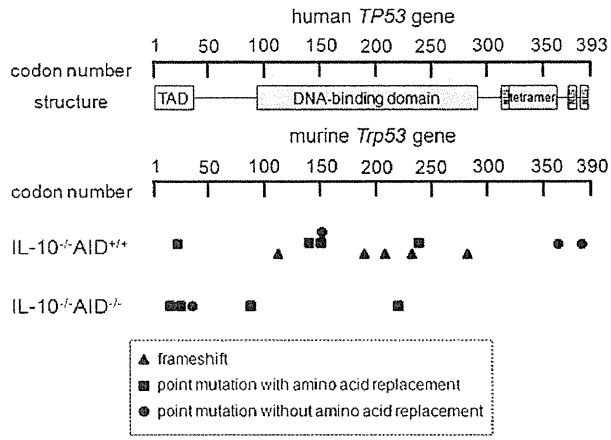

Figure 4 Distribution of murine *Trp53* mutations found in the cecal mucosa of IL-10$^{-/-}$AID$^{+/+}$ and IL-10$^{-/-}$AID$^{-/-}$ mice. Murine p53 codon numbers are shown with equivalent human p53 domain structure with transactivation (TAD), DNA-binding and tetramerization (tetramer) domains and nuclear localization signal (NLS).

mice, 7 (58.3%) of 12 genetic changes were single-base substitutions and 4 of these 7 alterations in the *Trp53* coding sequences resulted in amino-acid replacements with potential functional consequences. In contrast, the mutation frequency of the *Trp53* gene in the cecal epithelial cells of IL-10$^{-/-}$AID$^{-/-}$ mice (0.71 substitutions per 10^4 nucleotides) was significantly lower than that of IL-10$^{-/-}$AID$^{+/+}$ mice ($P<0.05$, Table 1). On the other hand, the *Apc*, *Ctnnb1* and *Kras* genes did not have remarkable numbers of nucleotide alterations in the inflamed cecal mucosa of IL-10$^{-/-}$AID$^{+/+}$ mice, and thus the incidence of nucleotide changes in the *Apc*, *Ctnnb1* and *Kras* genes was not significantly different between IL-10$^{-/-}$AID$^{+/+}$ and IL-10$^{-/-}$AID$^{-/-}$ mice (Table 1). These findings suggest that the *Trp53* gene is a specific target gene in chronically inflamed cecal mucosa in IL-10$^{-/-}$ mice, and the accumulation of genetic changes in the *Trp53* gene of the inflamed colonic mucosa was due to AID activity.

The incidence of colon cancer was reduced in IL-10$^{-/-}$ mice in the absence of endogenous AID

The findings that AID deficiency in IL-10$^{-/-}$ mice had no significant impact on the levels of colonic inflammation but reduced the frequencies of somatic mutations in the tumor-suppressor *Trp53* gene led us to speculate that the knockout of endogenous AID might reduce the incidence of colonic cancer development irrespective of ongoing colonic inflammation. Thus, we compared the neoplastic phenotype of the IL-10$^{-/-}$AID$^{+/+}$ mice with that of IL-10$^{-/-}$AID$^{-/-}$ mice. The frequency and spectrum of colonic tumors that developed in IL-10$^{-/-}$AID$^{+/+}$ mice and IL-10$^{-/-}$AID$^{-/-}$ mice are summarized in Table 2. Dysplastic changes in the mucosa of the large intestine were equally observed in most of these mice. These dysplastic lesions more frequently developed in the cecum than in the proximal and distal colon. Interestingly, invasive adenocarcinomas were detected

Table 2 Incidence of colonic tumors observed in IL-10$^{-/-}$AID$^{+/+}$ and IL-10$^{-/-}$AID$^{-/-}$ mice

	IL-10$^{-/-}$AID$^{+/+}$ (n=22)	*IL-10$^{-/-}$AID$^{-/-}$ (n=23)*
Mean age (weeks)	54.5	51.2
Male/female	13/9	13/10
Tumor formation		
Adenoma	20 (90.9%)	21 (91.3%)
Cecum[a]	20	20
Proximal colon[a]	1	1
Distal colon[a]	13	16
Adenocarcinoma	6* (27.2%)	1* (4.3%)
Cecum[a]	6	0
Proximal colon[a]	0	0
Distal colon[a]	0	1

Abbreviations: AID, activation-induced cytidine deaminase; IL-10, interleukin-10.
[a]Number of animals that developed adenoma or adenocarcinoma is shown.
*$P<0.05$ (*P*-value is calculated using Fisher's test).
Values in parentheses show percentages of animals that developed adenoma or adenocarcinoma.

in 6 of 22 IL-10$^{-/-}$AID$^{+/+}$ mice, and all the tumors characteristically developed from the dysplastic mucosa in the cecum (Table 2). Histopathological analysis of colonic tumors revealed moderate to poorly differentiated adenocarcinomas and invasive tumor cells beyond the submucosa with strong β-catenin expression (Figure 5a, Supplementary Figure 5). In contrast, IL-10$^{-/-}$AID$^{-/-}$ mouse developed no tumors in the inflamed colonic mucosa except only one tumor in the distal colon (Table 2). The colonic tumor that developed in the IL-10$^{-/-}$AID$^{-/-}$ mouse showed a trabecular pattern of growth within the submucosa, consistent with the morphologic appearance of well-differentiated adenocarcinoma (Figure 5b). These findings suggest that the upregulation of endogenous AID in the cecal mucosa driven by the inflammatory response contributes to the development of colonic cancers.

Discussion

The causal association between colonic inflammation and carcinogenesis is now well recognized (Eaden *et al.*, 2001; Podolsky, 2002). A recent genetic-linkage analysis of patients with IBD revealed that loss-of function mutations in genes encoding the IL-10 receptor proteins are associated with severe, early-onset enterocolitis, a finding that underscores the pivotal role of IL-10 in mediating the signals that control inflammation in the human gut (Glocker *et al.*, 2009). Consistent with the clinical finding, a mouse model with targeted disruption of the *IL-10* gene invariably develops enterocolitis that eventually progresses to colon cancer under conventional housing conditions; this mouse model is thus extremely useful as a disease model of human IBD (Kuhn *et al.*, 1993; Berg *et al.*, 1996; Sturlan *et al.*, 2001). In the present study, expression of AID was most prominent in the inflamed cecal mucosa of IL-10$^{-/-}$ mice. Moreover, we demonstrated that a deficiency of

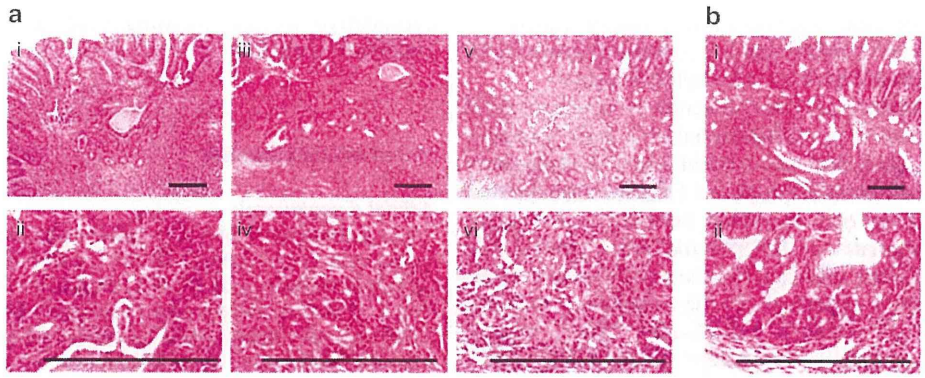

Figure 5 Colonic adenocarcinomas developed in IL-10$^{-/-}$ mice. (**a**) Microscopic images (hematoxylin and eosin (H&E) stain) of adenocarcinomas developed in the cecum of the IL-10$^{-/-}$AID$^{+/+}$ mice (i–vi). Scale bars are 500 μm. (**b**) Microscopic images (H&E stain) of adenocarcinoma developed in the distal colon of the IL-10$^{-/-}$AID$^{-/-}$ mouse (i, ii). Scale bars are 500 μm.

endogenous AID reduces the incidence of both the accumulation of somatic mutations in the *Trp53* gene and the development of colon cancer in inflamed colonic mucosa. *In vitro*, we previously demonstrated that aberrant AID expression is induced in response to proinflammatory cytokine stimulation, and colonic epithelial cells underlying chronic inflammation acquire the genetic mutations achieved by AID genotoxic activity (Endo *et al.*, 2008). Together, these findings suggest that inappropriate AID expression plays a pivotal role in the development of colorectal cancers via the accumulation of genetic alterations in the colonic mucosa of IBD.

We revealed here that endogenous AID is upregulated in inflamed colonic mucosa of elder IL-10$^{-/-}$ mice and the degree of AID expression paralleled extent of colonic inflammation. This observation is consistent with the findings that AID protein expression was detected in the colonic epithelium of inflammatory lesions from patients with IBD (Endo *et al.*, 2008). Colonic mucosal inflammation is usually mediated by either an excessive Th1 T-cell response associated with increased IFN-γ and IL-12 secretion or an excessive Th2 T-cell response associated with increased IL-4, IL-5 and IL-13 secretion (Fuss *et al.*, 1996). We previously found that the proinflammatory cytokine TNF-α, the Th2 cytokines IL-4 and IL-13 and Th1 cytokine IL-12 enhanced aberrant AID expression in cultured colonic epithelial cells (Endo *et al.*, 2008). On the other hand, TNF-α expression is elevated in colonic tissues of IL-10$^{-/-}$ mice (Berg *et al.*, 1996), and colitis in IL-10$^{-/-}$ mice is predominantly mediated by Th1-type T cells with increased production of IL-12 (Berg *et al.*, 1996; Davidson *et al.*, 1996). Consistent with these previous findings, in the present study, blockade of the activity of TNF-α or IL-12 suppressed AID expression in association with reduced production of various proinflammatory cytokines in the inflamed colonic mucosa of IL-10$^{-/-}$ mice. Thus, it is reasonable to assume that cytokine signalings, especially those mediated by TNF-α and IL-12, contribute to aberrant AID expression in the colonic cells of IL-10$^{-/-}$ mice.

A causal relationship between colonic inflammation and the accumulation of *TP53* tumor-suppressor gene mutations has been reported in human IBD (Yin *et al.*, 1993; Kern *et al.*, 1994; Hussain *et al.*, 2000; Leedham *et al.*, 2009). Alterations in the *TP53* gene, a late event in the pathogenesis of sporadic colorectal cancers, occur in dysplastic lesions with a background of ulcerative colitis (Yin *et al.*, 1993; Holzmann *et al.*, 1998) and are likely to proceed to dysplasia (Lashner *et al.*, 1999). Thus, the increased *TP53* mutation load in inflamed colonic epithelium of patients with IBD suggests that *TP53* mutations in noncancerous colon tissue of IBD patients specifically confer susceptibility to the development of colorectal cancers in an inflammatory microenvironment (Hussain *et al.*, 2000). In the present study, we found that high frequencies of nucleotide alterations had accumulated in *Trp53* gene mutation in inflamed mucosa of IL-10$^{-/-}$AID$^{+/+}$ mice. In addition, we demonstrated that a deficiency of endogenous AID in inflamed colonic mucosa resulted in a significantly reduced occurrence of somatic mutations in the *Trp53* genes, whereas no significant accumulation of somatic mutations appeared in the *Apc*, *Ctnnb1* and *Kras* genes in the inflamed colonic mucosa of IL-10$^{-/-}$AID$^{+/+}$ mice compared with IL-10$^{-/-}$AID$^{-/-}$ mice. It is unclear why the *Trp53* gene was more sensitive to AID-mediated genotoxic activity than the *Apc*, *Ctnnb1* and *Kras* genes in colonic epithelial cells of IL-10$^{-/-}$ mice. The present findings, however, are consistent with a previous observation that target gene selection for AID-induced somatic mutations varies among tissues and target cells (Morisawa *et al.*, 2008), and AID expression in cultured human colonic epithelial cells preferentially targets the *TP53* gene *in vitro* (Endo *et al.*, 2008). On the other hand, alterations in the *APC* and *KRAS* genes are also detected in dysplastic lesions and cancer tissues that develop in human IBD (Redston *et al.*, 1995). Therefore, we assume that the mutations in *APC*, *CTNNB1*, and *KRAS* genes were also present, but that their frequencies were below the detection limits of the present study. Further comprehensive sequencing analyses are required to determine how the AID-mediated genotoxic effects

achieve the target gene selection and whether IL-10$^{-/-}$ mice and human IBD share a similar process of mutational accumulation in tumor-related genes.

It is noteworthy that AID deficiency resulted in the reduced incidence of colitis-associated colon cancer development. AID deficiency caused the development of hyperplasia of isolated lymphoid follicles associated with an expansion of anaerobic flora in the small intestine (Fagarasan *et al.*, 2002; Suzuki *et al.*, 2004). We found no significant differences in the production levels of inflammatory cytokines in the colonic mucosa between the IL-10$^{-/-}$AID$^{+/+}$ and IL-10$^{-/-}$AID$^{-/-}$ mice. This observation might be consistent with the previous findings that activated B cells were not the primary mediator of inflammatory response in the colon of IL-10$^{-/-}$ mice, as evidenced by their ability to transfer colitis to immunodeficient RAG2$^{-/-}$ mice (Davidson *et al.*, 1996). In contrast to the similar levels of colonic inflammatory activity, the incidence of colon cancers was significantly lower in IL-10$^{-/-}$AID$^{-/-}$ mice compared with the IL-10$^{-/-}$AID$^{+/+}$ mice harboring endogenous AID. It may be emphasized that expression levels of endogenous AID in the cecal mucosa was significantly higher than those of the remaining sites of the colon, and all the colon cancers that developed in IL-10$^{-/-}$AID$^{+/+}$ mice were located at the cecum, whereas none of the IL-10$^{-/-}$AID$^{-/-}$ mice developed cancers in their cecum. Only one IL-10$^{-/-}$AID$^{-/-}$ mouse developed a tumor in the distal colon. Histological examination indicated that this tumor had the morphologic appearance of well-differentiated adenocarcinoma located within the submucosa, whereas all the cancers developed in IL-10$^{-/-}$AID$^{+/+}$ mice invaded the muscularis propria or adventitia with the characteristics of moderate to poorly differentiated adenocarcinomas. Based on the above discussion, ectopic AID expression in the inflamed colonic mucosa is an indispensable factor for the development of colon cancers in IL-10$^{-/-}$ mice.

