FIGURE LEGENDS

Fig.1 Number of amino acid substitutions per sample in the sustained viral responders (SVR) and the non-sustained viral responders (non-SVR) group.

The numbers of variations, relative to a population consensus, that were unique to either SVR or non-SVR patients are shown for the full open reading frame (ORF) (Fig.1, left) and for each HCV protein (Fig.1, right).

Fig.2a Different amino acid usages at each viral amino acid position between the sustained viral responders (SVR) and the non-sustained viral responders (non-SVR) patients.

Amino acid variation was determined between SVR and non-SVR patients by Fisher's exact probability test. The longitudinal axis shows the - logP value.

Fig.2b Sequence alignment in the core region.

Dashes indicate amino acids identical to the consensus sequence and substituted amino acids are shown by standard single letter codes.

Fig.2c Sliding window analysis.

Viral regions affecting treatment outcome are shown as red spots. There are four hot spots: at core amino acid 110, amino acids 400-403 (i.e. the hyper variable region) in Envelope2 (E2) region, amino acids 724-743 in E2 and amino acids 2258-2306 in the nonstructural (NS)5A.

Fig.2d Sequence alignment amino acids in the nonstructural (NS)5A around amino acids 2258 to 2306.

Dashes indicate amino acids identical to the consensus sequence and substituted amino acids are shown by standard single letter codes.

Fig.3 Correlation between pretreatment HCV RNA levels and the number of substitutions in the NS5A region aa 2258 to 2306.

Spearman's correlation coefficient by rank test is demonstrated.

Table

Table 1. Baseline Characteristics of All Patients (Group 1 and 2)

Characteristic	SVR (n = 58)			non-SVR (n = 20)			*P* value△
	Group 1 (n = 36)	Group 2 (n =22)	Combined (n = 58)	Group 1 (n = 7)	Group 2 (n = 13)	Combined (n = 20)	
Gender(Male/Female)	20 / 16	9 / 13	29 / 29	4 / 3	5 / 8	9 /11	0.80†
Age(yrs)	50.0 ± 12.5*	57.3 ± 10.0	52.4 ± 12.1	55.0 ± 9.7	59.8 ± 6.4	58.1 ± 7.8	0.058‡
ALT(IU/l)	86.6 ± 86.6	71.2 ± 50.4	80.5 ± 74.2	52.9 ± 29.3	88.1 ± 90.1	75.8 ± 75.5	0.81‡
Platelet($x10^4/mm^3$)	20.8 ± 6.2	19.0 ± 5.2	20.1 ± 5.8	14.7 ± 7.1	19.1 ± 4.9	17.6 ± 6.0	0.11‡
Fibrosis score(0-2 / ≥3)§	34 / 1	19 / 2	53 / 3	4 / 3	11 / 2	15 / 5	0.049†
HCV RNA(KIU/ml)	760(2-3100)**	340(54-3600)	550(12-3600)	1300 (350-30000)	1400 (180-5000)	1300 (180-30000)	0.002‖
IFN dose(≥80% / 60-80%)¶	28 / 4	21 / 1	49 / 5	4 / 3	11 / 2	15 / 5	0.12†
Ribavirin dose(≥80% / 60-80%)¶	27 / 5	17 / 5	44 / 10	4 / 3	5 / 8	9 / 11	0.003†
RVR rate (%)	87.5	54.5	74.1	33.3	46.1	42.1	0.022†
EVR rate (%)	100	100	100	66.7	100	89.4	0.07†

* : mean ± SD ** : median (range) † : Fisher's exact probability test ‡ : Student t test ‖ : Mann-Whitney's U test △: *P* values between all SVR (n = 58) vs. all non-SVR (n = 20)

Several clinical charcteristics listed below were unavailable in some patients

§ : SVR : n = 56 (35 in group1, 21 in group2), non-SVR : n = 17 (7 in group1, 10 in group2) ¶ : SVR : n = 54 (32 in group1, 22 in group2)

Table 2a. Variations in each Amino Acid Position and SVR rate

Position		Group 1 (n = 43)	*P* value	Group 2 (n = 35)	*P* value	Combined (n = 78)	*P* value
Core aa 110	T	100% (19 / 19)	0.01	92.9% (13 / 14)	0.004	97% (32 / 33)	5E - 05
	non T	70.8% (17 / 24)		42.9% (9 / 21)		57.8% (26 / 45)	
p7 aa 773	V	77.4% (24 / 31)	0.16	53.6% (15 / 28)	0.03	66.1% (39 / 59)	0.002
	non V	100% (12 /12)		100% (7 / 7)		100% (19 / 19)	
NS5A aa 2099	R	92.9% (13 / 14)	0.40	91.7% (11 / 12)	0.01	92.3% (24 / 26)	0.01
	non R	79.3% (23 / 29)		47.8% (11 / 23)		65.4% (34 / 52)	
NS5B aa 3013	L	78.9% (26 / 33)	0.17	47.8% (11 / 23)	0.01	66.1% (37 / 56)	0.008
	non L	100% (10 / 10)		91.7% (11 / 12)		95.5% (21 / 22)	

