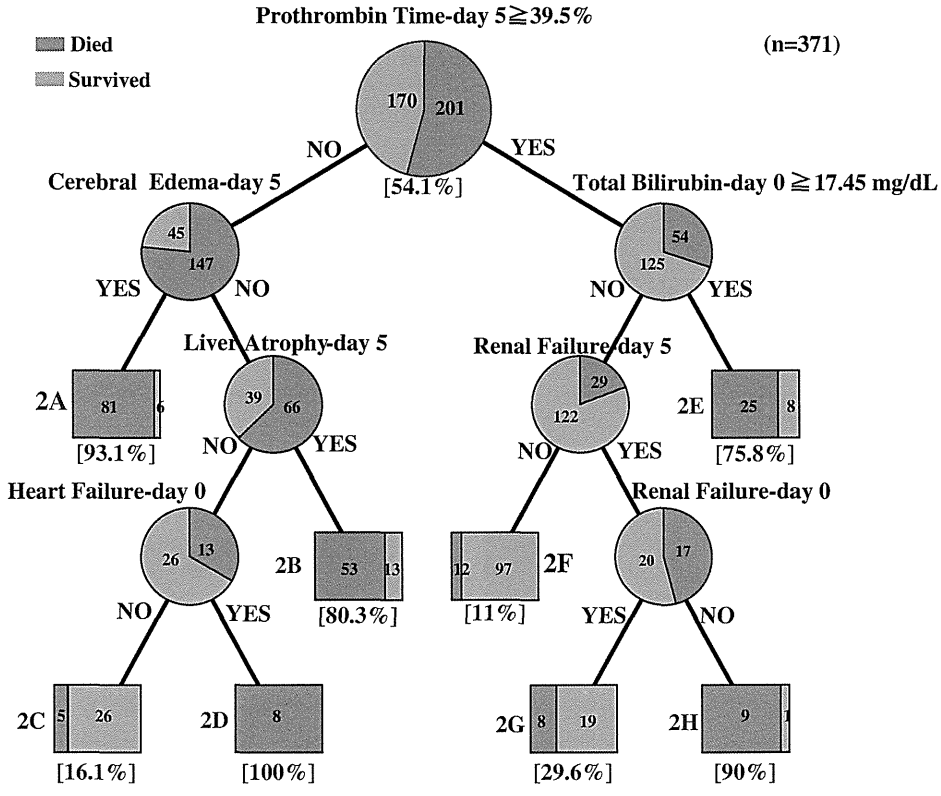

Fig. 2 The decision tree algorithm for outcome prediction at 5 days after the onset of grade II or more severe hepatic encephalopathy (day 5)

Table 3 The accuracy of the decision tree algorithms to predict the prognostic outcome of acute liver failure patients at the onset of hepatic encephalopathy and 5 days later

	At the onset of hepatic encephalopathy	At 5 days after the onset of hepatic encephalopathy
Patients for the formation of the algorithm 1998–2003 (n = 371)		
Accuracy	79.0	83.6
Sensitivity	77.6	82.6
Specificity	80.6	84.7
PPV	82.5	86.5
NPV	75.3	80.5
Patients for the validation of the algorithm 2004–2007 (n = 160)		
Accuracy	71.2	73.1
Sensitivity	75.0	63.6
Specificity	67.1	82.4
PPV	70.6	77.8
NPV	71.8	70.0

PPV positive predictive value, *NPV* negative predictive value

the analysis using the data set at the onset of hepatic encephalopathy, and 8 categories using the data set at 5 days after the onset of encephalopathy. The number of patients who died and the mortality rates of the patients in each category are shown in Table 4. The distribution of the patients and the mortality rates in each category were almost equivalent to those in the patients used for the formation of the algorithms both at the onset of hepatic encephalopathy and 5 days later, except for category-2C. The mortality rate in patients classified as category-2C was 16.1% in patients used for the formation of the algorithm, while the rate was 91.7% in those used for the validation (Table 4b).

The predictive accuracies assessed in patients for validation of the algorithms were 71 and 73%, respectively, at the onset of hepatic encephalopathy and 5 days later, similar to findings in the patients used for the formation of the algorithms (Table 3). The sensitivity, specificity, PPV, and NPV were 75, 67, 71, and 72%, respectively, at the onset of the encephalopathy, and 64, 82, 78, and 70%, respectively, at 5 days after the onset of encephalopathy.

Application of the algorithms for ALF patients receiving liver transplantation

When the data from the 211 patients who had received liver transplantation were applied for the established algorithms at the onset of hepatic encephalopathy, 141 patients (66.8%) were classified as category-1A, category-1B, category-1D, or category-1F, in which categories the

Table 4 The numbers of deaths and the mortality rates of patients in each category classified through decision tree analysis: comparison among patients used for the formation of the algorithm, those used for the validation of the algorithm, and those who received liver transplantation

Categories classified through decision tree analysis	Mortality rates of patients % (number of patients)		Number of patients
	Patients for algorithm formation 1998–2003 ($n = 371$)	Patients for algorithm validation 2004–2007 ($n = 160$)	Patients receiving liver transplantation 1998–2007 ($n = 211$)
(a) The algorithm for the patients at the onset of hepatic encephalopathy			
1A	89.0 (81/91)	83.9 (26/31)	95
1B	79.6 (39/49)	50.0 (16/32)	34
1C	22.6 (7/31)	37.5 (3/8)	10
1D	66.7 (20/30)	83.3 (10/12)	8
1E	30.8 (4/13)	18.2 (2/11)	7
1F	88.2 (15/17)	80.0 (8/10)	4
1G	25.0 (35/140)	30.2 (16/53)	53
(b) The algorithm for the patients at 5 days after the onset of hepatic encephalopathy			
2A	93.1 (81/87)	86.4 (19/22)	19
2B	80.3 (53/66)	71.4 (15/21)	36
2C	16.1 (5/31)	91.7 (11/12)	16
2D	100.0 (8/8)	– (0/0)	0
2E	75.8 (25/33)	72.7 (8/11)	18
2F	11.0 (12/108)	17.3 (9/52)	20
2G	29.6 (8/27)	25.0 (4/16)	1
2H	90.0 (9/10)	– (0/0)	2

mortality rates were greater than 50% in patients for the formation of the algorithm (Table 4a). In contrast, 53 patients (25.2%) were classified as category-1G, in which the mortality rates were 25.0 and 29.4%, respectively, in patients used for the formation and those used for the validation of the algorithm.

The outcome at 5 days after the onset of hepatic encephalopathy was assessed in 112 (53.1%) of the 211 patients who had received liver transplantation, because the transplantation was done within 5 days after the onset of hepatic encephalopathy in 99 patients (Table 4b). Consequently, 75 (67.0%) of the 112 patients were classified as category-2A, category-2B, category-2D, category-2E, or category-2H for the formation of the algorithm, in which categories the mortality rates were greater than 50%. Sixteen patients (14.3%) were classified as category-2C for validation of the algorithm, in which category the mortality rate was greater than 90%, despite the fact that the mortality in it was only 16.1% in the patients used for formation of the algorithm.

Discussion

In the present study, we established a predictive model to determine the outcome of patients with ALF through decision tree analysis, one of the data-mining methods. Data-mining has been applied to analysis in fields such as business intelligence, marketing, banking and finance, customer relationship management, and engineering, as well as various areas of science, including medicine. In clinical medicine, data-mining techniques are used to construct a predictive model, which supports clinical decisions for researchers as well as practitioners [17]. A decision tree algorithm is one of the most popular data-mining techniques, constructed through recursive data partitioning, where the data are split according to the values of a selected attribute in iteration. Decision trees have already been applied to the field of hepatology; for example, to analyze the characteristic features of hepatocellular carcinoma [18–20], and to evaluate the therapeutic efficacy of pegylated-interferon and ribavirin for patients with chronic hepatitis due to HCV infection [21, 22].

In the present study, algorithms of two types were established; an algorithm for use at the onset of hepatic encephalopathy and one for use 5 days later, because, in Japan, conservative medical care including artificial liver support is generally performed in most patients, including those receiving liver transplantation, following the onset of hepatic encephalopathy. In fact, as shown in Table 1c, plasma exchange and hemodiafiltration were carried out in more than 90 and 70%, respectively, of patients with ALF. Thus, the outcome of the patients could be evaluated 5 days after the onset of hepatic encephalopathy in 53% of patients receiving liver transplantation (Table 4). The data sets obtained from ALF patients seen between 1998 and

2003 were used for the formation of the algorithms and those from the patients seen between 2004 and 2007 for their validation, because the outcomes of the patients seen in the two periods were almost equivalent, although there were some differences between the two periods in the frequencies of the therapeutic procedures undertaken (Table 1c, d).

According to the established decision tree algorithms, the patients with ALF were classified into 7 categories through 6 items at the onset of hepatic encephalopathy and into 8 categories through 7 items at 5 days after the onset of hepatic encephalopathy. Serum bilirubin concentration was selected as the first split item in the former algorithm, and the patients were further classified based on peripheral blood platelet counts, age, presence or absence of ascites, and the etiology of liver injuries. In contrast, the prothrombin time at 5 days after the onset of encephalopathy was the first split item in the latter algorithm, and the patients were then classified based on the serum bilirubin concentration and presence or absence of cerebral edema, liver atrophy, and cardiac and renal failure at the onset of encephalopathy or 5 days later. The interval between the onset of disease symptoms and hepatic encephalopathy has been considered to be one of the most important factors to determine the prognosis of ALF patients [4], and this factor was selected as a parameter in the previous guidelines [5]. The prothrombin time and the ratio of the direct-to-total bilirubin concentration at the onset of hepatic encephalopathy were previously selected as parameters as well [5]. However, these factors were not chosen as items responsible for the prognosis of ALF patients in our novel model established through decision tree analysis. These decisions are in line with findings in our previous report [7], in which ALF patients could be classified into three clusters independent of the interval between the onset of disease symptoms and the onset of hepatic encephalopathy, and the prognosis of the patients differed markedly among the clusters. Moreover, among 7 items in the algorithms at 5 days after the onset of hepatic encephalopathy, the extent of cerebral edema, renal failure, and heart failure may vary depending on the therapeutic devices used, especially regarding methods for artificial liver support (ALS) [23–25]. High-flow continuous hemodiafiltration (CHDF) and on-line hemodiafiltration (HDF) are much more effective than conventional HDF and CHDF [26, 27]. In the present study, most of the patients received conventional CHDF and HDF (data not shown), and such therapeutic devices were not selected as factors affecting the prognoses of the patients.

Certain characteristic features in both our algorithms are deserving of inclusion in the algorithms. First, the categories can be divided into 2 types depending on their mortality rates; the mortality rates in patients used for the formation of the model were greater than 66.7% in 4 categories in both algorithms, while they were less than 33.3% in the remaining 3 and 4 categories, respectively, in the algorithm used at the onset of hepatic encephalopathy and that used 5 days later. Secondly, 341 of the 371 patients used for the establishment of decision trees (91.9%) were classified into 4 major categories, in which the number of patients belonging to each category was greater than 30 in the algorithm at the onset of hepatic encephalopathy. Also, 325 patients (87.6%) were classified into 5 major categories in the algorithm at 5 days after the onset of hepatic encephalopathy. Considering these characteristic features of both algorithms, the novel model constructed through the decision tree analysis seems to be useful for the prediction of the outcome of patients with ALF, because the first characteristic above allowed the analysis to achieve high accuracy rates when the outcomes of the patients were predicted qualitatively as "death" or "survival". In contrast, the second characteristic may enable us to obtain stable results for prediction even after the validation. In fact, as shown in Table 3, the predictive accuracies of both algorithms were high; 79.0 and 83.6%, respectively, in the algorithm at the onset of hepatic encephalopathy and that at 5 days later, when the outcome of patients belonging to the categories with mortality rates greater than 50% was predicted as "death". Moreover, the sensitivity, specificity, PPV, and NPV values were greater than 75% in the algorithm at the onset of hepatic encephalopathy, and greater than 80% in the algorithm at 5 days later. Also, the mortality rates in patients used for the algorithm formation were similar to those in the patients used for the validation in each category, except for category-2C. As a result, the predictive accuracies were also high in patients used for the validation algorithm; 71.2 and 73.1%, respectively, in the algorithm at the onset of hepatic encephalopathy and that at 5 days later, when the outcome of patients was assessed qualitatively. Thus, it is concluded that the present model, consisting of 2 algorithms, may be useful to predict the outcome of ALF patients both quantitatively and qualitatively. Clinicians can obtain the predictive mortality rates of the patients depending on the categories to which the patients belong, and they can also predict the outcome as "death" or "survival" with satisfactory accuracies.

However, there are several weak points in both algorithms to predict the outcome of the ALF patient. Although the reproducibility of the algorithm at the onset of hepatic encephalopathy was generally good in each category, a 29.6% difference in mortality rates was found between the formation and validation data sets in category-1B. Also, there was a 75.6% difference between the two data sets in category-2C. Moreover, the validation could not be done in categories-2D and -2H, because no patients were classified in these categories in the validation groups, and a similar

situation was found in the analysis of patients who had received liver transplantation. The significance of such minor terminal nodes (leaves) constructed with small numbers of patients should be further validated in patients enrolled in the nationwide survey since 2008.

