Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Immunology

Helicobacter pylori-induced activation-induced cytidine deaminase expression and carcinogenesis

Hiroyuki Marusawa and Tsutomu Chiba

Tumorigenesis is a multistep process in which the accumulation of genetic alterations drives the transformation of normal cells into malignant derivatives. Activation-induced cytidine deaminase (AID) contributes to immune system diversity by inducing somatic hypermutations and class-switch recombinations of human immunoglobulin genes. The mutagenic activity of AID, however, can also induce genetic changes in various genes and may lead to the development of cancer. *Helicobacter pylori*, a class 1 carcinogen for human gastric cancer, affects AID expression by two different mechanisms, introduction of bacterial virulence factors into host cells and induction of inflammatory responses, thereby contributing to the accumulation of mutations in tumor-related genes. Aberrant AID activity may therefore be a novel link between infection and carcinogenesis.

Address
Department of Gastroenterology and Hepatology, Graduate School of Medicine, Kyoto University, 54 Kawara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan

Corresponding author: Marusawa, Hiroyuki (maru@kuhp.kyoto-u.ac.jp)

Current Opinion in Immunology 2010, **22**:442–447

This review comes from a themed issue on
Host pathogens
Edited by Adolfo Garcia-Sastre and Philippe Sansonetti

0952-7915/$ – see front matter

DOI 10.1016/j.coi.2010.06.001

Introduction

Helicobacter pylori (*H. pylori*) is a gram-negative, spiral-shaped bacterium colonized in human populations for more than 58,000 years [1]. *H. pylori* infection is involved in the development of several human diseases, including gastro-duodenal ulcers, gastric cancer, and mucosa-associated lymphoid tissue (MALT) lymphoma of the stomach. *H. pylori* strains exhibit a high level of genetic diversity, and a striking difference among strains is the presence or absence of a 40-kb DNA segment, termed the cag pathogenicity island (PAI). The risk for developing *H. pylori* infection-mediated gastric disorders is closely associated with the strain [2]. The risk of developing gastric cancer is higher in patients infected with cagPAI-positive *H. pylori* compared with cagPAI-negative *H. pylori* [3,4], but how *H. pylori* infection contributes to gastric carcinogenesis remained unknown. Genetic changes in tumor-related genes are essential in the malignant transformation that leads to cancer cell development. How the intra-gastric residential bacteria induce the genetic changes required for tumorigenesis in host gastric epithelial cells is unclear, since the extracellular habitant *H. pylori* cannot directly access host genomic DNA located in the nucleus of gastric epithelial cells. Recent studies, however, revealed that cagPAI-positive *H. pylori* manipulates the host nucleotide editing enzymes to induce mutagenesis in human DNA sequences of the gastric epithelium [5].

Novel mechanism of active mutagenesis achieved by nucleotide editing enzymes

Genetic changes in tumor-related genes are essential for malignant transformation in cancer cell development. Mechanisms that account for genetic changes required for tumorigenesis are unknown, except for defects in the DNA repair system that are observed in certain human cancers. Several enzymes that induce nucleotide alterations were recently identified, providing a new avenue for understanding the mutagenesis mechanism. The apolipoprotein B mRNA editing enzyme catalytic polypeptide (APOBEC) family comprises nucleotide editing enzymes that insert nucleotide alterations in target DNA or RNA through cytidine deamination [6]. The human APOBEC family consists of APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3DE, APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4, and activation-induced cytidine deaminase (AID), and contributes to producing various favorable physiologic outcomes by modifying target gene sequences. For example, APOBEC1 participates in lipid metabolism through deaminating a specific cytidine to uridine in the Apo-B mRNA, resulting in the formation of a termination codon, which leads to the production of a half-length genomically encoded Apo-B100. APOBEC3G is an anti-viral molecule that induces hypermutation in viral DNA sequences and acts as a host defense factor against viruses such as HIV-1. Although the majority of APOBEC family members exhibit mutagenic activity against human RNA or exogenous viral genomes, only AID has the ability to induce nucleotide alterations and double-strand DNA breaks in human genomic sequences. Under physiologic conditions, AID is expressed in germinal center B cells and induces somatic hypermutation and class-switch recombination of immunoglobulin genes, thereby amplifying immune system diversity [7]. In sharp contrast to the favorable role of AID in the immune system, excessive AID activity might affect non-immunoglobulin

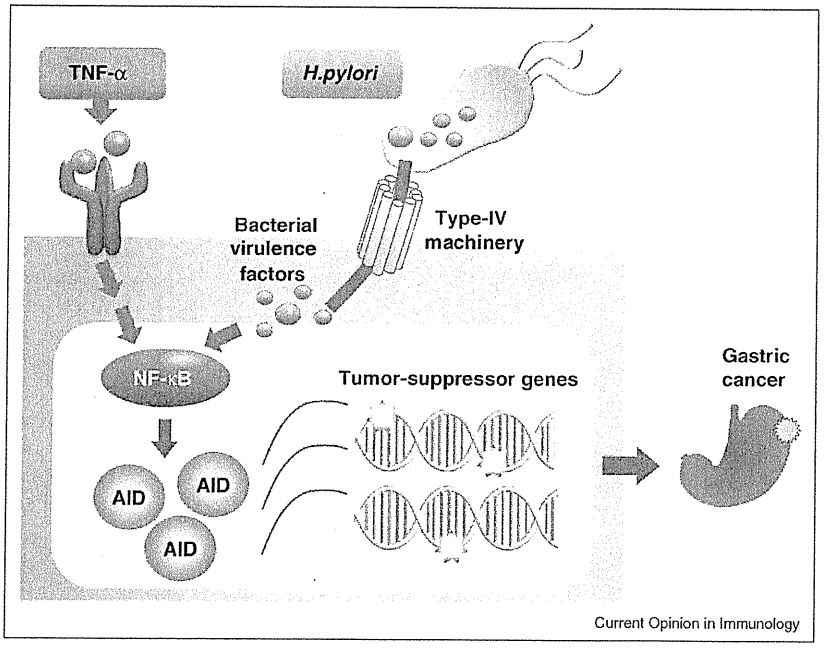
genes, including tumor-related genes in non-lymphoid cells [8].

CagPAI-positive *H. pylori* induces aberrant AID expression in gastric epithelial cells

AID can alter host genomic information, but there are safeguard mechanisms that restrict its potential tumorigenic activity, including post-transcriptional regulation by microRNA [9,10,11•], post-translational modification by protein phosphorylation or ubuiquitination [12–14], and regulation of subcellular localization [15–18]. Restriction of AID expression is also an important regulatory system that minimizes the aberrant mutagenic activity of AID. AID gene transcription is restricted mainly to activated germinal center B lymphocytes where editing of the immunoglobulin gene is required [19,20], while AID expression is not detected in normal epithelial cells under physiologic conditions. How then is AID expressed in epithelial cells under pathologic conditions, especially where the tumorigenic risk is unusually high? Strikingly, endogenous AID is expressed in the epithelial cells of *H. pylori*-infected stomach. Gastric epithelial cells and some infiltrating lymphocytes are immunoreactive for AID protein expression in the majority of chronic gastritis tissues infected with cagPAI-positive *H. pylori* [21]. Moreover, eradication of *H. pylori* infection by antibiotics substantially decreases AID protein expression in gastric mucosa. These findings suggest that cagPAI-positive *H. pylori* somehow upregulates AID protein in the gastric epithelium of the infected host.

