A model of liver carcinogenesis originating from hepatic progenitor cells with accumulation of genetic alterations.

Soo Ki Kim*[1], Akihiro Nasu*[1], Junji Komori[2], Takahiro Shimizu[1], Yuko Matsumoto[1], Yasuko Minaki[1], Kenji Kohno[3], Kazuharu Shimizu[4], Shinji Uemoto[2], Tsutomu Chiba[1] and Hiroyuki Marusawa[1]

1. Department of Gastroenterology and Hepatology,
Graduate School of Medicine, Kyoto University, Kyoto, Japan
2. Department of Surgery,
Graduate School of Medicine, Kyoto University, Kyoto, Japan
3. Graduate School of Biological Sciences,
Nara Institute of Science and Technology, Nara, Japan
4. Department of Nanobio Drug Discovery,
Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, Japan

*These authors contributed equally.

Corresponding Author: Hiroyuki Marusawa

Department of Gastroenterology and Hepatology,

Graduate School of Medicine, Kyoto University, Kyoto, Japan

Address; 54 Kawaharacho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan.

E-mail address; maru@kuhp.kyoto-u.ac.jp

Phone+; +81-75-751-4319

Fax; +81-75-751-4303

Short Title

Liver carcinogenesis originating from hepatic progenitor cells.

Key Words:

liver cancer, hepatic progenitor cells, activation-induced cytidine deaminase (AID), mutation, liver carcinogenesis

List of Abbreviations:

AFP, alpha-fetoprotein; AID, activation-induced cytidine demainase; hHB-EGF, human heparin binding epidermal growth factor-like growth factor; CK, cytokeratin; DT, diphtheria toxin; GFP, green fluorescent protein; HCC, hepatocellular carcinoma; ICC, intrahepatic holangiocarcinoma; MAPK, mitogen-activated protein kinase; PPAR, peroxisome proliferator-activated receptor; PCR, polymerase chain reaction; SNV, single nucleotide variant; Tg, transgenic; TRECK, toxin-receptor mediated conditional cell knockout.

Brief Description:

Activation-induced cytidine deaminase (AID), a DNA mutator enzyme, contributes to inflammation-associated carcinogenesis through its mutagenic activity. In the present study, taking advantage of the ability of AID to induce stepwise genetic aberrations, we established a novel model showing accumulation of genetic alterations in fetal hepatic progenitor cells progressed to liver tumors, including both hepatocellular carcinoma and cholangiocarcinoma. We also revealed the overall landscape of genetic alterations accumulated during tumorigenesis by whole exome sequencing.

Financial Support: This work was supported by JSPS KAKENHI, Research program of the Project for Development of Innovative Research on Cancer Therapeutics (P-Direct) from Ministry of Education, Culture, Sports, Science and Technology of Japan, Health and Labour Sciences Research Grants for Research from the Ministry of Health, Labour and Welfare, and Takeda Science Foundation.

Conflict of Interest

The authors declare no conflict of interest.

Abstract

Activation-induced cytidine deaminase (AID) contributes to inflammation-associated carcinogenesis through its mutagenic activity. In the present study, by taking advantage of the ability of AID to induce genetic aberrations, we investigated whether liver cancer originates from hepatic stem/progenitor cells that accumulate stepwise genetic alterations. For this purpose, hepatic progenitor cells enriched from the fetal liver of AID transgenic (Tg) mice were transplanted into recipient “toxin-receptor mediated conditional cell knockout” (TRECK) mice, which have enhanced liver regeneration activity under the condition of diphtheria toxin treatment. Whole exome sequencing was used to determine the landscape of the accumulated genetic alterations in the transplanted progenitor cells during tumorigenesis. Liver tumors developed in 7 of 11 (63.6%) recipient TRECK mice receiving enriched hepatic progenitor cells from AID Tg mice, while no tumorigenesis was observed in TRECK mice receiving hepatic progenitor cells of wild-type mice. Histologic examination revealed that the tumors showed characteristics of hepatocellular carcinoma and partial features of cholangiocarcinoma with expression of the AID transgene. Whole exome sequencing revealed that several dozen genes acquired single nucleotide variants in tumor tissues originating from the transplanted hepatic progenitor cells of AID Tg mice. Microarray analyses revealed that the majority of the mutations (>80%) were present in actively transcribed genes in the liver-lineage cells. These findings provided the evidence suggesting that accumulation of genetic alterations in fetal hepatic progenitor cells progressed to liver cancers, and the selection of mutagenesis depends on active transcription in the liver-lineage cells.

Introduction

Tumorigenesis comprises multiple processes with a stepwise accumulation of genetic alterations that drive the progressive transformation of normal cells into highly malignant derivatives [1]. Recent studies of a large number of genomes in human cancer tissues clarified that cancer cells generally possess hundreds of somatic mutations and dysregulated gene expression profiles [2-4]. Although the origin of cancer cells remains mostly unsolved at present, it might be difficult for fully differentiated cells to acquire these large numbers of nucleotide alterations during their limited life span to achieve malignant transformation. In contrast, stem/progenitor cells have a long lifetime in order to supply the differentiated progenies in each organ. Thus, it appears reasonable to assume that long-lived tissue stem/progenitor cells can accumulate genetic alterations and hence could be the origin of tumor cells. Consistent with this hypothesis, a number of studies have provided evidence that the mutations would most likely result in expansion of the altered stem cells, perpetuating and increasing the chances of additional mutations, leading to malignant transformation [5-8].

Several studies have provided evidence that hepatocellular carcinoma (HCC) might originate from hepatic stem/progenitor cells [9-12]. A histologic study of clinical specimens also revealed that a substantial number of human HCC tissues have bipotential characteristics with coexpression of biliary and hepatocytic markers such as cytokeratin 7 (CK7), CK19, alpha-fetoprotein (AFP), and albumin [13, 14]. Conversely, all cholangiocarcinoma tissues examined showed hepatocellular differentiation in part of the tumor and expression of hepatic progenitor cell markers [15]. Findings from a recent study also suggested that human HCC could arise as a consequence of the dysregulated proliferation of hepatic progenitor cells when the TGF-β and IL-6 signaling pathway was disrupted [16].

Activation-induced cytidine deaminase (AID) can induce genetic alterations in human genome DNA sequences [17, 18]. Under physiological condition, AID is expressed almost exclusively in B lymphocytes, and plays a critical role not only in class switch recombination but also in somatic hypermutation of immunoglobulin genes. We recently demonstrated that inflammatory stimulation triggers aberrant AID expression in epithelial cells and initiates and/or promotes oncogenic pathways by inducing genetic alterations in various tumor-related genes [19, 20]. Indeed, AID expression is induced by proinflammatory cytokine stimulation and/or hepatitis C virus infection through NF-kB activation in hepatocytes [21], and the resultant AID upregulation leads to the accumulation of somatic mutations in *TP53* and *c-MYC* genes, both of which are frequently mutated in human cancer tissues [21, 22]. These findings suggest that aberrant AID production induced by chronic inflammation in the liver contributes to hepatocarcinogenesis via the accumulation of genetic aberrations in tumor-related genes [23].

The fact that it usually takes over a year for AID transgenic (Tg) mice to accumulate the genetic aberrations required for carcinogenesis [24, 25] prompted us to speculate that constitutive expression of AID in the cells with long life-span might possess the higher risk for malignant transformation compared to that in the cells with the limited life-span. Therefore, in the present study, we took advantage of the AID-mediated stepwise genotoxicity that recapitulates human hepatitis-associated carcinogenesis to investigate whether liver cancer originates from fetal hepatic progenitor cells with constitutive AID expression. Accordingly, we separated hepatic progenitor cells enriched from the fetal liver of AID Tg mice followed by transplantation into recipient mice and examined whether recipient mice receiving AID-expressing hepatic progenitor cells develop liver tumors. Furthermore, to unveil the overall landscape of genetic

alterations that accumulate in hepatic progenitor cells during the process of malignant transformation, we applied whole exome sequencing and determined the whole picture of genetic aberrations that accumulated in liver cancer cells originating from hepatic stem/progenitor cells.

Materials and Methods

Animals

The “toxin-receptor mediated conditional cell knockout” mice, which are homozygous for the albumin enhancer/promoter driven-human heparin binding epidermal growth factor-like growth factor (hHB-EGF) alleles, achieve the specific and conditional ablation of hepatocytes under the treatment of diphtheria toxin (DT) [26]. AID Tg mice were previously described [24]. All animals were maintained in a specific pathogen-free facility at the Kyoto University Faculty of Medicine. All animal experiments were approved by the ethics committee for animal experiments and performed under the Guidelines for Animal Experiments of Kyoto University.

Isolation of enriched hepatic progenitor cells, cell transplantation, and administration of diphtheria toxin

Hepatic progenitor cells were obtained from the fetal liver of pregnant wild-type, AID Tg , and green fluorescent protein (GFP) Tg mice on gestational day 13.5 and were enriched through sphere formation as previously described [27]. Briefly, after the digestion of fetal liver tissues using a 0.5% collagenase solution (Invitrogen, Carlsbad, CA), fetal liver cells were subjected to floating culture to form spheres in RPMI 1640 (Invitrogen) supplemented with 10% fetal calf serum. After 16 h incubation, the formed spheres were selected by gravity sedimentation and inoculated on type-I collagen-coated culture plates (Asahi Glass, Co., Ltd., Chiba, Japan). After 24 h of incubation, floating hematopoietic cells were removed by washing and adhered cells

were collected using trypsin-ethylenediaminetetraacetic acid (EDTA) solution (Sigma-Aldrich Co., Ltd., St. Louis, MO) for 3 min. The dissociated cells were counted and suspended in a Ca^{2+}-free Hank's balanced salt solution (Invitrogen) with fetal calf serum at a density of 5.0 x 10^6 cells/ml as the enriched hepatic progenitor cells. To characterize the enriched hepatic progenitor cells, expression levels of fetal liver stem/progenitor markers, including albumin, AFP, DLK1, CK19, and CD133, were examined using both immunohistochemistry and RT-PCR. In addition, the lack of expression of the hematopoietic cell marker CD45 in sphere-derived hepatic progenitor cells was also confirmed by both immunostaining and RT-PCR.

To achieve efficient engraftment of transplanted hepatic progenitor cells to livers of the recipient mice, we used TRECK mice as a liver-specific regeneration model [26]. TRECK mice express DT receptor under control of the albumin promoter, and treatment with DT selectively and efficiently ablates the hepatocytes, resulting in enhanced liver regeneration and efficient colonization of transplanted hepatic progenitor cells [27]. The enriched hepatic progenitor cells were transplanted into 7- to 9-wk old TRECK mice using an intrasplenic approach [27, 28]. We injected 0.2 ml of a cell suspension containing 1.0 x 10^6 hepatic progenitor cells. The DT was purified as described previously [26] and a total of 75 ng/kg DT was administered by intraperitoneal injection into recipient mice twice a week for 25 wk from the day of cell transplantation.

Whole exome capture and massively-parallel sequencing

Massively-parallel sequencing was performed using the Illumina Genome Analyzer IIx (Illumina, San Diego, CA) as described [29]. End-repair of DNA fragments, addition of adenine to the 3' ends of DNA fragments, adaptor ligation, and PCR amplification were performed according to the instructions. Exome capture was performed according to the NimbleGen Arrays Users Guide

(Roche, Basal, Switzerland). The DNA library was hybridized to the custom designed NimbleGen Seq Cap arrays targeting a total of 17,089 genes, including 157,728 exons. These libraries were enriched independently using a minimal PCR amplification step of 18 cycles with Phusion High-Fidelity DNA polymerase. The concentration of enriched DNAs were measured by Quant-iT PicoGreen Reagent and Kits (Invitrogen) to make a working concentration of 10 nM. Cluster generation and sequencing was performed for 76 cycles on the Illumina Genome Analyzer IIx as described using the pair-end protocol and collecting 76 bases from each read [29]. The obtained images were analyzed and base-called using GA pipeline software version 1.4 with the default settings provided by Illumina. All sequence reads were deposited in the DNA Data Bank of Japan Sequence Read Archive; accession number DRA000601.

