrs8099917のminor allele(T/G+G/G)をもつ症例では，major allele(T/T)をもつ症例に比較して約38倍，無効となりやすいことが示された[10]．著者らの検討でも，ペグインターフェロン+リバビリンを標準期間治療した107例の検討では，IL28B SNPの頻度はT/Tが78%，T/G+G/Gが22%で，SVR率はそれぞれ59%，17%で有意にT/Tの場合，高率であった(p=0.001)．そこで，年齢・性別・肝線維化，肝機能検査値などとともに，IL28B遺伝子多型を含めて，治療効果を規定する因子の多変量解析を行ったところ，有意な因子は年齢(60歳未満)，ISDR変異数(≧1)，IRRDR変異数(≧4)，IL28B SNP(T/T)であり，ウイルス側因子であるISDRやIRRDRと，宿主因子であるIL28Bは独立して治療効果を規定する因子であることが明らかになった(表1)．

文献

1) Enomoto N, 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 334: 77-81, 1996.
2) Enomoto N, 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 96: 224-230, 1995.
3) El-Shamy A, et al: Sequence variation in hepatitis C virus nonstructural protein 5A predicts clinical outcome of pegylated interferon/ribavirin combination therapy. Hepatology 48: 38-47, 2008.
4) Blight KJ, et al: Efficient initiation of HCV RNA replication in cell culture. Science 290: 1972-1974, 2000.
5) Maekawa S, et al: Introduction of NS5A mutations enables subgenomic HCV replicon derived from chimpanzee-infectious HC-J4 isolate to replicate efficiently in Huh-7 cells. J Viral Hepat 11: 394-403, 2004.
6) 坂本 穣, 榎本信幸：Interferon sensitivity determining region: ISDR. 肝胆膵 57(5): 773-779, 2008.
7) Prouski V, Holmes EC: SWAN: sliding window analysis of nucleotide sequence variability. Bioinformatics 14: 467-468, 1998.
8) 坂本 穣, 榎本信幸：ウイルス性慢性肝炎：診断と治療の進歩. 日内会誌 97(1): 57-63, 2008.
9) 坂本 穣, 榎本信幸：遺伝子変異からみたC型慢性肝炎に対するインターフェロン治療効果予測. 日消誌 106: 485-492, 2009.
10) Tanaka Y, et al: Genome-wide association of IL28B with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C. Nat Genet 41: 1105-1109, 2009.

Journal of Medical Virology 83:445–452 (2011)

Sequences in the Interferon Sensitivity-Determining Region and Core Region of Hepatitis C Virus Impact Pretreatment Prediction of Response to PEG-Interferon Plus Ribavirin: Data Mining Analysis

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

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

The aim of the present study was to clarify the significance of viral factors for pretreatment prediction of sustained virological response to pegylated-interferon (PEG-IFN) plus ribavirin (RBV) therapy for chronic hepatitis C using data mining analysis. Substitutions in the IFN sensitivity-determining region (ISDR) and at position 70 of the HCV core region (Core70) were determined in 505 patients with genotype 1b chronic hepatitis C treated with PEG-IFN plus RBV. Data mining analysis was used to build a predictive model of sustained virological response in patients selected randomly (n = 304). The reproducibility of the model was validated in the remaining 201 patients. Substitutions in ISDR (odds ratio = 9.92, $P < 0.0001$) and Core70 (odds ratio = 1.92, $P = 0.01$) predicted sustained virological response independent of other covariates. The decision-tree model revealed that the rate of sustained virological response was highest (83%) in patients with two or more substitutions in ISDR. The overall rate of sustained virological response was 44% in patients with a low number of substitutions in ISDR (0–1) but was 83% in selected subgroups of younger patients (<60 years), wild-type sequence at Core70, and higher level of low-density lipoprotein cholesterol (LDL-C) (≥120 mg/dl). Reproducibility of the model was validated ($r^2 = 0.94$, $P < 0.001$). In conclusion, substitutions in ISDR and Core70 of HCV are significant predictors of response to PEG-IFN plus RBV therapy. A decision-tree model that includes these viral factors as predictors could identify patients with a high probability of sustained virological response. *J. Med. Virol. 83:445–452, 2011.*

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

INTRODUCTION

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

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

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

Accepted 26 October 2010
DOI 10.1002/jmv.22005
Published online in Wiley Online Library (wileyonlinelibrary.com).

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

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

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

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

MATERIALS AND METHODS

Patients

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

Laboratory Tests

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

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

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

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

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

Statistical Analysis

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

Student's *t*-test was used for the univariable comparison of quantitative variables and Fisher's exact test was used for the comparison of qualitative variables. For the multivariable analysis for factors associated with sustained virological response, logistic regression models with backward selection were used to identify independent predictors of sustained virological response. Variables that showed significant association with sustained virological response by univariable analysis were included in the multivariable analysis. IBM-SPSS software v.15.0 (SPSS, Inc., Chicago, IL) was used for these analyses. For the decision-tree analysis [Segal and Bloch, 1989], the data mining software IBM SPSS Modeler 13 (IBM SPSS, Inc.) was used, as reported previously [Kurosaki et al., 2010a,b]. In brief, the software searched for the optimal split variables to build a decision-tree structure. The entire study population was first evaluated to determine the variables and cut-off points for the most significant division into two subgroups having different probabilities of sustained virological response. Thereafter, analysis was repeated on all subgroups in the same way until either no additional significant variable was detected or the sample size was below 20.

RESULTS

Generation of the Decision-Tree Model

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

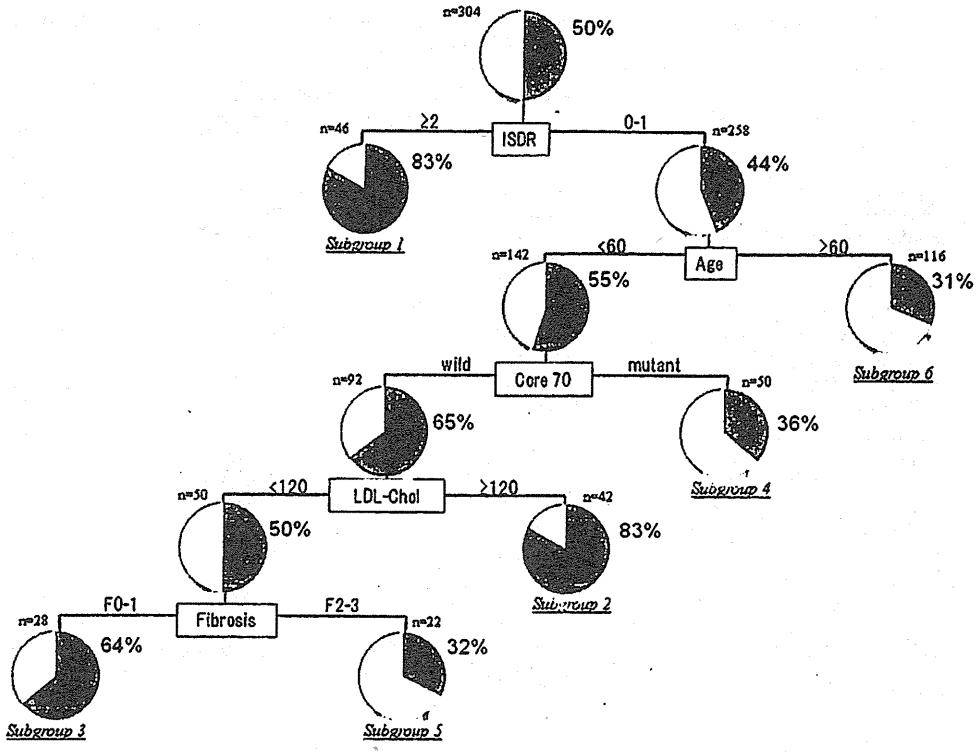


Fig. 1. Decision-tree model. Boxes indicate the factors used for splitting and the cutoff value for the split. Pie charts indicate the rate of sustained virological response for each group of patients after splitting. Terminal subgroups of patients discriminated by the analysis are numbered from 1 to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