CagPAI contain approximately 30 putative genes encoding various bacterial proteins such as cytotoxin-associated gene A (cagA) [22]. CagPAI-positive *H. pylori* introduces several bacterial virulence factors into gastric epithelial cells through a type-IV secretion apparatus, and cagPAI-positive *H. pylori*-derived peptidoglycans introduced into the host cells have been shown to be responsible for activating the transcription factor NF-κB [23]. The AID promoter region also includes sites for several transcription factors, such as NF-κB, STAT6, HoxC4, Sp1, Sp3, and Pax5 [24–27], and AID expression in B lymphocytes is induced in response to NF-κB activation through CD40 ligand signaling [28]. Together, these findings suggest that cagPAI-positive *H. pylori* induces AID expression via NF-κB activation by introducing bacterial virulence factors, and that the proinflammatory response caused by *H. pylori* infection also triggers AID expression via the activation of NF-κB in gastric epithelium, because proinflammatory cytokines such as tumor necrosis factor

Figure 1



Helicobacter pylori infection triggers AID expression in gastric epithelial cells via two distinct pathways. AID acts as a cytidine deaminase that is capable of inducing nucleotide alterations in human DNA sequences. Under physiologic conditions, no AID expression is detectable in normal gastric epithelium. *Helicobacter pylori* (*H. pylori*) infection, however, can induce AID expression in gastric epithelial cells via two distinct pathways. CagPAI-positive *H. pylori* strains possess type-IV machinery and can inject bacterial virulence factors into gastric epithelial cells, leading to the activation of the host transcriptional factor NF-κB. The host inflammatory response triggered by *H. pylori* infection also activates NF-κB in gastric epithelium. As a result, AID is transcriptionally upregulated in gastric epithelial cells, and can contribute to the production of unfavorable genetic changes in tumor-related genes, leading to gastric carcinogenesis.

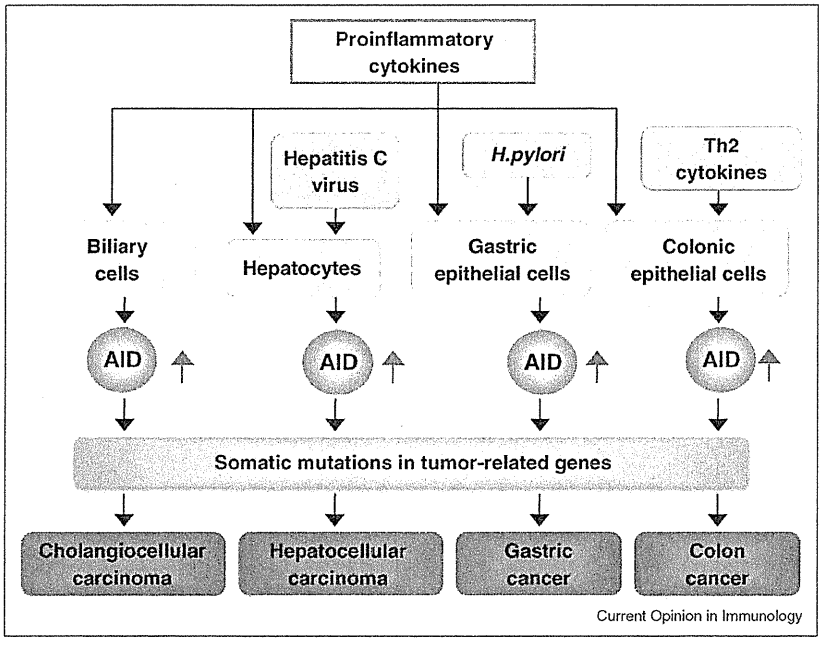
(TNF)-α and IL-1β can induce NF-κB activation in various types of cells. These hypotheses are supported by *in vitro* analyses showing that AID expression is induced in response to cagPAI-positive *H. pylori* infection or stimulation with the proinflammatory cytokine TNF-α via NF-κB signaling in cultured human gastric epithelial cells [21]. Based on the clinical course of *H. pylori*-infected individuals, both bacterial factors that are introduced into epithelial cells and the inflammatory response against *H. pylori* infection would be responsible for aberrant AID expression in gastric epithelium (Figure 1), and the direct action of the bacterial virulence factors contributes to activate AID in the early stage of *H. pylori* infection when the number of bacteria is high. In the late phase of chronic gastritis, when gastric atrophy has progressed and the number of *H. pylori* is decreased, the proinflammatory cytokine plays a central role in causing the constitutive expression of AID in gastric epithelial cells.

AID induces DNA mutations in tumor-related genes in gastric epithelial cells

The target of AID-mediated genotoxic effects is not restricted to immunoglobulin genes and several non-immunoglobulin genes are also targeted by AID in lymphocytes. Approximately 25% of expressed non-immunoglobulin genes analyzed, including Bcl6 and Cd83, accumulated AID-mediated mutations in germinal B cells [29••]. Moreover, AID produces double-strand DNA breaks throughout the genome, including *c-myc* in B cells [30,31•]. The impact of AID expression in non-lymphoid gastrointestinal epithelial cells was clarified using mouse model with constitutive and ubiquitous AID expression. AID transgenic mice accumulated somatic mutations in various tumor-related genes and developed tumors in both lymphoid and non-lymphoid tissues [32,33]. The tumors developed in the epithelial organs of AID transgenic mice included lung, liver, and gastric cancers.

The findings that AID mutagenic activity results in stomach cancer led us to speculate that *H. pylori* infection in association with aberrant AID expression contributes to human carcinogenesis via the accumulation of genetic alterations in gastric epithelial cells [21]. In *in vitro*-cultured gastric epithelial cells, cagPAI-positive *H. pylori* infection led to somatic mutations in the tumor-suppressor *TP53* gene. The number of nucleotide alterations observed in *H. pylori*-infected cells was significantly reduced by knockdown of

Figure 2



AID links inflammation and infection to cancer development in various gastrointestinal tissues. This figure is a model depicting the role of AID in the development of human cancers. Normal epithelial cells lack endogenous AID expression under physiologic conditions. *Helicobacter pylori* (*H. pylori*) infection and the resultant inflammatory stimulation, however, trigger aberrant AID expression in gastric epithelial cells. Similarly, hepatitis C virus infection and the resultant constitutive inflammation lead to AID expression in hepatocytes. In addition to proinflammatory cytokines, Th2 cytokine plays a role in the enhanced expression of AID in colonic epithelium. Constitutive AID activation in these epithelial cells results in the accumulation of somatic mutations in various target genes. If crucial nucleotide changes in the tumor-related genes are induced by AID activity, the resultant cells can acquire the transformation, leading to cancer development.

endogenous AID, indicating that the somatic mutations in the *TP53* gene in cells infected with cagPAI-positive *H. pylori* were due to the induction of endogenous AID expression in gastric cells. In wild-type mice, oral infection with cagPAI-positive *H. pylori* upregulated AID protein expression. Moreover, nucleotide alterations emerge in the *TP53* gene in stomach tissues after oral *H. pylori* infection in wild-type mice. These findings strongly suggest that *H. pylori* infection causes accumulation of somatic mutations in tumor-related genes such as *TP53* through aberrant upregulation of AID in gastric epithelial cells.

