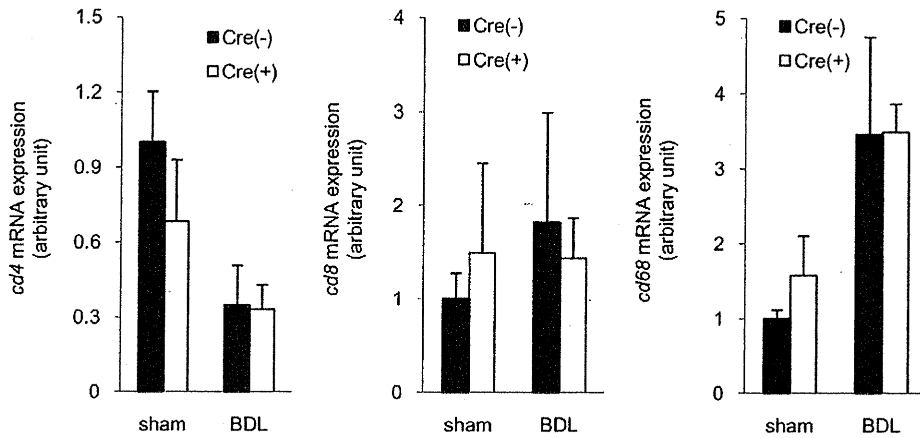
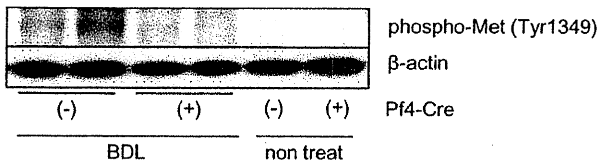
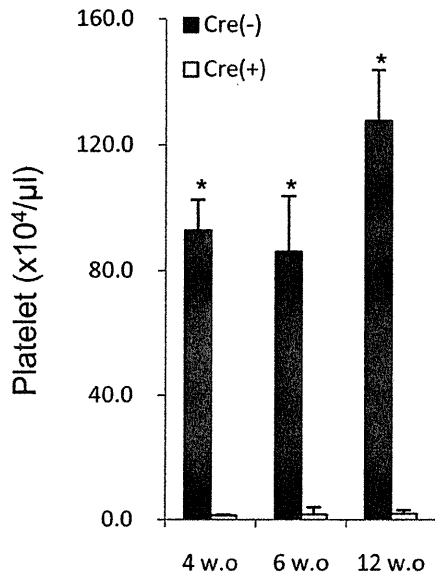


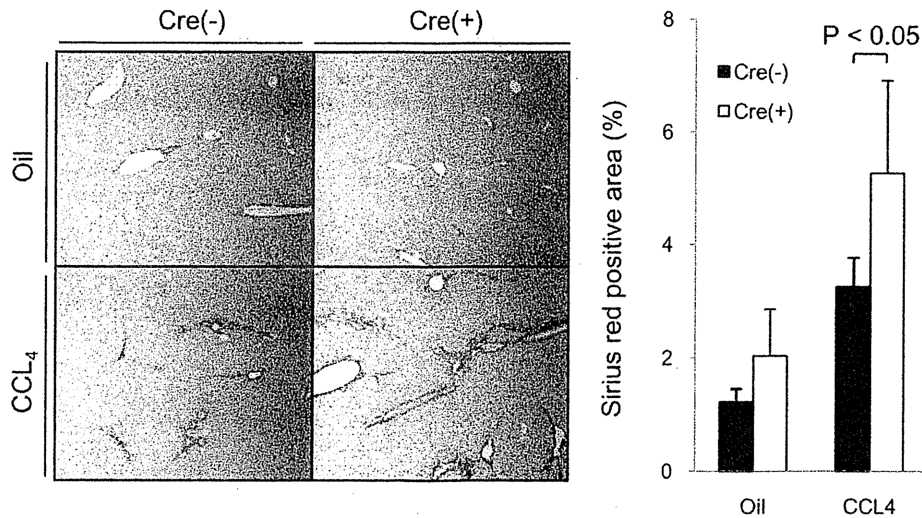
Supplementary Figure 1. Intrahepatic cell fractions upon bile duct ligation (BDL) treatment are not different between the thrombocytopenic mice and the control littermates. *bcl-x$^{flox/flox}$ Pf4-Cre* mice and *bcl-x$^{flox/flox}$* mice were sham operated or subjected to BDL and analyzed 10 days later (4–6 mice per group). Cre(+) and Cre(−) stand for *bcl-x$^{flox/flox}$ Pf4-Cre* and *bcl-x$^{flox/flox}$*, respectively. *cd4*, *cd8*, And *cd68* messenger RNA levels in the liver were determined by real-time reverse-transcription polymerase chain reaction.



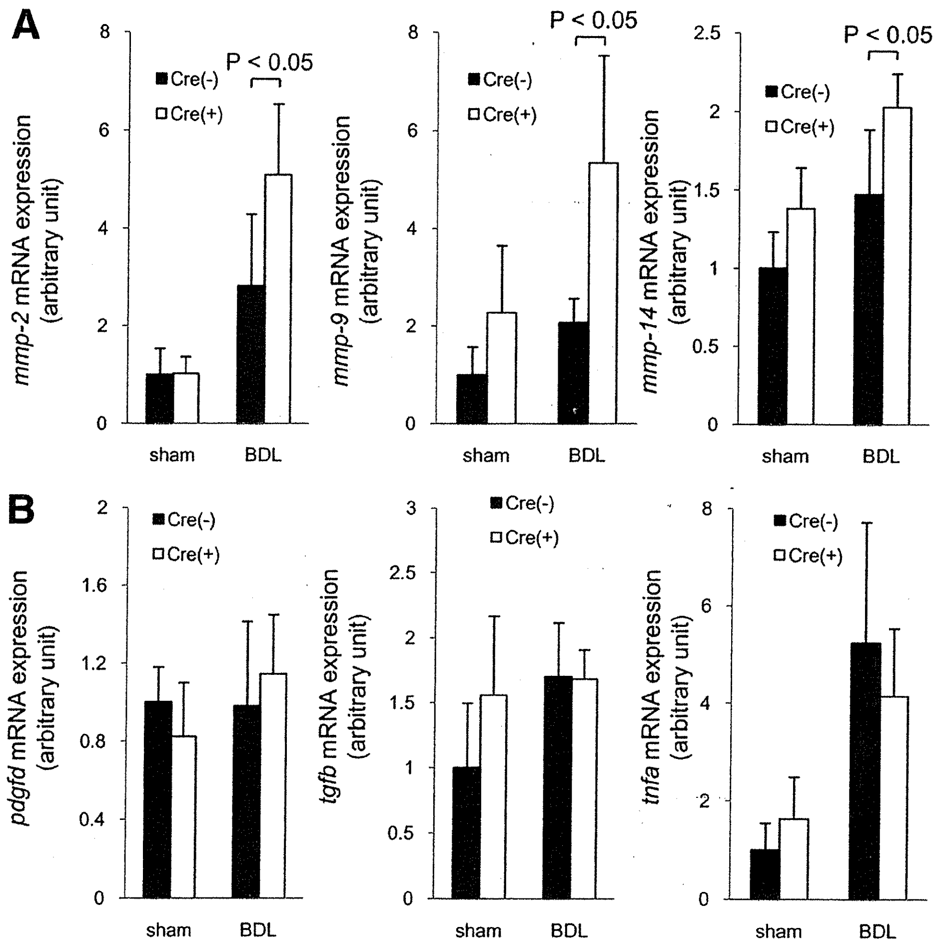
Supplementary Figure 2. Phosphorylation of Met protein in the liver is stronger in the control littermates than in the thrombocytopenic mice at 3 days after bile duct ligation (BDL) treatment. *bcl-x$^{flox/flox}$ Pf4-Cre* mice and *bcl-x$^{flox/flox}$* mice were subjected to BDL and analyzed 3 days later. Cre(+) and Cre(−) stand for *bcl-x$^{flox/flox}$ Pf4-Cre* and *bcl-x$^{flox/flox}$*, respectively. Phosphorylation of Met protein in the liver was determined by Western blotting.



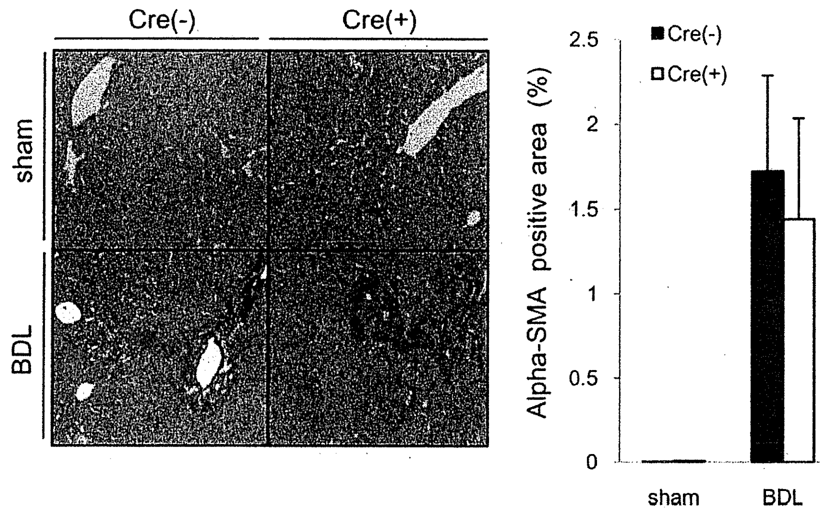
Supplementary Figure 3. *bcl-x^flox/flox^ Pf4-Cre* mice exhibit severe thrombocytopenia at as early as 4 weeks of age, and it persists for a longer time. Circulating platelet counts of *bcl-x^flox/flox^ Pf4-Cre* mice and *bcl-x^flox/flox^* mice at the age of 4, 6, and 12 weeks. Cre(+) and Cre(−) stand for *bcl-x^flox/flox^ Pf4-Cre* and *bcl-x^flox/flox^*, respectively. *$P < .05$ vs Cre(+). 5–8 Mice per group.



