

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/ μ l) 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 underlying 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.

- [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, Degot 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, Arafat 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, Greenon 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, Guernon 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.

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.

Identification of a functional variant in the *MICA* promoter which regulates *MICA* expression and increases HCV-related hepatocellular carcinoma risk

Paulisally Hau Yi Lo¹, Yuji Urabe^{1,2}, Vinod Kumar¹, Chizu Tanikawa¹, Kazuhiko Koike³, Naoya Kato⁴, Daiki Miki^{2,5}, Kazuaki Chayama^{2,5}, Michiaki Kubo⁵, Yusuke Nakamura^{1,6}, Koichi Matsuda¹

1 Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

2 Departments of Medical and Molecular Science, Division of Frontier Medical Science, Programs for Biomedical Research, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima, Japan

3 Department of Gastroenterology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

4 Unit of Disease Control Genome Medicine, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

5 Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan

6 Departments of Medicine and Surgery, and Center for Personalized Therapeutics, The
University of Chicago. USA

Corresponding Author

Koichi Matsuda M.D., Ph.D.

Associate Professor,

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: koichima@ims.u-tokyo.ac.jp

List of abbreviations: HCV, Hepatitis C virus; HCC, hepatocellular carcinoma; SNP, single nucleotide polymorphism; MICA, MHC class I polypeptide-related sequence A; EMSA, electrophoretic mobility shift assay; ChIP, chromatin immunoprecipitation; ELISA, enzyme-linked immunosorbent assay; SP1, Specificity Protein 1; HBV, Hepatitis B viruses; GWAS, genome-wide association study; MHC, major

histocompatibility complex; NK, natural killer; CHC, chronic hepatitis C; VNTRs, variable numbers of tandem repeats; CNV, copy number variation; LC, liver cirrhosis; HBx, hepatitis B virus X protein.

Abstract

Hepatitis C virus (HCV) infection is the major cause of hepatocellular carcinoma (HCC) in Japan. We previously identified the association of SNP rs2596542 in the 5' flanking region of the *MHC class I polypeptide-related sequence A (MICA)* gene with the risk of HCV-induced HCC. In the current study, we performed detailed functional analysis of 12 candidate SNPs in the promoter region and found that a SNP rs2596538 located at 2.8 kb upstream of the *MICA* gene affected the binding of a nuclear protein(s) to the genomic segment including this SNP. By electrophoretic mobility shift assay (EMSA) and chromatin immunoprecipitation (ChIP) assay, we identified that transcription factor Specificity Protein 1 (SP1) can bind to the protective G allele, but not to the risk A allele. In addition, reporter construct containing the G allele was found to exhibit higher transcriptional activity than that containing the A allele. Moreover, SNP rs2596538 showed stronger association with HCV-induced HCC ($P = 1.82 \times 10^{-5}$ and OR = 1.34) than the previously identified SNP rs2596542. We also found significantly higher serum level of soluble MICA (sMICA) in HCV-induced HCC patients carrying the G allele than those carrying the A allele ($P = 0.00616$). In summary, we have identified a functional SNP that is associated with the expression of MICA and the risk for HCV-induced HCC.

Introduction

Hepatocellular carcinoma (HCC) is one of the common cancers in the world. It is well-known to be associated with the chronic infection of Hepatitis B (HBV) and Hepatitis C (HCV) viruses. In Japan, nearly 70% of HCC patients are infected with HCV [1]. The annual rate of developing HCC among patients with HCV-related liver cirrhosis in Japan is estimated to be about 4-8 percent [2]. Recent analyses have identified various genetic factors that are related with viral induced liver diseases [3-5]. In our previous two-stage genome-wide association study (GWAS) using a total number of 1,394 cases and 5,486 controls, a SNP rs2596542 located on chromosome 6p21.33 was shown to be significantly associated with HCV-induced HCC ($P = 4.21 \times 10^{-13}$ and OR = 1.39) [6]. This SNP is located within the class I major histocompatibility complex (MHC) region and is at about 4.8 kb upstream of *MHC class I polypeptide-related sequence A (MICA)* gene. We also identified that the risk A allele of SNP rs2596542 was strongly associated with the low expression of soluble MICA (sMICA) in the serum of HCV-related HCC patients [6].