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

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

Validation of the Decision-Tree Model

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

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

Effect of Dose Reductions of PEG-IFN and RBV

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

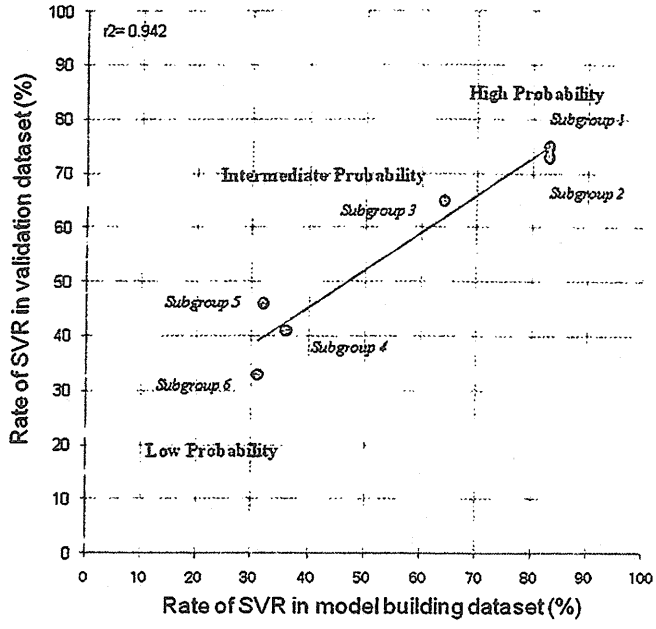


Fig. 2. Validation of the decision-tree analysis: Subgroup-stratified comparison of the rate of sustained virological response. Each patient in the validation set was allocated to subgroups 1–6 by following the flowchart form of the decision tree, and the rates of sustained virological response were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

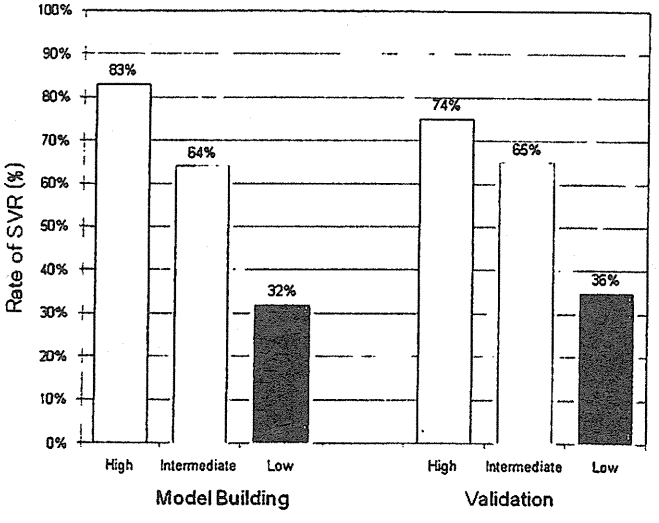


Fig. 3. Comparison of sustained virological response rates between groups divided by the decision tree. The rate of sustained virological response was compared between three groups of patients as divided by the decision-tree analysis. Black, gray, and white boxes indicate the low-probability group (subgroup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

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

Multivariable Logistic Regression Analysis

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

DISCUSSION

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

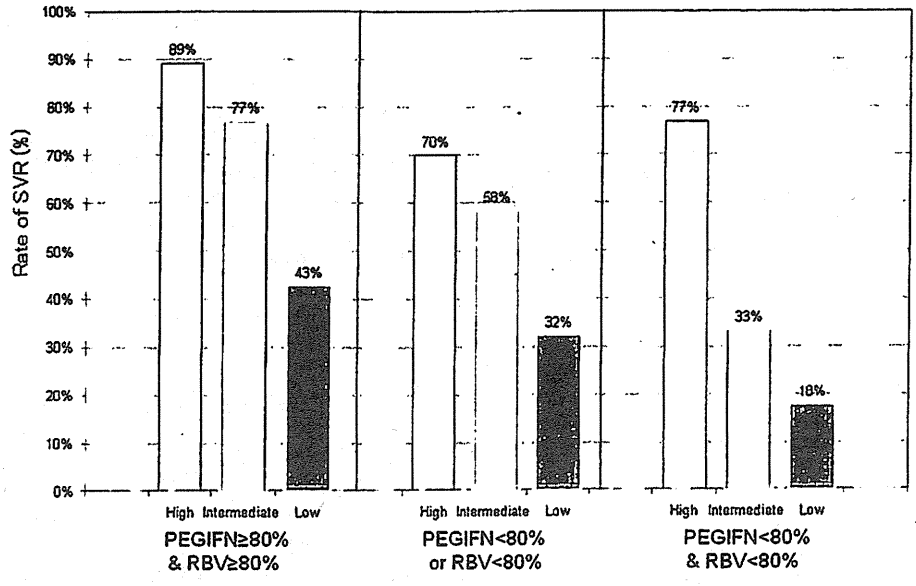


Fig. 4. Comparison of the rate of sustained virological response between the decision-tree groups stratified by drug adherence. The three groups of patients divided by the decision tree (black, gray, and white boxes indicating the low-, intermediate-, and high-probability groups, respectively) were further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

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

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

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

Parameter		Odds	95% CI	*P*-value
Age (years)	<60 vs. ≥60	2.28	1.31–3.94	0.003
Sex	Male vs. female	3.36	1.87–5.99	<0.0001
GGT (IU/L)	<40 vs. ≥40	2.65	1.45–4.85	0.002
LDL-C (mg/dl)	≥120 vs. <120	1.79	0.91–3.53	0.094
Platelets (109/L)	≥120 vs. <120	2.69	1.22–5.90	0.014
ISDR mutations	≥2 vs. 0–1	9.92	3.71–26.54	<0.0001
Core70	Wild vs. mutant	1.92	1.07–3.47	0.030

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

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

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

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

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

REFERENCES

Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, Someya T, Kobayashi M, Saitoh S, Watahiki S, Sato J, Matsuda M, Arase Y, Ikeda K, Kumada H. 2005. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. Intervirology 48:372–380.

Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Arase Y, Ikeda K, Kumada H. 2007a. Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: Amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. J Hepatol 46:403–410.

Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Arase Y, Ikeda K, Kumada H. 2007b. Predictors of viral kinetics to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b. J Med Virol 79:1686–1695.

Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Kumada H. 2009. A matched case-controlled study of 48 and 72 weeks of peginterferon plus ribavirin combination therapy in patients infected with HCV genotype 1b in Japan: Amino acid substitutions in HCV core region as predictor of sustained virological response. J Med Virol 81:452–458.

Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Chayama K, Nakamura Y, Kumada H. 2010. Amino acid substitution in hepatitis C virus core region and genetic variation near the interleukin 28B gene predict viral response to telaprevir with peginterferon and ribavirin. Hepatology 52:421–429.

Averbook BJ, Fu P, Rao JS, Mansour EG. 2002. A long-term analysis of 1018 patients with melanoma by classic Cox regression and tree-structured survival analysis at a major referral center: Implications on the future of cancer staging. Surgery 132:589–602.

Baquerizo A, Anselmo D, Shackleton C, Chen TW, Cao C, Weaver M, Gornbein J, Geevarghese S, Nissen N, Farmer D, Demetriou A, Busuttil RW. 2003. Phosphorus ans an early predictive factor in patients with acute liver failure. Transplantation 75:2007–2014.

Bedossa P, Poynard T. 1996. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. Hepatology 24:289–293.

Breiman LJH, Friedman RA, Olshen CJ, Stone CM. 1980. Classification and regression trees. CA: Wadsworth.

Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. 2003. Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. Hepatology 38:645–652.

Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Izumi N, Marumo F, Sato C. 1995. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. J Clin Invest 96:224–230.

Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Ogura Y, Izumi N, Marumo F, Sato C. 1996. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N Engl J Med 334:77–81.

Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL, , Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J. 2002. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med 347:975–982.

Garzotto M, Beer TM, Hudson RG, Peters L, Hsieh YC, Barrera E, Klein T, Mori M. 2005. Improved detection of prostate cancer using classification and regression tree analysis. J Clin Oncol 23:4322–4329.

Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, Heinzen EL, Qiu P, Bertelsen AH, Muir AJ, Sulkowski M, McHutchison JG, Goldstein DB. 2009. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 461:399–401.

Hezode C, Forestier N, Dusheiko G, Ferenci P, Pol S, Goeser T, Bronowicki JP, Bourliere M, Gharakhanian S, Bengtsson L, McNair L, George S, Kieffer T, Kwong A, Kauffman RS, Alam J, Pawlotsky JM, Zeuzem S. 2009. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. N Engl J Med 360: 1839–1850.