Table 2b. Number of Amino Acid Substitutions in each Region and SVR rate

Region		Group 1 (n = 43)	*P* value	Group 2 (n = 35)	*P* value	Combined (n = 78)	*P* value
E2 aa 400-403	mutation ≥2	89.3% (25 / 28)	0.22	100% (11 / 11)	0.002	92.3% (36 / 39)	0.0005
	mutation 0-1	73.3% (11 / 15)		45.8% (11 / 24)		56.4% (22 / 39)	
E2 aa 724-743	mutation ≥1	100% (28 / 28)	0.0002	72% (18 / 25)	0.12	86.8% (46 / 53)	0.0006
	no mutation	53.3% (8 / 15)		40% (4 / 10)		48% (12 / 25)	
ISDR(aa 2213-2248)	mutation ≥2	100% (15 / 15)	0.08	86.7% (13 / 15)	0.02	93.3% (28 / 30)	0.003
	mutation 0-1	75% (21 / 28)		45% (9 / 20)		62.5% (30 / 48)	
NS5A aa 2258-2306	mutation ≥5	100% (19 / 19)	0.01	84.2% (16 / 19)	0.006	92.1% (35 / 38)	0.0006
	mutation 0-4	70.8% (17 / 24)		37.5% (6 / 16)		57.5% (23 / 40)	

Table 3. Multivariate Logistic Regression Analysis

Factor	Odds (95% CI)	*P* value
Age	1.01 (0.91-1.13)	0.85
HCV RNA	1.00 (1.00-1.00)	0.09
Fibrosis score ≥3/0-2	2.37 (0.21-26.7)	0.48
RVR achievement	3.46 (0.54-22.1)	0.19
Ribavirin dose ≥ 80%	16.0 (1.66-153)	0.02
Core aa 110 T	24.7 (1.72-353)	0.02
NS5A aa 2258-2306 mutations 0-4/≥5	11.5 (1.23-108)	0.03

Table 4a. Baseline Characteristics of Patients with NS5A aa 2258-2306 mutations 0-4 or ≥5 (Group 1 and 2)

Characteristic	Mutation 0-4 (n = 40)	Mutation ≥5 (n = 38)	*P* value
Gender (Male/Female)	22 / 18	16 / 22	NS†
Age (yrs)	54.3 ± 11.4*	53.5 ± 11.5	NS‡
ALT (IU/l)	73.8 ± 70.3	85.3 ± 78.7	NS‡
Platelet ($x10^4/mm^3$)	18.0 ± 5.9	21.0 ± 5.7	0.03‡
Fibrosis score (0-2 / ≥3)§	33 / 5	33 / 2	NS†
HCV RNA (KIU/ml)	1100 (99 - 30000) **	380 (12 - 5000)	0.02‖
IFN dose (≥80% / 60-80%)¶	31 / 8	33 / 2	NS†
Ribavirin dose (≥80% / 60-80%)¶	25 / 14	28 / 7	NS†
RVR rate (%)	65.8	62.9	NS†
EVR rate (%)	94.7	100	NS†
Relapse rate (%)	35.9	7.9	0.002†
SVR rate (%)	57.5	92.1	0.0006†

*: mean ± SD, †: Fisher's exact probability test , ‡: Student t test, §: Mutation 0-4 : n = 38, mutation ≥5 : n = 35,

**: median (range), ‖: Mann-Whitney's U test, ¶: Mutation 0-4 : n = 39, mutation ≥5 : n = 35

Table 4b. Baseline Characteristics of Patients with Core 110 T or N/S (Group 1 and 2)

Characteristic	Core 110 T (n = 33)	Core 110 N/S (n = 45)	*P* value
Gender (Male/Female)	18 / 15	20 / 25	NS†
Age (yrs)	50.4 ± 13.0*	56.4 ± 9.5	0.032‡
ALT (IU/l)	64.5 ± 48.2	88.8 ± 86.2	NS‡
Platelet ($x10^4/mm^3$)	19.3 ± 4.9	19.5 ± 6.6	NS‡
Fibrosis score (0-2 / ≥3)§	30 / 1	36 / 6	NS†
HCV RNA (KIU/ml)	580 (54 - 3600) **	980 (12 - 30000)	NS‖
IFN dose (≥80% / 60-80%)¶	26 / 3	38 / 7	NS†
Ribavirin dose (≥80% / 60-80%)¶	23 / 6	30 / 15	NS†
RVR rate (%)	72.4	59.1	NS†
EVR rate (%)	100	95.5	NS†
Relapse rate (%)	3.0	38.6	9E-05†
SVR rate (%)	97.0	57.8	5E-05†

*: mean ± SD, †: Fisher's exact probability test , ‡: Student t test, §: Core 110 T : n = 31, Core 110 N/S : n = 42,