Recent studies revealed that AID is involved in regulating DNA methylation in certain systems (Rai *et al.*, 2008; Bhutani *et al.*, 2010; Guo *et al.*, 2011). Moreover, infiltrating leukocytes, including B cells, might modulate tumor cell properties via the production of certain chemokines or cytokines (Ammirante *et al.*, 2010). Therefore, further studies are necessary to examine the incidence of inflammation-associated cancers in mice in which AID is specifically deficient in the epithelial cells, and to clarify whether AID has a role in inflammation-associated tumorigenesis through the epigenetic modification of tumor-related genes.

In conclusion, we demonstrated that the proinflammatory cytokine TNF-α and the Th1 cytokine IL-12 are responsible for aberrant AID expression in the colonic mucosa of IL-10$^{-/-}$ mice with chronic inflammation. Aberrant AID expression in the inflamed colon is associated with the accumulation of somatic mutations in tumor-suppressor *Trp53* gene, and AID deficiency resulted in a reduced incidence of colitis-associated colon cancers. These findings may lead to a novel strategy for preventing carcinogenesis by targeting AID irrespective of the ongoing colonic inflammation in patients with IBD.

Materials and methods

Animal experiments

The generation of AID$^{-/-}$ mice was described previously (Fagarasan *et al.*, 2001). IL-10$^{-/-}$ mice (The Jackson Laboratory, Bar Harbor, ME, USA) and AID$^{-/-}$ mice were crossed on a C57BL/6 background to generate IL-10$^{-/-}$AID$^{-/-}$ mice. All mice were maintained in a specific pathogen-free facility at Kyoto University Faculty of Medicine. Cecal and colonic epithelium was purified as follows: cecum and colon were cut into 2.0 cm long and incubated with 2 mM EDTA in Hank's balanced salt solution without calcium and magnesium for 10 min at room temperature. Then, the tissues were tumbled for 6 min and the mucosa was selectively stripped from the submucosa. The stripped mucosa was washed with phosphate-buffered saline three times and the supernatant containing floating cells and debris were discarded. The obtained epithelial tissue samples and nonepithelial tissue samples were frozen in liquid nitrogen for nucleotide extraction. In some experiments, IL-10$^{-/-}$ mice were intraperitoneally injected with TNF antagonist etanercept and neutralizing antibody to murine IL-12p40 (Watanabe *et al.*, 2006). Accordingly, 40-week-old IL-10$^{-/-}$ mice were injected with etanercept at a dose of 3 mg/kg body weight over 5 days and killed on day 12. Other 40-week-old IL-10$^{-/-}$ mice were injected with IL-12p40 mAb at a dose of 0.5 mg/body weight on days 1 and 8, and killed at day 12. All animal experiments were approved by the ethics committee for animal experiments and performed under the Guidelines for Animal Experiments of Kyoto University.

Histopathological and immunohistochemical analyses

The entire colon was removed and washed with phosphate buffered saline. The cecum, the proximal colon and the distal colon were dissected transversely and fixed in 4% (w/v) formaldehyde. The fixed tissue was embedded in paraffin and sectioned at 3 μm in a random manner. In particular, two types of histological preparations were sectioned from the tissues of the cecum. These samples were stained with hematoxylin and eosin and analyzed histologically in a blind fashion by three readers. Immunohistochemical staining was performed according to a previously described protocol (Toda *et al.*, 1999). The polyclonal antibodies for phospho-NF-κB p65 (Ser276) and β-catenin were purchased from Cell Signaling Technology (Danvers, MA, USA) and BD Biosciences (Franklin Lakes, NJ, USA), respectively.

In situ *hybridization*

A digoxigenin-labeled RNA probe specific for murine AID was transcribed with digoxigenin-11-UTP according to the manufacturer's instructions (Roche, Basel, Switzerland) from a 1.7-kb complementary DNA (cDNA) amplified using the following primers: 5′-ATGGACAGCCTTCTGGTGATGAA-3′ and 5′-CTTGTTCCCAAGGTCGCAAGGAAAGG-3′. Similarly, an RNA probe for murine villin1 was transcribed from a 1.6-kb cDNA amplified using the following primers: 5′-TGAATGCC CAACTCAAAGGCTCTCTC-3′ and 5′-ACCTCAAAAGCC TTGGTGTTATCAGC-3′. *In situ* hybridization was performed as described previously (Nakatani *et al.*, 2004). The alkaline phosphatase chromogen reaction was performed using Fast Red (Roche) as the substrate at room temperature for

48 h. The sections were then washed with distilled water and mounted on glass slides in mounting medium.

Semiquantitative and quantitative RT–PCR
Total RNA was extracted from the tissues using QuickGene RNA Tissue Kit (Fuji, Tokyo, Japan). cDNA was synthesized using Transcriptor First-Strand cDNA Synthesis Kit (Roche). PCR amplification was performed using Takara Ex Taq DNA polymerase (Takara, Tokyo, Japan). The oligonucleotide primers for the semiquantitative RT–PCR are shown in Supplementary Table 1. Gene expression was quantified by quantitative real-time RT–PCR using LightCycler 480 System II (Roche). The oligonucleotide primers for the quantitative RT–PCR are shown in Supplementary Table 2. To assess the quantity of isolated RNA as well as the efficiency of cDNA synthesis, target cDNAs were normalized to the endogenous mRNA levels of the housekeeping reference gene *18S rRNA* (Matsumoto *et al.*, 2007). For simplicity, ratios are represented as relative values compared with expression levels in lysate from control specimens.

Subcloning and sequencing analyses of tumor-related genes
The oligonucleotide primers for the amplification of the murine *Trp53*, *Apc*, *Ctnnb1* and *Kras* genes are shown in Supplementary Table 3. Amplification of targeted sequences was performed using high-fidelity Phusion Taq Polymerase (Finnzymes, Espoo, Finland), and the products were subcloned using pGEM-T Easy Vector Systems (Promega, Madison, WI, USA). The resulting plasmids were subjected to sequence analysis using Applied Biosystems 3500 Genetic Analyzer (Life Technologies, Carlsbad, CA, USA).

Statistical analysis
Statistical analysis was performed using the Mann–Whitney *U* test, χ^2 test and Fisher's test. Differences were considered to be statistically significant if *P*-values were <0.05.

Abbreviations

AID, activation-induced cytidine deaminase; cDNA, complementary DNA; IBD, inflammatory bowel disease; IFN, interferon; IL, interleukin; mAb, monoclonal antibody; NF, nuclear factor; RT–PCR, reverse transcription–PCR; TNF, tumor necrosis factor; *TP53*, *tumor protein p53*; WT, wild type.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Dr Keiichiro Suzuki and Dr Tasuku Honjo for critical reading of this manuscript, and Dr Masamichi Muramatsu for providing information of siRNA for AID. This work was supported by grants-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science; a grant-in-aid for Scientific Research from the Ministry of Health, Labor, and Welfare, Japan; and a research grant of the Princess Takamatsu Cancer Research Fund.

References

Ammirante M, Luo JL, Grivennikov S, Nedospasov S, Karin M. (2010). B-cell-derived lymphotoxin promotes castration-resistant prostate cancer. *Nature* **464**: 302–305.

Berg DJ, Davidson N, Kuhn R, Muller W, Menon S, Holland G *et al.* (1996). Enterocolitis and colon cancer in interleukin-10-deficient mice are associated with aberrant cytokine production and CD4(+) TH1-like responses. *J Clin Invest* **98**: 1010–1020.

Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SY, Blau HM. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**: 1042–1047.

Chiba T, Marusawa H. (2009). A novel mechanism for inflammation-associated carcinogenesis; an important role of activation-induced cytidine deaminase (AID) in mutation induction. *J Mol Med* **87**: 1023–1027.

Chiba T, Seno H, Marusawa H, Wakatsuki Y, Okazaki K. (2006). Host factors are important in determining clinical outcomes of Helicobacter pylori infection. *J Gastroenterol* **41**: 1–9.

Davidson NJ, Leach MW, Fort MM, Thompson-Snipes L, Kuhn R, Muller W *et al.* (1996). T helper cell 1-type CD4+ T cells, but not B cells, mediate colitis in interleukin 10-deficient mice. *J Exp Med* **184**: 241–251.

Eaden JA, Abrams KR, Mayberry JF. (2001). The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* **48**: 526–535.

Endo Y, Marusawa H, Kinoshita K, Morisawa T, Sakurai T, Okazaki IM *et al.* (2007). Expression of activation-induced cytidine deaminase in human hepatocytes via NF-kappaB signaling. *Oncogene* **26**: 5587–5595.

Endo Y, Marusawa H, Kou T, Nakase H, Fujii S, Fujimori T *et al.* (2008). Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers. *Gastroenterology* **135**: 889–898, 898 e1–3.

Fagarasan S, Kinoshita K, Muramatsu M, Ikuta K, Honjo T. (2001). In situ class switching and differentiation to IgA-producing cells in the gut lamina propria. *Nature* **413**: 639–643.

Fagarasan S, Muramatsu M, Suzuki K, Nagaoka H, Hiai H, Honjo T. (2002). Critical roles of activation-induced cytidine deaminase in the homeostasis of gut flora. *Science* **298**: 1424–1427.

Fearon ER, Vogelstein B. (1990). A genetic model for colorectal tumorigenesis. *Cell* **61**: 759–767.

Fuss IJ, Neurath M, Boirivant M, Klein JS, de la Motte C, Strong SA *et al.* (1996). Disparate CD4+ lamina propria (LP) lymphokine secretion profiles in inflammatory bowel disease. Crohn's disease LP cells manifest increased secretion of IFN-gamma, whereas ulcerative colitis LP cells manifest increased secretion of IL-5. *J Immunol* **157**: 1261–1270.

Glocker EO, Kotlarz D, Boztug K, Gertz EM, Schaffer AA, Noyan F *et al.* (2009). Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N Engl J Med* **361**: 2033–2045.

Guo JU, Su Y, Zhong C, Ming GL, Song H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**: 423–434.

Holzmann K, Klump B, Borchard F, Hsieh CJ, Kuhn A, Gaco V *et al.* (1998). Comparative analysis of histology, DNA content, p53 and Ki-ras mutations in colectomy specimens with long-standing ulcerative colitis. *Int J Cancer* **76**: 1–6.

Hussain SP, Amstad P, Raja K, Ambs S, Nagashima M, Bennett WP *et al.* (2000). Increased p53 mutation load in noncancerous colon tissue from ulcerative colitis: a cancer-prone chronic inflammatory disease. *Cancer Res* **60**: 3333–3337.

Ikeda K, Marusawa H, Osaki Y, Nakamura T, Kitajima N, Yamashita Y *et al.* (2007). Antibody to hepatitis B core antigen

and risk for hepatitis C-related hepatocellular carcinoma: a prospective study. *Ann Intern Med* **146**: 649–656.

Inoue S, Nakase H, Matsuura M, Mikami S, Ueno S, Uza N *et al.* (2009). The effect of proteasome inhibitor MG132 on experimental inflammatory bowel disease. *Clin Exp Immunol* **156**: 172–182.

Kern SE, Redston M, Seymour AB, Caldas C, Powell SM, Kornacki S *et al.* (1994). Molecular genetic profiles of colitis-associated neoplasms. *Gastroenterology* **107**: 420–428.

Komori J, Marusawa H, Machimoto T, Endo Y, Kinoshita K, Kou T *et al.* (2008). Activation-induced cytidine deaminase links bile duct inflammation to human cholangiocarcinoma. *Hepatology* **47**: 888–896.

Kou T, Marusawa H, Kinoshita K, Endo Y, Okazaki IM, Ueda Y *et al.* (2007). Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis. *Int J Cancer* **120**: 469–476.

Kuhn R, Lohler J, Rennick D, Rajewsky K, Muller W. (1993). Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75**: 263–274.

Lashner BA, Shapiro BD, Husain A, Goldblum JR. (1999). Evaluation of the usefulness of testing for p53 mutations in colorectal cancer surveillance for ulcerative colitis. *Am J Gastroenterol* **94**: 456–462.

Leedham SJ, Graham TA, Oukrif D, McDonald SA, Rodriguez-Justo M, Harrison RF *et al.* (2009). Clonality, founder mutations, and field cancerization in human ulcerative colitis-associated neoplasia. *Gastroenterology* **136**: 542–550 e6.

Lin WW, Karin M. (2007). A cytokine-mediated link between innate immunity, inflammation, and cancer. *J Clin Invest* **117**: 1175–1183.

Liu R, Bal HS, Desta T, Behl Y, Graves DT. (2006). Tumor necrosis factor-alpha mediates diabetes-enhanced apoptosis of matrix-producing cells and impairs diabetic healing. *Am J Pathol* **168**: 757–764.

Mantovani A, Allavena P, Sica A, Balkwill F. (2008). Cancer-related inflammation. *Nature* **454**: 436–444.

Marra F, Valente AJ, Pinzani M, Abboud HE. (1993). Cultured human liver fat-storing cells produce monocyte chemotactic protein-1. Regulation by proinflammatory cytokines. *J Clin Invest* **92**: 1674–1680.

Matsumoto Y, Marusawa H, Kinoshita K, Endo Y, Kou T, Morisawa T *et al.* (2007). Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat Med* **13**: 470–476.