Ikeda H, Suzuki M, Okuse C, Yamada N, Okamoto M, Kobayashi M, Nagase Y, Takahashi H, Matsunaga K, Matsumoto N, Itoh F, Yotsuyanagi H, Koitabashi Y, Yasuda K, Iino S. 2009. Short-term prolongation of pegylated interferon and ribavirin therapy for genotype 1b chronic hepatitis C patients with early viral response. Hepatol Res 39:753–759.

Izumi N, Nishiguchi S, Hino K, Suzuki F, Kumada Y, Itoh Y, Asahina Y, Tamori A, Hiramatsu N, Hayashi N, Kudo M. 2010. Management of hepatitis C; Report of the consensus meeting at the 45th annual meeting of the Japan society of hepatology (2009). Hepatol Res 40:347–368.

Jensen DM, Morgan TR, Marcellin P, Pockros PJ, Reddy KR, Hadziyannis SJ, Ferenci P, Ackrill AM, Willems B. 2006. Early identification of HCV genotype 1 patients responding to 24 weeks peginterferon alpha-2a (40 kd)/ribavirin therapy. Hepatology 43:954–960.

Kurosaki M, Enomoto N, Murakami T, Sakuma I, Asahina Y, Yamamoto C, Ikeda T, Tozuka S, Izumi N, Marumo F, Sato C, Ogura Y. 1997. Analysis of genotypes and amino acid residues 2209 to 2248 of the NS5A region of hepatitis C virus in relation to the response to interferon-beta therapy. Hepatology 25:750–753.

Kurosaki M, Matsunaga K, Hirayama I, Tanaka T, Sato M, Yasui Y, Tamaki N, Hosokawa T, Ueda K, Tsuchiya K, Nakanishi H, Ikeda H, Itakura J, Takahashi Y, Asahina Y, Higak M, Enomoto N, Izumi N. 2010a. A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. Hepatol Res 40:251–260.

Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, Hiramatsu N, Sugauchi F, Yatsuhashi H, Izumi N. 2010b. Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis. J Gastroenterol DOI: 10.1007/s00535-010-0322-5.

Kurosaki M, Tanaka Y, Nishida N, Sakamoto N, Enomoto N, Honda M, Sugiyama M, Matsuura K, Sugauchi F, Asahina Y, Nakagawa M, Watanabe M, Sakamoto M, Maekawa S, Sakai A, Kaneko S, Ito K, Masaki N, Tokunaga K, Izumi N, Mizokami M. 2010c. Pretreatment prediction of response to pegylated-interferon plus ribavirin for chronic hepatitis C using genetic polymorphism in *IL28B* and viral factors. J Hepatol DOI: 10.1016/j.jhep.2010.07.037.

LeBlanc M, Crowley J. 1995. A review of tree-based prognostic models. Cancer Treat Res 75:113–124.

Lee SS, Ferenci P. 2008. Optimizing outcomes in patients with hepatitis C virus genotype 1 or 4. Antivir Ther 13:9–16.

Leiter U, Buettner PG, Eigentler TK, Garbe C. 2004. Prognostic factors of thin cutaneous melanoma: An analysis of the central malignant melanoma registry of the German Dermatological Society. J Clin Oncol 22:3660–3667.

Li JH, Lao XQ, Tillmann HL, Rowell J, Patel K, Thompson A, Suchindran S, Muir AJ, Guyton JR, Gardner SD, McHutchison JG, McCarthy JJ. 2010. Interferon-lambda genotype and low serum low-density lipoprotein cholesterol levels in patients with chronic hepatitis C infection. Hepatology 51:1904–1911.

Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, Goodman ZD, Koury K, Ling M, Albrecht JK. 2001. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: A randomised trial. Lancet 358:958–965.

McHutchison JG, Everson GT, Gordon SC, Jacobson IM, Sulkowski M, Kauffman R, McNair L, Alam J, Muir AJ. 2009. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. N Engl J Med 360:1827–1838.

Miyaki K, Takei I, Watanabe K, Nakashima H, Omae K. 2002. Novel statistical classification model of type 2 diabetes mellitus patients for tailor-made prevention using data mining algorithm. J Epidemiol 12:243–248.

Munoz de Rueda P, Casado J, Paton R, Quintero D, Palacios A, Gila A, Quiles R, Leon J, Ruiz-Extremera A, Salmeron J. 2008. Mutations in E2-PePHD, NS5A-PKRBD, NS5A-ISDR, and NS5A-V3 of hepatitis C virus genotype 1 and their relationships to pegylated interferon-ribavirin treatment responses. J Virol 82:6644–6653.

Segal MR, Bloch DA. 1989. A comparison of estimated proportional hazards models and regression trees. Stat Med 8:539–550.

Shirakawa H, Matsumoto A, Joshita S, Komatsu M, Tanaka N, Umemura T, Ichijo T, Yoshizawa K, Kiyosawa K, Tanaka E. 2008. Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors. Hepatology 48:1753–1760.

Shiratori Y, Omata M. 2000. Predictors of the efficacy of interferon therapy for patients with chronic hepatitis C before and during therapy: How does this modify the treatment course? J Gastroenterol Hepatol 15:E141–E151.

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

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

Valera VA, Walter BA, Yokoyama N, Koyama Y, Iiai T, Okamoto H, Hatakeyama K. 2007. Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. Ann Surg Oncol 14:34–40.

Yu ML, Dai CY, Huang JF, Chiu CF, Yang YH, Hou NJ, Lee LP, Hsieh MY, Lin ZY, Chen SC, Wang LY, Chang WY, Chuang WL. 2008. Rapid virological response and treatment duration for chronic hepatitis C genotype 1 patients: A randomized trial. Hepatology 47:1884–1893.

Zlobec I, Steele R, Nigam N, Compton CC. 2005. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. Clin Cancer Res 11:5440–5443.

Hepatol Int
DOI 10.1007/s12072-011-9307-6

ORIGINAL ARTICLE

Analysis of viral amino acids sequences and the IL28B SNP influencing the development of hepatocellular carcinoma in chronic hepatitis C

Mika Miura · Shinya Maekawa · Makoto Kadokura · Ryota Sueki · Kazuki Komase · Hiroko Shindo · Takako Ohmori · Asuka Kanayama · Kuniaki Shindo · Fumitake Amemiya · Yasuhiro Nakayama · Takatoshi Kitamura · Tomoyoshi Uetake · Taisuke Inoue · Minoru Sakamoto · Shunichi Okada · Nobuyuki Enomoto

Received: 23 April 2011 / Accepted: 22 July 2011

Abstract
Background and aims The association between hepatitis C virus (HCV) sequences with interleukin 28B (IL28B) single-nucleotide polymorphism (SNP) in the development of hepatocellular carcinoma (HCC) has not been well clarified.
Methods Complete HCV open-reading frame sequences were determined in 20 patients developing HCC and 23 non-HCC patients with HCV-1b infection in two distant time points. An additional 230 patients were studied cross-sectionally for core and NS5A sequences with HCC development. Among them, 98 patients with available samples were investigated for changes in viral core sequences over time. Finally, IL28B SNPs and HCC development were investigated in 228 patients.
Results During observation period (HCC for 10.8 years, and non-HCC for 11.1 years), changes in core a.a. 70 and three amino acid positions in NS5A were characteristics of the patients developing HCC. In 230 patients, Q (glutamine) or H (histidine) to R (arginine) ratio at core a.a. 70 was significantly higher in the HCC group (HCC group 43:22 vs. non-HCC group 66:99, $p = 0.001$). A change in core R70Q was observed over time in 11 patients associated with a decrease in platelets ($p = 0.005$) and albumin ($p = 0.005$), while a Q70R change was observed in 4 patients without associated changes in platelets (nonsignificant) and albumin (nonsignificant). IL28B SNP showed significant correlation with the core a.a. 70 residue. There was no evident link between IL28B SNPs and the occurrence of HCC.
Conclusions Hepatitis C virus core a.a. 70 residue is associated with liver disease progression and is independent factor for HCC development in genotype-1b infection. IL28B SNPs are related to core a.a. 70 residue, but not to HCC. The functional relevance of core a.a. 70 residue in hepatitis C pathogenesis should be further investigated.

