Gastroenterology 2014;146:222–232

BASIC AND TRANSLATIONAL—LIVER

Leptin Receptor Somatic Mutations Are Frequent in HCV-Infected Cirrhotic Liver and Associated With Hepatocellular Carcinoma

Atsuyuki Ikeda,[1] Takahiro Shimizu,[1] Yuko Matsumoto,[1] Yosuke Fujii,[1] Yuji Eso,[1] Tadashi Inuzuka,[1] Aya Mizuguchi,[1] Kazuharu Shimizu,[2] Etsuro Hatano,[3] Shinji Uemoto,[3] Tsutomu Chiba,[1] and Hiroyuki Marusawa[1]

Departments of [1]Gastroenterology and Hepatology and [3]Surgery, Graduate School of Medicine, and [2]Department of Nanobio Drug Discovery, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, Japan

See Covering the Cover synopsis on page 2.

BACKGROUND & AIMS: Hepatocellular carcinoma develops in patients with chronic hepatitis or cirrhosis via a stepwise accumulation of various genetic alterations. To explore the genetic basis of development of hepatocellular carcinoma in hepatitis C virus (HCV)-associated chronic liver disease, we evaluated genetic variants that accumulate in nontumor cirrhotic liver. **METHODS:** We determined the whole exome sequences of 7 tumors and background cirrhotic liver tissues from 4 patients with HCV infection. We then performed additional sequencing of selected exomes of mutated genes, identified by whole exome sequencing, and of representative tumor-related genes on samples from 22 cirrhotic livers with HCV infection. We performed in vitro and in vivo functional studies for one of the mutated genes. **RESULTS:** Whole exome sequencing showed that somatic mutations accumulated in various genes in HCV-infected cirrhotic liver tissues. Among the identified genes, the leptin receptor gene (*LEPR*) was one of the most frequently mutated in tumor and nontumor cirrhotic liver tissue. Selected exome sequencing analyses detected *LEPR* mutations in 12 of 22 (54.5%) nontumorous cirrhotic livers. In vitro, 4 of 7 (57.1%) *LEPR* mutations found in cirrhotic livers reduced phosphorylation of STAT3 to inactivate LEPR-mediated signaling. Moreover, 40% of Lepr-deficient (C57BL/KsJ-*db/db*) mice developed liver tumors after administration of thioacetamide compared with none of the control mice. **CONCLUSIONS:** Based on analysis of liver tissue samples from patients, somatic mutations accumulate in *LEPR* in cirrhotic liver with chronic HCV infection. These mutations could disrupt LEPR signaling and increase susceptibility to hepatocarcinogenesis.

Keywords: Liver Cancer; Whole Exome Sequencing; Genetics; STAT3.

Chronic inflammation plays an important role in the development of various human cancers. Indeed, many human cancers are closely associated with chronic inflammation, such as *Helicobacter pylori*–associated gastric cancer and inflammatory bowel disease–associated colorectal cancer.[1,2] On the other hand, tumor cells are believed to be generated by a stepwise accumulation of genetic alterations in various tumor-related genes during the process of inflammation-associated carcinogenesis.[3–6] Thus, it is reasonable to assume that somatic mutations latently accumulate in inflamed tissues, where the risk of tumorigenesis is high. Consistent with this hypothesis, several studies have shown frequent somatic mutations in nontumorous inflammatory tissues.[7,8] To clarify the mechanisms of inflammation-associated carcinogenesis, it is important to unveil the genetic alterations that occur in the inflamed tissues before tumor development. The diversity of mutated genes and the low frequency of genetic alterations compared with tumor tissues, however, are obstacles to revealing the landscape of accumulated genetic aberrations in chronically inflamed nontumorous tissues.

Several possible molecular mechanisms have been proposed for the genetic alterations occurring in the inflammatory condition.[9] We recently showed that the expression of activation-induced cytidine deaminase (AID), a DNA/RNA mutator enzyme family member, links inflammation to an enhanced susceptibility to genetic aberration during the development of various gastrointestinal and hepatobiliary cancers.[10–12] One clear example of inflammation-associated cancer is human hepatocellular carcinoma (HCC). HCC arises in the background of chronic inflammation caused by hepatitis C virus (HCV) infection.[13] We showed that aberrant AID expression triggered by HCV infection and the resultant inflammatory response leads to the generation of somatic mutations in various tumor-related genes in the inflamed liver tissues.[14,15] The target genes of AID-mediated mutagenesis in the inflamed hepatocytes, however, remain unclear.

Recent advances in sequencing technology have enabled us to reveal the whole picture of human genome sequences in association with the risk of development of a variety of human diseases, including cancers.[16,17] Whole exome capture

Abbreviations used in this paper: **AID, activation-induced cytidine deaminase; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; Ig, immunoglobulin; TAA, thioacetamide.**

0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2013.09.025

has identified several candidate driver genes in various human cancers.[18–20] Although deep sequencing on tumor tissues provides the most comprehensive analysis of the cancer genome, the genetic alterations accumulated in chronically inflamed tissues might provide an additional opportunity to clarify the early genetic changes required for carcinogenesis. In the present study, we applied whole exome sequencing to not only the tumor but also nontumorous liver tissues infected with HCV and found that somatic mutations of the leptin receptor gene (*LEPR*) latently underlie a subset of the cirrhotic liver tissues, providing the putative genetic basis for HCV-associated hepatocarcinogenesis.

Materials and Methods

Whole Exome Capture and Massively Parallel Sequencing

Massively parallel sequencing was performed as described previously.[21,22] Fragmented DNA (more than 5 μg) was used to prepare each DNA sequencing library. The DNA libraries were prepared according to the instructions provided with the Illumina Preparation Kit (Illumina, San Diego, CA). Whole exome sequence capture was then performed using SeqCap EZ Human Exome Library v2.0 (Roche, Madison, WI) according to the manufacturer's instructions. Cluster generation was performed on the Illumina cluster station (using their TruSeq PE Cluster Kit v5). Paired-end sequence for 2×76 base pairs was performed on the Illumina Genome Analyzer IIx (using their SBS Kits v5). Data collection and base calling were performed using SCS v2.9/RTA 1.9, and the resultant data files were converted to the FASTQ format.

Selected Exome Capture and Massively Parallel Sequencing

Fragmented DNA (1 μg) was used to prepare each DNA sequencing library. The DNA libraries were prepared using TruSeq DNA Sample Prep Kits (Illumina) according to the manufacturer's protocol. Selected gene capture (*TP53*, *CTNNB1*, *LEPR*) was performed using the SeqCap EZ Choice library (Roche) according to the manufacturer's recommendations. Cluster generation and multiplexed paired-end sequencing for $2 \times 71 + 7$ base pairs was performed as described previously. Data collection and base calling were performed as described previously and demultiplexed using CASAVA version 1.8.2 software (Illumina) with the default settings.

Sequence data analysis and variant filtering, patients, cell culture and transfection, immunoblotting analysis, and animal experiments are described in Supplementary Methods and Supplementary Figures 1 and 2.

Results

Whole Exome Sequencing Identified the Mutation Signature of Synchronous HCCs in Patients With Chronic HCV Infection

To explore the genetic basis of HCV-associated hepatocarcinogenesis, we first determined the whole exome sequences in matched pairs of HCC and background liver tissues obtained from 4 patients with chronic HCV infection (Supplementary Table 1, patients 1–4). Three of these patients had multiple HCCs, and one had a solitary HCC in the liver. To compare the mutation signature in synchronous HCCs that developed in the same background liver, we determined the whole exome sequences of 2 representative HCCs in 3 cases and a solitary HCC in the remaining case (Figure 1). These 7 HCCs from 4 patients comprised 2 well-differentiated and 5 moderately differentiated HCCs, and the background liver tissue showed the histological characteristics of cirrhosis. To subtract the normal variants of each individual from the somatic mutations, we also determined the whole exome sequences of matched peripheral lymphocytes in each patient.

On average, we generated approximately 3.1 gigabases of sequence per sample, 80.1% of which were aligned with the human reference genome (Human Genome Build 37.3), and the mean coverage in the targeted regions was 33.8-fold (Supplementary Table 2). The variant filtering process is summarized in Supplementary Figure 1, and the overall error rate in our current platform was confirmed to be less than 0.2%, as described previously.[21] Overall, a total of 970 nucleotide positions in 768 different genes were mutated at a frequency of more than 20% of reads in the 7 HCC tissues (Supplementary Table 3). Among them, 79 genes were recurrently mutated in 2 or more tumor tissues (data not shown). These genes included representative tumor-related genes associated with HCC such as *TP53* (mutated in 2 of 7 tumors). Pathway analyses using Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/) revealed that metabolic pathway–related genes were most frequently damaged in HCC tissues (5 of 7 tumors) (Supplementary Table 4).

Interestingly, the mutation signature was remarkably different between the synchronously developed HCCs in each patient (Figure 1). In patient 3, none of the genes were commonly mutated in the 2 tumors examined, while 29 and 225 genes acquired independent somatic mutations in each tumor, respectively. In contrast, 32 genes (64.0% of mutated genes of HCC 1 in patient 1) and 9 (24.3% of mutated genes of HCC 1 in patient 2) were commonly mutated in the synchronously developed HCCs of those patients, indicating that the synchronous HCCs that developed in patient 1 or 2 shared a common pattern of genetic aberrations. These findings may suggest that the synchronous tumors in patients 1 and 2 were derived from common tumor-precursor cells or developed through intrahepatic metastasis, whereas the tumors in patient 3 developed independently in a multicentric manner.

Somatic Mutations Accumulated in the Cirrhotic Liver With HCV Infection

Whole exome sequencing also revealed a large number of nucleotide alterations in nontumorous cirrhotic liver tissues. In some cases, the total number of mutated genes in nontumorous liver was higher than that in tumor tissues, while the mutation frequency in nontumorous tissues tended to be lower than that in the matched tumor tissues (Figure 2). Sorting Intolerant From Tolerant (SIFT) functional impact predictions (http://provean.jcvi.org/index.php) revealed that the mean percentage of somatic

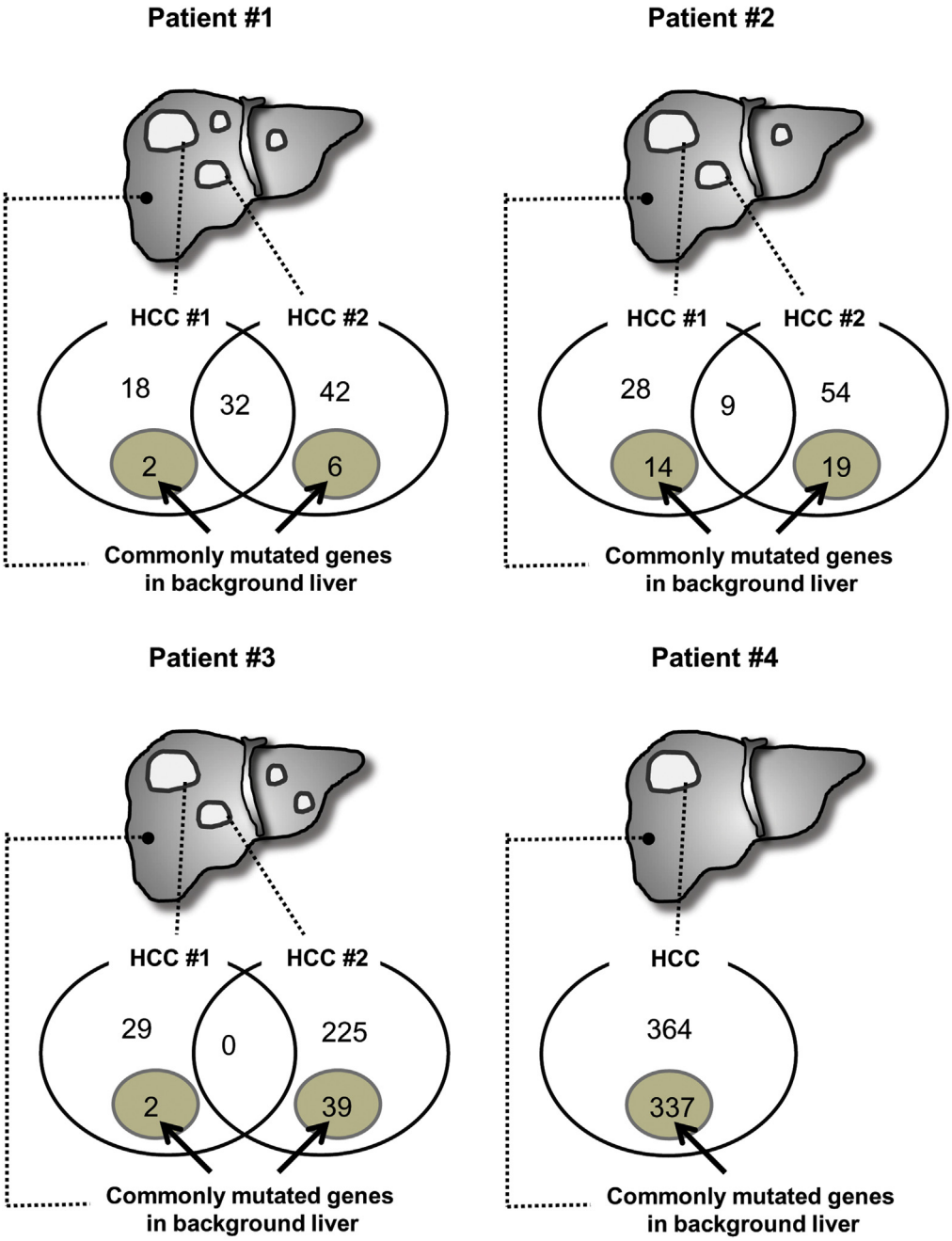

Figure 1. Schematic diagram showing the number of mutated genes in tumors and the number of genes commonly mutated in both tumor and the matched nontumorous liver tissues. Patients 1, 2, and 3 had synchronously developed HCCs, and patient 4 had a solitary HCC. Venn diagrams represent the number of mutated genes in each HCC tissue determined by whole exome sequencing. The numbers of genes commonly mutated in the synchronously developed multiple HCCs were 32, 9, and 0 in patients 1, 2, and 3, respectively. Among the mutated genes in HCC (at a frequency of more than 20% of reads), the number of genes commonly mutated in both HCC and matched nontumorous background liver (at a frequency of more than 5% of reads) is shown in *shaded circles*.

mutations predicted to be "damaging" in tumorous and nontumorous tissues was 20.4% and 13.1%, respectively, suggesting that somatic mutations that accumulated in nontumorous tissues included "passenger" mutations with less functional significance more frequently than those that accumulated in tumor tissues (Supplementary Table 3). We also identified a total of 448 indels in 7 HCC tissues (Supplementary Table 5), while fewer indels were detected in all of the nontumorous cirrhotic liver tissues examined (Supplementary Table 6). Consistent with previous studies,[19] we found that one-third of the mutations that accumulated in the exome sequences of HCC tissues were enriched as C>T; G>A transition, followed by A>G; T>C. Similar to tumor tissues, C>T; G>A transition mutations were most frequently detected in nontumorous cirrhotic tissues (Supplementary Figure 3).

The aim of this study was to identify the somatic mutations in the nontumorous HCV-positive cirrhotic liver that may contribute to tumorigenesis. Therefore, we focused on the genes commonly mutated in both tumor and nontumorous liver tissues from the same patient. Because few genes commonly acquired somatic mutations with a frequency of more than 20% both in the tumor and the matched nontumorous liver tissues, we selected potential somatic mutations in nontumorous tissues that represented more than 5% of the total reads for further evaluation

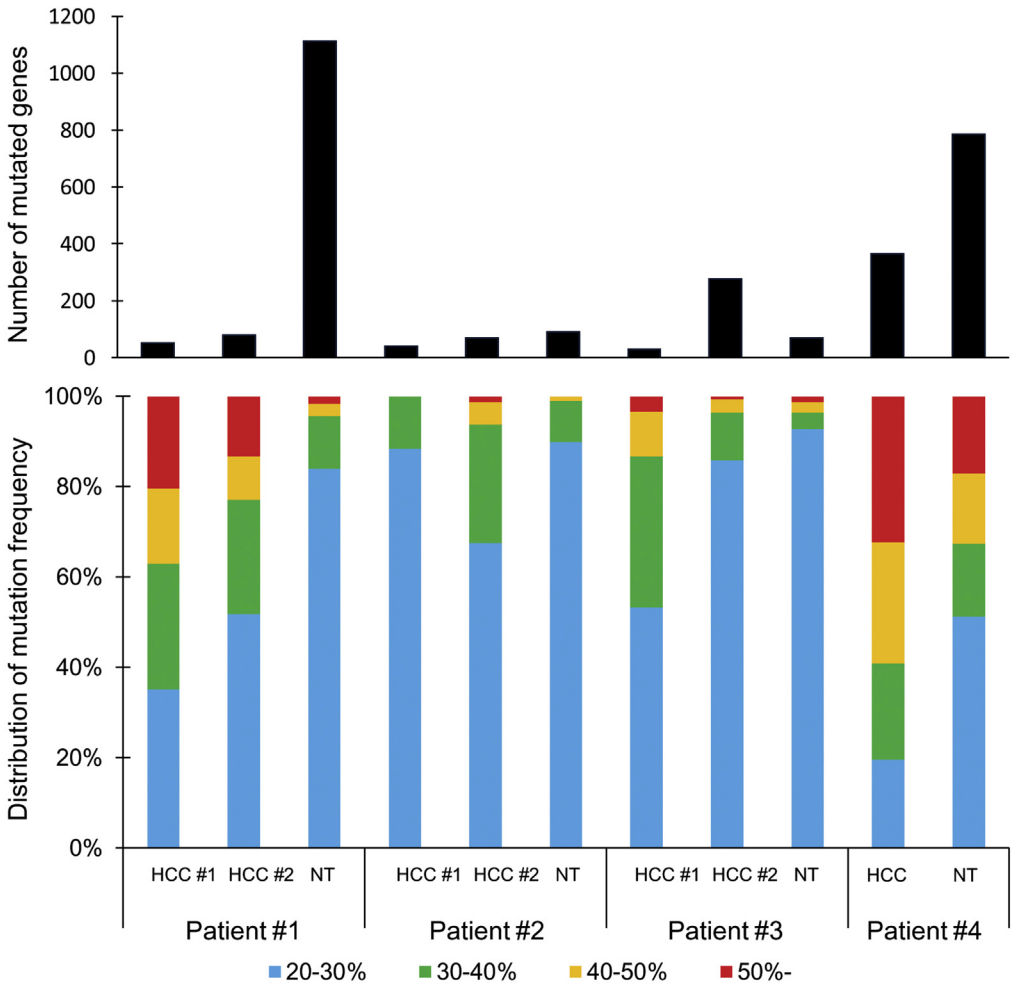

Figure 2. The number of mutated genes and the distribution of mutation frequency in tumor and nontumorous cirrhotic liver tissues. The number of mutated genes (*upper panel*) and the distribution of mutation frequency (*lower panel*) detected by whole exome sequencing in each sample are shown (at a frequency of more than 20% of reads). Patients 1, 2, and 4 had more mutated genes in nontumorous liver tissue than those in HCC, while the mutation frequency at each nucleotide position in the majority of nontumorous cirrhotic liver tissues was <30%. NT, nontumorous cirrhotic liver.

(Supplementary Figure 1). The 5% threshold in nontumorous liver was chosen because common polymorphisms in each patient were excluded by determining the nucleotide changes with a frequency of more than 5% in the matched normal samples, such as peripheral lymphocytes.[18,23]

Based on these criteria, nucleotide positions that were commonly mutated in both the tumor (at a frequency of more than 20% of reads) and the matched background liver (at a frequency of more than 5% of reads)of each patient were detected (Figure 1). Among them, we focused on 40 mutations that result in amino acid changes (Supplementary Table 7) and found that only 2 genes, *LEPR* and *ZNF408*, were recurrently mutated with a frequency greater than 5% of reads in nontumorous cirrhotic livers from 2 of the 4 patients (listed as the top 2 genes in Supplementary Table 7). Of these 2 genes, we focused on *LEPR*, which has mutations that have been correlated with various human diseases, such as obesity and metabolic disorders.[24]

Identification of LEPR as the Recurrently Mutated Gene in Cirrhotic Livers With HCV Infection

We designed a selected sequence capture system that enabled us to enrich the whole exonic sequences of the *LEPR* followed by deep sequencing. In addition, selected exonic capture of *TP53* and *CTNNB1*, the representative driver genes for hepatocarcinogenesis,[19,20,25] was performed on the same cohort. Accordingly, the selected exonic sequencing was applied to 22 additional HCV-positive cirrhotic liver tissues, 10 HCC tissues, and matched peripheral lymphocytes from 22 patients (Supplementary Table 1, patients 5–26). Selected exome sequencing generated a mean coverage of 996-, 1656-, and 2348-fold on *LEPR*, *TP53*, and *CTNNB1*, respectively (Supplementary Table 8). The variant filtering process is summarized in Supplementary Figure 2, and we detected both high-frequency (at a frequency of more than 20% of reads) and low-frequency (at a frequency of 1%–20% of reads) mutations separately.