Morisawa T, Marusawa H, Ueda Y, Iwai A, Okazaki IM, Honjo T *et al.* (2008). Organ-specific profiles of genetic changes in cancers caused by activation-induced cytidine deaminase expression. *Int J Cancer* **123**: 2735–2740.

Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**: 553–563.

Nakatani T, Mizuhara E, Minaki Y, Sakamoto Y, Ono Y. (2004). Helt, a novel basic-helix-loop-helix transcriptional repressor expressed in the developing central nervous system. *J Biol Chem* **279**: 16356–16367.

Neurath MF, Pettersson S, Meyer zum Buschenfelde KH, Strober W. (1996). Local administration of antisense phosphorothioate oligonucleotides to the p65 subunit of NF-kappa B abrogates established experimental colitis in mice. *Nat Med* **2**: 998–1004.

Podolsky DK. (2002). Inflammatory bowel disease. *N Engl J Med* **347**: 417–429.

Popivanova BK, Kitamura K, Wu Y, Kondo T, Kagaya T, Kaneko S *et al.* (2008). Blocking TNF-alpha in mice reduces colorectal carcinogenesis associated with chronic colitis. *J Clin Invest* **118**: 560–570.

Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**: 1201–1212.

Redston MS, Papadopoulos N, Caldas C, Kinzler KW, Kern SE. (1995). Common occurrence of APC and K-ras gene mutations in the spectrum of colitis-associated neoplasias. *Gastroenterology* **108**: 383–392.

Sturlan S, Oberhuber G, Beinhauer BG, Tichy B, Kappel S, Wang J *et al.* (2001). Interleukin-10-deficient mice and inflammatory bowel disease associated cancer development. *Carcinogenesis* **22**: 665–671.

Suzuki K, Meek B, Doi Y, Muramatsu M, Chiba T, Honjo T *et al.* (2004). Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proc Natl Acad Sci USA* **101**: 1981–1986.

Takai A, Toyoshima T, Uemura M, Kitawaki Y, Marusawa H, Hiai H *et al.* (2009). A novel mouse model of hepatocarcinogenesis triggered by AID causing deleterious p53 mutations. *Oncogene* **28**: 469–478.

Toda Y, Kono K, Abiru H, Kokuryo K, Endo M, Yaegashi H *et al.* (1999). Application of tyramide signal amplification system to immunohistochemistry: a potent method to localize antigens that are not detectable by ordinary method. *Pathol Int* **49**: 479–483.

Watanabe T, Kitani A, Murray PJ, Wakatsuki Y, Fuss IJ, Strober W. (2006). Nucleotide binding oligomerization domain 2 deficiency leads to dysregulated TLR2 signaling and induction of antigen-specific colitis. *Immunity* **25**: 473–485.

Yin J, Harpaz N, Tong Y, Huang Y, Laurin J, Greenwald BD *et al.* (1993). p53 point mutations in dysplastic and cancerous ulcerative colitis lesions. *Gastroenterology* **104**: 1633–1639.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc)

UICC
global cancer control

IJC
International Journal of Cancer

Excessive activity of apolipoprotein B mRNA editing enzyme catalytic polypeptide 2 (APOBEC2) contributes to liver and lung tumorigenesis

Shunsuke Okuyama, Hiroyuki Marusawa, Tomonori Matsumoto, Yoshihide Ueda, Yuko Matsumoto, Yoko Endo, Atsushi Takai and Tsutomu Chiba

Department of Gastroenterology and Hepatology, Graduate School of Medicine, Kyoto University, Shogoin, Sakyo-Ku, Kyoto, Japan

Apolipoprotein B mRNA editing enzyme catalytic polypeptide 2 (APOBEC2) was originally identified as a member of the cytidine deaminase family with putative nucleotide editing activity. To clarify the physiologic and pathologic roles, and the target nucleotide of APOBEC2, we established an APOBEC2 transgenic mouse model and investigated whether APOBEC2 expression causes nucleotide alterations in host DNA or RNA sequences. Sequence analyses revealed that constitutive expression of APOBEC2 in the liver resulted in significantly high frequencies of nucleotide alterations in the transcripts of eukaryotic translation initiation factor 4 gamma 2 (*Eif4g2*) and phosphatase and tensin homolog (*PTEN*) genes. Hepatocellular carcinoma developed in 2 of 20 APOBEC2 transgenic mice at 72 weeks of age. In addition, constitutive APOBEC2 expression caused lung tumors in 7 of 20 transgenic mice analyzed. Together with the fact that the proinflammatory cytokine tumor necrosis factor-α induces ectopic expression of APOBEC2 in hepatocytes, our findings indicate that aberrant APOBEC2 expression causes nucleotide alterations in the transcripts of the specific target gene and could be involved in the development of human hepatocellular carcinoma through hepatic inflammation.

The number of coding sequences in the genome is limited, but the genomic information encoded in DNA or RNA sequences can be manipulated to produce a wide range of expression products in cells.[1] Apolipoprotein B mRNA editing enzyme catalytic polypeptide (APOBEC) family members are nucleotide-editing enzymes capable of inserting somatic mutations in DNA and/or RNA through their cytidine deaminating activity.[2] The APOBEC family comprises APOBEC1, -2, -3A, -3B, -3C, -3DE, -3F, -3G, -3H, -4, activation-induced cytidine deaminase (AID) in humans, and APOBEC1, -2, -3, and AID in mice, and contribute to producing various physiologic outcomes by modifying target gene sequences.[3–5] For example, APOBEC1 participates in lipid metabolism by deaminating a specific cytidine to uridine in Apolipoprotein (Apo-) B transcript sequences. The nucleotide change induced by APOBEC1 activity results in the formation of a termination codon in an Apo-B48 mRNA, leading to the production of molecules about half the size of a full-length genomically encoded Apo-B100.[6,7] APOBEC3G is a cytidine deaminase that induces hypermutation in viral DNA sequences and acts as a host defense factor against various viruses, including HIV-1 and hepatitis B viruses.[8–15] On the other hand, AID is expressed in germinal center B-cells and induces somatic hypermutation and class switch recombination of the immunoglobulin genes encoded in human DNA sequences, resulting in the amplification of immune diversity.[16,17] APOBEC1, APOBEC3G and AID thus create nucleotide changes in their preferential target DNA or RNA structures. In contrast to these APOBEC proteins, little is known about the function and editing activity of APOBEC2. Although previous reports indicate that murine APOBEC2 mRNA and protein are expressed exclusively in heart and skeletal muscle, the substrate and function of APOBEC2 and whether APOBEC2 has nucleotide editing activity remain unknown.[18,19]

Accumulating evidence suggests that excessive or aberrant activity of APOBEC family members leads to tumorigenesis through their nucleotide editing of tumor-related genes.

Key words: APOBEC2, hepatocellular carcinoma, lung cancer

Abbreviations: APOBEC: Apolipoprotein B mRNA editing enzyme catalytic polypeptide; EIF4G2: Eukaryotic translation initiation factor 4 gamma 2; AID: activation-induced cytidine deaminase; Apo-: Apolipoprotein; Tg: transgenic; NF-κB: nuclear factor-κB; HCC: hepatocellular carcinoma; TNF: tumor necrosis factor; cDNA: Complimentary DNA; RT-PCR: real-time reverse-transcription polymerase chain reaction; ER: estrogen receptor

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

Grant sponsors: Japan Society for the Promotion of Science (JSPS), a Grant from the Ministry of Health, Labor, and Welfare, Japan, the Takeda Science Foundation

DOI: 10.1002/ijc.26114

History: Received 8 Jan 2011; Accepted 25 Mar 2011; Online 5 Apr 2011

Correspondence to: Hiroyuki Marusawa, MD, PhD, Department of Gastroenterology and Hepatology, Graduate School of Medicine, Kyoto University, 54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan, Tel.: +81-75-751-4302, Fax: +81-75-751-4303, E-mail: maru@kuhp.kyoto-u.ac.jp

Transgene expression of APOBEC1 causes dysplasia and carcinoma in mouse and rabbit liver due to its aberrant editing of the eukaryotic translation initiation factor 4 gamma 2 (Eif4g2).[20,21] A more striking tumor phenotype is observed in mice with constitutive and ubiquitous AID expression. We previously demonstrated that AID transgenic (Tg) mice developed tumors in various organs, including liver, lung, stomach and lymphoid organs, accompanied by the accumulation of somatic mutations on several tumor-related genes such as *Tp53* and *Myc*.[22,23] Interestingly, we also found that proinflammatory cytokine stimulation induces a substantial upregulation of APOBEC2 transcription *via* the activation of the transcriptional factor nuclear factor-κB (NF-κB) in hepatoma-derived cells, whereas only trace amounts of endogenous APOBEC2 expression are detectable in normal hepatocytes.[24] On the basis of the fact that most human hepatocellular carcinoma (HCC) arises in the setting of chronic liver disease with the features of chronic hepatitis or liver cirrhosis, we hypothesized that APOBEC2 enzyme activity has a role in the accumulation of genetic alterations in tumor-related genes under conditions of hepatic inflammation, thereby contributing to the development of HCC. In this study, we investigated the putative nucleotide editing ability of APOBEC2 on the host genes in hepatocytes, and its relevance to carcinogenesis by establishing Tg mice that constitutively express APOBEC2.

Material and Methods

APOBEC2 Tg mice

Total RNA was extracted from murine liver using Sepasol-RNA 1 Super (Nacalai Tesque, Kyoto, Japan) according to the manufacturer's protocol. Complimentary DNA (cDNA) was synthesized from total RNA with random hexamer primers using a Transcriptor First Strand cDNA Synthesis Kit (Roche, Mannheim, Germany). After amplification of the murine APOBEC2 gene using high-fidelity Phusion Taq polymerase (Finnzymes, Espoo, Finland) with oligonucleotide primers, 5′-GCAGAATTCCACCATGGCTCAGAAGGAAG AGGC-3′(forward) and 5′-ACTCTCGAGCCTACTTCAGGA TGTCTGCC-3′ (reverse), murine APOBEC2 cDNA (1.2 kbp) was cloned downstream of the chicken β-actin (CAG) promoter. The purified fragment of the CAG promoter and APOBEC2 transgene was microinjected into fertilized eggs of the Slc:BDF1, the hybrid of C57BL/6CrSlc and DBA/2CrSlc (Japan SLC, Shizuoka, Japan), to generate APOBEC2 Tg mice. Tg mice were maintained in specific pathogen-free conditions at the Institute of Laboratory Animals of Kyoto University. Control mice were littermates carrying no transgene. Tissue samples from Tg mice were removed and fixed in 4% (w/v) formaldehyde, embedded in paraffin, stained with hematoxylin and eosin and examined for histologic abnormalities. Tissue samples were also frozen immediately in liquid nitrogen for nucleotide extraction. The mice received humane care according to the "Guide for the Care and Use of Laboratory Animals" prepared by the National Academy of Sciences and published by the National Institutes of Health, USA (NIH publication 86-23).

Quantitative real-time reverse transcription PCR

Quantitative real-time reverse-transcription polymerase chain reaction (RT-PCR) for murine *APOBEC1* and *APOBEC2* amplification was performed using a LightCycler® 480 instrument (Roche). cDNA was synthesized from 1 μg of total RNA isolated from the cells with random hexamer primers in a total volume of 20 μL using Transcriptor First Strand cDNA Synthesis Kit (Roche). Real-time PCRs were set up in 20 μL of FastStart Universal SYBR Green (Roche) with the RT product and the following oligonucleotide primers: APOBEC1, 5′-CGAAGCTTATTGGCCAAGGT-3′ (forward) and 5′-AAGGAGATGGGGTGGTATCC-3′ (reverse); APOBEC2, 5′-CCCTTCGAGATTGTCACTGG-3′ (forward) and 5′-TGTTCATCCTCCAGGTAGCC-3′ (reverse). Target cDNAs were normalized to the endogenous RNA levels of the housekeeping reference gene for *18S ribosomal RNA* (*18S rRNA*).[25] For simplicity, the expression levels of *APOBEC2* are represented as relative values compared with the control specimen in each experiment.

Immunoblotting

Homogenates of murine specimens were diluted in 2× sodium dodecyl sulfate sample buffer (62.5 mM Tris-HCl, pH 6.8; 2% SDS; 5% β-mercaptoethanol; 10% glycerol, and 0.002% bromophenol blue) and boiled for 5 min. Protein samples were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis on 12% (w/v) polyacrylamide gels and subjected to immunoblotting analysis.[26] A polyclonal antibody against human and murine APOBEC2 was generated using purified recombinant APOBEC2 protein as an immunogen. A mouse monoclonal antibody against α-tubulin was purchased from Calbiochem (San Diego, CA).

Cell culture and transfection

Human hepatoma-derived cell lines HepG2 and Huh7 were maintained in Dulbecco's modified Eagle's medium (Gibco-BRL) containing 10% fetal bovine serum. Trans-IT 293 transfection reagent (Mirus Bio Corporation, Madison, WI) was used for plasmid transfection. To generate stable cell lines, pcDNA3-ERT2 was made by inserting the ERT2 fragment, which was cut out from pERT2[27] with *Bam*HI and *Eco*RI. pcDNA3-APO2-ERT2 was made by inserting the PCR-amplified coding sequence of human *APOBEC2*, which was synthesized by RT-PCR with the oligonucleotide primers 5′-ATAGG TACCATGGCCCAGAAGGAAGAGGC-3′ (forward) and 5′-ATAGGATCCAGCTTCAGGATGTCTGCCAAC-3′ (reverse), into the *KpnI-BamHI* site of pcDNA3-ERT2. HepG2 cells were transfected with a *Sca*I-linearized pcDNA3-APO2-ERT2 vector encoding the active form of APOBEC2 fused with the hormone-binding domain of the human estrogen receptor (ER), designated APOBEC2-ER, and cultured in medium

containing G418 (Roche) until colonies of stably transfected clones arose.