MICA is a membrane protein which is up-regulated in various tumor cells and also induced in response to various cellular stresses such as infection, hypoxia, and heat shock [7]. It is an important component of the innate immune response, as MICA can

bind to the NKG2D receptor and subsequently activate natural killer (NK) cells, CD8+ cells, and $\gamma\delta$ T cells [8,9]. Moreover, membrane MICA can be shed by metalloproteinases, including MMP9, ADAM10, and ADAM17, and secreted into serum as a soluble form [10,11]. Since these metalloproteinases are often activated in HCC, the expressions of both membrane-bound MICA and sMICA are increased [12,13]. SNP rs2596542 was found to be associated with the progression from chronic hepatitis C (CHC) to HCC and also with serum sMICA level. Hence, both rs2596542 and sMICA would be possible prognostic biomarkers for CHC patients. However, their underlying molecular mechanisms were not fully elucidated so far.

We hypothesize that *MICA* variations could affect sMICA level by either one or both of the following two possible mechanisms: (1) the genetic variation(s) in the coding region affecting the protein stability and (2) the transcriptional regulation. Previously, variable numbers of tandem repeats (VNTRs) in exon 5 of *MICA* were identified to affect MICA subcellular localization and serum MICA level [14]. The exon 5 of *MICA* encodes the transmembrane domain and the insertion of an extra G nucleotide in the domain would result in a premature stop codon that would generate MICA protein without a transmembrane domain and subsequently affect sMICA level [14]. However, our previous results indicated that MICA VNTR was not significantly associated with the

sMICA level or HCC risk [6]. Therefore, in the current study, we have tried to investigate whether the *MICA* variations would affect the *MICA* transcription in the liver cancer cells. Through the functional analysis of genetic variations in the *MICA* promoter region, we here report a causative SNP rs2596538 that increases the binding affinity of the transcription factor Specificity Protein 1 (SP1) and the risk of progression of the disease.

Materials and Methods

Samples and genotyping

DNA samples for direct sequencing (50 HCV-related HCC cases), imputation analysis (721 HCV-related HCC cases and 5,486 HCV-negative controls), and serum samples for sMICA ELISA (246 HCV-related HCC) were obtained from BioBank Japan [15,16]. Genotyping of SNPs from 1,394 HCC patients and measurement of sMICA expression by ELISA were performed in the previous study [6]. Genotyping of SNP rs2596542 in 1,043 CHC was performed previously in RIKEN using Illumina HumanHap610-Quad BeadChip [17]. All CHC subjects had abnormal levels of serum alanine transaminase for more than 6 months and were positive for both HCV antibody and serum HCV RNA. The SNP rs2596542 in liver cirrhosis samples without hepatocellular carcinoma from

BioBank Japan (n = 420) and the University of Tokyo (n = 166) were genotyped using Illumina HumanHap610-Quad BeadChip or invader assay [18]. All subjects were either subjected to liver biopsy or diagnosed by non-invasive methods including hepatic imaging, biochemical data, and the presence/ absence of clinical manifestations of portal hypertension [18]. The samples used in the current project were listed in Supplementary Table 1. Case samples with HBV co-infection were excluded from this study. The subjects with cancers, chronic hepatitis B, diabetes or tuberculosis were excluded from non-HCV controls. All subjects were Japanese origin and provided written informed consent. This research project was approved by the ethical committees of the University of Tokyo and RIKEN.

Imputation study

The imputation study was performed by using a hidden Markov model programmed in MACH [19] and haplotype information from 1000 genomes database [20]. The imputation results were confirmed by direct DNA sequencing in 50 randomly selected samples.

Cell culture

Human liver cancer cell lines HLE and HepG2 were purchased from JHSF (Osaka, Japan) and ATCC. These cells were grown in Dulbecco's modified Eagle's medium (Invitrogen) with 10% fetal bovine serum. Cells were cultured at 37°C with 5% CO₂.

EMSA

HLE cells were grown in 15 cm culture plate until they reached 95% confluency. The plate was then sealed with parafilm and immersed in a water bath at 42.5°C for 1.5 hours [21]. Nuclear extracts from these cells were prepared according to the standard protocol [22]. EMSA was carried out using DIG Gel Shift Kit, 2nd Generation (Roche) according to the manufacturer's instructions. The sequences of the 12 probes were listed in the Supplementary Table 2. In brief, 30 fmol of labeled probes were hybridized with 5 µg nuclear extract for 15 minutes at room temperature. The mixtures were then loaded into a 6% TBE gel, separated by electrophoresis at 4°C and transferred onto a nylon membrane. The membrane was then hybridized with anti-digoxigenin-AP antibody and developed by CSPD solution. For competition study, nuclear extracts were incubated with non-labeled oligonucleotides first before adding labeled probe. For supershift assay, SP1 antibody (SC-59X, Santa Cruz Biotechnology) was added into the nuclear extract and incubated on ice for 30 minutes first before adding labeled probe. The mixtures

were then separated by electrophoresis using 4% TBE gel. All EMSAs were repeated twice for reconfirmation of the results.