Liver transplantation was performed in 221 (21.6%) of the 1,022 patients enrolled in the study. These patients were excluded from the subjects used for the formation and validation of the decision tree algorithms. However, we evaluated the possible outcomes of these patients using the established algorithms. To our surprise, as shown in Table 4, 33% of the transplanted patients were classified into the categories showing a predictive mortality rate of less than 50% both at the onset of hepatic encephalopathy and at 5 days later. We note particularly that there existed 53 of 211 transplanted patients (25.1%) belonging to category-1G, with predictive mortality rates of 25.0 and 29.0%, respectively, in patients used for the formation and those used for the validation of the algorithms. Thus, the clinical features of transplanted patients should, in the future, be evaluated retrospectively with reference to peripheral blood platelet counts and the etiology of liver injury, as well as serum bilirubin concentration, the items responsible for classification as category-1G. Also, it should be noted that 16 of 112 patients (14.3%) were classified as category-2C at 5 days after the onset of hepatic encephalopathy. The significance of category-2C, characterized by items such as cerebral edema, liver atrophy, and cardiac failure, should be investigated further.

In Europe and the United State, the indications for liver transplantation in patients with ALF have been evaluated based on the guidelines proposed by O'Grady et al. [28], in which the prognosis was estimated differently in patients with liver failure due to acetaminophen intoxication and those with liver failure caused by viral hepatitis and drug allergy-induced liver injury. In the former category of patients, the prognosis was estimated based on three parameters: arterial blood pH, peak prothrombin time, and the serum creatinine level. In contrast, in the latter category of patients, the prognosis was determined based on 5 parameters: etiology of the disease, age of the patient, the duration of jaundice before the onset of hepatic encephalopathy, peak prothrombin time, and the serum bilirubin level. Thus, the usefulness of our novel model based on the decision tree analysis should also be evaluated in ALF patients in Europe and the United States, especially in those with acute liver failure due to viral hepatitis and drug allergy-induced liver injury, in comparison with the guidelines proposed by O'Grady et al. [28]. However, it should be kept in mind that the purpose of our model is to predict the possible mortality rates of ALF patients, but not to determine the indication for liver transplantation automatically. In our model, cerebral edema and cardiac failure, which may disallow the patients from receiving liver transplantation, are included as split items. Liver transplantation cannot be performed for patients showing high mortality rates due to complications caused by ALF that correspond to items that are contra-indications for surgical procedures.

In conclusion, we have developed a novel model consisting of two algorithms for predicting the outcome of ALF patients at the onset of hepatic encephalopathy and at 5 days later, through decision tree analysis. This system may be useful to determine the indication for liver transplantation, because the mortality rates can be estimated by the algorithms with high accuracy rates, which were similarly high both before and after validation.

Acknowledgments This study was supported by a Health Labor Sciences Research Grant, Research on Measures for Intractable Diseases, from the Ministry of Health, Labor and Welfare of Japan.

Conflict of interest Hirohito Tsubouchi received research grants and lecture fees from Chugai Pharmaceutical Co., Ltd., and MSD. Satoshi Mochida received research grants from Chugai Pharmaceutical Co., Ltd., MSD, and Toray Industries, Inc., and lecture fees from MSD.

References

1. Mochida S, Takikawa Y, Nakayama N, Oketani M, Naiki T, Yamagishi Y, et al. Diagnostic criteria of acute liver failure: a report by the Intractable Hepato-Biliary Diseases Study Group of Japan. Hepatol Res. 2011;41:805–12.
2. Inuyama Symposium Kiroku Kanko-Kai. The proceedings of the 12th Inuyama symposium. Hepatitis type A and fulminant hepatitis. Chugai Igaku-sha, Tokyo. 1982 (in Japanese).
3. Mochida S, Fujiwara K. Symposium on clinical aspects in hepatitis virus infection. 2. Recent advances in acute and fulminant hepatitis in Japan. Intern Med. 2001;40:175–7.
4. Fujiwara K, Mochida S, Matsui A, Nakayama N, Nagoshi S, Toda G, et al. Fulminant hepatitis and late onset hepatic failure in Japan: Summary of 698 patients between 1998 and 2003 analyzed in annual nationwide survey. Hepatol Res. 2008;38:646–57.
5. Sugihara J, Naito T, Ishiki Y, Murakami N, Naiki T, Koshino Y, *et al.* A multicenter study on the prognosis and indication of liver transplantation for fulminant hepatitis in Japan: details of decision of the guideline for liver transplantation in Japanese Acute Hepatic Failure Study Group (1996). Acta Hepatol Japonica. 2001;42:543–57 (in Japanese).
6. Mochida S, Nakayama N, Matsui A, Nagoshi S, Fujiwara K. Re-evaluation of the Guideline published by the Acute Liver Failure Study Group of Japan in 1996 to determine the indications of liver transplantation in patients with fulminant hepatitis. Hepatol Res. 2008;38:970–9.
7. Nakayama N, Oketani M, Kawamura Y, Inao M, Nagoshi S, Fujiwara K, et al. Novel classification of acute liver failure

through clustering using a self-organizing map: usefulness for prediction of the outcome. J Gastroenterol. 2011;46:1127–35.
8. Kohonen T. Self-organizing maps. Berlin: Springer; 2001.
9. Talbi ML, Charef A. PVC discrimination using the QRS power spectrum and self-organizing maps. Comput Methods Programs Biomed. 2009;94:223–31.
10. Basara HG, Yuan M. Community health assessment using self-organizing maps and geographic information systems. Int J Health Geogr. 2008;30(7):67.
11. Tsunedomi R, Iizuka N, Hamamoto Y, Uchimura S, Miyamoto T, Tamesa T, et al. Patterns of expression of cytochrome P450 genes in progression of hepatitis C virus-associated hepatocellular carcinoma. Int J Oncol. 2005;27:661–7.
12. Haydon GH, Hiltunen Y, Lucey MR, Collett D, Gunson B, Murphy N, et al. Self-organizing maps can determine outcome and match recipients and donors at orthotopic liver transplantation. Transplantation. 2005;79:213–8.
13. Omori K, Terai S, Ishikawa T, Aoyama K, Sakaida I, Nishina H, et al. Molecular signature associated with plasticity of bone marrow cell under persistent liver damage by self-organizing-map-based gene expression. FEBS Lett. 2004;578:10–20.
14. Gebbinck MS, Verhoeven JT, Thijssen JM, Schouten TE. Application of neural networks for the classification of diffuse liver disease by quantitative echography. Ultrason Imaging. 1993;15:205–17.
15. Takasaki S, Kawamura Y, Konagaya A. Selecting effective siRNA sequences based on the self-organizing map and statistical techniques. Comput Biol Chem. 2006;30:169–78.
16. Gando S, Iba T, Eguchi Y, Ohtomo Y, Okamoto K, Koseki K, et al. A multicenter, prospective validation of disseminated intravascular coagulation diagnostic criteria for critically ill patients: comparing current criteria. Crit Care Med. 2006;34:625–31.
17. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform. 2008;77:81–97.
18. Luk JM, Lam BY, Lee NP, Ho DW, Sham PC, Chen L, et al. Artificial neural networks and decision tree model analysis of liver cancer proteomes. Biochem Biophys Res Commun. 2007;361:68–73.
19. Kawaguchi T, Kakuma T, Yatsuhashi H, Watanabe H, Saitsu H, Nakao K, et al. Data mining reveals complex interactions of risk factors and clinical feature profiling associated with the staging of non-hepatitis B virus/non-hepatitis C virus-related hepatocellular carcinoma. Hepatol Res. 2011;41:564–71.
20. Kurosaki M, Hiramatsu N, Sakamoto M, Suzuki Y, Iwasaki M, Tamori A, 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 (Epub ahead of print).
21. Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, Hiramatsu N, et al. Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis. J Gastroenterol. 2011;46:401–9.
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;83:445–52.
23. Yoshiba M, Inoue K, Sekiyama K, Koh I. Favorable effect of new artificial liver support on survival of patients with fulminant hepatic failure. Artif Organs. 1996;20:1169–72.
24. Sekido H, Matsuo K, Takeda K, Ueda M, Morioka D, Kubota T, et al. Usefulness of artificial liver support for pretransplant patients with fulminant hepatic failure. Transplant Proc. 2004;36:2355–6.
25. Inoue K, Watanabe T, Maruoka N, Kuroki Y, Takahashi H, Yoshiba M. Japanese-style intensive medical care improves prognosis for acute liver failure and the perioperative management of liver transplantation. Transplant Proc. 2010;42:4109–12.
26. Kubota T, Sekido H, Takeda K, Morioka D, Tanaka K, Endo I, et al. Acute hepatic failure with deep hepatic coma treated successfully by high-flow continuous hemodiafiltration and living-donor liver transplantation: a case report. Transplant Proc. 2003;35:394–6.
27. Inoue K, Kourin A, Watanabe T, Yamada M, Yasuda H, Yoshiba M. Plasma exchange in combination with online-hemodiafiltration as a promising method for purifying the blood of fulminant hepatitis patients. Hepatol Res. 2008;38:S46–51.
28. O'Grady JG, Alexander GJ, Hayllar KM, Williams R. Early indicators of prognosis in fulminant hepatic failure. Gastroenterology. 1989;97:439–45.

Research Article

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

A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region

Yuji Urabe[1,2], Hidenori Ochi[2], Naoya Kato[4], Vinod Kumar[1,3], Atsushi Takahashi[3], Ryosuke Muroyama[4], Naoya Hosono[3], Motoyuki Otsuka[5], Ryosuke Tateishi[5], Paulisally Hau Yi Lo[1], Chizu Tanikawa[1], Masao Omata[5], Kazuhiko Koike[5], Daiki Miki[2], Hiromi Abe[2], Naoyuki Kamatani[3], Joji Toyota[6], Hiromitsu Kumada[7], Michiaki Kubo[3], Kazuaki Chayama[2], Yusuke Nakamura[1,3], Koichi Matsuda[1,*]

[1]*Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan;* [2]*Department of Medical and Molecular Science, Division of Frontier Medical Science, Programs for Biomedical Research, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima, Japan;* [3]*Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan;* [4]*Unit of Disease Control Genome Medicine, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan;* [5]*Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan;* [6]*Department of Gastroenterology, Sapporo Kosei General Hospital, Hokkaido, Japan;* [7]*Department of Hepatology, Toranomon Hospital, Tokyo, Japan*

Background & Aims: We performed a genome-wide association study (GWAS) of hepatitis C virus (HCV)-induced liver cirrhosis (LC) to identify predictive biomarkers for the risk of LC in patients with chronic hepatitis C (CHC).

Methods: A total of 682 HCV-induced LC cases and 1045 CHC patients of Japanese origin were genotyped by Illumina Human Hap 610-Quad bead Chip.

Results: Eight SNPs which showed possible associations ($p < 1.0 \times 10^{-5}$) at the GWAS stage were further genotyped using 936 LC cases and 3809 CHC patients. We found that two SNPs within the major histocompatibility complex (MHC) region on chromosome 6p21, rs910049 and rs3135363, were significantly associated with the progression from CHC to LC ($p_{combined} = 9.15 \times 10^{-11}$ and 1.45×10^{-10}, odds ratio (OR) = 1.46 and 1.37, respectively). We also found that *HLA-DQA1*0601* and *HLA-DRB1*0405* were associated with the progression from CHC to LC ($p = 4.53 \times 10^{-4}$ and 1.54×10^{-4} with OR = 2.80 and 1.45, respectively). Multiple logistic regression analysis revealed that rs3135363, rs910049, and *HLA-DQA1*0601* were independently associated with the risk of HCV-induced LC. In addition, individuals with four or more risk alleles for these three loci have a 2.83-fold higher risk for LC than those with no risk allele, indicating the cumulative effects of these variations.

Conclusions: Our findings elucidated the crucial roles of multiple genetic variations within the MHC region as prognostic/predictive biomarkers for CHC patients.

Introduction

Two million people in Japan and 210 million people worldwide are estimated to be infected with the hepatitis C virus (HCV), which is known to be a major cause of chronic viral liver disease [1]. Patients with chronic hepatitis C (CHC) usually exhibit mild inflammatory symptoms, but are at a significantly high risk for developing liver cirrhosis (LC) and hepatocellular carcinoma [2]. More than 400,000 people at present suffer from LC, which is ranked as the 9th major cause of death in Japan. In addition, liver cancer causes approximately 32,000 deaths per year, making it the 4th most common cause of death from malignant diseases. Thus, HCV-related diseases are important public health problems [3].

Clinical outcomes after the exposure to HCV vary enormously among individuals. Approximately 70% of infected persons will develop chronic hepatitis [4], and about 20–30% of CHC patients will develop cirrhosis, but others can remain asymptomatic for decades [2]. The annual death rate of patients with decompensated cirrhosis is as high as 15–30% [5]. Moreover, more than 7% of LC patients develop hepatocellular cancer in Japan and Taiwan, while the frequencies are less than 1.6% among other ethnic groups [6,7]. These inter-individual and inter-ethnic differences have been attributed to various factors such as viral genotypes,

Keywords: Genome-wide association study; Hepatitis C virus; Liver cirrhosis; Major histocompatibility complex.
Received 22 April 2012; received in revised form 15 December 2012; accepted 24 December 2012

* Corresponding author. Address: Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato, Tokyo 108-8639, Japan. Tel.: +81 3 5449 5376; fax: +81 3 5449 5123.
E-mail address: koichima@ims.u-tokyo.ac.jp (K. Matsuda).