High-frequency mutations in *TP53* and *CTNNB1* were detectable in 1 of 10 (10%) and 1 of 10 (10%) of the HCCs, respectively (Table 1), and these rates in the HCCs were consistent with recent deep-sequencing studies.[19,20] None of the nontumorous liver tissues possessed high-frequency mutations in *TP53* or *CTNNB1*; however, low-frequency mutations of *TP53* and *CTNNB1* were detected in 17 of 22 (77.3%) and 12 of 22 (54.5%) of the nontumorous livers, respectively. These findings indicated that somatic mutations in the representative cancer driver genes latently

Table 1. Number of Tumor Tissues and Nontumorous Cirrhotic Liver Tissues With Somatic Mutations at High and Low Frequencies in *TP53*, *CTNNB1*, and *LEPR*

	LEPR	*TP53*	*CTNNB1*
High-frequency mutations (>20%)			
Tumor (n = 10)	0	1	1
Nontumor (n = 22)	1[a]	0	0
Low-frequency mutations (1%–20%)			
Tumor (n = 10)	9	8	9
Nontumor (n = 22)	12[a]	17	12

[a]One patient had both high-frequency and low-frequency mutations in *LEPR*.

accumulated with a relatively low frequency in the cirrhotic livers with HCV infection.

Interestingly, we also found high- and/or low-frequency mutations in *LEPR* in both tumor and nontumorous liver tissues. Indeed, 9 of 10 (90%) tumors and 12 of 22 (54.5%) nontumorous cirrhotic livers possessed high- and/or low-frequency mutations in *LEPR* (Table 1). Notably, some somatic mutations were commonly detected in different positions of the same patient's liver. For example, C1084T (reference position: 65557165) mutations of *LEPR* were detected in the right, left, and caudate lobes of one patient (Supplementary Table 1, patient 11), suggesting that some of the hot spots of the acquired somatic mutations in the *LEPR* gene are commonly present in hepatocytes of the same liver underlying HCV infection. On the other hand, no mutations in *LEPR* were identified by deep sequencing analysis of noncirrhotic liver tissues from patients with chronic HCV infection or liver tissue from patients without HCV infection. (Supplementary Table 9). To confirm the somatic mutations present in *LEPR* in the nontumorous liver, we validated the candidate mutations by Sanger sequencing. For this purpose, we determined the sequences of exons 9 and 10 of *LEPR* of at least 50 randomly picked clones that were amplified from the nontumorous liver tissues of each patient. Although it was difficult to detect all the low-frequency mutations using the conventional cloning-sequencing method, we confirmed that somatic mutations were recurrently accumulated in *LEPR* of nontumorous cirrhotic liver tissues (Supplementary Figure 4).

LEPR Mutations Found in HCV-Positive Cirrhotic Liver Resulted in the Disruption of Downstream Signaling

Selected exome sequencing detected low-frequency mutations at a total of 650 nucleotide positions of *LEPR* in 12 of 22 (54.5%) HCV-positive cirrhotic liver tissues. Although the nucleotide changes were unevenly distributed throughout the whole *LEPR* exonic sequences, we detected 67 nucleotide alterations at the immunoglobulin (Ig) domain of *LEPR*, 38 of which (56.7%) were recurrently mutated in 2 or more patients (Figure 3*A*). Among them, nonsynonymous mutations that caused the amino acid changes were detected at 62 of the 67 (92.5%) nucleotide positions, and 10 of the 62 were also mutated in at least one HCC tissue examined in this study. Histological examination revealed no significant association between the presence of *LEPR* mutations and the level of fatty changes in the liver tissue (data not shown).

To explore the functional relevance of *LEPR* mutations detected in HCV-positive cirrhotic liver tissues, we randomly selected 7 *LEPR*s with a mutated Ig domain from 62 nonsynonymously mutated *LEPR*s and examined the downstream signaling properties of the mutated *LEPR* in vitro.

Accordingly, we subcloned the mutated *LEPR*s and constructed expression plasmids encoding those mutant *LEPR*s (Figure 3*B*). We first confirmed that only a small amount of endogenous LEPR expression was observed in both HEK293 and HepG2 cells (Supplementary Figure 5) and the induction of the phosphorylation of STAT3 by wild-type *LEPR* in the presence of recombinant human leptin (Figure 3*C*). In contrast, 4 of 7 (57.1%) mutations in the Ig domain of *LEPR* resulted in the reduction or loss of STAT3 phosphorylation in vitro (Figure 3*C*). To clarify the functional significance of *LEPR* mutations, the cell proliferation rate was determined in HepG2 cells expressing either wild-type or mutated *LEPR*s that were identified in HCV-positive cirrhotic liver tissues using the lentivirus system.[26] Up-regulation of cyclin D1 and/or E transcripts as well as enhanced cell proliferation were observed in the cells with expression of the mutated *LEPR* gene compared with wild-type cells, while there was no difference in the expression levels and subcellular localization between wild-type and mutated LEPR protein (Supplementary Figure 6). These findings indicate that some of the somatic mutations that latently accumulated in the Ig domain of *LEPR* of the cirrhotic liver tissue might cause dysfunction of LEPR-mediated signaling in the cells with those somatic mutations.

LEPR Dysfunction Enhanced Susceptibility to Tumorigenesis

To determine the functional relevance of LEPR dysfunction on development of liver cancer, we examined whether disruption of the LEPR gene contributes to liver tumorigenesis using a genetically altered mouse model, the Lepr-deficient C57BL/KsJ-*db/db* mouse (hereafter referred to as *db/db* mouse).[27] Thioacetamide (TAA), a putative carcinogen, is well established to induce liver fibrosis and tumorigenesis in a murine model.[28] Thus, we conducted an assay to evaluate whether LEPR insufficiency alters the effects of TAA-mediated tumorigenesis. Accordingly, TAA was prepared at a concentration of 0.02%, a relatively low dose compared with the carcinogenic dose,[29] and administered in drinking water to mice for 24 weeks. The body weight of the *db/db* mice was about twice that of their lean littermates, and *db/db* mice had hepatomegaly even after normalizing the liver weight to the body weight (Figure 4*A*). Histological

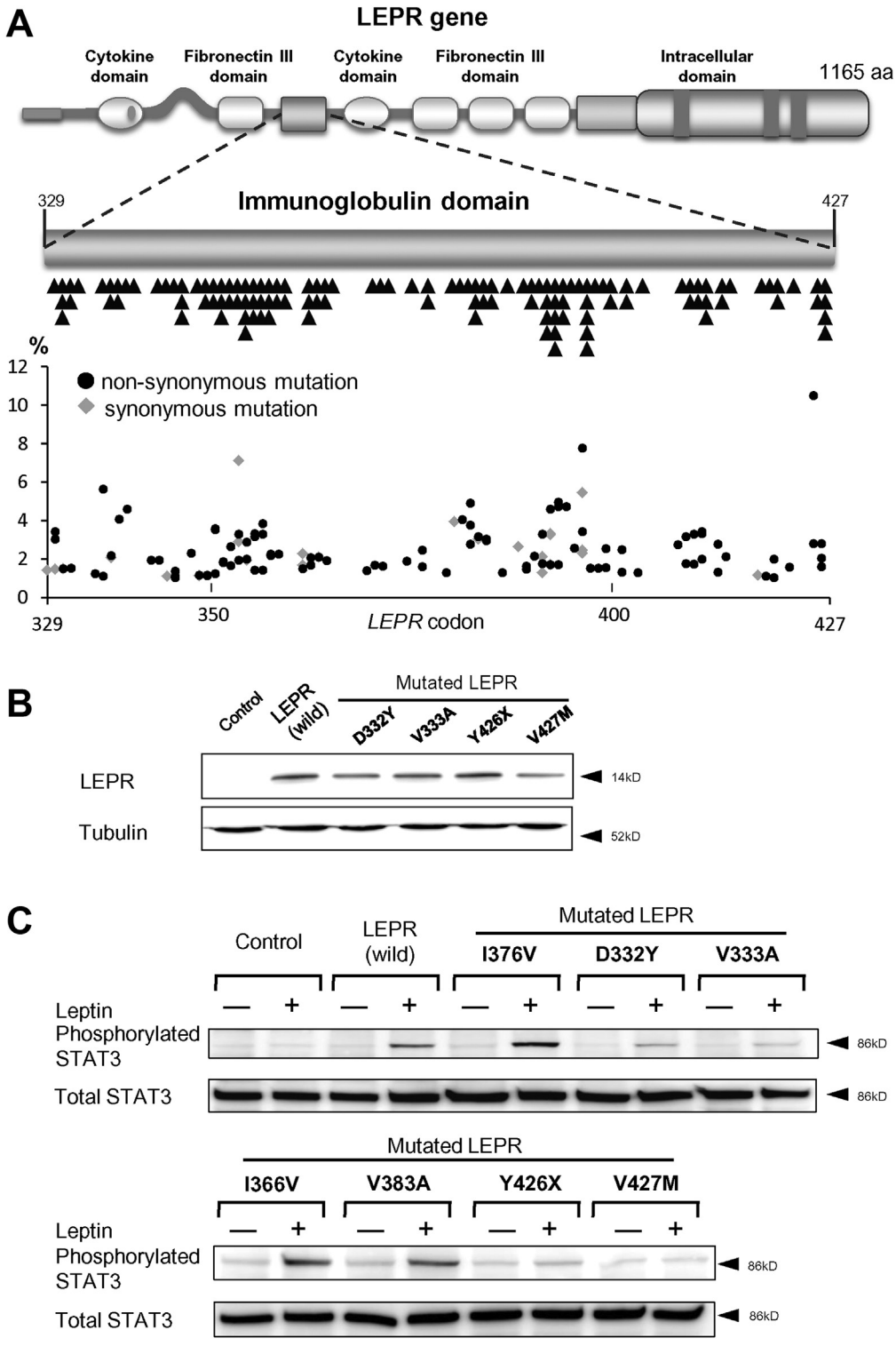

Figure 3. Distribution of mutations in the *LEPR* sequence in HCV-positive cirrhotic liver tissues. (*A*) Schematic diagram of the LEPR gene (*top panel*) and the Ig domain (*middle panel*). Mutated positions in the Ig domain are indicated by *black triangles*. A total of 38 of 67 (56.7%) mutated nucleotide positions of the Ig domain were recurrently mutated in 2 or more HCV-positive cirrhotic liver tissues. Frequencies of non-synonymous (*black circles*) and synonymous (*gray diamonds*) mutations at each nucleotide position of the Ig domain of each sample are shown (*lower panel*). Nonsynonymous mutations were detected at 62 of the 67 nucleotide positions. (*B* and *C*) HEK293 cells were transfected with constructs encoding wild-type or representative various mutated *LEPR*s that were identified in HCV-positive cirrhotic liver tissues. Control: empty vector. (*B*) Immunoblotting was performed on the lysate of the cells expressing either wild-type or a mutated Ig domain (D332Y, V333A, Y426X, and V427M) of the *LEPR* gene using anti-Myc antibodies. (*C*) After transfection, the cells were treated with or without recombinant leptin protein. Total protein was isolated and immunoblot analysis was performed using anti–phospho-STAT3 (*upper panel*) and anti–total STAT3 (*lower panel*).

examination revealed the accumulation of lipids within individual hepatocytes in the *db/db* mouse liver, a typical feature of steatosis (Figure 4*A*).

After administering TAA, blood levels of alanine aminotransferase were substantially elevated in *db/db* mice compared with control mice (Supplementary Table 10). Consistently, histological examination revealed that inflammatory activity was more severe in the liver of *db/db* mice than in the liver of control mice (Figure 4*B*). None of the control mice treated with TAA showed tumorigenesis 24 weeks after administration of TAA. In contrast, macroscopic liver nodules developed in 4 of 10 (40%) *db/db* mice

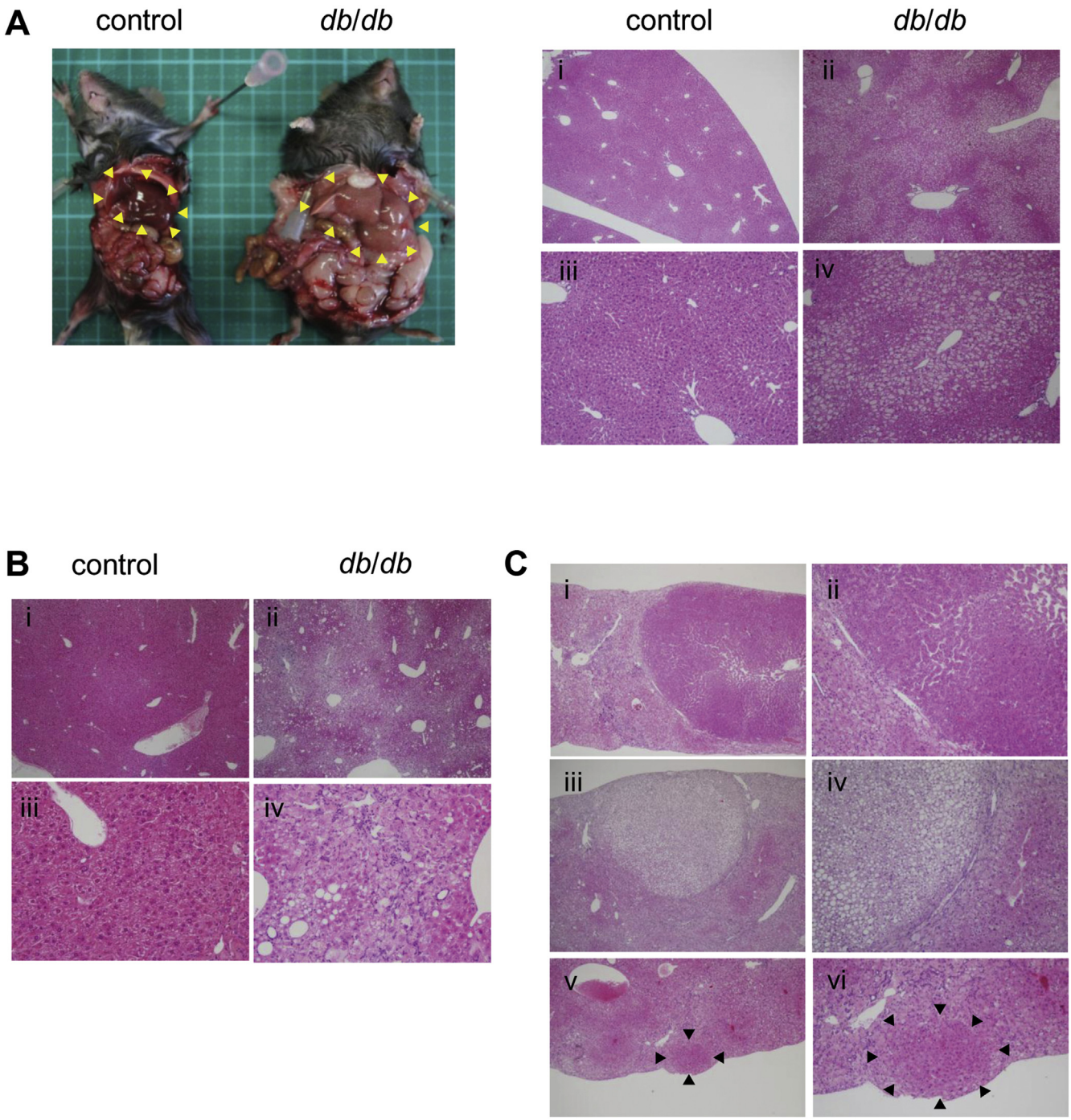

Figure 4. Tumors developed in *db/db* mice treated with TAA. (*A*) Representative macroscopic (*left panel*) and microscopic (*right panel*) images (H&E stain) of the liver from a *db/db* mouse and a littermate control mouse without administration of TAA. The liver of the *db/db* mouse is enlarged and yellowish compared with the control (*yellow arrowheads*). Histological analysis of the liver tissue of (*ii* and *iv*) *db/db* mouse and (*i* and *iii*) control mouse. Original magnification: 4× (*upper panels*) and 10× (*lower panels*). (*B*) Microscopic images (H&E stain) of (*i* and *iii*) control mice and (*ii* and *iv*) *db/db* mice treated with TAA for 24 weeks. In the *db/db* mice, inflammatory cell infiltration was extensively observed in the liver tissues underlying prominent steatosis (*ii* and *iv*). Original magnification: 4× (*upper panels*) and 20× (*lower panels*). (*C*) Microscopic images (H&E stain) of nodules that developed in *db/db* mice treated with TAA for 24 weeks (*i–vi*). Liver cancers developed in 2 *db/db* mice (*i–iv*). *Arrowheads* indicate hepatocyte hyperplasia (*v* and *vi*). Original magnification: 4× (*left panels*) and 10× (*right panels*).

that received the same dose of TAA during the same observation period (Table 2). Histological examination revealed that 2 *db/db* mice with liver nodules developed well-differentiated HCC (Figure 4*C*). In addition, the remaining nodules that developed in *db/db* mouse liver showed features of hepatocyte hyperplasia. These findings

Table 2. Incidence of Hepatic Nodules in C57BL/KsJ-*db/db* (*db/db*) and Misty (Control) Mice After 24 or 30 Weeks of Treatment With TAA

	db/db	Control
24 wk	n = 10	n = 10
Male/female	8/2	8/2
Tumor formation		
Total (%)	4[a] (40)	0 (0)
HCC	2	0
Hepatocyte hyperplasia	3	0
30 wk	n = 7	n = 10
Male/female	3/4	7/3
Tumor formation		
Total (%)	6[a] (86)	4 (40)
HCC	1	0
Hepatocyte hyperplasia	6	4

NOTE. The numbers of animals that developed hepatocyte hyperplasia and/or HCC are shown.
[a]One *db/db* mouse developed both HCC and hepatocyte hyperplasia.

suggest that Lepr-deficient *db/db* mice had high susceptibility to TAA-induced liver tumorigenesis.

Discussion

Tumor cells are considered to be generated by a stepwise accumulation of genetic alterations in tumor-related genes during the process of inflammation-associated carcinogenesis. Several studies have reported that epithelial tissues exposed to chronic inflammation accumulate genetic alterations in tumor-related genes before the onset of tumorigenesis.[7,8] Given that chronic inflammation induces somatic mutations, it is reasonable to assume that critical genetic alterations that contribute to tumorigenesis might emerge in chronically inflamed epithelial cells. Using whole exome sequencing, we showed here that considerable levels of somatic mutations accumulate not only in tumors but also in the nontumorous liver of patients with HCV-related cirrhosis.

Whole exome sequencing on synchronously developed HCCs showed a remarkable difference in the mutation signature in each case. In 2 cases, more than 20% of the mutated genes were commonly present in 2 tumors that developed in the same background liver, suggesting that these tumors were derived from a common origin or developed through intrahepatic metastasis. In contrast, the tumors that developed in the remaining case shared no common mutations, suggesting independent development in a multicentric manner. The data obtained from the latter case are consistent with those of a recent study in which no common somatic mutations were identified in the 2 pairs of multicentric HCCs that developed in the same background livers.[19] Taken together, these findings suggest that comprehensive whole exome sequencing on synchronously developed HCCs would permit distinction of the carcinogenic process between tumors that develop in a multicentric manner and that develop through intrahepatic metastasis.

Interestingly, we found that in some cases the total number of mutated genes of nontumorous liver tissues was larger than that of the matched tumor tissues, possibly due to the abundance of heterogeneous accumulation of passenger mutations in the nontumorous liver tissues.[30] The observation that the frequency of mutations at each nucleotide position in the nontumorous tissues tended to be lower than that in the matched tumor tissues may lend support to such a possibility. Notably, somatic mutations in the representative tumor-related genes, *TP53* and *CTNNB1*, were also latently accumulated in the cirrhotic liver tissues. It is unknown whether the *TP53* and/or *CTNNB1* mutations detected in nontumorous tissues were derived from the clinically undetectable small nest of cancer cells or premalignant hepatocytes; however, it is possible that these latent genetic alterations in tumor driver genes contribute to the development of HCC in the background of chronic liver disease.

Among the various mutated genes in the cirrhotic liver tissue, we identified *LEPR* as one of the most recurrently mutated genes. Indeed, we confirmed a total of 650 low-frequency mutations of the *LEPR* gene in 12 of 22 patients (54.5%) with HCV infection using selected exome sequencing. At present, it is not clear why a large number of mutations accumulate in the *LEPR* gene of nontumorous cirrhotic liver in patients with chronic HCV infection. One possibility may be that the *LEPR* gene is highly sensitive to AID-mediated mutagenesis in hepatocytes, because we recently observed that AID activation in cultured hepatoma-derived cells preferentially caused somatic mutations in the *LEPR* gene (Supplementary Table 11). On the other hand, close attention must be paid to the fact that only low-frequency mutations were detected in the *LEPR* gene in tumor tissues, consistent with the reported cancer genome database (International Cancer Genome Consortium data set version 12; http://dcc.icgc.org/web/). In general, tumor-specific driver mutations in tumor tissues are characterized by high-frequency mutations (eg, more than 20% nucleotide changes of total reads[18,23,31]). In this regard, the frequency of any mutation in the *LEPR* gene observed in the tumor tissues was more than 20% in our cases. Thus, the genetic changes in LEPR are unlikely to be direct driver mutations for HCC, but rather might play some role in the development of HCC in HCV-infected inflamed liver by providing a pathophysiological background for hepatocarcinogenesis by modifying the cell proliferation activity.