** : median (range), ‖ : Mann-Whitney's U test, ¶: Core 110 T : n = 29

Figure 1
Click here to download high resolution image

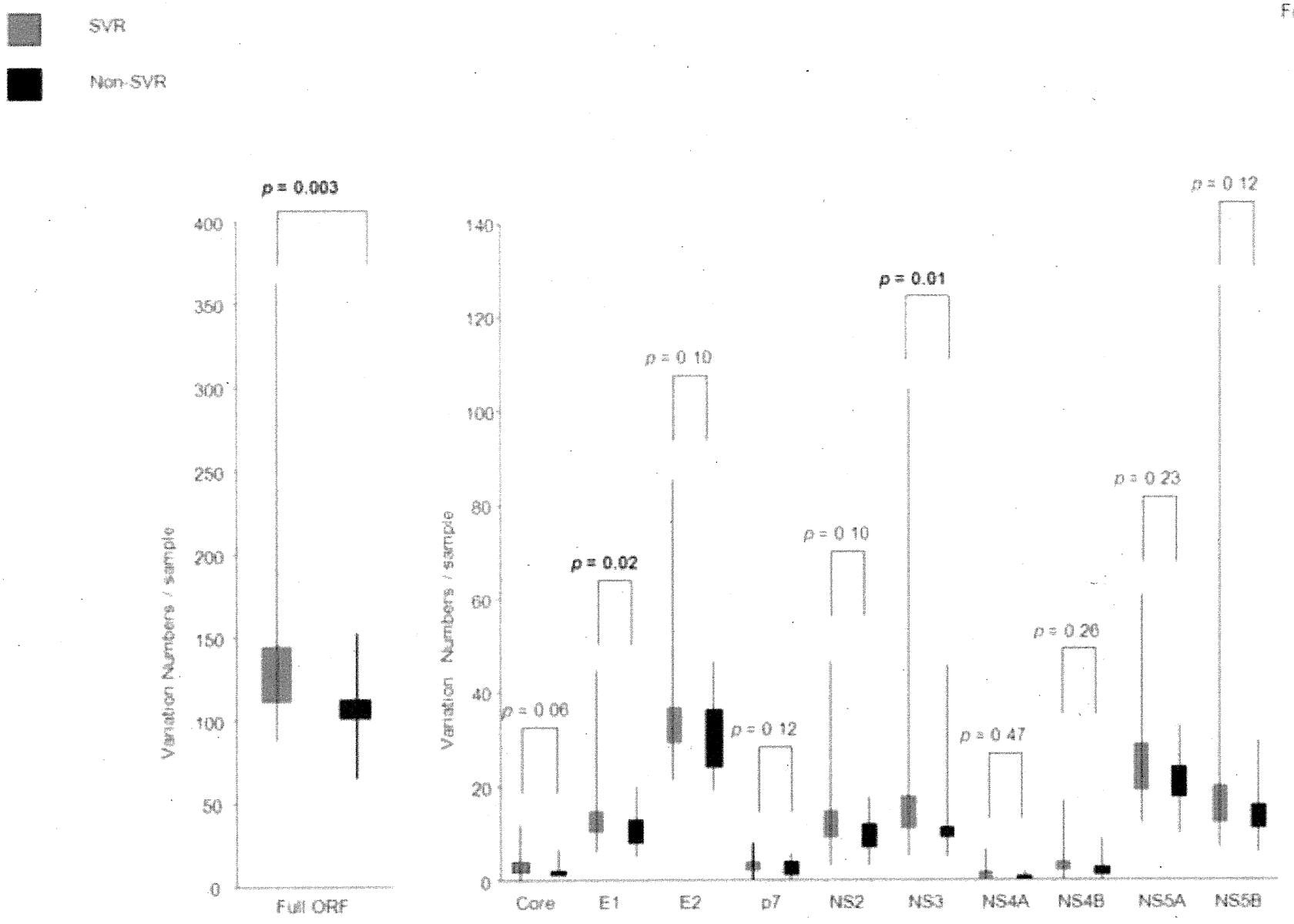