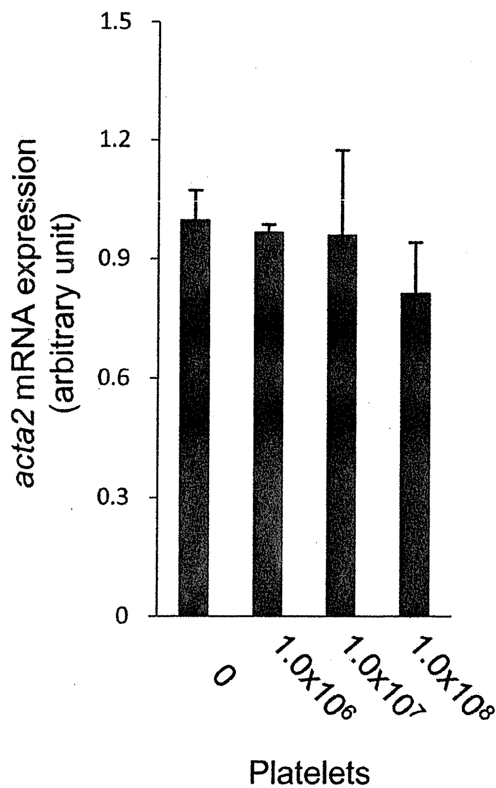
Supplementary Figure 4. Thrombocytopenia exacerbates liver fibrosis induced by chronic CCl_4 administration. *bcl-x^flox/flox^ Pf4-Cre* mice and *bcl-x^flox/flox^* mice were administered intraperitoneal injection of CCl_4 (0.3 mL/kg) or oil 2 times per week and examined 6 weeks later (7 mice per group). Cre(+) and Cre(−) stand for *bcl-x^flox/flox^ Pf4-Cre* and *bcl-x^flox/flox^*, respectively. Liver fibrosis was evaluated by picrosirius red staining of liver sections.



Supplementary Figure 5. (*A*) Gene expression of matrix metalloproteases is up-regulated in thrombocytopenic mice upon bile duct ligation (BDL) treatment. *bcl-x^flox/flox^ Pf4-Cre* mice and *bcl-x^flox/flox^* mice were sham operated or subjected to BDL and analyzed 10 days later (4–6 mice per group). Cre(+) and Cre(−) stand for *bcl-x^flox/flox^ Pf4-Cre* and *bcl-x^flox/flox^*, respectively. *mmp-2*, *mmp-9*, And *mmp-14* messenger RNA levels in the liver were determined by real-time reverse-transcription polymerase chain reaction. (*B*) Gene expression of fibrosis-related cytokines in the liver is not different between the thrombocytopenic mice and their control littermates. *bcl-x^flox/flox^ Pf4-Cre* mice and *bcl-x^flox/flox^* mice were sham operated or subjected to BDL and analyzed 10 days later (4–6 mice per group). Cre(+) and Cre(−) stand for *bcl-x^flox/flox^ Pf4-Cre* and *bcl-x^flox/flox^*, respectively. *pdgfd, tgfb, and tnfa* messenger RNA levels in the liver were determined by real-time reverse-transcription polymerase chain reaction.



Supplementary Figure 6. HSCs are similarly activated in the thrombocytopenic mice and the control mice upon bile duct ligation (BDL). *bcl-x$^{flox/flox}$ Pf4-Cre* mice and *bcl-x$^{flox/flox}$* mice were sham operated or subjected to BDL and analyzed 10 days later (4 or 5 mice per group). Cre(+) and Cre(−) stand for *bcl-x$^{flox/flox}$ Pf4-Cre* and *bcl-x$^{flox/flox}$*, respectively. To assess HSC activation, liver sections were stained with monoclonal anti-α-smooth muscle actin (*α-SMA*) (Dako, Glostrup, Denmark).



Supplementary Figure 7. Coculture with platelets does not affect messenger RNA expression of α-SMA in activated HSCs. HSCs (1.0 × 10^5) were cocultured with indicated dosages of platelets for 6 hours. *acta2* Messenger RNA levels in HSCs were determined by real-time reverse-transcription polymerase chain reaction. N = 3/group.

Supplementary Table 1. Antibodies Used for Western Blotting

Antibody	Manufacturer
Rabbit polyclonal antibody to Bcl-xL	Santa Cruz Biotechnology, Santa Cruz, CA
Rat monoclonal antibody to mouse integrin-α2B/CD41	R&D Systems, Minneapolis, MN
Mouse monoclonal antibody to Met	Cell Signaling Technology, Beverly, MA
Rabbit monoclonal antibody to phospho-Met (Tyr1234)	Cell Signaling Technology
Rabbit monoclonal antibody to phospho-Met (Tyr1349)	Cell Signaling Technology
Mouse monoclonal antibody to β-actin	Sigma-Aldrich, St Louis, MO
Rabbit polyclonal antibody to type I collagen	Rockland, Gilbertsville, PA
Rabbit polyclonal antibody to GAPDH	Trevigen, Gaithersburg, MD
Rabbit monoclonal antibody to stat3	Cell Signaling Technology
Rabbit monoclonal antibody to Erk1/2	Cell Signaling Technology
Rabbit monoclonal antibody to Akt	Cell Signaling Technology
Rabbit monoclonal antibody to phospho-stat3	Cell Signaling Technology
Rabbit monoclonal antibody to phospho-Ekr1/2	Cell Signaling Technology
Rabbit monoclonal antibody to phospho-Akt	Cell Signaling Technology

Supplementary Table 2. Clinicopathologic Features of HCC Patients

Target gene	Assay ID
col1a1	Mm00801666_g1
col1a2	Mm01165187_m1
met	Mm01156980_m1
mmp-2	Mm00439506_m1
mmp-9	Mm00600164_g1
mmp-14	Mm01318969_g1
acta2	Mm01546133_m1
actb	Mm02619580_g1
cd4	Mm01182108_m1
cd8	Mm00442754_m1
cd68	Mm03047343_m1
Tnfa	Mm01178820_m1
tgfb	Mm00546829_m1
pdgfd	Mm01135193_m1

Association of Gene Expression Involving Innate Immunity and Genetic Variation in Interleukin 28B With Antiviral Response

Yasuhiro Asahina,[1] Kaoru Tsuchiya,[1] Masaru Muraoka,[1,2] Keisuke Tanaka,[1,2] Yuichiro Suzuki,[1,2] Nobuharu Tamaki,[1] Yoshihide Hoshioka,[1] Yutaka Yasui,[1] Tomoji Katoh,[1] Takanori Hosokawa,[1] Ken Ueda,[1] Hiroyuki Nakanishi,[1] Jun Itakura,[1] Yuka Takahashi,[1] Masayuki Kurosaki,[1] Nobuyuki Enomoto,[2] Sayuri Nitta,[3] Naoya Sakamoto,[3] and Namiki Izumi[1]

Innate immunity plays an important role in host antiviral response to hepatitis C viral (HCV) infection. Recently, single nucleotide polymorphisms (SNPs) of *IL28B* and host response to peginterferon α (PEG-IFNα) and ribavirin (RBV) were shown to be strongly associated. We aimed to determine the gene expression involving innate immunity in *IL28B* genotypes and elucidate its relation to response to antiviral treatment. We genotyped *IL28B* SNPs (rs8099917 and rs12979860) in 88 chronic hepatitis C patients treated with PEG-IFNα-2b/RBV and quantified expressions of viral sensors (*RIG-I, MDA5*, and *LGP2*), adaptor molecule (*IPS-1*), related ubiquitin E3-ligase (*RNF125*), modulators (*ISG15* and *USP18*), and *IL28* (*IFNλ*). Both *IL28B* SNPs were 100% identical; 54 patients possessed rs8099917 TT/rs12979860 CC (*IL28B* major patients) and 34 possessed rs8099917 TG/rs12979860 CT (*IL28B* minor patients). Hepatic expressions of viral sensors and modulators in *IL28B* minor patients were significantly up-regulated compared with that in *IL28B* major patients (≈3.3-fold, $P < 0.001$). However, expression of *IPS-1* was significantly lower in *IL28B* minor patients (1.2-fold, $P = 0.028$). Expressions of viral sensors and modulators were significantly higher in nonvirological responders (NVR) than that in others despite stratification by *IL28B* genotype (≈2.6-fold, $P < 0.001$). Multivariate and ROC analyses indicated that higher *RIG-I* and *ISG15* expressions and *RIG-I/IPS-1* expression ratio were independent factors for NVR. IPS-1 down-regulation in *IL28B* minor patients was confirmed by western blotting, and the extent of IPS-1 protein cleavage was associated with the variable treatment response. *Conclusion:* Gene expression involving innate immunity is strongly associated with *IL28B* genotype and response to PEG-IFNα/RBV. Both *IL28B* minor allele and higher *RIG-I* and *ISG15* expressions and *RIG-I/IPS-1* ratio are independent factors for NVR. (HEPATOLOGY 2012;55:20-29)

Infection with hepatitis C virus (HCV) is a common cause of chronic hepatitis, which progresses to liver cirrhosis and hepatocellular carcinoma in many patients.[1] Pegylated interferon α (PEG-IFNα) and ribavirin (RBV) combination therapy has been used to treat chronic hepatitis C (CH-C) to alter the natural course of this disease. However, 20% patients are nonvirological responders (NVR) whose HCV-RNA does not become negative during the 48 weeks of PEG-IFNα/RBV combination therapy.[2] In a recent genome-wide association study, single nucleotide polymorphisms (SNPs) located near interleukin 28B

Abbreviations: CH-C, chronic hepatitis C; γ-GTP, γ-glutamyl transpeptidase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; HCV, hepatitis C virus; HMBS, hydroxymethylbilane synthase; IL28, interleukin 28; IPS-1, IFNβ promoter stimulator 1; ISG15, interferon-stimulated gene 15; MDA5, melanoma differentiation associated gene 5; NVR; nonvirological responders; PEG-IFNα, pegylated interferonα; SNP, single nucleotide polymorphism; RIG-I, retinoic acid-inducible gene I; RBV, ribavirin; RNF125, ring-finger protein 125; ROC, receiver operator characteristic; SVR, sustained viral responder; TVR, transient virological responder; USP18, ubiquitin-specific protease 18; VR, virological responder.

From the [1]Department of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan; [2]First Department of Internal Medicine, Faculty of Medicine, University of Yamanashi, Yamanashi, Japan; [3]Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Tokyo, Japan.

Received May 14, 2011; accepted August 16, 2011.

Supported by grants from the Japanese Ministry of Education, Culture, Sports, Science and Technology and the Japanese Ministry of Welfare, Health and Labor. The funding source had no role in the collection, analysis, or interpretation of the data, or in the decision to submit the article for publication.