Subcloning and sequencing of the target genes

The oligonucleotide primers for the amplification of the human *EIF4G2*, *PTEN*, and *TP53*, and murine *Eif4G2*, *Pten*, *Bcl6* and *Tp53*, genes are shown in Supporting Information Table S1. Amplification of the target sequences was performed using high-fidelity Phusion Taq polymerase (Finnzymes, Espoo, Finland), and the products were subcloned into a pcDNA3 vector (Invitrogen, Carlsbad, CA) using pGEM(R) -T Easy Vector System (Promega, Madison, WI) according to the manufacture's instruction. The resulting plasmids were subjected to sequence analysis as described.[28]

Results

Detection of endogenous APOBEC2 protein expression in hepatocytes

We previously reported that transcription of *APOBEC2* is induced by the proinflammatory cytokine tumor necrosis factor (TNF)-α through the activation of NF-κB. To confirm whether endogenous APOBEC2 protein is elevated in response to TNF-α stimulation in human hepatocytes, we generated a rabbit polyclonal antibody against a common amino-acid sequence to human and murine APOBEC2. Using this anti-APOBEC2 antibody, we first confirmed that plasmid-derived exogenous APOBEC2 protein was efficiently detected by immunoblotting analysis (Fig. 1*a*). We then examined whether endogenous APOBEC2 protein was upregulated by TNF-α stimulation in Huh-7 cells. Immunoblotting analysis using the APOBEC2 antibody revealed that endogenous APOBEC2 protein expression was strongly induced after TNF-α stimulation, suggesting that APOBEC2 protein has a role in hepatocyte function under inflammatory conditions (Fig. 1*b*).

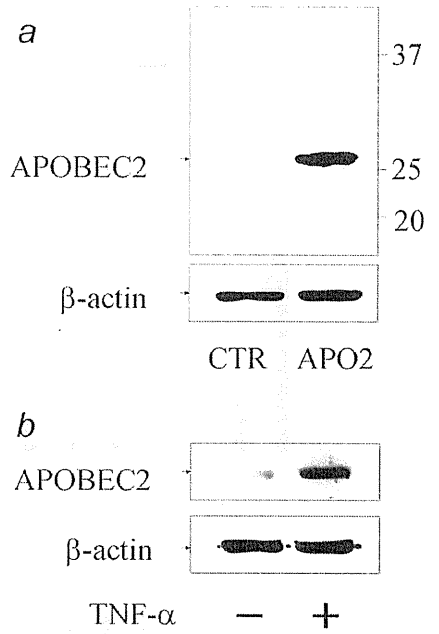

Figure 1. Detection of human APOBEC2 protein in hepatocytes by a specific anti-APOBEC2 antibody. (*a*) Huh7 cells were transfected with plasmid to induce the expression of human APOBEC2 (APO2) or control vector (CTR). After 48 hr, lysates of transfected cells were immunoblotted with anti-APOBEC2 antibody (upper panel) or anti-β-actin antibody (lower panel). (*b*) Huh7 cells were treated with tumor necrosis factor-α (100 ng/ml) for 48 hr followed by immunoblotting using anti-APOBEC2 antibody (upper panel) or anti-β-actin antibody (lower panel).

Establishment of a Tg mouse model constitutively expressing APOBEC2

To investigate the enzymatic activity of APOBEC2 *in vivo*, we generated a Tg mouse model with constitutive and ubiquitous expression of APOBEC2 under the control of CAG promoter. APOBEC2 Tg mice were born healthy and with a body weight similar to that of their wild-type littermates. The expression level of APOBEC2 in various organs of the Tg mice was examined by quantitative RT-PCR and compared with that in the wild-type mice. In wild-type mice, endogenous APOBEC2 transcript was expressed at high levels in heart and skeletal muscle, whereas little or no APOBEC2 expression was detected in the liver, gastrointestinal tracts, lung, spleen and kidney. In contrast, high expression of *APOBEC2* mRNA was ubiquitously detected in the Tg mice, but the expression levels of *APOBEC2* in the liver or lung of the Tg mice were relatively lower than those of the wild-type heart or skeletal muscle (Fig. 2*a*). Immunoblotting analysis using the specific antibodies against APOBEC2 also revealed widespread expression of APOBEC2 protein in various epithelial organs of the Tg mice, with relatively low expression in kidney and spleen (Fig. 2*b*).

Constitutive expression of APOBEC2 resulted in the accumulation of nucleotide alterations in RNA sequences of Eif4g2 and Pten genes in hepatocytes

To clarify whether APOBEC2 targets DNA or RNA, we first extracted total RNA from the nontumor liver tissues of 2 APOBEC2 Tg mice that developed HCC (described below) and their 3 APOBEC2 Tg littermates without any tumor phenotypes, and subjected them to sequence analyses. We chose 2 representative tumor-suppressor genes that are frequently mutated in human cancers, *Pten*, and *Tp53*. The *Bcl6* and *Eif4g2* genes were also included because they are the preferential targets for AID- and APOBEC1-mediated mutagenesis, respectively. We first confirmed that the transcription levels of the genes analyzed for RNA sequencing did not differ between the liver tissues of APOBEC2 Tg mice and wild-type littermates (Supporting Information Fig. S1). In addition, there was no difference in the quantitative levels of APOBEC1 expression between the APOBEC2-expressing liver and

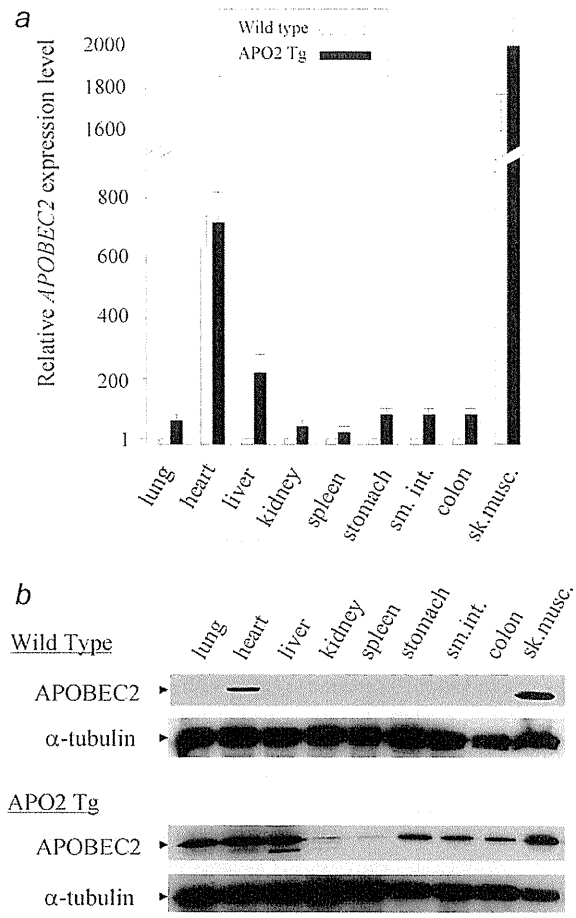

Figure 2. Expression analyses of APOBEC2 Tg mice. (*a*) Relative expression levels of *APOBEC2* transcripts calibrated by the amount of *18S rRNA* for indicated organs of adult APOBEC2 Tg mice (48-week-old) and their wild-type littermates. Data shown are mean results of quantitative real-time RT-PCR analyses for the indicated mouse groups ($n = 6$). Filled bar, APOBEC2 Tg mice; open bar, wild-type mice; sm.int, small intestine; sk.musc, skeletal muscle. (*b*) Results of immunoblot analysis using anti-APOBEC2 (upper panel) or anti-α-tubulin (lower panel) antibody for the lysates of the indicated organs of 48-week-old APOBEC2 Tg mice and their littermates.

normal liver of the wild-type mice (Supporting Information Fig. S2). Sequence analysis revealed a mean of 98,000 and 55,400 base reads per each gene transcript derived from the nontumor liver tissues of the APOBEC2 Tg and control mice, respectively. The total number of amplified clones and RNA sequence reads, and the frequency of nucleotide alterations detected in the nontumor liver tissues of 2 APOBEC2 Tg mice with HCC and the wild-type littermate of the same mouse line are shown in Table 1. The mutation frequencies were highest in the *Eif4g2* transcripts among the genes analyzed in APOBEC2-Tg mice, and were significantly greater compared with those in control tissues (mutation frequencies were 2.75 and 2.36 *vs.* 0.58 substitutions per 1×10^4 nucleotides; $p < 0.05$). Moreover, the nucleotide alteration frequency was significantly higher in the *Pten* gene transcripts from a APOBEC2-expressing liver (Tg-1) than in the control tissues (mutation frequencies were 2.43 *vs.* 0.44 substitutions per 1×10^4 nucleotides, respectively; $p < 0.01$). The *Pten* mRNA of a liver derived from another APOBEC2 Tg mouse (Tg-2; mutation frequency was 1.36 substitutions per 1×10^4 nucleotides) also had a higher nucleotide alteration frequency than that in the control mice, although the difference was not statistically significant ($p = 0.16$ *vs.* control). For the *Eif4g2* and *Pten* transcripts, nucleotide alterations were distributed over the sequences examined and all the alterations detected were different among clones (Fig. 3). Similar results were obtained from the analyses on the liver of 3 APOBEC2 Tg mice that lacked any tumor phenotypes. Indeed, several nucleotide changes had accumulated in both *Eif4g2* and *Pten* transcripts in the liver of all 3 APOBEC2 Tg mice examined (Supporting Information Table S2). In contrast, the mutation frequencies of *Tp53* and *Bcl6* genes of the liver of the APOBEC2 Tg mice were comparable with those of the wild-type mice.

APOBEC2 expression in the liver induced no nucleotide changes in DNA sequences

To clarify whether the nucleotide alterations that emerged in *Eif4g2* and *Pten* transcripts were due to DNA or RNA sequence changes, we determined the DNA sequences of both genes derived from the liver tissues of APOBEC2 Tg and control mice. DNA sequences with an average base length of 0.7 k containing exonic and intronic sequences were amplified, followed by sequence analyses. The total number of amplified clones and DNA sequences read, and the frequency of nucleotide alterations are shown in Supporting Information Table S3. In contrast to the analyses on the RNA sequences, there were no significant differences between the mutation frequency of APOBEC2 Tg mice and that of the wild-type mice of the DNA sequences of the *Eif4g2* and *Pten* genes in the liver. Indeed, no nucleotide alterations were observed in the DNA sequences of the *Eif4g2* gene in the liver of the APOBEC2 Tg mice. Similarly, no mutation was accumulated in the *Pten* DNA sequences of the APOBEC2-expressing liver, suggesting that constitutive expression of the APOBEC2 transgene had no effect on the DNA sequences of the examined regions in the *Eif4g2 and Pten* genes in hepatocytes.

APOBEC2 transgenic mice developed liver and lung tumors

Although most Tg mice were viable at 72 weeks, macroscopic liver and lung tumors developed in some of the APOBEC2 Tg mice. At 72 weeks of age, liver tumors were observed in 2 of 20 Tg male mice, and lung nodules were detected in 7 Tg mice. In contrast to the APOBEC2 Tg mice, none of the wild-type mice developed any tumors at the same age, except 1 with a very small adenoma in the lung. Histopathologic

Table 1. Summary of sequence analysis on the RNA extracted from the liver of the wild-type and APOBEC2 Tg mice

				Nucleotide alterations		
Gene	Mice	Clone	Sequence reads	Number	Frequency(/10^4)	APO2/Wt*
Eif4g2	Wt	82	50,949	3	0.58	
	Tg-1	83	50,835	14	2.75	4.7**
	Tg-2	90	54,986	13	2.36	4.1**
Pten	Wt	92	67,352	3	0.44	
	Tg-1	79	57,599	14	2.43	5.5***
	Tg-2	69	51,323	7	1.36	3.1
Bcl6	Wt	48	41,776	3	0.72	
	Tg-1	59	51,414	1	0.19	0.3
	Tg-2	48	42,413	4	0.94	1.3
Tp53	Wt	84	61,705	2	0.32	
	Tg-1	51	42,285	3	0.71	2.2
	Tg-2	50	40,880	3	0.73	2.3

*Frequency of nucleotide alteration in APOBEC2 Tg mice / in wild type mice. **$p < 0.05$, *vs.* Wt. ***$p < 0.01$, *vs.* Wt.
Abbreviations: Tg, APOBEC2 Tg mice; WT, wild type mice.

analysis of hepatic tumors developed in the APOBEC2 Tg mice revealed nodular aggregates of neoplastic hepatocytes and permeation of tumor cells into residual normal lobules (Fig. 4). Tumor cells had enlarged and hyperchromatic nuclei with chromatin clumping and occasional prominent nucleoli, which were similar to the morphologic characteristics of typical human HCC. On the other hand, lung tumors showed various degrees of cellular atypia, from adenoma to adenocarcinoma (Fig. 5*a*). In addition, monotonous atypical lymphocytes with cytologic features of lymphoblastic lymphoma, such as enlarged round nuclei, irregular nuclear contours, and frequent mitotic figures, massively invaded the spleens of 2 Tg mice (Fig. 5*b*). These findings suggest that constitutive expression of APOBEC2 causes the development of neoplasia in the epithelial organs, including the liver and the lung.