Figure 2a
Click here to download high resolution image

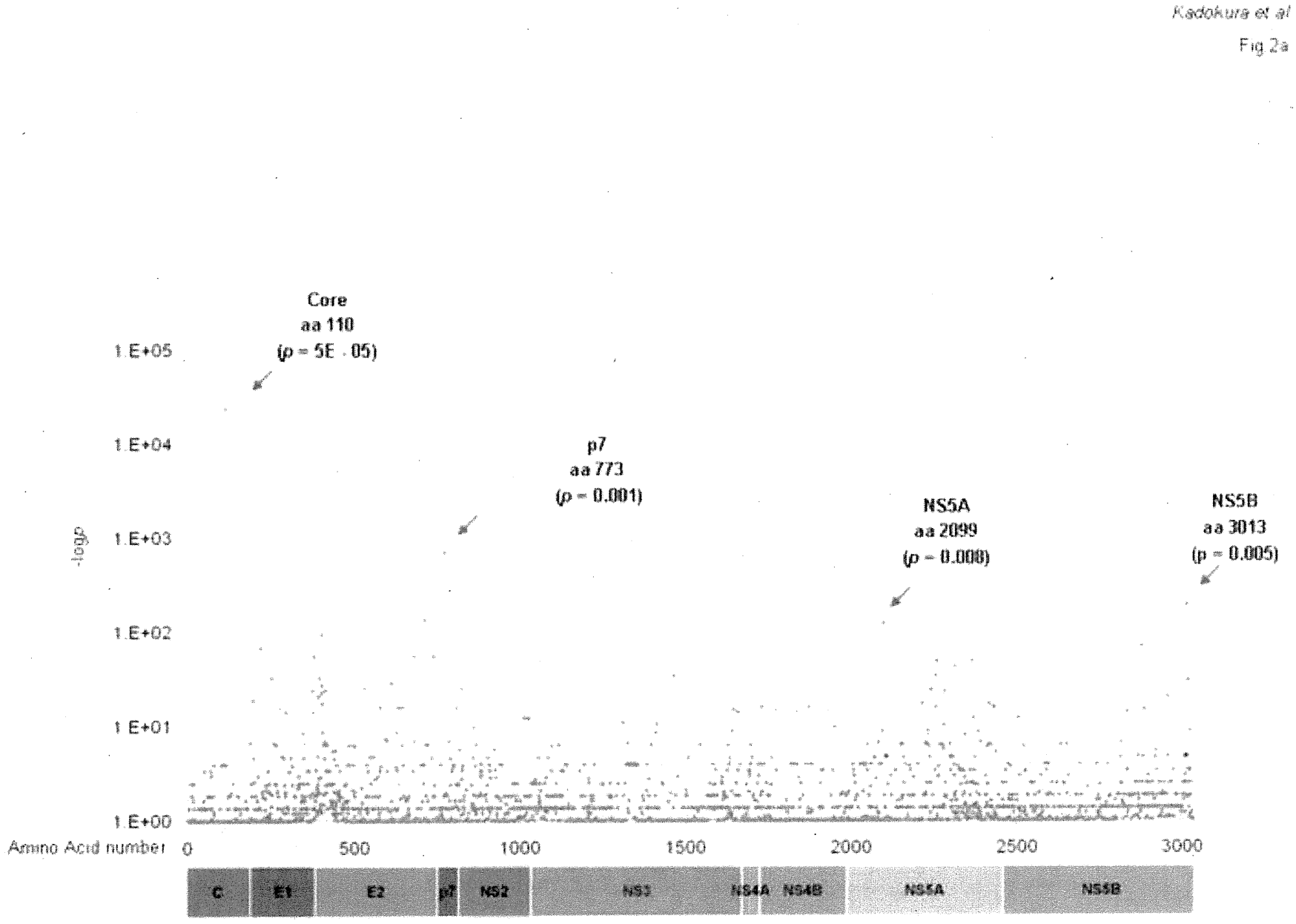


Figure 2c
Click here to download high resolution image