ChIP

The HLE cells (G allele homozygote) and HepG2 cells (heterozygote) were used in the ChIP assay. The plasmid pCAGGS-SP1 was transfected into both cells by using FuGENE6 Transfection Reagent (Roche). The ChIP assays were carried out using Chromatin Immunoprecipitation Assay Kit (Millipore) according to the manufacturer's protocol. In brief, the cells were treated with formaldehyde to crosslink DNA-protein complexes at 48 hours post-transfection. DNA-protein complexes were then sheared by sonication and immunoprecipitated by rabbit polyclonal anti-SP1 antibody (SC-59X, Santa Cruz Biotechnology). The resulting DNAs were analyzed by PCR (Supplementary Table 2). In order to determine the binding specificity of SP1 to the SNP rs2596538 allele, the PCR products from HepG2 cells were further sub-cloned into pCR 2.1 vector and sequenced to assess G to A ratio in both input DNA and immunoprecipitant.

Dual luciferase reporter assay

Three copies of 31 bp DNA fragments equivalent to the EMSA oligonucleotides of SNP rs2596538 were cloned into pGL3-promoter vector (Promega). The plasmids were co-transfected with pCAGGS-SP1 and pRL-TK plasmids (Promega) into HLE cells by FuGENE6 Transfection Reagent (Roche). The pCAGGS-SP1 plasmid provided the expression of transcription factor SP1, and pRL-TK plasmid served as internal control for transfection efficiency [23]. The cells were lysed at 48 hours post-transfection, and relative luciferase activities were measured by Dual Luciferase Assay System (Toyo B-Net).

Western blotting

Cancer cell lysates were prepared by using pre-chilled RIPA buffer, and 25 µg of each lysate was loaded into the gel and separated by SDS-PAGE. Western blotting was performed according to the standard protocol. Rabbit anti-MICA antibody (ab63709, abcam: 1/1000) and rabbit anti-SP1 antibody (17-601, Upstate Biotechnology: 1/500) were used in the experiment.

Statistical analysis

The case-control association was analyzed by Student's *t*-test and Fisher's exact test as

appropriate. The association of allele dependent sMICA expression was studied by Kruskal-Wallis test using R statistical environment version 2.8.1. The LD and coefficients (D' and r^2) were calculated by Haploview version 4.2 [24].

Results

Analyses of SNP rs2596542 in HCV-infected patients at different disease stages

Since the development of HCC consists of multiple steps, we investigated the role of SNP rs2596542 with disease progression. SNP rs2596542 was genotyped in patients at three different disease categories of CHC (chronic hepatitis C) without liver cirrhosis (LC) or HCC, LC without HCC, and HCC. The statistical analysis indicated that SNP rs2596542 was significantly associated with disease progression from CHC to LC with P-value of 0.048 and odds ratio of 1.17 (Table 1). The risk allele frequency among HCC patients (40.1%) was higher than that among LC patients (38.0%), but the association was not statistically significant (P-value of 0.203 and odds ratio of 1.09). These results suggested the involvement of *MICA* with both liver fibrosis and hepatocellular carcinogenesis.

HCV-HCC risk is not associated with MICA copy number variation

A previous report has indicated the deletion of the entire *MICA* locus in 3.2% of Japanese population [25] and this deletion was shown to be associated with the risk of nasopharyngeal carcinoma (NPC), especially in male [26]. To identify the functional SNP that may affect *MICA* mRNA expression, we analyzed the relation between the *MICA* copy number variation (CNV) and the HCC susceptibility. We quantified this CNV by real-time PCR in 375 HCV-related HCC patients and 350 HCV-negative controls. As shown in Supplementary Table 3, we found no difference in the copy numbers between HCC cases and controls, indicating that this CNV is unlikely to be causative genetic variation for the risk of HCC.

Direct sequencing of 5' flanking region of *MICA*

We then focused on the variations in the 5' flanking region of the *MICA* gene which may be associated with its promoter activity. We had conducted direct DNA sequencing of the 5-kb promoter region which included the marker SNP rs2596542 using genomic DNAs of 50 HCC subjects and identified 11 SNPs showing strong linkage disequilibrium with the marker SNP rs2596542 ($D' > 0.953$ and $r^2 > 0.832$) (Supplementary Fig. 1, Table 2).