Keywords HCV · HCC Core IL28B

Introduction

Hepatitis C virus (HCV) infection is a major risk factor for hepatocellular carcinoma (HCC). Chronic HCV infection can result in liver cirrhosis (LC) and HCC over the course of 20–30 years [1]. However, the rate of progression is variable; some patients remain for a long time with persistently normal ALT values, while others progress rapidly to LC and HCC.

Viral factors, host factors, and their interplay appear to play an important role in determining the progression of chronic hepatitis C to LC and HCC. In terms of viral factors, most previous clinical studies have focused on searching for HCV regions correlated with the response to interferon (IFN)-based therapy. In those analyses, correlation between amino acid substitutions and treatment response have been reported for the IFN sensitivity

M. Miura and S. Maekawa have contributed equally to this study.

Electronic supplementary material The online version of this article (doi:10.1007/s12072-011-9307-6) contains supplementary material, which is available to authorized users.

M. Miura S. Maekawa (✉) M. Kadokura R. Sueki
K. Komase H. Shindo T. Ohmori A. Kanayama
K. Shindo F. Amemiya Y. Nakayama T. Kitamura
T. Uetake T. Inoue M. Sakamoto S. Okada N. Enomoto
First Department of Internal Medicine, Faculty of Medicine, University of Yamanashi, 1110, Shimokato, Chuo, Yamanashi 409-3898, Japan
e-mail: maekawa@yamanashi.ac.jp

Published online: 17 August 2011

determining region in nonstructural (NS)5A [2], a.a. 70 and 91 in core [3], the PKR/eIF-2alpha phosphorylation homology domain (PePHD) in envelope (E)2 [4], and the IFN–ribavirin resistance determining region (IRRDR) in NS5A [5].

Regarding the viral factors related to disease progression, among the various HCV proteins, the core protein has been thought widely to contribute because it has been shown experimentally to affect multiple cellular functions, in addition to the evidence from clinical studies [6–10]. Core protein modifies cellular apoptosis, oncogenic signaling, reactive oxygen species formation, lipid metabolism, transcriptional activation, transformation, and immune reactivity. Core protein has oncogenic potential in transgenic mice [11]. In contrast, fewer clinical studies to date have systematically investigated the correlation between the variability of HCV regions and disease progression. However, some of those limited clinical studies reported a correlation between amino acid substitutions in core or NS5A with disease progression [12–15]. Despite those reports, few studies to date support the correlation. Moreover, it is unclear whether those viral sequences change during disease progression or how the disease activity is modified by those viral sequences in the long course of chronic hepatitis.

On the other hand, regarding host factors, recent reports disclosed a significant correlation between polymorphisms in the IL28B gene and responses to pegylated-IFN plus ribavirin therapy for HCV patients [16–19]. This single-nucleotide polymorphism (SNP) also showed significant correlation with natural HCV clearance [20]. However, it remains unknown whether the IL28B SNP is related to disease progression or the development of HCC

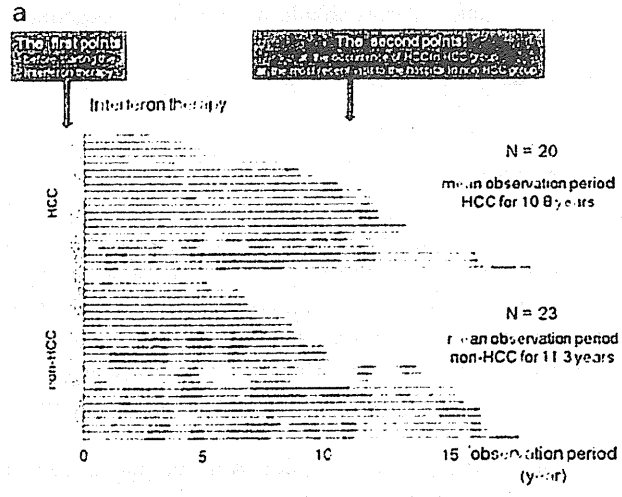
In this study, we first undertook the analysis to identify the viral regions related to disease progression and HCC development through the analysis of complete HCV open-reading frame (ORF) sequences. Because some regions in HCV core and NS5A showed characteristic changes over time in patients developing HCC during the observation period, we proceeded further to analyze the contribution of those regions to disease progression, in association with time and with the IL28B SNP.

Patients and methods

Patients

This study is based on the analysis of two groups of patients, 43 in Group 1 and 230 patients in Group 2.

In the first part, we tried to characterize and extract viral sequences specific to disease progression through the analysis of complete HCV ORFs (Fig. 1a). In particular, we focused our investigation on the changes in viral sequences over time in association with disease progression by comparing HCV sequences of two sufficiently distant time points. With this aim, we determined to investigate patients with a history of IFN therapy, because those patients often were followed long-term with preservation of old and recent sera. However, we excluded sustained



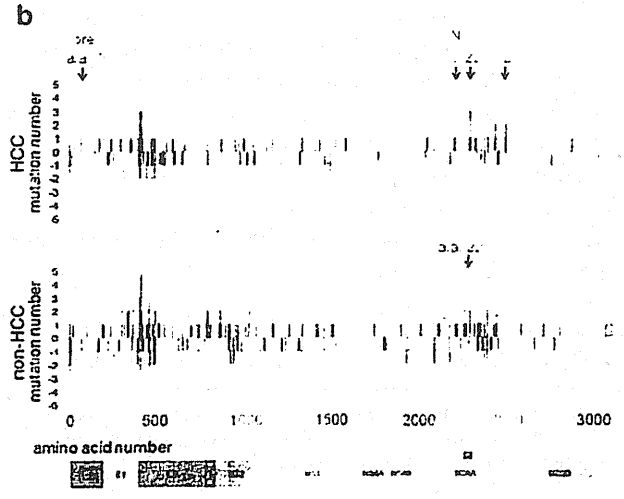


Fig. 1 a A total of 43 patients were analyzed for complete HCV ORF sequences. They were all non-responders to the previous IFN therapy. Twenty patients developed HCC during the observation period, while the 23 patients did not. HCV ORF sequences were determined for the paired samples and the predicted amino acid changes were compared in each patient. b Specific HCV amino acid changes associated with disease progression was evaluated by the analysis of the full-length viral ORF during the observation period for each patient according to the following rules: *1* +1 point for consensus to non-consensus, *2* −1 point for non-consensus to consensus, *3* 0 point for non-consensus to non-consensus. These points were added together and are shown for HCC and non-HCC patients. Patient with later HCC development (*upper panel*). Non-HCC patients (*lower panel*)

virologic response (SVR) patients because we thought that the viral clearance leads to improvement of the liver disease and, therefore, viral regions influencing the IFN response would be extracted as affecting the course of disease. Between March 1992 and April 2004, 273 consecutive patients with HCV-1b infection were given IFN monotherapy at Yamanashi University Hospital, and 133 were followed long-term. A total of 65 patients showed SVR, while 68 showed non-SVR. Among these 68 non-SVRs, 43 patients were included in the study because laboratory data and sera were available from the two distant time points (Group 1). Twenty patients developed HCC during the observation period, while the remaining 23 did not. Regarding the sera, the first time point for both groups was before starting IFN therapy, while the second points were at the occurrence of HCC for the HCC group and at the most recent visit to the hospital for the non-HCC group (Fig. 1a).

An additional 230 HCV-1b patients were recruited (Group 2) as the second study group. They were mainly outpatients at Yamanashi University Hospital and were selected randomly from those with stored sera at the time of disease diagnosis. Sixty-five had HCC, while the remaining 165 did not. They all were positive for HCV RNA at the time of study, although 74 patients had a history of IFN therapy. Parts of the core and NS5A sequences were determined at HCC onset in 65 HCC patients and at the most recent visit to the hospital in 165 non-HCC patients. Because historical sera around 10 years before were also available for 55 of these 230 patients, HCV sequence analysis was also performed for core and NS5A at those previous time points in those patients.

From these two study groups, 228 patients (68 HCCs and 160 non-HCCs) with available genomic DNAs were examined to determine the IL28B SNP

All the patients studied all fulfilled following criteria: (1) Negative for hepatitis B surface antigen. (2) No other forms of hepatitis, such as primary biliary cirrhosis, autoimmune liver disease, or alcoholic liver disease. (3) Free of coinfection with human immunodeficiency virus. (4) A signed consent was obtained for the study protocol that had been approved by Human Ethics Review Committee of Yamanashi University Hospital.