Abbreviations: CHC, chronic hepatitis C; GWAS, genome-wide association study; HCV, hepatitis C virus; LC, liver cirrhosis; MHC, major histocompatibility complex; OR, odds ratio; PBC, primary biliary cirrhosis; SNPs, single nucleotide polymorphisms.

Please cite this article in press as: Urabe Y et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. J Hepatol (2013), http://dx.doi.org/10.1016/j.jhep.2012.12.024

Research Article

Table 1. Characteristics of samples and methods used in this study.

Stage	Source	Platform	Number of samples	Female (%)	Age, yr (mean ± SD)
GWAS					
Liver cirrhosis	BioBank Japan	Illumina Human Hap 610	682	313 (46.3)	67.1 ± 9.7
Chronic hepatitis C[a]	Hiroshima University	Illumina Human Hap 610	1045	371 (35.5)	55.2 ± 11.0
Replication					
Liver cirrhosis	Tokyo University	Invader assay	716	334 (46.8)	64.4 ± 10.4
	Hiroshima University		220	98 (44.5)	64.7 ± 8.98
Chronic hepatitis C[a]	BioBank Japan	Invader assay	1670	780 (46.8)	59.7 ± 12.6
	Hiroshima University		2139	1061 (51.8)	58.8 ± 9.20

[a] Number of samples that qualified. CHC patients with severe liver fibrosis (F3 or F4) or lower platelet counts (<160,000) were excluded.

alcohol consumption, age at infection, co-infection of HIV or HBV [8–10], insulin resistance, steatosis, and metabolic syndrome [11]. Previous gene expression analyses also identified various genes associated with liver fibrosis among patients with CHC [12–14]. In addition, miRNAs such as mir-21 and mir-122 were shown to be correlated with live fibrosis [15,16].

Currently, the genome-wide association study is the most common method to identify genetic variations associated with disease risk [17–20]. In addition, the roles of genetic factors in HCV-related diseases have been elucidated. *IL28B* is associated with spontaneous clearance of HCV [21] as well as with the clinical response to the combination therapy of pegylated interferon and ribavirin [22,23]. Recently, our group has shown that SNP rs2596542 on *MICA* [24] and SNP rs1012068 on *DEPDC5* [25] are significantly associated with HCV-induced liver cancer. Although liver cirrhosis is the major risk factor of liver cancer, a fraction of CHC patients will develop HCC without accompanying LC. Therefore, the underlying genetic background would be different between HCV-induced LC and HCV-induced HCC. Previous studies identified the association of genetic variants in *HLA-DQ/DR/B* [26–28], *2–5AS* [29], *TLR3* [30], and *PNPLA3* [31] with the risk of liver fibrosis among patients with CHC. However, a comprehensive approach for HCV-induced LC has not been conducted so far. Here we performed GWAS of HCV-induced LC to identify predictive biomarkers for the risk of LC in patients with CHC.

Materials and methods

Ethics statement

All subjects provided written informed consent. This project was approved by the ethical committees at University of Tokyo, Hiroshima University, Sapporo Kosei General Hospital, Toranomon Hospital, and Center for Genomic Medicine, Institutes of Physical and Chemical Research (RIKEN).

Study population

The characteristics of each cohort are shown in Table 1. In this study, we conducted GWAS and replication analysis on a total of 1618 HCV-induced LC and 4854 CHC patients. All subjects had abnormal levels of serum alanine transaminase for more than 6 months and were positive for both HCV antibody and serum HCV RNA. Among 1618 LC and 4854 CHC samples, 342 LC patients (21.14%) and 2997 CHC patients (61.70%) underwent liver biopsy. The remaining 1276 LC and 1857 CHC patients were diagnosed by non-invasive methods including hepatic imaging (e.g., ultrasonography, computed tomography, arteriography or magnetic resonance imaging), biochemical data (serum bilirubin, serum albumin, platelet, or prothrombin time), and the presence/absence of clinical manifestations of portal hypertension (e.g., varices, encephalopathy or ascites). The patients with CHC or LC were recruited for this study regardless of their treatment history. We excluded from the analysis the followings CHC patients: (1) advanced liver fibrosis (F3 or F4 by New Inuyama classification) [32], (2) platelet count under 160,000 for patients without liver fibrosis staging, and (3) HBV co-infection. Characteristics of each study cohort are shown in Table 1. In brief, DNA of HCV-induced LC and CHC patients was obtained from Biobank Japan (http://biobankjp.org/) [33], the Hiroshima Liver Study Group (http://home.hiroshima-u.ac.jp/naika1/researchprofile/pdf/liverstudygroupe.pdf), Toranomon Hospital, and the University of Tokyo. All subjects were of Japanese origin.

SNP genotyping

Genomic DNA was extracted from peripheral blood leukocytes using a standard method. In GWAS, we genotyped 682 LC and 1045 CHC samples using Illumina Human Hap 610-Quad bead Chip (Supplementary Fig. 1). Samples with low call rate (<0.98) were excluded from our analysis (six LC and two CHC samples). We then applied SNP quality control as follows: call rate $\geqslant 0.99$ in LC and CHC samples, Hardy–Weinberg $p \geqslant 1 \times 10^{-6}$ in LC and CHC samples. Consequently, 461,992 SNPs on the autosomal chromosomes passed the quality control filters. SNPs with minor allele frequency of <0.01 in both LC and CHC samples were excluded from further analyses, considering statistical power in the replication analysis. Finally, we analyzed 431,618 SNPs in GWAS. Among the top ten SNPs showing $p < 1.0 \times 10^{-5}$, we selected nine SNPs for further analysis with LD threshold of $r^2 = 0.95$. In the replication stage, we genotyped 936 LC and 3809 CHC using multiplex PCR-based Invader assay (Third Wave Technologies).

Statistical analysis

The association of SNPs with the phenotype in the GWAS, replication stage, and combined analyses was tested by logistic regression analysis, upon adjusting for age at recruitment (continuous) and gender, by assuming additive model using PLINK [34]. In the GWAS, the genetic inflation factor λ was derived by applying logistic regressed p values for all the tested SNPs. The quantile–quantile plot was drawn using R program. The odds ratios were calculated using the non-susceptible allele as reference, unless stated otherwise. The combined analysis of GWAS and replication stage was verified by using the Mantel–Haenszel method. We set the significance threshold as follows; $p = 1 \times 10^{-5}$ in the GWAS stage (first stage) and $p = 6.25 \times 10^{-3}$ (=0.05/8) in the replication analysis. We considered $p < 5 \times 10^{-8}$ as threshold of GWAS significance in the combined analysis, which is the Bonferroni-corrected threshold for the number of independent SNPs genotyped in HapMap Phase II [35]. The heterogeneity across two stages was examined by using the Breslow–Day test [36]. We used Haploview software to analyze the association of haplotypes and LD values between SNPs. Quality control for SNPs was applied as follows: call rate $\geqslant 0.95$ in LC and CHC samples, and Hardy–Weinberg $p \geqslant 1 \times 10^{-6}$ in CHC samples in replication stage. The statistical power was 19.51% in GWAS (the first stage) ($p = 1.00 \times 10^{-5}$), 97.98% in replication ($p = 0.05/8$), and 74.76% in the combined stage ($p = 5.00 \times 10^{-8}$) at minor allele frequency of 0.3 and OR of 1.3.

Imputation-based association analysis of HLA class I and class II alleles

We obtained an SNP or a combination of SNPs which could tag the HLA alleles in the Japanese population from a previous study [37]. Genotypes of tagging SNPs were imputed in the GWAS samples by using a Hidden Markov model programmed in MACH [38] and haplotype information from HapMap JPT samples

Please cite this article in press as: Urabe Y et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. J Hepatol (2013), http://dx.doi.org/10.1016/j.jhep.2012.12.024

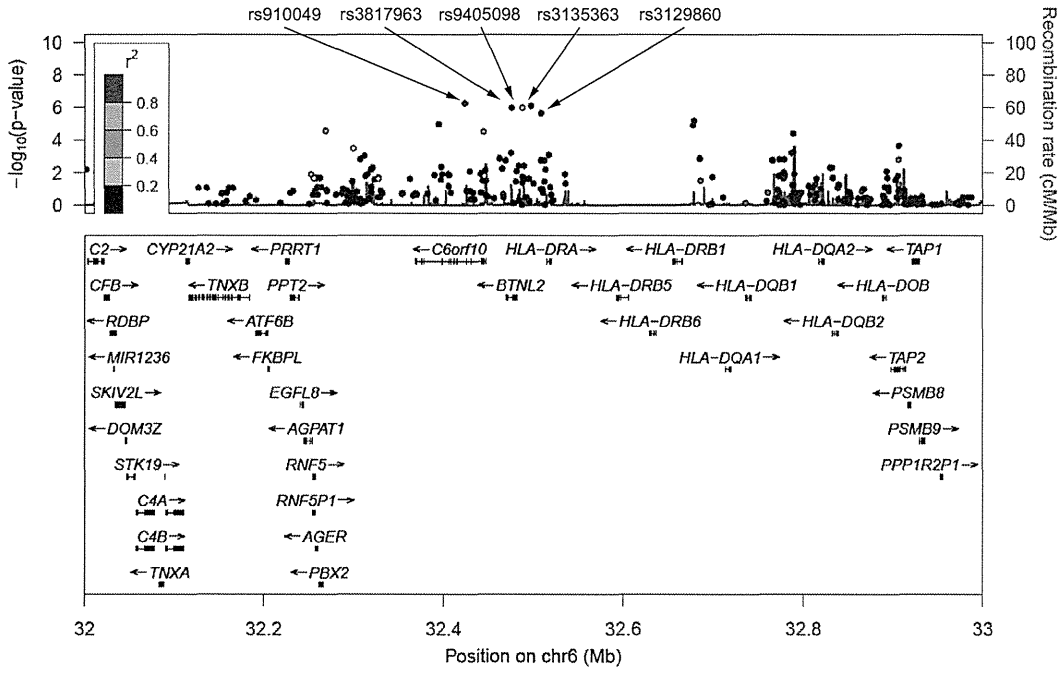

Fig. 1. Regional association plot at 6p21.3. (Upper panel) *p* Values of genotyped SNPs (circle) and imputed SNPs (cross) are plotted (as $-\log_{10} p$ value) against their physical position on chromosome 6 (NCBI Build 36). The *p* value for rs910049 at GWAS is represented by a purple diamond. Estimated recombination rates from HapMap JPT show the local LD structure. Inset; the color of the other SNPs indicates LD with rs3135363 according to a scale from $r^2 = 0$ to $r^2 = 1$ based on pair-wise r^2 values from HapMap JPT. (Lower panel) Gene annotations from the University of California Santa Cruz genome browser.

and 1000 genome imputation samples [39]. We applied the same SNP quality criteria as in GWAS, to select SNPs for the analysis. We employed the logistic regression analysis upon age and gender adjustment to assess the associations between HCV-induced LC and HLA alleles.

Software

For general statistical analysis, we employed R statistical environment version 2.9.1 (cran.r-project.org) or plink-1.06 (pngu.mgh.harvard.edu/~purcell/plink/). The Haploview software version 4.2 [40] was used to calculate LD and to draw Manhattan plot. Primer3 -web v0.3.0 (http://frodo.wi.mit.edu) web tool was used to design primers. We employed LocusZoom (http://csg.sph.umich.edu/locuszoom/) for regional plots. We used SNP Functional Prediction web tool for functional annotation of SNPs (http://snpinfo.niehs.nih.gov/snpfunc.htm) [41]. We used "Gene Expression Analysis Based on Imputed Genotypes" (http://www.sph.umich.edu/csg/liang/imputation) [42] for eQTL analysis. We used MACTH [43] web tool for searching potential binding sites for transcription factors (http://www.gene-regulation.com/index.htm).