H. pylori-associated lymphoid tumorigenesis and AID expression

Low-grade lymphomas originating from MALT develop in the stomach, salivary and thyroid glands, bronchi, and small intestine, and are classified as MALT lymphoma [34,35]. The acquisition of MALT is induced before the development of lymphoma as a response to a persistent antigenic stimulation [36]. The development of gastric MALT lymphoma, a representative gastric lymphoma, is strongly associated with *H. pylori* infection [37]. The seroprevalence of *H. pylori* is higher in patients with gastric MALT lymphomas than in control patients without MALT lymphoma [38], and eradication of *H. pylori* leads to complete regression of the lymphoma in nearly 80% of patients with early-stage disease [39,40]. On the contrary, several studies have aimed to clarify the role of AID in the development of MALT lymphoma, because AID is required for the development of germinal center-derived non-Hodgkin's lymphomas [41,42•]. AID mRNA was, however, expressed in only some extranodal marginal zone B-cell MALT lymphomas [41]. More recent studies demonstrate that neoplastic marginal zone B cells did not express detectable AID, whereas AID expression was confined to reactive areas within MALT lymphomas [43,44]. In addition to the low frequency of AID upregulation in MALT lymphoma tissues, it remains unknown whether *H. pylori* infection enhances the aberrant mutagenic activity of AID in gastric B cells. Further analyses are required to determine the role of AID in the development of *H. pylori*-associated MALT lymphomas.

Conclusions

The discovery of AID was a seminal finding that greatly advanced our understanding of the molecular mechanisms involved in immunoglobulin diversification [45]. Now, AID is central to our understanding of how inflammation and infection underlie the genetic alterations required for carcinogenesis in epithelial cells [46]. Indeed, proinflammatory cytokine induction of AID expression via the NF-κB pathway is not limited to gastric epithelial cells. AID expression is mediated by the inflammatory response in a variety of epithelial cells, including human hepatocytes [47,48••], and biliary [49] and colonic epithelial cells [50•]. Aberrant AID expression in these gastrointestinal organs results in somatic mutations in various tumor-related genes. Thus, AID may have a central role in genetic susceptibility to mutagenesis, which leads to cancers in these gastrointestinal tissues upon exposure to certain inflammation or infection (Figure 2).

A characteristic of *H. pylori*-associated gastric cancer is multicentric tumor development. Patients with a history of *H. pylori*-related gastric cancer are at high risk for subsequent development of gastric cancers [51], suggesting that each epithelial cell of the *H. pylori*-infected stomach possesses sufficient genetic damage for malignant transformation. Efficient strategies to restrict aberrant AID activity might help to prevent carcinogenesis in gastric epithelial cells inflamed by *H. pylori* infection.

Acknowledgements

We are grateful to K. Kinoshita for his useful suggestions and critical reading of our manuscript. HM is supported by the grant from the Ministry of Education, Culture, Sports, Science, and Research Grant of the Princess Takamatsu Cancer Research and Takeda Research Fund.

Conflict of interest

None.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, van der Merwe SW *et al.*: **An African origin for the intimate association between humans and *Helicobacter pylori***. *Nature* 2007, **445**:915-918.
2. Cover TL, Blaser MJ: ***Helicobacter pylori* in health and disease**. *Gastroenterology* 2009, **136**:1863-1873.
3. Huang JQ, Zheng GF, Sumanac K, Irvine EJ, Hunt RH: **Meta-analysis of the relationship between cagA seropositivity and gastric cancer**. *Gastroenterology* 2003, **125**:1636-1644.
4. Basso D, Zambon CF, Letley DP, Stranges A, Marchet A, Rhead JL, Schiavon S, Guariso G, Ceroti M, Nitti D *et al.*: **Clinical relevance of *Helicobacter pylori* cagA and vacA gene polymorphisms**. *Gastroenterology* 2008, **135**:91-99.
5. Chiba T, Marusawa H: **A novel mechanism for inflammation-associated carcinogenesis; an important role of activation-induced cytidine deaminase (AID) in mutation induction**. *J Mol Med* 2009, **87**:1023-1027.
6. Conticello SG: **The AID/APOBEC family of nucleic acid mutators**. *Genome Biol* 2008, **9**:229.
7. Shivarov V, Shinkura R, Doi T, Begum NA, Nagaoka H, Okazaki IM, Ito S, Nonaka T, Kinoshita K, Honjo T: **Molecular mechanism for generation of antibody memory**. *Philos Trans R Soc Lond B Biol Sci* 2009, **364**:569-575.
8. Kinoshita K, Nonaka T: **The dark side of activation-induced cytidine deaminase: relationship with leukemia and beyond**. *Int J Hematol* 2006, **83**:201-207.
9. Dorsett Y, McBride KM, Jankovic M, Gazumyan A, Thai TH, Robbiani DF, Di Virgilio M, San-Martin BR, Heidkamp G, Schwickert TA *et al.*: **MicroRNA-155 suppresses**

activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity* 2008, **28**:630-638.

10. Teng G, Hakimpour P, Landgraf P, Rice A, Tuschl T, Casellas R, Papavasiliou FN: **MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase**. *Immunity* 2008, **28**:621-629.

11. • de Yebenes VG, Belver L, Pisano DG, Gonzalez S, Villasante A, Croce C, He L, Ramiro AR: **miR-181b negatively regulates activation-induced cytidine deaminase in B cells**. *J Exp Med* 2008, **205**:2199-2206.

References [9] and [10] demonstrate that microRNA-155 (miR-155) is upregulated in B lymphocytes undergoing class-switch recombination and regulate the transcription of AID by targeting the 3′ untranslated region of the AID mRNA, suggesting the possible role of miR-155 as a tumor suppressor by inhibiting the genotoxic effect of AID. Similarly, reference 11 showed the negative regulation of AID transcription by another microRNA-181b.

12. McBride KM, Gazumyan A, Woo EM, Schwickert TA, Chait BT, Nussenzweig MC: **Regulation of class switch recombination and somatic mutation by AID phosphorylation**. *J Exp Med* 2008, **205**:2585-2594.

13. Cheng HL, Vuong BQ, Basu U, Franklin A, Schwer B, Astarita J, Phan RT, Datta A, Manis J, Alt FW *et al.*: **Integrity of the AID serine-38 phosphorylation site is critical for class switch recombination and somatic hypermutation in mice**. *Proc Natl Acad Sci USA* 2009, **106**:2717-2722.

14. Aoufouchi S, Faili A, Zober C, D'Orlando O, Weller S, Weill JC, Reynaud CA: **Proteasomal degradation restricts the nuclear lifespan of AID**. *J Exp Med* 2008, **205**:1357-1368.

15. Patenaude AM, Orthwein A, Hu Y, Campo VA, Kavli B, Buschiazzo A, Di Noia JM: **Active nuclear import and cytoplasmic retention of activation-induced deaminase**. *Nat Struct Mol Biol* 2009, **16**:517-527.