Complete and partial HCV ORF sequence determination by direct sequencing from sera

HCV RNA extraction, complementary DNA synthesis and amplification by two-step nested PCR from serum samples were done using the specific primers for full HCV ORF or partial viral regions as described previously [15]. PCR amplicons were sequenced directly by Big Dye Terminator Version 3.1 (ABI, Tokyo, Japan) with universal M13 forward and reverse primers using an ABI prism 3130 sequencer (ABI). Generated sequence files were assembled using Vector NTI software (Invitrogen, Tokyo, Japan) and base-calling errors were corrected following inspection of the chromatogram.

IL28B SNP analysis

Human genomic DNA was extracted from peripheral blood using a blood DNA extraction kit (QIAGEN, Tokyo, Japan) according to the manufacturer's protocol. The allele typing of each DNA sample was performed by real-time PCR with a model 7500 (ABI) using FAM-labeled SNP primer for the locus rs8099917 (ABI).

Statistical analysis

Statistical differences in the parameters, including all available patients' demographic, biochemic, hematologic, and virologic data, were determined between different groups of patients by Student's t test for numerical variables and Fisher's exact probability test for categorical variables. Odds ratios and their 95% confidence intervals were used to quantify the level of association. All p values of <0.05 by the two-tailed test were considered significant throughout. Multiple logistic regression analyses were used to identify the independent variables influencing core a.a. 70 residue and HCC development. Because most variables used for the analyses were generally considered to correlate with the disease progression, we entered all the variables into the multiple logistic regression analysis even if some of them did not reach significant differences in individual univariate analysis.

Results

Comparing complete HCV amino acid sequences between patients with and without HCC

The clinical characteristics of the 43 patients (Group 1) analyzed for HCV ORF changes over time are shown in Table 1. At the start of observation, clinical characteristics did not differ significantly between the HCC group and the non-HCC group. The mean observation period was comparable between the two groups and was 10.8 years for the HCC group and 11.3 years for non-HCC group ($p = 0.745$). On the other hand, platelets ($p < 0.001$), albumin ($p < 0.001$), and AFP ($p = 0.001$) became significantly lower or higher in the HCC group at the end of observation (Table 1).

We proceeded to investigate viral amino acid changes during the course of disease in each patient to determine

Table 1 Patient characteristics in Group 1

	At the start of observation			At the end of observation		
	HCC ($N = 20$)	Non-HCC ($N = 23$)	p value	HCC ($N = 20$)	Non-HCC ($N = 23$)	p value
Observation period (years)				10.8 ± 3.6	11.3 ± 3.8	0.745
Sex (male/female)	11/9	12/11	0.999	11/9	12/11	0.999
Age (years)	51.5 ± 8.0	50.0 ± 9.9	0.604	61.7 ± 10.0*	61.0 ± 10.9*	0.818
Stage of fibrosis (F1/2/3/4)	1/7/6/6	5/11/4/3	0.190	N/A	N/A	–
AST (IU/L)	102 ± 114	74 ± 40	0.695	71 ± 36*	51 ± 30	0.048
ALT (IU/L)	124 ± 86	104 ± 71	0.411	69 ± 47*	52 ± 31*	0.159
Platelets (10^4/mm^3)	16.2 ± 4.8	18.3 ± 6.2	0.217	9.7 ± 3.9*	15.3 ± 5.1	<0.001
Albumin (g/dL)	4.1 ± 0.4	4.1 ± 0.2	0.639	3.6 ± 0.4	4.1 ± 0.5	<0.001
γGTP (IU/L)	90 ± 60	71 ± 46	0.275	69 ± 59	45 ± 38*	0.114
T.Chol (mg/dL)	169 ± 28	156 ± 22	0.110	146 ± 21	164 ± 5.108	0.086
Alpha-fetoprotein (ng/mL)	10.5 ± 6.8	9.3 ± 10.8	0.695	42.4 ± 41.1	4.7 ± 2.7	0.001
HCV RNA concentration (kIU/mL)	706 ± 696	614 ± 1,181	0.760	3,325 ± 415*	4,508 ± 5,108*	0.426

* Factors with significant changes over time (<0.05)

whether specific amino acid changes related to disease progression could be identified. First, the consensus amino acid was determined at each amino acid position in the HCV ORF after determination of all sequences in these 43 patients. Amino acid changes were determined according to the following rules to highlight directional changes according to disease progression: When an amino acid changed from the consensus to the non-consensus during the observation period, we scored +1 point. Conversely, a change from the non-consensus to the consensus scored −1 point. We scored 0 point for a change from one non-consensus amino acid to another. As shown in Fig. 1b, directional amino acid changes were observed throughout the HCV genome to some degree both in patients with and without HCC development during the clinical course of almost 10 years, and frequent substitutions in E2 hypervariable region were common in both groups. On the other hand, in patients with HCC development, as many as four directional changes were observed at core a.a. 70 and at three amino acid positions of NS5A (Fig. 1b, upper panel). In contrast, in patients without HCC, the significant change ($n = 4$) was observed at a.a. 2,218 of NS5A when E2 hypervariable region was excluded (Fig. 1b, lower panel).

Core and NS5A sequences in patients with and without HCC

Because the first analysis suggested that the patients with later HCC development might accumulate specific mutations in core and NS5A at the time of HCC occurrence, additional sequences were analyzed from 230 HCV-1b patients to confirm the result. The clinical backgrounds of the additional 230 patients are shown in Table 2 (Group 2).

Table 2 Patient characteristics in Group 2

	HCC ($N = 65$)	Non-HCC ($N = 165$)	p value
Observation period (years)			
Sex (male/female)	42/23	76/89	0.018
Age (years)	68.2 ± 9.2	62.4 ± 11.7	<0.001
AST (IU/L)	66 ± 35	41 ± 21	<0.001
ALT (IU/L)	67 ± 47	44 ± 47	<0.001
Platelets (10^4/mm^3)	11.3 ± 5.8	15.3 ± 6.2	<0.001
Albumin (g/dL)	3.6 ± 0.5	4.4 ± 2.9	0.025
γGTP (IU/L)	59 ± 53	38 ± 40	0.001
T.Chol (mg/dL)	153 ± 30	165 ± 31	0.004
Alpha-fetoprotein (ng/mL)	302 ± 1,670	10 ± 25	0.025
HCV RNA concentration (kIU/mL)	5,400 ± 13,574	7,990 ± 8,512	0.104

All patients were positive for HCV RNA. Between the HCC (65 patients) and non-HCC (165 patients) groups, HCC patients were older ($p < 0.001$) and more frequently tended to be males ($p = 0.018$). Moreover, AST, ALT, γGTP, and AFP were significantly higher, and platelets, albumin, and cholesterol were significantly lower in the HCC group. Different predicted amino acids in the core and NS5A regions, between the two groups, are demonstrated in Fig. 2a. The ratio of the core a.a. 70Q (glutamine) or H (histidine) to R (arginine) was significantly higher with the existence of HCC as demonstrated in Fig. 2a (left panel). On the other hand, evident correlations were not confirmed between mutations in NS5A and disease progression (Fig. 2a, right panel). The ratio of Q or H

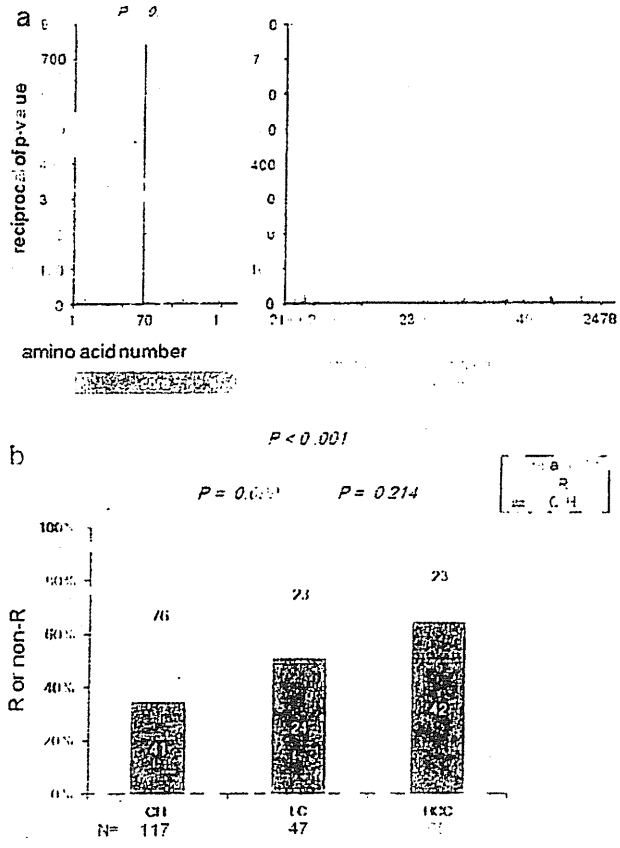


Fig. 2 Core and NS5A sequences in additional patients were studied. a Using sera from 230 additional patients at the time of diagnosis, amino acid usage was compared between HCC and non-HCC patients for the part of core and the NS5A region, and this difference is shown as the *bar height* expressed as reciprocal *p* values. b In 230 patients, the association between polymorphisms of core a.a. 70 and the state of liver disease (chronic hepatitis, LC, or HCC) is shown

to R progressively increased in patients in the three major groups of disease activity: chronic hepatitis, cirrhosis, and HCC (Fig. 2b). The association between the disease progression and core a.a. 70 polymorphism also was observed irrespective of IFN-based therapy (data not shown).