APOBEC2 induced the accumulation of nucleotide alterations of specific target RNA sequences in hepatocytes in vitro

To confirm whether APOBEC2 exerts genotoxic effects on RNA transcripts of the specific target genes, we investigated the alteration frequencies of RNA sequences in cells with constitutive APOBEC2 expression. For this purpose, we established a conditional expression system that allowed for APOBEC2 activation in the cells in response to an estrogen analogue, 4-hydroxytamoxifen (OHT). OHT treatment triggered a posttranslational conformational change and prompt activation of APOBEC2 in APOBEC2-ER expressing cells.[29] We analyzed 3 genes including *PTEN*, *TP53* and *EIF4G2* for the sequence analysis of APOBEC2-mediated mutagenesis *in vitro*. Total RNA was extracted from the APOBEC2-ER expressing HepG2 cells treated with OHT for 8 weeks and the coding RNA sequences of the selected genes were determined by sequence analyses. The total number of amplified clones and RNA sequence reads, and the frequency of nucleotide alterations are shown in Supporting Information Table S4. We found that the emergence of nucleotide alterations in the *PTEN* and *EIF4G2* transcripts was detected at higher frequencies in the cells with APOBEC2 activation compared with control cells treated with OHT, while these differences were not statistically significant ($p = 0.23$ *vs.* control, and $p = 0.39$ *vs.* control, respectively). In contrast, the frequency of nucleotide alterations in the transcripts of the *TP53* in the cells with APOBEC2 activation was comparable with that in the control cells. Similar to the findings obtained from the APOBEC2 Tg mice liver tissues, there were no significant differences between APOBEC2-expressing hepatocytes and control cells in the incidence of nucleotide alterations in the *PTEN* and *EIF4G2* genes (Supporting Information Table S5). These data further suggest that APOBEC2 exerts mutagenic activity in hepatocytes and preferentially achieves nucleotide substitutions in the coding sequences of the specific target genes.

Discussion

Among the APOBEC family members, APOBEC2 and AID homologs can be traced back to bony fish, whereas APOBEC1 and APOBEC3s are restricted to mammals.[30,31] The broad preservation of the APOBEC2 homolog among vertebrates suggests that APOBEC2 has a critical role in the physiology of many species. Little is currently known, however, about the biologic activity of APOBEC2 in any type of cells. Moreover, it is not known whether APOBEC2 possesses nucleotide editing activities like other APOBEC family member proteins. In the present study, we demonstrated for the first time that APOBEC2 expression triggered nucleotide alterations in RNA sequences of the specific genes in hepatocytes. In addition, our findings suggest that APOBEC2 could

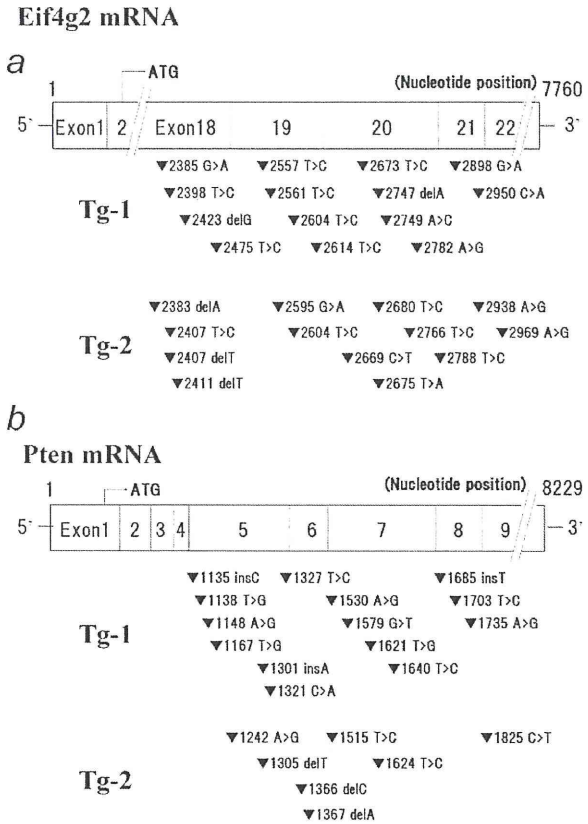

Figure 3. Distribution of nucleotide alterations in the Eif4g2 and Pten trascripts in the APOBEC2-expressing hepatocytes. The mRNA sequences between exon 18 and exon 21 of the Eif4g2 gene (*a*), and the mRNA sequences between exon 5 and exon 8 of the Pten gene (*b*) were determined in the nontumor liver tissues of 2 APOBEC2 Tg mice. The nucleotide positions of the mutations emerged in the Eif4g2 and Pten mRNA of APOBEC2-expressing liver are shown.

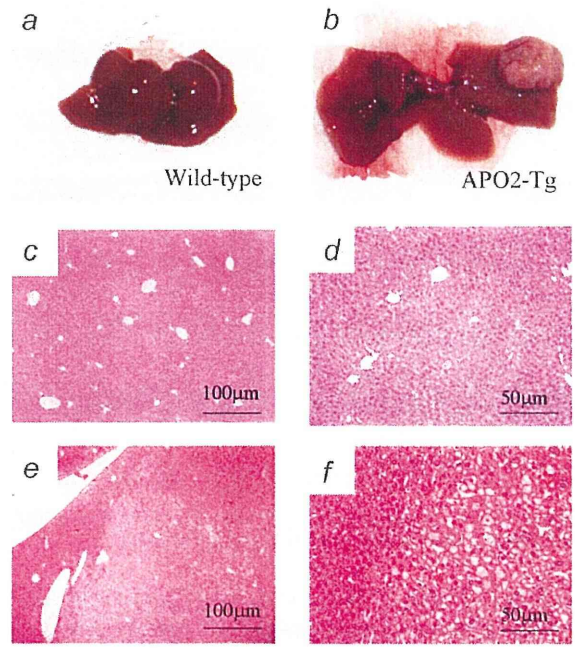

Figure 4. Tumors developed in the liver of APOBEC2 Tg mice. Macroscopic (*b*) and microscopic (haematoxylin and eosin) images (*e, f*) of the HCC that developed in a 72-week-old APOBEC2 Tg mouse and the non-cancerous liver of the same animal (*c,d*). Macroscopic image of the liver of a wild-type littermate is also shown (*a*). (Original magnifications: 3*c,e* ×40; 3*d,f* ×100).

contribute to tumorigenesis via the nucleotide alterations of RNA sequences of the target genes.

On the basis of the close sequence homology of APOBEC2 with other APOBEC proteins, APOBEC2 is thought to exhibit deamination activity to achieve nucleotide editing. Indeed, crystal structure analysis indicates that APOBEC2 contains amino acid residues with 4 monomers in each asymmetric unit that form a tetramer with an atypical elongated shape, and this prominent feature of the APOBEC2 tetramer suggests that the active sites are accessible to large RNA or DNA substrates.[32] In the present study, in a mouse model with constitutive APOBEC2 expression, nucleotide alterations were induced in RNA sequences of the *Eif4g2* and possibly the *Pten* genes in hepatocytes. Similar to its effect *in vivo*, aberrant APOBEC2 expression in cultured hepatocyte-derived cells induced nucleotide alterations in the *EIF4G2* transcripts. Although our findings demonstrate potential mutator activity of the APOBEC2 protein, it is unclear why the *EIF4G2* transcripts were more sensitive to APOBEC2 activity than other genes in hepatocytes. APOBEC1 expression in hepatocytes also induced somatic mutations in the transcripts of the *EIF4G2* gene.[21] Thus, the sequences of the *EIF4G2* gene might be a common target for the nucleotide editing effects of both the APOBEC1 and APOBEC2 proteins. Further analysis is required to identify the specific target genes of APOBEC2-mediated nucleotide editing in hepatocytes.

An intriguing finding was that the mouse model with constitutive and ubiquitous APOBEC2 expression spontaneously developed epithelial neoplasia in the lung and liver tissues as well as lymphoma. Similar phenotypic findings are observed in mouse models expressing APOBEC1 or AID. Tg mice with RNA-editing enzyme APOBEC1 expression develop HCC at high frequencies with an accumulation of somatic mutations at multiple sites on *Eif4g2* mRNA.[20,21] We also demonstrated that AID Tg mice develop tumors in several organs, including the liver, lung, stomach, and the lymphoid tissues through the accumulation of genetic changes induced by the genotoxic effect of AID.[22,23,28] The molecular mechanisms underlying the contribution of constitutive APOBEC2

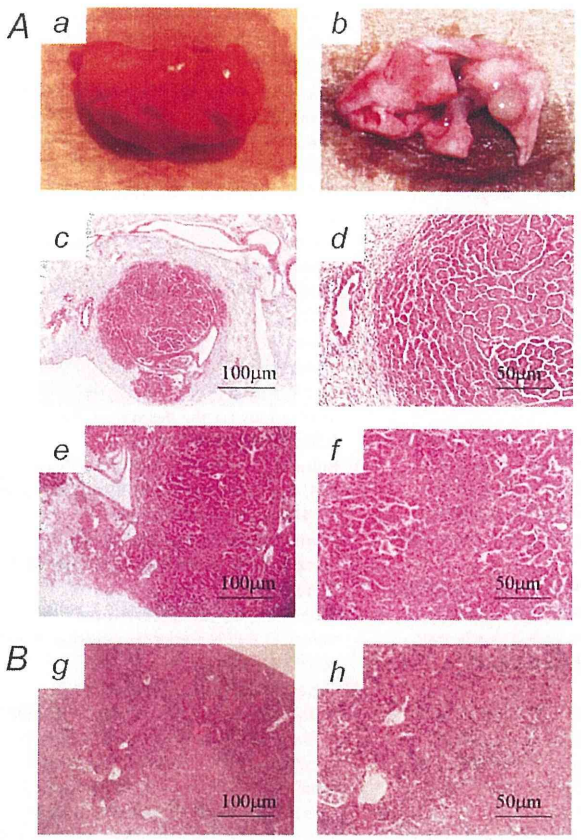

Figure 5. Lung tumors and lymphoma developed in APOBEC2 Tg mice. (*A*) Macroscopic view of a lung tumor that developed in a 72-week-old APOBEC2 Tg mouse (*b*). Microscopic view of a lung adenoma (*c,d*) and adenocarcinoma (*e,f*) that developed in a 72-week-old APOBEC2 Tg mouse. Macroscopic view of the lung of the wild-type littermate (*a*). (*B*) Histologic findings for lymphoma detected in the spleen of APOBEC2 Tg mice. (Original magnifications: 4*c,e,g* ×40; 4*d,f, h* ×100).

expression to tumorigenesis remain unknown. The number of mRNA mutations observed in the *Eif4g2 and Pten* genes in the liver of APOBEC2 Tg mice suggests that these genetic alterations by APOBEC2 have a role in the development of HCC. Indeed, the *EIF4G2* gene is a candidate molecule responsible for oncogenesis caused by the overexpression of APOBEC1,[21] and is frequently downregulated in human cancer tissues.[33] In addition, *PTEN* is one of the most frequently mutated tumor-suppressor genes in human cancers.[34] Thus, the tumorigenesis caused by constitutive APOBEC2 expression might be a consequence of promiscuous nucleotide editing.

Recent studies revealed that the expression of a subset of APOBEC family members is induced by cytokine stimulation in liver tissues. For example, we and other investigators demonstrated that APOBEC3G expression is triggered by interferon-α in hepatocytes, suggesting that APOBEC3G acts as a host defense in response to interferon signaling against viral infection.[35–37] In this study, we showed that TNF-α induced APOBEC2 protein expression in human hepatocytes. Considering the fact that chronic inflammation has important roles in human HCC development,[38,39] the finding that APOBEC2 is induced by proinflammatory cytokine stimulation and induces nucleotide alterations in tumor-related genes in hepatocytes provides a novel idea that aberrant expression of APOBEC2 in epithelial cells acts as a genotoxic factor linking inflammation and cancer development. The tumorigenic phenotype of the APOBEC2-Tg mice further suggests that APOBEC2 is involved in carcinogenesis of the liver tissue under conditions of chronic inflammation, the typical procancerous background of human HCC.

In conclusion, our findings provide the first direct evidence that APOBEC2 induces nucleotide changes preferentially in the *Eif4g2* and possibly the *Pten* genes, and the constitutive expression of APOBEC2 in epithelial tissues contributes to the development of various tumors including HCC and lung cancers. Understanding the pathologic role of APOBEC2 provides new insight into the mechanisms of cancer development in the liver underlying chronic inflammation. During our manuscript preparation, Sato *et al.* reported that they could not find the evidence of APOBEC2's affinity for RNA or high-stoichiometry association with a partner which usually associated with the known RNA editing enzymes.[40] Thus, further analyses would be required to clarify whether APOBEC2 dose possess an RNA-editing activity against specific target genes or overexpression of APOBEC2 causes nucleotide alterations in genome senquences in a promiscuous manner in hepatocytes.

References

1. Wedekind JE, Dance GS, Sowden MP, Smith HC. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet* 2003;19:207–16.
2. Cascalho M. Advantages and disadvantages of cytidine deamination. *J Immunol* 2004; 172:6513–8.
3. Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 2005;22:367–77.
4. OhAinle M, Kerns JA, Malik HS, Emerman M. Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. *J Virol* 2006;80:3853–62.
5. Pham P, Bransteitter R, Goodman MF. Reward versus risk: DNA cytidine deaminases triggering immunity and disease. *Biochemistry* 2005;44:2703–15.
6. Chen SH, Habib G, Yang CY, Gu ZW, Lee BR, Weng SA, Silberman SR, Cai SJ, Deslypere JP, Rosseneu M, *et al.* Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. *Science* 1987;238:363–6.
7. Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ, Scott J. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* 1987; 50:831–40.

8. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 2003;424:99–103.
9. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* 2003;424:94–8.
10. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH. DNA deamination mediates innate immunity to retroviral infection. *Cell* 2003;113:803–9.
11. Shindo K, Takaori-Kondo A, Kobayashi M, Abudu A, Fukunaga K, Uchiyama T. The enzymatic activity of CEM15/Apobec-3G is essential for the regulation of the infectivity of HIV-1 virion but not a sole determinant of its antiviral activity. *J Biol Chem* 2003;278:44412–6.
12. Takaori-Kondo A. APOBEC family proteins: novel antiviral innate immunity. *Int J Hematol* 2006;83:213–6.
13. Noguchi C, Ishino H, Tsuge M, Fujimoto Y, Imamura M, Takahashi S, Chayama K. G to A hypermutation of hepatitis B virus. *Hepatology* 2005;41:626–33.
14. Suspene R, Guetard D, Henry M, Sommer P, Wain-Hobson S, Vartanian JP. Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. *Proc Natl Acad Sci USA* 2005;102:8321–6.
15. Noguchi C, Hiraga N, Mori N, Tsuge M, Imamura M, Takahashi S, Fujimoto Y, Ochi H, Abe H, Maekawa T, Yatsuji H, Shirakawa K, *et al.* Dual effect of APOBEC3G on Hepatitis B virus. *J Gen Virol* 2007;88:432–40.
16. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem* 1999;274:18470–6.
17. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 2000;102:553–63.
18. Liao W, Hong SH, Chan BH, Rudolph FB, Clark SC, Chan L. APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochem Biophys Res Commun* 1999;260:398–404.
19. Mikl MC, Watt IN, Lu M, Reik W, Davies SL, Neuberger MS, Rada C. Mice deficient in APOBEC2 and APOBEC3. *Mol Cell Biol* 2005;25:7270–7.
20. Yamanaka S, Balestra ME, Ferrell LD, Fan J, Arnold KS, Taylor S, Taylor JM, Innerarity TL. Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc Natl Acad Sci USA* 1995;92:8483–7.
21. Yamanaka S, Poksay KS, Arnold KS, Innerarity TL. A novel translational repressor mRNA is edited extensively in livers containing tumors caused by the transgene expression of the apoB mRNA-editing enzyme. *Genes Dev* 1997;11:321–33.
22. Morisawa T, Marusawa H, Ueda Y, Iwai A, Okazaki IM, Honjo T, Chiba T. Organ-specific profiles of genetic changes in cancers caused by activation-induced cytidine deaminase expression. *Int J Cancer* 2008;123:2735–40.
23. Okazaki IM, Hiai H, Kakazu N, Yamada S, Muramatsu M, Kinoshita K, Honjo T. Constitutive expression of AID leads to tumorigenesis. *J Exp Med* 2003;197:1173–81.
24. Matsumoto T, Marusawa H, Endo Y, Ueda Y, Matsumoto Y, Chiba T. Expression of APOBEC2 is transcriptionally regulated by NF-kappaB in human hepatocytes. *FEBS Lett* 2006;580:731–5.
25. Kou T, Marusawa H, Kinoshita K, Endo Y, Okazaki IM, Ueda Y, Kodama Y, Haga H, Ikai I, Chiba T. Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis. *Int J Cancer* 2007;120:469–76.
26. Iwai A, Marusawa H, Matsuzawa S, Fukushima T, Hijikata M, Reed JC, Shimotohno K, Chiba T. Siah-1L, a novel transcript variant belonging to the human Siah family of proteins, regulates beta-catenin activity in a p53-dependent manner. *Oncogene* 2004;23:7593–600.
27. Schroeder T, Just U. Notch signalling via RBP-J promotes myeloid differentiation. *Embo J* 2000;19:2558–68.
28. Endo Y, Marusawa H, Kinoshita K, Morisawa T, Sakurai T, Okazaki IM, Watashi K, Shimotohno K, Honjo T, Chiba T. Expression of activation-induced cytidine deaminase in human hepatocytes via NF-kappaB signaling. *Oncogene* 2007;26:5587–95.
29. Doi T, Kinoshita K, Ikegawa M, Muramatsu M, Honjo T. De novo protein synthesis is required for the activation-induced cytidine deaminase function in class-switch recombination. *Proc Natl Acad Sci USA* 2003;100:2634–8.
30. Zhao Y, Pan-Hammarstrom Q, Zhao Z, Hammarstrom L. Identification of the activation-induced cytidine deaminase gene from zebrafish: an evolutionary analysis. *Dev Comp Immunol* 2005;29:61–71.
31. Saunders HL, Magor BG. Cloning and expression of the AID gene in the channel catfish. *Dev Comp Immunol* 2004;28:657–63.
32. Prochnow C, Bransteitter R, Klein MG, Goodman MF, Chen XS. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* 2007;445:447–51.
33. Buim ME, Soares FA, Sarkis AS, Nagai MA. The transcripts of SFRP1,CEP63 and EIF4G2 genes are frequently downregulated in transitional cell carcinomas of the bladder. *Oncology* 2005;69:445–54.
34. Salmena L, Carracedo A, Pandolfi PP. Tenets of PTEN tumor suppression. *Cell* 2008;133:403–14.
35. Tanaka Y, Marusawa H, Seno H, Matsumoto Y, Ueda Y, Kodama Y, Endo Y, Yamauchi J, Matsumoto T, Takaori-Kondo A, Ikai I, Chiba T. Anti-viral protein APOBEC3G is induced by interferon-alpha stimulation in human hepatocytes. *Biochem Biophys Res Commun* 2006;341:314–9.
36. Bonvin M, Achermann F, Greeve I, Stroka D, Keogh A, Inderbitzin D, Candinas D, Sommer P, Wain-Hobson S, Vartanian JP, Greeve J. Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology* 2006;43:1364–74.
37. Komohara Y, Yano H, Shichijo S, Shimotohno K, Itoh K, Yamada A. High expression of APOBEC3G in patients infected with hepatitis C virus. *J Mol Histol* 2006;37:327–32.
38. Thorgeirsson SS, Grisham JW. Molecular pathogenesis of human hepatocellular carcinoma. *Nat Genet* 2002;31:339–46.
39. Llovet JM, Burroughs A, Bruix J. Hepatocellular carcinoma. *Lancet* 2003;362:1907–17.
40. Sato Y, Probst HC, Tatsumi R, Ikeuchi Y, Neuberger MS, Rada C. Deficiency in APOBEC2 leads to a shift in muscle fiber-type, diminished body mass and myopathy. *J Biol Chem* 2009;285:7111–18.

Genetic Heterogeneity of Hepatitis C Virus in Association with Antiviral Therapy Determined by Ultra-Deep Sequencing

Akihiro Nasu[1], Hiroyuki Marusawa[1]*, Yoshihide Ueda[1], Norihiro Nishijima[1], Ken Takahashi[1], Yukio Osaki[2], Yukitaka Yamashita[3], Tetsuro Inokuma[4], Takashi Tamada[5], Takeshi Fujiwara[6], Fumiaki Sato[6], Kazuharu Shimizu[6], Tsutomu Chiba[1]

1 Department of Gastroenterology and Hepatology, Graduate School of Medicine, Kyoto University, Kyoto, Japan, **2** Department of Gastroenterology and Hepatology, Osaka Red Cross Hospital, Osaka, Japan, **3** Department of Gastroenterology and Hepatology, Wakayama Red Cross Hospital, Wakayama, Japan, **4** Department of Gastroenterology, Kobe City Medical Center General Hospital, Kobe, Japan, **5** Department of Gastroenterology and Hepatology Takatsuki Red Cross Hospital, Osaka, Japan, **6** Department of Nanobio Drug Discovery, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, Japan

Abstract

Background and Aims: The hepatitis C virus (HCV) invariably shows wide heterogeneity in infected patients, referred to as a quasispecies population. Massive amounts of genetic information due to the abundance of HCV variants could be an obstacle to evaluate the viral genetic heterogeneity in detail.

Methods: Using a newly developed massive-parallel ultra-deep sequencing technique, we investigated the viral genetic heterogeneity in 27 chronic hepatitis C patients receiving peg-interferon (IFN) α2b plus ribavirin therapy.

Results: Ultra-deep sequencing determined a total of more than 10 million nucleotides of the HCV genome, corresponding to a mean of more than 1000 clones in each specimen, and unveiled extremely high genetic heterogeneity in the genotype 1b HCV population. There was no significant difference in the level of viral complexity between immediate virologic responders and non-responders at baseline ($p = 0.39$). Immediate virologic responders ($n = 8$) showed a significant reduction in the genetic complexity spanning all the viral genetic regions at the early phase of IFN administration ($p = 0.037$). In contrast, non-virologic responders ($n = 8$) showed no significant changes in the level of viral quasispecies ($p = 0.12$), indicating that very few viral clones are sensitive to IFN treatment. We also demonstrated that clones resistant to direct-acting antivirals for HCV, such as viral protease and polymerase inhibitors, preexist with various abundances in all 27 treatment-naïve patients, suggesting the risk of the development of drug resistance against these agents.

Conclusion: Use of the ultra-deep sequencing technology revealed massive genetic heterogeneity of HCV, which has important implications regarding the treatment response and outcome of antiviral therapy.

Citation: Nasu A, Marusawa H, Ueda Y, Nishijima N, Takahashi K, et al. (2011) Genetic Heterogeneity of Hepatitis C Virus in Association with Antiviral Therapy Determined by Ultra-Deep Sequencing. PLoS ONE 6(9): e24907. doi:10.1371/journal.pone.0024907

Editor: Yoshio Yamaoka, Veterans Affairs Medical Center (111D), United States of America

Received June 17, 2011; **Accepted** August 19, 2011; **Published** September 22, 2011

Funding: This study was supported by Japan Society for the Promotion of Science (JSPS) Grants-in-aid for Scientific Research, and Health and Labour Sciences Research Grants for Research, and Research on Hepatitis from the Ministry of Health, Labour and Welfare, Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: maru@kuhp.kyoto-u.ac.jp

Introduction

Hepatitis C virus (HCV) is classified as a member of the Flaviviridae family [1] and has an approximately 9.6-kb single-stranded RNA genome. This RNA genome encodes a large precursor polyprotein, which is cleaved by viral and host proteases to generate at least 10 functional viral proteins; core, envelope (E)-1, E2, p7, nonstructural protein (NS)-2, NS3, NS4A, NS4B, NS5A, and NS5B [2,3]. A strong characteristic of HCV infection is its significant genetic diversity, the consequence of the absence of proofreading activity in RNA-dependent RNA polymerase [4], and the high level of viral replication during its life cycle [5]. The mean frequency of nucleotide alterations occurring in HCV RNA is calculated to be between 1.4×10^3 and 1.9×10^3 substitutions per nucleotide per year [6,7]. As a result, the infecting HCV clones in each patient invariably show population diversity with a high degree of genetic heterogeneity. The collection of viruses in a population of closely related but non-identical genomes is referred to as a quasispecies [8,9], and the dominant viral population may be evolving as a result of its viral replicative fitness and concurrent immune selection pressures that drive clonal selection.

It is reasonable to assume that the viral pathogenesis and sensitivity to treatment are affected by the generation of escape mutants through immune evasion and the modification of virulence characteristics by anti-viral treatment [10]. Thus, certain viral mutations have important implications for the pathogenesis of the viral disease and the sensitivity to antiviral therapy. Several studies have attempted to associate genetic heterogeneity or

number of mutations with pathogenesis and treatment outcome. However, the abundant diversity and complexity of the chronically-infected HCV has been an obstacle to evaluate the viral genetic heterogeneity in detail. In this respect, the recent introduction of ultra-deep sequencing technology, capable of producing millions of DNA sequence reads in a single run, is rapidly changing the landscape of genome research [11,12]. One application of ultra-deep sequencing was the identification of rare minority drug resistant clones of human immunodeficiency virus, which are not detectable by standard sequencing techniques [13–15]. Moreover, the recent study using 454/Roche pyrosequencing technology clarified the transmission bottlenecks by measuring the population structure within patients with HCV infection [16].

In this study, we used for the first time ultra-deep sequencing with Illumina Genome Analyzer II (Illumina, San Diego, CA) and determined the pictures of viral quasispecies of genotype 1b HCV in patients receiving peg-interferon (IFN) α2b plus ribavirin (RBV) to clarify the significance of the viral genetic complexity in the pathophysiology of HCV infection and the treatment outcome of the current IFN-based therapy for HCV-infected patients. Because our main objective was to determine whether the HCV sequence variation itself is responsible for the sensitivity or resistance to antiviral therapy, we compared the composition of the HCV population complexity 1 week after IFN administration in patients who showed a prompt decrease in HCV viremia with those in whom there was no reduction in the serum HCV RNA levels after the initiation of IFN treatment. We also examined the prevalence of drug-resistant mutations to direct-acting antivirals (DAAs) for HCV in treatment-naïve HCV-infected patients, based on the fact that drug-resistant mutations already exist in treatment-naïve patients with various pathogenic virus infections, such as human immunodeficiency viruses [14,17].

Results

Validation of multiplex ultra-deep sequencing of the HCV genome

We performed a massive parallel ultra-deep sequencing run on the Illumina Genome Analyzer II platform using multiplex tagging methods. First, we conducted a control experiment to validate the efficacy and error rates in ultra-deep sequencing of the viral genome. For this purpose, we used a plasmid encoding full-length HCV [18] as a template and determined the plasmid-derived whole HCV sequence. The ultra-deep sequencing platform provided us the full-length HCV genome information derived from the plasmids with a mean coverage of 1674.3 at each nucleotide site (Table 1). Errors comprised insertions (1.0%), deletions (4.2%), and nucleotide mismatches (94.8%) and the overall error rates by multiplex ultra-deep sequencing were determined to be a mean of 0.0010 per bp. Next we confirmed that the high-fidelity PCR amplification with HCV-specific primer sets followed by multiplex ultra-deep sequencing resulted in no significant increase in the error rates in the viral sequencing data (ranging from 0.0012 to 0.0013 per bp; per-nucleotide error rate, 0.12%–0.13%).