Leptin is a circulating hormone that is secreted by adipocytes and regulates energy homeostasis.[32] Leptin acts through binding to the extracellular domain of specific membrane receptor LEPR, which belongs to a family of class I cytokine receptors.[33] The extracellular domain of LEPR comprises 2 canonical cytokine receptor homology domains, Ig and fibronectin III domains, and the Ig domain is essential for the formation of the hexameric complex and for receptor activation.[34] In the present study, we confirmed that 67 mutations were present in the Ig domain

of *LEPR* in cirrhotic liver, and more than half of the mutations were recurrently mutated in 2 or more patients. Notably, more than 90% of those nucleotide alterations that accumulated in the Ig domain of *LEPR* were nonsynonymous mutations. Furthermore, we revealed that several nonsynonymous mutations that appeared in the Ig domain of *LEPR* impaired signaling to STAT3 in response to leptin, causing dysregulation of leptin signaling in the cells with those mutations. Sequencing the *LEPR* gene in patients with severe early-onset obesity revealed that the extracellular region of the LEPR has a variety of mutations in those patients.[35] A functional study of missense mutations in the *LEPR* found in severely obese patients also revealed that mutated LEPR has impaired signaling to STAT3, which is consistent with their inability to activate pathways involved in the reduction of food intake.[36] Together, these findings suggested that somatic mutations in the *LEPR* gene might provide the genetic basis for developing metabolic dysregulation in hepatocytes during hepatocarcinogenesis.

In the present study, we showed for the first time that *db/db* mice with disruption of the Lepr gene were more susceptible to developing hepatic inflammation as well as TAA-mediated tumorigenesis than wild-type mice. Consistent with our findings, a previous study reported an increased incidence of hepatocyte hyperplasia in leptin-deficient *ob/ob* mice, a model for nonalcoholic fatty liver disease.[37] Taken together, it is strongly suggested that dysregulation of LEPR signaling has a role in hepatic tumor development, but the mechanism of how the leptin signaling deficiency contributes to an enhanced inflammatory response and tumorigenesis is currently unknown. It should be noted that both *ob/ob* mice and *db/db* mice are characterized by hepatic steatosis, and steatosis is well recognized as a common histopathologic feature of the chronic HCV-infected liver. Epidemiological studies have revealed that fatty liver disease may be a common underlying pathology in patients with HCC,[38,39] and steatosis is an important cofactor in accelerating the development of hepatic fibrosis and inflammatory activity,[40,41] contributing to the progression of HCC in HCV-related chronic liver disease.[42] We found no correlation between the prevalence of *LEPR* mutations and the histological feature of fatty changes in HCV-positive cirrhotic liver tissues. On the other hand, previous studies have shown that leptin can oppose the action of insulin-induced signaling by reducing the phosphorylation of insulin receptor substrate 1 in human hepatic cells.[43,44] In addition, it was shown that leptin suppresses HCC via activation of the immune response, suggesting the tumor-suppressing function of leptin-mediated signaling.[45] Thus, we speculate that dysregulation of leptin signaling in the liver might be involved in the neoplastic process of patients with HCV-related chronic liver damage. Because somatic mutations in *LEPR* are limited to a small proportion of cells in cirrhotic liver tissue and the TAA-mediated liver inflammation model does not fully recapitulate HCV-associated chronic liver disease, further analysis is required to determine whether dysregulation of LEPR-mediated signaling caused by *LEPR* mutations contributes to the enhanced inflammatory response or tumorigenesis in patients with HCV-related chronic liver damage.

In conclusion, we showed that various somatic mutations latently accumulate in the nontumorous cirrhotic liver of patients with HCV infection. The findings that the *LEPR* gene was recurrently mutated in cirrhotic liver provide a novel putative link between the inflammation-mediated genetic aberrations, the dysregulation of leptin signaling, and the development of HCC in patients with HCV-related chronic liver disease. The gene catalogue identified in the HCV-infected chronically damaged liver might contain the putative driver gene associated with tumor initiation as well as the gene that provides the genetic basis for the development of HCC. Thus, further studies are required to identify the genetic alterations that contribute to tumor development in chronically inflamed liver underlying chronic HCV infection.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://dx.doi.org/10.1053/j.gastro.2013.09.025.

References

1. Coussens LM, Werb Z. Inflammation and cancer. Nature 2002;420:860–867.
2. Chiba T, Marusawa H, Ushijima T. Inflammation-associated cancer development in digestive organs: mechanisms and roles for genetic and epigenetic modulation. Gastroenterology 2012;143:550–563.
3. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature 1998;396:643–649.
4. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100:57–70.
5. Hussain SP, Schwank J, Staib F, et al. TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer. Oncogene 2007; 26:2166–2176.
6. Loeb LA, Bielas JH, Beckman RA. Cancers exhibit a mutator phenotype: clinical implications. Cancer Res 2008;68:3551–3557.
7. Barrett MT, Sanchez CA, Prevo LJ, et al. Evolution of neoplastic cell lineages in Barrett oesophagus. Nat Genet 1999;22:106–109.
8. Leedham SJ, Graham TA, Oukrif D, et al. Clonality, founder mutations, and field cancerization in human ulcerative colitis-associated neoplasia. Gastroenterology 2009;136:542–550.
9. **Hussain SP, Hofseth LJ**, Harris CC. Radical causes of cancer. Nat Rev Cancer 2003;3:276–285.
10. Matsumoto Y, Marusawa H, Kinoshita K, et al. Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat Med 2007;13:470–476.
11. Komori J, Marusawa H, Machimoto T, et al. Activation-induced cytidine deaminase links bile duct inflammation

to human cholangiocarcinoma. Hepatology 2008; 47:888–896.
12. Endo Y, Marusawa H, Kou T, et al. Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers. Gastroenterology 2008;135:889–898.
13. Ikeda K, Marusawa H, Osaki Y, et al. Antibody to hepatitis B core antigen and risk for hepatitis C-related hepatocellular carcinoma: a prospective study. Ann Intern Med 2007;146:649–656.
14. Endo Y, Marusawa H, Kinoshita K, et al. Expression of activation-induced cytidine deaminase in human hepatocytes via NF-kappaB signaling. Oncogene 2007; 26:5587–5595.
15. Kou T, Marusawa H, Kinoshita K, et al. Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis. Int J Cancer 2007; 120:469–476.
16. Wei X, Walia V, Lin JC, et al. Exome sequencing identifies GRIN2A as frequently mutated in melanoma. Nat Genet 2011;43:442–446.
17. Wang L, Tsutsumi S, Kawaguchi T, et al. Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency. Genome Res 2012;22:208–219.
18. **Wang K, Kan J**, Yuen ST, et al. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. Nat Genet 2011;43: 1219–1223.
19. **Fujimoto A, Totoki Y**, Abe T, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nat Genet 2012;44:760–764.
20. **Guichard C, Amaddeo G, Imbeaud S**, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet 2012;44:694–698.
21. Nasu A, Marusawa H, Ueda Y, et al. Genetic heterogeneity of hepatitis C virus in association with antiviral therapy determined by ultra-deep sequencing. PLoS One 2011;6:e24907.
22. Nishijima N, Marusawa H, Ueda Y, et al. Dynamics of hepatitis B virus quasispecies in association with nucleos(t)ide analogue treatment determined by ultra-deep sequencing. PLoS One 2012;7:e35052.
23. Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 2011; 469:539–542.
24. **Clément K, Vaisse C**, Lahlou N, et al. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. Nature 1998;392:398–401.
25. Laurent-Puig P, Zucman-Rossi J. Genetics of hepatocellular tumors. Oncogene 2006;25:3778–3786.
26. **Morita S, Matsumoto Y**, Okuyama S, et al. Bile acid-induced expression of activation-induced cytidine deaminase during the development of Barrett's oesophageal adenocarcinoma. Carcinogenesis 2011;32:1706–1712.
27. Lee GH, Proenca R, Montez JM, et al. Abnormal splicing of the leptin receptor in diabetic mice. Nature 1996; 379:632–635.
28. Becker FF. Thioacetamide hepatocarcinogenesis. J Natl Cancer Inst 1983;71:553–558.
29. Schnur J, Nagy P, Sebestyén A, et al. Chemical hepatocarcinogenesis in transgenic mice overexpressing mature TGF beta-1 in liver. Eur J Cancer 1999; 35:1842–1845.
30. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. Nature 2007;446:153–158.
31. **Yan XJ, Xu J, Gu ZH**, et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet 2011; 43:309–315.
32. Schwartz MW, Woods SC, Porte D, et al. Central nervous system control of food intake. Nature 2000;404:661–671.
33. Tartaglia LA. The leptin receptor. J Biol Chem 1997; 272:6093–6096.
34. **Peelman F, Iserentant H**, De Smet AS, et al. Mapping of binding site III in the leptin receptor and modeling of a hexameric leptin.leptin receptor complex. J Biol Chem 2006;281:15496–15504.
35. Farooqi IS, Wangensteen T, Collins S, et al. Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. N Engl J Med 2007;356:237–247.
36. **Kimber W, Peelman F**, Prieur X, et al. Functional characterization of naturally occurring pathogenic mutations in the human leptin receptor. Endocrinology 2008; 149:6043–6052.
37. Yang S, Lin HZ, Hwang J, et al. Hepatic hyperplasia in noncirrhotic fatty livers: is obesity-related hepatic steatosis a premalignant condition? Cancer Res 2001; 61:5016–5023.
38. Marrero JA, Fontana RJ, Su GL, et al. NAFLD may be a common underlying liver disease in patients with hepatocellular carcinoma in the United States. Hepatology 2002;36:1349–1354.
39. Kodama Y, Brenner DA. c-Jun N-terminal kinase signaling in the pathogenesis of nonalcoholic fatty liver disease: multiple roles in multiple steps. Hepatology 2009;49:6–8.
40. Hourigan LF, Macdonald GA, Purdie D, et al. Fibrosis in chronic hepatitis C correlates significantly with body mass index and steatosis. Hepatology 1999;29:1215–1219.
41. Lonardo A, Adinolfi LE, Loria P, et al. Steatosis and hepatitis C virus: mechanisms and significance for hepatic and extrahepatic disease. Gastroenterology 2004; 126:586–597.
42. Ohata K, Hamasaki K, Toriyama K, et al. Hepatic steatosis is a risk factor for hepatocellular carcinoma in patients with chronic hepatitis C virus infection. Cancer 2003;97:3036–3043.
43. Cohen B, Novick D, Rubinstein M. Modulation of insulin activities by leptin. Science 1996;274:1185–1188.
44. Wang Y, Kuropatwinski KK, White DW, et al. Leptin receptor action in hepatic cells. J Biol Chem 1997; 272:16216–16223.

45. Elinav E, Abd-Elnabi A, Pappo O, et al. Suppression of hepatocellular carcinoma growth in mice via leptin, is associated with inhibition of tumor cell growth and natural killer cell activation. J Hepatol 2006;44:529–536.

Author names in bold designate shared co-first authorship.

Received February 11, 2013. Accepted September 10, 2013.

Reprint requests
Address requests for reprints 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. e-mail: maru@kuhp.kyoto-u.ac.jp; fax: (81) 75-751-4303.

Acknowledgments
Data profiling: Sequence reads with Genome Analyzer were deposited in the DNA Data Bank of Japan Sequence Read Archive (http://trace.ddbj.nig.ac.jp/dra/index_e.shtml) under accession no. DRA000867.

The authors thank Dr K. Terasawa, Dr T. Fujiwara, Dr K. Takahashi, and Dr Y. Ueda for helpful suggestions; Dr N. Nishijima, Dr A. Takai, Dr A. Nasu, Dr Y. Endo, Y. Nakagawa, and C. Kakimoto for help with the analyses; and Dr H. Kokuryu and Dr T. Kusaka at Kyoto Katsura Hospital for their support.

Conflicts of interest
The authors disclose no conflicts.

Funding
Supported by Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research, KAKENHI on Innovative Areas, and Health and Labour Sciences Research Grants for Research on Intractable Disease and Research on Hepatitis from the Ministry of Health, Labour and Welfare, Japan.

Supplementary Methods

Patients

The study group comprised patients who underwent living donor liver transplantation or potentially curative resection of primary HCC at Kyoto University Hospital from 2000 to 2010. The selection of patients enrolled in this study was based on the availability of a sufficient amount of tissue for analysis. Patients included 17 men and 9 women, with a mean age at the time of surgery of 54.9 ± 7.7 years (mean ± SD; range, 37–76 years). Among them, whole exome sequencing was applied to 7 tumors, 4 nontumorous cirrhotic livers, and matched peripheral lymphocytes from 4 patients (Supplementary Table 1, patients 1–4). Furthermore, we performed selected exome sequencing of 22 nontumorous cirrhotic livers, 10 tumors, and matched peripheral lymphocytes from 22 other affected patients (Supplementary Table 1, patients 5–26). All patients were positive for serum anti-HCV and/or HCV RNA. Written informed consent for the use of resected tissue was obtained from all patients in accordance with the Declaration of Helsinki, and the Kyoto University Graduate School and Faculty of Medicine Ethics Committee approved the study.

Sequence Data Analysis and Variant Filtering

Using NextGENe v2.2 software (SoftGenetics, State College, PA), the obtained reads were aligned with the reference sequences of the Human Genome Build 37.3. Reads with 96% or more bases matching a particular position of the reference sequences were aligned. Furthermore, reads with a median quality value score of more than 20 and no more than 3 uncalled nucleotides were allowed anywhere in one read. Only sequences that passed the quality filters were analyzed, and each position of the genome was assigned a coverage depth representing the number of times the nucleotide position was sequenced. To identify somatic mutations, we used a number of scores to provide an empirical estimation of the likelihood that a given mutation was real and not an artifact of sequencing or alignment errors.

In the whole exome sequencing analysis, candidates of somatic mutations were selected according to the variant filtering process (Supplementary Figure 1). We defined nucleotide alterations that appeared in more than 20% of reads as somatic mutations.[1–3] When detecting the genes commonly mutated in both tumor and nontumorous liver tissues of the same subjects, we also selected potential nucleotide alterations that appeared between 5% and 20% of the total reads in nontumorous liver tissues for further evaluation. We excluded potential somatic mutations that represented more than 5% of the reads in peripheral lymphocytes of the same patient as common variants in each individual. Candidate nucleotide alterations were tested using standard Sanger sequencing on an Applied Biosystems 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA) to validate the presence of each mutation.

In selected exome sequencing analysis, candidates of somatic mutations were selected according to the variant filtering process (Supplementary Figure 2). We defined somatic mutations with more than 20% of reads as high-frequency mutations and those that appeared between 1% and 20% of total reads as low-frequency mutations. We excluded potential somatic mutations that represented more than 1% of the reads in peripheral lymphocytes of the same patient. In cases in which we could not obtain lymphocyte DNA, candidates of somatic mutations found in the lymphocytes of 2 or more different patients were excluded in consideration of possible Japanese polymorphisms.

We compared our variants against common and germline polymorphisms present in the dbSNP135 to discard known germline single nucleotide polymorphisms.

All sequence reads were deposited in the DNA Data Bank of Japan Sequence Read Archive (accession no. DRA000867).

Score

SoftGenetics developed the overall mutation score to provide an empirical estimation of the likelihood that a given mutation is real and not an artifact of sequencing or alignment errors. The overall mutation score of NextGENe can be used like Phred scores, in which the scores are logarithmically linked to error probabilities. The overall mutation score of NextGENe is obtained as the product of the "coverage score," which is calculated from the depth of coverage at the position of the mutation and with a value ranging from 0 (where depth of coverage is 1) to an unlimited number, multiplied by each of the 4 types of additional penalty scores, such as the read balance score, allele balance score, mismatch score, and wrong allele score, with values less than 1 but positive (the calculating formula for each score is not shown). These scores are described in the NextGENe User Manual in detail (http://www.softgenetics.com/NextGENe.html).

Overall mutation score. SoftGenetics developed the overall mutation score to provide an empirical estimation of the likelihood that a given mutation is real and not an artifact of sequencing or alignment errors. A low overall mutation score, however, does not mean that the mutation is more than likely a false mutation. The low score implies only that the mutation cannot be called a true mutation with absolute certainty. As a general guideline, if the coverage is high (500 to several thousand reads) and the data are bidirectional, then scores that are $\leq$5 indicate that the mutation is most likely false, whereas scores of $\geq$25 indicate that the mutation is most likely true.

Mismatch score. Several variations from the reference sequence that occur very close together often indicate a region where mutation calls are less reliable. The mismatch score penalizes a specific mutation if other mismatched bases are found nearby. The software first looks for mismatches that occur in a minimum percentage of reads in the 10–base pair region that is found on either side of the mutation that is being scored.

Wrong allele score. Mismatches that are different from the consensus are referred to as wrong mismatches. These wrong mismatches most likely result from sequencing errors. For example, A, C, G, T, and insertions represent wrong mismatches when a deletion was called at a position.

Cell Culture and Transfection

The complementary DNA encoding the wild-type and the mutated LEPR were generated by reverse-transcription polymerase chain reaction from the messenger RNA of the liver tissues, followed by polymerase chain reaction amplification using Phusion high-fidelity DNA polymerase (Finnzymes, Espoo, Finland) and the following oligonucleotide primers: 5′-CGCGGATCCATGATTTGTCAAAAATTC-3′ (sense) and 5′- AAGGAAAAAAGCGGCCGCTTACACAGTTAGGTCA CACA-3′ (antisense). The resulting polymerase chain reaction fragments were inserted into the *Bam*HI-*Not*I sites of pcDNA3 for HEK293 and the *Bam*HI-*Apa*I sites of lentivirus for HepG2, as described previously.[4]

HEK293 and HepG2 cells were maintained in Dulbecco's modified Eagle medium (Gibco BRL, Rockville, MD) containing 10% fetal bovine serum. For transfection of plasmids into HEK293 cells, we used Lipofectamine2000 transfection reagent (Invitrogen, Carlsbad, CA). At 40 hours after transfection, the cells were serum starved for 8 hours and then either left unstimulated or stimulated with 100 ng/mL recombinant human leptin (Sigma-Aldrich, St Louis, MO) for 10 minutes. Expression of either wild-type or mutant LEPR in HepG2 cells was performed using a lentiviral vector–mediated wild-type and mutated LEPR expression system as described previously.[5] In brief, LEPR complementary DNA fragments were inserted into the viral vectors, followed by the production of lentiviral stocks in HEK293 cells. HepG2 cells were cultured in virus-containing medium for 48 hours, starved for 8 hours, treated with 100 ng/mL recombinant human leptin (Sigma-Aldrich) for 10 minutes, and then subjected to immunoblotting, immunostaining, quantitative reverse-transcription polymerase chain reaction, or a cell proliferation (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide [MTT]) assay.

Immunoblotting Analysis

Immunoblotting was performed using anti-STAT3 and anti–phospho-STAT3 antibody (Cell Signaling Technology, Danvers, MA) according to the manufacturer's protocol.

Animal Experiments

C57BL/KsJ-*db/db* mice (*db/db* mice), which possess homozygous deletion of the *Lepr*, *Ob-R* gene, and misty mice, which are wild-type with a normal *Lepr*, were purchased from Japan SLC (Shizuoka, Japan). TAA (Sigma-Aldrich) was prepared at a concentration of 0.02% and administered in drinking water to mice for 24 weeks or 30 weeks beginning at 5 weeks of age. These mice were then killed for analysis of the development of liver tumors. All animal experiments were approved by the Ethics Committee for Animal Experiments and performed under the Guidelines for Animal Experiments of Kyoto University.

Supplementary References

1. **Wang K, Kan J**, Yuen ST, et al. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. Nat Genet 2011;43:1219–1223.
2. Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 2011;469:539–542.
3. **Yan XJ, Xu J, Gu ZH**, et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet 2011; 43:309–315.
4. Endo Y, Marusawa H, Kinoshita K, et al. Expression of activation-induced cytidine deaminase in human hepatocytes via NF-kappaB signaling. Oncogene 2007; 26:5587–5595.
5. **Morita S, Matsumoto Y**, Okuyama S, et al. Bile acid-induced expression of activation-induced cytidine deaminase during the development of Barrett's oesophageal adenocarcinoma. Carcinogenesis 2011;32:1706–1712.
6. Aly HH, Watashi K, Hijikata M, et al. Serum-derived hepatitis C virus infectivity in interferon regulatory factor-7-suppressed human primary hepatocytes. J Hepatol 2007;46:26–36.

Supplementary Table 1. Clinical Features of 4 Patients Who Underwent Whole Exome Sequencing and 22 Patients Who Underwent Selected Exome Sequencing

Patient no.	Age (*y*)	Sex	Body mass index (kg/m^2)	α-Fetoprotein (*ng/mL*)	Des-γ-carboxy prothrombin (*mAU/mL*)	HCC	Histological grade
Whole exome sequencing							
1	51	Male	23.3	16	185	M	Well
2	58	Female	22.3	103	7	M	Mod
3	55	Female	26.7	779	881	M	Mod
4	53	Male	22.3	34	85	S	Mod
Selected exome sequencing							
5	65	Male	25.2	17	7	M	Mod
6	49	Female	21.6	149	107	M	Mod
7	40	Male	25.7	24	50	M	Mod
8	50	Male	25.0	16	23	M	Mod
9	57	Female	23.4	8	30	M	Mod
10	56	Female	22.8	5	929	M	Mod
11	53	Male	18.6	30	31	M	Mod
12	65	Female	29.7	6	1877	S	Mod
13	57	Male	19.0	19	167	S	Well
14	76	Male	21.8	75,363	37,784	M	Poor
15	64	Male	18.7	177	8	—	—
16	57	Male	25.5	45	68	—	—
17	54	Female	25.9	<3	10	—	—
18	50	Male	22.3	585	61	—	—
19	60	Female	21.3	434	72	—	—
20	57	Male	25.0	15	8310	—	—
21	56	Male	19.0	15	383	—	—
22	49	Female	21.8	38	227	—	—
23	59	Male	25.6	6	12	—	—
24	49	Male	23.2	4	320	—	—
25	37	Male	22.2	4	13	—	—
26	51	Male	20.5	3	90	—	—

M, multiple; Well, well-differentiated HCC; Mod, moderately differentiated HCC; S, solitary; Poor, poorly differentiated HCC.