Results

Genome-wide association study for HCV-induced liver cirrhosis

We performed a two-stage GWAS using a total of 1618 cases and 4854 controls (Supplementary Fig. 1). In the first stage, a whole genome scan was performed on 682 Japanese patients with HCV-induced LC and 1045 Japanese patients with CHC, using Illumina Human Hap 610-Quad bead Chip. The genotyping results of 431,618 single nucleotide polymorphisms (SNPs) obtained after our standard quality control were used for further analysis. CHC patients with severe liver fibrosis (F3 or F4 according to the New Inuyama classification [32]) or lower platelet counts (<160,000) were excluded from the control group. As progression from CHC to LC is strongly affected by age and gender, we performed logistic regression analyses including age and gender as covariates at all tested loci in our analyses. The genetic inflation factor lambda was 1.051, indicating that there is little or no evidence of population stratification (Supplementary Fig. 2A). Although no SNPs cleared the GWAS significance threshold ($p < 5 \times 10^{-8}$) at the first stage, we selected ten candidate SNPs showing suggestive association of $p < 1 \times 10^{-5}$ (Supplementary Fig. 2B and Supplementary Table 1). After excluding SNP rs6891116 due to almost absolute linkage with SNP rs10252674 ($r^2 = 0.99$), the remaining nine SNPs were further genotyped using an independent cohort, consisting of 936 LC and 3809 CHC cases, by multiplex PCR-based Invader assay as the second stage. We could successfully obtain genotype results for eight SNPs after the QC filter (call rate $\geqslant 0.95$ in LC and CHC samples, Hardy–Weinberg of $p \geqslant 1 \times 10^{-6}$ in CHC samples). The logistic regression analysis adjusted by age and gender revealed that five SNPs on chromosome 6q21.3 indicated a significant association with progression from CHC to LC after the Bonferroni correction ($p < 0.05/8 = 6.25 \times 10^{-3}$, Supplementary Table 2). A meta-analysis of the two stages with a fixed-effects model revealed that all of the five SNPs significantly associated with progression from CHC to LC (*p* values of 9.15×10^{-11}–1.28×10^{-8} with odds ratios (OR) of 1.30–1.46, Fig. 1 and Table 2). These five SNPs were located in the HLA class II region and were in strong linkage disequilibrium with each other (D' >0.75, Sup-

Please cite this article in press as: Urabe Y et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. J Hepatol (2013), http://dx.doi.org/10.1016/j.jhep.2012.12.024

Research Article

Table 2. Summary of GWAS and replication analyses.

SNP	Stage	Allele (1/2)	Gene	Liver cirrhosis				Chronic hepatitis C				OR (95% CI)[b]	*p* value[c]	*p* value$_{het}$[d]
				11	12	22	RAF[a]	11	12	22	RAF[a]			
rs910049														
	GWAS	a/g	*C6orf10* (6p21.3)	24	217	435	0.196	25	224	794	0.131	1.73 (1.40-2.15)	5.39×10^{-7}	
	Replication			38	259	631	0.180	66	952	2790	0.142	1.37 (1.20-1.58)	7.59×10^{-6}	
	Combined[e]											1.46 (1.28-1.62)	9.15×10^{-11}	0.075
rs3817963														
	GWAS	a/g	*BTNL2* (6p21.3)	92	343	241	0.390	101	437	505	0.306	1.53 (1.29-1.81)	9.50×10^{-7}	
	Replication			130	395	395	0.356	409	1573	1816	0.315	1.22 (1.10-1.36)	2.66×10^{-4}	
	Combined[e]											1.30 (1.18-1.42)	1.28×10^{-8}	0.029
rs9405098														
	GWAS	a/g	No gene (6p21.3)	75	293	308	0.328	70	365	608	0.242	1.54 (1.30-1.84)	1.10×10^{-6}	
	Replication			100	361	462	0.304	249	1429	2129	0.253	1.30 (1.16-1.46)	5.64×10^{-6}	
	Combined[e]											1.37 (1.23-1.50)	1.04×10^{-10}	0.105
rs3135363														
	GWAS	c/t	No gene (6p21.3)	35	258	383	0.757	89	447	507	0.700	1.58 (1.32-1.90)	7.89×10^{-7}	
	Replication			73	322	540	0.750	389	1486	1929	0.702	1.30 (1.16-1.46)	7.94×10^{-6}	
	Combined[e]											1.37 (1.24-1.51)	1.45×10^{-10}	0.069
rs3129860														
	GWAS	a/g	No gene (6p21.3)	58	294	324	0.303	57	348	638	0.221	1.55 (1.29-1.82)	6.45×10^{-6}	
	Replication			88	339	507	0.276	208	1341	2246	0.231	1.28 (1.14-1.44)	2.53×10^{-5}	
	Combined[e]											1.36 (1.22-1.49)	1.07×10^{-9}	0.085

1618 (682 in GWAS and 936 in replication) liver cirrhosis and 4854 (1045 in GWAS and 3809 in replication) chronic hepatitis C samples were analyzed.
[a]RAF, risk allele frequency.
[b]OR, odds ratios; CI, confidence interval.
[c]*p* Values obtained by logistic regression analysis adjusted for age and gender under additive model.
[d]*p* Values of heterogeneities (Phet) across three stages were examined by using the Breslow–Day test.
[e]Combined odds ratio and *p* values for independence test were calculated by Mendel-hauzen and Laird method in the meta-analysis.

plementary Fig. 3). To further evaluate the effect of each variation on the progression from CHC to LC, we performed multiple logistic regression analyses. As a result, rs910049 (p of 1.91×10^{-3} with OR of 1.25) and rs3135363 (p of 1.49×10^{-4} with OR of 1.23) remained significantly associated with the progression risk from CHC to LC, while the remaining three SNPs failed to show significant associations ($p > 0.05$) (Supplementary Table 3). Thus, two SNPs, rs910049 and rs3135363, seem to be independent risk factors for HCV-induced LC.

Since reduced platelet level is associated with a poor prognosis among CHC patients [44] we excluded patients with platelet level of less than 160,000 from CHC groups to increase the risk of type 2 error in this study. We also conducted the analysis using only CHC patients diagnosed with liver biopsy. As a result, both SNPs reached genome-wide significance ($p < 5 \times 10^{-8}$), although the associations were reduced due to the smaller sample size (Supplementary Table 4).

Subgroup analyses, stratified by IFN treatment status, amount of alcohol consumption, and gender, were also performed, since these factors were shown to be associated with the prognosis of CHC patients [45–47]. A total of 334 LC patients (35.83%) and 2325 CHC (82.4%) were treated with IFN therapy. Although the frequency of IFN treatment was different between CHC and LC groups, these variations associated with the LC risk regardless of IFN treatment as well as gender and alcohol consumption (Supplementary Fig. 4A–C). When we included these factors as covariates, the association of these variations with HCV-induced LC was sustained, with OR of 1.48 and 1.56, and SNP rs3135363 still reached genome-wide significance ($p = 3.95 \times 10^{-9}$) (Supplementary Table 5).

The association of previously reported variations with HCV-induced LC

Non-synonymous SNP rs738409 (I148M) in the *PNPLA3* gene was shown to be associated with progression of LC in the previous prospective study in Caucasians [31]. SNP rs738409 was also associated with the severity of non-alcoholic fatty liver disease in Japanese [48]. Therefore, we analyzed SNP rs738409 in our case-control cohort, but rs738409 did not significantly associate with HCV-induced LC ($p = 0.24$ and OR = 1.10), although the risk G allele was more frequent among LC than CHC (Supplementary Table 6). Our result is similar to what observed among Caucasians in the previous study, in which rs738409 increased liver cancer risk among alcoholic cirrhosis but did not among hepatitis C cirrhosis [49]. Since biological studies demonstrated that its risk allele (G) abolishes the triglyceride hydrolysis activity of *PNPLA3* [50] *PNPLA3* variation would have a strong impact on non-viral cirrhosis.

Recently, GWAS in the Caucasian population identified the association of SNPs rs4374383, rs16851720 and rs9380516 with the progression of liver fibrosis after HCV infection [51]. However, SNPs rs4374383 and rs16851720 did not exhibit significant association ($p = 0.654$ and 0.231, respectively) in our sample set. Although SNP rs9380516 exhibited the association with p-value of 0.015, the risk allele showed an opposite result

Please cite this article in press as: Urabe Y et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. J Hepatol (2013), http://dx.doi.org/10.1016/j.jhep.2012.12.024

Table 3. Results of three associated variations from candidate gene analyses.

Gene	Tagging SNP	Haplotype frequency		OR (95% CI)[a]		*p* value[b]
		Liver cirrhosis	Chronic hepatitis C			
DQA1**0601*	rs2736182(T) + rs2071293(A)	0.038	0.019	2.80	1.38-3.32	4.53×10^{-4}
DRB1**0405*	rs411326(C) + rs2395185(A) + rs4599680(A)	0.324	0.266	1.45	1.15-1.56	1.54×10^{-4}

Association was tested by comparing haplotype distribution between 682 liver cirrhosis and 1045 chronic hepatitis C samples in GWAS.
[a] OR, odds ratio; CI, confidence interval.
[b] *p* Values were obtained by case-control analysis of GWAS stage (*p* for haplotype were obtained by score test, implemented in R) (*DQA1*0601* and *DRB1*0405*). The *p* values obtained by logistic regression analysis adjusted for age and gender under additive model.

(Supplementary Table 6). Taken together, these SNPs would not be associated with liver fibrosis in the Japanese population.

Genes related to extracellular matrix turnover or immune response (*KRT 19*, *COL1A1*, *STMN2*, *CXCL6*, *CCR2*, *TIMP1*, *IL8*, *IL1A*, *ITGA2*, *CLDN 4*, and *IL2*) were shown to be implicated in liver fibrosis of chronic hepatitis C [14]. To further characterize these loci, we conducted imputation analyses in the GWAS sample set (682 cases and 1045 controls), using data from HAPMAP phase II (JPT), and found 163 SNPs in 9 loci. However, none of these SNPs indicated significant association with *p*-value of less than 0.01 (Supplementary Table 7). Thus, variations of these genes did not associate with progression from chronic hepatitis C to liver cirrhosis.

Imputation-based fine mapping of HLA region

The most significantly associated SNP rs3135363 is located within an intergenic region between *BTNL2* and *HLA-DRA*, and rs910049 is located in intron 7 of *C6orf10* gene (Supplementary Figs. 5 and 6). To further characterize these loci, we conducted imputation-based association analysis for the GWAS samples (682 LC and 1045 CHC samples) using data from HAPMAP Phase II (JPT), and could obtain the results of nearly 6000 SNPs in a 4-Mb genomic region. The regional association plots revealed that all modestly-associated SNPs are confined within a 700-kb region containing 21 genes, namely *TNXB*, *ATF6B*, *FKBPL*, *PRRT1*, *PPT2*, *EGFL8*, *AGPAT1*, *RNF5*, *RNF5P1*, *AGER*, *PBX2*, *C6orf10*, *BTNL2*, *HLA-DRA*, *HLA-DRB5*, *HLA-DRB6*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQA2*, *HLA-DQB1* and *HLA-DQB2* (Supplementary Fig. 5). Although 640 SNPs, including ten non-synonymous SNPs within the 4-Mb region, showed very modest associations ($p < 0.01$) with HCV-induced LC, none of these SNPs in this region revealed strong association with HCV-induced LC, after adjustment with the two SNPs, rs910049 and rs3135363 (Supplementary Fig. 7). Taken together, the associations observed in this region would reflect the association with rs910049 and rs3135363.

Previous reports indicated the association of *HLA-DRB1* and *HLA-DQ* alleles with HCV-induced chronic hepatitis in the Japanese population [26]. To investigate the association of HLA alleles with HCV-induced LC, we estimated the genotypes at the HLA region by applying the imputation results of HLA-tagging SNPs [37]. We could successfully determine 53 alleles of *HLA-A*, *B*, *C*, *DQA*, *DQB*, and *DRB* genes and find that *HLA-DQA1*0601* and *HLA-DRB1*0405* were strongly associated with HCV-induced LC (p values of 4.53×10^{-4} and 1.54×10^{-4} with ORs of 2.80 and 1.45) even after the Bonferroni correction ($p < 0.05/53 = 9.43 \times 10^{-4}$) (Table 3 and Supplementary Table 8A–E) [37].

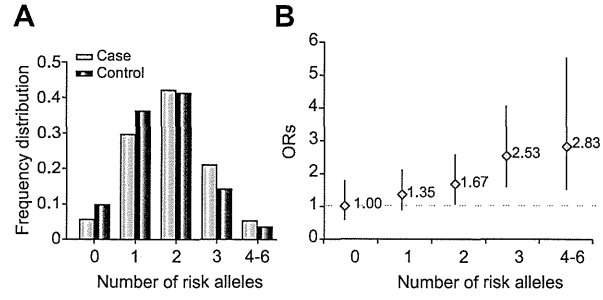

Fig. 2. Cumulative effects of liver cirrhosis risk alleles. (A) Frequency distribution divided by risk allele numbers (rs910049, rs3135353, and *HLA-DQA0601*) among liver cirrhosis (light blue bars) and chronic hepatitis C (dark blue bars) patients. (B) Plot of the increase odds ratio (OR) for liver cirrhosis according to the number of risk alleles. The ORs are relative to the subjects with no risk alleles (rs910049, rs3135353, and *HLA-DQA0601*). Vertical bars correspond to 95% confidence intervals. Horizontal line marks the null value (OR = 1).

Cumulative effect of multiple loci within the HLA region

SNPs rs3135363 and rs910049, *HLA-DQA1*0601*, and *HLA-DRB1*0405* are located within a 300-kb segment in the HLA class II region and show moderate linkage disequilibrium (Supplementary Fig. 8). To further evaluate these genetic factors, we performed multiple logistic regression analyses and found that rs910049 (p of 9.40×10^{-3} with OR of 1.38), rs3135363 (p of 3.94×10^{-4} with OR = 1.41), and *HLA-DQA1*0601* (p of 7.79×10^{-3} with OR of 1.54) were significantly associated with HCV-induced LC (Supplementary Table 9), indicating these three variations were independent risk factors for progression of CHC to LC.