RNA preparation and hybridization to the microarray

Total RNA was extracted from adult mice (12 wk old) liver tissues, bone marrow, and the fetal liver at day 13.5 of gestation using RNeasy Mini Kit (Qiagen, Valencia, CA). The details of the procedures for hybridization to the microarray were described previously [30]. RNA amplification and labeling were performed according to the manufacturer's instructions (Agilent Technologies, Palo Alto, CA). Array image acquisition and feature extraction were performed using an Agilent G2505C scanner with feature extraction software (Agilent Technologies). Microarray data were deposited in the GEO database; accession number GSE39213.

Genome Analyzer sequence data analysis and variant filtering.

Semiquantitative reverse transcription PCR and quantitative real-time genomic and reverse transcription-PCR

Histology and immunohistochemistry

Southern blot analysis

Statistical analysis

These procedures are described in Supplementary Materials and Methods [31-33].

Results

Enrichment of hepatic progenitor cells derived from fetal mouse liver

Enriched hepatic progenitor cells were obtained from the fetal liver of wild-type, AID Tg and GFP Tg mice through the formation of cell spheres, and the dissociated cells were cultured, counted, and then transplanted into recipient mice (Figure 1A). To characterize the sphere-derived hepatic cells used for the transplantation procedure, we first examined the expression of various marker genes in the fetal liver of wild-type mice. Immunohistochemistry revealed that expression of both the liver cell marker albumin and the hematopoietic cell marker CD45 were detectable in the fetal liver tissues (Figure 1B). Cells expressing DLK1, a cell surface marker for hepatic stem/progenitor cells, comprised ~10% of the total cells of the fetal liver parenchyma (Supplementary Figure 2A). The enriched cell population specifically contained cells expressing the hepatocyte-lineage cell markers such as albumin and AFP, but no expression of CD45 was detectable in these sphere-forming cells (Figure 1C). In addition, we

confirmed that almost the entire enriched sphere-derived cell population expressed E-cadherin and DLK1, and a subset of those enriched cells expressed CK19 and CD133 (Supplementary Figure 2B). On the other hand, the floating cells that did not form spheres strongly expressed CD45 (Figure 1D). RT-PCR also revealed that the sphere-forming cells prepared for the transplantation procedure expressed albumin, AFP, DLK1, CK19, and CD133 transcripts, but not CD45 (Figure 1E). Similar results were obtained in the fetal liver of AID Tg mice (data not shown). These expression profiles of the collected sphere-derived cells were consistent with those found in previous studies [34] and indicated that the enriched cells derived from the fetal liver fully contained hepatic lineage progenitor cells.

Efficient engraftment of transplanted hepatic progenitor cells in the recipient liver

To enhance engraftment of the transplanted cells in the liver, we used TRECK mice as a liver-specific regeneration model. These mice express hHB-EGF precursor, which functions as a DT receptor, under the control of an albumin promoter, and thus the hepatocytes of these mice are selectively ablated by the administration of DT [26]. We confirmed that the transcripts of hHB-EGF were specifically detectable in the liver of the TRECK mice (Figure 2A), and immunohistochemistry also revealed that hHB-EGF protein expression was present in the TRECK mouse liver tissues (Figure 2B). Serum alanine aminotransferase levels of a TRECK mouse were increased at 24 h after 75ng/kg of DT administration, peaked at 48 to 72 h, and subsequently returned to basal levels after 120 h (Figure 2C). After repeated trials, we found that twice-weekly DT administration maintained the sublethal liver injury, resulting in the constitutive hepatic regeneration process. Under these experimental conditions, DT-mediated ablation of hepatocytes in TRECK mice resulted in the expansion of cells expressing E-cadherin, EpCAM, and HNF4α accompanied by an increased number of the Ki67-positive cells,

suggesting enhanced proliferation activity of hepatocyte-lineage cells including hepatic progenitor cells and mature hepatocytes in the TRECK liver tissues (Figure 2D, and data not shown).

To examine the repopulation of the transplanted cells in the recipient liver, the hepatic progenitor cells of the GFP Tg fetal livers obtained in a similar way were introduced into the TRECK mice, followed by the repeated DT administration. At day 7, the GFP-positive cells were observed as clusters, and at day 30 the cluster of the GFP-positive cells was large enough to view macroscopically (Figure 2E). Moreover, the cluster of hepatocytes derived from the transplanted GFP-positive enriched hepatic progenitor cells was detectable in the recipient liver even 90 days after the transplantation while no such cells were observed in the liver of mice without DT administration (Figure 2F). These findings indicated that the transplanted cells efficiently engrafted and continued to proliferate in the recipient livers treated with DT as time progressed.

Transplanted hepatic progenitor cells with constitutive AID expression progressed to liver cancers.

Next, the enriched hepatic progenitor cells from AID Tg mice were transplanted into 13 recipient (TRECK) mice, and the DT was administered to the recipient mice for 25wk. Two mice died in a week after transplantation, while the remaining 11 mice were viable and thus subjected to phenotypic analyses. We found that liver tumors developed in 7 of 11 (63.6%) recipient mice that received the enriched hepatic progenitor cells of the AID Tg mice 90 wk after cell transplantation (Figure 3A). Among them, four mice developed multiple tumors and three developed a single large nodule. On the other hand, none of the 13 recipient mice receiving hepatic progenitor cells from wild-type or GFP Tg mice showed tumorigenesis during the same observation period, while

only one mouse developed a tumor with the characteristics of lipoma. Moreover, all the five recipient mice examined that received the mature hepatocytes of adult AID Tg mice at 6 months of age showed no phenotypic changes in the liver tissues. Histologic examination revealed that all the tumors examined showed the characteristics of well-to-moderately differentiated HCC. Interestingly, one tumor showed not only the enhanced AFP expression but also the ductal formation of tumor cells accompanied by the expression of CK19, indicating the features of intrahepatic cholangiocarcinoma (ICC) (Figure 3B, upper and middle panel). In addition, partial positivity for MUC1 immunostaining in the tumor indicated that the tumor contained the mucin-producing area (data not shown). On the other hand, no histologic changes were observed in the non-tumorous region of liver tissues receiving the AID-expressing hepatic progenitor cells (Figure 3B, lower panel).

To examine whether the cancers that developed in recipient mice liver were derived from the transplanted hepatic progenitor cells, we examined the expression of the AID Tg mice-specific transgene in three randomly selected tumors that developed in the recipient livers. Southern blotting analyses revealed strong signals of the AID transgene in the tumor tissues (Figure 3C). Weak signal of the AID transgene was also detected in the non-tumorous region, suggesting continuous engraftment of the transplanted hepatic progenitor-derived cells in the recipient mouse liver. In contrast, there were no detectable signals of the AID transgene in organs other than the liver of recipient mice, such as kidney, or in liver tissues of the TRECK mice without receiving the transplantation. Quantitative genomic PCR analyses also confirmed that all tumor tissues examined strongly expressed the AID transgene (Figure 3D). Moreover, the expression level of hHB-EGF in the tumor tissue was significantly lower than that in the surrounding non-tumorous liver tissue (Supplementary Figure 2C). These findings suggested that the

transplanted hepatic progenitor cells with constitutive AID expression achieved the malignant transformation and progressed to either HCC or cholangiocarcinoma.

Landscape of genetic alterations accumulated in the transplanted hepatic progenitor cells during the process of malignant transformation.

To unveil the landscape of genetic alterations that accumulated in the transplanted hepatic progenitor cells during the process of tumorigenesis, we determined the sequences of the whole exome in two independent liver cancers from two different recipient mice and the corresponding hepatic progenitor cells of the same AID Tg mice from which they originated (Table1). As a control, we also determined the whole exome sequences of the livers of their littermates with a wild-type phenotype. A total of 94.2% of the reads were properly aligned to the reference mouse genome and accordingly we obtained about 4.4 Gb of the aligned sequence data per sample on average after exome enrichment. 77.6% of the captured target exons were covered by 20X or more coverage depth read with a high quality genotype call. The variant filtering process is summarized in Supplementary Figure 1. We identified 24 (23 single nucleotide variants [SNVs] and 1 indel) and 162 (160 SNVs and 2 indels) somatic mutations in HCC#1 and HCC#2, of which the number of mutated genes with SNVs were 23 (HCC#1) and 105 (HCC#2), respectively (Table 2, and Supplementary Table 4). As shown in Supplementary Figure 2D, C/G to T/A substitution pattern was dominant, consistent with the previous finding that AID induces C/G to T/A transition into the genome [23, 24]. The candidate variants were then validated by conventional direct population Sanger sequencing (Supplementary Figure 3), and we finally confirmed that 20 (HCC#1) and 87 (HCC#2) SNVs were nonsynonymous variants. Among them, there were no genes commonly mutated in both tumors. Interestingly, 19 of 23 (82.6% in HCC#1) and 80 of 105 (76.2% in HCC #2) genes with SNVs were those reported in human liver

cancer tissues (International Cancer Genome Consortium; http://www.icgc.org/). Although tumor-suppressor Trp 53 gene also acquired mutations in both tumors, the nucleotide alteration rate was less than 20%. Pathway analyses using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (http://www.genome.jp/kegg/) revealed that 11 (HCC#1) and 66 (HCC#2) genes were categorized into the well-known signaling pathways, including peroxisome proliferator-activated receptor (PPAR) and mitogen-activated protein kinase (MAPK) signaling, and cell adhesion function (Table 3).

Although it is widely recognized that the mutational profiles of the tumor-related genes differ between different tissues, the mechanisms of those organ-specific differences in the mutated genes during the process of tumorigenesis remain unclear. We speculated that the genes that acquired mutations in HCC tissues might be preferentially and actively transcribed in hepatic lineage cells, because it has been shown that AID-induced mutagenic activity is directly proportional to the transcription levels of the target gene [35-37]. Therefore, we analyzed the gene expression profiles in the fetal and adult liver using microarray, and examined whether the mutated genes in HCC tissues were transcribed at relatively higher levels in liver-lineage cells compared with hematopoietic lineage cells. Among the mutated genes identified, transcription levels of 95.4% and 85.8% of the genes in HCC#1 and HCC#2, respectively, were higher in fetal and/or adult liver tissues than in bone marrow-derived hematopoietic cells (Table 2, and Supplementary Table 4), indicating that the genes actively transcribed in fetal and/or adult liver cells might have preferentially acquired the mutations through the genotoxic activity of AID. Consistently, quantitative RT-PCR analyses revealed that all the mutated genes analyzed were actively transcribed in adult liver tissues (Supplementary Figure 4). In contrast, representative genes that are actively transcribed in hematopoietic tissues [38], such as *Cd4*, *Cd5*, and *Tgfbr2*,

showed no mutations in liver tumors and less or no transcription in the liver compared with other organs (Supplementary Figure 4). We also confirmed that 19 (82.6% in HCC#1) and 93 (88.6% in HCC#2) of the mutated genes were actively transcribed in the liver tissues based on the mouse whole transcriptome analysis [39].Together, these findings suggest that the acquisition of mutations during hepatocarcinogenesis strongly depends on the transcription of target genes in the liver-lineage cells.