16. Geisberger R, Rada C, Neuberger MS: **The stability of AID and its function in class-switching are critically sensitive to the identity of its nuclear-export sequence**. *Proc Natl Acad Sci USA* 2009, **106**:6736-6741.

17. Casellas R, Yamane A, Kovalchuk AL, Potter M: **Restricting activation-induced cytidine deaminase tumorigenic activity in B lymphocytes**. *Immunology* 2009, **126**:316-328.

18. Ito S, Nagaoka H, Shinkura R, Begum N, Muramatsu M, Nakata M, Honjo T: **Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1**. *Proc Natl Acad Sci USA* 2004, **101**:1975-1980.

19. 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**:18470-18476.

20. Muto T, Muramatsu M, Taniwaki M, Kinoshita K, Honjo T: **Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene**. *Genomics* 2000, **68**:85-88.

21. Matsumoto Y, Marusawa H, Kinoshita K, Endo Y, Kou T, Morisawa T, Azuma T, Okazaki IM, Honjo T, Chiba T: ***Helicobacter pylori* infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium**. *Nat Med* 2007, **13**:470-476.

22. Hatakeyama M: ***Helicobacter pylori* and gastric carcinogenesis**. *J Gastroenterol* 2009, **44**:239-248.

23. Viala J, Chaput C, Boneca IG, Cardona A, Girardin SE, Moran AP, Athman R, Memet S, Huerre MR, Coyle AJ *et al.*: **Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori* cag pathogenicity island**. *Nat Immunol* 2004, **5**:1166-1174.

24. Crouch EE, Li Z, Takizawa M, Fichtner-Feigl S, Gourzi P, Montano C, Feigenbaum L, Wilson P, Janz S, Papavasiliou FN *et al.*: **Regulation of AID expression in the immune response**. *J Exp Med* 2007, **204**:1145-1156.

25. Yadav A, Olaru A, Saltis M, Setren A, Cerny J, Livak F: **Identification of a ubiquitously active promoter of the murine activation-induced cytidine deaminase (AICDA) gene**. *Mol Immunol* 2006, **43**:529-541.

26. Park SR, Zan H, Pal Z, Zhang J, Al-Qahtani A, Pone EJ, Xu Z, Mai T, Casali P: **HoxC4 binds to the promoter of the cytidine deaminase AID gene to induce AID expression, class-switch DNA recombination and somatic hypermutation**. *Nat Immunol* 2009, **10**:540-550.

27. Tran TH, Nakata M, Suzuki K, Begum NA, Shinkura R, Fagarasan S, Honjo T, Nagaoka H: **B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers**. *Nat Immunol* 2010, **11**:148-154.

28. Dedeoglu F, Horwitz B, Chaudhuri J, Alt FW, Geha RS: **Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB**. *Int Immunol* 2004, **16**:395-404.

29. •• Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, Schatz DG: **Two levels of protection for the B cell genome during somatic hypermutation**. *Nature* 2008, **451**:841-845.

By extensive sequencing of murine B cell genes, the authors demonstrate that AID acts broadly on the genome and approximately 25% of expressed genes analyzed in this study accumulated mutations possibly induced by AID activity in germinal center B cells.

30. Robbiani DF, Bothmer A, Callen E, Reina-San-Martin B, Dorsett Y, Difilippantonio S, Bolland DJ, Chen HT, Corcoran AE, Nussenzweig A *et al.*: **AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations**. *Cell* 2008, **135**:1028-1038.

31. • Robbiani DF, Bunting S, Feldhahn N, Bothmer A, Camps J, Deroubaix S, McBride KM, Klein IA, Stone G, Eisenreich TR *et al.*: **AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations**. *Mol Cell* 2009, **36**:631-641.

Demonstration of the capacity of AID to produce paired DNA double-strand break throughout the genome that can lead to lymphoma-associated chromosome translocation in mature B cells.

32. 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**:1173-1181.

33. 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**:2735-2740.

34. Isaacson P, Wright DH: **Malignant lymphoma of mucosa-associated lymphoid tissue. A distinctive type of B-cell lymphoma**. *Cancer* 1983, **52**:1410-1416.

35. Zucca E, Bertoni F, Roggero E, Cavalli F: **The gastric marginal zone B-cell lymphoma of MALT type**. *Blood* 2000, **96**:410-419.

36. Suarez F, Lortholary O, Hermine O, Lecuit M: **Infection-associated lymphomas derived from marginal zone B cells: a model of antigen-driven lymphoproliferation**. *Blood* 2006, **107**:3034-3044.

37. Isaacson PG, Du MQ: **MALT lymphoma: from morphology to molecules**. *Nat Rev Cancer* 2004, **4**:644-653.

38. Parsonnet J, Hansen S, Rodriguez L, Gelb AB, Warnke RA, Jellum E, Orentreich N, Vogelman JH, Friedman GD: ***Helicobacter pylori* infection and gastric lymphoma**. *N Engl J Med* 1994, **330**:1267-1271.

39. Zullo A, Hassan C, Andriani A, Cristofari F, De Francesco V, Ierardi E, Tomao S, Morini S, Vaira D: **Eradication therapy for *Helicobacter pylori* in patients with gastric MALT lymphoma: a pooled data analysis**. *Am J Gastroenterol* 2009, **104**:1932-1937 quiz 1938.

40. Wotherspoon AC, Doglioni C, Diss TC, Pan L, Moschini A, de Boni M, Isaacson PG: **Regression of primary low-grade B-cell gastric lymphoma of mucosa-associated lymphoid**

tissue type after eradication of *Helicobacter pylori*. *Lancet* 1993, **342**:575-577.

41. Greeve J, Philipsen A, Krause K, Klapper W, Heidorn K, Castle BE, Janda J, Marcu KB, Parwaresch R: **Expression of activation-induced cytidine deaminase in human B-cell non-Hodgkin lymphomas**. *Blood* 2003, **101**:3574-3580.

42. • Pasqualucci L, Bhagat G, Jankovic M, Compagno M, Smith P, Muramatsu M, Honjo T, Morse HC 3rd, Nussenzweig MC, Dalla-Favera R: **AID is required for germinal center-derived lymphomagenesis**. *Nat Genet* 2008, **40**:108-112.
By crossing the mouse model of B cell lymphoma with the AID deficient mouse, the authors showed that AID deficiency prevents Bcl6-dependent, germinal center-derived B cell lymphomas.

43. Bombardieri M, Barone F, Humby F, Kelly S, McGurk M, Morgan P, Challacombe S, De Vita S, Valesini G, Spencer J *et al.*: **Activation-induced cytidine deaminase expression in follicular dendritic cell networks and interfollicular large B cells supports functionality of ectopic lymphoid neogenesis in autoimmune sialoadenitis and MALT lymphoma in Sjogren's syndrome**. *J Immunol* 2007, **179**:4929-4938.

44. Deutsch AJ, Aigelsreiter A, Staber PB, Beham A, Linkesch W, Guelly C, Brezinschek RI, Fruhwirth M, Emberger W, Buettner M *et al.*: **MALT lymphoma and extranodal diffuse large B-cell lymphoma are targeted by aberrant somatic hypermutation**. *Blood* 2007, **109**:3500-3504.