Fig 2c

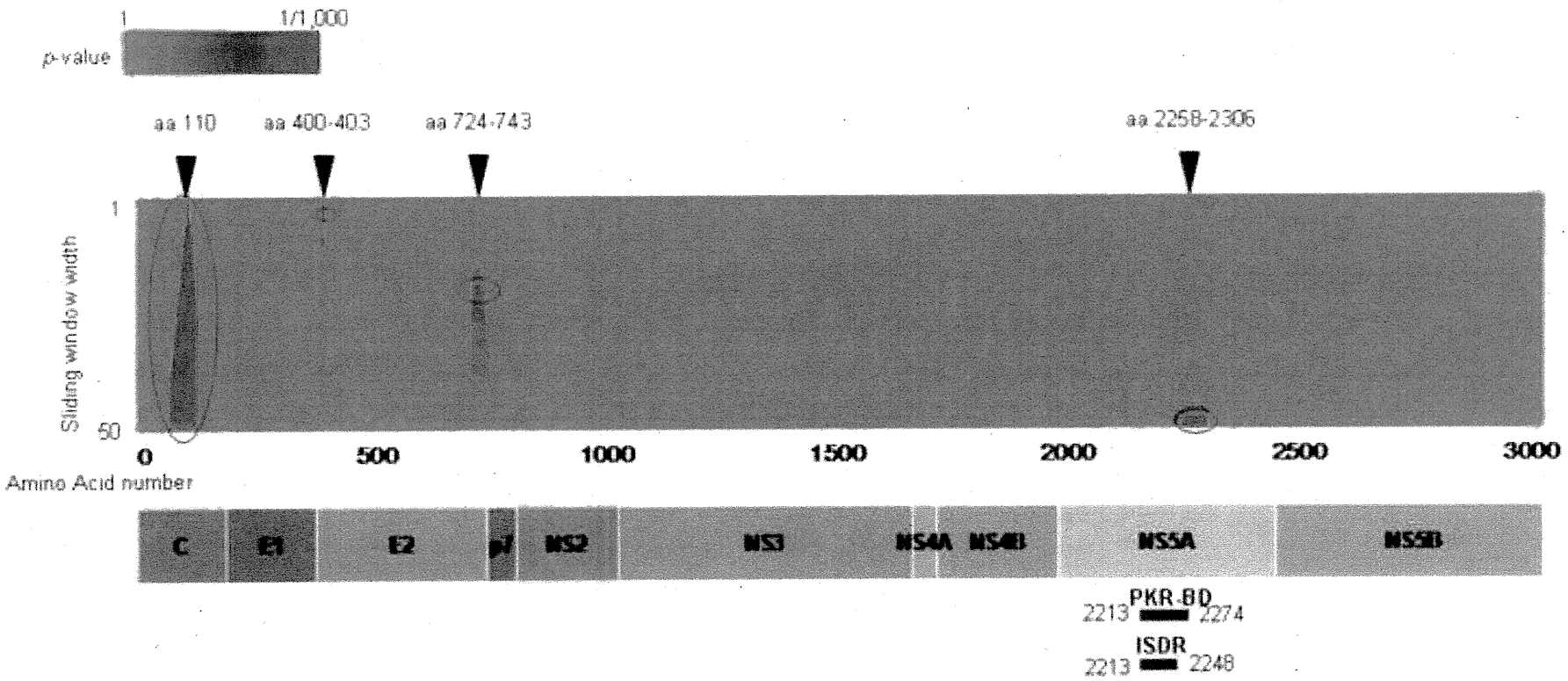


Figure 2d
Click here to download high resolution image

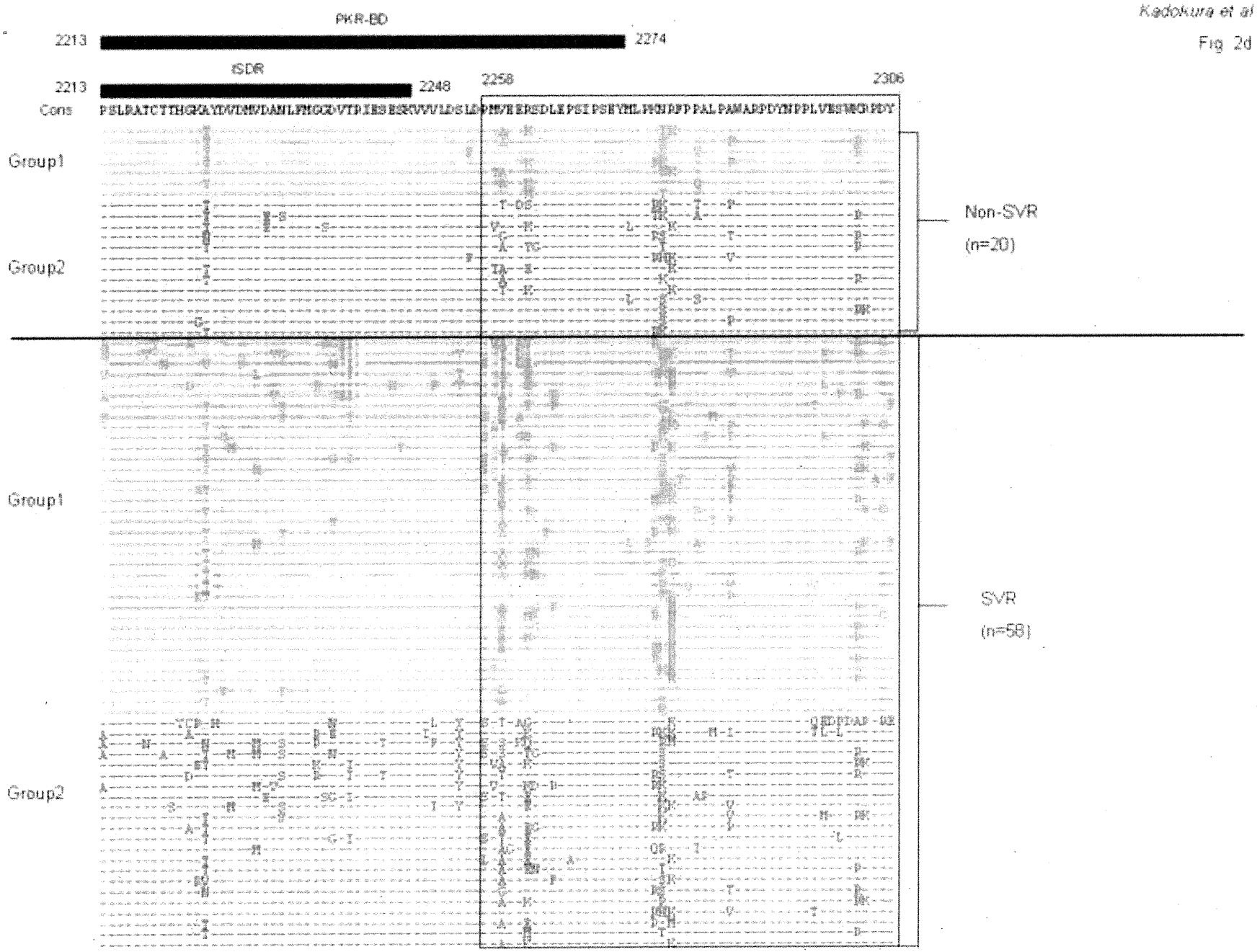