Changes in core a.a. 70 over time in patients with and without HCC

We then examined changes in core a.a. 70 over time in association with disease progression (Fig. 3). For this analysis, 55 patients from Group 2, for whom sera from two distant time points were available, were added to the 43 patients in Group 1 and a total of 98 patients were enrolled. When they were classified into two groups according to later HCC onset, the mean observation period was comparable between the groups, 10.4 years for the HCC group and 12 years for the non-HCC group. The occurrence of core a.a. 70Q was 61% (22/36) at the time of

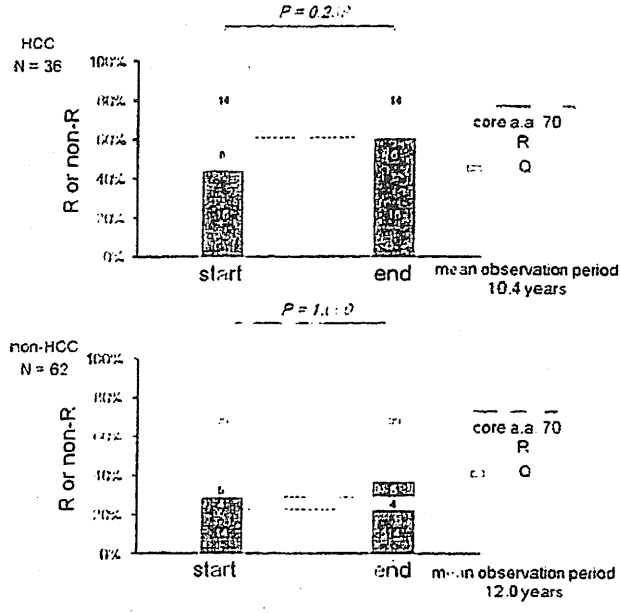


Fig. 3 Changes in core a.a. 70 over time were studied. In 98 patients available for the analysis of 2, sufficiently distant time points, amino acid changes of core a.a. 70 were investigated over time in each patient group. Patients with later HCC development (*upper panel*). Non-HCC patients (*lower panel*)

HCC onset in the HCC group (Fig. 3, upper panel), but 31% (19/62) for the non-HCC group (Fig. 3, lower panel). In contrast, it was 44% (16/36) at the start of observation in the HCC group (Fig. 3, upper panel) and 29% (18/62) for the non-HCC group (Fig. 3, lower panel). Regarding the core a.a. 70 changes over time, R was dominant throughout the observation period in the non-HCC group (71% at the start and 69% at the end), while the dominant amino acid changed from R (56%) to Q (61%) in the HCC group, so that the core a.a. 70 residues of 17% of the HCC patients changed from R to Q during the course of HCC development. In other words, 6 of 22 patients with 70Q at the onset of HCC had 70R originally (6/22, 27%), while 0 of 14 (0%) with 70R at the most recent observation time had 70Q at the beginning. There were no patients with core a.a. 70H throughout the observation period in this population. The result demonstrates that the relationship between 70Q and HCC development is significant at the time of HCC development. At the start of observation, there was also a tendency that the patients with 70Q compared with 70R develop HCC. However, this difference did not reach statistical significance as shown in Supplementary Fig. 1.

Core a.a. 70 changes over time and their association with disease progression

These 98 patients were classified into four groups according to the pattern of core a.a. 70 change (Table 3) and their

Table 3 Progression of liver disease in 98 patients categorized by core a.a. 70 changes over time

	R → R ($N = 53$)			Q → R ($N = 4$)		
	Start	End	p value	Start	End	p value
HCC rate (HCC/non-HCC)	–	26.4% (14/39)	–	–	0% (0/4)	
Sex (male/female)	–	25/28	–	–	4/0	
Observation period (years)	–	11.1 ± 3.4	–	–	12.9 ± 3.5	
Age (years)	51.3 ± 11.7	62.4 ± 12.1	<0.001	48.0 ± 11.6	61.0 ± 9.1	0.128
AST (IU/L)	68 ± 73	48 ± 26	0.066	56 ± 32	83 ± 61	0.456
ALT (IU/L)	80 ± 71	48 ± 35	0.003	114 ± 71	96 ± 42	0.678
Platelets ($10^{-4}/mm^3$)	17.0 ± 5.8	15.0 ± 6.7	0.104	21.3 ± 3.9	17.2 ± 5.2	0.251
Albumin (g/dL)	4.1 ± 0.4	3.9 ± 0.6	0.225	4.4 ± 0.4	4.3 ± 0.4	0.647
γGTP (IU/L)	56 ± 51	38 ± 40	0.052	95 ± 51	61 ± 46	0.371
T.Chol (mg/dL)	172 ± 36	158 ± 33	0.032	152 ± 14	175 ± 32	0.222
Alpha-fetoprotein (ng/mL)	8.3 ± 9.5	12.5 ± 22.1	0.202	6.0 ± 6.0	5.2 ± 2.2	0.816
HCV RNA concentration (kIU/mL)	4,634 ± 8,509	7,070 ± 14,159	0.291	5,798 ± 7,970	13,676 ± 1,881	0.162
	R → Q ($N = 11$)			Q → Q ($N = 30$)		
	Start	End	p value	Start	End	p value
HCC rate (HCC/non-HCC)		54.5% (6/5)			53.3% (16/14)	
Sex (male/female)		6/5			13/17	
Observation period (years)		13.7 ± 1.65			10.8 ± 3.5	
Age (years)	56.4 ± 7.5	69.3 ± 9.3	0.002	54.6 ± 8.5	64.9 ± 9.9	<0.001
AST (IU/L)	62 ± 47	46 ± 12	0.285	79 ± 51	60 ± 31	0.087
ALT (IU/L)	100 ± 69	37 ± 15	0.008	95 ± 58	59 ± 36	0.006
Platelets ($10^{-4}/mm^3$)	17.7 ± 3.9	11.8 ± 4.8	0.005	16.3 ± 6.5	11.9 ± 5.6	0.007
Albumin (g/dL)	4.2 ± 0.2	3.8 ± 0.4	0.005	4.1 ± 0.3	3.8 ± 0.5	0.009
γGTP (IU/L)	73 ± 53	33 ± 16	0.025	101 ± 55	71 ± 65	0.065
T.Chol (mg/dL)	157 ± 21	144 ± 27	0.245	163 ± 28	150 ± 32	0.100
Alpha-fetoprotein (ng/mL)	7.1 ± 4.3	97.8 ± 63.6	0.267	20.8 ± 50.0	35.1 ± 54.7	0.295
HCV RNA concentration (kIU/mL)	2,415 ± 3,163	2,349 ± 1,851	0.957	2,869 ± 3,984	3,229 ± 4,026	0.731

clinical characteristics were investigated. Significant decreases of platelets ($p = 0.007$) and albumin ($p = 0.009$) were observed in the Q unchanged group during the observation period, but not in the R unchanged group ($p = 0.104$ and 0.225, respectively). Because platelets and albumin are markers of liver disease progression, it was considered that the Q unchanged group progressed rapidly with frequent HCC occurrence (53%, 16/30) while the R unchanged group showed stable disease with less frequent HCC occurrence (26%, 14/53). In contrast, the R to Q group showed progressive disease ($p = 0.005$ and 0.005, respectively) similar to the Q unchanged group, while the Q to R group showed stable disease similar to the R unchanged group ($p = 0.251$ and 0.647, respectively), demonstrating that amino acid changes of core a.a. 70 were significantly associated with disease progression.