Supplementary Table 2. Overview of Whole Exome Sequencing Data From 4 Patients With HCC Who Had HCV Infection

	Tumor (n = 7)	Nontumor (n = 4)	Lymphocytes (n = 4)
Total reads	44,323,036	41,920,372	38,661,394
Aligned reads	40,046,800	33,742,449	31,595,571
Aligned sequence (*base pairs*)	2,824,088,514	2,384,058,470	2,221,753,713
Median read depth	40.2	31.9	27.4
Coverage			
1×	31,560,125	32,343,635	30,935,484
8×	24,724,702	23,432,758	23,549,909
20×	17,707,636	15,000,474	16,272,508
30×	13,599,418	11,752,775	12,527,511

NOTE. Whole exome sequencing was performed on tumor tissues, nontumorous cirrhotic liver tissues, and matched peripheral lymphocytes from each patient. Total reads, aligned reads, aligned sequences (base pairs), median read depth, and number of target regions, which were 1×, 8×, 20×, and 30× or more coverage depth read, are shown.

Supplementary Table 3. List of 970 Nucleotide Positions in 768 Genes That Were Mutated at a Frequency of More Than 20% of Reads in 7 HCC Tumors From 4 Patients

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
AGRN	875083	1	26	20	A < C	NS	D	2
LOC728661	1487244	1	8	18	G < T	NS	N	3
CDC2L2	1540787	1	3	43	T < C	NS	N	4
PANK4	2331358	1	18	32	T < C	NS	N	4
KIAA0562	3645675	1	7	67	T < C	S	N	4
CHD5	5928578	1	24	54	C < T	S	N	2
PTCHD2	11319504	1	7	23	G < C	NS	N	4
PLOD1	11750469	1	4	22	G < T	NS	N	4
PRAMEF1	12595752	1	3	170	G < A	S	N	4
PRAMEF1	12596087	1	3	93	C < T	NS	D	4
PRAMEF11	12625168	1	5	48	G < A	S	N	4
PRAMEF11	12628397	1	3	38	C < T	S	D	4
PRAMEF11	12628415	1	3	36	C < T	S	N	4
HNRNPCL1	12647885	1	1	143	T < C	S	N	4
PRAMEF7	12717626	1	1	27	A < G	S	N	4
PRAMEF9[a]	13064237	1	1	26	G < A	NS	N	2
PRAMEF9[a]	13064255	1	1	35	G < A	NS	D	2
PRAMEF18	13117381	1	1	27	G < A	NS	N	4
ARHGEF10L	17547108	1	1	109	T < G	S	N	4
PLA2G2D	20082054	1	3	56	T < C	NS	N	4
HSPG2	21856574	1	5	77	C < A	NS	N	4
CELA3A	21973988	1	6	105	T < G	NS	N	4
CELA3A	21976308	1	7	49	G < A	S	N	4
LOC100289113	22086886	1	1	28	A < C	NS	D	1
LUZP1	23059855	1	1	48	T < C	S	N	4
TRIM63	26025003	1	5	90	T < C	NS	N	4
SLC9A1	27120757	1	1	56	A < G	S	N	4
PHC2	33310033	1	8	132	C < T	S	N	4
CSMD2	33528214	1	51	83	T < C	NS	N	4
SLC2A1	42884612	1	8	74	T < C	S	N	4
TIE1	43269564	1	14	55	T < C	S	N	4
MAST2	45983460	1	17	45	T < G	NS	N	4
LRP8	53222315	1	9	143	G < T	S	N	4
ANGPTL3	62554389	1	2	24	A < T	NS	D	3
LEPR	65548341	1	4	31	C < A	S	N	3
RPE65	68386987	1	12	33	A < C	NS	N	1
ZNF644	90894104	1	2	18	G < A	NS	N	3
RBM15	110372981	1	1	17	A < C	S	N	1
RBM15	110373546	1	1	39	T < C	NS	N	3
CHI3L2	111273982	1	9	79	C < T	NS	N	4
CSDE1	114765324	1	8	29	C < A	NS	N	3
CSDE1	114765325	1	8	29	C < A	NS	N	3
IGSF3	116648924	1	2	69	G < A	S	NO	2
NBPF20	122618548	1	15	62	G < A	S	N	4
NBPF20	122618618	1	15	140	C < T	NS	N	4
NBPF20	122618624	1	15	174	A < T	NS	N	4
PDE4DIP	122663887	1	31	88	C < T	S	N	4
PDE4DIP	122667176	1	28	71	C < T	NS	N	4
NBPF10	123083515	1	1	83	A < G	NS	N	4
NBPF10	123092695	1	8	17	C < T	NS	N	3
NBPF10	123094578	1	10	100	A < C	NS	N	3
NBPF10	123094595	1	10	217	A < G	NS	N	4
NBPF10	123158473	1	86	408	G < C	NS	N	4
ANKRD35	123351469	1	10	14	A < T	NS	N	3
GPR89C	123673973	1	1	23	T < G	NS	D	4
BCL9	124884100	1	6	18	G < A	S	N	3
NBPF14	125797806	1	18	56	C < T	NS	N	1
NBPF14	125799375	1	16	76	T < C	S	N	2
NBPF14	125799402	1	16	71	G < A	S	N	2

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
NBPF15	126071852	1	4	159	A < G	S	N	4
HRNR	129676605	1	2	214	G < A	S	N	4
HRNR	129676984	1	2	52	G < C	NS	NO	3
HRNR	129677003	1	2	88	C < T	NS	N	4
FLG[a]	129766583	1	2	89	C < G	NS	N	2
FLG	129767962	1	2	102	C < T	NS	N	1
FLG	129768306	1	2	39	T < C	NS	N	4
FLG	129768312	1	2	59	C < G	NS	N	4
FLG	129771039	1	2	449	G < A	NS	N	2
FLG	129771228	1	2	298	C < G	NS	N	4
FLG	129773236	1	2	135	T < C	NS	N	2
FLG	129773862	1	2	222	G < C	NS	N	4
FLG	129774814	1	2	232	T < C	NS	N	4
PGLYRP3	130769598	1	2	118	C < T	S	N	4
CLK2	132724561	1	8	17	C < T	NS	D	3
CLK2	132724562	1	8	17	G < T	S	N	3
MSTO1	133072971	1	11	33	T < G	S	N	3
GON4L	133214185	1	27	50	C < A	NS	D	2
IQGAP3	134016387	1	12	66	C < G	NS	N	4
PEA15	137673244	1	3	25	A < T	NS	N	3
HSPA6	138985040	1	1	34	C < T	NS	D	4
NUF2	140800188	1	8	38	C < A	NS	NO	3
FAM78B	143529898	1	2	204	C < G	S	N	4
F5	147009112	1	10	162	C < T	NS	N	4
FAM5C[a]	167558142	1	7	33	G < A	NS	N	1
ZBTB41	174618823	1	10	13	A < C	NS	NO	2
KIF21B[a]	178450152	1	18	56	T < C	S	N	1
TMEM9	178602981	1	4	127	A < G	S	N	4
ELF3	179471218	1	2	54	C < G	S	N	4
PPP1R12B	180023641	1	21	22	C < A	NS	N	3
KDM5B	180267325	1	1	13	G < C	NS	D	1
CHI3L1	180642801	1	5	141	T < C	NS	D	4
FAM71A	189989294	1	1	133	T < C	NS	N	3
MIA3	200015587	1	13	40	T < C	S	N	4
JMJD4	205110357	1	6	53	C < T	S	N	4
OBSCN[a]	205602418	1	8	22	T < C	NS	D	1
RHOU	206063445	1	2	52	C < G	S	N	4
GNPAT	208576822	1	2	47	C < T	NS	D	3
LYST	213162183	1	3	14	G < T	NS	N	2
ADSS	221776216	1	7	22	A < C	NS	D	3
ADSS	221776218	1	7	22	C < T	S	N	3
KIF26B	223037622	1	11	110	C < T	S	N	4
LOC391343	227830117	2	1	41	T < G	NS	NR	4
LOC391343	227830313	2	1	16	G < C	S	NR	3
PXDN	228577682	2	17	144	G < C	NS	N	4
ODC1	237355517	2	10	27	G < A	S	N	4
APOB	247947498	2	16	24	A < C	NS	N	3
APOB	247947499	2	16	24	A < T	NS	N	3
ALK	256154817	2	15	59	A < G	S	N	4
FSHR	275890251	2	10	21	G < T	NS	NO	3
C2orf63	282104344	2	10	39	G < A	NS	N	4
CYP26B1	299058969	2	6	23	G < A	S	N	1
CCDC142	301407728	2	4	37	G < A	NS	NO	2
ST3GAL5	312787861	2	3	17	T < C	NS	N	4
KIAA1310	319723935	2	13	14	G < T	NS	N	3
ACTR1B	320724551	2	6	276	C < T	S	N	2
CHST10	323459632	2	5	57	G < C	S	N	4
MAP4K4	324943015	2	26	37	G < C	NS	D	3
SLC9A4	325569668	2	3	69	A < T	NS	D	3
TGFBRAP1	328335511	2	10	51	T < C	NS	N	4

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
RGPD3	329534289	2	1	37	A < G	S	N	4
LIMS1	331725649	2	1	96	G < A	NS	D	4
10-Sep	332632987	2	6	104	G < A	S	N	4
LOC645529	336733665	2	3	118	T < C	NS	NR	4
POTEF	353150970	2	13	98	T < A	NS	N	4
POTEF	353185302	2	1	71	T < C	NS	N	4
TUBA3E	353260275	2	3	94	C < T	NS	N	2
ACTBL3	354657198	2	1	108	T < G	S	NR	4
THSD7B	360341133	2	11	14	G < C	S	N	2
GALNT5	380364795	2	7	19	T < G	NS	D	3
SCN9A	389348565	2	11	38	T < C	NS	N	3
SCN9A	389352545	2	9	39	T < C	S	N	4
ABCB11	391996481	2	23	27	G < A	NS	NO	3
OLA1	397151333	2	9	14	A < T	S	N	3
TTN	401781960	2	95	24	T < G	NS	D	3
TTN	401781961	2	95	24	C < T	S	N	3
SESTD1	402189020	2	14	28	G < T	NS	D	3
DUSP19	406151282	2	1	41	C < G	NS	D	3
ZNF804A	408011183	2	4	20	A < G	S	N	3
LOC200726	429716908	2	1	28	C < T	NS	NR	3
ERBB4	434456096	2	28	29	C < T	NS	N	3
RNF25	441737431	2	8	18	G < A	S	N	3
C2orf24	442244994	2	8	51	G < A	NS	N	4
C2orf24	442245306	2	8	31	A < G	NS	N	4
TUBA4A	442323330	2	4	103	C < T	NS	D	4
OBSL1	442639564	2	4	59	G < A	S	N	4
SERPINE2	447057057	2	5	24	G < A	S	N	3
DOCK10	447917888	2	20	54	G < A	NS	NO	1
DIS3L2	455310852	2	10	90	G < A	NS	N	1
ALPP	455451136	2	1	40	C < T	NS	D	4
LRRFIP1	460828821	2	11	16	A < G	S	N	1
HDAC4	462131420	2	20	77	G < A	S	N	4
ITPR1	469948734	3	21	68	A < C	S	N	4
WNT7A	479128207	3	3	47	C < T	S	N	4
ZFYVE20	480358270	3	5	13	C < T	NS	N	3
OXNAD1	481544487	3	1	101	C < T	S	N	4
RARB	490854079	3	5	14	T < C	S	N	3
EOMES	492992244	3	4	84	C < T	NS	N	1
SCN10A	504030094	3	9	56	C < T	S	N	4
SCN11A[a]	504168069	3	15	110	T < A	NS	NO	1
CX3CR1	504539250	3	1	36	T < C	NS	N	3
CTNNB1	506498027	3	2	39	G < A	NS	D	1
CCR5	511646366	3	1	20	C < A	NS	NO	3
COL7A1	513857189	3	21	50	T < C	S	N	4
RBM6	515335643	3	16	16	G < A	NS	D	3
RBM5	515386440	3	22	46	G < A	NS	D	3
GLYCTK	517558462	3	4	64	C < A	S	N	2
KBTBD8	532186608	3	2	48	G < T	NS	NO	3
FOXP1	536153715	3	13	189	A < C	NS	D	3
FOXP1	536153718	3	13	190	C < T	S	N	3
LOC100288801	540845478	3	1	220	G < A	NS	N	4
LOC100288801[a]	540846776	3	2	228	G < A	S	N	2
EPHA3	554308286	3	2	14	T < C	S	N	3
DCBLD2	560650518	3	16	37	G < A	NS	N	3
DCBLD2	560650520	3	16	35	A < C	NS	NO	3
BOC	575123882	3	6	56	C < T	S	N	4
GPR156	582017947	3	9	64	A < T	S	N	2
HEG1	586864008	3	6	41	T < C	NS	N	3
MCM2[a]	589457039	3	5	53	A < G	NS	N	1
RUVBL1	589951418	3	6	33	T < A	NS	D	3

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
RUVBL1	589951420	3	6	30	T < G	NS	D	3
C3orf25	591272422	3	2	44	T < C	NS	D	4
PLXND1	591434995	3	7	116	G < A	S	N	4
COL6A6	592486478	3	27	23	G < T	NS	D	3
SLCO2A1	595793430	3	11	60	A < G	NS	D	1
RYK	596026424	3	13	102	C < T	NS	N	1
ZBTB38[a]	603296515	3	1	39	A < G	NS	N	1
PLOD2	607970905	3	3	11	T < C	S	N	2
PLSCR2	608303872	3	4	55	A < T	NS	D	3
TMEM183B	611832538	3	1	44	A < C	NS	NR	3
TSC22D2	612260917	3	1	14	T < G	S	N	4
MYNN	631624024	3	1	54	C < T	S	N	4
TNIK[a]	633027074	3	8	63	C < G	NS	D	1
IL1RAP	652454036	3	2	73	C < A	S	N	2
MUC4	657634114	3	3	76	A < G	NS	N	4
MUC4	657640844	3	2	214	T < C	NS	N	4
FGFRL1	661098114	4	6	16	C < A	NS	N	3
TNIP2	662767650	4	6	33	G < A	NS	N	4
LOC100288212	680710684	4	2	44	G < A	S	N	3
GPR125	682521774	4	1	21	C < A	NS	N	2
TBC1D1	698120919	4	19	29	G < A	NS	N	4
SCFD2	711062849	4	1	22	A < C	NS	D	3
SCFD2	711062850	4	1	24	G < T	NS	N	3
KIAA1211	714010832	4	4	25	A < G	NS	D	3
UGT2B28	726937878	4	5	40	A < G	S	N	3
UGT2B28	726937879	4	5	39	A < T	NS	D	3
SULT1B1	727380561	4	5	28	A < C	NS	D	3
ENAM	728278770	4	2	31	G < T	NS	N	3
ANKRD17	730738780	4	29	56	T < A	S	N	3
ANKRD17	730738782	4	29	58	A < G	NS	N	3
FRAS1	735932986	4	6	28	T < A	NS	NO	3
FRAS1	735932989	4	6	28	C < T	S	N	3
AFF1	744724579	4	3	25	G < A	S	N	3
SPARCL1	745172651	4	2	66	G < T	NS	N	4
HERC6	746068457	4	4	52	T < A	NS	N	4
CXXC4	762168791	4	1	23	C < T	NS	N	3
PDE5A	777231322	4	8	81	A < G	S	N	4
FAT4	783168335	4	17	26	A < T	NS	D	3
FAT4	783168337	4	17	25	G < T	NS	N	3
INPP4B[a]	799948368	4	7	144	C < A	NS	N	1
EDNRA	805163608	4	1	58	C < T	S	N	3
RBM46[a]	812505537	4	4	102	T < C	NS	N	1
ACCN5	813543803	4	1	41	G < A	S	N	4
1-Mar	821206546	4	4	89	C < T	S	N	4
DDX60	825930266	4	26	15	C < T	NS	N	3
DDX60	825930267	4	26	15	A < T	NS	N	3
MFAP3L	827669812	4	2	33	G < C	NS	D	3
AGA	835115128	4	5	41	T < A	S	N	3
IRF2	842095828	4	4	19	T < C	NS	D	3
SORBS2[a]	843292510	4	13	232	C < T	NS	D	1
FAM149A	843833669	4	4	50	A < G	NS	D	4
FAT1	844287485	4	14	146	A < G	NS	N	4
FAT1	844298124	4	9	19	G < A	NS	D	3
TRIML2	845769469	4	7	40	C < T	NS	N	2
FRG2	847704735	4	1	72	T < A	NS	N	4
MAFIP	848044252	4	7	56	C < G	S	NR	4
MAFIP	848046049	4	4	33	G < A	S	NR	4
SLC6A18	849416422	5	10	180	C < T	NS	N	4
NDUFS6	849988000	5	4	70	G < T	S	N	3
TRIO	862540906	5	17	113	C < T	S	N	2

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
ANKH	862913981	5	8	116	T < C	S	N	4
FBXL7	864100358	5	3	48	A < G	NS	N	2
RNASEN	879698242	5	2	141	G < A	S	N	4
ADAMTS12	881707057	5	23	45	G < A	NS	N	4
EGFLAM	886579974	5	9	30	T < G	NS	NO	3
EGFLAM	886610442	5	17	63	G < A	S	N	2
CD180	911650805	5	3	38	G < A	S	N	3
MARVELD2	913887307	5	1	47	C < T	NS	N	4
WDR41	921906081	5	10	129	C < T	NS	N	4
GDF9	977321994	5	1	37	G < T	NS	D	3
C5orf15	978417392	5	2	18	G < A	S	N	3
KIF20A	982639408	5	3	29	T < C	S	N	3
KDM3B	982876692	5	14	17	T < C	S	N	4
LOC202051	983854576	5	6	61	C < T	S	N	4
PCDHB11	985652769	5	1	24	T < C	S	N	4
HMHB1	988272050	5	2	125	C < T	NS	D	4
ABLIM3[a]	993692261	5	13	55	A < G	S	N	1
PDGFRB	994581527	5	9	24	A < C	NS	D	1
NDST1	994979638	5	2	57	C < T	S	N	4
NDST1	994984515	5	3	14	C < G	NS	N	3
KIF4B	999468839	5	1	50	C < A	S	N	3
KIF4B	999468844	5	1	48	A < C	NS	N	3
ADAM19	1001937293	5	21	26	C < A	NS	N	3
FBXW11[a]	1016325320	5	8	98	T < C	NS	D	1
C5orf47	1018438284	5	1	27	T < G	S	N	3
FGFR4	1021542240	5	8	26	G < A	NS	N	4
FLT4	1025068341	5	19	33	G < C	NS	D	4
BTNL3	1025454701	5	8	38	G < T	S	N	3
TUBB2A	1029022083	6	4	28	A < G	S	N	3
LRRC16A	1051287607	6	3	142	A < C	S	N	4
SLC17A4	1051637752	6	3	191	C < T	S	N	4
BTN3A2	1052237964	6	3	68	T < C	S	N	4
HLA-G	1055664896	6	5	85	C < T	S	N	4
HLA-A	1055777815	6	2	33	T < A	NS	N	2
HLA-A[a]	1055777819	6	2	85	A < C	S	N	2
C4A	1057829599	6	21	20	T < G	NS	D	2
TNXB	1057902726	6	17	35	G < A	S	N	4
BTNL2[a]	1058229998	6	6	218	C < T	NS	N	2
BTNL2[a]	1058230002	6	6	220	G < A	NS	N	2
HLA-DRB1	1058415838	6	4	83	A < G	S	N	4
HLA-DQB1	1058497104	6	3	50	A < G	S	N	2
HLA-DPB1	1058920866	6	4	94	G < A	NS	N	4
GRM4	1059927081	6	2	94	C < T	NS	D	2
C6orf127	1061622915	6	3	78	A < G	S	N	4
SRPK1	1061704552	6	11	58	G < T	S	N	3
SLC26A8	1061790503	6	16	180	T < C	NS	N	4
TREML2	1067029775	6	3	241	T < C	NS	N	4
TTBK1	1069098172	6	12	37	G < C	NS	N	2
HSP90AB1	1070084817	6	3	24	G < A	NS	N	2
GPR116	1072716510	6	7	52	A < C	S	N	3
CD2AP	1073430855	6	12	58	T < A	NS	N	3
PKHD1	1077359171	6	65	80	T < A	NS	N	3
GSTA2	1078484988	6	4	89	C < G	NS	N	4
GFRAL	1081063844	6	2	32	C < T	NS	N	4
PRIM2	1083334432	6	10	62	A < G	S	NR	4
PRIM2	1083379733	6	13	50	G < A	NS	NR	4
PRIM2	1083379822	6	13	79	T < C	NS	NR	4
EYS	1087198904	6	40	25	G < T	S	N	3
EYS	1087198906	6	40	19	G < C	NS	N	3
IMPG1	1099519071	6	2	37	T < C	NS	N	3