Figure 2b
Click here to download high resolution image

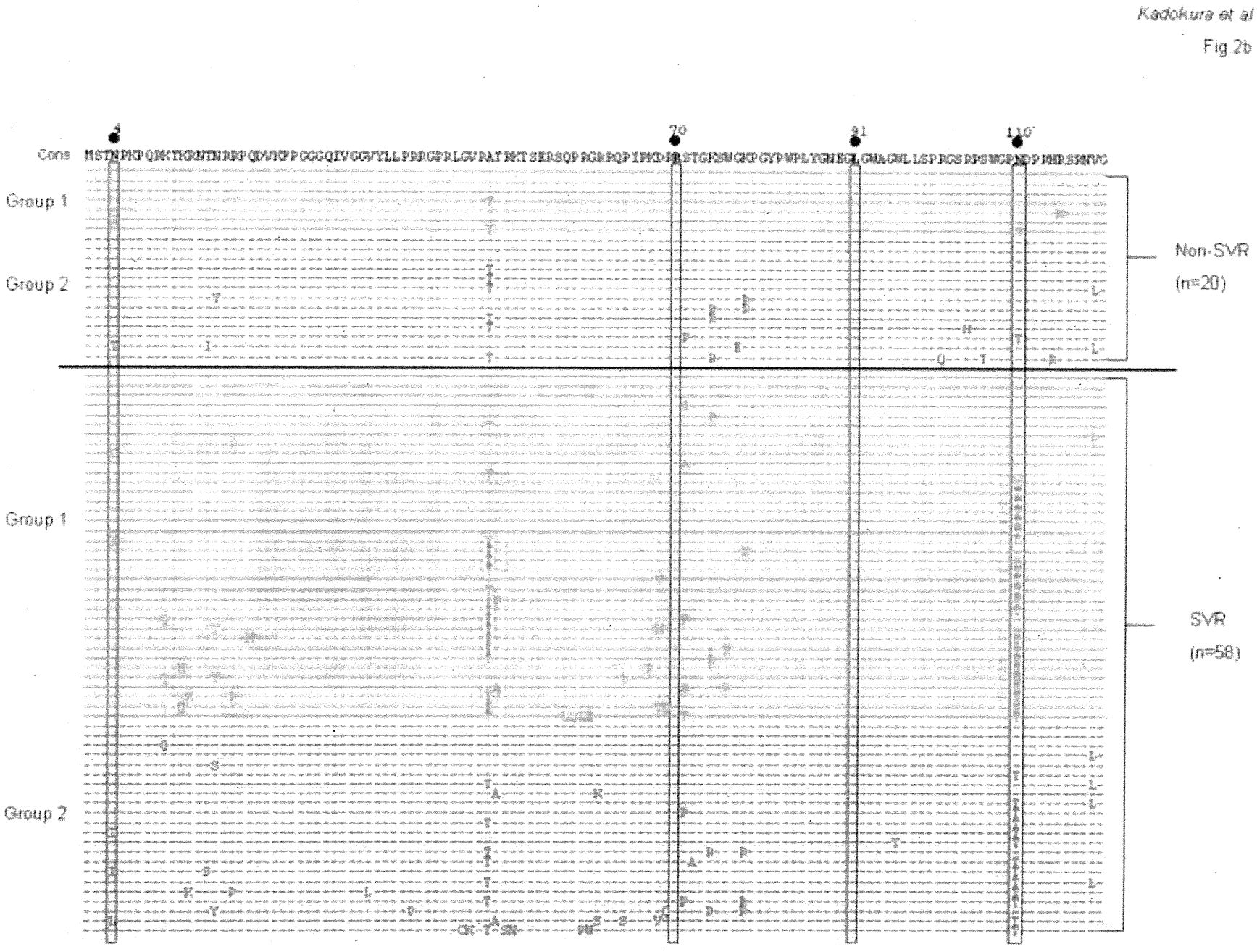
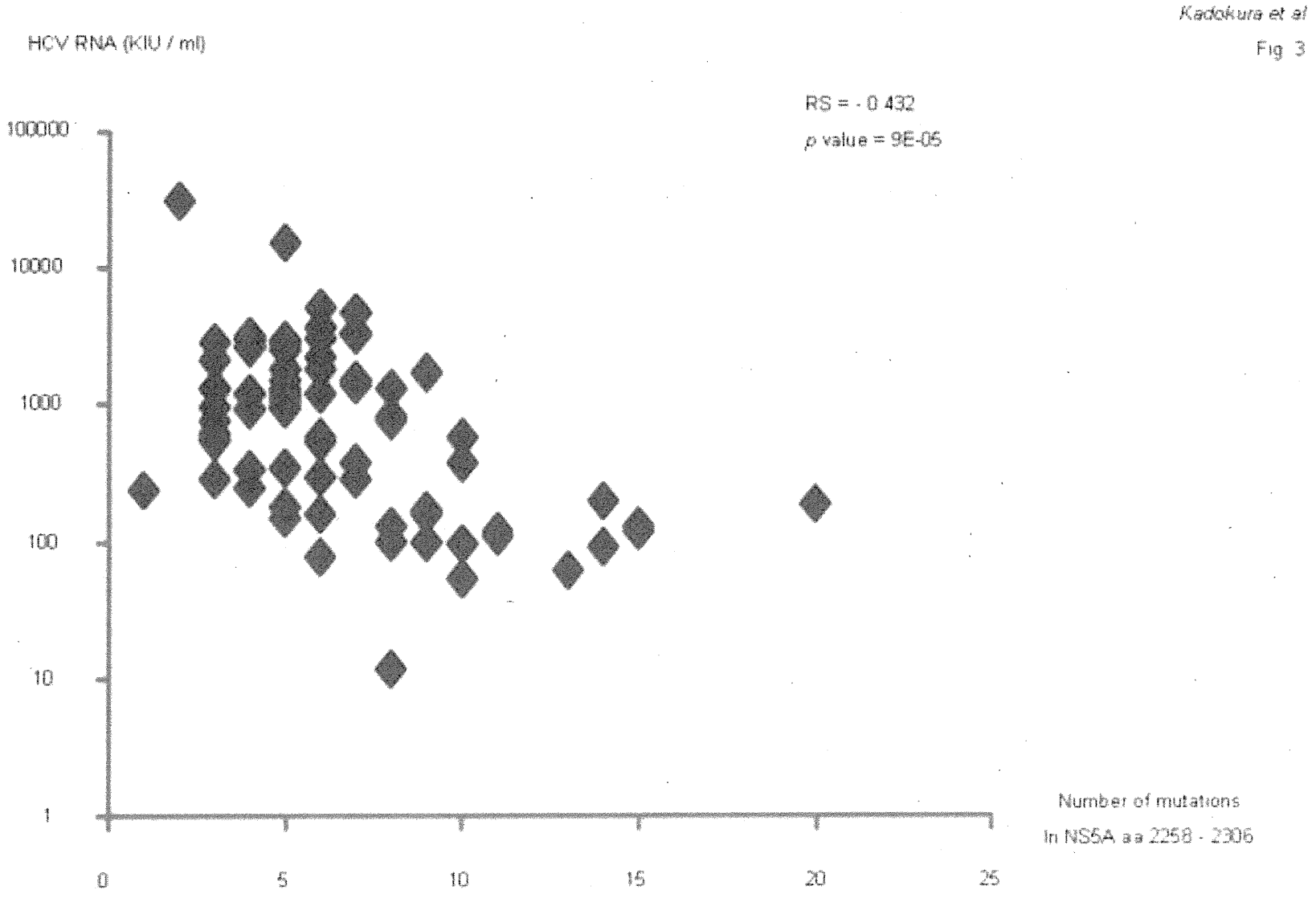