The IL28B SNP and its association with core a.a. 70 and disease progression

Next, the association between the state of liver disease and IL28B SNP was analyzed for a total of 228 patients through the analysis of the rs8099917 locus. Among them, 162 patients (71%) had the major homozygous TT alleles, while 66 patients (29%) had the minor homozygous or heterozygous alleles (GG/TG). Although some patients had a history of IFN therapy, all patients were positive for HCV RNA at the time of study. The clinical characteristics related to disease progression were compared, as shown in Table 4. Each group consisted of patients with similar distributions of age and sex. Though most clinical factors showed no evident differences in these groups, γGTP was high ($p = 0.020$) and HCV RNA concentration was apt to be low ($p = 0.085$) in TG/GG group. Moreover the ratio

Table 4 Patient characteristics classified by IL28B SNPs at the time of diagnosis

	TT ($N = 162$)	TG/GG ($N = 66$)	p value
Sex (male/female)	81/81	32/34	0.951
Age (years)	63.3 ± 10.7	63.8 ± 11.9	0.735
Platelets (10^{-4}/mm^3)	13.8 ± 5.8	14.3 ± 6.6	0.632
Albumin (g/dL)	4.2 ± 3.0	4.0 ± 0.5	0.528
γGTP (IU/L)	41 ± 39	55 ± 49	0.020
T.Chol (mg/dl)	162 ± 31	157 ± 31	0.237
HCV RNA concentration (kIU/ml)	7,576 ± 10,292	5,069 ± 6,701	0.085
Alpha-fetoprotein (ng/ml)	39.0 ± 152.2	27.3 ± 48.3	0.555
AST (IU/L)	49.6 ± 29.6	51.9 ± 30.5	0.607
ALT (IU/L)	54.5 ± 53.9	51.6 ± 37.7	0.689
Core a.a. 70R/(Q/H)	106/56	17/49	<0.001
Core a.a. 91L/(M/C)	108/54	38/28	0.253
ISDR	1.2 ± 2.0	0.9 ± 1.5	0.164
IRRDR	5.1 ± 2.4	4.7 ± 2.2	0.207
HCC −/+	115/47	45/21	0.788
IFN −/+	95/67	34/32	0.402

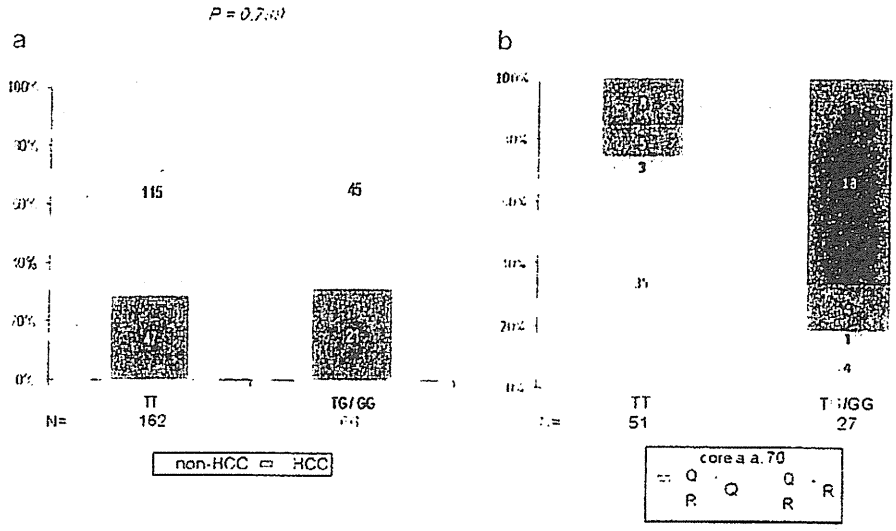


Fig. 4 a Association between the state of liver disease and IL28B SNP. b Time-dependent core a.a. 70 changes and its relation to IL28B SNP was investigated in 78 patients

of R/(Q/H) at core a.a. 70 was significantly higher in those with the TT alleles than in those with TG/GG ($p < 0.001$). In association of IL28B SNP with HCC development, there was no evident relationship as demonstrated in Fig. 4.

IL28B SNP and time-dependent core a.a. 70 changes

In Fig. 4b, it is demonstrated that the direction of time-dependent core a.a. 70 change was influenced by IL28B SNPs. In IL28B TG/GG patients, 4 (50%) out of 8 patients with the initial core a.a. 70R changed into 70Q, while only 1 (5%) out of 19 patients with the initial core a.a. 70Q changed into 70R, demonstrating that core a.a. 70 tended to change into Q over time in IL28B TG/GG patients ($p = 0.034$). On the other hand, there was no evident changing direction in IL28B TT patients; 5 (13%) out of 40 patients with the initial core a.a. 70R changed into 70Q, while 3 (27%) out of 11 patients with the initial core a.a. 70Q changed into 70R ($p = 0.45$).

Multivariate analysis for independent factors influencing core a.a. 70

To investigate further the relationship between core a.a. 70, the IL28B SNP, and HCC development, we divided the patients according to the specification of core a.a. 70 and

Table 5 Factors related to polymorphism of core a.a. 70

Variables	Univariate analysis ($N = 228$)		Multivariate analysis ($N = 228$)	
	Odds ratio (95% CI)	p value	Odds ratio (95% CI)	p value
Sex				
Female	1	0.415	1	0.812
Male	1.23 (0.74–2.01)		1.08 (0.58–1.99)	
Age (years)				
<65	1	0.216	1	0.855
≥65	1.39 (0.82–2.35)		1.06 (0.57–1.96)	
Platelets (10^{-4}/mm^3)				
>12	1	0.004	1	0.844
<12	1.76 (1.03–2.99)		1.07 (0.53–2.16)	
Albumin (g/dL)				
>4	1	0.002	1	0.300
≤4	2.28 (1.33–3.91)		1.46 (0.71–3.00)	
γGTP (IU/L)				
<41	1	0.003	1	0.299
≥41	2.32 (1.32–4.09)		1.42 (0.73–2.79)	
ALT (IU/L)				
<41	1	0.040	1	0.573
≥41	1.74 (1.03–2.94)		1.22 (0.62–2.39)	
IL28B				
TT	1	<0.001	1	<0.001
TG or GG	5.46 (2.88–10.30)		5.74 (2.91–11.31)	
HCC				
−	1	<0.001	1	0.046
+	2.98 (1.65–5.37)		2.21 (1.01–4.83)	
Previous IFN therapy				
−	1	0.874	1	0.644
+	0.96 (0.57–1.62)		0.87 (0.47–1.59)	

those factors, as well as clinical factors, were compared in univariate and multivariate analyses. In Table 5, it may be seen that platelets, albumin, γGTP, ALT, the IL28B SNP, and number of patients with HCC development differed significantly between the two groups in univariate analysis. In contrast, successive multivariate analysis demonstrated that the number of patients with HCC development ($p = 0.046$) and the IL28B SNP ($p < 0.001$) were extracted as independent variables correlated with the core a.a. 70 residue (Table 5).

Multivariate analysis for independent factors influencing HCC development

To disclose factors influencing HCC development, multivariate analysis was performed. As shown in Table 6, age, albumin, and core a.a. 70 residue were extracted as independent factors. On the other hand, IL28B SNP was not extracted as one of those factors.

Discussion

In this study, we have documented several important findings. Through the investigation of HCV sequences, including complete HCV ORFs analysis, we have shown that the core a.a. 70 residue and its changes over time are associated with the disease progression as well as HCC development in genotype-1b HCV infection. Specifically, core a.a. 70Q/H was associated with HCC development and disease progression; core a.a. 70 often changed with time and R70Q substitutions were associated with progressive disease, while Q70R substitutions were associated with the stable disease. Moreover, we have shown that the IL28B SNP and core a.a. 70 showed significant linkage. In contrast, we have also shown that HCC development and disease progression were not apparently correlated with the IL28B SNP.