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
ME1	1106705882	6	10	19	T < A	NS	D	3
ME1	1106705883	6	10	17	T < A	NS	NO	3
GABRR1	1112675095	6	5	172	A < G	NS	N	2
MDN1	1113138459	6	88	90	G < A	S	N	4
WISP3	1134999625	6	2	157	T < G	NS	D	4
HS3ST5	1136996498	6	2	16	C < T	NS	N	3
COL10A1	1139059549	6	2	97	A < G	S	N	2
NKAIN2[a]	1147293746	6	3	247	C < G	NS	NO	1
PERP[a]	1161034772	6	2	61	T < C	NS	N	1
HEBP2	1161351286	6	4	56	A < C	S	N	4
NHSL1	1161385538	6	4	32	G < T	NS	D	3
HIVEP2	1165711211	6	1	34	C < T	NS	D	3
SAMD5	1170447330	6	1	31	C < T	S	N	4
PCMT1	1172688307	6	1	41	T < C	NS	D	3
ZBTB2	1174303968	6	2	18	T < A	NS	N	3
SYNE1	1175082156	6	136	30	C < T	S	N	3
SYNJ2	1181103116	6	11	37	T < G	NS	N	3
TULP4	1181540258	6	13	24	A < C	NS	N	3
RSPH3	1182019083	6	6	43	A < T	NS	D	3
IGF2R	1183085535	6	16	66	A < G	S	N	4
AGPAT4	1184192475	6	3	85	C < T	NS	N	2
MLLT4	1190935073	6	19	75	A < C	S	N	4
FAM120B	1193244908	6	1	53	G < A	S	N	4
ADAP1	1194606192	7	6	69	G < A	S	N	4
MICALL2	1195144333	7	7	58	G < C	S	N	4
SDK1	1197713015	7	15	37	G < T	NS	D	3
RSPH10B	1199630320	7	18	54	C < G	S	N	3
VWDE	1206072045	7	12	19	G < T	NS	N	3
HDAC9[a]	1212495378	7	16	176	A < T	NS	D	1
TMEM196	1213427541	7	3	117	G < T	S	N	2
ITGB8[a]	1214065640	7	2	100	G < C	NS	D	1
C7orf10	1234451360	7	14	36	T < A	NS	N	3
C7orf10	1234451361	7	14	38	T < G	NS	NO	3
AEBP1	1237814514	7	18	22	A < T	NS	N	1
MYO1G	1238671649	7	11	87	G < A	S	N	4
C7orf65	1241361009	7	3	60	A < G	S	N	4
ABCA13	1242074221	7	33	18	A < C	NS	D	2
ABCA13	1242105751	7	39	33	C < T	S	N	3
LOC100289307	1263515207	7	2	24	G < T	NS	NR	2
MLXIPL	1263542408	7	7	13	C < T	NS	D	1
SPDYE5	1265649637	7	3	130	G < A	S	NR	4
POR	1266124069	7	2	11	G < T	S	N	3
HGF	1271854360	7	18	27	T < A	NS	D	3
SEMA3E	1273551762	7	11	83	C < T	S	N	4
SEMA3A	1274113176	7	17	291	T < C	S	N	4
FZD1	1281417809	7	1	72	G < T	S	N	3
SAMD9	1283257403	7	1	18	T < A	NS	N	1
PVRIG	1290339880	7	1	17	C < T	NS	N	1
MUC12	1291135269	7	1	14	C < T	S	N	1
MUC12	1291159789	7	5	42	C < T	S	N	4
MUC12	1291160135	7	5	46	G < A	NS	N	4
MUC12	1291160603	7	5	15	T < C	NS	N	4
MUC12	1291161436	7	5	14	G < C	S	N	4
MUC12	1291161644	7	5	52	C < T	NS	N	4
MUC12	1291165668	7	5	25	G < T	NS	D	4
MUC12	1291165672	7	5	29	A < C	S	N	4
MUC12	1291166156	7	5	79	C < G	NS	N	4
MUC12	1291166192	7	5	26	C < A	NS	N	4
MUC17	1291200140	7	3	165	G < A	NS	N	4
MUC17	1291201640	7	3	120	T < A	NS	N	4

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
MUC17	1291203834	7	3	37	T < C	S	N	2
PLOD3	1291376235	7	13	27	G < C	NS	D	4
LOC100132214	1292658130	7	12	31	A < G	NS	D	4
IFRD1	1302624679	7	8	44	T < G	S	N	4
POT1	1314997694	7	11	13	C < T	NS	N	2
POT1	1314997698	7	11	12	A < G	S	N	2
CALD1	1325140404	7	3	19	A < T	NS	N	3
TMEM140	1325371901	7	1	64	A < C	S	N	4
JHDM1D	1330288269	7	20	22	A < T	S	N	3
JHDM1D	1330288270	7	20	24	A < T	NS	D	3
TRBV7-7	1332617173	7	2	155	T < C	S	NR	4
TRBV20-1	1332996157	7	5	110	A < T	NS	NR	4
TRBV20-1	1332996161	7	5	114	C < T	NS	NR	4
LOC441294	1333766303	7	1	41	G < A	S	NR	3
LOC441294	1333766306	7	1	41	A < T	S	NR	3
CTAGE4	1334380174	7	1	29	A < T	NS	N	3
ARHGEF5L	1334381935	7	1	16	T < G	NS	D	2
EZH2	1339023228	7	5	106	C < G	NS	N	4
AGAP3	1341312932	7	7	15	G < T	S	N	1
MLL3	1342375605	7	36	28	G < A	NS	D	3
DLGAP2	1351315601	8	5	24	A < G	S	N	4
MCPH1	1356177925	8	13	40	C < T	S	N	4
FAM90A15	1356815370	8	4	30	C < G	NS	NR	2
TNKS	1359086641	8	2	19	A < T	NS	D	2
TNKS	1359086644	8	2	17	G < C	NS	D	2
RP1L1	1360114631	8	3	25	T < C	NS	N	4
RP1L1	1360116520	8	3	23	T < C	NS	N	2
C8orf74	1360204184	8	3	75	G < T	NS	D	4
MTUS1	1367102384	8	14	33	G < A	S	N	4
DOCK5	1374758768	8	10	113	G < A	S	N	4
C8orf41	1382968877	8	1	13	A < G	S	N	4
CHRNA6	1392210295	8	5	36	G < T	S	N	1
KCNB2	1420442160	8	2	17	A < C	NS	D	3
JPH1	1421750684	8	4	41	T < C	NS	N	3
ZFHX4	1424211672	8	1	17	C < T	S	N	3
CA2	1432983132	8	6	24	G < A	NS	D	3
REXO1L1	1433167938	8	1	21	G < C	NS	D	4
LOC100289448	1433170685	8	1	17	G < C	NS	NR	4
RNF19A	1447714815	8	9	14	C < A	NS	N	2
ANGPT1	1454778129	8	4	54	G < T	NS	N	3
ANGPT1	1454778131	8	4	45	T < C	NS	D	3
COL14A1	1467659774	8	8	87	T < C	S	N	4
ZHX1	1470709988	8	1	42	G < T	NS	N	3
TG	1480325843	8	3	113	G < A	NS	N	2
COL22A1	1486144992	8	36	63	G < T	NS	N	4
FLJ43860	1488920394	8	19	104	G < A	S	NR	4
CYP11B2	1490439544	8	5	202	C < T	S	N	4
LY6H	1490684040	8	3	122	C < G	S	N	4
KIAA0020	1495552364	9	1	35	C < T	NS	N	4
CNTLN	1510054815	9	11	13	A < T	NS	NO	3
LINGO2	1520664425	9	1	19	G < T	S	N	3
PRSS3	1526511755	9	3	163	A < G	NS	N	4
PRSS3	1526512490	9	4	108	T < C	S	N	4
VCP	1527776116	9	8	30	T < C	NS	D	3
FAM75A1	1532072293	9	4	27	C < T	NS	N	2
ALDH1A1[a]	1548388439	9	11	85	G < A	NS	D	1
TLE1	1557099343	9	9	23	G < T	NS	N	3
TLE1	1557099344	9	9	21	T < C	NS	N	3
WNK2	1568934596	9	27	77	A < G	NS	D	3
C9orf129	1568961550	9	2	47	T < C	NS	N	3

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
PTCH1[a]	1571085902	9	17	111	T < C	NS	N	1
GRIN3A	1577199512	9	9	20	A < T	NS	N	3
MUSK	1586295091	9	1	57	G < A	NS	N	3
FKBP15	1588814600	9	13	33	A < C	NS	N	4
COL27A1	1589933932	9	59	118	A < G	NS	N	4
ORM2	1589957833	9	4	37	G < C	NS	D	4
GSN	1596958694	9	17	159	T < C	S	N	4
MAPKAP1	1601110716	9	8	14	A < T	S	N	3
CDK9	1603414267	9	4	23	A < G	NS	D	3
C9orf78	1605455403	9	8	32	A < G	S	N	4
C9orf98	1608412608	9	9	23	T < A	NS	N	1
C9orf98	1608412610	9	9	24	G < A	S	N	1
GFI1B	1608580179	9	6	19	T < A	NS	D	1
GFI1B	1608580182	9	6	19	G < T	NS	NO	1
ABO	1608845331	9	7	39	C < T	NS	NR	4
ABO	1608845366	9	7	43	A < T	NS	NR	4
SARDH	1609287306	9	10	36	G < A	S	N	4
OLFM1	1610646021	9	2	93	T < C	S	N	4
PAEP	1611120049	9	4	108	C < A	NS	N	4
CACNA1B	1613566398	9	28	22	G < A	S	N	3
CACNA1B	1613566402	9	28	24	A < C	NS	N	3
PFKP	1617341440	10	9	103	C < T	S	N	4
AKR1CL2	1619063490	10	2	31	A < T	NS	D	3
ITIH2	1621971113	10	16	62	C < G	NS	N	4
BEND7	1627671894	10	7	32	T < A	S	N	4
ARMETL1	1629060660	10	2	54	T < A	NS	D	3
CUBN	1631072959	10	62	38	T < A	S	N	3
CUBN	1631276445	10	26	49	T < A	NS	N	3
MRC1L1	1632130724	10	24	11	G < T	NS	D	4
PIP4K2A	1637047251	10	6	18	T < A	NS	D	3
ARHGAP21	1639064800	10	25	109	C < T	NS	N	1
TMEM72	1656420929	10	5	46	C < T	S	N	4
ANUBL1	1657125773	10	5	19	C < T	S	N	4
ANXA8L2	1658602609	10	12	48	T < C	S	N	4
AGAP9	1658906406	10	1	38	G < T	NS	D	4
AGAP9	1658906463	10	1	30	T < G	NS	N	4
MSMB[a]	1662146277	10	2	105	A < G	NS	D	1
PCDH15	1666172531	10	34	14	G < T	NS	N	2
PCDH15[a]	1666177751	10	32	70	A < G	S	N	1
TMEM26	1673760923	10	6	22	C < T	NS	N	3
HKDC1	1681600842	10	12	23	C < T	S	N	4
ADAMTS14	1683108476	10	21	79	A < G	NS	N	4
USP54	1685873867	10	15	93	G < A	NS	N	1
DLG5	1690161340	10	23	67	G < T	NS	N	4
FAM22B	1692061925	10	7	25	A < C	S	N	4
BMPR1A	1699271849	10	9	83	C < T	S	N	4
FAM25A	1699372551	10	2	93	A < G	S	N	4
MYOF	1705731605	10	20	24	G < C	NS	D	3
MYOF	1705731606	10	20	24	C < T	S	N	3
TLL2	1708736294	10	15	21	T < G	NS	N	3
TLL2	1708736297	10	15	22	A < T	NS	D	3
MMS19[a]	1709816131	10	18	118	G < A	S	N	1
MMS19[a]	1709816132	10	18	116	C < A	S	N	1
BTRC	1713888546	10	13	19	C < G	NS	D	2
POLL	1713935693	10	3	174	G < T	NS	N	4
PNLIPRP1	1728959073	10	12	57	T < C	NS	D	4
DMBT1	1734932783	10	13	13	G < T	NS	NO	1
DMBT1	1734934333	10	14	38	A < T	S	N	3
DMBT1	1734942466	10	20	106	T < C	S	N	4
CTBP2	1737305458	10	1	13	G < A	S	N	3

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
MMP21	1738045784	10	7	33	G < A	NS	D	3
JAKMIP3	1744395894	10	10	74	A < G	NS	N	4
C10orf93	1745195014	10	4	39	G < A	S	N	2
KNDC1	1745437947	10	5	104	C < T	S	N	4
SYCE1	1745809057	10	13	80	T < C	NS	N	4
FRG2B	1745879355	10	4	105	C < T	S	N	4
SCGB1C1	1746098936	11	2	28	A < G	S	N	4
B4GALNT4	1746279024	11	8	44	G < A	S	N	4
CHID1	1746775660	11	11	37	G < A	NS	N	4
MUC2	1746998500	11	30	33	C < G	S	N	4
MUC2	1746998519	11	30	39	C < A	S	N	4
MUC5AC	1747163509	11	40	33	A < G	S	N	4
MUC5AC	1747176291	11	50	197	G < T	NS	N	4
MUC5AC	1747183167	11	59	41	G < A	NS	N	4
KRTAP5-3	1747534374	11	1	73	C < T	NS	N	3
TNNT3	1747860451	11	10	88	C < A	NS	D	2
ART1	1749586329	11	2	22	A < G	S	N	4
DCHS1	1752558224	11	6	111	C < T	NS	D	2
SOX6	1761982617	11	9	24	T < C	S	N	3
SAAL1	1764013726	11	9	29	T < C	NS	N	4
SAAL1	1764016154	11	7	38	T < A	NS	N	4
MRGPRX3	1764064511	11	1	277	G < A	S	NO	1
MRGPRX3	1764064669	11	1	70	T < C	NS	D	4
NAV2	1765806846	11	5	40	G < A	S	N	4
NAV2	1765972037	11	14	74	A < G	NS	D	3
FANCF	1768551685	11	1	24	C < T	NS	N	3
SLC5A12	1772648382	11	1	29	C < A	NS	NO	3
SLC5A12	1772648383	11	1	29	C < A	NS	N	3
MPPED2	1776462933	11	2	21	G < T	S	N	3
MPPED2	1776462935	11	2	20	C < G	NS	N	3
ZNF408	1792629936	11	4	54	T < A	NS	N	2
GLYAT	1800978539	11	3	36	G < C	NS	D	3
PGA3	1803475551	11	6	166	T < G	NS	D	4
AHNAK	1804789193	11	3	55	T < C	NS	N	4
SIPA1	1807916289	11	15	12	G < T	S	N	1
CATSPER1	1808286286	11	7	52	C < T	NS	N	4
RBM4B	1808942484	11	1	34	A < G	NS	D	3
TPCN2	1811349628	11	19	316	T < C	NS	N	4
FADD	1812550653	11	2	45	G < T	NS	N	2
C11orf30	1818755429	11	19	54	T < C	S	N	4
GDPD4	1819477756	11	8	38	C < T	NS	D	3
ALG8	1820310415	11	13	26	G < A	S	N	3
GAB2	1820434380	11	5	79	A < G	S	N	4
FAT3	1835072125	11	17	53	G < T	NS	N	4
PANX1	1836411162	11	4	21	C < A	NS	N	3
PIWIL4	1836824979	11	9	76	G < C	NS	N	4
CWC15	1837197727	11	5	30	A < G	NS	NR	2
TMEM133	1843211533	11	1	84	A < C	NS	N	4
TRPC6[a]	1843723722	11	2	65	C < T	S	N	1
TMEM123	1844621025	11	3	11	G < A	NS	D	2
ZC3H12C	1852355676	11	2	23	G < T	NS	N	3
LAYN	1853779209	11	7	40	G < A	NS	N	2
ZW10	1855955650	11	15	49	A < C	NS	N	3
CEP164	1859631014	11	31	58	G < T	S	N	4
DSCAML1	1859657009	11	25	62	G < A	S	N	4
DSCAML1	1859751449	11	4	51	G < T	NS	N	4
IL10RA	1860212261	11	4	49	A < G	S	N	4
TMPRSS4	1860336319	11	12	79	C < T	NS	D	4
BCL9L	1861117811	11	8	14	G < T	NS	D	1
CCDC84	1861234319	11	10	45	C < G	S	N	2

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
ZNF202	1865948687	11	2	14	C < A	NS	NO	3
HSN2	1878252896	12	1	49	T < C	S	N	1
VWF	1883406765	12	25	19	T < C	NS	D	2
ACSM4	1884701983	12	11	32	G < T	NS	D	3
PRB2	1888770834	12	3	58	T < C	S	N	4
PRB2	1888770905	12	3	100	G < T	NS	N	2
PRB2	1888771519	12	3	100	C < T	NS	N	4
PIK3C2G	1895940574	12	25	19	C < T	NS	D	3
SLCO1C1	1898110764	12	9	41	A < G	S	N	4
ARID2	1920469949	12	15	35	T < C	S	N	3
AMIGO2	1921696779	12	1	29	C < A	NS	N	3
KRT86	1926923866	12	5	158	G < A	NS	N	4
KRT86	1926923877	12	5	167	C < G	S	N	4
KRT2	1927263989	12	9	24	C < A	NS	D	3
KRT2	1927263991	12	9	24	C < A	NS	D	3
NCKAP1L	1929139765	12	18	111	C < T	S	N	4
RDH16	1931575957	12	1	144	G < T	NS	D	2
LRP1	1931814492	12	54	29	C < T	S	N	4
LRIG3	1933499258	12	13	28	T < A	NS	D	3
TMEM5	1938398588	12	1	72	G < A	S	N	4
RASSF9	1960454767	12	1	16	T < A	NS	NO	3
C12orf12	1965572650	12	1	74	C < A	NS	N	3
NUP37	1976695411	12	7	27	A < C	NS	D	3
USP30	1983694022	12	8	37	C < T	S	N	4
C12orf51	1986840857	12	36	47	T < C	S	N	2
DDX54	1987789501	12	7	85	G < A	S	N	4
PLBD2	1987987438	12	5	27	C < T	NS	NO	3
SDSL	1988046794	12	4	40	G < T	NS	N	3
MED13L	1990588227	12	24	18	C < A	NS	D	3
CIT	1994346828	12	24	71	T < C	S	N	4
ORAI1	1996254022	12	2	53	C < T	S	N	4
B3GNT4	1996814014	12	1	20	C < G	NS	N	4
CLIP1	1996970528	12	4	25	G < A	S	N	2
SBNO1	1997954707	12	3	23	T < C	NS	N	3
SETD8	1998000144	12	3	24	C < T	S	N	2
GPR133	2005609817	12	9	29	A < G	S	N	3
POLE	2007276853	12	12	66	T < C	NS	D	3
PGAM5	2007320186	12	6	59	C < T	NS	N	4
TPTE2	2008847341	13	18	64	A < G	S	N	4
TPTE2	2008847342	13	18	61	T < A	NS	N	4
PARP4	2013890787	13	15	79	G < A	S	N	1
SLC7A1	2018953795	13	2	58	G < C	S	N	4
NBEA	2024476994	13	7	190	T < C	S	N	4
DCLK1	2025231759	13	11	53	G < T	S	N	4
KBTBD6	2030552300	13	1	29	C < T	NS	N	4
MED4	2037507221	13	3	16	T < A	NS	D	3
RB1	2037880563	13	20	49	G < A	S	D	1
RCBTB1	2038988090	13	1	88	G < A	S	N	4
PCDH9	2056646696	13	1	23	T < A	NS	NO	3
PCDH9	2056646999	13	1	39	C < A	NS	N	3
KLF12	2063234134	13	4	43	C < G	NS	N	3
COL4A1	2099524302	13	37	135	T < A	S	N	4
C13orf16	2100677265	13	2	112	C < T	S	N	4
ATP11A	2102082860	13	29	44	T < C	S	N	4
RASA3	2103108926	13	21	105	C < T	S	N	4
POTEG	2104010042	14	1	119	C < T	NS	N	4
P704P	2104476800	14	1	60	C < T	S	N	4
NDRG2	2105942492	14	15	51	G < C	NS	D	4
HAUS4	2107873469	14	7	11	G < T	S	N	2
HOMEZ	2108202700	14	2	14	A < T	NS	N	3