To investigate the pathophysiological roles of rs910049 and rs3135363 in disease progression, we searched the eQTL database (http://www.sph.umich.edu/csg/liang/imputation) and found that risk alleles of rs910049 (A) and rs3135363 (T) were associated with lower expression of *HLA-DQA* (LOD of $\geqslant$6.86 and 17.31, respectively) and *DRB1* (LOD of $\geqslant$12.01 and 18.96, respectively), and with higher expression of *HLA-DQB1* (LOD of $\geqslant$6.76 and 4.46, respectively) (Supplementary Table 10). Thus, rs910049 and rs3135363 are likely to affect the expression of HLA class II molecules and subsequently alter the risk of HCV-induced LC.

Finally, we examined the cumulative effects of rs910049, rs3135363, and *HLA-DQA1*0601*. Individuals with four or more risk alleles (8.8% of general population) have 2.83-fold higher risk of HCV-induced LC compared with those with no risk allele (15.0% of general population, Fig. 2).

Please cite this article in press as: Urabe Y et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. J Hepatol (2013), http://dx.doi.org/10.1016/j.jhep.2012.12.024

Research Article

Discussion

We here demonstrated that multiple genetic variations in the MHC region were significantly associated with the risk of disease progression from CHC to LC, using a total of 1618 HCV-LC and 4854 CHC cases. Since a substantial proportion of patients with CHC show progression to LC in a certain time period, exclusion of CHC patients who have a high risk for LC from control subjects is essential to reduce the risk of false negative association. In this study, CHC patients with advanced fibrosis (F3 or F4 in stage) or with reduced platelet level (less than 160,000/μl) were excluded from the control samples, since these alterations are well-known risk factors for LC development [9,32]. Consequently, we were successfully able to identify the HCV-induced LC loci.

HLA genes are known to play critical roles in the regulation of our immune responses through controlling the antigen presentation to CD8 (class I) and CD4 (class II) T cells. Although previous studies indicated the association of HLA class I alleles such as *HLA-B57*, *HLA-A11*, and *HLA-C04* with persistent HCV infection [52,53], no SNPs in the HLA class I region exhibited strong association with HCV-induced LC. Here we identified three variations (rs910049, rs3135363, *HLA-DQA1*0601*) in the HLA class II region to be significantly associated with the progression risk from CHC to LC. Since two SNPs, rs910049 and rs3135363, had been indicated to affect expression levels of *HLA-DRB1* and *DQ*, our findings indicated the significant pathophysiological roles of HLA class II molecules in the development of HCV-induced liver fibrosis. Considering the function of *HLA-DQ* and *HLA-DR*, we suggest that the antigen presentation by HLA class II molecules is likely to play a critical role in the elimination of HCV-infected liver cells and subsequently prevent HCV-induced LC.

Direct acting antiviral drugs for HCV can cure up to 75% of patients infected with HCV genotype 1, and the lifetime risk of developing LC and HCC among HCV carriers was decreased during the two recent decades [54,13]. However, the amino acid sequence of the NS3 protease domain varies significantly between HCV genotypes and the antiviral efficacy differs in different HCV genotypes [55]. Moreover, protease inhibitors increased the incidence of adverse reactions such as anemia and skin rash [56]. Therefore, estimation of liver cirrhosis risk and prediction of treatment response would be essential to provide a personalized treatment and to achieve the optimal results. Due to the recent advances in pharmacogenetic studies, genetic factors associated with efficacy and adverse effects of anti-HCV treatment were identified. *IL-28B* is a powerful predictor of treatment outcome of pegylated interferon and ribavirin therapy [22], while a genetic variation in the *ITPA* gene was shown to be associated with ribavirin-induced anemia [57]. Since we conducted a retrospective study, and the majority of LC patients did not receive IFN treatment, we could not evaluate the treatment responses in our study design. However, SNPs identified in this study were associated with the LC risk independent of IFN treatment. Although the impact of each SNP was relatively weak compared with viral factors (HCV genotype, core and NS5A mutation [58]) and host factors (age, gender, obesity, and insulin resistance), we found that individuals with three or more risk alleles have a nearly three-fold higher risk of LC than those with no risk allele. Since lifetime risk of HCC development among HCV carriers is as high as about 27% for male and 8% for female [59], these three loci would have the strong effect on the clinical outcome of CHC patients. In general, the progression from chronic hepatitis C to liver cirrhosis usually takes more than 20–30 years. Therefore, a large scale prospective cohort study with more than 10-year follow-up is essential to evaluate the role of these variations as a prognostic biomarker. We would like to perform prospective analysis in future studies. We hope that our findings would contribute to clarify the underling molecular mechanism of HCV-induced liver cirrhosis.

Financial support

This work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government.

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.

Authors' contributions

Y. U., K. K., K. C., and K.M. conceived and designed the study; Y. U., H. O., N. K., Y. K., R. M., N. H., and M. K. performed genotyping; A. T., P. H. Y. L., C. T., and N. K. performed quality control at genome-wide phase; M. O., R. T., M. O., K. K., D. M., H. A., J. T., H. K., Y. N., K. M. and M. K. managed DNA samples; Y. U. analyzed and summarized the whole results; Y. U., Y. N., and K. M. wrote the manuscript; Y. N. obtained funding for the study.

Acknowledgments

We thank Ayako Matsui and Hiroe Tagaya (the University of Tokyo), and the technical staff of the Laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN, for their technical support.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhep.2012.12.024.

References

[1] Shepard CW, Finelli L, Alter MJ. Global epidemiology of hepatitis C virus infection. Lancet Infect Dis 2005;5:558–567.
[2] Seeff LB. Natural history of chronic hepatitis C. Hepatology 2002;36:S35–S46.
[3] Thomas DL, Seeff LB. Natural history of hepatitis C. Clin Liver Dis 2005;9:383–398, vi.
[4] Freeman AJ, Dore GJ, Law MG, Thorpe M, Von Overbeck J, Lloyd AR, et al. Estimating progression to cirrhosis in chronic hepatitis C virus infection. Hepatology 2001;34:809–816.
[5] Hoofnagle JH. Hepatitis C: the clinical spectrum of disease. Hepatology 1997;26:15S–20S.
[6] Tanaka H, Imai Y, Hiramatsu N, Ito Y, Imanaka K, Oshita M, et al. Declining incidence of hepatocellular carcinoma in Osaka, Japan, from 1990 to 2003. Ann Intern Med 2008;148:820–826.
[7] Global burden of disease (GBD) for hepatitis C. J Clin Pharmacol 2004;44:20–29.

Please cite this article in press as: Urabe Y et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. J Hepatol (2013), http://dx.doi.org/10.1016/j.jhep.2012.12.024

[8] Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet 1997;349:825–832.
[9] Yoshida H, Shiratori Y, Moriyama M, Arakawa Y, Ide T, Sata M, et al. Interferon therapy reduces the risk for hepatocellular carcinoma: national surveillance program of cirrhotic and noncirrhotic patients with chronic hepatitis C in Japan. IHIT Study Group. Inhibition of hepatocarcinogenesis by interferon therapy. Ann Intern Med 1999;131:174–181.
[10] Zhang Q, Tanaka K, Sun P, Nakata M, Yamamoto R, Sakimura K, et al. Suppression of synaptic plasticity by cerebrospinal fluid from anti-NMDA receptor encephalitis patients. Neurobiol Dis 2012;45:610–615.
[11] Aghemo A, Prati GM, Rumi MG, Soffredini R, D'Ambrosio R, Orsi E, et al. A sustained virological response prevents development of insulin resistance in chronic hepatitis C patients. Hepatology 2012;56:549–556.
[12] Bièche I, Asselah T, Laurendeau I, Vidaud D, Degot C, Paradis V, et al. Molecular profiling of early stage liver fibrosis in patients with chronic hepatitis C virus infection. Virology 2005;332:130–144.
[13] Estrabaud E, Vidaud M, Marcellin P, Asselah T. Genomics and HCV infection: progression of fibrosis and treatment response. J Hepatol 2012;57:1110–1125.
[14] Asselah T, Bièche I, Laurendeau I, Paradis V, Vidaud D, Degott C, et al. Liver gene expression signature of mild fibrosis in patients with chronic hepatitis C. Gastroenterology 2005;129:2064–2075.
[15] Marquez RT, Bandyopadhyay S, Wendlandt EB, Keck K, Hoffer BA, Icardi MS, et al. Correlation between microRNA expression levels and clinical parameters associated with chronic hepatitis C viral infection in humans. Lab Invest 2010;90:1727–1736.
[16] Morita K, Taketomi A, Shirabe K, Umeda K, Kayashima H, Ninomiya M, et al. Clinical significance and potential of hepatic microRNA-122 expression in hepatitis C. Liver Int 2011;31:474–484.
[17] Cui R, Okada Y, Jang SG, Ku JL, Park JG, Kamatani Y, et al. Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. Gut 2011;60:799–805.
[18] Kumar V, Matsuo K, Takahashi A, Hosono N, Tsunoda T, Kamatani N, et al. Common variants on 14q32 and 13q12 are associated with DLBCL susceptibility. J Hum Genet 2011;56:436–439.
[19] Tanikawa C, Urabe Y, Matsuo K, Kubo M, Takahashi A, Ito H, et al. A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population. Nat Genet 2012;44:430–434.
[20] Urabe Y, Tanikawa C, Takahashi A, Okada Y, Morizono T, Tsunoda T, et al. A genome-wide association study of nephrolithiasis in the Japanese population identifies novel susceptible loci at 5q35.3, 7p14.3, and 13q14.1. PLoS Genet 2012;8:e1002541.
[21] Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 2009;461:798–801.
[22] Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 2009;41:1105–1109.
[23] Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate ML, et al. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. Nat Genet 2009;41:1100–1104.
[24] Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, Hosono N, et al. Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nat Genet 2011;43:455–458.
[25] Miki D, Ochi H, Hayes CN, Abe H, Yoshima T, Aikata H, et al. Variation in the DEPDC5 locus is associated with progression to hepatocellular carcinoma in chronic hepatitis C virus carriers. Nat Genet 2011;43:797–800.
[26] Kuzushita N, Hayashi N, Moribe T, Katayama K, Kanto T, Nakatani S, et al. Influence of HLA haplotypes on the clinical courses of individuals infected with hepatitis C virus. Hepatology 1998;27:240–244.
[27] Singh R, Kaul R, Kaul A, Khan K. A comparative review of HLA associations with hepatitis B and C viral infections across global populations. World J Gastroenterol 2007;13:1770–1787.
[28] Mosaad YM, Farag RE, Arafa MM, Eletreby S, El-Alfy HA, Eldeek BS, et al. Association of human leucocyte antigen Class I (HLA-A and HLA-B) with chronic hepatitis C virus infection in Egyptian patients. Scand J Immunol 2010;72:548–553.
[29] Li CZ, Kato N, Chang JH, Muroyama R, Shao RX, Dharel N, et al. Polymorphism of OAS-1 determines liver fibrosis progression in hepatitis C by reduced ability to inhibit viral replication. Liver Int 2009;29:1413–1421.
[30] Mozer-Lisewska I, Sikora J, Kowala-Piaskowska A, Kaczmarek M, Dworacki G, Zeromski J. The incidence and significance of pattern-recognition receptors in chronic viral hepatitis types B and C in man. Arch Immunol Ther Exp (Warsz) 2010;58:295–302.
[31] Trépo E, Pradat P, Potthoff A, Momozawa Y, Quertinmont E, Gustot T, et al. Impact of patatin-like phospholipase-3 (rs738409 C>G) polymorphism on fibrosis progression and steatosis in chronic hepatitis C. Hepatology 2011;54:60–69.
[32] Romero-Gómez M, Gómez-González E, Madrazo A, Vera-Valencia M, Rodrigo L, Pérez-Alvarez R, et al. Optical analysis of computed tomography images of the liver predicts fibrosis stage and distribution in chronic hepatitis C. Hepatology 2008;47:810–816.
[33] Nakamura Y. The BioBank Japan project. Clin Adv Hematol Oncol 2007;5:696–697.
[34] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575.
[35] Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851–861.
[36] Breslow NE, Day NE. Statistical methods in cancer research. The design and analysis of cohort studies. IARC Sci Publ 1987;II:1–406.
[37] de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 2006;38:1166–1172.
[38] Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316:1341–1345.
[39] Consortium GP. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–1073.
[40] Barrett J, Fry B, Maller J, Daly M. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–265.
[41] Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res 2009;37:W600–W605.
[42] Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, et al. A genome-wide association study of global gene expression. Nat Genet 2007;39:1202–1207.
[43] Kel AE, Gössling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: a tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res 2003;31:3576–3579.
[44] Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. Hepatology 2003;38:518–526.
[45] Cammà C, Di Bona D, Schepis F, Heathcote EJ, Zeuzem S, Pockros PJ, et al. Effect of peginterferon alfa-2a on liver histology in chronic hepatitis C: a meta-analysis of individual patient data. Hepatology 2004;39:333–342.
[46] Marcellin P, Asselah T, Boyer N. Fibrosis and disease progression in hepatitis C. Hepatology 2002;36:S47–S56.
[47] Silini E, Bottelli R, Asti M, Bruno S, Candusso ME, Brambilla S, et al. Hepatitis C virus genotypes and risk of hepatocellular carcinoma in cirrhosis: a case-control study. Gastroenterology 1996;111:199–205.
[48] Kawaguchi T, Sumida Y, Umemura A, Matsuo K, Takahashi M, Takamura T, et al. Genetic polymorphisms of the human PNPLA3 gene are strongly associated with severity of non-alcoholic fatty liver disease in japanese. PLoS One 2012;7:e38322.
[49] Nischalke HD, Berger C, Luda C, Berg T, Müller T, Grünhage F, et al. The PNPLA3 rs738409 148M/M genotype is a risk factor for liver cancer in alcoholic cirrhosis but shows no or weak association in hepatitis C cirrhosis. PLoS One 2011;6:e27087.
[50] He S, McPhaul C, Li JZ, Garuti R, Kinch L, Grishin NV, et al. A sequence variation (I148M) in PNPLA3 associated with nonalcoholic fatty liver disease disrupts triglyceride hydrolysis. J Biol Chem 2010;285:6706.
[51] Patin E, Kutalik Z, Guergnon J, Bibert S, Nalpas B, Jouanguy E, et al. Genome-wide association study identifies variants associated with progression of liver fibrosis from HCV infection. Gastroenterology 2012;143:124–152, e1–12.
[52] Kim AY, Kuntzen T, Timm J, Nolan BE, Baca MA, Reyor LL, et al. Spontaneous control of HCV is associated with expression of HLA-B 57 and preservation of targeted epitopes. Gastroenterology 2011;140:e681.
[53] Fanning LJ, Kenny-Walsh E, Shanahan F. Persistence of hepatitis C virus in a white population: associations with human leukocyte antigen class 1. Hum Immunol 2004;65:745–751.
[54] Imhof I, Simmonds P. Genotype differences in susceptibility and resistance development of hepatitis C virus to protease inhibitors telaprevir (VX-950) and danoprevir (ITMN-191). Hepatology 2011;53:1090–1099.