45. Delker RK, Fugmann SD, Papavasiliou FN: **A coming-of-age story: activation-induced cytidine deaminase turns 10**. *Nat Immunol* 2009, **10**:1147-1153.

46. Marusawa H: **Aberrant AID expression and human cancer development**. *Int J Biochem Cell Biol* 2008, **40**:1399-1402.

47. 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**:5587-5595.

48. •• Takai A, Toyoshima T, Uemura M, Kitawaki Y, Marusawa H, Hiai H, Yamada S, Okazaki IM, Honjo T, Chiba T *et al.*: **A novel mouse model of hepatocarcinogenesis triggered by AID causing deleterious p53 mutations**. *Oncogene* 2008.
Pioneering study demonstrating the crucial role of mutational accumulation in immature stem cells on the development of cancer cells. The authors show that constitutive AID expression in cells producing a marker of immature hepatocytes resulted in the appearance of hepatocellular carcinoma.

49. Komori J, Marusawa H, Machimoto T, Endo Y, Kinoshita K, Kou T, Haga H, Ikai I, Uemoto S, Chiba T: **Activation-induced cytidine deaminase links bile duct inflammation to human cholangiocarcinoma**. *Hepatology* 2008, **47**:888-896.

50. • Endo Y, Marusawa H, Kou T, Nakase H, Fujii S, Fujimori T, Kinoshita K, Honjo T, Chiba T: **Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers**. *Gastroenterology* 2008, **135**: 889–898, e881–883.
The authors show that AID expression in human colonic epithelial cells is induced in response to Th2-driven cytokines IL-4 and IL-13, which are activated in inflammatory bowel disease. This is the first study to provide the possible link between colonic inflammation, AID expression and colorectal cancer development.

51. Aoi T, Marusawa H, Sato T, Chiba T, Maruyama M: **Risk of subsequent development of gastric cancer in patients with previous gastric epithelial neoplasia**. *Gut* 2006, **55**:588-589.

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-

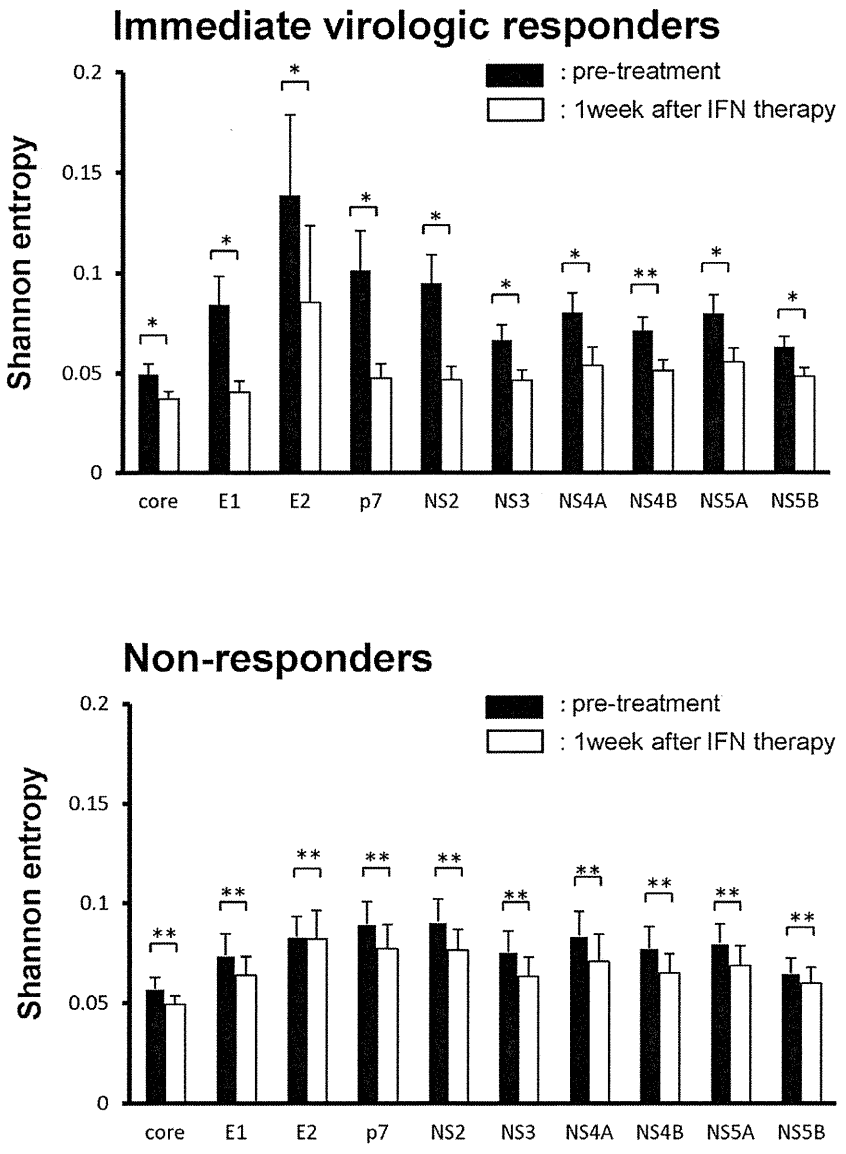


Figure 1. Changes in the genetic complexity of each HCV genomic region before and after the administration of peg-IFNα2b plus RBV. Shannon entropy values at baseline (black bar) and 1 week after initiation of treatment with peg-IFNα2b plus RBV (white bar) in 8 immediate virologic responders (A) and in 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-