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
DHRS4	2108891643	14	4	116	C < T	NS	NO	4
DHRS4L2	2108914889	14	1	132	G < T	NS	D	3
DHRS4L2	2108916099	14	2	35	G < A	S	N	4
GZMH	2109533483	14	3	196	G < C	NS	N	4
GZMH	2109533541	14	3	38	C < A	NS	D	3
C14orf182	2134929075	14	1	27	A < T	NS	D	3
MAP4K5	2135379887	14	13	28	G < A	NS	D	3
PPM1A	2145206418	14	2	80	G < T	NS	N	3
SYNE2	2148878083	14	7	85	A < G	NS	D	1
PLEKHG3	2149666422	14	14	36	C < A	NS	D	2
PLEKHG3	2149666423	14	14	38	C < T	NS	D	2
GPHN	2151839368	14	6	21	G < A	S	N	3
SIPA1L1	2156512152	14	1	22	A < T	S	N	3
DIO2	2165125971	14	3	63	C < T	S	N	3
FLRT2	2170544935	14	1	20	G < T	NS	N	3
DDX24	2178978135	14	6	101	A < C	NS	D	2
BEGAIN	2185466810	14	4	97	T < C	S	N	4
C14orf73	2188025488	14	2	16	T < G	S	N	2
TMEM179	2189527477	14	1	28	C < G	NS	N	2
ADSSL1	2189666079	14	10	107	G < A	NS	D	2
AHNAK2	2189862548	14	7	57	G < A	S	N	4
AHNAK2	2189862844	14	7	26	A < C	NS	N	4
LOC727832	2192485885	15	8	14	A < G	NS	N	1
C15orf2	2196518354	15	1	17	C < G	S	N	3
GOLGA8G	2200364560	15	8	34	T < A	NS	D	3
GOLGA8G	2200368232	15	3	23	T < G	NS	D	4
CHRNA7	2203996850	15	7	134	G < A	S	N	4
RYR3	2205498123	15	33	72	T < G	S	N	4
RYR3	2205507714	15	37	44	T < G	NS	D	1
SRP14	2211874811	15	5	170	G < A	S	N	4
STARD9	2214531200	15	23	19	A < T	NS	N	3
DMXL2	2223337705	15	18	32	A < G	NS	N	4
RNF111	2230919345	15	7	17	C < T	S	N	3
ANXA2	2232187466	15	12	29	G < A	NS	D	2
ITGA11	2240170436	15	14	59	G < A	S	N	4
GOLGA6B	2244497986	15	4	49	G < A	S	N	4
GOLGA6	2245910065	15	15	18	G < T	NS	N	1
CYP1A2	2246588818	15	1	41	A < G	NS	D	3
GOLGA6C	2247104814	15	11	23	A < T	NS	N	4
GOLGA6C	2247104889	15	11	35	G < A	NS	N	4
GOLGA6C	2247106801	15	13	22	G < A	NS	N	2
CSPG4	2247528218	15	3	39	G < A	NS	D	2
SGK269	2248972147	15	3	11	G < C	NS	N	3
KIAA1024	2251295855	15	1	14	C < A	NS	N	3
AP3B2	2254828471	15	20	55	A < G	S	N	1
LOC100288732	2260124462	15	5	24	G < C	NS	NR	3
LOC100288732	2260124464	15	5	27	T < A	NS	NR	3
KIF7	2261638402	15	3	16	T < G	NS	D	2
SEMA4B	2262210365	15	5	49	G < A	NS	N	4
FURIN	2262865860	15	3	115	T < G	NS	D	4
MEF2A	2271698884	15	9	28	A < C	NS	N	3
ADAMTS17	2271960760	15	22	65	T < C	NS	N	4
HBA2	2274131104	16	3	26	T < G	S	N	3
PDIA2	2274242118	16	2	51	G < C	S	N	4
JMJD8	2274641266	16	4	23	T < A	NS	N	3
PRSS22	2276813235	16	4	38	C < T	NS	N	4
CLDN9	2276971003	16	1	43	C < T	S	N	1
ALG1	2279037309	16	9	58	T < C	S	N	4
TMEM114	2282529635	16	1	33	A < T	NS	D	4
TEKT5	2284677496	16	5	142	T < C	NS	N	4

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
NOMO2	2292457809	16	10	19	T < A	NS	N	4
TMC5	2293409286	16	16	94	C < T	NS	NO	1
ACSM5[a]	2294342868	16	5	162	C < G	NS	N	1
ACSM5	2294348591	16	7	60	C < G	NS	D	4
NPIPL3	2295321650	16	8	32	A < G	S	N	4
OTOA	2295663836	16	23	35	C < T	NS	D	4
VWA3A	2296051856	16	20	40	C < T	NS	N	4
PRKCB	2298073558	16	10	19	C < T	NS	NO	3
KIAA0556	2301696816	16	27	72	G < A	NS	N	4
TUFM[a]	2302763330	16	6	66	C < A	S	N	2
INO80E	2303915676	16	3	14	A < G	S	N	3
POL3S	2305006549	16	3	23	T < C	NS	N	4
ERAF	2305447472	16	2	18	G < T	S	N	4
LOC100287647	2307849444	16	2	151	G < A	S	NR	1
ABCC12	2310775449	16	28	22	T < G	NS	D	3
ABCC12	2310775450	16	28	22	C < A	NS	D	3
ABCC12	2310796723	16	19	112	G < A	S	N	1
BRD7	2313015114	16	12	42	A < C	NS	NO	4
SALL1	2313830540	16	2	19	A < C	NS	D	3
CETP	2319673630	16	14	59	G < A	NS	N	4
SETD6	2321208022	16	5	48	G < A	S	N	4
CDH5	2329078423	16	2	105	C < T	S	N	4
PDPR	2332847939	16	17	49	C < T	S	N	4
PKD1L2	2343889813	16	7	21	G < A	NS	NR	4
MPHOSPH6	2344839892	16	5	61	G < A	S	N	4
CRISPLD2	2347537002	16	2	20	A < G	NS	N	4
FAM38A	2351390771	16	35	14	C < T	NS	N	4
FAM38A	2351393656	16	33	16	A < C	NS	D	1
WDR81	2354433283	17	1	18	G < A	S	N	4
WDR81	2354443084	17	10	67	C < T	S	N	4
TSR1	2355042000	17	1	34	G < A	NS	D	3
TRPV3	2356238371	17	7	155	C < T	S	N	4
TRPV3	2356249176	17	4	23	T < C	NS	N	4
ITGAE	2356435127	17	24	49	G < A	NS	D	4
ITGAE	2356463343	17	9	31	G < A	S	D	4
ZZEF1	2356772099	17	28	13	C < A	NS	N	3
GGT6	2357265990	17	1	13	G < A	NS	N	4
CXCL16	2357444046	17	3	22	C < T	S	N	4
TEKT1	2359518533	17	5	87	G < A	NS	D	4
AMAC1L3	2360187655	17	2	35	G < A	NS	N	4
AMAC1L3	2360187863	17	2	20	G < A	NS	N	4
AMAC1L3	2360188508	17	2	179	T < C	NS	N	4
TP53[a]	2360379830	17	6	38	G < A	NS	D	1
MYH13	2363018801	17	28	98	G < A	S	N	4
COX10	2366807730	17	4	62	G < A	S	N	4
COX10	2366897810	17	6	105	C < T	S	N	4
FAM18B2	2368251449	17	5	57	A < G	NS	N	4
FAM18B2	2368259332	17	3	141	A < C	S	N	4
TBC1D26	2368443097	17	3	80	A < C	NS	D	4
TBC1D26	2368443106	17	3	90	A < G	NS	N	4
SHMT1	2371041192	17	7	30	C < T	S	N	3
LGALS9C	2371193284	17	4	95	C < T	S	N	4
LGALS9C	2371193294	17	4	83	G < A	NS	N	4
ULK2	2372501833	17	19	21	G < T	S	N	3
KCNJ12	2374121073	17	1	157	G < A	NS	D	4
KCNJ12	2374121499	17	1	55	C < T	NS	D	4
KCNJ12	2374121521	17	1	67	G < C	NS	N	4
KIAA0100	2376657621	17	24	46	A < C	NS	N	4
SUPT6H	2376730797	17	36	14	C < T	S	N	1
CCL8[a]	2382349668	17	2	51	A < G	NS	N	1

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
TBC1D3B	2384202011	17	5	27	C < T	NS	D	4
TBC1D3E	2384401830	17	9	36	A < T	NS	D	4
TBC1D3D	2385938140	17	2	61	A < G	NS	N	4
TBC1D3D	2385940014	17	4	85	C < A	NS	N	4
TBC1D3D	2385940944	17	6	147	G < A	NS	N	4
ERBB2	2387531879	17	17	51	A < G	NS	N	4
TOP2A	2388219990	17	9	19	T < C	S	N	3
KRT25	2388559828	17	4	159	G < A	S	N	2
KRT26	2388580603	17	1	18	A < T	S	N	3
KRT40	2388787496	17	6	48	A < G	S	N	4
KRTAP3-2	2388808375	17	1	95	T < C	NS	N	4
KRTAP1-1	2388849768	17	1	37	G < C	NS	D	4
KRTAP4-1	2388993082	17	2	33	G < C	NS	N	4
KRTAP4-1	2388993086	17	2	30	A < G	S	N	4
KRTAP9-4	2389058283	17	1	217	C < T	NS	D	4
KRTAP9-4	2389058336	17	1	39	A < C	NS	N	3
KRTAP9-9	2389063986	17	1	44	A < C	NS	N	4
TUBG1	2390418855	17	10	14	C < A	S	N	1
BRCA1	2390878791	17	13	31	C < T	NS	NO	3
NAGS	2391737375	17	6	45	G < A	NS	N	3
CDK5RAP3	2395703185	17	3	61	T < G	NS	D	2
FAM117A	2397445898	17	7	17	G < A	S	N	3
ITGA3	2397801147	17	6	33	T < C	S	N	4
NOG	2404324293	17	1	67	A < C	NS	N	3
MTMR4	2406238200	17	6	13	A < T	NS	NO	3
CSH2	2411602334	17	4	55	C < T	NS	N	4
GH2	2411610386	17	4	18	G < C	NS	N	1
TEX2	2411943294	17	1	21	A < T	NS	D	3
COG1	2420849730	17	7	78	C < T	S	N	4
GPR142[a]	2422019077	17	3	57	A < G	NS	D	1
UNK	2423468317	17	14	19	A < G	S	N	4
QRICH2	2423941144	17	4	44	T < G	NS	N	4
HRNBP3	2426764023	17	1	22	G < A	S	N	4
CBX4	2427461157	17	5	66	C < T	NS	N	2
RNF213	2427979649	17	9	73	G < A	S	N	4
MYL12A	2434375274	18	1	40	G < A	NS	N	3
MYL12A	2434375275	18	1	41	A < T	NS	D	3
AMAC1L1	2442732047	18	1	22	G < A	S	N	2
C18orf1	2444767403	18	5	65	A < G	S	D	3
LOC729774	2445483471	18	2	38	G < T	NS	NR	3
POTEC	2445664872	18	1	68	T < C	NS	N	2
CTAGE1	2448017831	18	1	35	C < T	NS	N	1
KCTD1	2452149054	18	1	18	T < G	NS	N	3
DSG4	2457014977	18	15	41	G < T	NS	NO	3
FAM59A	2457889818	18	4	30	G < T	S	N	3
FAM59A	2457889821	18	4	29	A < C	NS	D	3
MOCOS	2461870479	18	15	143	T < C	NS	N	4
SLC14A2	2471234230	18	4	39	G < A	NS	N	3
KIAA1632	2471505962	18	25	17	A < T	NS	N	3
KIAA1632	2471505963	18	25	17	A < T	NS	NO	3
FUSSEL18	2472796815	18	1	35	A < T	NS	D	1
ZBTB7C	2473578004	18	2	78	T < G	NS	D	4
ZBTB7C	2473588900	18	1	208	T < C	S	N	4
KIAA0427	2474259908	18	6	22	T < A	NS	N	3
CXXC1	2475832249	18	10	67	A < G	S	N	4
TCF4	2480799098	18	12	33	C < A	NS	D	3
TCF4	2481003232	18	3	19	C < A	S	N	3
CCBE1	2485008619	18	4	99	C < T	S	N	2
NETO1[a]	2498404004	18	3	115	A < T	NS	D	1
C19orf6	2506845912	19	4	13	T < G	NS	D	1

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
ABCA7	2506887468	19	25	15	A < G	S	N	4
REXO1	2507659336	19	3	14	C < T	S	N	4
FAM108A1	2507714358	19	2	13	T < C	NS	N	4
PIP5K1C	2509476658	19	13	27	G < A	NS	N	3
MAP2K2	2509934448	19	6	47	C < T	NS	D	4
ACER1	2512145698	19	3	26	T < C	NS	N	4
LASS4	2514105354	19	7	55	T < C	S	N	4
MUC16	2514735984	19	51	113	T < C	NS	N	4
MUC16	2514735995	19	51	104	C < T	NS	N	4
MUC16	2514783318	19	5	66	C < T	NS	N	4
MUC16	2514793493	19	3	59	C < T	NS	N	4
MUC16	2514797580	19	3	46	T < C	S	N	3
ICAM3	2516182922	19	2	22	T < C	NS	N	4
MAST1	2518717997	19	26	18	G < A	S	N	3
CYP4F12	2521526643	19	5	51	T < C	NS	N	4
OR10H2	2521572582	19	1	89	C < T	S	N	4
AP1M1	2522077812	19	11	81	C < G	NS	D	3
CPAMD8	2522819358	19	17	55	G < A	NS	N	3
KIAA1683	2524110904	19	2	62	C < G	NS	N	4
ISYNA1	2524280086	19	7	57	T < C	S	N	4
KIAA0892	2525185534	19	7	61	G < A	S	N	3
ZNF536	2533672403	19	3	28	T < C	S	N	4
GPI	2537503303	19	7	71	C < G	S	N	4
CD22	2538462717	19	5	17	C < G	S	N	3
C19orf15	2541461368	19	1	88	T < C	NS	N	4
MAP4K1	2541732174	19	14	25	G < A	NS	N	4
CAPN12	2541857821	19	18	71	A < G	S	N	4
LGALS4[a]	2541932950	19	3	30	C < A	NS	D	1
ECH1	2541939937	19	9	29	C < T	NS	D	4
PLEKHG2	2542544839	19	12	22	C < T	S	N	3
FCGBP	2543017507	19	21	76	G < A	S	N	4
FCGBP	2543053201	19	6	13	G < T	S	N	1
SNRPA	2543896811	19	2	60	A < G	S	N	4
CYP2F1	2544255597	19	1	31	G < A	S	N	4
ERF	2545386691	19	4	32	G < A	S	N	4
PSG3	2545867483	19	4	156	C < A	S	N	4
PSG8	2545901763	19	2	60	C < A	NS	D	3
CEACAM20	2547650657	19	7	58	T < C	NS	NR	4
ERCC2	2548501717	19	6	71	T < G	S	N	4
EMP3	2551464282	19	2	40	G < T	NS	NO	1
TMEM143	2551479358	19	6	19	A < C	NS	D	3
PTH2	2552559200	19	2	22	G < C	NS	N	3
SHANK1	2553853011	19	2	18	T < C	NS	N	1
ZNF808[a]	2555691982	19	3	39	G < A	NS	N	1
ZNF765	2556544681	19	3	19	C < G	NS	N	1
ZNF765	2556544684	19	3	19	T < C	S	N	1
ZNF761	2556586232	19	2	16	G < A	S	NR	4
LILRB3[a]	2557359732	19	3	208	G < C	NS	N	2
LILRA1	2557740721	19	5	61	T < C	NS	N	2
KIR2DL4	2557949666	19	3	17	C < G	NS	N	3
KIR3DL1	2557963191	19	3	109	A < G	S	N	4
KIR2DS4	2557982650	19	3	37	T < G	S	NR	2
KIR2DS4	2557982701	19	3	36	G < T	NS	NR	2
KIR2DS4	2557982728	19	3	21	G < C	NS	NR	2
RDH13	2558201492	19	1	28	C < T	NS	N	4
RDH13	2558201493	19	1	29	G < T	NS	N	4
ZFP28	2559692350	19	3	33	C < A	S	N	3
ZNF550[a]	2560701027	19	1	67	C < T	NS	D	1
ZSCAN22	2561483213	19	2	17	T < G	NS	N	3
KIR2DS1	2561871082	19	3	153	A < G	S	NR	2

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
SIGLEC1	2565630685	20	3	19	C < T	NS	N	4
PAK7	2571487871	20	4	131	C < T	NS	N	4
FLRT3	2576252200	20	1	17	T < G	NS	D	3
CST9L	2585490888	20	2	216	T < G	NS	N	4
BPI	2595650275	20	11	148	A < G	NS	N	4
LBP	2595672219	20	2	23	G < A	S	N	4
LBP[a]	2595691968	20	10	25	G < T	S	N	1
KIAA1219	2595847726	20	10	14	T < G	NS	D	3
PTPRT	2599404902	20	31	37	C < A	NS	D	3
SEMG2	2602544774	20	2	42	A < G	S	N	4
ZNF335	2603273210	20	21	32	A < G	S	N	4
PCK1	2614832083	20	3	52	A < G	S	N	4
CTSZ	2616266012	20	5	51	A < G	S	N	4
OGFR	2620085220	20	4	20	G < A	S	N	4
KCNQ2	2620689776	20	14	14	C < G	NS	N	4
LOC100132288	2622007697	21	2	24	C < T	NS	NR	4
LOC100288017	2623267208	21	1	18	G < A	NS	NR	2
POTED	2623681322	21	1	60	G < A	NS	N	2
KRTAP13-2	2640442822	21	1	81	A < T	NS	D	3
C21orf66	2642816546	21	12	14	A < C	NS	N	1
C21orf66	2642816547	21	12	14	C < T	NS	N	1
WRB	2649461227	21	2	73	G < T	NS	NO	4
WRB	2649461228	21	2	73	A < T	NS	D	4
DSCAM	2650145634	21	27	116	G < C	S	N	4
PRDM15	2651870131	21	31	65	G < A	S	N	4
PFKL	2654380692	21	4	19	C < T	S	N	4
KRTAP10-6	2654660390	21	1	137	G < A	S	N	4
KRTAP12-2	2654734983	21	1	59	C < T	NS	N	4
KRTAP12-2	2654735333	21	1	71	G < A	S	N	4
KRTAP12-2	2654735334	21	1	69	C < T	NS	N	4
COL6A2	2656200935	21	27	63	C < G	S	N	3
FTCD	2656222648	21	2	55	T < A	NS	D	2
CECR5	2658218162	22	6	14	T < G	NS	D	3
CECR2	2658624712	22	16	25	C < T	S	N	4
LOC100288065	2658662354	22	4	63	A < G	NS	N	4
TBX1	2660347982	22	4	51	C < T	S	N	4
ZNF280B[a]	2663338671	22	1	78	G < A	S	N	2
C22orf30	2672604693	22	3	17	C < T	NS	NO	3
ISX	2675974809	22	2	60	G < A	S	N	1
HMGXB4	2676157665	22	4	42	A < T	S	N	3
APOL1	2677147183	22	2	27	C < T	NS	N	1
TMPRSS6	2677959079	22	17	144	G < A	S	N	4
TMPRSS6	2677959089	22	17	133	A < G	NS	N	4
SSTR3	2678099174	22	1	18	G < A	S	N	4
APOBEC3A	2679853734	22	3	53	C < T	S	N	4
L3MBTL2	2682109341	22	5	24	C < T	S	N	4
NAGA	2682959967	22	3	46	C < T	S	N	4
TTLL12	2684071796	22	5	53	C < T	S	N	3
SCUBE1	2684110469	22	15	48	C < G	S	N	4
LOC100289317	2686220100	22	1	25	A < G	NS	NR	4
CELSR1	2687425845	22	1	58	A < G	NS	N	4
MAPK8IP2	2691494874	22	11	52	G < A	NS	N	4
CD99	2693971425	X	6	16	G < A	NS	D	4
PRKX	2694923444	X	2	88	G < A	S	N	4
ARHGAP6[a]	2702537703	X	4	81	G < A	NS	D	2
DMD	2723711682	X	37	92	T < C	NS	N	4
WAS	2739828502	X	11	13	G < T	NS	N	3
GATA1	2739931290	X	2	14	A < C	NS	N	3
GAGE12E	2740547782	X	1	62	G < T	NS	D	4
PAGE1[a]	2740686656	X	3	37	T < C	S	N	1

Supplementary Table 3. Continued

Gene	Reference position	Chromosome	Coding sequence	Coverage	Allele change	Amino acid change	Functional predictions by SIFT	Patient no.
USP27X	2740875896	X	1	39	G < C	NS	NR	3
TSPYL2	2744246047	X	6	76	A < G	NS	N	3
FAM120C	2745238526	X	14	66	C < A	NS	D	2
ITIH5L	2745914425	X	8	22	G < A	NS	N	3
MSN	2752987418	X	9	47	G < A	S	N	2
OPHN1	2755314567	X	20	47	T < C	NS	N	3
DGAT2L6	2757452592	X	5	20	C < T	S	N	3
LPAR4	2765991350	X	1	46	G < T	NS	D	3
LPAR4	2765991352	X	1	43	C < T	S	N	3
PCDH11X	2779854213	X	7	171	C < A	NS	N	3
SYTL4	2787924130	X	9	24	G < A	S	N	3
SYTL4	2787924131	X	9	24	T < C	NS	D	3
NXF5	2789077402	X	3	81	T < C	NS	D	1
NXF2	2789554971	X	10	19	C < T	S	N	3
CLDN2	2794152808	X	1	26	C < A	NS	N	3
CLDN2	2794152809	X	1	26	C < T	NS	N	3
TRPC5	2799176319	X	1	24	G < A	NS	D	3
TRPC5	2799176320	X	1	24	G < T	NS	N	3
RHOXF2B	2807087244	X	4	29	A < G	NS	N	3
PLAC1	2821580905	X	1	20	A < G	NS	N	3
RBMX	2823837181	X	8	40	G < C	NS	D	4
SLITRK4	2830598721	X	1	56	T < C	NS	N	3
NSDHL	2839816887	X	6	109	A < G	NS	D	3
MPP1	2841799005	X	5	36	T < C	S	N	3
MPP1	2841799006	X	5	36	T < C	NS	N	3
RBMY1D	2863471281	Y	11	23	T < C	S	N	1

SIFT, Sorting Intolerant From Tolerant; N, nonsynonymous mutation; D, deleterious; N, neutral; S, synonymous mutation; NO, nonsense mutation; NR, no record found.
[a]These genes were commonly mutated in the synchronously developed HCCs from patients 1 and 2.