To estimate the accuracy of detecting nucleotide alterations using reads filtered by average base quality and mapping quality, we introduced the plasmid with single point mutations within the wild-type viral sequences with the ratio of 1:99 and 1:999 and assessed the sensitivity and accuracy of quantification with the high-fidelity PCR amplification followed by multiplex ultra-deep sequencing. Duplicate control experiments revealed that mutations present at an input ratio of 0.10% ranged between 0.09 and 0.19%, and the results could be reproducibly quantified (data not shown). Based on these results, we picked up the low abundant mutations that presented at frequency of more than 0.20% among the total viral clones, a level that could rule out putative errors caused by massively-parallel sequencing, in the current platform used in this study.

Table 1. Error frequency of ultra-deep sequencing for the plasmid encoding full-genome HCV sequence.

	PCR amplification	
	(−)*	(+)*
Total read nucleotides	15,118,929	24,158,372
Mean coverage	1674.3	5562.6
Type of errors		
mismatches	14,629 (94.8%)	26,243 (88.6%)
deletions	640 (4.2%)	2510 (8.5%)
insertions	147 (1.0%)	859 (2.9%)
Overall error rate (%)	0.102	0.123

*(−); Ultra-deep sequencing of HCV encoding plasmid
(+); Ultra-deep sequencing of PCR-amplified HCV encoding plasmid.
doi:10.1371/journal.pone.0024907.t001

Large heterogeneity of viral clones in HCV-infected patients

HCV infection comprises a heterogeneous mixture of viral clones with various mutations. To clarify the landscape of HCV heterogeneity as a quasispecies, we determined the viral full-genome sequences derived from 27 HCV-infected patients by multiplex ultra-deep sequencing and compared the results with those obtained by the direct population Sanger sequencing method. All sequence reads by multiplex ultra-deep sequencing have been deposit in DNA Data Bank of Japan Sequence Read Archive (http://www.ddbj.nig.ac.jp/index-e.html) under accession number DRA000366.

HCV nucleotide sequence reads by ultra-deep sequencing were aligned to the consensus viral sequences in the same serum specimen that were determined by direct population Sanger sequencing. A mean number of 1705-fold coverage on average was achieved at each nucleotide site of the HCV sequences in each specimen. The average frequencies of altered sequences detected in each viral genomic region are summarized in Table 2. Compared with the representative sequence of the population average clone, the mutation frequency was 1.04% of the total viral genomic sequences and 16.1% of the total nucleotide positions on average. Most of the genomic changes observed in viral variants were single base substitutions and unevenly distributed throughout the region of the HCV genome.

Among the viral genomic regions, the nucleotide sequence complexity expressed as the Shannon entropy was smallest in the core region. In contrast, the viral sequence complexity in the E2 region was highest among the HCV genomic regions and significantly greater than the average mutation frequency of the remaining HCV genome ($p = 0.0026$). Similarly, the ratio of the number of mutated nucleotides to the total number of nucleotides analyzed in the E2 region was significantly higher than that of the remaining HCV genome ($p = 5.66 \times 10^{-6}$). These findings clearly confirmed that the quasispecies complexity in E2, which contains hypervariable region1 (HVR1) and HVR2, was prominently larger than that of other viral genomic regions [19].

Table 2. Mean genetic complexity of the genotype1b HCV in chronically infected 27 patients.

Viral genomic Region	Mean number of aligned nucleotides	Mean number of mutated nucleotides	Mean coverage	Mutation frequency (%)	Mean Shannon entropy
Core	779,839	5027	1361	0.61	0.045926
E1	739,220	7902	1360	0.99	0.064884
E2	1,382,907	19,724	1265	1.37	0.088584
p7	217,000	3237	1148	1.44	0.075829
NS2	673,579	8702	1073	1.19	0.075333
NS3	4,958,188	52,204	2619	0.93	0.060767
NS4A	427,677	5604	2640	1.32	0.072217
NS4B	1,209,000	17,485	1544	1.26	0.063190
NS5A	2,034,626	28,820	1518	1.28	0.067398
NS5B	2,720,417	27,449	1681	0.90	0.054805
Total	14,875,801	172,327	1705	1.04	0.062624

doi:10.1371/journal.pone.0024907.t002

Early dynamic changes of viral complexity after the administration of peg-IFNα2b plus RBV

Among 27 patients enrolled in this study, 8 showed a prompt decrease in their serum HCV RNA levels and 8 showed no significant changes 1 week after initiating treatment with peg-IFNα2b plus RBV. To clarify the changes in the viral quasispecies in response to antiviral therapy, we determined the early dynamic changes in viral complexity before and after 1 week of peg-IFNα2b plus RBV administration in these 8 immediate virologic responders and 8 non-responders. All cases were infected with genotype 1b viruses, and the clinical features, including serum HCV RNA level at baseline, did not significantly differ between immediate virologic responders and non-responders (Table 3). A mean coverage of 1798-fold and 2416-fold were mapped to each reference sequence in immediate virologic responders before and after peg-IFNα2b plus RBV administration, respectively. Similarly, a mean coverage of 1780-fold and 2461-fold were determined in non-responders before and after peg-IFNα2b plus RBV administration, respectively (Table 4 and Table S1).

We then estimated the genomic complexity by calculating the Shannon entropy for each nucleotide position before and after the administration of peg-IFNα2b plus RBV (Table 4). There was no significant difference in the level of viral complexity between immediate virologic responders and non-responders at a baseline (mean Shannon entropy value 0.072 vs 0.075, $p = 0.39$). Immediate virologic responders, however, showed a significant reduction in the nucleotide sequence complexity after the administration of peg-IFNα2b plus RBV (mean Shannon entropy value 0.072 vs 0.049, $p = 0.037$), indicating that the viral quasispecies nature after the peg-IFNα2b plus RBV treatment

Table 3. Characteristics of patients that showed immediate virologic response or non-response to PEG-IFNα2b plus ribavirin combination therapy.

	Immediate virologic responders	Non-responders	P-value
Age†	50.5 (45–68)	60 (55–69)	0.12
Sex (male/female)	5/3	5/3	1
Alanine aminotransaminase† (IU/l)	54 (15–198)	72 (30–143)	0.51
Total bilirubin† (mg/dl)	0.6 (0.4–1.8)	0.8 (0.4–1.4)	0.34
Platelet count† ($\times 10^4/mm^3$)	18.9 (7.1–27.2)	16.7 (11.6–22.5)	0.68
HCV genotype	1b	1b	
HCV viral load† (log IU/ml)			
pre-treatment	6.6 (6.2–7.5)	6.9 (6.1–7.6)	0.43
after treatment	4.6 (4.0–5.2)	6.5 (6.1–6.8)	**0.028**
Final outcome			**0.025**
sustained viral response	6	0	
Relapse	1	1	
non-response	0	6	
withdraw*	1	1	

† Values are median (range).

* The treatment was discontinued in one immediate virologic responder and one non-responder, due to the side effect of IFN and the development of liver cancer, respectively.

doi:10.1371/journal.pone.0024907.t003

Table 4. Genetic complexity at pre-treatment and 1 week after PEG-IFNα2b plus ribavirin combination therapy in immediate virologic responders and non-responders.

	Immediate virologic responders (N = 8)		Non-responders (N = 8)	
	Pre-treatment	**1 week after IFN therapy**	**Pre-treatment**	**1 week after IFN therapy**
Mean number of aligned reads	263,452	356,963	256615	354,398
Mean number of aligned nucleotides	16,632,186	22,438,125	16,248,820	22,379,922
Mean coverage	1798	2416	1780	2461
Mutation frequency (%)	0.96	0.63	1.13	1.11
Shannon entropy	0.072*	0.049*	0.075**	0.066**

Wilcoxon rank sum test.
* $p = 0.037$.
** $p = 0.12$.
doi:10.1371/journal.pone.0024907.t004

became relatively more homogeneous than at baseline status in this group. In contrast, no significant changes in the nucleotide sequence complexity were observed in non-responder patients before and after treatment with peg-IFNα2b plus RBV (mean Shannon entropy value 0.075 vs 0.066, $p = 0.12$). We then examined whether specific nucleotide position might be associated with the response to peg-IFNα2b plus RBV treatment in immediate virologic responders, but complexity was not commonly shared at any specific nucleotide position that changed by more than 50% after peg-IFNα2b plus RBV administration (data not shown), indicating no association between the specific nucleotide position and the response to peg-IFNα2b plus RBV treatment.

Elimination of minor viral clones by peg-IFNα2b plus RBV therapy

Next, we compared the nucleotide complexity in each viral genomic region of the immediate virologic responders with that of non-responders before and after peg-IFNα2b plus RBV administration (Figure 1 and Table S2). In immediate virologic responders, the peg-IFNα2b plus RBV therapy induced a significant reduction in the nucleotide sequence complexity in all viral genomic regions except NS4B. In contrast, non-responders showed no significant change in the viral sequence complexity in any viral genomic region. For example, there was no significant difference in the mean complexity in the E2 region at baseline between the immediate virologic responders and non-responders. The administration of peg-IFNα2b plus RBV significantly reduced the levels of nucleotide sequence complexity in the E2 region in all the immediate virologic responders (mean Shannon entropy value 0.139 vs 0.085, respectively. $p = 0.012$, Figure 1 and Table S2). In contrast, no significant changes in the sequence complexity were observed in the E2 (mean Shannon entropy value 0.083 vs 0.082, respectively. $p = 0.89$) regions in non-responder cases after treatment with peg-IFNα2b plus RBV.

To examine whether certain viral clones in non-responders showed sensitivity to IFN therapy, we investigated the sequence complexity in HVR1 in the E2 region in detail before and after peg-IFNα2b plus RBV therapy, because the HVR1 region possessed one of the highest complexities among viral genomic regions. In immediate virologic responders, the heterogeneity at each nucleotide position was reduced in response to peg-IFNα2b plus RBV administration (representative nucleotide changes are shown in Figure 2A). In contrast, the ratio of mutated clones among the total sequence reads determined at each nucleotide site in HVR1 showed no significant change before and after the administration of peg-IFNα2b plus RBV in the majority of non-responders (Figure 2B), suggesting that very few viral clones showed sensitivity to peg-IFNα2b plus RBV and were eliminated after the administration of peg-IFNα2b plus RBV.

Detection of viral clones with drug-resistant mutations

Because none of the DAAs for HCV were approved by Japanese health coverage at the time of this study, all patients enrolled into this study were naïve to DAAs for HCV including protease and polymerase inhibitors. Thus, we determined whether the reported drug-resistant mutants exist spontaneously in nature among treatment-naïve HCV-infected patients. For this purpose, we examined the naturally prevalent mutations against HCV protease and polymerase inhibitors in the 27 patients. The drug-resistant mutations examined here included 9 mutations resistant to NS3/4 protease inhibitors, including Telaprevir, Boceprevir, TMC435350, ITMN191/R7227, MK-7009, and BI-201335, and 5 mutations resistant to NS5B polymerase inhibitors, including Filibuvir, BI-207127, and R7128 [20].

The mean number of sequence reads at the nucleotide position comprising mutations resistant to NS3/4A protease and NS5B polymerase inhibitors among the 27 cases were obtained with 1179-fold and 1972-fold coverage, respectively. Based on the detection rate of the low-level viral clones determined by the control experiments, we picked up the drug-resistant mutants that presented at a frequency of more than 0.2% among the total viral clones. Based on these criteria, at least one resistant mutation was detected in all subjects (Table 5). The mean prevalence of the 14 drug-resistant mutations ranged from 0.20% to 99.1% indicating that the proportion of resistant mutations substantially differed in each case. The T54S/A mutation resistant to Teraprevir and Boceprevir in genotype 1b HCV [21] was the most commonly detected (20 of 27 cases, 74.1%). The proportion of T54S/A mutations among the total clones ranged from 0.21% to 86.9% and thus substantially differed between cases. Other mutations resistant to the NS3/4A protease-inhibitor were detected in 16 of 27 cases (59.3%) at V55A and Q80R/K, and 12 of 27 cases (44.4%) at V36A/M. In contrast, no D168A/V/T/H mutation resistant to ITMN191/R7227, MK-7009, TMC435350, and BI-201335 was detectable. Regarding NS5B polymerase inhibitors, the V499A mutation resistant to BI-207127, was most frequently detected and 20 of 27 (74.1%) of subjects possessed the resistant-mutant clones at levels 0.20% to 99.1% at baseline. Only one case had the BI-207127-resistant P496A mutant clones and none had the R7128-resistant S282T clones. Of the 27 subjects, 16 (59.3%) harbored mutations resistant to at least four kinds of NS5B polymerase inhibitors and/or NS3/4A protease-inhibitors. More-