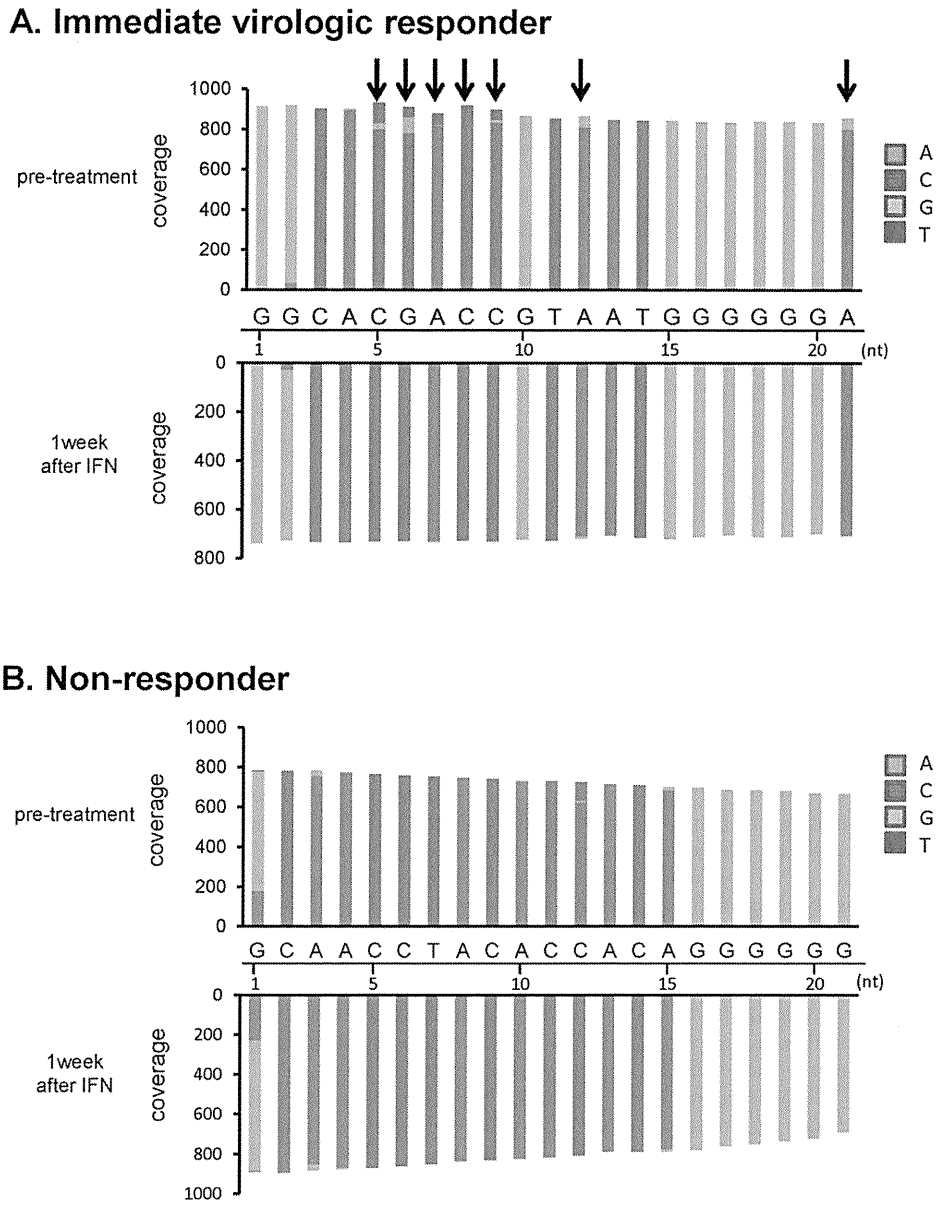


Figure 2. Ratio of mutated nucleotides in the HVR1 region before and after administration of peg-IFNα2b plus RBV therapy. Representative results of a immediate virologic responder (Patient#3) (**A**) and a non-responder (Patient#9) (**B**) are shown. The read numbers (coverage) at each nucleotide position 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

that M414T mutants preexisting at a frequency of 0.22% and 0.18% in the treatment-naïve replicon population rapidly increased upon treatment with DAAs in a dose-dependent manner, reaching frequencies of 25% and 60% after 4 days of treatment. These findings suggest that those preexisting minor mutants might cause resistance against DAAs through the selection of dominant mutations. Thus, the significance of low-abundance variants in treatment-naïve patients requires further exploration.

The present study raises two limitations of ultra-deep parallel sequencing technology in the analyses of viral quasispecies. First, because the massive parallel ultra-deep sequencing platform is based on multitudinous short reads, it is difficult to separately evaluate the association between nucleotide sites mapped to different viral genome regions in a single viral clone. Indeed, it is difficult to clarify the potential mutational linkage between different viral genomic regions because of the short read length of the shotgun sequencing approach. Second, it is difficult to accurately analyze highly polymorphic regions such as the HVR by ultra-deep sequencing, because mutation findings strongly depend on mapping to the reference genome sequences. Thus, utilization of both conventional and ultra-deep sequencing technology might be necessary to fully clarify the significance and clinical relevance of the prominent HCV genomic heterogeneity.

In summary, using ultra-deep sequencing technology, we clearly demonstrated the extremely large genetic complexity in the genotype1b HCV derived from chronically infected patients. Although there was no significant difference in the level of viral complexity between immediate virologic responders and non-responders at baseline, immediate virologic responders, but not non-responders, showed a rapid reduction in the viral sequence variability at an early phase of peg-IFNα2b plus RBV administration. We also showed that drug-resistant mutants were widely present in treatment-naïve HCV-infected patients, indicating a putative risk for the expansion of resistant clones to DAAs. Further studies with a large number of patients are needed to fully elucidate the significance of viral heterogeneity in the clinical outcome of patients receiving anti-viral therapy.

Materials and Methods

Patients

The participants comprised 27 Japanese adult chronic hepatitis patients with genotype 1b HCV infection and the mean baseline level of serum HCV RNA determined by TaqMan RT-PCR (Applied Biosystems, Foster City, CA) was 6.9 log IU/ml. All patients received conventional peg-IFNα2b plus RBV combination therapy (Schering-Plough, Kenilworth, NJ) at Kyoto University and affiliated hospitals from February 2007 to December 2008. Indications for IFN-based combination therapy included high serum values of alanine aminotransferase and positivity for serum HCV RNA. Patients were treated with peg-IFNα2b (1.5 μg/kg) once per week, combined with daily oral RBV for 48 weeks [38]. The RBV dose was 600 mg/day in patients weighing less than 60 kg, 800 mg/day in those weighing at least 60 kg but less than 80 kg, and 1000 mg/day in those weighing 80 kg or more.

In this study, immediate virologic responders were defined as patients whose serum HCV RNA levels declined by more than 2 log IU/mL after 1 week of treatment with peg-IFNα2b plus RBV, while non-responders were defined as those whose serum HCV RNA levels declined less than 2 log IU/mL after peg-IFNα2b plus RBV administration. Of the original 27 patients, the serum before and 1 week after initiating treatment with peg-IFNα2b plus RBV of 16 cases was available for further analyses, and 8 of these cases were defined as immediate virologic responders and 8 cases were defined as non-responders. Among these non-responder cases, the serum HCV RNA levels in 6 of 8 (75.0%) patients changed by less than 1 log IU/mL after 1 week of treatment. The decline in HCV RNA levels in the remaining 2 cases was slightly over 1 log IU/mL (1.2 and 1.4 log IU/mL).

The ethics committee at Kyoto University approved the studies, and written informed consent for participation in this study was obtained from all patients.

Direct population Sanger sequencing

To define the representative reference sequences of full-length HCV in each clinical specimen, all samples were first subjected to direct population Sanger sequencing using Applied Biosystems 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA) [39]. Serum samples were obtained before the start and at 1 week after initiation of peg-IFNα2b and RBV combination therapy. Total RNA was extracted from 140 μL of serum using a QIAamp Viral RNA Mini kit (QIAGEN, Valencia, CA) and reverse-transcribed in a volume of 20 μL with the One step RNA PCR Kit AMV (Takara Bio, Ohtsu, Japan).

HCV genomes were amplified using Phusion High-Fidelity DNA polymerase (FINZYMES, Espoo, Finland). Oligonucleotide primers were designed to amplify the first-half (~5,000 bps) and the latter-half (~4,500 bps) of the genotype 1b HCV genome sequences (Table S3).

PCR products purified by the QIAquick Gel Extraction kit (Qiagen) were assayed for direct sequencing [40]. Nucleotide sequences of PCR products were determined using an ABI Prism Big Dye Terminator Ready Reaction Kit (Applied Biosystems). The serum of a healthy volunteer was used as a negative control.

Massive-parallel ultra-deep sequencing

Paired-end sequencing with multiplexed tags was carried out using the Illumina Genome Analyzer II. End-repair of DNA fragments, addition of adenine to the 3′ ends of DNA fragments, adaptor ligation, and PCR amplification by Illumina-paired end PCR primers were performed as described previously [41].