Discussion

Recently, recognition of the role of tissue stem/progenitor cells in the carcinogenesis process led to a new hypothesis that cancer arises from tissue stem/progenitor cells [40]. Indeed, genetically-engineered fetal progenitor cells lacking the tumor-suppressor gene function have been shown to play a role as the origin of liver cancer [9, 11, 41]. Whether the stepwise accumulation of genetic alterations on hepatic stem/progenitor cells contributes to the development of tumor cells, however, remains unknown. In the present study, we demonstrated that engrafted hepatic progenitor cells originated from the AID Tg mice progressed to liver tumors, including both HCC and cholangiocarcinoma, through the accumulation of somatic mutations in a variety of target genes.

Several previous studies demonstrated that the transplanted putative fetal liver stem/progenitor cells are capable of repopulating the liver that encounter extensive liver injury favoring the proliferation and survival of transplanted hepatocytes [42-44]. The DT receptor has been identified as a membrane-anchored form of the HB-EGF precursor [26]. Recently, it was shown that transplanted hepatic progenitor cells derived from the fetal liver were efficiently engrafted and repopulated in the liver of recipient HB-EGF-expressing mice with DT stimulation [27, 28]. Using

this model, efficient engraftment of the transplanted cells in recipient mice with HB-EGF expression in the liver enabled us to examine the fate of transplanted hepatic progenitor cells with constitutive AID expression. Notably, liver tumors with histologic features of human HCC developed in the recipient mice that received the hepatic progenitor cells derived from the AID Tg mice, while no tumorigenesis was observed in the recipient mice transplanted with hepatic progenitor cells of control mice. The findings that the tumors contained the AID transgene indicated that these tumors were derived from the transplanted hepatic progenitor cells accompanied with the AID-induced genetic aberrations. Interestingly, one of those tumors showed both the characteristics of HCC and cholangiocarcinoma in a single nodule, suggesting that the hepatic progenitor cells with the accumulation of genetic aberration could possess the potential to progress both HCC- and cholangiocarcinoma- lineage tumor cells. Alternatively, it might be possible that AID-mediated genetic alterations contribute to modifying the differentiation status of tumor cells, leading to either HCC or bile duct cancers from common progenitor cells.

Sequencing of whole genomes, whole exomes, and whole transcriptomes of cancer samples has recently become feasible using deep sequencing technologies. In this study, to obtain the overall picture of genetic alterations accumulated in the hepatic progenitor cells of the AID Tg mice that achieved malignant transformation, we performed whole exome sequencing of the transplanted progenitor cells and the resultant tumor tissues, and unveiled the landscape of genetic alterations that accumulated during tumorigenesis. We found that various genetic aberrations, mainly SNVs, were highly accumulated in the tumors, further supporting the putative involvement of aberrant AID activity in the development of HCC. One thing to be noted is that approximately 80% of mutated genes detected in the liver cancer tissues developed in the recipient mice have been

reported to be mutated in human HCC tissues (International Cancer Genome Consortium; http://www.icgc.org/), although it is not possible to draw a definitive conclusion from analyses of the limited number of HCCs that developed in the recipient mice. Functional annotation analyses revealed that many of the genes that acquired genetic aberrations are categorized into several important signaling pathways, including those involved in the regulation of cell proliferation, cell metabolism, and cell adhesion. Thus, it could be suggested the stepwise dysregulation of cell function caused by the accumulation of genetic aberrations in hepatic progenitor cells appears to play a pivotal role in the development of tumor cells.

We previously revealed that genetic changes induced by the genotoxic activity of AID show organ-specific profiles and suggested the possibility that the target preference of AID-induced mutagenesis contributes to the diversity of tissue-specific oncogenic pathways [23]. One possible explanation for the target selection for mutagenesis is that AID preferentially induces mutations in the actively-transcribed genes in each cell, because AID likely induces somatic mutations on the single-strand DNA exposed during the transcription process [35-37]. Consistent with this hypothesis, we confirmed in this study that the majority of genes with SNVs were the actively transcribed genes in liver lineage cells. However, we also observed that the transcription level of the gene is not solely responsible for the acquisition of AID-mediated genotoxicity, because one of the most actively transcribed hepatotrophic genes, albumin, did not accumulate SNVs in liver tumor cells (data not shown). Consistently, extensive sequencing of various genes in B lymphocytes revealed that only 25% of the transcribed genes accumulated SNVs in an AID-dependent manner [45]. Mutational hotspots preferentially attacked by AID genotoxicity frequently possess unique sequence characteristics, so-called RGYW/WRCY motifs (where W=A or T, R=A or G, and Y=C or T) in transcribed targets [46]. Moreover, a recent study reported

clusters of various types of repeat sequences in the vicinity of cleaved sites in AID target genes [47]. Thus, target selection of AID-mediated mutagenesis might require both active transcription and sequence characteristics of the genes.

In conclusion, the findings in the present study suggested that mutagenic activity of AID might contribute to the malignant transformation of hepatic progenitor cells to liver cancer cells via the induction of genetic alterations. Some of the actively transcribed genes in the liver-lineage cells preferentially accumulated SNVs and might contribute to the development of tumor cells. However, based on the model used in the present study, we could not fully determine whether the developed tumors derived directly from the fetal hepatic progenitor cells or via mature hepatocytes, because the transplanted fetal progenitor cells differentiated into mature hepatocytes in the recipient liver [27]. Moreover, the truly significant driver mutations responsible for hepatocarcinogenesis remain unclear. Thus, further elucidation of the precise step of the AID-induced accumulation of genetic aberrations will be required to identify the genetic alterations that possess the key to the carcinogenesis process. In addition, the fractionation by fluorescence-activated cell sorting would be essential to identify the subset of hepatic stem/progenitor cells that play a role in the origin of tumor cells.

Acknowledgements

We are grateful for the support of Research Fellowships of the Japan Society for the Promotion of Science (JSPS) for Young Scientists. We thank Dr. I. Ikai, Dr. T. Machimoto, and Dr. R. Kamimura for helpful suggestion and Dr. K. Terasawa, Dr. Y. Fujii, Ms. K. Matsubara and Ms. C. Kakimoto for help with the analyses.

References

1. Hanahan D, Weinberg RA. The Hallmarks of Cancer. Cell 2000;100:57-70.

2. Thorgeirsson SS, Grisham JW. Molecular pathogenesis of human hepatocellular carcinoma. . Nat Genet 2002;31:339-46.

3. Hussain SP, Schwank J, Staib F, Wang XW, Harris CC. TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer. Oncogene 2007;26:2166-76.

4. Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, Calderaro J, Bioulac-Sage P, Letexier M, Degos F, Clement B, Balabaud C, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet 2012;44:694-8.

5. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature 2001;414:105-11.

6. Pardal R, Clarke MF, Morrison SJ. Applying the principles of stem-cell biology to cancer. Nat Rev Cancer 2003;3:895-902.

7. Burkert J, Wright NA, Alison MR. Stem cells and cancer: an intimate relationship. J Pathol 2006;209:287-97.

8. Mishra L, Banker T, Murray J, Byers S, Thenappan A, He AR, Shetty K, Johnson L, Reddy EP. Liver stem cells and hepatocellular carcinoma. Hepatology 2009;49:318-29.

9. Dumble ML, Croager EJ, Yeoh GCT, Quail EA. Generation and characterization of p53 null transformed hepatic progenitor cells: oval cells give rise to hepatocellular carcinoma. Carcinogenesis 2002;23:435-45.

10. Zender L, Spector MS, Xue W, Flemming P, Cordon-Cardo C, Silke J, Fan ST, Luk JM, Wigler M, Hannon GJ, Mu D, Lucito R, et al. Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. Cell 2006;125:1253-67.

11. Xue W, Zender L, Miething C, Dickins RA, Hernando E, Krizhanovsky V, Cordon-Cardo C, Lowe SW. Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. Nature 2007;445:656-60.

12. Yamashita T, Ji J, Budhu A, Forgues M, Yang W, Wang HY, Jia H, Ye Q, Qin LX, Wauthier E, Reid LM, Minato H, et al. EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features. Gastroenterology 2009;136:1012-24.

13. Durnez A, Verslype C, Nevens F, Fevery J, Aerts R, Pirenne J, Lesaffre E, Libbrecht L, Desmet V, Roskams T. The clinicopathological and prognostic relevance of cytokeratin 7 and 19 expression in hepatocellular carcinoma. A possible progenitor cell origin. Histopathology 2006;49:138-51.

14. Roskams T. Liver stem cells and their implication in hepatocellular and cholangiocarcinoma. Oncogene 2006;25:3818-22.

15. Komuta M, Spee B, Vander Borght S, De Vos R, Verslype C, Aerts R, Yano H, Suzuki T, Matsuda M, Fujii H, Desmet VJ, Kojiro M, et al. Clinicopathological study on cholangiolocellular carcinoma suggesting hepatic progenitor cell origin. Hepatology 2008;47:1544-56.

16. Tang Y, Kitisin K, Jogunoori W, Li C, Deng CX, Mueller SC, Ressom HW, Rashid A, He AR, Mendelson JS, Jessup JM, Shetty K, et al. Progenitor/stem cells give rise to liver cancer due to aberrant TGF-beta and IL-6 signaling. Proc Natl Acad Sci U S A 2008;105:2445-50.

17. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 2000;102:553-63.

18. Honjo T, Kinoshita K, Muramatsu M. Molecular mechanism of class switch recombination: linkage with somatic hypermutation. Annu Rev Immunol 2002;20:165-96.

19. Matsumoto Y, Marusawa H, Kinoshita K, Endo Y, Kou T, Morisawa T, Azuma T, Okazaki IM, Honjo T, Chiba T. Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat Med 2007;13:470-6.

20. Endo Y, Marusawa H, Kou T, Nakase H, Fujii S, Fujimori T, Kinoshita K, Honjo T, Chiba T. Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers. Gastroenterology 2008;135:889-98.

21. Endo Y, Marusawa H, Kinoshita K, Morisawa T, Sakurai T, Okazaki IM, Watashi K, Shimotohno K, Honjo T, Chiba T. Expression of activation-induced cytidine deaminase in human hepatocytes via NF-kappaB signaling. Oncogene 2007;26:5587-95.

22. Kou T, Marusawa H, Kinoshita K, Endo Y, Okazaki IM, Ueda Y, Kodama Y, Haga H, Ikai I, Chiba T. Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis. Int J Cancer 2007;120:469-76.

23. Chiba T, Marusawa H, Ushijima T. Inflammation-associated cancer development in digestive organs: mechanisms and roles for genetic and epigenetic modulation. Gastroenterology 2012;143:550-63.

24. Okazaki IM, Hiai H, Kakazu N, Yamada S, Muramatsu M, Kinoshita K, Honjo T. Constitutive expression of AID leads to tumorigenesis. J Exp Med 2003;197:1173-81.

25. Morisawa T, Marusawa H, Ueda Y, Iwai A, Okazaki IM, Honjo T, Chiba T. Organ-specific profiles of genetic changes in cancers caused by activation-induced cytidine deaminase expression. Int J cancer 2008;123:2735-40.

26. Saito M, Iwawaki T, Taya C, Yonekawa H, Noda M, Inui Y, Mekada E, Kimata Y, Tsuru A, Kohno K. Diphtheria toxin receptor-mediated conditional and targeted cell ablation in transgenic mice. Nat Biotechnol 2001;19:746-50.

27. Machimoto T, Yasuchika K, Komori J, Ishii T, Kamo N, Shimoda M, Konishi S, Saito M, Kohno K, Uemoto S, Ikai I. Improvement of the survival rate by fetal liver cell transplantation in a mice lethal liver failure model. Transplantation 2007;84:1233-9.