Please cite this article in press as: Urabe Y et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. J Hepatol (2013), http://dx.doi.org/10.1016/j.jhep.2012.12.024

Research Article

[55] Asselah T, Marcellin P. Direct acting antivirals for the treatment of chronic hepatitis C: one pill a day for tomorrow. Liver Int 2012;32:88–102.
[56] Ozeki I, Akaike J, Karino Y, Arakawa T, Kuwata Y, Ohmura T, et al. Antiviral effects of peginterferon alpha-2b and ribavirin following 24-week monotherapy of telaprevir in Japanese hepatitis C patients. J Gastroenterol 2011;46:929–937.
[57] Ochi H, Maekawa T, Abe H, Hayashida Y, Nakano R, Kubo M, et al. ITPA polymorphism affects ribavirin-induced anemia and outcomes of therapy – a genome-wide study of Japanese HCV virus patients. Gastroenterology 2010;139:1190–1197.
[58] Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N Engl J Med 1996;334:77–82.
[59] Huang YT, Jen CL, Yang HI, Lee MH, Su J, Lu SN, et al. Lifetime risk and sex difference of hepatocellular carcinoma among patients with chronic hepatitis B and C. J Clin Oncol 2011;29:3643–3650.

Please cite this article in press as: Urabe Y et al. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. J Hepatol (2013), http://dx.doi.org/10.1016/j.jhep.2012.12.024

HEPATOLOGY
Official Journal of the American Association for the Study of Liver Diseases

MicroRNA-140 Acts as a Liver Tumor Suppressor by Controlling NF-κB Activity by Directly Targeting DNa Methyltransferase 1 (Dnmt1) Expression

Akemi Takata,[1] Motoyuki Otsuka,[1] Takeshi Yoshikawa,[1] Takahiro Kishikawa,[1] Yohko Hikiba,[2] Shuntaro Obi,[3] Tadashi Goto,[1] Young Jun Kang,[4] Shin Maeda,[1] Haruhiko Yoshida,[1] Masao Omata,[1] Hiroshi Asahara,[5,6,7] and Kazuhiko Koike[1]

MicroRNAs (miRNAs) are small RNAs that regulate the expression of specific target genes. While deregulated miRNA expression levels have been detected in many tumors, whether miRNA functional impairment is also involved in carcinogenesis remains unknown. We investigated whether deregulation of miRNA machinery components and subsequent functional impairment of miRNAs are involved in hepatocarcinogenesis. Among miRNA-containing ribonucleoprotein complex components, reduced expression of DDX20 was frequently observed in human hepatocellular carcinomas, in which enhanced nuclear factor-κB (NF-κB) activity is believed to be closely linked to carcinogenesis. Because DDX20 normally suppresses NF-κB activity by preferentially regulating the function of the NF-κB-suppressing miRNA-140, we hypothesized that impairment of miRNA-140 function may be involved in hepatocarcinogenesis. DNA methyltransferase 1 (Dnmt1) was identified as a direct target of miRNA-140, and increased Dnmt1 expression in DDX20-deficient cells hypermethylated the promoters of metallothionein genes, resulting in decreased metallothionein expression leading to enhanced NF-κB activity. MiRNA-140-knockout mice were prone to hepatocarcinogenesis and had a phenotype similar to that of DDX20 deficiency, suggesting that miRNA-140 plays a central role in DDX20 deficiency-related pathogenesis. ***Conclusion*****: These results indicate that miRNA-140 acts as a liver tumor suppressor, and that impairment of miRNA-140 function due to a deficiency of DDX20, a miRNA machinery component, could lead to hepatocarcinogenesis. (HEPATOLOGY 2013;57:162-170)**

Hepatocellular carcinoma (HCC) is the third most common cause of cancer-related mortality worldwide.[1] Although multiple major risk factors have been identified, such as infection with hepatitis viruses B or C, the molecular mechanisms underlying HCC development remain poorly understood, hindering the development of novel therapeutic approaches. Therefore, a better understanding of the molecular pathways involved in hepatocarcinogenesis is critical for the development of new therapeutic options.

Nuclear factor-κB (NF-κB) is one of the best-characterized intracellular signaling pathways. Its activation is a common feature of human HCC.[2-4] It acts as an inhibitor of apoptosis and as a tumor promoter[4,5] and is associated with the acquisition of a transformed phenotype during hepatocarcinogenesis.[6] In fact, studies using patient samples suggest that NF-κB activation in the liver leads to the development of HCC.[7] Although there are conflicting reports,[8] activation of the NF-κB pathway in the liver is crucial for the initiation and promotion of HCC.[4]

Abbreviations: DEN, diethylnitrosamine; Dnmt1, DNA methyltransferase 1; HCC, hepatocellular carcinoma; miRNA; microRNA; miRNP, miRNA-containing ribonucleoprotein; MT, metallothionein; NF-κB, nuclear factor-κB; RT-PCR, reverse-transcription polymerase chain reaction; TNF-α, tumor necrosis factor-α; TRAIL, TNF-related apoptosis-inducing ligand; UTR, untranslated region.

From the [1]Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; the [2]Division of Gastroenterology, Institute for Adult Diseases, Asahi Life Foundation, Tokyo, Japan; the [3]Department of Hepatology, Kyoundo Hospital, Tokyo, Japan; the [4]Department of Immunology and Microbial Science, and the [5]Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA; the [6]Department of Systems BioMedicine, Tokyo Medical and Dental University, Tokyo, Japan; and [7]CREST, Japan Science and Technology Agency, Tokyo, Japan.

Received March 30, 2012; accepted July 18, 2012.

Supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology, Japan (#22390058, #23590960, and #20390204) (M. O., T. G., and K. K.); Health Sciences Research Grants from the Ministry of Health, Labor and Welfare of Japan (Research on Hepatitis) (to K. K.); National Institutes of Health Grant R01AI088229 (to Y. J. K.); the Miyakawa Memorial Research Foundation (to A. T.); and grants from the Sagawa Foundation for Promotion of Cancer Research, the Astellas Foundation for Research on Metabolic Disorders, and the Cell Science Research Foundation (to M. O.).

MicroRNAs (miRNAs) are small RNA molecules that regulate the expression of target genes and are involved in various biological functions.[9-12] Although specific miRNAs can function as either suppressors or oncogenes in tumor development, a general reduction in miRNA expression is commonly observed in human cancers.[13-22] In this context, it can be hypothesized that deregulation of the machinery components involved in miRNA function may be related to the functional impairment of miRNAs and the pathogenesis of carcinogenesis.

In this study, we show that the expression of DDX20, an miRNA-containing ribonucleoprotein (miRNP) component, is frequently decreased in human HCC. Because DDX20 is required for both the preferential loading of miRNA-140 into the RNA-induced silencing complex and its function,[23] we hypothesized that DDX20 deficiency would lead to hepatocarcinogenesis via impaired miRNA-140 function. MiRNA-140 knockout mice were indeed more prone to hepatocarcinogenesis, and we identified a possible molecular pathway from DDX20 deficiency to liver cancer.

Materials and Methods

Mouse and Liver Tumor Induction. MiRNA-140$^{-/-}$ mice have been described.[24] Recombinant murine tumor necrosis factor-α (TNF-α) (25 μg/kg; Wako, Osaka, Japan) was injected into the tail vein, and the mice were sacrificed 1 hour later. To induce liver tumors, 15-day-old mice received an intraperitoneal injection of diethylnitrosamine (DEN) (25 mg/kg body weight), and were sacrificed 32 weeks later. All animal experiments were performed in compliance with the regulations of the Animal Use Committee of the University of Tokyo and the Institute for Adult Disease, Asahi Life Foundation.

Plasmids. FLAG-tagged human DDX20-expressing plasmids were as described.[23] The pGL3-based reporter plasmid containing Dnmt1 3′ untranslated region (UTR) sequences was provided by G. Marucucci.[25]

Detailed Materials and Methods. The detailed experimental procedures of clinical samples, cells, plasmids, reporter assays, reverse-transcription polymerase chain reaction (RT-PCR) analysis, antibodies, western blotting, cell assays, immunohistochemistry, microarray analysis, methylation analysis, and electrophoretic mobility-shift assay are described in the Supporting Information.

Statistical Analysis. Statistically significant differences between groups were determined using a Wilcoxon rank-sum test. A Wilcoxon signed-rank test was used for statistical comparisons of protein expression levels between HCC and surrounding noncancerous tissues.

Table 1. Cases with Differential Expression Levels of miRNP Components in HCC (n = 10)

Gene ID	Gene Symbol	Decreased	Increased	No Change
23405	Dicer1	2	1	7
27161	EIF2C2 (AGO2)	1	1	8
6895	TARBP2(TRBP2)	2	0	8
11218	DDX20 (GEMIN3)	8	0	2
50628	GEMIN4	1	0	9

The expression levels of each miRNP component were determined via immunohistochemistry.

The numbers indicate the number of cases that had the differential expression levels (decreased, increased, or no change) in HCC tissues compared with those in surrounding liver tissues.

Results

DDX20 Expression Is Frequently Decreased in HCC. The expression levels of proteins reported to be miRNP components (Dicer, Ago2, TRBP2, DDX20 [also known as Gemin3], and Gemin4)[26] were initially determined via immunohistochemistry in HCC and noncancerous background liver tissues from 10 patients. DDX20 expression was lower in HCC tissue compared with the surrounding noncancerous tissue in 8 of 10 cases, whereas expression of the other genes was unchanged (Table 1 and Supporting Fig. 1). Therefore, and because DDX20 was recently identified as a possible liver tumor suppressor in mice,[27] we determined its role as a human HCC suppressor.

DDX20 protein expression was lower in several HCC cell lines, such as Huh7 and Hep3B (Fig. 1A), compared with normal hepatocytes. DDX20 protein levels were also lower in human HCC needle biopsy specimens than in surrounding noncancerous liver tissue (Fig. 1B). Immunohistochemical analysis

Address reprint requests to: Motoyuki Otsuka, M.D., Department of Gastroenterology, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. E-mail: otsukamo-tky@umin.ac.jp; fax: (81)-3-3814-0021.

View this article online at wileyonlinelibrary.com.

DOI 10.1002/hep.26011

Potential conflict of interest: Nothing to report.

Additional Supporting Information may be found in the online version of this article.