Briefly, the viral genome sequences were amplified with high-fidelity PCR and sheared by nebulization using 32 psi N2 for 8 min and the sheared fragments were purified and concentrated using QIAquick PCR purification Kit (Qiagen). The overhangs resulting from fragmentation were then converted into blunt ends using T4 DNA polymerase and Klenow enzymes, followed by the addition of terminal 3′ adenine-residues. Next, one of the adaptors containing six unique base pair (bp) tags, such as "ATCACG" and "CGATGT" (Multiplexing Sample Preparation Oligonucleotide Kit, Illumina), was ligated to each fragment using DNA ligase. Adaptor-ligated DNAs in the range of 200 to 350 bp were then size-selected by agarose gel electrophoresis. These libraries were amplified independently using a minimal PCR amplification step of 18 cycles with Phusion High-Fidelity DNA polymerase and then purified using a QIAquick PCR purification Kit for a downstream assay. Cluster generation and sequencing was performed for 64 cycles on the Illumina Genome Analyzer II following the manufacturer's instructions. Obtained images were analyzed and base-called using GA pipeline software version 1.4 with default settings provided by Illumina.

Genome Analyzer sequence data analysis

Using the high performance alignment software "NextGene" (SoftGenetics, State College, PA), the 64 base tags obtained from the Genome Analyzer II reads were aligned to the reference HCV RNA sequences of ~9200 bp that were determined by direct population Sanger sequencing in each clinical specimen. Entire reads were removed from the analysis when the median quality value score was below 20 and when containing more than 3 uncalled nucleotides. The low quality bases were trimmed from reads when more than 3 consecutive bases fell below a quality value score of 16. Based on the above criteria, reads with 90% or more bases matching a particular position of the reference sequence were aligned. Each position of the viral genome was assigned a coverage depth, representing the number of times the nucleotide position was sequenced.

Statistical analysis

Results are expressed as mean or median values and range (minimum and maximum). Pretreatment values were compared using the Mann–Whitney U-test. Categorical variables were analyzed by Fisher's exact test. P values of less than 0.05 were considered statistically significant. The viral quasispecies nature was evaluated by analyzing the genetic complexity based on the number of different sequences present in the population. Genetic complexity was determined by Shannon entropy values calculated as follows:

$$Sn = -\frac{\sum_{i=1}^{n} fi(\ln fi)}{N}$$

where n is the number of different species identified, fi is the observed frequency of the particular variant in the quasispecies, and N is the total number of clones analyzed [23,42]. Statistical comparisons of complexity between two groups were made using the Wilcoxon rank sum test or the Mann–Whitney U-test.

Supporting Information

Figure S1 Relationship between serum HCV RNA levels and the number of resistant mutant. No correlation was observed between serum HCV RNA levels (log IU/ml) and the number of resistant mutations against direct-acting antivirals in 27 cases in this study.
(TIF)

Table S1 Aligned reads, nucleotides, and mean coverage of each reference sequence in all patients.
(DOC)

Table S2 Mean genetic complexity in each viral genomic region of the 8 immediate virologic responders and 8 non-responders at pre-treatment and 1 week after IFN therapy.
(DOC)

Table S3 The oligonucleotide primers for PCR amplifying the whole HCV sequences.
(DOC)

Acknowledgments

The following institutions and investigators also participated in the study: Dr. M. Umeda, Dr. T. Kimura, Hyogo Prefectural Amagasaki Hospital; Dr. S. Yazumi, Kitano Hospital; Prof. K. Okazaki, Kansai Medical University; Dr. T. Nakamura, Kansai Denryoku Hospital; Dr. H. Yamada, Shinko Hospital; Dr. H. Kokuryu, Kyoto Katsura Hospital; Dr. A. Nakamura, Sanda City Hospital; Dr. C. Kawanami, Dr. N. Miyake, Ohtsu Red Cross Hospital; and Dr. K. Kajimura, Kishiwada City Hospital.

We thank Prof. A. Sekine for helpful advice, and Dr. Y. Matsumoto, Dr. Y. Endo, Dr. M. Wada, Dr. A. Takai and Ms. C. Hirano for ultra-deep sequencing analysis.

Author Contributions

Conceived and designed the experiments: AN HM. Performed the experiments: AN HM. Analyzed the data: AN HM NN TF FS KS TC YU. Contributed reagents/materials/analysis tools: AN HM YO YY TI TT. Wrote the paper: AN HM KT TC.

References

1. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, et al. (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 244: 359–362.
2. Grakoui A, Wychowski C, Lin C, Feinstone S, Rice C (1993) Expression and identification of hepatitis C virus polyprotein cleavage products. J Virol 67: 1385–1395.
3. Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Shimotohno K (1991) Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. Proc Natl Acad Sci USA 88: 5547–5551.
4. Steinhauer DA, Domingo E, Holland JJ (1992) Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. Gene 122: 281–288.
5. Neumann A, Lam N, Dahari H, Gretch D, Wiley T, et al. (1998) Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. Science 282: 103–107.
6. Ogata N, Alter H, Miller R, Purcell R (1991) Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. Proc Natl Acad Sci USA 88: 3392–3396.
7. Okamoto H, Kojima M, Okada S, Yoshizawa H, Iizuka H, et al. (1992) Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability. Virology 190: 894–899.
8. Martell M, Esteban J, Quer J, Genescà J, Weiner A, et al. (1992) Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. J Virol 66: 3225–3229.
9. Pawlotsky J (2006) Hepatitis C virus population dynamics during infection. Curr Top Microbiol Immunol 299: 261–284.
10. Argentini C, Genovese D, Dettori S, Rapicetta M (2009) HCV genetic variability: from quasispecies evolution to genotype classification. Future Microbiol 4: 359–373.
11. Margulies M, Egholm M, Altman W, Attiya S, Bader J, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–380.
12. Mardis E (2009) New strategies and emerging technologies for massively parallel sequencing: applications in medical research. Genome Med 1: 40.
13. Hedskog C, Mild M, Jernberg J, Sherwood E, Bratt G, et al. (2010) Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. PLoS One 5: e11345.
14. Simen BB, Simons JF, Hullsiek KH, Novak RM, Macarthur RD, et al. (2009) Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. J Infect Dis 199: 693–701.
15. Tsibris AM, Korber B, Arnaout R, Russ C, Lo CC, et al. (2009) Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. PLoS One 4: e5683.
16. Wang GP, Sherrill-Mix SA, Chang KM, Quince C, Bushman FD (2010) Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. J Virol 84: 6218–6228.
17. Margeridon-Thermet S, Shulman NS, Ahmed A, Shahriar R, Liu T, et al. (2009) Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients. J Infect Dis 199: 1275–1285.
18. Marusawa H, Hijikata M, Chiba T, Shimotohno K (1999) Hepatitis C virus core protein inhibits Fas- and tumor necrosis factor alpha-mediated apoptosis via NF-kappaB activation. J Virol 73: 4713–4720.
19. Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Ohkoshi S, et al. (1991) Hypervariable regions in the putative glycoprotein of hepatitis C virus. Biochem Biophys Res Commun 175: 220–228.
20. Sarrazin C, Zeuzem S (2010) Resistance to direct antiviral agents in patients with hepatitis C virus infection. Gastroenterology 138: 447–462.
21. Kuntzen T, Timm J, Berical A, Lennon N, Berlin AM, et al. (2008) Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naïve patients. Hepatology 48: 1769–1778.
22. Lauring AS, Andino R (2010) Quasispecies theory and the behavior of RNA viruses. PLoS Pathog 6: e1001005.
23. Fishman SL, Branch AD (2009) The quasispecies nature and biological implications of the hepatitis C virus. Infect Genet Evol 9: 1158–1167.