28. Ishii T, Yasuchika K, Machimoto T, Kamo N, Komori J, Konishi S, Suemori H, Nakatsuji N, Saito M, Kohno K, Uemoto S, Ikai I. Transplantation of embryonic stem cell-derived endodermal cells into mice with induced lethal liver damage. Stem cells 2007;25:3252-60.

29. Nasu A, Marusawa H, Ueda Y, Nishijima N, Takahashi K, Osaki Y, Yamashita Y, Inokuma T, Tamada T, Fujiwara T, Sato F, Shimizu K, et al. Genetic Heterogeneity of Hepatitis C Virus in Association with Antiviral Therapy Determined by Ultra-Deep Sequencing. PLoS ONE 2011;6:e24907.

30. Fujiwara M, Marusawa H, Wang HQ, Iwai A, Ikeuchi K, Imai Y, Kataoka A, Nukina N, Takahashi R, Chiba T. Parkin as a tumor suppressor gene for hepatocellular carcinoma. Oncogene 2008;27:6002-11.

31. Ishii T, Yasuchika K, Fujii H, Hoppo T, Baba S, Naito M, Machimoto T, Kamo N, Suemori H, Nakatsuji N, Ikai I. In vitro differentiation and maturation of mouse embryonic stem cells into hepatocytes. Exp Cell Res 2005;309:68-77.

32. Nakatani T, Mizuhara E, Minaki Y, Sakamoto Y, Ono Y. Helt, a novel basic-helix-loop-helix transcriptional repressor expressed in the developing central nervous system. J Biol Chem 2004;279:16356-67.

33. Toda Y, Kono K, Abiru H, Kokuryo K, Endo M, Yaegashi H, Fukumoto M. Application of tyramide signal amplification system to immunohistochemistry: a potent method to localize antigens that are not detectable by ordinary method. Pathol Int 1999;49:479-83.

34. Yasuchika K, Hirose T, Fujii H, Oe S, Hasegawa K, Fujikawa T, Azuma H, Yamaoka Y. Establishment of a Highly Efficient Gene Transfer System for Mouse Fetal Hepatic Progenitor Cells. Hepatology 2002;36:1488-97.

35. Yoshikawa K, Okazaki IM, Eto T, Kinoshita K, Muramatsu M, Nagaoka H, Honjo T. AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. Science 2002;296:2033-6.

36. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature 2003;422:726-30.

37. Ramiro AR, Stavropoulos P, Jankovic M, Nussenzweig MC. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. Nat Immunol 2003;4:452-6.

38. Kotani A, Okazaki IM, Muramatsu M, Kinoshita K, Begum NA, Nakajima T, Saito H, Honjo T. A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. Proc Natl Acad Sci U S A 2005;102:4506-11.

39. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5:621-8.

40. Sell S, Leffert HL. Liver cancer stem cells. J Clin Oncol 2008;26:2800-5.

41. Katz SF, Lechel A, Obenauf AC, Begus-Nahrmann Y, Kraus JM, Hoffmann EM, Duda J, Eshraghi P, Hartmann D, Liss B, Schirmacher P, Kestler HA, et al. Disruption of Trp53 in livers of mice induces formation of carcinomas with bilineal differentiation. Gastroenterology 2012;142:1229-39.

42. Nierhoff D, Ogawa A, Oertel M, Chen YQ, Shafritz DA. Purification and characterization of mouse fetal liver epithelial cells with high in vivo repopulation capacity. Hepatology 2005;42:130-9.

43. Oertel M, Menthena A, Chen YQ, Teisner B, Jensen CH, Shafritz DA. Purification of fetal liver stem/progenitor cells containing all the repopulation potential for normal adult rat liver. Gastroenterology 2008;134:823-32.

44. Duncan AW, Dorrell C, Grompe M. Stem cells and liver regeneration. Gastroenterology 2009;137:466-81.

45. Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, Schatz DG. Two levels of protection for the B cell genome during somatic hypermutation. Nature 2008;451:841-5.

46. Rogozin IB, Pavlov YI, Bebenek K, Matsuda T, Kunkel Ta. Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. Nat Immunol 2001;2:530-36.

47. Kato L, Begum NA, Burroughs AM, Doi T, Kawai J, Daub CO, Kawaguchi T, Matsuda F, Hayashizaki Y, Honjo T. Nonimmunoglobulin target loci of activation-induced cytidine deaminase (AID) share unique features with immunoglobulin genes. Proc Natl Acad Sci U S A 2012;109:2479-84.

Figure legends

Figure 1. Enrichment of hepatic stem/progenitor cells from the fetal liver.

(A) Schematic diagram showing the transplantation of the enriched hepatic stem/progenitor cells of AID Tg mice or control (CTR) mice into the recipient TRECK mice. DT was administered intraperitoneally twice a week to recipient (TRECK) mice for 25 wk from the day of cell transplantation. The phenotypes were examined 90wk after transplantation. (B) Microscopic image (H&E staining) of the fetal liver tissues. Immunohistochemical staining for both the liver cell marker albumin and the hematopoietic cell marker CD45 are shown. (C) Immunohistochemical staining of the enriched cell population from the fetal liver via sphere formation for albumin, AFP, and CD45. (D) Immunohistochemical staining of floating cells that did not form spheres for CD45. (E) Representative RT-PCR for the various phenotypic expression: albumin, AFP, DLK1, CK19, CD133, CD45 and control Actb (β-actin). Total RNA was extracted from the spheres of the enriched cell population from the fetal liver, adult liver tissue, bone marrow, and fetal liver tissue.

Figure 2. Efficient engraftment of the transplanted hepatic progenitor cells in the recipient liver.

(A) The hHB-EGF expression in the liver, kidney, and spleen of a TRECK mouse and the liver of a wild-type mouse determined by semiquantitative RT-PCR analysis. Upper, hHB-EGF expression; Lower, control Actb expression. (B) Immunohistochemical staining for human hHB-EGF in the liver of the TRECK and wild-type mice. Upper, hHB-EGF immunofluorescence; Lower, DAPI staining. (C) Time-course changes in ALT values of the TRECK and wild-type mice after the first DT administration. Vertical bars show SD. (D) Immunostaining analysis of liver tissue specimens of a TRECK mouse with (DT [+]) or without

(DT [-]) DT administration. Upper, H&E staining; Middle, E-cadherin immunofluorescence; Lower, Ki-67 immunofluorescence. (E) Macroscopic image of a representative liver receiving GFP-positive hepatic progenitor cells at 30 days after transplantation. (F) Histologic analysis of liver tissue specimens receiving GFP-positive hepatic progenitor cells at 90 days after transplantation. Upper, H&E staining; Middle, GFP immunofluorescence; Lower, DAPI staining.

Figure 3. Development of tumors in livers receiving hepatic progenitor cells from AID Tg mice.

(A) Macroscopic images of tumors that developed in recipient mice receiving progenitor cells from AID Tg mice. (B) Microscopic images of a liver tumor that developed in a recipient mouse receiving hepatic progenitor cells from an AID Tg mouse. Upper, AFP-positive part; Middle, CK19-positive part; Lower, non-tumorous liver tissue. Immunohistochemical staining for H&E, AFP, and CK19 are shown. (C) Southern blot analysis for the AID transgene. DNA was extracted from three liver tumor tissues (Tumor #1, 2, 3), a non-tumor liver tissue (Non-tumor), the kidney of the corresponding animal, a liver of a TRECK mouse (Negative control; NC), and a liver of an AID Tg mouse (Positive control; PC), followed by the amplification and hybridization to the probe specific for the AID transgene. (D) Results of quantitative genomic PCR for AID transgene in three liver tumor tissues, a non-tumor liver tissue, and the kidney of the corresponding animal. N.D. means not detected.

Figure 1. Kim et al.

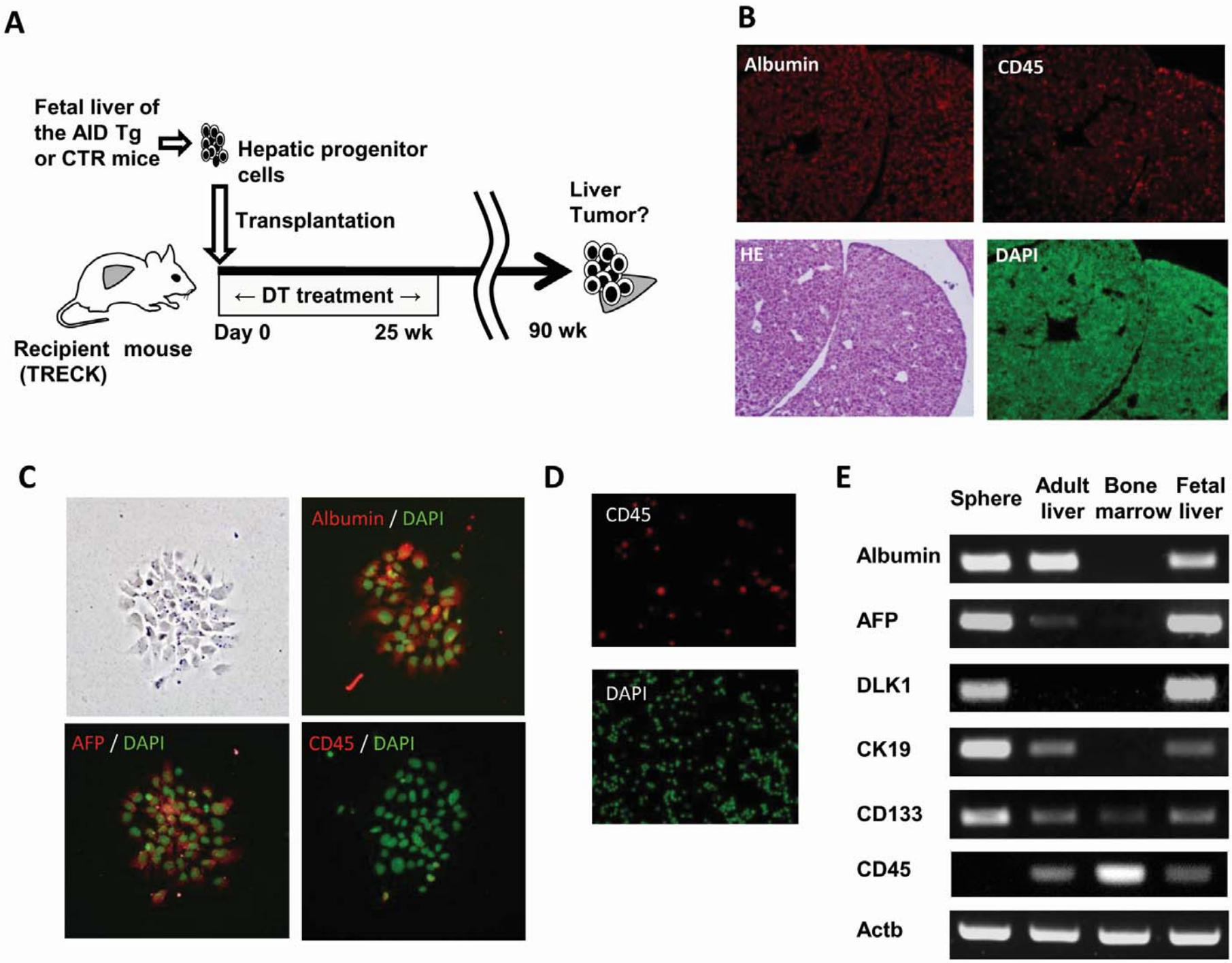


Figure 2. Kim et al.