Figure 3
Click here to download high resolution image



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

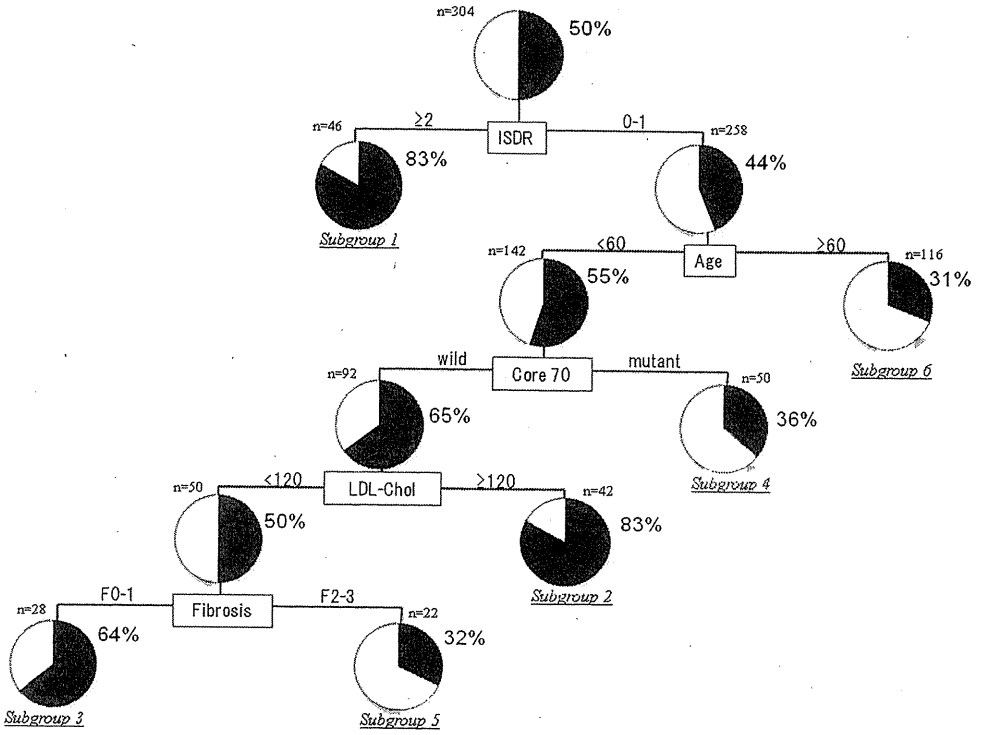


Fig. 1. Decision-tree model. Boxes indicate the factors used for splitting and the cutoff value for the split. Pie charts indicate the rate of sustained virological response for each group of patients after splitting. Terminal subgroups of patients discriminated by the analysis are numbered from 1 to 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)