Supplementary Table 4. Functional Relevance of Mutations Detected in HCC Tumors

Patient 1					
HCC 1	No. of mutated genes	51			
	Amino acid change (+)	38 (74.5%)			
	Functional loss[a] (+)	20 (39.2%)			
	KEGG pathway				
	Pathways in cancer	CTNNB1	PDGFRB	TP53	
	Wnt signaling pathway	FBXW11			
	PI3K-Akt signaling pathway	ITGB8			
	Others	ALDH1A1	HDAC9	SORBS2	RYR3
HCC 2	No. of mutated genes	79			
	Amino acid change (+)	58 (73.4%)			
	Functional loss[a] (+)	23 (29.1%)			
	KEGG pathway				
	Viral carcinogenesis	HDAC9	RB1	TP53	
	Wnt signaling pathway	FBXW11			
	PI3K-Akt signaling pathway	ITGB8			
	Others	ALDH1A1	NXF5		
Patient 2					
HCC 1	No. of mutated genes	39			
	Amino acid change (+)	20 (51.3%)			
	Functional loss[a] (+)	10 (25.6%)			
	KEGG pathway				
	Metabolic pathways	DBH			
	Others	AGRN			
HCC 2	No. of mutated genes	70			
	Amino acid change (+)	40 (57.1%)			
	Functional loss[a] (+)	20 (28.6%)			
	KEGG pathway				
	Metabolic pathways	ADSSL1	FTCD	RDH16	
	Others	ABCA13	BTRC	VWF	C4A
		GRM4			
Patient 3					
HCC 1	No. of mutated genes	30			
	Amino acid change (+)	20 (66.7%)			
	Functional loss[a] (+)	6 (20.0%)			
	KEGG pathway				
	Metabolic pathways	CYP1A2			
HCC 2	No. of mutated genes	276			
	Amino acid change (+)	208 (75.3%)			
	Functional loss[a] (+)	90 (32.6%)			
	KEGG pathway				
	Metabolic pathways	ACSM4	ADSS	UGT2B28	DHRS4L2
		GALNT5	ME1	POLE	NSDHL
		PIK3C2G			
	PI3K-Akt signaling pathway	COL6A6	HGF	ANGPT1	LPAR4
	Neuroactive ligand receptor	GLRA2			
	Others	CDK9	CA2	ABCC12	AP1M1
		GNPAT	GLYAT	RUVBL1	GDF9
		MYL12A	MLL3	SLC18A2	MAP4K4
		PRPF8	PIP4K2A	SLC9A4	NUP37
		VCP	TTN		

Supplementary Table 4. Continued

Patient 4					
HCC	No. of mutated genes	364			
	Amino acid change (+)	177 (48.6%)			
	Functional loss[a] (+)	46 (12.6%)			
	KEGG pathway				
	Metabolic pathways	ACSM5	ALPP	PNLIPRP1	
	MAPK signaling pathway	HSPA6	MAP2K2		
	PI3K-Akt signaling pathway	FLT4			
	Others	ECH1	CHI3L1	FURIN	CD99
		KCNJ12	ITGAE	TMPRSS4	REXO1L1
		RBMX	PLOD3	TUBA4A	PGA3

[a]The number of mutated genes predicted to be "damaging (deleterious)" by Sorting Intolerant From Tolerant (SIFT) functional impact predictions (http://provean.jcvi.org/index.php). The genes categorized in multiple pathways are shown in only one representative pathway.

Supplementary Table 5. List of 448 Indels in 409 Genes at a Frequency of >20% of Reads in 7 HCC Tumors From 4 Patients

Reference position	Gene	Chromosome	Coding sequence	Coverage	Allele change	Patient no.
7813482	ERRFI1	1	3	17	insA	3
12628605[a]	PRAMEF11	1	3	27	insC	2
12718307	PRAMEF7	1	2	47	insT	3
17358674[a]	PADI6	1	9	41	delG	4
17358674[a]	PADI6	1	9	62	delGT	2
26696284	ARID1A	1	2	19	delC	4
31395890[a]	SERINC2	1	9	17	insG	3
46770747	CYP4B1	1	8	42	delAT	4
46770748	CYP4B1	1	8	42	delT	4
52949277[a]	LOC100133211	1	1	42	delG	4
53189215	MAGOH	1	3	25	insA	3
54095320[a]	CDCP2	1	4	21	insC	4
62557547	ANGPTL3	1	4	17	insT	3
78876012	ELTD1	1	10	13	insA	3
89014597	GBP1	1	4	17	insA	2
90670304	BARHL2	1	2	20	insC	3
108974106	CLCC1	1	6	84	insG	4
122705624[a]	PDE4DIP	1	14	80	delG	4
122713730[a]	PDE4DIP	1	6	465	delT	1
131401399	DENND4B	1	12	43	insG	3
131474740	NUP210L	1	34	13	insG	3
131951749	SHE	1	3	112	insA	2
133072970	MSTO1	1	11	33	insA	3
133795656	CCT3	1	2	34	insT	3
133844355[a]	RHBG	1	9	40	delC	4
134043234[a]	TTC24	1	3	12	delC	3
136505109	IFI16	1	7	31	insT	3
146999551	F5	1	13	23	insT	3
151907797	GPR52	1	1	18	insA	2
162596811	C1orf25	1	10	26	insT	3
163815430	TPR	1	15	43	delT	3
174374155	CFHR4	1	5	18	insC	3
177507300	NR5A2	1	5	20	insT	3
201685738	NVL	1	6	32	insA	3
203543660	ACBD3	1	2	18	insG	3
208315870[a]	ARV1	1	3	146	delCT	1
208315871[a]	ARV1	1	3	144	delT	1
213162772	LYST	1	3	19	insG	3
214770350	RYR2	1	11	22	insG	3
219641707	PLD5	1	2	102	insA	1
224245114	AHCTF1	1	14	19	insG	3
230467928	RNASEH1	2	8	13	insT	3
236404886	ADAM17	2	19	29	insG	3
251086730[a]	LOC375190	2	8	40	insC	4
254371933	IFT172	2	38	18	insT	3
258505433[a]	SRD5A2	2	1	28	insG	2
267355164	SLC8A1	2	1	15	insG	3
282849120	EFEMP1	2	1	14	insA	3
287714203	PAPOLG	2	15	20	delT	3
288274645	USP34	2	15	14	insC	3
301408194	CCDC142	2	2	27	insC	3
302628924	C2orf3	2	3	22	insA	3
312276713	RETSAT	2	4	55	delC	2
314781292	RGPD2	2	5	27	insT	3
317672121[a]	LOC391405	2	4	43	delA	4
318709414	TRIM43	2	1	22	insA	3
325574912	SLC9A4	2	6	27	insT	3
329490991	RGPD3	2	20	29	insT	3
331565693	GCC2	2	22	18	insC	2
332910299	RGPD5	2	21	25	insA	3
333616906	RGPD7	2	8	21	delA	3

Supplementary Table 5. Continued

Reference position	Gene	Chromosome	Coding sequence	Coverage	Allele change	Patient no.
335725611	SLC20A1	2	8	14	insC	3
336700192	RABL2A	2	4	76	delG	1
340887331	DDX18	2	7	18	insT	3
350555029	IWS1	2	11	62	delG	3
355383031[a]	ZNF806	2	3	50	delC	4
355383457[a]	ZNF806	2	3	56	insA	4
355383669[a]	ZNF806	2	3	52	delA	4
361736116	NXPH2	2	2	48	insT	3
385464478	KCNH7	2	10	41	insT	3
387807820	COBLL1	2	2	23	insA	3
388452944	SCN2A	2	26	57	insT	3
400689016	TTC30A	2	1	32	insA	3
401650351	TTN	2	270	19	insG	3
401670059	TTN	2	242	25	insT	3
401800064	TTN	2	64	26	insA	3
402189048	SESTD1	2	14	31	insG	3
403038177	CWC22	2	11	16	insA	3
418958541	DNAH7	2	34	27	insC	3
439213578	XRCC5	2	13	22	delC	3
446013793	ACSL3	2	14	26	insT	3
446670417	SCG2	2	1	17	insA	3
446671148	SCG2	2	1	30	insC	3
456401277[a]	SAG	2	10	76	delA	2
463724350	AQP12B	2	1	26	delC	2
479793553[a]	GRIP2	3	10	55	insG	4
504388108	TTC21A	3	6	13	insA	3
509772719	ZNF852	3	3	20	delTC	4
509772720	ZNF852	3	3	19	delC	4
511646368	CCR5	3	1	24	insT	3
515483759[a]	SLC38A3	3	2	18	insG	4
538148646	GLT8D4	3	6	14	insC	3
538243406[a]	FLJ10213	3	1	12	insA	2
538564462	PDZRN3	3	10	24	insG	3
540846731[a]	LOC100288801	3	2	39	delG	2
540918687	ZNF717	3	4	18	delC	1
570208682	HHLA2	3	4	18	insT	3
570487430	DZIP3	3	10	45	insT	3
574780156	CD200R1	3	4	24	insA	3
587083495	ZNF148	3	6	14	insG	3
591946859[a]	LOC644974	3	6	36	delC	3
595459401	TOPBP1	3	26	16	insT	3
608303870	PLSCR2	3	4	59	insG	3
611616867	C3orf16	3	5	22	delCT	3
611616868	C3orf16	3	5	21	delT	3
612474548	SELT	3	4	38	insT	4
631967149	PHC3	3	10	13	insT	3
648156647	DGKG	3	2	44	insC	3
652237997[a]	CLDN16	3	1	271	delG	2
658673909	PAK2	3	12	16	insT	3
660166378	ZNF595	4	4	17	insA	1
662098266[a]	POLN	4	23	65	delG	3
696277331	FLJ16686	4	3	53	delC	1
725715810	TMPRSS11F	4	7	30	insA	3
728128121	LOC100129410	4	3	13	insC	2
752860649	UNC5C	4	14	17	insT	3
767688857[a]	EGF	4	24	16	insC	1
779934903	KIAA1109	4	39	17	insT	3
782346647	ANKRD50	4	3	37	insT	3
827268418	NEK1	4	4	17	insT	3
841124368	CDKN2AIP	4	3	16	insA	3
853638502	KIAA0947	5	14	24	insT	3

Supplementary Table 5. Continued

Reference position	Gene	Chromosome	Coding sequence	Coverage	Allele change	Patient no.
889026155[a]	CARD6	5	3	32	insT	2
891727995	PAIP1	5	2	16	insG	2
901348948	MAP3K1	5	13	24	insC	3
909928153	ADAMTS6	5	3	23	insG	3
914509435	SERF1B	5	3	15	insG	3
914509482	SERF1B	5	3	49	insA	3
915509411[a]	GTF2H2	5	13	31	insT	2
919186632	HEXB	5	11	43	insA	3
922917852	SCAMP1	5	7	37	insA	1
928534269	EDIL3	5	7	25	insA	3
931867232	CCNH	5	7	26	insT	3
956697390	EPB41L4A	5	11	13	insT	1
966610635	ZNF474	5	1	39	delT	1
972596345	SLC12A2	5	8	35	insT	3
980635083[a]	SMAD5	5	6	105	insC	1
985314450[a]	LOC100288105	5	1	14	delC	4
985640033[a]	PCDHB9	5	1	32	insA	1
985844899[a]	PCDHGA8	5	1	27	delC	3
992330379	SCGB3A2	5	1	14	delA	3
994446878[a]	TIGD6	5	1	136	delT	1
994476149	HMGXB3	5	6	14	delA	3
998157358	GRIA1	5	11	19	insC	3
1020539380	FAM153B	5	4	23	insC	3
1039337531	C6orf114	6	1	30	insA	3
1052252527	BTN2A2	6	2	44	insG	3
1054107191[a]	ZNF187	6	1	33	insG	4
1056096293[a]	FLJ45422	6	2	18	insT	2
1057247419[a]	MICA	6	5	27	delG	3
1082305754	DST	6	45	18	insT	3
1088830738	EYS	6	6	19	insT	3
1093406718	COL19A1	6	5	16	insA	3
1113248546	MDN1	6	15	14	insC	3
1113280524	MDN1	6	2	40	insA	3
1131602437[a]	FOXO3	6	2	64	insG	3
1133380782	SLC22A16	6	4	20	insA	3
1135037748	C6orf225	6	1	17	delC	3
1153093327	SAMD3	6	7	14	delC	3
1154647636	LOC643854	6	1	26	insT	3
1154648098	LOC643854	6	1	20	insC	3
1159216432	BCLAF1	6	2	13	delT	2
1161156444	PBOV1	6	1	36	insG	3
1182019086	RSPH3	6	6	43	insA	3
1200500350	RSPH10B2	7	19	23	insG	3
1206053594	VWDE	7	19	26	insA	4
1221518316	TAX1BP1	7	13	14	insA	3
1222659823[a]	KIAA0644	7	1	90	delC	4
1222659922[a]	KIAA0644	7	1	26	insC	3
1226974977	BBS9	7	7	19	insT	1
1228643735[a]	DPY19L1	7	18	22	delAT	4
1228643736[a]	DPY19L1	7	18	50	delT	1
1262731853	TYW1B	7	8	142	delA	4
1262954278	TRIM74	7	2	24	insA	2
1265555276	TRIM73	7	2	84	insT	2
1266437261[a]	FLJ37078	7	14	43	insC	2
1266593512[a]	ZP3	7	8	51	insG	1
1266763110[a]	POMZP3	7	5	83	delA	4
1278946055	C7orf62	7	1	20	insC	3
1283360469	HEPACAM2	7	4	29	insT	3
1283589759	CALCR	7	9	26	insT	3
1290893801[a]	ZAN	7	30	28	insG	3
1291366094	MOGAT3	7	2	24	insA	3

Supplementary Table 5. Continued

Reference position	Gene	Chromosome	Coding sequence	Coverage	Allele change	Patient no.
1291722996[a]	EMID2	7	13	20	insG	4
1292538685	LOC100289561	7	1	14	insA	3
1295252926	MLL5	7	12	21	insG	3
1298402792	NRCAM	7	1	17	insT	3
1319055841[a]	KCP	7	10	62	insC	1
1319073009	KCP	7	1	30	delC	2
1333766379	LOC441294	7	1	46	insA	4
1334380185	CTAGE4	7	1	39	insA	3
1334381975	ARHGEF5L	7	1	19	insA	1
1339923632[a]	KRBA1	7	12	76	insC	2
1339973995[a]	SSPO	7	9	44	insC	1
1340003537[a]	SSPO	7	60	15	insC	4
1340012514	SSPO	7	76	23	delA	2
1340015859	SSPO	7	83	14	delC	2
1340525483	C7orf29	7	1	24	delC	1
1341211228[a]	ATG9B	7	10	49	insC	1
1341434558	SMARCD3	7	10	21	delC	3
1342197228	GALNTL5	7	5	71	delT	4
1342442397[a]	MLL3	7	14	208	insT	4
1356372261[a]	XKR5	8	6	55	delAG	1
1374409954[a]	NEFL	8	3	38	delG	4
1380219728[a]	UBXN8	8	7	83	insT	1
1380304215	TEX15	8	1	23	insA	3
1388426070[a]	PLEKHA2	8	11	28	delC	2
1395399601[a]	PRKDC	8	31	17	insG	1
1398930064	PXDNL	8	14	27	insA	3
1410692513[a]	YTHDF3	8	4	24	insG	1
1415952398	C8orf34	8	2	32	insG	3
1445261384	LAPTM4B	8	2	16	insC	3
1490189877[a]	JRK	8	1	12	delCA	3
1490189878[a]	JRK	8	1	19	delA	2
1491176363	ZNF623	8	1	29	insT	3
1492082552[a]	RECQL4	8	14	20	delG	3
1498992866	LOC645969	9	1	155	insT	4
1527437913[a]	C9orf144B	9	4	20	delC	4
1543290663[a]	FOXD4L5	9	1	39	delG	1
1546032104	TRPM3	9	22	19	insT	3
1552818643	VPS13A	9	48	29	insG	3
1574648314	COL15A1	9	13	29	insC	3
1586295095	MUSK	9	1	62	insT	3
1608846803[a]	ABO	9	6	117	insC	4
1620006324	GDI2	10	7	14	insG	3
1620254092	IL2RA	10	4	14	insC	3
1621795546[a]	ITIH5	10	14	23	delC	1
1633127998	NSUN6	10	2	26	insA	3
1647389817	ITGB1	10	13	22	insA	3
1652560241[a]	LOC340947	10	2	25	delT	1
1653671683[a]	LOC642424	10	3	117	delT	1
1657313101	AGAP4	10	7	23	delT	2
1658942495	FAM25G	10	3	48	insC	3
1662197526	LOC100287932	10	6	22	insA	4
1662338998[a]	AGAP6	10	1	50	insC	2
1666373362	PCDH15	10	19	56	insC	3
1673760921	TMEM26	10	6	25	insT	3
1685560504	FAM149B1	10	7	26	insT	3
1701949711	PANK1	10	3	30	insA	3
1708407108	CCNJ	10	3	18	insC	3
1708510568	ZNF518A	10	1	30	insC	1
1708668598	DNTT	10	2	18	insA	3
1709332414	C10orf12	10	1	18	insG	3
1728973932[a]	PNLIPRP2	10	3	52	insG	1

Supplementary Table 5. Continued

Reference position	Gene	Chromosome	Coding sequence	Coverage	Allele change	Patient no.
1733216553	BRWD2	10	8	28	insT	3
1737221823	ZRANB1	10	1	17	insA	3
1738045786	MMP21	10	7	33	insG	3
1748226082	C11orf21	11	4	63	insG	1
1750053615	RRM1	11	14	17	insT	3
1753342600	SYT9	11	4	18	insG	3
1760006709[a]	SPON1	11	5	72	insC	4
1764016157	SAAL1	11	7	40	delT	4
1771005317	LUZP2	11	12	34	insA	3
1782417168	TRAF6	11	6	24	delG	3
1782519665	RAG2	11	1	21	insA	3
1792247476[a]	CREB3L1	11	12	40	insG	1
1802120506	TCN1	11	7	13	insA	3
1802663567[a]	MS4A14	11	2	61	delTT	4
1802663568[a]	MS4A14	11	2	22	delT	3
1803663946[a]	TMEM216	11	3	54	insA	4
1804797590	AHNAK	11	3	12	insG	3
1805556025	SLC22A10	11	1	17	insC	3
1810263379[a]	UNC93B1	11	7	53	insG	3
1810284280[a]	ALDH3B1	11	2	63	insC	2
1810287509[a]	ALDH3B1	11	6	18	insC	1
1810293595[a]	ALDH3B1	11	9	28	insC	4
1814065554	LOC729523	11	1	22	delT	3
1826743977	DLG2	11	5	23	insT	3
1832107207	LOC642446	11	1	33	delT	4
1837197723[a]	CWC15	11	5	152	insT	1
1837299118[a]	SFRS2B	11	1	36	insC	4
1850549218	ATM	11	49	24	insT	3
1852355678	ZC3H12C	11	2	25	insC	3
1854201323[a]	DIXDC1	11	7	16	insC	1
1860877259[a]	TREH	11	15	28	insG	2
1861246651[a]	SLC37A4	11	3	37	delC	1
1861288156	VPS11	11	2	13	insC	4
1867800518	EI24	11	9	14	insC	4
1867851321	CHEK1	11	5	44	insC	3
1888645169[a]	PRB3	12	4	34	delG	4
1888731023[a]	PRB1	12	3	136	delC	1
1891856090	ATF7IP	12	11	19	insG	3
1893735417[a]	MGST1	12	2	12	delAA	3
1893735418[a]	MGST1	12	2	18	delA	3
1898574937	SLCO1B1	12	7	17	insC	3
1902256413	BCAT1	12	5	22	insG	3
1913975525	KIF21A	12	10	20	insT	3
1914378775	SLC2A13	12	10	17	insA	3
1927092176	KRT6C	12	1	15	insG	2
1930622534	SUOX	12	3	14	insG	3
1931678522	TMEM194A	12	9	23	insG	3
1932337710	OS9	12	12	17	insA	3
1959863488	LRRIQ1	12	26	12	delA	3
1962616654	C12orf50	12	3	28	insA	3
1978598568[a]	TDG	12	3	14	insA	3
1986789153	LOC100287839	12	9	35	insC	3
1997115077	RSRC2	12	10	28	insG	3
1999523126	UBC	12	1	29	delT	3
2009256491	ZMYM5	13	5	14	insC	3
2012756904	SACS	13	9	20	insT	3
2012761230	SACS	13	9	23	insT	3
2017859185	FLT1	13	4	36	insA	3
2022550582	STARD13	13	5	85	delT	1
2026525487	CSNK1A1L	13	1	13	insC	2
2038965626	RCBTB1	13	8	17	insG	3