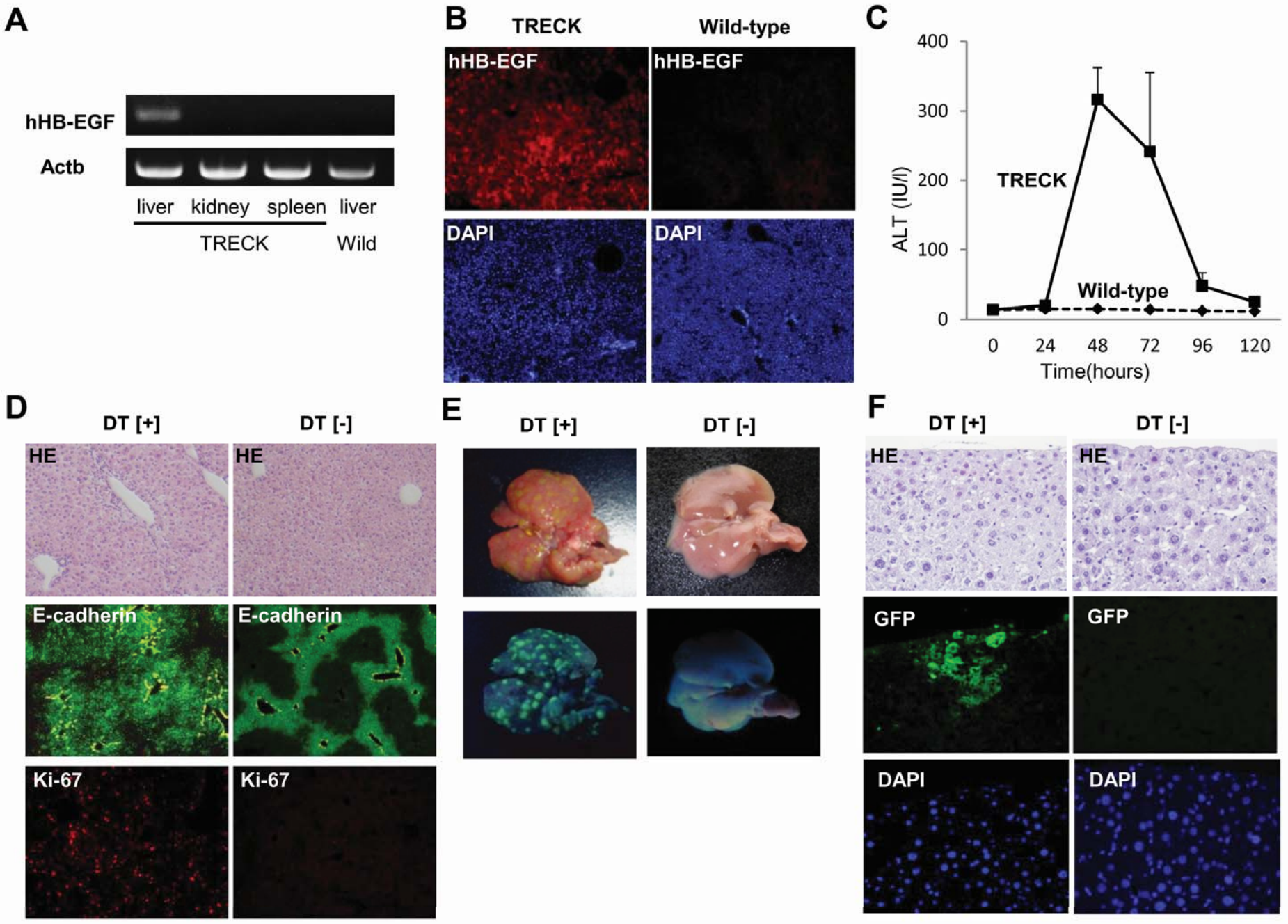
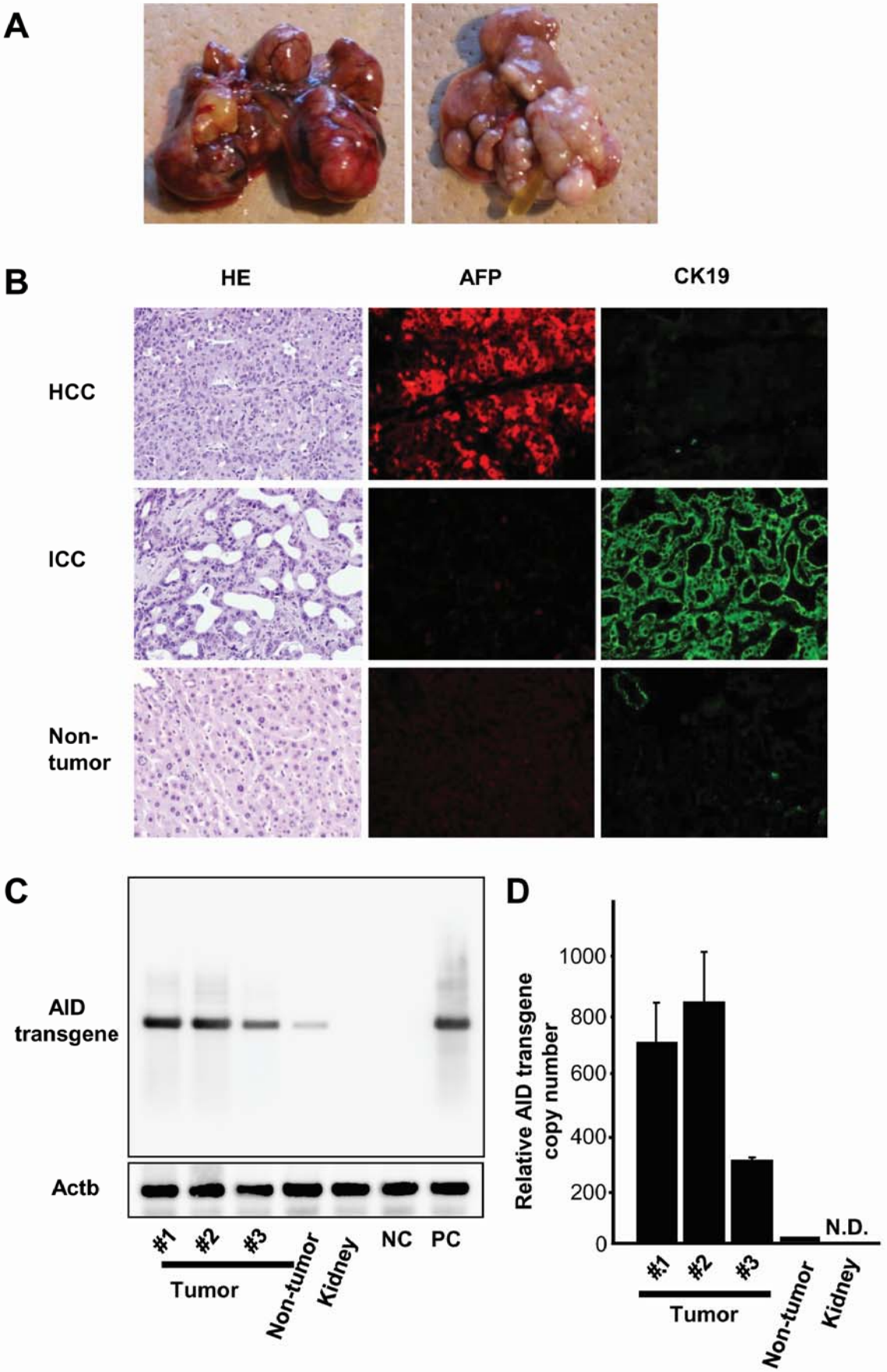


Figure 3. Kim et al.



Supplementary Materials and Methods

Genome Analyzer sequence data analysis and variant filtering.

Using the high performance alignment software "NextGENe v2.1" (SoftGenetics, State College, PA), the 76 base-pair reads obtained from the Genome Analyzer IIx were aligned with the reference sequences of Mus musculus whole genome derived from the NCBI build 37.1 annotation. Reads with 96% or more bases matching a particular position of the reference sequences were aligned. Furthermore, two quality filters were used for sequencing reads: reads with a median quality value score of more than 20 and no more than 3 uncalled nucleotides were allowed anywhere in the 76 bases. Only sequences that passed the quality filters, rather than raw sequences, 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 developed by SoftGenetics to provide an empirical estimation of the likelihood that a given mutation is real and not an artifact of sequencing or alignment errors. This score is based on the concept of Phred scores, where quality scores are logarithmically linked to error probabilities. The Overall Mutation score is calculated according to the following equation: Overall Mutation score = (Coverage score) x (four optional scores). The four optional scores are the Read balance score, Allele balance score, Mismatch score, and Wrong Allele score. These scores are described in "Next GENe v2.1" in detail. The candidates of somatic mutations were selected according to the variant filtering process (Supplementary Figure 1). As we further required that common variants in mice be excluded, the mutations in liver tumors must not be found in more than 5% of the reads in the liver of wild-type C57BL/6 mouse. In addition, the

mutations in liver tumors should not be observed in the original corresponding hepatic progenitor cells of the AID Tg mice, therefore nucleotide alterations could not appear in more than 5% of reads in the originating cells. Candidate single nucleotide variants 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 HCCs and the absence of each in the liver of wild-type mice.

Semiquantitative reverse transcription PCR and quantitative real-time genomic and reverse transcription-PCR

Total RNA was extracted from the tissues using RNeasy Mini Kit (Qiagen). cDNA was synthesized using Transcriptor First-Strand cDNA Synthesis Kit (Roche). The oligonucleotide primers for the semiquantitative reverse transcription-PCR (RT-PCR) are shown in Supplementary Table 1. Quantification of gene copy numbers or gene expression was performed by quantitative real-time genomic PCR or RT–PCR using LightCycler 480 System II (Roche). The oligonucleotide primers for the quantitative real-time genomic PCR and RT-PCR are shown in Supplementary Table 2. To assess the quantity of isolated DNA, target DNAs were normalized to the DNA levels of the housekeeping reference gene *Actb*. Similarly, to assess the quality of isolated RNA as well as the efficiency of cDNA synthesis, target cDNAs were normalized to the endogenous mRNA levels of the housekeeping reference gene *18S rRNA*. For simplicity, ratios are presented as relative values compared with expression levels in lysate from control specimens.

Histology and immunohistochemistry

The details of the immunohistochemistry procedures were described previously [31-33]. The primary and secondary antibodies used for immunostaining are shown in Supplementary Table 3.

Southern blot analysis

Southern blot analysis was performed using AlkPhos Direct Labelling Reagents (GE Healthcare, Buckinghamshire, England), with DNA probes labeled using alkaline phosphatase using the following primer and probe sets: AID transgene SB sense; GGACAGCCTTCTGATGAAGC, AID transgene SB antisense; TGGCATATGTTGCCAAACTC, AID transgene probe sense; GGACAGCCTTCTGATGAAGC, AID transgene probe antisense; GAAGTTGTCTGGTTAGCCGG, Actb SB sense; TGTACGTAGCCATCCAG, Actb SB antisense; CCTTCACCGTTCCAGT, Actb probe sense; TGAGCTGCCTGACGG, Actb probe antisense; GCCACCGATCCACACA.

Statistical analysis

Statistical differences in the gene expression levels were analyzed using the Kruskal-Wallis H-test with Bonferroni correction.

References

31. Ishii T, Yasuchika K, Fujii H, Hoppo T, Baba S, Naito M, Machimoto T, Kamo N, Suemori H, Nakatsuji N, Ikai I. In vitro differentiation and maturation of mouse embryonic stem cells into hepatocytes. Exp Cell Res 2005;309:68-77.