24. Xu Z, Fan X, Xu Y, Di Bisceglie AM (2008) Comparative analysis of nearly full-length hepatitis C virus quasispecies from patients experiencing viral breakthrough during antiviral therapy: clustered mutations in three functional genes, E2, NS2, and NS5a. J Virol 82: 9417–9424.
25. Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, et al. (1996) Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N Engl J Med 334: 77–81.
26. Pascu M, Martus P, Höhne M, Wiedenmann B, Hopf U, et al. (2004) Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences. Gut 53: 1345–1351.
27. Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, et al. (2007) Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. J Hepatol 46: 403–410.
28. Farci P, Strazzera R, Alter HJ, Farci S, Degioannis D, et al. (2002) Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome. Proc Natl Acad Sci USA 99: 3081–3086.
29. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, et al. (2009) Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 461: 399–401.
30. Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, et al. (2009) IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. Nat Genet 41: 1100–1104.
31. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, et al. (2009) Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 41: 1105–1109.
32. Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, et al. (2009) Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 461: 798–801.
33. Thompson AJ, McHutchison JG (2009) Antiviral resistance and specifically targeted therapy for HCV (STAT-C). J Viral Hepat 16: 377–387.
34. López-Labrador FX, Moya A, Gonzàlez-Candelas F (2008) Mapping natural polymorphisms of hepatitis C virus NS3/4A protease and antiviral resistance to inhibitors in worldwide isolates. Antivir Ther 13: 481–494.
35. Colson P, Brouk N, Lembo F, Castellani P, Tamalet C, et al. (2008) Natural presence of substitution R155K within hepatitis C virus NS3 protease from a treatment-naïve chronically infected patient. Hepatology 47: 766–767.
36. Bartels DJ, Zhou Y, Zhang EZ, Marcial M, Byrn RA, et al. (2008) Natural prevalence of hepatitis C virus variants with decreased sensitivity to NS3.4A protease inhibitors in treatment-naive subjects. J Infect Dis 198: 800–807.
37. Lu L, Mo H, Pilot-Matias TJ, Molla A (2007) Evolution of resistant M414T mutants among hepatitis C virus replicon cells treated with polymerase inhibitor A-782759. Antimicrob Agents Chemother 51: 1889–1896.
38. Iwai A, Marusawa H, Takada Y, Egawa H, Ikeda K, et al. (2006) Identification of novel defective HCV clones in liver transplant recipients with recurrent HCV infection. J Viral Hepat 13: 523–531.
39. Matsumoto Y, Marusawa H, Kinoshita K, Endo Y, Kou T, et al. (2007) Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat Med 13: 470–476.
40. Marusawa H, Uemoto S, Hijikata M, Ueda Y, Tanaka K, et al. (2000) Latent hepatitis B virus infection in healthy individuals with antibodies to hepatitis B core antigen. Hepatology 31: 488–495.
41. Ruike Y, Imanaka Y, Sato F, Shimizu K, Tsujimoto G (2010) Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. BMC Genomics 11: 137.
42. Wolinsky SM, Korber BT, Neumann AU, Daniels M, Kunstman KJ, et al. (1996) Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. Science 272: 537–542.

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.

Cancer Cell Biology

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*).

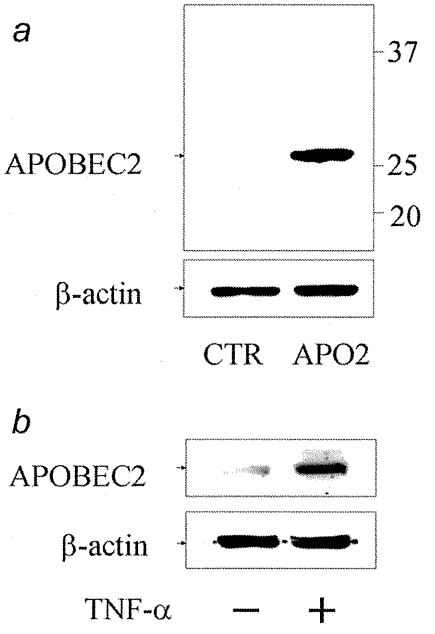


Figure 1. Detection of human APOBEC2 protein in hepatocytes by a specific anti-APOBEC2 antibody. (*a*) Huh7 cells were transfected with plasmid to induce the expression of human APOBEC2 (APO2) or control vector (CTR). After 48 hr, lysates of transfected cells were immunoblotted with anti-APOBEC2 antibody (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

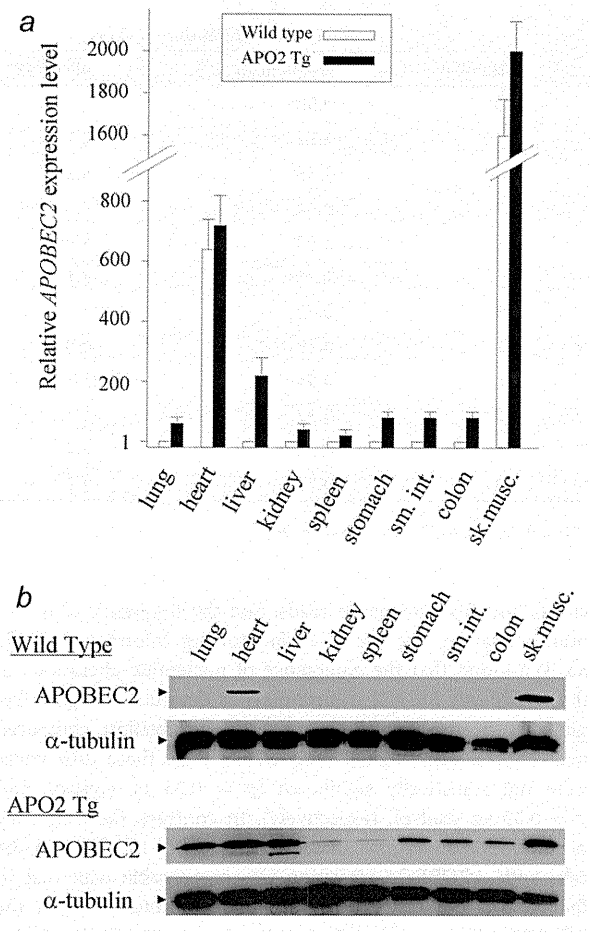


Figure 2. Expression analyses of APOBEC2 Tg mice. (*a*) Relative expression levels of *APOBEC2* transcripts calibrated by the amount of *18S rRNA* for indicated organs of adult APOBEC2 Tg mice (48-week-old) and their wild-type littermates. Data shown are mean results of quantitative real-time RT-PCR 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