(*IL28B*) that encodes for type III IFNλ3 were shown to be strongly associated with a virological response to PEG-IFNα/RBV combination therapy.[3-5] In particular, the rs8099917 TG and GG genotypes were shown to be strongly associated with a null virological response to PEG-IFNα/RBV.[3] However, mechanisms involving resistance to PEG-IFNα/RBV have not been completely elucidated.

The innate immune system has an essential role in host antiviral defense against HCV infection.[6] The retinoic acid-inducible gene I (RIG-I), a cytoplasmic RNA helicase, and related melanoma differentiation associated gene 5 (MDA5) play essential roles in initiating the host antiviral response by detecting intracellular viral RNA.[7,8] The IFNβ promoter stimulator 1 (IPS-1)—also called the caspase-recruiting domain adaptor inducing IFNβ, mitochondrial antiviral signaling protein, or virus-induced signaling adaptor—is an adaptor molecule. IPS-1 connects RIG-I sensing to downstream signaling, resulting in *IFNβ* gene activation.[9-12] RIG-I sensing of incoming viral RNA has been shown to be modified by LGP2,[8,13] a helicase related to RIG-I and MDA5 lacking caspase-recruiting domain. The ubiquitin ligase ring-finger protein 125 (RNF125) has been shown to conjugate ubiquitin to RIG-I, MDA5, and IPS-1 and this suppresses the functions of these proteins.[14] Further, these molecules are ISGylated by the IFN-stimulated gene 15 (ISG15), a ubiquitin-like protein,[15] and ISG15 is specifically removed from ISGylated protein by ubiquitin-specific protease 18 (USP18) to regulate the RIG-I/IPS-1 system.[16,17] Moreover, the NS3/4A protease of HCV specifically cleaves IPS-1 as part of its immune-evasion strategy.[9,18] Therefore, the RIG-I/IPS-1 system and its regulatory systems have essential roles in the innate antiviral response.

Recently, we demonstrated that baseline intrahepatic gene expression levels of the RIG-I/IPS-1 system were prognostic biomarkers of the final virological outcome in CH-C patients who were treated with PEG-IFNα/RBV combination therapy.[19] We found that up-regulation of *RIG-I* and *ISG15* and a higher expression ratio of *RIG-I/IPS-1* could predict NVR for subsequent treatment with PEG-IFNα/RBV combination therapy.[19] However, association of gene expression involving innate immunity and genetic variation of *IL28B* has not yet been elucidated. Hence, the aim of this study was to determine gene expression involving the innate immune system in different genetic variations of *IL28B* and elucidate the relation of gene expression to final virological outcome of PEG-IFNα/RBV combination therapy in CH-C patients.

Patients and Methods

Patients. Among histologically proven CH-C patients admitted at the Musashino Red Cross Hospital, 88 patients with HCV genotype 1b and a high viral load (>5 log IU/mL by TaqMan HCV assay; Roche Molecular Diagnostics, Tokyo, Japan) were included in the present study (Table 1). Patients with decompensated liver cirrhosis, autoimmune hepatitis, or alcoholic liver injury were excluded. No patient had tested positive for hepatitis B surface antigen or anti-human immunodeficiency virus antibody or had received immunomodulatory therapy before enrollment. Forty-two patients had been enrolled in a previous study that determined hepatic gene expression involving innate immunity.[19] Written informed consent was obtained from all patients and the study was approved by the Ethical Committee of Musashino Red Cross Hospital in accordance with the Declaration of Helsinki.

Treatment Protocol. The patients were administered subcutaneous injections of PEG-IFNα-2b (PegIntron, MSD, Whitehouse Station, NJ) at a dose of 1.5 μg kg^{-1} week^{-1} for 48 weeks. RBV (Rebetol, MSD) was administered concomitantly over this treatment period, administered orally twice daily at 600 mg/day for patients who weighed less than 60 kg and 800 mg/day for patients who weighed between 60-80 kg. The dose of PEG-IFNα-2b was reduced to 0.75 μg kg^{-1} week^{-1} when either neutrophil count was less than 750/mm^3 or platelet count was less than 80 $\times$ 10^3/mm^3. The dose of RBV was reduced to 600 mg/day when the hemoglobin concentration decreased to 10 g/dL. More than 80% adherence was achieved in all patients.

Measurement of Hepatic Gene Expression. Liver biopsy was performed immediately before initiating

Address reprint requests to: Namiki Izumi, M.D., Ph.D., Chief, Department of Gastroenterology and Hepatology, Musashino Red Cross Hospital, 1-26-1 Kyonancho 1-26-1, Musashinoshi, Tokyo 180-8610, Japan. E-mail: nizumi@musashino.jrc.or.jp; fax: +81-422-32-9551.
Copyright © 2011 by the American Association for the Study of Liver Diseases.
View this article online at wileyonlinelibrary.com.
DOI 10.1002/hep.24623
Potential conflict of interest: Nothing to report.
Additional Supporting Information may be found in the online version of this article.

Table 1. Patient Characteristics and *IL28B* Genotype

	IL28B Major*	*IL28B* Minor†	*P*-value‡
Patients, n	54	34	
Age (SD), year	58.8 (10.0)	59.1 (10.3)	0.918§
Sex, n (%)			0.051‖
Male	13 (24.1)	15 (44.1)	
Female	41 (75.9)	19 (55.9)	
BMI (SD), kg/m^2	22.7 (3.5)	23.5 (3.6)	0.193§
ALT (SD), IU/L	61.3 (50.7)	62.4 (44.7)	0.962§
γ-GTP (SD), IU/L	36.7 (25.9)	57.3 (52.4)	0.010§
LDL-cholesterol (SD), mg/dL	103.3 (29.8)	91.8 (26.9)	0.067§
Hemoglobin (SD), g/dL	14.1 (1.4)	14.4 (1.3)	0.186§
Platelet count (SD), $\times 10^3 3/\mu L$	161 (6.4)	163 (4.4)	0.489§
Fibrosis stage, n (%)			0.532‖
F1, 2	38 (70.4)	26 (76.5)	
F3, 4	16 (29.6)	8 (23.5)	
Viral load (SD), $\times 10^{63}$ IU/mL	1.7 (1.4)	1.9 (2.0)	0.788§
%HCV core 70 & 91 a.a. double mutation¶	8.9	43.5	0.001‖
%ISDR wild**	43.5	51.7	0.486‖
Viral response, n (%)			<0.001‖
SVR	17 (31.5)	13 (38.2)	
TVR	26 (48.1)	3 (8.8)	
NVR	11 (20.4)	18 (52.9)	

Unless otherwise indicated, data are given as mean (SD).
*rs8099917 TT and rs12979860 CC.
†rs8099917 TG and rs12979860 CT.
BMI, body mass index; ALT, alanine aminotransferase; γ-GTP, γ-glutamyl transpeptidase; LDL-C, low-density lipoprotein cholesterol; HCV, hepatitis C virus; ISDR, interferon sensitivity determining region; SVR, sustained virological response; TVR, transient virological response; NVR, nonvirological response.
‡Comparison between *IL28B* major and minor genotypes.
§Mann-Whitney *U* test.
‖Chi-square test.
¶HCV core mutation was determined in 68 patients.
**ISDR was determined in 75 patients.

the therapy. After extraction of total RNA from liver biopsy specimens, the messenger RNA (mRNA) expression of the positive and negative cytoplasmic viral sensor (*RIG-I, MDA5*, and *LGP2*), the adaptor molecule (*IPS-1*), the related ubiquitin E3-ligase (*RNF125*), the modulators of these molecules (*ISG15* and *USP18*), and *IFNλ* (*IL28A/B*) was quantified by real-time quantitative polymerase chain reaction (PCR) using target gene-specific primers. In brief, total RNA was extracted by the acid-guanidinium-phenol-chloroform method using Isogen reagent (Nippon Gene, Toyama, Japan) from the liver biopsy specimen, which was 0.2-0.4 cm in length and 13G in diameter. Complementary DNA (cDNA) was transcribed from 2 μg of total RNA template in a 140-μL reaction mixture using the SYBR RT-PCR Kit (Takara Bio, Otsu, Japan) with random hexamer. Real-time quantitative PCR was performed using Smart Cycler version II (Takara Bio) with the SYBR RT-PCR Kit (Takara Bio) according to the manufacturer's instructions. Assays were performed in duplicate and the expression levels of target genes were normalized to the expressions of glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene and hydroxymethylbilane synthase (*HMBS*), an enzyme that is stable in the liver, as quantified using real-time quantitative PCR as internal controls. For accurate normalization, a set of two housekeeping genes was used in the present study. Sequences of the primer sets were as follows: *RIG-I*, 5′-AAAGCATGCATGGTGTTCCAGA-3′, 5′-TCATTCGTGCATGCTCACTGATAA-3′; *MDA5*, 5′-ACATAACAGCAACATGGGCAGTG-3′, 5′-TTTGGTAAGGCCTGAGCTGGAG-3′; *LGP2*, 5′-ACAGCCTTGCAAACAGTACAACCTC-3′, 5′-GTCCCAAATTTCCGGCTCAAC-3′; *IPS-1*, 5′-GGTGCCATCCAAAGTGCCTACTA-3′, 5′-CAGCACGCCAGGCTTACTCA-3′; *RNF125*, 5′-AGGGCACATATTCGGACTTGTCA-3′, 5′-CGGGTATTAAACGGCAAAGTGG-3′; *ISG15*, 5′-AGCGAACTCATCTTTGCCAGTACA-3′, 5′-CAGCTCTGACACCGACATGGA-3′; *USP18*, 5′-TGGTTCTGCTTCAATGACTCCAATA-3′, 5′-TTTGGGCATTTCCATTAGCACTC-3′; *IFNλ*: 5′-CAGCTGCAGGTGAGGGA-3′, 5′-GGTGGCCTCCAGAACCTT-3′; *GAPDH*, 5′-GCACCGTCAAGGCTGAGAAC-3′, 5′-ATGGTGGTGAAGACGCCAGT-3′; *HMBS*, 5′-AAGCGGAGCCATGTCTGGTAAC-3′, 5′-GTACCCACGCGAATCACTCTCA-3′.