32. Nakatani T, Mizuhara E, Minaki Y, Sakamoto Y, Ono Y. Helt, a novel basic-helix-loop-helix transcriptional repressor expressed in the developing central nervous system. J Biol Chem 2004;279:16356-67.

33. Toda Y, Kono K, Abiru H, Kokuryo K, Endo M, Yaegashi H, Fukumoto M. Application of tyramide signal amplification system to immunohistochemistry: a potent method to localize antigens that are not detectable by ordinary method. Pathol Int 1999;49:479-83.

Supplementary Figure legends

Supplementary Fig 1.

Variant filtering process for somatic mutations determined by whole exome sequencing.

Supplementary Figure 2.

(A) Immunohistochemical staining for liver tissue derived from ED13.5 fetal mice DAPI staining, DLK1 fluorescence and merge.

(B) Immunohisotochemical staining of the enriched cell population from the fetal liver via sphere formation for E-cadherin, DLK1, CK19 and CD133.

(C) hHB-EGF expression of liver tumor and non-tumor liver tissue. Total RNA was extracted from liver tumor tissue (Tumor), non-tumor liver tissue (Non-tumor), kidney of the corresponding mouse (Kidney), liver of an AID Tg mouse (Negative control; NC), and liver of a TRECK mouse (Positive control; PC), and subjected to semiquantitative RT-PCR analysis for hHB-EGF expression.

(D) Frequency of indels and nucleotide substitution patterns of HCCs.

Supplementary Figure 3.

Representative Sanger sequencing trace files of *Nob1*, *Pck1* and *Mc3r*.

Comparisons of the sequenced regions between the liver of wild-type mouse (Upper) and HCC samples from TRECK mice transplanted with hepatic progenitor cells of the AID Tg mice (Lower) are shown. Specific mutations are observed in the trace images of lower panels and highlighted by the arrows.

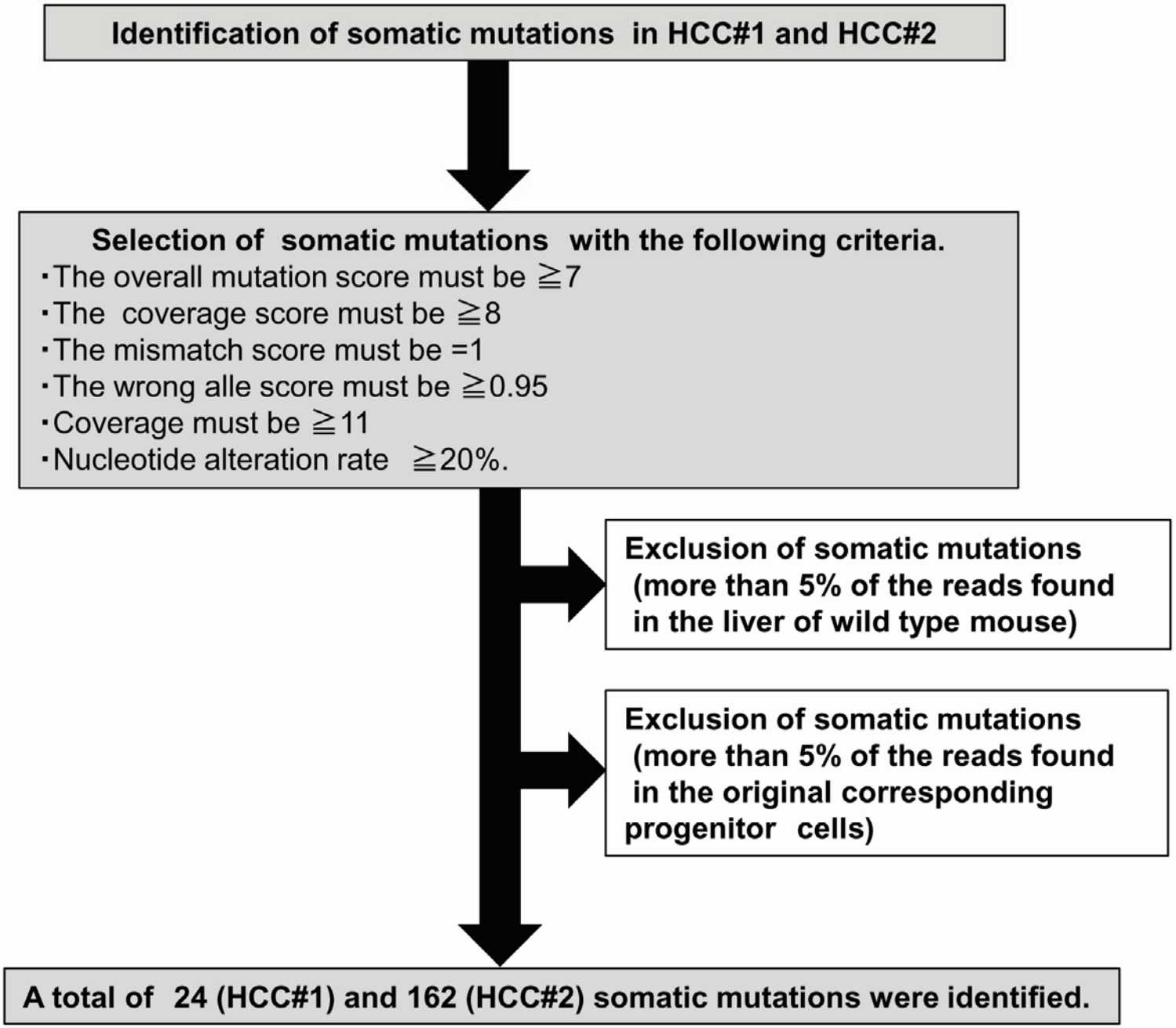
c.: cDNA. p: protein.

Supplementary Figure 4.

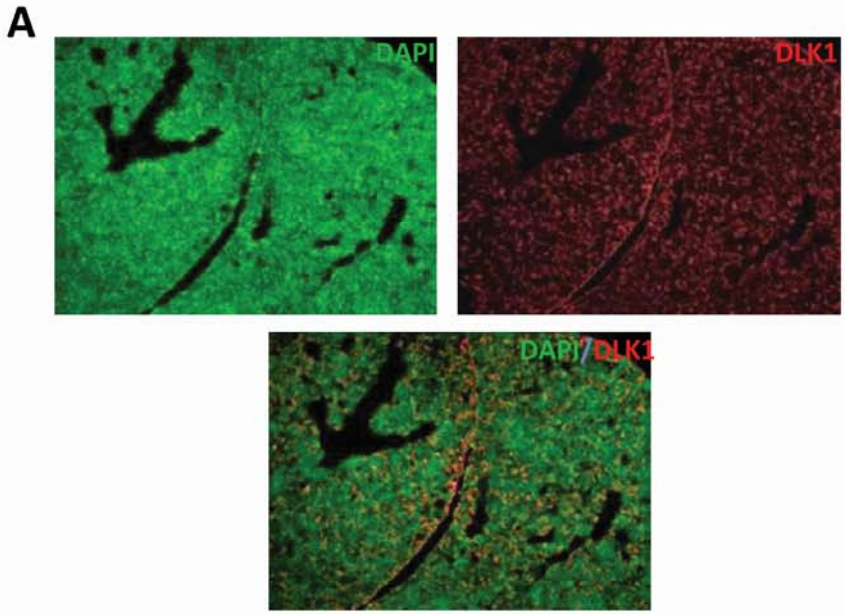
Expression profiles of the representative genes in various tissues of wild-type mice.

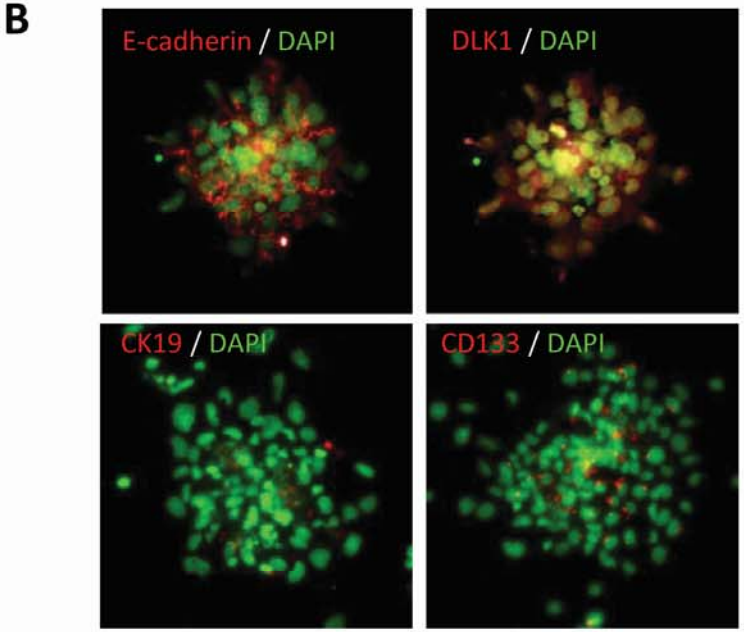
Total RNA was extracted from the liver, kidney, lung, brain and spleen of the wild-type mice and subjected to quantitative RT-PCR. The expression levels of the mutated (A) and non-mutated (B) genes in the liver cancer tissues are shown. Values shown in the graphs were normalized relative to the expression of the spleen (mean±SE; n=3).

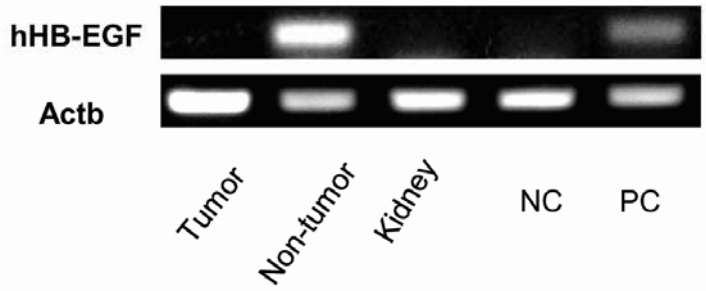
Supplementary Figure 1. Kim et al.

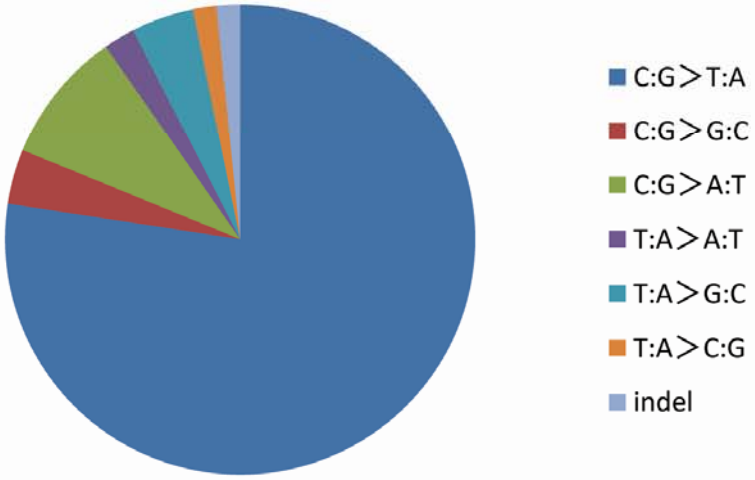


Supplementary Figure 2. Kim et al.