Allele specific binding of nuclear protein to genomic region including SNP rs2596538

To investigate whether these genetic variations would affect the binding affinity of some transcription factors, we had conducted the electrophoretic mobility shift assay (EMSA) using the nuclear extract of HLE human hepatocellular carcinoma cells. Since MICA is a stress-inducible protein [21], we first treated the cells with heat shock treatment at 42°C for 90 minutes and confirmed significant induction of MICA expression as shown in Fig. 1a. Then we performed EMSA using 24 labeled-oligonucleotides corresponding to each allele of the 12 candidates' SNPs. The results of EMSA demonstrated that an oligonucleotide corresponding to a G allele of SNP rs2596538 exhibited stronger binding affinity to a nuclear protein(s) than that to an A allele (Fig. 1b). We then confirmed the specific binding of nuclear proteins to the G allele by competitor assay using non-labeled oligonucleotides (Fig. 1c). The self (G allele) oligonucleotides inhibited the formation of DNA-protein complex in a dose-dependent manner, but the non-self (A allele) oligonucleotides showed no inhibition effect. Taken together, some nuclear protein(s) in hepatocellular carcinoma cells would interact with a DNA fragment including the G allele of SNP rs2596538.

SNP rs2596538 regulates the binding of SP1

Since *in silico* analysis identified a putative GC box in a protective G allele but not in a risk A allele (Fig. 2a), the transcription factor SP1 might preferentially bind to the G allele. Based on this information, we further performed competitor assay using non-labeled oligonucleotides (Supplementary Table 1) and found that among seven tested oligonucleotides, only SP1-consensus oligonucleotides could effectively inhibit the binding of the nuclear protein(s) to the labeled G allele (Fig. 2b). In addition, we identified that the addition of anti-SP1 antibody caused a supershift of a band corresponding to the DNA-protein complex while control IgG did not cause the band shift (Fig. 2c). This result clearly indicated that the SP1 protein is very likely to be a component of the DNA-protein complex.

Furthermore, we performed chromatin immunoprecipitation (ChIP) assay to confirm the binding of SP1 to this genomic region *in vivo*. We had used two cell lines with different genetic backgrounds at SNP rs2596538 locus: HLE cells carrying the only G allele, while HepG2 cells harboring both A and G alleles. After the introduction of SP1 expression vector (pCAGGS-SP1) into these cell lines, the cell extracts were subjected to ChIP assay using anti-SP1 antibody (Fig. 2d). Subsequent PCR experiments indicated that SP1 bound to a genomic fragment containing the G allele of SNP rs2596538 *in vivo*,

while 3' UTR region of *MICA* (negative control) was not immunoprecipitated with anti-SP1 antibody. To further evaluate the binding ability of SP1 to each allele *in vivo*, we sub-cloned the DNA fragment that amplified from genomic DNA of HepG2 cells before and after immunoprecipitation by anti-SP1 antibody. The subsequent sequencing results showed that 26 out of 29 tested clones contained the G allele, demonstrating the preferential binding of SP1 to the G allele (Fig. 2e).

SP1 over-expression preferentially up-regulates MICA expression at G allele

To further investigate the physiological role of the interaction between SP1 and this genomic region, we performed reporter gene assay. Three copies of 31-bp DNA fragments flanking the candidate functional SNP rs2596538 were subcloned into the multiple cloning sites of the pGL3 promoter vector. The relative luciferase activity of the plasmid including the G allele was significantly higher than that including the A allele (Fig. 3a). Furthermore, over-expression of SP1 in the cells could significantly enhance the luciferase activity of the G-allele vector, while the enhancement of the A-allele vector was relatively modest (Fig. 3a). We also evaluated the effect of ectopically expressed SP1 on the MICA expression in HLE cells. Western-blot analysis showed that MICA protein expression was significantly increased after the SP1

over-expression (Fig. 3b). These results provided a strong evidence that the G allele has higher transcriptional potential that can be inducible by SP1.

Association of SNP rs2596538 with HCC risk and sMICA level in HCV-induced HCC patients

To further investigate the role of SNP rs2596538 in human carcinogenesis, we investigated the association of SNP rs2596538 with HCV-induced HCC in 721 HCV-HCC cases and 5,486 HCV-negative controls that had been genotyped using Illumina HumanHap610-Quad Genotyping BeadChip in our previous study [6]. We performed imputation analysis by using haplotype data from 1000 genome database [20] and found that an A allele of SNP rs2596538 was considered to be a risk allele for HCV-related HCC (Table 3, odds ratio = 1.343, $P = 1.82 \times 10^{-5}$). The functional SNP rs2596538 exhibited a stronger association with the HCC risk than the marker SNP rs2596542 (2.46×10^{-5}). We also analyzed the relationship between the SNP rs2596538 and the sMICA level among 246 HCV-induced HCC patients and found a significant association with the P-value of 0.00616 (Fig. 4). These results were concordant with our functional analyses in which the G allele exhibited a higher affinity to SP1 and revealed