Recently, core amino acids have been reported in several studies to be associated with HCC [12, 21, 25]. In

Table 6 Factors related to influencing HCC development

Variables	Univariate analysis ($N = 228$)		Multivariate analysis ($N = 228$)	
	Odds ratio (95% CI)	p value	Odds ratio (95% CI)	p value
Sex				
Female	1	0.161	1	0.190
Male	1.50 (0.85–2.67)		1.69 (0.77–3.71)	
Age (years)				
<65	1	0.006	1	0.004
≥65	2.30 (1.28–4.16)		3.26 (1.46–7.25)	
Platelets (10^{-4}/mm^3)				
>12	1	<0.001	1	0.021
<12	5.82 (3.11–10.88)		2.59 (1.16–5.82)	
Albumin (g/dL)				
>4	1	<0.001	1	<0.001
≤4	13.75 (6.69–28.24)		7.73 (3.53–16.94)	
γGTP (IU/L)				
<41	1	<0.001	1	0.122
≥41	3.09 (1.70–5.62)		1.87 (0.85–4.13)	
ALT (IU/L)				
<41	1	<0.001	1	0.109
≥41	3.88 (2.06–7.31)		1.98 (0.86–4.56)	
IL28B				
TT	1	0.626	1	0.290
TG or GG	1.17 (0.13–2.17)		0.63 (0.27–1.49)	
Core a.a. 70				
R	1	<0.001	1	0.029
Q/H	2.91 (1.61–5.26)		2.44 (1.09–5.44)	
Previous IFN therapy				
−	1	0.949	1	0.331
+	0.98 (0.55–1.74)		1.46 (0.68–3.16)	

these studies, patients with core a.a. 70Q/H frequently developed HCC with exacerbation of liver damage. In this analysis, we confirmed the previous findings. However, because this association might be a reflection of the core-dependent IFN sensitivity differences often reported in recent studies [12, 22, 25], we restricted the analysis to patients, who were unable to clear HCV RNA previously through IFN-based therapy. Moreover, we also confirmed the relationship of the core sequences and disease development among the populations without a previous history of IFN therapy (data not shown). These findings strongly confirmed the role of core a.a. 70 in disease progression, independent of any IFN response.

It is a focus of interest how the core sequence evolves with time or with the course of disease. If the core sequences were fixed throughout the course of disease, HCV with core 70Q might be an "oncogenic" virus, while HCV with core 70R might be "non-oncogenic", and the initial viral sequence might forecast future liver disease. In this study, we have demonstrated that core sequences changed in 15% (15/98) of patients during the observation period of around 10 years. Among these changes, R70Q ($N = 11$) was more common than Q70R ($N = 4$). Interestingly, changes in this region were significantly associated with disease activity or HCC development, although patients with R70Q substitutions were significantly more likely to have exacerbation of the disease and Q70R substitutions were associated with the stable disease. These results demonstrate that the core a.a. 70 residue is not fixed, but often changes with time during the course of disease in close association with disease progression and HCC development. Although the molecular mechanism of their interaction needs further exploration, this result highlights the important clinical and basic implications for the association between host and virus.

The importance of the IL28B SNP has been demonstrated recently in HCV infection in terms of a correlation with treatment outcome of pegylated-IFN plus ribavirin

therapy [16–19]. The contribution of the IL28B SNP to the outcome of therapy was confirmed in successive studies, although the mechanism remains under investigation. On this basis, we sought to investigate the impact of the IL28B SNP on disease progression and HCC development, separate from the IFN-based treatment response. As shown in Table 4, we compared the clinical features between the two groups (IL28B GG/TG vs. IL28B TT). Importantly, this comparison disclosed a significant correlation between the core a.a. 70 polymorphisms and the IL28B SNP ($p < 0.001$) and confirmed the existence of a complex interaction between the host and the virus in chronic HCV infection. According to the result, patients with IL28B TG/GG were more likely to be infected with HCV with core a.a. 70Q/H than with core a.a. 70R and vice versa. Although the molecular mechanisms of their relationship remain unknown, it could be speculated that the IL28B SNP has an influence on the viral core sequences, because the host IL28B SNP remains fixed and cannot be influenced by the viral core sequence.

On the other hand, we observed no evident association between the IL28B SNP and HCC development. This was rather unexpected because it is considered that the IL28B SNP has a significant influence on the core a.a. 70 residue. Therefore, to clarify the correlation among core a.a. 70, IL28B, and HCC development, we undertook multivariate analysis to extract the independent variables affecting the core 70 residue. As demonstrated in Table 5, the IL28B SNP and the development of HCC were extracted as variables independently correlated with the core a.a. 70 residue. The result indicates that the core a.a. 70 residue was not only influenced by the IL28B SNP, but also by factors strongly related to HCC development, independent of the IL28B SNP. When considering the result, it is not strange if there is no direct relationship between IL28B SNP and HCC development. In contrast, multivariate analysis undertaken for disclosing factors influencing HCC development revealed that core a.a. 70 residue was a variable independently associated with HCC development other than age, albumin, or platelets even though the IL28B SNP was not extracted (Table 6). However, further comprehensive studies are warranted to disclose the molecular mechanisms for the complicated relationships among core a.a. 70, IL28B, and HCC development.

In conclusion, we have shown that core a.a. 70 was closely associated with disease progression and, often, changes of that residue were accompanied by temporal changes in liver damage, in close relationship with the IL28B SNP.

Acknowledgements 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).

References

1. Kiyosawa K, Sodeyama T, Tanaka E, et al. Interrelationship of blood transfusion, non-A, non-B hepatitis and hepatocellular carcinoma: analysis by detection of antibody to hepatitis C virus. Hepatology 1990;12:671–675
2. Enomoto N, Sakuma I, Asahina Y, et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N Engl J Med 1996;334:77–81
3. Akuta N, Suzuki F, Sezaki H, et al. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. Intervirology 2005;48:372–380
4. Hung CH, Chen CH, Lee CM, et al. Association of amino acid variations in the NS5A and E2-PePHD region of hepatitis C virus 1b with hepatocellular carcinoma. J Viral Hepat 2008;15:58–65
5. El-Shamy A, Nagano-Fujii M, Sasase N, Imoto S, Kim SR, Hotta H. Sequence variation in hepatitis C virus nonstructural protein 5A predicts clinical outcome of pegylated interferon/ribavirin combination therapy. Hepatology 2008;48:38–47
6. Moriya K, Nakagawa K, Santa T, et al. Oxidative stress in the absence of inflammation in a mouse model for hepatitis C virus-associated hepatocarcinogenesis. Cancer Res 2001;61:4365–4370
7. Okuda M, Li K, Beard MR, et al. Mitochondrial injury, oxidative stress, and antioxidant gene expression are induced by hepatitis C virus core protein. Gastroenterology 2002;122:366–375
8. Nishina S, Hino K, Korenaga M, et al. Hepatitis C virus-induced reactive oxygen species raise hepatic iron level in mice by reducing hepcidin transcription. Gastroenterology 2008;134(1):226–238
9. Bartosch B, Thimme R, Blum HE, Zoulim F. Hepatitis C virus-induced hepatocarcinogenesis. J Hepatol 2009;51:810–820
10. Tsai WL, Chung RT. Viral hepatocarcinogenesis. Oncogene 2010;29:2309–2324
11. Moriya K, Fujie H, Shintani Y, et al. The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. Nat Med 1998;4:1065–1067
12. Akuta N, Suzuki F, Kawamura Y, et al. Amino acid substitutions in the hepatitis C virus core region are the important predictor of hepatocarcinogenesis. Hepatology 2007;46:1357–1364
13. Franco S, Gimenez-Barcons M, Puig-Basagoiti F, et al. Characterization and evolution of NS5A quasispecies of hepatitis C virus genotype 1b in patients with different stages of liver disease. J Med Virol 2003;71:195–204
14. Gimenez-Barcons M, Franco S, Suarez Y, et al. High amino acid variability within the NS5A of hepatitis C virus (HCV) is associated with hepatocellular carcinoma in patients with HCV-1b-related cirrhosis. Hepatology 2001;34:158–167
15. Nagayama K, Kurosaki M, Enomoto N, Miyasaka Y, Marumo F, Sato C. Characteristics of hepatitis C viral genome associated with disease progression. Hepatology 2000;31:745–750
16. Fukuhara T, Taketomi A, Motomura T, et al. Variants in IL28B in liver recipients and donors correlate with response to peginterferon and ribavirin therapy for recurrent hepatitis C. Gastroenterology 2010;139:1577–1585
17. Tanaka Y, Nishida N, Sugiyama M, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 2009;41:1105–1109