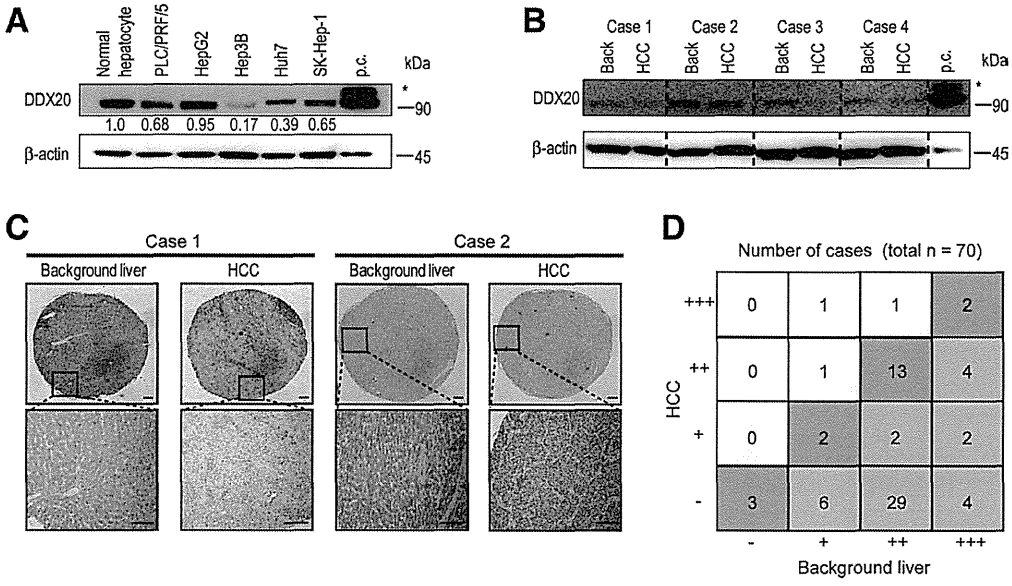

Fig. 1. Reduced DDX20 expression levels in hepatocellular carcinoma. (A) DDX20 protein expression in HCC cell lines. Numbers between the panels indicate DDX20 protein levels normalized to β-actin levels. Lysates of 293T cells transiently transfected with a FLAG-tagged DDX20-expressing plasmid yielded two DDX20 bands corresponding to the endogenous DDX20 protein and the transfected FLAG-tagged DDX20 protein (*) as a positive control (p.c.; far right lane). Data represent the results of three independent determinations. (B) DDX20 protein expression in four HCC needle biopsy specimens and in the surrounding noncancerous background liver tissue (Back). *Positive control. (C) Immunohistochemical analysis of DDX20 protein expression in HCC and surrounding tissues (background liver). Two representative cases are shown. Scale bars, 500 μm. The lower panels display magnified images of the boxed areas in the upper panels. (D) Grid summarizing DDX20 immunohistochemical staining data from 70 cases. In 47 cases (pink shading), DDX20 protein levels were lower in the HCC tissues than in the surrounding tissues ($P < 0.05$; Wilcoxon signed-rank test).

confirmed that DDX20 expression was frequently lower in HCC than in surrounding noncancerous liver tissue (Fig. 1C,D). Specifically, 47 of 70 cases examined showed reduced DDX20 protein expression in HCC versus background noncancerous liver tissue (Fig. 1D and Supporting Table 1). These results indicate that the expression of DDX20, an miRNP component, is frequently reduced in human HCC, and suggest that this reduced DDX20 expression might be involved in the pathogenesis of a subset of HCC cases.

NF-κB Activity Is Enhanced by DDX20 Deficiency. Because DDX20 knockout mice are embryonic-lethal,[28] DDX20 has been suggested to have important biological roles. DDX20, a DEAD-box protein,[29] was originally found to interact with survival motor neuron protein.[30] Later, it was identified as a major component of miRNPs,[31] which may mediate miRNA function. As we have reported, DDX20 is preferentially involved in miRNA-140-3p function,[23] acting as a suppressor of NF-κB activity in the liver.[32] DDX20-knockdown PLC/PRF/5 cells exhibit enhanced NF-κB activity[23] (Fig. 2A). Whereas the proliferation rates of DDX20-knockdown cells and control cells were comparable (Fig. 2B), apoptotic cell death after stimulation with TNF-related apoptosis-inducing ligand (TRAIL), which induces both cell apoptosis and NF-κB activation,[33] was significantly reduced in DDX20-knockdown cells (Fig. 2C). Similar results were obtained using DDX20-knockdown HepG2 cells (Supporting Fig. 2A-D). Conversely, NF-κB activity was reduced, but cell proliferation remained unchanged, in Hep3B cells stably overexpressing DDX20 (Fig. 2D,E). Sensitivity to TRAIL-induced apoptosis was restored in these cells (Fig. 2F). Similar results were also obtained using Huh7 cells (Supporting Fig. 2E-H). These data confirm a previous report that DDX20 deficiency enhances NF-κB activity and the downstream events of this pathway.

Metallothionein Expression Is Decreased by DDX20 Deficiency. Next, to investigate the biological consequences of DDX20 deficiency, we examined the changes in transcript levels in DDX20-knockdown cells using microarrays (GEO accession number: GSE28088). The expression of genes driven by NF-κB that are related to carcinogenesis, such as FASLG, IRAK1, CARD9, and Galectin-1, were enhanced significantly in DDX20-knockdown cells, as expected (Table 2). To determine the mechanism underlying the enhanced NF-κB activation in DDX20-deficient cells, we searched for candidate genes and noticed that the

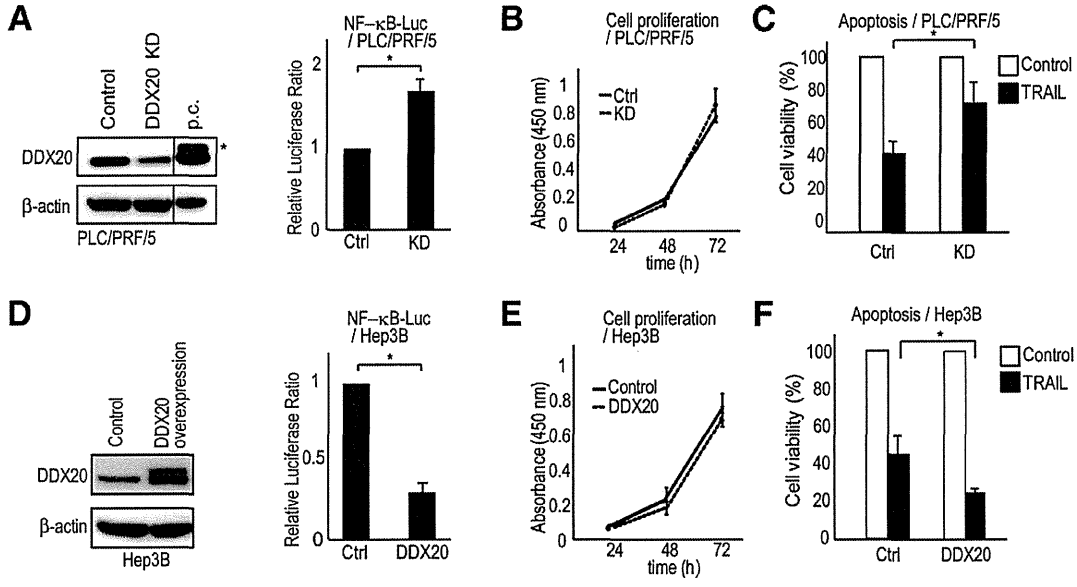

Fig. 2. Modulation of downstream events of the nuclear factor-κB pathway by DDX20. (A) Left: Establishment of stable DDX20-knockdown (DDX20 KD) PLC/PRF/5 cells. *Positive control (p.c.). Right: DDX20 deficiency enhances TNF-α–induced NF-κB activity. NF-κB reporter plasmids were transiently transfected into control (Ctrl) or DDX20-knockdown (KD) PLC/PRF/5 cells. The cells were then treated with TNF-α (5 ng/mL) or vehicle for 6 hours. $*P < 0.05$. Data are presented as the mean ± SD of three independent determinations. (B) Cell proliferation rates were comparable for control (Ctrl) and DDX20-knockdown (KD) PLC/PRF/5 cells. Data are presented as the mean ± SD of three determinations. (C) DDX20 deficiency reduces TRAIL-induced apoptotic cell death. Control (Ctrl) and DDX20-knockdown (KD) PLC/PRF/5 cells were incubated with 25 ng/mL TRAIL. Data represent cell viability after TRAIL stimulation (gray bars) relative to the number of vehicle-treated cells (white bars). $*P < 0.05$. Data are presented as the mean ± SD of triplicate determinations. (D) Left: Establishment of stable DDX20-overexpressing cells. Hep3B cells were infected with control or FLAG-tagged DDX20-overexpressing lentiviruses and selected on puromycin. Western blot analysis confirmed increased expression of DDX20 protein. Right: DDX20 overexpression suppresses TNF-α–induced NF-κB activity. NF-κB reporter plasmids were transiently transfected into Hep3B control (Ctrl) and DDX20-overexpressing (DDX20) cells treated with TNF-α for 6 hours. Data are presented as the mean ± SD of three independent determinations. $*P < 0.05$. (E) Proliferation of control (Ctrl) and DDX20-overexpressing (DDX20) Hep3B cells was measured as described in (B). (F) DDX20 overexpression reduces TRAIL-induced apoptotic cell death. Data for control (Ctrl) and DDX20-overexpressing (DDX20) Hep3B cells are shown. $*P < 0.05$.

Table 2. Increased Expression of NF-κB–Related Genes in DDX20-Knockdown HepG2 Cells Compared with Wild-Type Cells

RefSeq ID	Symbol	Description	Ratio	Representative Gene Function
NM_000639	FASLG	Fas ligand	3.5	NF-κB target, apoptosis
NM_052813	C9orf151	CARD9	2.5	NF-κB cascade, NF-κB target
NM_014959	CARD8	Tumor up-regulated CARD-containing antagonist of CASP9 (TUCAN)	2.2	NF-κB target
NM_131917	FAF1	FAS-associated factor 1 (hFAF1)	1.9	Cytoplasmic sequestering of NF-κB, NF-κB target
NM_020644	TMEM9B	Transmembrane protein 9B precursor	1.9	Positive regulation of NF-κB transcription factor activity
NM_017544	NKRF	ITBA4 protein	1.9	Negative regulation of transcription
NM_006247	PPP5C	Protein phosphatase T	1.8	Positive regulation of NF-κB cascade
NM_020345	NKIRAS1	KappaB-Ras1	1.8	NF-κB cascade
NM_001569	IRAK1	IRAK-1	1.7	Positive regulation of NF-κB transcription factor activity
NM_177951	PPM1A	Protein phosphatase 1A	1.7	Positive regulation of NF-κB cascade
NM_018098	ECT2	Epithelial cell-transforming sequence 2 oncogene	1.6	Positive regulation of NF-κB cascade
NM_002305	LGALS1	Galectin-1 (putative MAPK-activating protein MP12)	1.6	Positive regulation of NF-κB cascade
NM_015093	TAB2	TAK1-binding protein 2	1.6	Positive regulation of NF-κB cascade
NM_004180	TANK	TRAF-interacting protein	1.5	NF-κB cascade
NM_014976	PDCD11	Programmed cell death protein 11	1.5	rRNA processing
NM_015336	ZDHHC17	Putative NF-κB–activating protein 205	1.5	Positive regulation of NF-κB cascade
NM_002503	NFKBIB	IKB-β	1.5	Cytoplasmic sequestering of NF-κB
NM_138330	ZNF675	Zinc finger protein 675	1.5	Negative regulation of NF-κB transcription factor activity

The genes were identified as NF-κB–related based on the Gene Ontology and the GeneCodis Databases.

Table 3. Decreased Expression Levels of MT Genes in DDX20 Knockdown HepG2 Cells Compared with Wild-Type Cells

Symbol	Description	Ratio
MT1E	Metallothionein-1E	**0.12**
MT1F	Metallothionein-1F	**0.36**
MT1H	Metallothionein-1H	**0.16**
MT1G	Metallothionein-1G	**0.06**
MT1M	Metallothionein-1M	**0.24**
MT1X	Metallothionein-1X	**0.27**
MT2A	Metallothionein-2	**0.28**
MT3	Metallothionein-3	0.84
MTL5	Metallothionein-like 5 (Tesmin)	1.12

Numbers in boldface type indicate values <0.5.

expression levels of a group of metallothioneins (MTs), such as MT1E, MT1F, MT1G, MT1M, MT1X, and MT2A, were all significantly decreased when DDX20 was deficient (Table 3). The decreased expression of MTs in DDX20-knockdown HepG2 and PLC/PRF/5 cells was confirmed via quantitative RT-PCR (Fig. 3a and Supporting Fig. 3). Expression of MT-3, which was not altered in the microarray analysis, was similarly unaltered in quantitative RT-PCR analysis. Notably, it was already known that MTs are frequently silenced in human primary liver cancers.[34-36] In addition, MT knockout mice have enhanced NF-κB activity, likely due to reactive oxygen species, and these mice are more prone to hepatocarcinogenesis.[37] These results suggest that DDX20 deficiency enhances NF-κB activity by decreasing the expression of MTs, which could facilitate the development of liver cancer.