Genotyping for* IL28B *(rs8099917 and rs12979860) Polymorphism. Genetic polymorphism in a tagged SNP located near the *IL28B* gene (rs8099917 and rs12979860) was determined by direct sequencing of PCR-amplified DNA. In brief, after extraction from whole blood samples, genomic DNA was amplified by PCR. Sequences of the primer sets were: rs8099917, 5′-ATCCTCCTCTCATCCCTCATC-3′, 5′-GGTATCAACCCCACCTCAAAT-3′; rs12979860, 5′-GGACGAGAGGGCGTTAGAG-3′, 5′-AGGGACCGCTACGTAAGTCAC-3′.

Both strands of the PCR products were sequenced by the dye terminator method using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Chiba, Japan); nucleotide sequences were determined by a capillary DNA sequencer ABI3730xl (Applied Biosystems). Homozygosity (rs8099917 GG and rs12979860 TT) or heterozygosity (rs8099917 TG and rs12979860 CT) of the minor sequence was defined as having the *IL28B* minor allele, whereas homozygosity for the major sequence (rs8099917 TT and rs12979860 CC) was defined as having the *IL28B* major allele.

Western Blotting. Western blotting was performed using samples from 14 patients (six from *IL28B* major patients and eight from *IL28B* minor patients) as described.[19] In brief, liver biopsy specimens of

approximately 10 mg were homogenized in 100 μL of Complete Lysis-M (Roche Applied Science, Penzberg, Germany). Next, 30 μg of protein was separated by NuPAGE 4%-12% Bis-Tris gels (Invitrogen, Carlsbad, CA) and blotted on polyvinylidene difluoride membranes. The membranes were immunoblotted with anti-RIG-I (Cell Signaling Technology, Danvers, MA) or anti-IPS-1 (Enzo Life Science, Farmingdale, NY), followed by anti-β-actin (Sigma Aldrich, St. Louis, MO). After immunoblotting with horseradish peroxidase-conjugated secondary antibody, signals were detected by chemiluminescence (BM Chemiluminescence Blotting Substrate, Roche Applied Science, Mannheim, Germany). Optical densitometry was performed using ImageJ software (NIH, Bethesda, MD). Naive Huh7 cells were used for a positive control for full-length IPS-1, and cells transfected with HCV-1b subgenomic replicon[20] were used for a positive control for cleaved IPS-1.

Definitions of Response to Therapy. A patient negative for serum HCV-RNA during the first 6 months after completing PEG-IFNα-2b/RBV combination therapy was defined as a sustained viral responder (SVR), and a patient for whom HCV-RNA became negative at the end of therapy and reappeared after completion of therapy was defined as a transient virological responder (TVR). A patient for whom HCV-RNA became negative at the end of therapy (SVR + TVR) was defined as a virological responder (VR). A patient whose HCV-RNA did not become negative during the course of therapy was defined as an NVR. HCV-RNA was determined by TaqMan HCV assay (Roche Molecular Diagnostics).

Statistical Analysis. Categorical data were compared using the chi-square test and Fisher's exact test. Distributions of continuous variables were analyzed by the Mann-Whitney U test for two groups. All tests of significance were two-tailed and $P < 0.05$ was considered statistically significant.

Results

Patient Characteristics and* IL28B *Genotype. Table 1 shows patient characteristics according to *IL28B* genotype. SNPs at rs8099917 and rs12979860 were 100% identical; 54 patients were identified as having the major alleles (rs8099917 TT/rs12979860 CC; *IL28B* major patients) and the remaining 34 had the minor alleles (rs8099917 TG/rs12979860 CT; *IL28B* minor patients). Patients having a minor homozygote (rs8099917 GG or rs12979860 TT) were not found in this study, which is consistent with a recent report of the rarity of a minor homozygote in Japanese patients.[3] *IL28B* minor patients were significantly associated with a higher γ-glutamyl transpeptidase (γ-GTP) level and higher frequency of mutations at amino acid positions 70 and 91 of the HCV core region (glutamine or histidine mutation at amino acid position 70; methionine mutation at amino acid position 91). NVR rate was significantly higher in *IL28B* minor patients than in *IL28B* major patients.

Gene Expression Involving Innate Immunity and* IFNλ *in the Liver. Hepatic expression levels of cytoplasmic viral sensors (*RIG-I, MDA5*, and *LGP2*) were significantly higher in *IL28B* minor patients than in *IL28B* major patients (Fig. 1). Similarly, expressions of *ISG15* and *USP18* were significantly higher in *IL28B* minor patients than in *IL28B* major patients (Fig. 1). In contrast, the hepatic expression of the adaptor molecule (*IPS-1*) was significantly lower in *IL28B* minor patients than that in *IL28B* major patients (Fig. 1). Hepatic expression of *RNF125* was similar among *IL28B* genotypes (Fig. 1). *IFNλ (IL28A/B)* expression was higher in *IL28B* minor patients, but not statistically significant (Fig. 1). Because expression of *RIG-I* and *IPS-1* were negatively correlated, the expression ratio of *RIG-I/IPS-1* in *IL28B* minor patients was significantly higher than in *IL28B* major patients (Fig. 1).

Next, to assess the relationship between baseline hepatic gene expression and treatment efficacy, we compared levels of gene expression involving innate immunity and *IFNλ* based on the final virological response (Fig. 2). Overall, hepatic expressions of cytoplasmic viral sensors and the ISG15/USP18 system in NVR patients were significantly higher than those in VR patients. In a similar but opposite manner, hepatic expressions of *IPS-1* and *RNF125* in NVR patients were significantly lower than that in VR patients, and the expression of *IFNδ* was higher in NVR patients, but the differences were not statistically significant. Expression ratio of *RIG-I/IPS-1* was significantly higher in NVR patients than that in VR patients.

Because hepatic expressions of the RIG-I/IPS-1 and ISG15/USP18 systems were significantly related both to *IL28B* minor and NVR patients, *RIG-I* and *ISG15* expression levels and the *RIG-I/IPS-1* ratio between VR and NVR patients were further stratified by *IL28B* genotype (Fig. 3). Even in the subgroup of *IL28B* minor patients, the expressions of *RIG-I* and *ISG15* were significantly higher in NVR patients than those in VR patients. Similar tendencies were observed in a subgroup of *IL28B* major patients, in whom the *RIG-I/IPS-1* expression ratio was significantly higher in

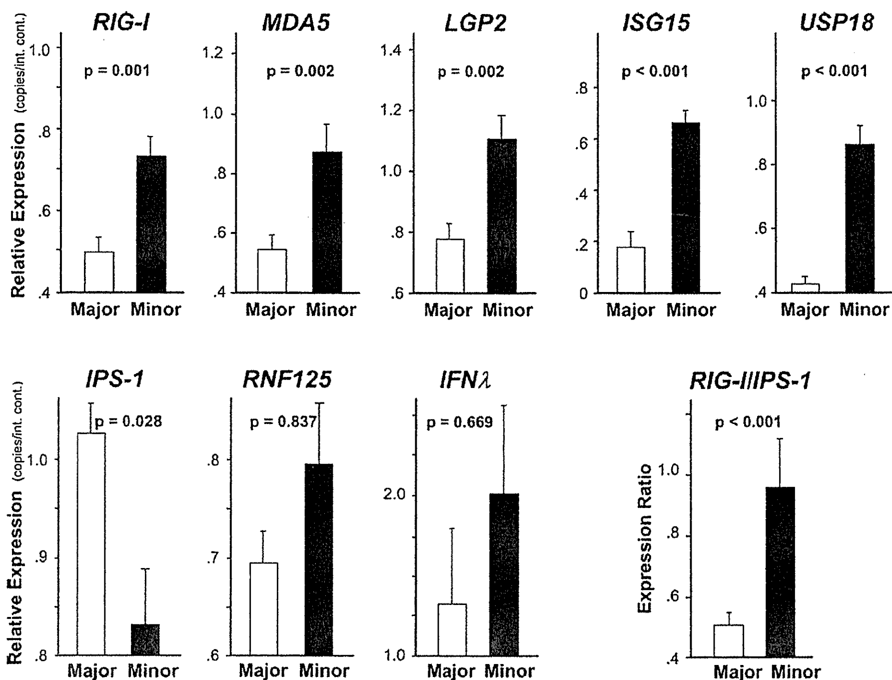


Fig. 1. Comparison of hepatic gene expression levels between *IL28B* major (rs8099917 TT/rs12979860 CC, n = 54) and *IL28B* minor patients (rs8099917 TG/rs12979860 CT, n = 34). Expression levels of cytoplasmic viral sensors (*RIG-I, MDA5,* and *LGP2*), modulators (*ISG15* and *USP18*), an adaptor (*IPS-1*), negative regulators (*RNF125*) and *IFNλ*, and expression ratio of the *RIG-I/IPS-1* are shown. Error bars indicate standard error. The *P*-values were determined by the Mann-Whitney *U* test.

NVR patients than in VR patients. However, in patients of the same virological response subgroup, *RIG-I* and *ISG15* expression levels and *RIG-I/IPS-1* ratio were higher in *IL28B* minor patients, and the difference in *ISG15* expression in subgroup of VR and NVR patients and that in *RIG-I/IPS-1* ratio in subgroup of VR patients was statistically significant between *IL28B* genotypes (Fig. 3).