Supplementary Table 5. Continued

Reference position	Gene	Chromosome	Coding sequence	Coverage	Allele change	Patient no.
2046563396	PRR20	13	2	28	delC	2
2063234131	KLF12	13	4	47	insT	3
2066482633	MYCBP2	13	75	17	insT	3
2066508358	MYCBP2	13	62	14	insC	3
2066632819	MYCBP2	13	22	23	delC	3
2066717508	MYCBP2	13	2	33	delAA	3
2066717509	MYCBP2	13	2	33	delA	3
2088603872	GPR18	13	1	20	insT	3
2105948032	NDRG2	14	1	34	delG	1
2106009532	FLJ10357	14	18	14	delG	3
2108927297	DHRS4L2	14	6	41	insA	4
2109139875[a]	MDP-1	14	6	13	delA	1
2117359342	AKAP6	14	1	20	insA	3
2117747539	AKAP6	14	12	21	insA	3
2137979011	DDHD1	14	10	22	insC	3
2148241015[a]	GPHB5	14	1	18	insG	4
2158414589[a]	C14orf169	14	1	19	insC	3
2159993929[a]	FAM164C	14	1	14	insA	1
2160606560	TTLL5	14	4	17	insA	3
2179419547	SERPINA12	14	2	54	insC	3
2179491154	SERPINA4	14	3	14	insG	3
2181450460	PAPOLA	14	5	33	insC	3
2202211427[a]	CHRFAM7A	15	4	191	delCA	1
2202211428[a]	CHRFAM7A	15	4	252	delA	4
2203996021[a]	CHRNA7	15	6	166	delTG	1
2203996022[a]	CHRNA7	15	6	50	delG	2
2204534873	SCG5	15	5	24	insC	3
2212460825	CASC5	15	10	14	insA	3
2220067652	SLC12A1	15	5	21	insA	3
2237036677	LOC100287371	15	3	32	insG	3
2243652079[a]	NR2E3	15	6	34	delC	1
2251295853	KIAA1024	15	1	14	insT	3
2252413491	ARNT2	15	14	24	insC	3
2256610001	ZSCAN2	15	2	14	insC	3
2257065252	PDE8A	15	4	20	delT	3
2261248094	FANCI	15	2	19	insC	3
2261584966	C15orf42	15	7	21	insT	3
2270957952[a]	LOC145814	15	4	23	insC	4
2271092254[a]	SYNM	15	1	19	insG	3
2274046312[a]	C16orf35	16	12	89	insG	4
2274304546	AXIN1	16	1	20	delC	3
2277509768[a]	NLRC3	16	7	81	delG	1
2285935013[a]	LOC729978	16	4	20	delAT	4
2285935014[a]	LOC729978	16	4	44	delT	1
2292434768	NOMO2	16	24	22	insG	3
2294397443	ACSM2A	16	9	23	delA	3
2294883814	DNAH3	16	53	18	insC	3
2304906998	HSD3B7	16	6	48	delC	2
2332868773	CLEC18C	16	3	24	insA	3
2333553555[a]	HYDIN	16	68	29	delA	4
2338969142[a]	CNTNAP4	16	1	82	insT	1
2351412465[a]	LOC100289580	16	2	67	delC	2
2354387432	PRPF8	17	4	11	insG	3
2356396572[a]	P2RX5	17	3	13	delG	1
2359357840	C17orf100	17	1	14	insG	2
2360272579[a]	SENP3	17	6	20	delA	4
2361527508[a]	PIK3R6	17	16	42	insG	1
2363416732	C17orf48	17	3	19	insA	3
2371198121	LGALS9C	17	9	16	insA	4
2376394518[a]	SEBOX	17	1	29	insG	4
2376430014[a]	SLC46A1	17	4	15	delA	1

Supplementary Table 5. Continued

Reference position	Gene	Chromosome	Coding sequence	Coverage	Allele change	Patient no.
2382300534	CCL7	17	2	24	insT	3
2383802642[a]	MMP28	17	4	28	insC	4
2384283858	TBC1D3C	17	13	31	insG	1
2388631071	KRT10	17	1	14	delC	3
2392844842[a]	PLCD3	17	10	24	insC	1
2393016586[a]	MAP3K14	17	4	16	insG	2
2407377091	CLTC	17	3	28	insT	3
2409792732	MED13	17	2	24	insA	3
2411313186[a]	WDR68	17	5	42	delG	1
2412151377	DDX5	17	8	23	insT	3
2434290839	MYOM1	18	8	16	insA	3
2448603454	RBBP8	18	14	22	insC	3
2451555232	LOC100287386	18	2	31	insA	1
2471234235	SLC14A2	18	4	39	delC	3
2492044315	CDH19	18	11	24	insA	3
2501962862	ZNF516	18	2	27	delG	2
2508173565[a]	SPPL2B	19	7	26	insC	2
2510788089	UHRF1	19	14	13	insC	3
2514792389	MUC16	19	3	18	insA	3
2514803399	MUC16	19	3	29	insT	3
2518236406	ZNF799	19	4	25	insA	3
2521463907[a]	CYP4F8	19	4	79	insC	1
2522001621[a]	HSH2D	19	5	71	delA	2
2538892348	C19orf55	19	9	17	delG	2
2543188059	ZNF780B	19	2	24	insC	3
2543756504[a]	LTBP4	19	24	14	insG	1
2544255517[a]	CYP2F1	19	1	53	insC	4
2544853028	CEACAM5	19	4	26	insT	3
2547650400[a]	CEACAM20	19	8	54	delT	1
2547930257[a]	CBLC	19	8	18	insC	3
2552076265[a]	DHDH	19	4	55	insG	2
2552600822	ALDH16A1	19	10	73	insC	2
2554469302[a]	LOC147645	19	10	37	insG	4
2555437083[a]	ZNF480	19	1	51	delG	1
2555750854	ZNF83	19	1	26	insG	3
2559350849	ZSCAN5C	19	1	50	insA	2
2560590155	ZNF749	19	3	34	insA	1
2560866399	ZNF671	19	4	14	insA	3
2561351770[a]	ZNF274	19	4	78	insG	2
2562070563	DEFB126	20	2	20	delCC	3
2562070564	DEFB126	20	2	20	delC	3
2567847490[a]	CHGB	20	4	28	delGA	2
2567847491[a]	CHGB	20	4	70	delA	2
2580083985	CSRP2BP	20	4	17	insG	3
2583130413[a]	NCRNA00153	20	7	49	insG	1
2606534089	DDX27	20	4	21	insA	3
2608270909	MOCS3	20	1	23	insG	3
2640900547[a]	KRTAP7-1	21	1	27	delA	4
2643647262[a]	SON	21	12	40	insA	1
2643647273[a]	SON	21	12	33	delA	4
2654166830	TRAPPC10	21	21	24	insT	3
2656193953[a]	LOC100288508	21	5	14	insC	1
2670991039	HORMAD2	22	2	23	insG	3
2681753989	DNAJB7	22	1	66	insA	1
2683020374	CYP2D6	22	5	18	insG	4
2701397266	WWC3	X	7	15	insG	3
2708217926	RBBP7	X	2	16	insC	3
2709924320	CDKL5	X	4	26	insC	3
2711314268	CXorf23	X	3	22	delG	3
2713575280	PHEX	X	19	29	insG	3
2736210225	KDM6A	X	17	17	insC	3

Supplementary Table 5. Continued

Reference position	Gene	Chromosome	Coding sequence	Coverage	Allele change	Patient no.
2736225869	KDM6A	X	24	16	insA	3
2737802282	SLC9A7	X	7	21	insC	3
2739406485	SSX1	X	6	45	insT	3
2739444455[a]	SSX9	X	2	11	delC	2
2741301867[a]	DGKK	X	22	55	insG	1
2743920170	SSX2B	X	6	35	insC	3
2745406637	WNK3	X	16	40	insA	3
2755460714	OPHN1	X	8	18	insC	3
2757668459	KIF4A	X	28	24	insG	3
2758547967	NONO	X	6	22	insT	3
2761842615	RLIM	X	3	18	insG	3
2771580011	HDX	X	5	18	insA	3
2779112744	PCDH11X	X	2	18	insT	3
2788398590	CENPI	X	20	19	insC	3
2789376503	TCEAL6	X	1	24	insG	2
2789554171	NXF2	X	7	14	insT	3
2789554906	NXF2	X	10	29	insA	3
2802179110	IL13RA2	X	4	18	insT	3
2823639589	ARHGEF6	X	18	16	insA	3
2840930428[a]	LCAP	X	1	57	insC	2
2841794077	MPP1	X	7	14	insG	3

[a]These indels commonly occurred in more than one HCC.

Supplementary Table 6. List of 81 Nucleotide Positions in 77 Genes With Indels at a Frequency of >20% of Reads in 4 Nontumorous Tissues From 4 Patients

Reference position	Gene	Chromosome	Coding sequence	Coverage	Allele change	Patient no.
36247083	THRAP3	1	4	18	delG	1
75174421	SLC44A5	1	16	19	delT	1
114430296	TRIM33	1	20	23	delC	1
133132499	YY1AP1	1	7	37	insT	1
133844355	RHBG	1	9	48	delC	4
201121914	CAPN2	1	3	14	delC	1
201173637	TP53BP2	1	13	22	delG	1
247664301	C2orf43	2	4	18	delA	2
319394229	SNRNP200	2	37	20	delA	1
322653290	AFF3	2	14	35	delA	1
331834049	RANBP2	2	20	16	delG	1
332901065	RGPD5	2	20	22	delT	1
374789589	NEB	2	4	37	insT	4
382950536	LY75	2	5	19	delA	4
401635744	TTN	2	274	25	delA	1
409835043	FAM171B	2	8	17	delT	1
412064501	COL3A1	2	14	21	insA	4
454784716	PTMA	2	4	14	delT	1
463724350[a]	AQP12B	2	1	27	delC	3
463734336	AQP12A	2	2	14	delG	2
503335742	DLEC1	3	4	20	delT	1
503335743	DLEC1	3	4	20	delA	1
735406533	CNOT6L	4	10	21	delG	1
785877214	LARP2	4	14	23	delA	1
798021293	SCOC	4	1	18	insC	1
810971969	TRIM2	4	5	18	delC	1
883256725	PRLR	5	3	29	delG	4
939146286	ANKRD32	5	16	15	insC	1
985314450[a]	LOC100288105	5	1	27	delC	3
1033746568	BMP6	6	5	24	delC	1
1068664364	KIAA0240	6	4	17	insT	4
1193244025	FAM120B	6	1	43	insA	1
1222659823	KIAA0644	7	1	463	delC	4
1282877394	CDK6	7	3	21	delA	1
1289880816	CYP3A4	7	12	45	delG	1
1333766765	LOC441294	7	1	13	delA	4
1340012514	SSPO	7	76	53	delA	3
1356372261	XKR5	8	6	130	delA	4
1490189877[a]	JRK	8	1	29	delC	3
1490189878[a]	JRK	8	1	15	delA	2
1492082552[a]	RECQL4	8	14	43	delG	3
1505961686	MPDZ	9	2	28	insG	4
1526015264	NFX1	9	3	36	delT	1
1573925487	GABBR2	9	17	59	insT	4
1580509237	ABCA1	9	4	24	insT	1
1637516682	ARMC3	10	18	19	delT	1
1637516683	ARMC3	10	18	19	delT	1
1657313100[a]	AGAP4	10	7	19	delT	2
1657313101[a]	AGAP4	10	7	14	delT	3
1807397040	SYVN1	11	7	15	insA	1
1832107207	LOC642446	11	1	18	delT	4
1855967268	ZW10	11	8	40	delC	1
1861246651	SLC37A4	11	3	143	delC	4
1884320486	ATN1	12	4	15	delA	1
1929584670	KIAA0748	12	6	31	delC	4
1955959709	PPFIA2	12	18	52	delA	4
1994379801	CIT	12	17	28	delG	1
1995110709	DYNLL1	12	2	14	delG	1
1997144069	KNTC1	12	2	31	delC	4
2105624179	RNASE4	14	1	15	delC	1

Supplementary Table 6. Continued

Reference position	Gene	Chromosome	Coding sequence	Coverage	Allele change	Patient no.
2162358340	C14orf133	14	13	15	delT	1
2243652079	NR2E3	15	6	129	delC	4
2256057594	ADAMTSL3	15	12	26	delT	1
2277509768[a]	NLRC3	16	7	12	delG	2
2302633200	EIF3C	16	4	18	delG	4
2303380808	SULT1A4	16	3	24	delA	1
2351412465	LOC100289580	16	2	103	delC	4
2356396572	P2RX5	17	3	40	delG	4
2376621991	SPAG5	17	3	13	delC	1
2386619109	CCDC49	17	5	14	delT	1
2413869089	APOH	17	5	24	delC	1
2501962862	ZNF516	18	2	29	delG	3
2507200605	MUM1	19	8	36	delG	1
2538892348	C19orf55	19	9	20	delG	2
2565046537	UBOX5	20	2	15	delG	1
2587599923	ZNF337	20	4	19	delT	1
2598525448	ZHX3	20	1	19	delT	1
2625038622	NRIP1	21	1	24	delG	1
2661518554	FAM108A5	22	2	13	delG	3
2748277559	SPIN2B	X	1	13	delG	2
2792445004	TEX13A	X	2	18	delC	3

[a]These indels commonly occurred in more than one HCC.

Supplementary Table 7. List of 40 Somatic Mutations With Amino Acid Changes Commonly Detected in Both the Tumor (at a Frequency of More Than 20% of Reads) and Matched Nontumorous Cirrhotic Liver (at a Frequency of More Than 5% of Reads) of the Same Patient

					Tumor		Nontumor	
Gene	Reference position	Chromosome	Reference nucleotide	Mutation nucleotide	Mutation frequency (%)	Patient no.	Mutation frequency (%)	Patient no.
LEPR	65548341	1	C	A	25.8	3	15.0	3
							21.9	1
ZNF408	1792629936	11	T	A	20.4	2	16.0	2
							15.8	4
HRNR	129676984	1	G	C	28.9	3	5.4	3
PXDN	228577682	2	G	C	45.1	4	47.2	4
POTEF	353150970	2	T	A	41.8	4	31.0	4
ALPP	455451136	2	C	T	32.5	4	37.5	4
GPR125	682521774	4	C	A	38.1	2	40.0	2
HERC6	746068457	4	T	A	36.5	4	44.9	4
EGFLAM	886579974	5	T	G	23.3	3	5.3	3
C4A	1057829599	6	T	G	25.0	2	11.5	2
WISP3	1134999625	6	T	G	43.3	4	64.3	4
C7orf10	1234451360	7	T	A	25.0	3	8.3	3
PVRIG	1290339880	7	C	T	23.5	1	21.3	1
MUC17	1291200140	7	G	A	21.2	4	12.5	4
PLOD3	1291376235	7	G	C	48.2	4	51.7	4
COL27A1	1589933932	9	A	G	56.8	4	54.6	4
AGAP9	1658906463	10	T	G	36.7	4	16.2	4
POLL	1713935693	10	G	T	44.8	4	38.6	4
MUC5AC	1747183167	11	G	A	43.9	4	43.8	4
MRGPRX3	1764064669	11	T	C	40.0	4	42.5	4
TMEM133	1843211533	11	A	C	59.5	4	83.3	4
TMEM123	1844621025	11	G	A	27.3	2	7.3	2
TMPRSS4	1860336319	11	C	T	54.4	4	41.3	4
DHRS4L2	2108914889	14	G	T	20.5	3	11.5	3
GOLGA6C	2247104814	15	A	T	21.7	4	9.6	4
PRSS22	2276813235	16	C	T	50.0	4	36.7	4
FAM38A	2351390771	16	C	T	21.4	4	54.3	4
GGT6	2357265990	17	G	A	92.3	4	41.7	4
COX10	2366897810	17	C	T	55.2	4	36.3	4
KIAA0100	2376657621	17	A	C	47.8	4	60.0	4
TBC1D3B	2384202011	17	C	T	63.0	4	27.4	4
TBC1D3D	2385938140	17	A	G	45.9	4	21.0	4
ERBB2	2387531879	17	A	G	66.7	4	54.6	4
CSH2	2411602334	17	C	T	90.9	4	79.5	4
QRICH2	2423941144	17	T	G	50.0	4	60.4	4
MOCOS	2461870479	18	T	C	72.0	4	62.2	4
CPAMD8	2522819358	19	G	A	21.8	3	15.2	3
MAP4K1	2541732174	19	G	A	36.0	4	54.3	4
PSG8	2545901763	19	C	A	28.3	3	9.5	3
KRTAP12-2	2654734983	21	C	T	59.3	4	43.5	4

NOTE. The first 2 genes listed were recurrently mutated in the nontumorous inflamed livers of 2 patients.

Supplementary Table 8. Overview of Selected Exome Sequencing Data From 22 Patients With HCV Infection

		Aligned reads	Aligned sequence (*base pairs*)	Median read depth
TP53	Tumor	29,334	2,035,570	1476.2
	Nontumor	31,848	2,200,641	1575.3
	Lymphocytes	36,690	2,539,944	1917.2
CTNNB1	Tumor	90,022	6,215,000	2344.3
	Nontumor	75,785	5,282,450	1991.2
	Lymphocytes	100,430	7,013,325	2710.8
LEPR	Tumor	34,328	2,390,335	538.3
	Nontumor	60,128	4,219,089	1025.6
	Lymphocytes	86,830	6,085,511	1423.0

NOTE. Selected exome sequencing of *TP53*, *CTNNB1*, and *LEPR* was performed for 22 nontumorous cirrhotic liver tissues, 10 HCC tissues, and matched peripheral lymphocytes from each patient. Aligned reads, aligned sequences (base pairs), and median read depth are shown for each sample.

Supplementary Table 9. Clinical Features and Overview of Deep Sequencing Data of Patients Who Underwent Deep Sequencing of the *LEPR* Gene

	Chronic hepatitis (n = 15)	Normal liver (n = 9)
Age (*y*)	59.3	55.9
Sex (male/female)	6/9	7/2
Aligned reads	4290	3956
Aligned sequence (*base pairs*)	1,044,737	1,275,068
Median read depth	2838	3440
No. of mutations in the *LEPR* gene	0	0

NOTE. We determined the sequences of the *LEPR* gene in the liver of 15 noncirrhotic patients with HCV-associated chronic hepatitis. In addition, normal liver tissues were obtained from 9 liver donors at the time of the operation. Age, sex, aligned reads, aligned sequences (base pairs), median read depth, and numbers of mutations are shown.

Supplementary Table 10. Mean Body Weights and Serum Levels of Insulin, Triglyceride, Total Cholesterol, and Alanine Aminotransferase of C57BL/KsJ-*db*/*db* (*db*/*db*) Mice and Misty (Control) Mice After 4 Weeks of Treatment With TAA

	db/*db*	Control
Body weight (*g*)	46.5 ± 0.6	23.5 ± 0.4
Insulin (*ng*/*mL*)	30.6 ± 28.3	1.6 ± 0.2
Triglyceride (*mg*/*dL*)	95.0 ± 5.0	50.0 ± 20.0
Total cholesterol (*mg*/*dL*)	215.0 ± 15.0	95.0 ± 15.0
Alanine aminotransferase (*IU*/*L*)	1325.0 ± 1085.0	75.0 ± 35.0

NOTE. All data are presented as mean ± SD.

Supplementary Table 11. Categorization of the Mutated Genes Detected by Whole Exome Sequencing of the AID-Expressing Hepatocyte Cell Line Using the Kyoto Encyclopedia of Genes and Genomes Database

Pathway				
Metabolic pathways	ATP6V0A4	DMGDH	HSD17B3	PGD
	ATP6V1C2	GALNT1	HYAL2	PHGDH
	BCMO1	GATM	NDST1	POLR3B
	CPS1	HKDC1	PAH	
PI3K-Akt signaling pathway	BCL2L11	IBSP NOS3	PRKCZ	TEK
	COL27A1			
MAPK signaling pathway	FLNB	SP1	CACNA1F	PTPN7
Cytokine-cytokine receptor interaction	LEPR	TNFRSF8	TNFRSF10A	
Transcriptional misregulation in cancer	EYA1	GZMB	JMJD7-PLA2G4B	
Proteoglycans in cancer	FLN	ITGB3	TIMP3	VTN
PPAR signaling pathway	CPT1B	CYP4A22	PPARD	
Cell cycle	E2F2	ESPL1	MCM7	
Pathways in cancer	FLT3	TRAF4	PDGFA	
Hedgehog signaling pathway	GLI3	LRP2	CSNK1A1L	
Others	95 genes			

NOTE. The genes categorized in multiple pathways are shown in only one representative pathway. Constitutive AID expression resulted in the accumulation of nucleotide alterations in various genes, including LEPR, of the cultured hepatocyte–derived cells. Whole exome sequencing was performed on DNA derived from established non-neoplastic human primary hepatocyte cells[6] with constitutive AID expression. AID expression in the cultured hepatocytes was performed using a lentiviral system.[5] After 8 weeks of AID expression, the DNA was extracted and subjected to whole exome sequencing as described in Materials and Methods. Overall, a total of 460 nucleotide positions in 380 different genes were defined as mutated in the AID-expressing cultured hepatocytes through the variant filtering process. Among them, pathway analyses by the Kyoto Encyclopedia of Genes and Genomes revealed that many genes, including LEPR, were categorized into well-known signaling pathways: the metabolic pathway, PI3K-Akt signaling pathway, MAPK signaling pathway, cytokine-cytokine receptor interaction pathway, and transcriptional misregulation in cancer pathway. Only categorized genes are shown.