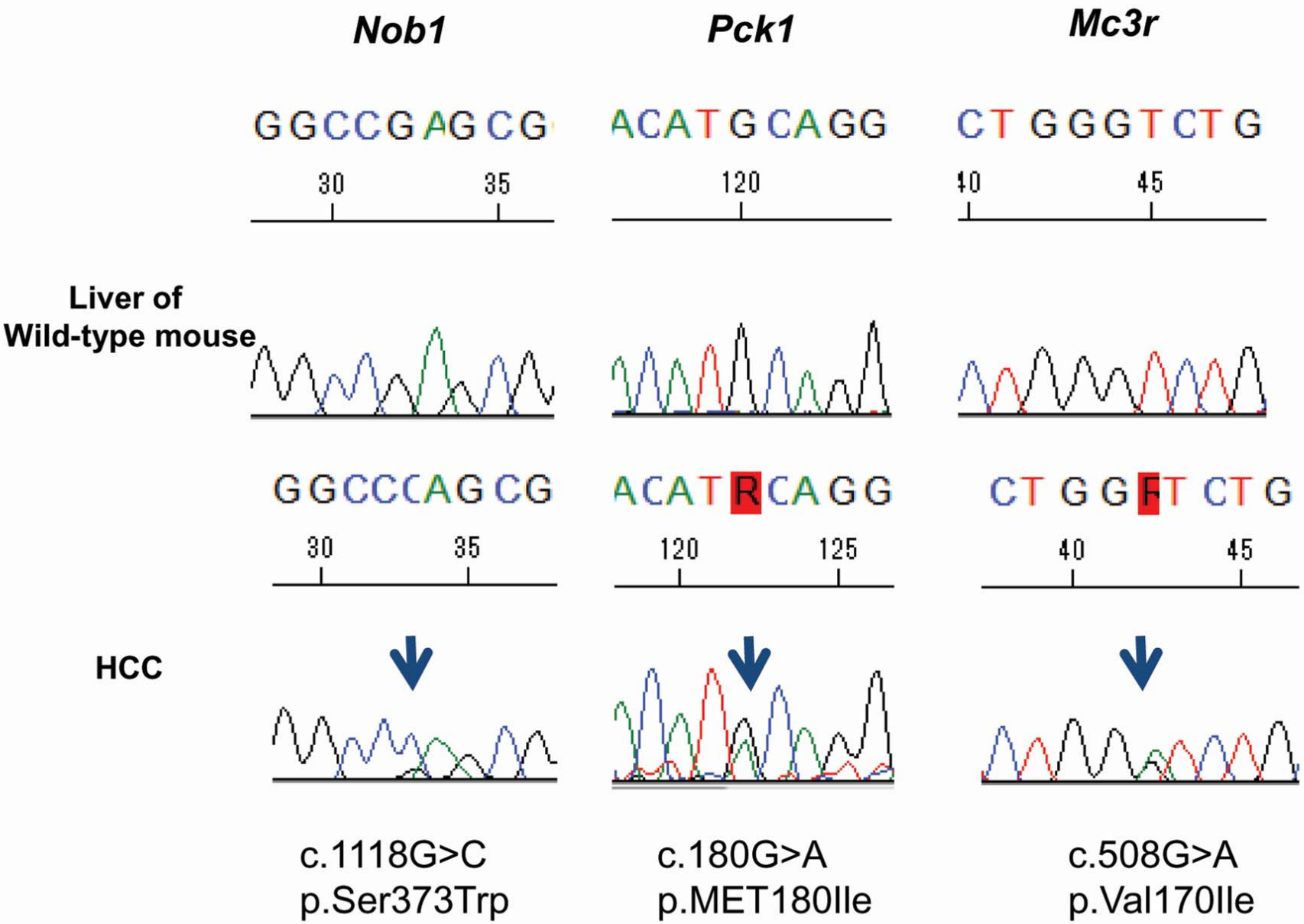




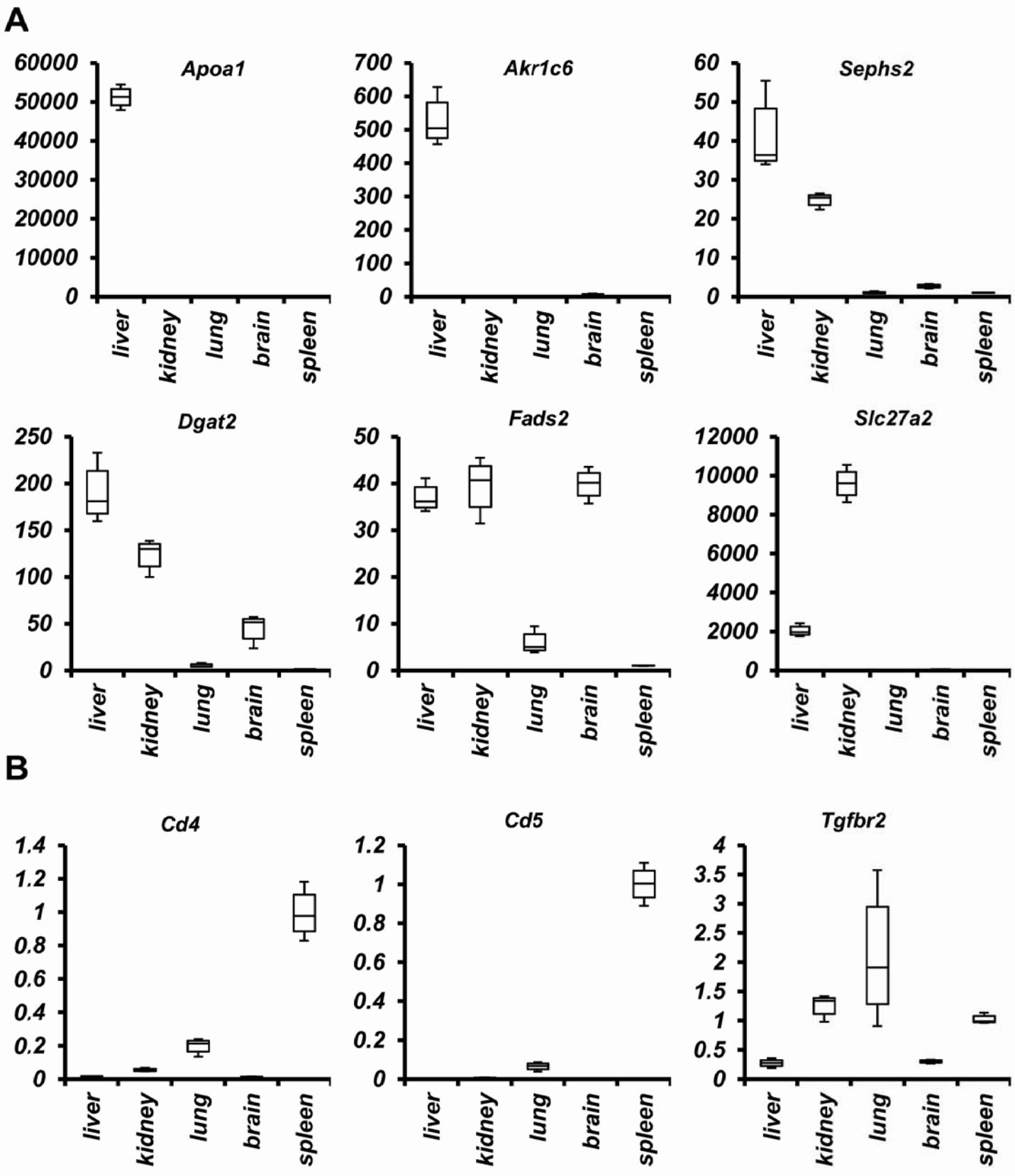




Supplementary Figure 3. Kim et al.



Supplementary Figure 4. Kim et al.



Supplementary Table 1.

Primer sequences used for semiquantitative RT-PCR amplification.

	forward primer	reverse primer
albumin	5’-GACAAGGAAAGCTGCCTGAC-3’	5’-TTCTGCAAAGTCAGCATTGG-3’
AFP	5’-AGCAAAGCTGCGCTCTCTAC-3’	5’-GAGTTCACAGGGCTTGCTTC-3’
DLK1	5’-TGTGACCCCCAGTATGGATT-3’	5’-CCTTGCAGACTCCATTGACA-3’
CK19	5’-CTCGGATTGAGGAGCTGAAC-3’	5’-TCACGCTCTGGATCTGTGAC-3’
CD133	5’-GAAAAGTTGCTCTGCGAACC-3’	5’-TCTCAAGCTGAAAAGCAGCA-3’
CD45	5’-GTGCCTTGTTCAATCTCTTGG-3’	5’-CAGTTAGCATCCTGCTTGCC-3’
hHB-EGF	5’-AGTCCGTGACTTGCAAGAGG-3’	5’-GTCCCTCTTCTTCCCTAGCC-3’
Actb	5’-GTGGGCCGCTCTAGGCACCAA-3’	5’-CTCTTTGATGTCACGCACGAT-3’

Supplementary Table 2.

Primer sequences used for quantitative genomic PCR and RT-PCR amplification.

	forward primer	reverse primer
AID	5'-CGTGGTGAAGAGGAGAGATAGTG-3'	5'-CAGTCTGAGATGTAGCGTAGGAA-3'
Actb	5’-GTGTGATGGTGGGAATGGGT-3’	5’-CTGGGTCATCTTTTCACGGTT-3’
Sephs2	5’-AAGAGACGGTGCAGGAAGG-3’	5’-AGTCCATGCCAATGCTCAG-3’
Dgat2	5’-AGTGGCAATGCTATCATCATCGT-3’	5’-AAGGAATAAGTGGGAACCAGATCA-3’
Fads2	5’-TCCCTTTCTACGGCATCTTG-3’	5’-TGTGACCCACACAAACCAGT-3’
Slc27a2	5’-TCAAAGTCCCCAAAGGTGAG-3’	5’-TTTCCTCCAGCATAGCCAAT-3’
Cd4	5’-GCAGCATGGCAAAGGTGTAT-3’	5’-GATGATGAGAGGAAACGATCC-3’
Cd5	5’-CCAGAGAACGACACCTCCAC-3’	5’-ACCACACCTGTGCACCTCA-3’
Tgfbr2	5’-AGGACCATCCATCCACTGAA-3’	5’-GGACAGTCTCACATCGCAAA-3’
18S rRNA	5’-TAGAGTGTTCAAAGCAGGCCC-3’	5’-CCAACAAAATAGAACCGCGGT-3’

Supplementary Table 3.
Antibodies used for immunostaining.

Primary antibodies

Antigen	Species	Dilution	Supplier
AFP	Rbt	1/200	DAKO
Albumin	Rbt	1/500	Nordic
CD45	rat	1/500	R&D
CD133	rat	1/250	eBioscience
CK19	goat	1/100	Santa Cruz
DLK1	Rbt	1/150	Abcam
E-cadherin	rat	1/100	Takara
GFP	chicken	1/1000	Abcam
HB-EGF	goat	1/50	R&D
Ki67	Rbt	1/1000	Novocastra

Secondary antibodies

Conjugate	Antigen	Species	Dilution	Supplier
Alexa Fluor 488	Chicken IgG	Goat	1/200	Molecular Probes
Alexa Fluor 488	Rabbit IgG	Goat	1/200	Molecular Probes
Dylight 488	Goat IgG	Donkey	1/100	Jackson
FITC	Rat IgG	Donkey	1/100	Jackson
Dylight 549	Rabbit IgG	Donkey	1/200	Jackson
Dylight 549	Rat IgG	Donkey	1/200	Jackson
Cy3	Goat IgG	Donkey	1/100	Jackson

Supplementary Table 4.