Receiver Operator Characteristic (ROC) Analysis. To determine the usefulness of these gene quantifications and *IL28B* genotyping as predictors of NVR, an ROC analysis was conducted (Fig. 4A). The area under the ROC curve for *RIG-I* and *ISG15* expressions and *RIG-I/IPS-1* expression ratio was 0.712, 0.782, and 0.732, respectively, suggesting that quantification of these gene transcripts is useful for

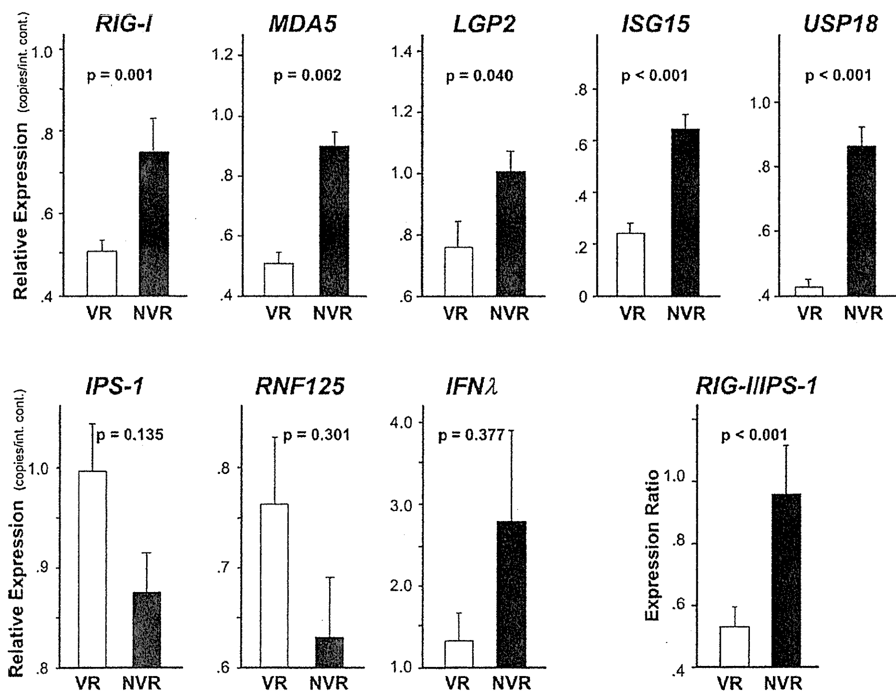


Fig. 2. Comparison of hepatic gene expression levels between virological responders (VR, n = 60) and nonvirological responders (NVR, n = 28). Expression levels of cytoplasmic viral sensors (*RIG-I, MDA5,* and *LGP2*), modulators (*ISG15* and *USP18*), an adaptor (*IPS-1*), negative regulators (*RNF125*) and *IFNλ*, and *RIG-I/IPS-1* expression ratio are shown. Error bars indicate standard error. The *P*-values were determined by the Mann-Whitney *U* test.

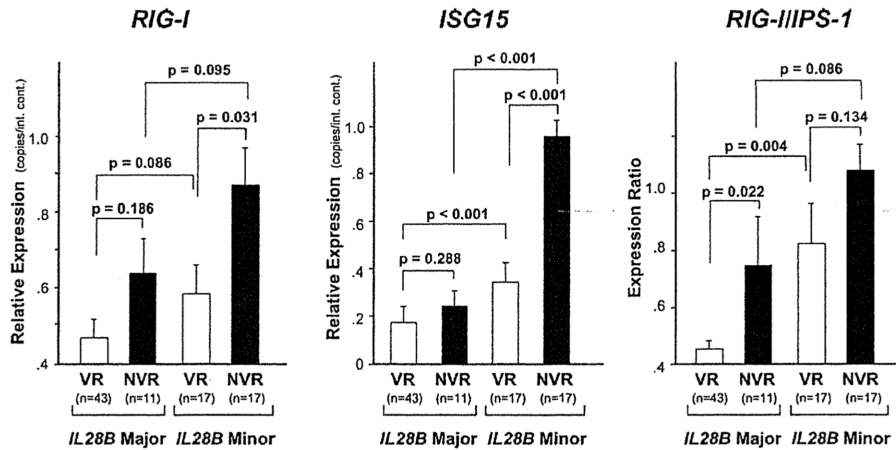


Fig. 3. Comparison of hepatic gene expression levels between virological responders (VR) and nonvirological responders (NVR) in subgroups of the *IL28B* genotype (*IL28B* Major, rs8099917 TT/rs12979860 CC; *IL28B* Minor, rs8099917 TG/rs12979860 CT). Expressions of *RIG-I* and *ISG15* as well as the *RIG-I/IPS-1* expression ratio are shown. Error bars indicate standard error. The numbers of patients in each subgroup are shown in the bottom of the figure.

prediction of NVR (Table 2). The area under the ROC curve for *IL28B* genotype was 0.662, which was lower compared with that for *RIG-I* and *ISG15* expressions and *RIG-I/IPS-1* ratio.

When we stratified the patients by the cutoff value for *RIG-I* and *ISG15* expressions and *RIGI/IPS-1* ratio, no statistically significant difference was found in NVR rates among *IL28B* genotypes within the same subgroup (Fig. 4B).

Factors Associated with NVR. In univariate analysis, age, platelet counts, double mutation at amino acid positions 70 and 91 of the HCV core region, *IL28B* minor allele, and hepatic expressions of *RIG-I, MDA5, LGP2, ISG15*, and *USP18*, and *RIG-I/IPS-1* ratio were significantly

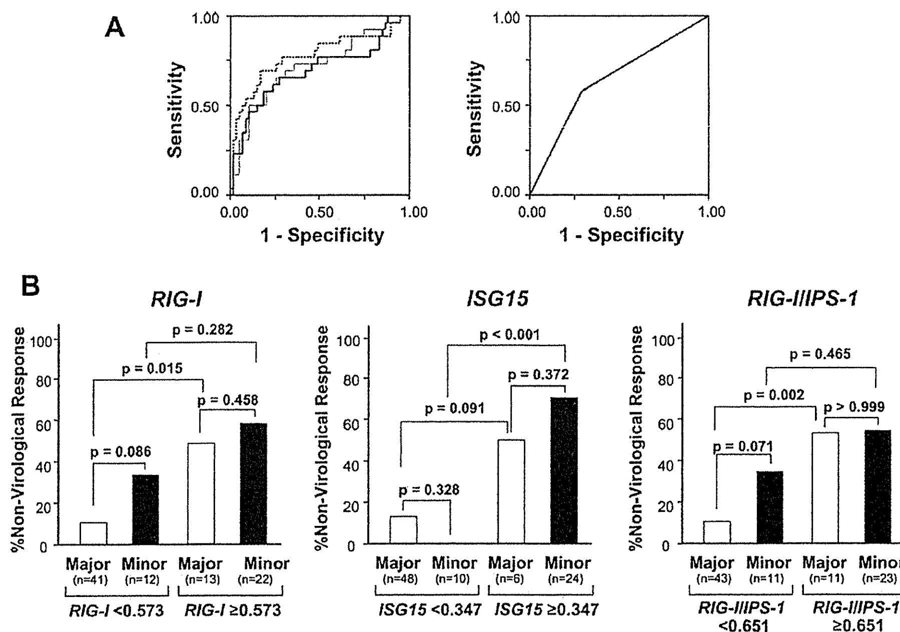


Fig. 4. (A) Receiver operator characteristics (ROC) curve for prediction of nonvirological response. ROC curves were generated to compare *RIG-I* (black line), *ISG15* (dotted line), and *RIG-I/IPS-1* ratio (gray line) (all in the left panel), and *IL28B* genotype (in the right panel). (B) Nonvirological response rate in *IL28B* major (rs8099917 TT/rs12979860 CC) and minor patients (rs8099917 TG/rs12979860 CT) in subgroups divided by the cutoff value of *RIG-I* and *ISG15* expression and the *RIG-I/ISG15* ratio determined by ROC analysis. Cutoff values of *RIG-I* and *ISG15* expression are expressed as expression copy number normalized to the expression of an internal control. The numbers of patients in each subgroup are shown in the bottom of the figure.

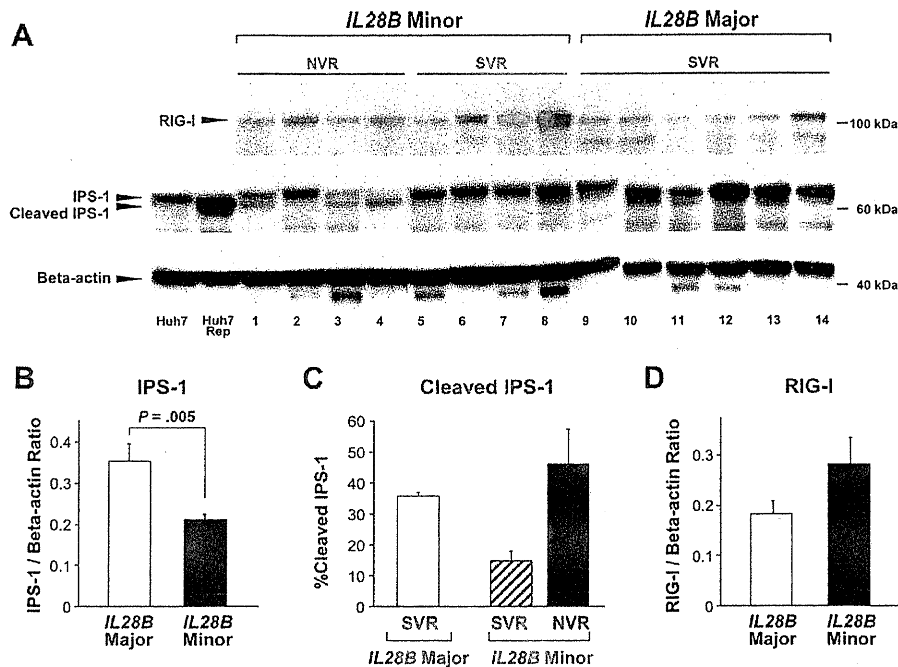


Fig. 5. (A) Western blotting for IPS-1 and RIG-I protein expression levels. Eight lanes contain samples from *IL28B* minor patients (lanes 1-8) and six lanes contain samples from *IL28B* major patients (lanes 9-14). Four lanes contain samples from nonvirological responders (NVR, lanes 1-4) and 10 lanes contain samples from sustained virological responders (SVR, lanes 5-14). Specific bands for RIG-I, full-length IPS-1, cleaved IPS-1, and β-actin are indicated by arrows. Naive Huh7 cells were used for a positive control for full-length IPS-1 (lane Huh7), and cells transfected with HCV-1b subgenomic replicon (Reference #20) were used for a positive control for cleaved IPS-1 (lane Huh7 Rep). (B) Total IPS-1 protein expression levels normalized to β-actin according to *IL28B* genotype. Error bars indicate standard error. *P*-value was determined by Mann-Whitney *U* test. (C) Percentage of cleaved IPS-1 products in total IPS-1 protein according to treatment responses stratified by *IL28B* genotype. Error bars indicate standard error. (D) RIG-I protein expression levels normalized to β-actin according to *IL28B* genotype. Error bars indicate standard error.

expression, total hepatic IPS-1 protein expression was significantly lower in *IL28B* minor patients than in *IL28B* major patients (Fig. 5B). With regard to *IL28B* minor patients, the percentage of cleaved IPS-1 protein in total IPS-1 in SVR was lower than that in NVR (Fig. 5C). In contrast to IPS-1 protein expression, hepatic RIG-I protein expression was higher in *IL28B* minor patients than that in *IL28B* major patients (Fig. 5D).