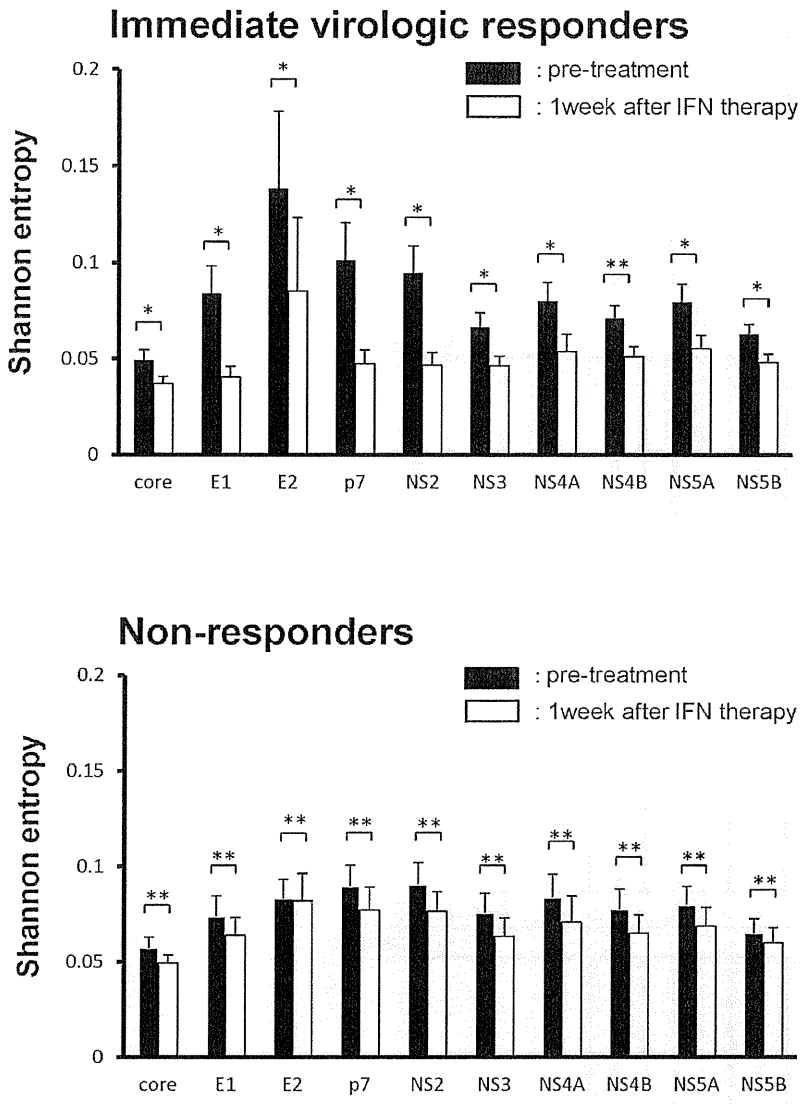

Figure 1. Changes in the genetic complexity of each HCV genomic region before and after the administration of peg-IFNα2b plus RBV. Shannon entropy values at baseline (black bar) and 1 week after initiation of treatment with peg-IFNα2b plus RBV (white bar) in 8 immediate virologic responders (A) and in 8 non-responders (B) are shown. * $p<0.05$, ** not significant. (Mean values ± SD; n = 8)
doi:10.1371/journal.pone.0024907.g001

over, 5 subjects (18.5%) harbored resistance to 6 antiviral drugs. Notably, 3 subjects harbored resistance to 8 of 9 antiviral drugs. There was no significant association between the frequency of drug-resistant mutations and the serum viral load ($r = 0.0678$) (Figure S1).

These findings indicate that drug-resistant HCV variants are present in a considerable proportion among the chronically HCV-infected, DAAs-naïve patients.

Discussion

Sequence heterogeneity, so-called quasispecies, is a common feature of RNA viruses, including HCV [22]. Previous studies of the viral genome with conventional Sanger sequencing methods revealed that HCV infection comprises a cloud of closely related sequence variants differing by as little as one nucleotide from a population average sequence [23]. A number of studies have aimed to clarify the significance of viral mutations in association with clinical features, including viral persistency and chronicity, degree of liver damage, response to treatment, and selection of mutants resistant to anti-viral therapy. The quasispecies nature of HCV, however, represents a major obstacle in determining the significance of the viral clone with specific sequence characteristics. Newly developed ultra-deep sequencing analysis allowed us to clarify the whole picture of viral quasispecies present in chronically HCV-infected patients. In the present study, ultra-deep sequencing determined a mean total of more than 10 million nucleotides of the viral genome in each specimen, representing more than 1000 clones infecting each patient, thus demonstrating the abundant genetic complexity of HCV.

It is well recognized that the HCV genome is heterogeneous at the intra-individual level [9,10]. The current ultra-deep sequenc-

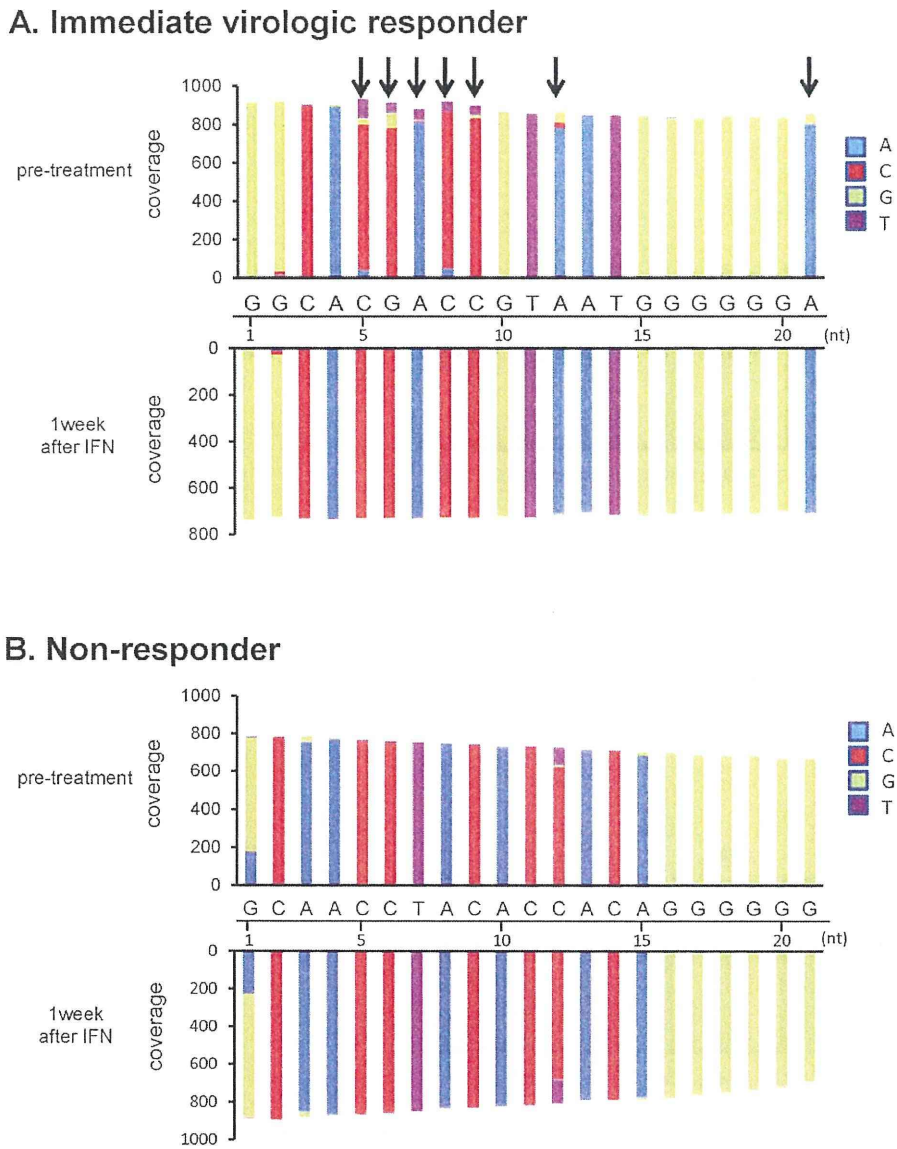

Figure 2. Ratio of mutated nucleotides in the HVR1 region before and after administration of peg-IFNα2b plus RBV therapy. Representative results of a immediate virologic responder (Patient#3) (**A**) and a non-responder (Patient#9) (**B**) are shown. The read numbers (coverage) at each nucleotide position of the HVR1 (from 1st nucleotide to 21st nucleotide in E2 region) at pre-treatment (upper graphs) and 1 week after initiating treatment with peg-IFNα2b plus RBV (lower graphs) are shown. Arrows indicate the nucleotide positions that showed the elimination of minor mutant clones after administration of peg-IFNα2b plus RBV.
doi:10.1371/journal.pone.0024907.g002

ing analyses revealed that the E2 region had the highest sequence heterogeneity, while the core region had the lowest sequence heterogeneity among the viral genomic regions encoding different functional viral proteins. More than 15% of nucleotides in the E2 region were mutated in all cases examined. These findings are consistent with previous conventional Sanger sequencing-based studies showing that HVR1 and HVR2 possess the highest sequence diversity among the HCV genomic regions [19] and that the highest values of mean Shannon entropy at the HCV 1a population level are in the E2 region [24].

Various mutations in the HCV genome are associated with the therapeutic response. For example, a number of mutations within a so-called IFNα sensitivity determining region of NS5A are closely associated with sensitivity to IFN-based anti-viral therapy [25,26]. A recent study also showed that amino acid substitution in the HCV core region could be a useful predictor of the virologic response to peg-IFNα plus RBV combination therapy [27]. Although the findings of these studies suggested that certain mutations in the representative HCV clone could predict treatment outcome, it is unknown whether the specific viral clone comprising those mutations directly displays sensitivity or resistance to anti-viral therapy. In the present study, sequential comparison of the HCV1b genome derived at baseline and at 1 week after the administration of peg-IFNα2b plus RBV demon-

Table 5. Prevalence of anti-HCV drug resistant mutations among the treatment-naïve patients.

Residue and Position	Drugs	Number of patients with mutated clones (%)	Frequency of the mutated clones (%)*
Resistant mutation to NS3/4A protease inhibitor			
T54S/A	Telaprevir Boceprevir	20/27 (74.1%)	0.49 (0.21–86.9)
V55A	Boceprevir	16/27 (59.3%)	0.4 (0.23–1.53)
Q80R/K	TMC435350	16/27 (59.3%)	0.36 (0.24–1.37)
V36A/M	Telaprevir Boceprevir	12/27 (44.4%)	0.47 (0.20–0.88)
V170A/T	Boceprevir	11/27 (40.7%)	0.52 (0.20–1.03)
A156T/V	Telaprevir	7/27 (25.9%)	0.35 (0.20–0.80)
R155K/T/Q	Telaprevir Boceprevir ITMN191/R7227 MK-7009 TMC435350 BI-201335	5/27 (18.5%)	0.42 (0.22–0.62)
A156S	Telaprevir Boceprevir	3/27 (11.15)	0.35 (0.24–0.83)
D168A/V/T/H	ITMN191/R7227 MK-7009 TMC435350 BI-201335	0/27 (0%)	
Resistant mutation to NS5B polymerase inhibitor			
V499A	BI-207127	20/27 (74.1%)	0.59 (0.20–99.1)
M423T/I/V	Filibuvir	12/27 (44.4%)	0.41 (0.21–1.48)
P495S/L/A/T	BI-207127	9/27 (33.3%)	0.37 (0.21–0.87)
P496A/S	BI-207127	1/27 (3.7%)	0.32
S282T	R7128	0/27 (0%)	

* Values are median (range).

doi:10.1371/journal.pone.0024907.t005

strated that IFN treatment resulted in no selective decrease of the viral clones comprising the previously defined mutational changes that were associated with a response to anti-viral therapy. Moreover, immediate virologic responders showed no common baseline nucleotide alterations that are efficiently eliminated in response to the administration of peg-IFNα2b plus RBV. Thus, our data suggest that an HCV sequence variation itself at a specific single nucleotide position does not directly reflect the virologic features regarding the sensitivity to IFN therapy in each viral clone, at least at the early stage of IFN administration. In contrast, several studies have provided evidence of the pre-existence of viral strains with an inherent resistance to IFN in patients who subsequently experienced a viral breakthrough or relapse [24,28]. Thus, there is room for further investigation to identify IFN-resistant clones by comparing the viral clones at baseline with those at the point of relapse using ultra-deep sequencing technology.

Notably, a distinct pattern of dynamic changes of HCV quasispecies was present between immediate responders and non-responders. Immediate responders showed a significant decrease of genetic complexity spanning all the viral genetic regions, resulting in a more homogeneous viral population after 1 week of peg-IFNα2b plus RBV administration. In contrast, non-responders showed no significant change in the genetic complexity in any of the HCV genomic regions. Our findings are consistent with the previous study showing that the early changes in HCV quasispecies determined by E1/E2 sequences provided prognostic information as early as the first 2 weeks after starting IFN therapy [28]. Moreover, the findings that there is no difference in the level of genetic complexity between early responders and non-responders at baseline and that almost none of the pre-existed HCV clones were eliminated in non-responder cases might suggest that the absence of sensitivity to IFN treatment in non-responders is due to host factors. Consistent with this hypothesis, recent studies revealed that host genetic variations at the IL28B gene are associated with a virologic response to peg-IFNα plus RBV combination therapy [29-32]. Alternatively, it is possible that a particular HCV protein of certain HCV mutants contributed to the strong inhibition of IFN-mediated anti-viral response in the liver of non-responders. Although dynamic changes in HVR1 sequences revealed that the minor viral clones were promptly eliminated in immediate virologic responders, the originally-inhabited major viral clones persisted 1 week after peg-IFNα2b plus RBV administration. Thus, further analyses are required to clarify how viral heterogeneity might be associated with the response to anti-viral therapy.

DAAs are promising drugs that could be more effective than peg-IFNα plus RBV therapy [33]. These DAAs include HCV NS3/4A protease and NS5B RNA-dependent RNA polymerase inhibitors, both of which have currently advanced to phase 1-3 trials. Increasing evidence, however, has clearly revealed that monotherapy with DAAs poses a high risk for the selection of resistant variants because of the high genetic heterogeneity of HCV [20]. Several studies reported the low prevalence of DAAs resistant mutants as the dominant clones in treatment-naïve cases [21,34-36]. For example, Kuntzen et al showed that drug-resistant mutations were detectable by conventional sequencing at individual frequencies between 0.3% and 2.8% in a treatment-naïve genotype 1 HCV-infected population [21]. In sharp contrast, ultra-deep sequencing identified that DAAs-resistant variants are common among treatment-naïve patients. Indeed, ultra-deep sequencing showed that 26 of 27 (96%) treatment-naive Japanese patients enrolled in this study possessed at least two clones resistant to DAAs, while 70.2% of the mutants presented as a very minor population (less than 1%) in each individual. It remains unclear whether these minor drug-resistant mutations have clinical significance, because the DAAs are not yet approved here in Japan. Recent *in vitro* findings, however, showed that minor but preexisting resistant mutants in HCV replicon cells were selected and expanded after DAAs therapy [37]. Lu et al revealed