MiRNA-140 Directly Targets Dnmt1. Because MT expression is regulated principally by CpG island methylation in their promoter regions,[38,39] we examined the quantitative methylation status of MT promoters in DDX20-knockdown cells. The CpG islands of the MT1E, MT1G, MT1M, MT1X, and MT2A promoters, and the CpG shores of the MT1F promoters, were significantly more highly methylated under DDX20-deficient conditions, as determined by the comprehensive Illumina Quantitative Methylation BeadChip method (Table 4, Supporting Table 2, and GSE 37633). A crucial step in DNA methylation involves DNA methyltransferase (Dnmt), which catalyzes the methylation of CpG dinucleotides in genomic DNA.[40] The methylation status of MT promoters is mediated specifically by Dnmt1.[41] Because Dnmt1 contains a predicted miRNA-140-3p target site in its 3′ UTR, with a perfect match to its seed sequences (Fig. 3B), and because the effects of miRNA-140-3p activity were impaired in DDX20-knockdown cells,[23] it was hypothesized that whereas miRNA-140 normally targets and suppresses Dnmt1 protein expression, miRNA-140-3p dysfunction due to DDX20 deficiency results in enhanced Dnmt1 expression, leading to hypermethylation of MT promoters. Consistent with this hypothesis, Dnmt1 expression was increased significantly in DDX20-knockdown cells (Fig. 3C). miRNA-140 precursor overexpression suppressed activity of the Dnmt1 3′ UTR reporter construct, the effect of which was lost when two mutations were introduced into its seed sequences (Fig. 3D). MiRNA-140 precursor overexpression suppressed Dnmt1 protein expression (Fig. 3E). These results indicate that miRNA-140 directly targets Dnmt1 and suppresses its expression in the normal state. Consistently, decreased DDX20, increased Dnmt1, and decreased MT expression were detected together in human clinical HCC samples, as determined via immunohistochemistry (Fig. 3F). By contrast, miRNA-140 precursor-overexpressing Huh7 cells showed increased expression of MTs and reduced NF-κB activity *in vitro* (Supporting Fig. 4A,B). Moreover, the increase in the number of spheres formed from PLC/PRF/5 cells due to DDX20 knockdown was antagonized by treatment with an NF-κB inhibitor or a demethylating agent (Supporting Fig. 5). Taken together, these results suggest that the up-regulated Dnmt1 protein expression caused by functional impairment of miRNA-140-3p due to DDX20 deficiency results in decreased expression of MTs *via* enhanced methylation at the CpG sites in their promoters. This may lead to enhanced NF-κB activity and cellular transformation at least *in vitro*.

MiRNA-140 Is a Liver Tumor Suppressor. To further examine the biological consequences of functional impairment of miRNA-140 due to DDX20 deficiency, we determined the phenotypes of miRNA-140 knockout (miRNA-140$^{-/-}$) mice (Fig. 4A). Similar to the *in vitro* DDX20 knockdown results, Dnmt1 expression was increased and MT levels decreased in the liver tissue of these mice (Fig. 4B). NF-κB–DNA binding activity was enhanced in the livers of miRNA-140$^{-/-}$ mice after tail-vein injection of TNF-α, a crucial cytokine that induces NF-κB activity and hepatocarcinogenesis (Fig. 4C). As was found in MT knockout mice, phosphorylation of p65 at serine 276, which is critical for p65 activation, was significantly increased in the livers of miRNA-140$^{-/-}$ mice after DEN exposure, which induces NF-κB activation and liver tumors[37] (Fig. 4D). Notably, the size and number of liver tumors that developed 8 months after DEN exposure were markedly elevated in miRNA-140$^{-/-}$ mice compared with control mice (Fig. 4E,F). These results indicate that miRNA-140$^{-/-}$ mice are indeed

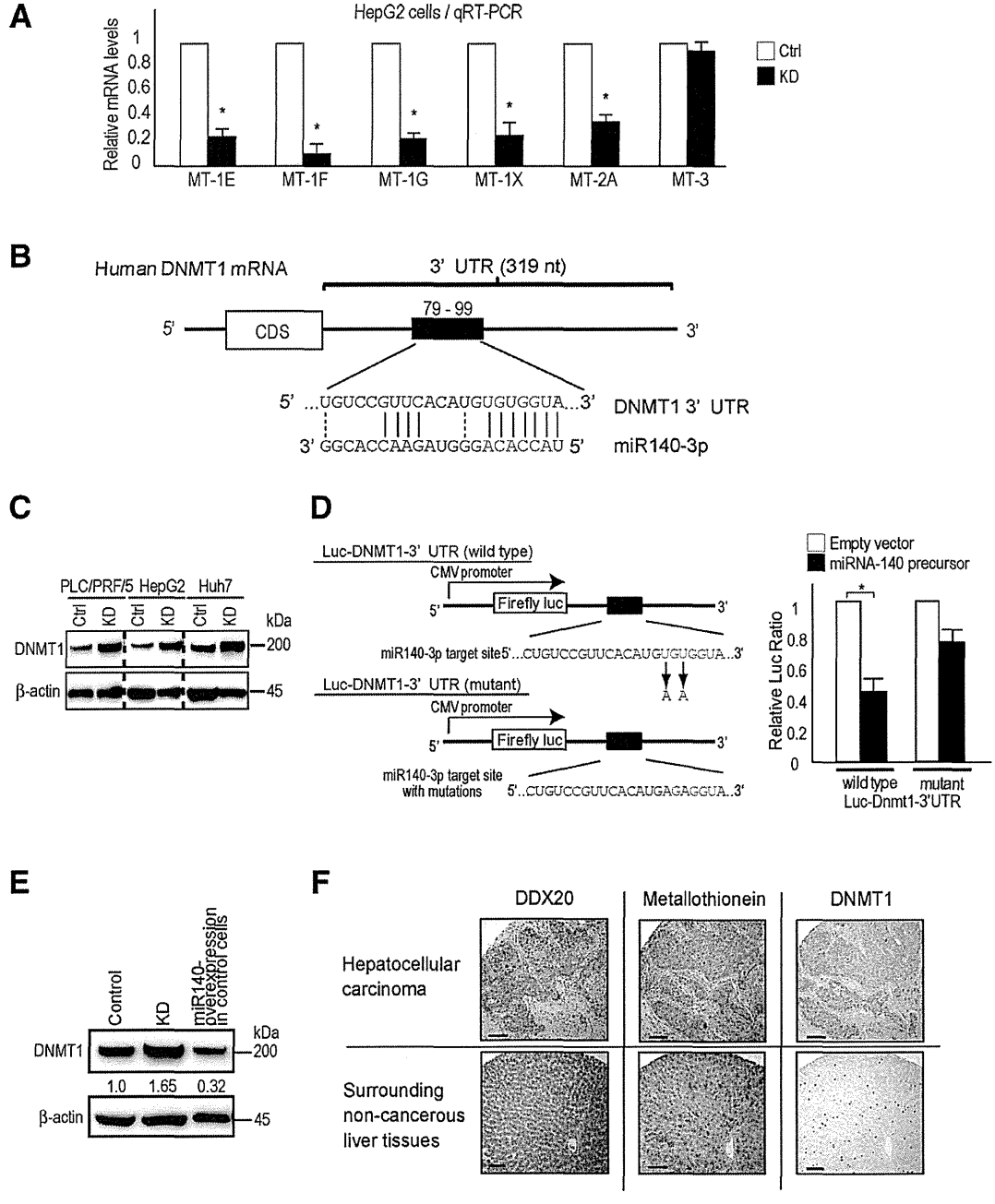

Fig. 3. Targeting of Dnmt1 by miRNA-140-3p and reduced MT expression. (A) The expression levels of MTs were determined using quantitative reverse-transcriptase polymerase chain reaction. The relative expression ratios of the MTs in control (white bars) and DDX20-knockdown (black bars) HepG2 cells were calculated by normalizing control cell values to 1.0. The data represent the mean ± SD of three independent determinations. $*P < 0.05$. (B) Putative miRNA-140-3p target sites in the 3′ UTR of human Dnmt1. Seed sequences are indicated in red. (C) Dnmt1 expression was increased in DDX20-knockdown cells. Ctrl, control cells; KD, DDX20-knockdown cells. (D) Left: Schematic diagrams of wild-type (upper) and mutant (lower) luciferase reporter constructs (Luc-Dnmt1-3′ UTRs) carrying the Dnmt1 3′ UTR region harboring the putative miRNA-140-3p target site. The mutant seed sequence contained two nucleotide substitutions. Right: The Dnmt1 3′ UTR is targeted directly by miRNA-140-3p. Cells were cotransfected with Luc-Dnmt1-3′ UTR (wild-type or mutant) plus either an empty vector (white bars) or a plasmid expressing the miRNA-140 precursor (black bars). Data are the mean ± SD of three independent determinations. (E) Overexpression of miRNA-140 reduces Dnmt1 expression in control cells. Values between the panels indicate Dnmt1 protein levels normalized to those of β-actin. KD, DDX20 knockdown cells. (F) Representative histochemical images showing expression of DDX20, Dnmt1, and MT proteins in HCC (upper three panels) and surrounding tissue (lower panels). Compared with adjacent noncancerous liver tissue, HCCs exhibited decreased DDX20 and MT expression and increased Dnmt1 expression. Note that adjacent sections were stained for each protein. Scale bar, 50 μm.

Table 4. Methylation Levels in CpG Islands of the MT Genes in DDX20-Knockdown HepG2 Cells Compared with Control Cells

Symbol	CpG Island Methylation Ratio	Target ID
MT1E	**1.14**	cg00178359
	1.29	cg06463589
	3.65	cg02512505
	1.02	cg15134649
MT1G	**2.14**	cg16452857
	1.03	cg27367960
	1.00	cg03566142
	0.99	cg07791866
MT1M	**1.16**	cg02132560
	0.98	cg02160530
	1.03	cg04994964
MT1X	**1.24**	cg05596720
	1.05	cg26802333
	1.06	cg09147880
	1.01	cg08872713
MT2A	**2.06**	cg07395075
	0.94	cg20430434

Values were determined using the quantitative Illumina Human Methylation BeadsChip. Boldface values indicate increased methylation levels in DDX20 knockdown cells.

more prone to liver cancer development and suggest that miRNA-140 acts as a liver tumor suppressor, probably by suppressing NF-κB activity, although we cannot completely exclude other molecular mechanisms. Nonetheless, these results also suggest that the impairment of miRNA-140 function due to DDX20 deficiency may lead to hepatocarcinogenesis in humans, as we have observed in miRNA-140$^{-/-}$ mice (Supporting Figs. 6 and 7).

Discussion

Here, we report that miRNA-140$^{-/-}$ mice have increased NF-κB activity and are more prone to HCC development. In addition, we show that DDX20, an miRNP component, is frequently decreased in human HCC tissues. Because DDX20 deficiency preferentially causes impaired miRNA-140 function,[23] the functional impairment of miRNA-140 may result in phenotypes similar to those of miRNA-140$^{-/-}$ mice and may lead to hepatocarcinogenesis. In support of the hypothesis that DDX20 dysfunction is involved in hepatocarcinogenesis, DDX20 is located at 1p21.1-p13.2, a frequently deleted chromosomal region in human HCC,[27] and DDX20 was recently identified as a possible liver tumor suppressor in a functional screen in mice.[27] Although the possibility that intracellular signaling pathways other than miRNA-140 may also be involved in the biological consequences of DDX20 deficiency cannot be denied, we believe that functional impairment of miRNA-140 plays a major role in the phenotypes induced by DDX20 deficiency, based on the phenotypic similarities.

Changes in miRNA expression levels have been reported in various tumors.[7,12,42] However, in this study, we found that reduced expression of an miRNA machinery component might lead to carcinogenesis, at least in part, through functional impairment of miRNAs. Recent studies have shown that components of the RNA interference machinery are associated with the outcome of ovarian cancer patients,[43] and that single-nucleotide polymorphisms in miRNA machinery genes can be used as diagnostic risk markers.[44,45] Therefore, the impairment of miRNA function caused by deregulated miRNA machinery components may also be involved in carcinogenesis.

Our study identified Dnmt1 as a critical target of miRNA-140. The decreased MT expression due to the CpG promoter methylation induced by Dnmt1 resulted in enhanced NF-κB activity. This finding was consistent with the results obtained using MT gene knockout mice, in which enhanced NF-κB activation promoted hepatocarcinogenesis.[37] The decrease in MT expression that results from increased Dnmt1 expression caused by functional impairment of miRNA-140, together with increased NF-κB activation and hepatocarcinogenesis in MT knockout mice,[37] supports the concept that the DDX20/miRNA-140/Dnmt1/MT/NF-κB pathway may play a crucial role in hepatocarcinogenesis. However, we cannot fully exclude the possibility that other intracellular signaling pathways are also involved in the induction of hepatocarcinogenesis by miRNA-140 or DDX20 deficiency, because the precise role of NF-κB in hepatocarcinogenesis has not been clearly defined,[8] although constitutive activation of NF-κB signaling has been frequently detected in human HCCs.[46] The mechanisms by which DDX20 expression is initially decreased and the reason its locus is frequently deleted in HCC remain to be elucidated. However, because DDX20 expression is also regulated by methylation of its CpG promoter,[47] once this pathway is deregulated, decreased DDX20 expression could be maintained by a positive feedback mechanism, even without deletion of its locus.[27]

In conclusion, this study shows that miRNA-140 acts as a liver tumor suppressor. We show that DDX20, an miRNP component, is frequently decreased in human HCC, which may induce hepatocarcinogenesis via impairment of miRNA-140 function. These results suggest the importance of investigations of not only aberrant miRNA expression levels,[12,14,17,48] but also deregulation of miRNP