Discussion

In the present study we found that the baseline expression levels of intrahepatic viral sensors and related regulatory molecules were significantly associated with the genetic variation of *IL28B* and final virological outcome in CH-C patients treated with PEG-IFNα/RBV combination therapy. Although the relationship between the *IL28B* minor allele and NVR in PEG-IFNα/RBV combination therapy is evident, mechanisms responsible for this association remain unknown. *In vitro* studies have suggested that cytoplasmic viral sensors, such as RIG-I and MDA5, play a pivotal role in the regulation of IFN production and augment IFN production through an amplification circuit.[7,8] Our results indicate that expressions of *RIG-I* and *MDA5* and a related amplification system may be up-regulated by endogenous IFN at a higher baseline level in *IL28B* minor patients. However, HCV elimination by subsequent exogenous IFN is insufficient in these patients, as reported,[19] suggesting that *IL28B* minor patients may have adopted a different equilibrium in their innate immune response to HCV. Our data are further supported by recent reports of an association between intrahepatic levels of IFN-stimulated gene expression and PEG-IFNα/RBV response as well as with *IL28B* genotype.[21-23]

In contrast to cytoplasmic viral sensor (*RIG-I*, *MDA5*, and *LGP2*) and modulator (*ISG15* and *USP18*) expression, the adaptor molecule (*IPS-1*) expression was significantly lower in *IL28B* minor patients. Moreover, western blotting further confirmed IPS-1 protein downregulation in *IL28B* minor patients by revealing decreased protein levels. Because IPS-1 is one of the main target molecules of HCV evasion,[9,18]

transcriptional and translational *IPS-1* expression are probably suppressed by HCV with resistant phenotype, which may be more adaptive in *IL28B* minor patients than in *IL28B* major patients. When we analyzed the proportion of full-length or cleaved IPS-1 to the total IPS-1 protein in a subgroup of *IL28B* minor patients, cleaved IPS-1 product was less dominant in SVR than in NVR, whereas uncleaved full-length IPS-1 protein was more dominant in SVR than in NVR. Therefore, the ability of HCV to evade host innate immunity by cleaving IPS-1 protein and/or host capability of protection from IPS-1 cleavage is probably responsible for the variable treatment responses in *IL28B* minor patients.

Our results indicated a close association between *IL28B* minor patients with higher γ-GTP level and higher frequency of HCV core double mutants, which are known factors for NVR. In contrast, no significant association was observed between *IL28B* genotype and age, gender, or liver fibrosis, which are also known to be unfavorable factors for virological response to PEG-IFNα/RBV. Therefore, certain factors other than the *IL28B* genotype may independently influence virological response. To elucidate whether gene expression involving innate immunity independently associates with a virological response from the *IL28B* genotype, we performed further analysis in a subgroup and conducted a multivariate regression and ROC analyses. Our multivariate and ROC analyses demonstrate that higher expressions of *RIG-I* and *ISG15* as well as a higher ratio of *RIG-I/IPS-1* are independently associated with NVR, and quantification of these values is more useful in predicting final virological response to PEG-IFNα/RBV than determination of *IL28B* genotype in each individual patients. However, the SVR rates in our patients were similar among *IL28B* genotypes, which suggests more SVR patients with the *IL28B* minor allele were included in the present study than those in the general CH-C population. Hence, our data did not necessarily exclude the possibility of the *IL28B* genotype in predicting NVR, although our multivariate analysis could not identify the *IL28B* minor allele as an independent factor for NVR. Interestingly, an association between *IL28B* genotype and expressions of *RIG-I* and *ISG15* as well as *RIG-I/IPS-1* expression ratio is still observed even in patients with the same subgroup of virological response (Fig. 3).

In the present study, although hepatic *IFNλ* expression was observed to be higher in *IL28B* minor and NVR patients, it was not statistically significant. Because *IL28B* shares 98.2% homology with *IL28A*, our primer could not distinguish the expression of *IL28B* from that of *IL28A*, and moreover, we could not specify which cell expresses *IFNλ* (i.e., hepatocytes or other immune cells that have infiltrated the liver). Therefore, the precise mechanisms underlying *IL28B* variation and expression of *IFNλ* in relation to treatment response need further clarification by specifying type of IFNλ and uncovering the producing cells.

In the present study we included genotype 1b patients because it is imperative to designate a virologically homogenous patient group to associate individual treatment responses with different gene expression profiles that direct innate immune responses. We have reported that the *RIG-I/IPS-1* ratio was significantly higher in NVR with HCV genotype 2.[19] However, our preliminary results indicated that baseline hepatic *RIG-I* and *ISG15* expression and the *RIG-I/IPS-1* expression ratio is not significantly different among *IL28B* genotypes in patients infected with genotype 2 (Supporting Figure). This may be related to the rarity of NVR with HCV genotype 2 and the lower effect of *IL28B* genotype on virological responses in patients infected with HCV genotype 2.[24] The association among treatment responses in all genotypes, the different status of innate immune responses, and *IL28B* genotype needs to be examined further.

Differences in allele frequency for *IL28B* SNPs among the population groups has been reported. The frequency of *IL28B* major allele among patients with Asian ancestry is higher than that among patients with European and African ancestry.[25] Because *IL28B* polymorphism strongly influences treatment responses within each population group,[5] our data obtained from Japanese patients can be applied to other population groups. However, the rate of SVR having African ancestry was lower than that having European ancestry within the same *IL28B* genotype.[5] Hence, further study is required to clarify whether this difference among the population groups with the same *IL28B* genotype could be explained by differences in expression of genes involved in innate immunity.

In a recent report, an SVR rate of telaprevir with PEG-IFNα/RBV was only 27.6% in *IL28B* minor patients.[26] Because new anti-HCV therapy should still contain PEG-IFNα/RBV as a platform for the therapy, our findings regarding innate immunity in addressing the mechanism of virological response and predicting NVR remain important in this new era of directly acting anti-HCV agents, such as telaprevir and boceprevir.

In conclusion, this clinical study in humans demonstrates the potential relevance of the molecules involved in innate immunity to the genetic variation

of *IL28B* and clinical response to PEG-IFNα/RBV. Both the *IL28B* minor allele and higher expressions of *RIG-I* and *ISG15* as well as higher *RIG-I/IPS-1* ratio are independently associated with NVR. Innate immune responses in *IL28B* minor patients may have adapted to a different equilibrium compared with that in *IL28B* major patients. Our data will advance both understanding of the pathogenesis of HCV resistance and the development of new antiviral therapy targeted toward the innate immune system.

References

1. Kiyosawa K, Sodeyama T, Tanaka E, Gibo Y, Yoshizawa K, Nakano Y, 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. Zeuzem S, Pawlotsky JM, Lukasiewicz E, von Wagner M, Goulis I, Lurie Y, et al. DITTO-HCV Study Group. International, multicenter, randomized, controlled study comparing dynamically individualized versus standard treatment in patients with chronic hepatitis C. J Hepatol 2005;43:250-257.
3. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, et al. Genome-wide association of IL28B with response to pegylated IFN-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 2009;10:1105-1109.
4. Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate ML, et al. IL28B is associated with response to chronic hepatitis C IFN-alpha and ribavirin therapy. Nat Genet 2009;10:1100-1104.
5. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 2009;461:399-401.
6. Biron CA. Initial and innate responses to viral infections—pattern setting in immunity or disease. Curr Opin Microbiol 1999;2:374-381.
7. Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M, et al. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nat Immunol 2004;5:730-737.
8. Yoneyama M, Kikuchi M, Matsumoto K, Imaizumi T, Miyagishi M, Taira K, et al. Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. J Immunol 2005;175:2851-2858.
9. Meylan E, Curran J, Hofmann K, Moradpour D, Binder M, Bartenschlager R, et al. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. Nature 2005;437:1167-1172.
10. Kawai T, Takahashi K, Sato S, Coban C, Kumar H, Kato H, et al. IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. Nat Immunol 2005;6:981-988.
11. Seth RB, Sun L, Ea CK, Chen ZJ. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. Cell 2005;122:669-682.
12. Xu LG, Wang YY, Han KJ, Li LY, Zhai Z, Shu HB. VISA is an adapter protein required for virus-triggered IFN-beta signaling. Mol Cell 2005;19:727-740.
13. Rothenfusser S, Goutagny N, DiPerna G, Gong M, Monks BG, Schoenemeyer A, et al. The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. J Immunol 2005;175:5260-5268.
14. Arimoto K, Takahashi H, Hishiki T, Konishi H, Fujita T, Shimotohno K. Negative regulation of the RIG-I signaling by the ubiquitin ligase RNF125. Proc Natl Acad Sci U S A 2007;104:7500-7505.
15. Zhao C, Denison C, Huibregtse JM, Gygi S, Krug RM. Human ISG15 conjugation targets both IFN-induced and constitutively expressed proteins functioning in diverse cellular pathways. Proc Natl Acad Sci U S A 2005;102:10200-10205.
16. Schwer H, Liu LQ, Zhou L, Little MT, Pan Z, Hetherington CJ, et al. Cloning and characterization of a novel human ubiquitin-specific protease, a homologue of murine UBP43 (Usp18). Genomics 2000;65:44-52.
17. Malakhov MP, Malakhova OA, Kim KI, Ritchie KJ, Zhang DE. UBP43 (USP18) specifically removes ISG15 from conjugated proteins. J Biol Chem 2002;277:9976-9981.
18. Li XD, Sun L, Seth RB, Pineda G, Chen ZJ. Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. Proc Natl Acad Sci U S A 2005;102:17717-17722.
19. Asahina Y, Izumi N, Hirayama I, Tanaka T, Sato M, Yasui Y, et al. Potential relevance of cytoplasmic viral sensors and related regulators involving innate immunity in antiviral response. Gastroenterology 2008;134:1396-1405.
20. Tanabe Y, Sakamoto N, Enomoto N, Kurosaki M, Ueda E, Maekawa S, et al. Synergistic inhibition of intracellular hepatitis C virus replication by combination of ribavirin and interferon-alpha. J Infect Dis 2004;189:1129-1139.
21. Honda M, Sakai A, Yamashita T, Nakamoto Y, Mizukoshi E, Sakai Y, et al. Hepatic ISG expression is associated with genetic variation in interleukin 28B and the outcome of IFN therapy for chronic hepatitis C. Gastroenterology 2010;139:499-509.
22. Urban TJ, Thompson AJ, Bradric SS, Fellay J, Schuppan D, Cronin KD, et al. *IL28B* genotype is associated with differential expression of intrahepatic interferon-stimulated genes in patients with chronic hepatitis C. HEPATOLOGY 2010;52:1888-1896.
23. Dill MT, Duong FHT, Vogt JE, Bibert S, Bochud PY, Terracciano L, et al. Interferon-induced gene expression is a stronger predictor of treatment response than *IL28B* genotype in patients with hepatitis C. Gastroenterology 2011;140:1021-1031.
24. Yu ML, Huang CF, Huang JF, Chang NC, Yang JF, Lin ZY, et al. Role of interleukin-28B polymorphism in the treatment of hepatitis C virus genotype 2 infection in Asian patients. HEPATOLOGY 2011;53:7-13.
25. Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'hUigin C, Kidd J, et al. Genetic variation in *IL28B* and spontaneous clearance of hepatitis C virus. Nature 2009;461:798-802.
26. Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, et al. Amino acid substitution in hepatitis C virus core region and genetic variation near the interleukin 28B gene predict viral response to telaprevir with pegIFN and ribavirin. HEPATOLOGY 2010;52:421-429.