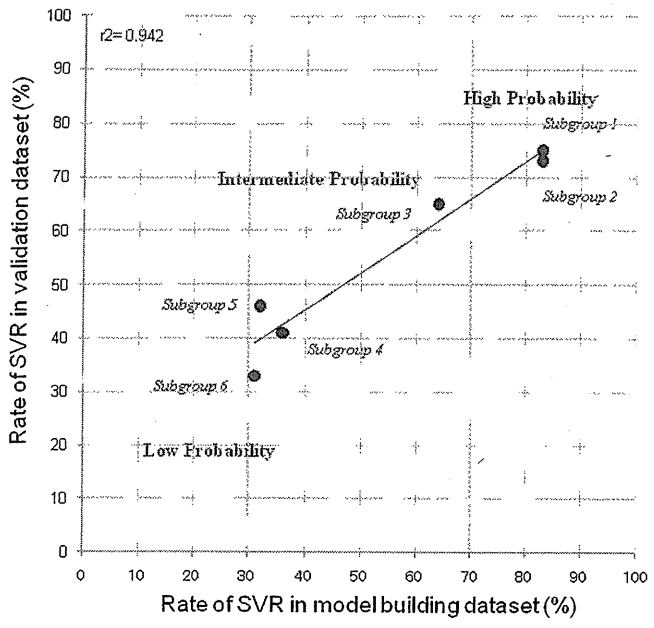


Fig. 2. Validation of the decision-tree analysis: Subgroup-stratified comparison of the rate of sustained virological response. Each patient in the validation set was allocated to subgroups 1–6 by following the flowchart form of the decision tree, and the rates of sustained virological response were 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: $r^2 = 0.94$).

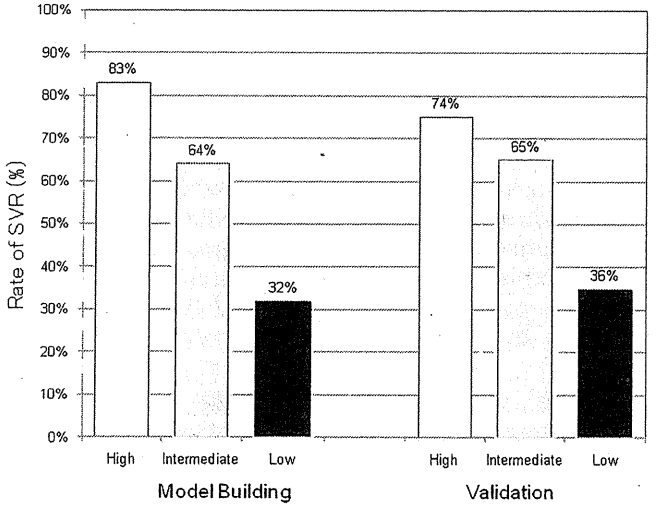


Fig. 3. Comparison of sustained virological response rates between groups divided by the decision tree. The rate of sustained virological response was compared between three groups of patients as divided by the decision-tree analysis. Black, gray, and white boxes indicate the low-probability group (subgroup 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.

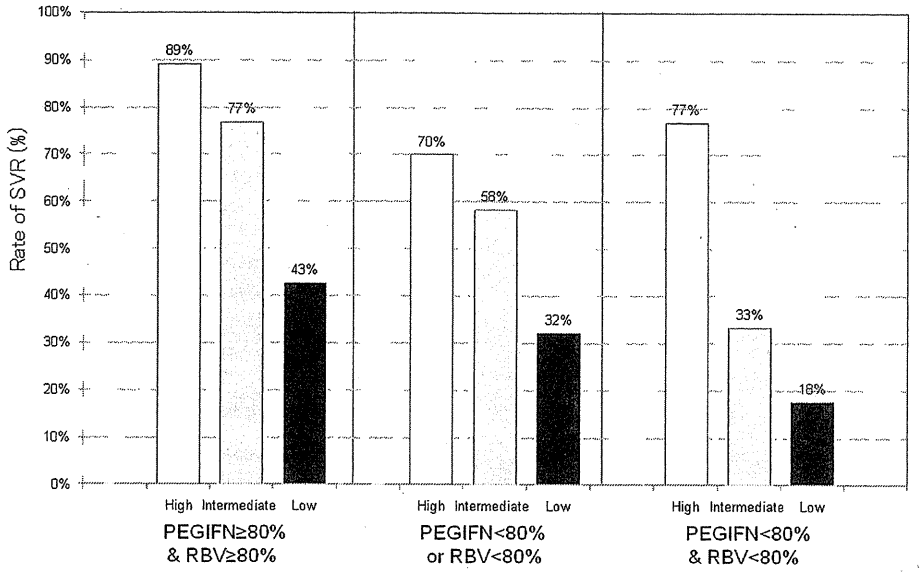


Fig. 4. Comparison of the rate of sustained virological response between the decision-tree groups stratified by drug adherence. The three groups of patients divided by the decision tree (black, gray, and white boxes indicating the low-, intermediate-, and high-probability groups, respectively) were further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