Research Article

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

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

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

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

Background & Aims: Assessment of the risk of hepatocellular carcinoma (HCC) development is essential for formulating personalized surveillance or antiviral treatment plan for chronic hepatitis C. We aimed to build a simple model for the identification of patients at high risk of developing HCC.
Methods: Chronic hepatitis C patients followed for at least 5 years (n = 1003) were analyzed by data mining to build a predictive model for HCC development. The model was externally validated using a cohort of 1072 patients (472 with sustained virological response (SVR) and 600 with nonSVR to PEG-interferon plus ribavirin therapy).
Results: On the basis of factors such as age, platelet, albumin, and aspartate aminotransferase, the HCC risk prediction model identified subgroups with high-, intermediate-, and low-risk of HCC with a 5-year HCC development rate of 20.9%, 6.3–7.3%, and 0–1.5%, respectively. The reproducibility of the model was confirmed through external validation ($r^2 = 0.981$). The 10-year HCC development rate was also significantly higher in the high- and intermediate-risk group than in the low-risk group (24.5% *vs.* 4.8%; $p < 0.0001$). In the high-and intermediate-risk group, the incidence of HCC development was significantly reduced in patients with SVR compared to those with nonSVR (5-year rate, 9.5% *vs.* 4.5%; $p = 0.040$).
Conclusions: The HCC risk prediction model uses simple and readily available factors and identifies patients at a high risk of HCC development. The model allows physicians to identify patients requiring HCC surveillance and those who benefit from IFN therapy to prevent HCC.

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

Introduction

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

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

Cancer

Keywords: Decision tree; Prediction; Pegylated interferon; Ribavirin; Risk.
Received 27 May 2011; received in revised form 8 August 2011; accepted 4 September 2011
* Corresponding author. Address: Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, 1-26-1 Kyonan-cho, Musashino-shi, Tokyo 180-8610, Japan. Tel.: +81 422 32 3111; fax: +81 422 32 9551.
E-mail address: nizumi@musashino.jrc.or.jp (N. Izumi).

ELSEVIER

Please cite this article in press as: Kurosaki M et al. Data mining model using simple and readily available factors could identify patients at high risk for hepatocellular carcinoma in chronic hepatitis C. J Hepatol (2011), doi:10.1016/j.jhep.2011.09.011

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

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

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

Materials and methods

Patients

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

HCC surveillance and diagnosis

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

Statistical analysis

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

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

Results

Univariate and multivariate analysis of factors associated with HCC development

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

HCC risk prediction model by data mining analysis

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

Please cite this article in press as: Kurosaki M et al. Data mining model using simple and readily available factors could identify patients at high risk for hepatocellular carcinoma in chronic hepatitis C. J Hepatol (2011), doi:10.1016/j.jhep.2011.09.011

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

	Model derivation (n = 1003)	External cohort, non-SVR (n = 600)	External cohort, SVR (n = 472)
Sex: Male/Female*	463 (46%)/540 (54%)	306 (51%)/294 (49%)	299 (63%)/173 (37%)
Age (yr)	57.3 (11.1)	55.9 (9.6)	51.4 (10.6)
Body mass index (kg/m²)	23.5 (3.2)	23.4 (3.3)	23.3 (3.1)
Albumin (g/dl)	4.1 (0.3)	4.0 (0.4)	4.0 (0.3)
AST (IU/L)	64.2 (36.5)	67.3 (43.8)	62.5 (48.3)
ALT (IU/L)	80.6 (55.1)	81.2 (62.3)	88.6 (82.1)
GGT (IU/L)	59.3 (50.5)	67.6 (65.1)	55.7 (71.2)
Total cholesterol (mg/dl)	172.1 (31.5)	168.2 (31.0)	174.3 (33.7)
Platelet (10^9/L)	154.0 (53.0)	153.7 (53.2)	176.6 (49.7)
Hemoglobin (g/dl)	13.3 (1.5)	14.2 (1.5)	14.4 (1.4)
HCC development within 5 years: n (%)*	62 (6.2%)	36 (6.0%)	10 (2.0%)

Data expressed as mean (standard deviation) unless otherwise indicated.
AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma-glutamyltransferase; HCC, hepatocellular carcinoma; SVR, sustained virological response.
*Data expressed as number of patients (percentage).

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

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

Six hundred nonSVR patients from an external cohort were fitted into the HCC risk prediction model and classified into the seven subgroups. The 5-year HCC development rate of these subgroups was 0–17.9%. The HCC development rate in the individual subgroups of the model derivation group was closely correlated to that in the corresponding subgroups of the external validation group (Fig. 2; correlation coefficient r^2 = 0.981). The HCC development rate in the subgroup of patients with the highest risk of HCC development (high-risk group) according to the model older age (⩾60 years) with a lower platelet count (<150 × 10^9/L) and lower albumin levels (<4.0 g/dl) was 20.9% in the model derivation group and 17.9% in the external validation group. The intermediate-risk group or the patients with an HCC development rate of at least 5% consisted of the following three subgroups: (1) older age (⩾60 years), lower platelet count (<150 × 10^9/L), higher albumin levels (⩾4.0 g/dl), and higher AST levels (⩾40 IU/L); (2) older age (⩾60 years), higher platelet count (⩾150 × 10^9/L), and lower albumin levels (<3.75 g/dl); and (3) younger age (<60 years) and lower platelet count (<150 × 10^9/L). In these intermediate-risk groups, the 5-year HCC development rate was 6.3–7.3% in the model derivation group and 5.3–7.9% in the external validation group. The low-risk group consisted of the following three subgroups: (1) younger age (<60 years) and higher platelet count (⩾150 × 10^9/L); (2) older age (⩾60 years), lower platelet count (<150 × 10^9/L), higher albumin levels (⩾4.0 g/dl), and lower AST levels (<40 IU/L); and (3) older age (⩾60 years), higher platelet count (⩾150 × 10^9/L), and higher albumin levels (⩾3.75 g/dl). In these low-risk groups, the 5-year HCC development rate was 0–1.5% in the model derivation group and 0–2.9% in the external validation group.

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

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

CI, confidence interval.

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

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

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

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

Cancer

Please cite this article in press as: Kurosaki M et al. Data mining model using simple and readily available factors could identify patients at high risk for hepatocellular carcinoma in chronic hepatitis C. J Hepatol (2011), doi:10.1016/j.jhep.2011.09.011

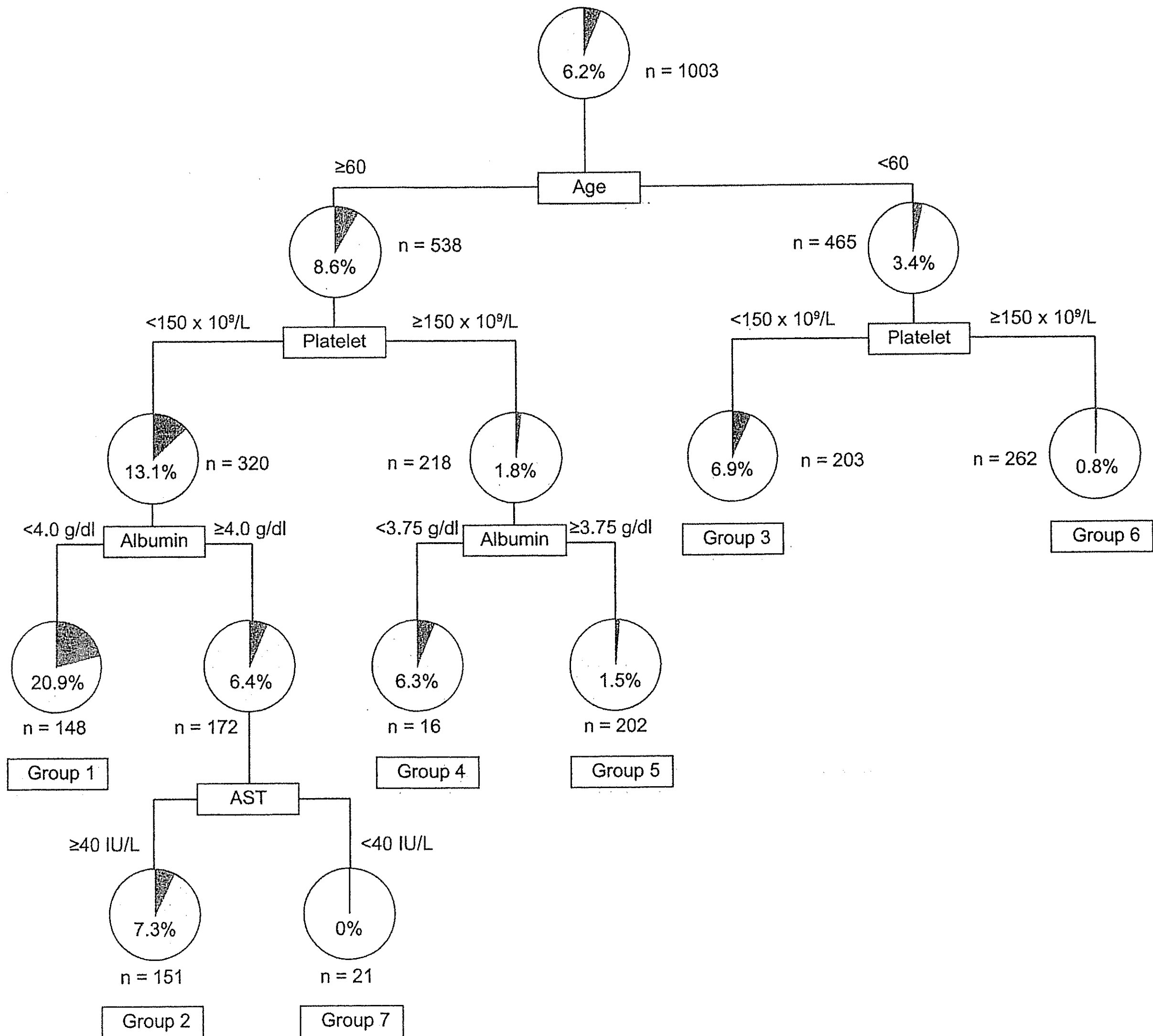


Fig. 1. The decision tree model of HCC development within 5 years. Boxes indicate the factors used to differentiate patients and the cutoff values for those different groups. Pie charts indicate the HCC development rate within 5 years for each group of patients after differentiation. Terminal groups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

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

Discussion

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

Please cite this article in press as: Kurosaki M et al. Data mining model using simple and readily available factors could identify patients at high risk for hepatocellular carcinoma in chronic hepatitis C. J Hepatol (2011), doi:10.1016/j.jhep.2011.09.011

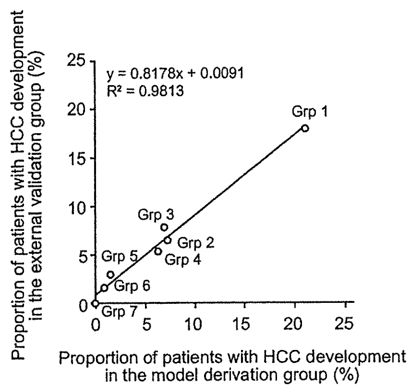


Fig. 2. External validation of the decision tree model with an independent cohort. Each patient in the external validation group was allocated to groups 1–7 following the flowchart of the decision tree. The HCC development rates were then calculated for each group and the graph plotted. The *x*-axis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

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

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

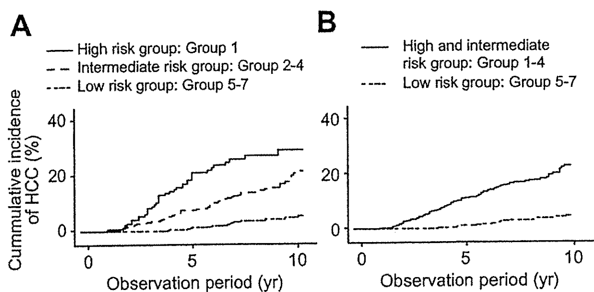


Fig. 3. Cumulative incidence of HCC development beyond 5 years in subgroups of patients defined by the decision tree model. Cumulative incidences of HCC in the groups classified by the decision tree model are compared. (A) The cumulative HCC development rate beyond 5 years is higher in the high- (group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

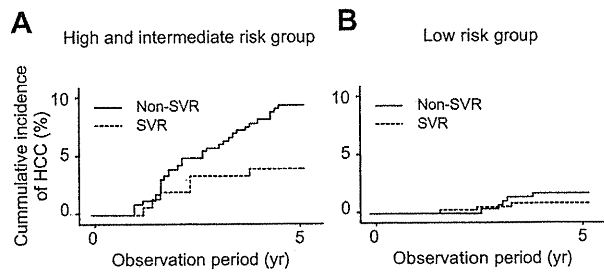


Fig. 4. Sustained virological response to PEG-IFN plus RBV therapy reduces the incidence of HCC development after stratification by the HCC risk. The 600 nonSVR patients and the 472 SVR patients in the external cohort were fitted into the HCC risk prediction model and classified into the high and intermediate-risk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

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

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

Cancer

Please cite this article in press as: Kurosaki M et al. Data mining model using simple and readily available factors could identify patients at high risk for hepatocellular carcinoma in chronic hepatitis C. J Hepatol (2011), doi:10.1016/j.jhep.2011.09.011

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

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

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

Conflict of interest

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

Financial support

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

References

[1] Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55:74–108.

[2] Bosch FX, Ribes J, Diaz M, Cleries R. Primary liver cancer: worldwide incidence and trends. Gastroenterology 2004;127:S5–S16.

[3] Bruix J, Sherman M. Management of hepatocellular carcinoma. Hepatology 2005;42:1208–1236.

[4] Fattovich G, Giustina G, Degos F, Tremolada F, Diodati G, Almasio P, et al. Morbidity and mortality in compensated cirrhosis type C: a retrospective follow-up study of 384 patients. Gastroenterology 1997;112:463–472.

[5] Niederau C, Lange S, Heintges T, Erhardt A, Buschkamp M, Hurter D, et al. Prognosis of chronic hepatitis C: results of a large, prospective cohort study. Hepatology 1998;28:1687–1695.

[6] Degos F, Christidis C, Ganne-Carrie N, Farmachidi JP, Degott C, Guettier C, et al. Hepatitis C virus related cirrhosis: time to occurrence of hepatocellular carcinoma and death. Gut 2000;47:131–136.

[7] Yoshida H, Tateishi R, Arakawa Y, Sata M, Fujiyama S, Nishiguchi S, et al. Benefit of interferon therapy in hepatocellular carcinoma prevention for individual patients with chronic hepatitis C. Gut 2004;53:425–430.

[8] Kurosaki M, Hosokawa T, Matsunaga K, Hirayama I, Tanaka T, Sato M, et al. Hepatic steatosis in chronic hepatitis C is a significant risk factor for developing hepatocellular carcinoma independent of age, sex, obesity, fibrosis stage and response to interferon therapy. Hepatol Res 2010;40:870–877.

[9] Velazquez RF, Rodriguez M, Navascues CA, Linares A, Perez R, Sotorrios NG, et al. Prospective analysis of risk factors for hepatocellular carcinoma in patients with liver cirrhosis. Hepatology 2003;37:520–527.

[10] Lok AS, Seeff LB, Morgan TR, di Bisceglie AM, Sterling RK, Curto TM, et al. Incidence of hepatocellular carcinoma and associated risk factors in hepatitis C-related advanced liver disease. Gastroenterology 2009;136:138–148.

[11] Cardoso AC, Moucari R, Figueiredo-Mendes C, Ripault MP, Giuily N, Castelnau C, et al. Impact of peginterferon and ribavirin therapy on hepatocellular carcinoma: incidence and survival in hepatitis C patients with advanced fibrosis. J Hepatol 2010;52:652–657.

[12] Craxi A, Camma C. Prevention of hepatocellular carcinoma. Clin Liver Dis 2005;9:329–346, viii.

[13] Breiman LJH, Friedman RA, Olshen CJ, Stone CM. Classification and regression trees. Calif: Wadsworth; 1980.

[14] Garzotto M, Park Y, Mongoue-Tchokote S, Bledsoe J, Peters L, Blank BH, et al. Recursive partitioning for risk stratification in men undergoing repeat prostate biopsies. Cancer 2005;104:1911–1917.

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

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

[17] Leiter U, Buettner PG, Eigentler TK, Garbe C. Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the german dermatological society. J Clin Oncol 2004;22:3660–3667.

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

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

[20] Baquerizo A, Anselmo D, Shackleton C, Chen TW, Cao C, Weaver M, et al. Phosphorus as an early predictive factor in patients with acute liver failure. Transplantation 2003;75:2007–2014.

[21] Kurosaki M, Matsunaga K, Hirayama I, Tanaka T, Sato M, Yasui Y, et al. A predictive model of response to peginterferon ribavirin in chronic hepatitis c using classification and regression tree analysis. Hepatol Res 2010;40:251–260.

[22] Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, Hiramatsu N, et al. Sequences in the interferon sensitivity-determining region and core region of hepatitis C virus impact pretreatment prediction of response to PEG-interferon plus ribavirin: data mining analysis. J Med Virol 2011;83:445–452.

[23] Izumi N, Asahina Y, Kurosaki M. Predictors of virological response to a combination therapy with pegylated interferon plus ribavirin including virus and host factors. Hepat Res Treat 2010;2010:703602.

[24] Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, Hiramatsu N, et al. Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis. J Gastroenterol 2011;46:401–409.

[25] Kurosaki M, Tanaka Y, Nishida N, Sakamoto N, Enomoto N, Honda M, et al. Pre-treatment prediction of response to pegylated-interferon plus ribavirin for chronic hepatitis C using genetic polymorphism in IL28B and viral factors. J Hepatol 2011;54:439–448.

[26] Kurosaki M, Hiramatsu N, Sakamoto M, Suzuki Y, Iwasaki M, Tamori A, et al. Age and total ribavirin dose is an independent predictor of relapse among early virological responders to peg-interferon plus ribavirin therapy in chronic hepatitis C revealed by data mining analysis. Antivir Ther, in press.

[27] Hiramatsu N, Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, et al. Pretreatment prediction of anemia progression by pegylated interferon alpha-2b plus ribavirin combination therapy in chronic hepatitis C infection: decision-tree analysis. J Gastroenterol 2011;46:1111–1119.

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

[29] Lu SN, Wang JH, Liu SL, Hung CH, Chen CH, Tung HD, et al. Thrombocytopenia as a surrogate for cirrhosis and a marker for the identification of patients at high-risk for hepatocellular carcinoma. Cancer 2006;107:2212–2222.

[30] Bruno S, Stroffolini T, Colombo M, Bollani S, Benvegnu L, Mazzella G, et al. Sustained virological response to interferon-alpha is associated with improved outcome in HCV-related cirrhosis: a retrospective study. Hepatology 2007;45:579–587.

[31] Bonis PA, Tong MJ, Blatt LM, Conrad A, Griffith JL. A predictive model for the development of hepatocellular carcinoma, liver failure, or liver transplantation for patients presenting to clinic with chronic hepatitis C. Am J Gastroenterol 1999;94:1605–1612.

Please cite this article in press as: Kurosaki M et al. Data mining model using simple and readily available factors could identify patients at high risk for hepatocellular carcinoma in chronic hepatitis C. J Hepatol (2011), doi:10.1016/j.jhep.2011.09.011