List of somatic mutations identified in HCC#1 and HCC#2;

Gene expression profiles obtained by microarray analysis on the adult liver (AL), fetal liver (FL), bone marrow (BM) of wild-type mouse are also shown.

HCC#1

Gene	Mutation type	#NS or S	Chr.	Position	Nucleotide change	Amino Acid change	*AL/BM	**FL/BM
Alox12b	SNV	NS	11	68980425	G>GT	505V>LV	2.38	1.46
Ccrl1	SNV	NS	9	104001408	G>GT	226T>NT	1.98	1.80
Drd3	SNV	NS	16	43762622	A>AT	90E>EV	2.26	1.39
Erbb2	SNV	NS	11	98281423	A>AC	32K>KT	2.12	5.41
Etv3	SNV	S	3	87340528	A>AC	499G>GG	1.35	0.68
Gpr149	SNV	NS	3	62407876	C>AC	208V>FV	2.44	1.43
Kcnk12	SNV	S	17	88145738	C>AC	278L>LL	0.30	0.19
Mapk8ip3	SNV	NS	17	25040950	G>GT	648Y>XY	2.28	1.36
Mast2	SNV	NS	4	115985621	A>AC	686V>VG	2.34	1.44
Mettl13	SNV	NS	1	164474500	T>AT	310E>VE	–	–
Myo1d	SNV	NS	11	80506554	T>GT	44E>ED	2.28	1.45
Nrap	SNV	NS	19	56452652	G>GT	331Y>YXY	2.47	1.22
Olfr30	indel	NS	11	58269229	delA;A>AG	Frameshift	(2.56)	(1.55)
Olfr979	SNV	S	9	39808465	A>AC	115T>TT	0.74	0.79
Prkar1b	SNV	NS	5	139584476	T>GT	103E>ED	2.34	1.41
Rnf31	SNV	NS	14	56220168	T>TG	882F>LFV	1.20	1.13
Slc9a9	SNV	S	9	94710357	C>CT	182Y>YY	2.39	1.45
Sorbs1	SNV	NS	19	40439589	G>GT	177P>HP	2.93	1.53
Stk36	SNV	NS	1	74670485	A>AG	581T>TA	2.41	1.62
Tbc1d5	SNV	NS	17	50881420	T>AT	655K>NK	1.67	1.09
Tox3	SNV	NS	8	92782013	G>GT	145Y>XY	2.16	1.51
Ttrap	SNV	NS	13	24923642	C>CA	15P>TPA	1.13	1.20
Uhrf1	SNV	S	17	56452509	G>GT	300R>RR	0.01	2.19
Vcpip1	SNV	NS	1	9715251	T>TA	992R>XRR	3.01	4.01

HCC #2

Gene	Mutation type	#NS or S	Chr.	Position	Nucleotide change	Amino Acid change	*AL/BM	**FL/BM
Aacs	SNV	S	5	125956363	G>AG	6R>RR	6.42	1.04
Aacs	SNV	S	5	125956417	G>AG	24K>KK	6.42	1.04
Actn1	SNV	NS	12	81361038	G>AG	14Q>XQ	0.28	0.27
Adamts3	SNV	NS	5	90204375	C>CT	199V>VI	2.42	1.47
Ahsg	SNV	NS	16	22892223	C>CT	13L>LF	7602.86	1829.35
Ahsg	SNV	NS	16	22892301	C>CT	39Q>QX	7602.86	1829.35
Ahsg	SNV	S	16	22892225	C>AC	13L>LL	7602.86	1829.35
Akr1c6	SNV	NS	13	4435630	G>AG	64A>TA	5080.98	3.94
Alb	SNV	NS	5	90889997	C>CT	13S>SF	15204.99	2731.27
Alb	SNV	NS	5	90891774	C>CT	63H>HY	15204.99	2731.27
Apoa1	SNV	NS	9	46037935	G>GA	82Q>QHQ	1518.68	158.09
Apoa1	SNV	S	9	46037938	G>AG	83L>LL	1518.68	158.09
Apoa2	SNV	S	1	173155489	C>CT	15S>SS	6878.26	504.23
Apoe	SNV	NS	7	20281829	C>CG	279V>VL	12.09	1.25
Apoe	SNV	NS	7	20282221	G>AG	148S>FS	12.09	1.25
Apoe	SNV	S	7	20282958	C>CT	27Q>QQ	12.09	1.25
Apof	SNV	S	10	127706366	G>AG	111Q>QQ	1273.68	53.42
Apof	SNV	S	10	127706663	G>AG	210K>KK	1273.68	53.42
Apon	SNV	S	10	127691969	C>CT	160N>NN	611.10	1.48
Arrdc2	SNV	NS	8	73363073	C>CT	76S>SN	1.96	3.04
Aurkaip1	SNV	NS	4	155206603	C>CT	67P>PS	1.39	1.39
Car3	SNV	S	3	14864331	G>AG	39K>KK	2182.34	9.80
Cbx4	SNV	S	11	118947307	C>CT	8E>EE	0.80	0.20
Cd163	SNV	S	6	124261711	T>CT	361G>GG	1.88	0.08
Cebpa	SNV	S	7	35905297	G>AG	287R>RR	12.28	0.80
Chka	SNV	S	19	3852137	C>CT	21S>SS	5.08	2.47
Creb3l2	SNV	NS	6	37391783	C>CT	6S>SN	2.50	1.92
Creb3l2	SNV	S	6	37391785	C>CT	5E>EE	2.50	1.92
Csnk1g1	SNV	S	9	65806460	C>CG	38L>LL	0.30	0.73
Cspp1	SNV	NS	1	10124184	C>AC	1086L>ML	1.05	1.58
Cyfip2	SNV	NS	11	46036116	G>AG	968S>FS	0.10	0.06
Cyp2e1	SNV	NS	7	147950701	G>AG	61A>TA	2536.61	1.43
Dcn	SNV	S	10	96957666	C>AC	76P>PP	2.45	1.45

Dgat2	SNV	NS	7	106330954	C>CT	23S>SN	22.46	1.78
Dnajb11	SNV	S	16	22858150	C>CT	13L>LL	1.93	1.58
Dusp4	SNV	S	8	35870934	G>AG	51L>LL	0.20	0.73
Egr1	SNV	NS	18	35021266	G>AG	42S>NS	5.12	7.25
Egr1	SNV	NS	18	35022195	C>CT	126P>PS	5.12	7.25
Egr1	SNV	NS	18	35022756	C>CT	313L>LF	5.12	7.25
Egr1	SNV	S	18	35021216	C>CT	25H>HH	5.12	7.25
Egr1	SNV	S	18	35021181	C>CT	14L>LL	5.12	7.25
Epha1	SNV	NS	6	42315843	G>CG	257L>VL	92.18	16.39
Espn	indel	NS	4	151503731	delC;C>CT	Frameshift	(2.35)	(1.45)
Fabp5	SNV	NS	3	10015063	G>AG	58S>NS	1.92	6.34
Fads2	SNV	NS	19	10175779	G>GA	47T>IST	30.86	3.48
Fgg	SNV	S	3	82812657	C>CT	131S>SS	9478.53	896.52
Fos	SNV	NS	12	86815104	G>AG	39S>NS	0.05	0.03
Fos	SNV	S	12	86816132	G>AG	130E>EE	0.05	0.03
Foxq1	SNV	NS	13	31651848	C>CT	355A>AV	3.35	0.77
Fzd7	SNV	S	1	59540012	G>AG	70L>LL	2.51	1.76
Fzd7	SNV	S	1	59540237	G>AG	145R>RR	2.51	1.76
G0s2	SNV	S	1	195098972	G>AG	8S>SS	5.20	0.22
G0s2	SNV	NS	1	195098727	G>AG	90A>VA	5.20	0.22
G0s2	SNV	NS	1	195098751	G>AG	82A>VA	5.20	0.22
G0s2	SNV	NS	1	195098773	C>CG	75A>AP	5.20	0.22
G0s2	SNV	NS	1	195098827	G>AG	57Q>XQ	5.20	0.22
Gadd45g	indel	NS	13	51942996	insT;C>CT	Frameshift	(3.41)	(1.15)
Gdf15	SNV	S	8	73155343	C>CT	56E>EE	9.89	3.10
Gpx1	SNV	NS	9	108241856	C>CT	72L>LF	1.58	1.31
H3f3b	SNV	S	11	115885064	G>AG	52I>II	0.32	0.75
Hes1	SNV	NS	16	30065775	C>CT	34H>HY	2.79	1.01
Hes1	SNV	NS	16	30066009	G>AG	68D>ND	2.79	1.01
Hes1	SNV	NS	16	30066249	G>AG	69S>NS	2.79	1.01
Hist1h1c	SNV	S	13	23831235	C>CT	173S>SS	1.03	1.25
Hist1h1d	SNV	NS	13	23647312	C>CT	119P>PS	1.71	2.20
Hist1h1d	SNV	NS	13	23647441	C>CT	162P>PS	1.71	2.20
Hist1h1d	SNV	NS	13	23647504	C>CT	183P>PSA	1.71	2.20
Hist1h1d	SNV	NS	13	23647654	C>CT	233R>RW	1.71	2.20
Hist1h1e	SNV	NS	13	23713850	G>AG	193P>SP	0.10	1.73

Hist1h1e	SNV	NS	13	23713882	C>CT	182S>SN	0.10	1.73
Hist1h1e	SNV	NS	13	23713916	G>AG	171H>YH	0.10	1.73
Hist1h1e	SNV	NS	13	23714126	G>AG	101Q>XQ	0.10	1.73
Hist1h1e	SNV	S	13	23713857	G>AG	190S>SS	0.10	1.73
Hist1h2bb	SNV	NS	13	23838810	C>CT	50H>HY	–	–
Hist1h2bk	SNV	S	13	22128073	C>CT	107L>LL	0.92	1.99
Hist1h3a	SNV	NS	13	23853962	C>CT	98S>SN	0.21	1.73
Hist1h3a	SNV	S	13	23853961	G>AG	98S>SS	0.21	1.73
Hist1h4c	SNV	NS	13	23790295	C>CT	8G>VGD	–	–
Hist2h2ac	SNV	S	3	96024476	C>CT	97L>LL	0.12	1.36
Hist2h2ac	SNV	S	3	96024488	C>CT	93E>EE	0.12	1.36
Hist4h4	SNV	NS	6	136752629	G>AG	91L>FL	2.34	1.43
Hnrnpa1	SNV	NS	15	103071540	C>CT	16L>LF	0.27	0.66
Ier2	SNV	S	8	87186325	G>GT	142R>RR	1.68	0.97
Ier2	SNV	S	8	87186580	G>AG	57C>CC	1.68	0.97
Ier5	SNV	NS	1	156945649	G>AG	373P>SP	0.17	0.47
Igfbp1	SNV	NS	11	7098022	G>GA	21G>DAG	770.63	659.29
Igfbp1	SNV	S	11	7098110	C>CT	50C>CC	770.63	659.29
Il17rc	SNV	NS	6	113432885	G>AG	600V>MV	77.84	10.63
Insig1	SNV	S	5	28398081	C>CT	35G>GG	36.23	3.07
Insig1	SNV	S	5	28398246	G>AG	90Q>QQ	36.23	3.07
Insig1	SNV	NS	5	28398015	C>CT	13S>RSS	36.23	3.07
Insig1	SNV	NS	5	28398158	G>AG	61S>NS	36.23	3.07
Insig1	SNV	NS	5	28398323	G>AG	116G>DG	36.23	3.07
Insig1	SNV	NS	5	28398328	G>AG	118A>TA	36.23	3.07
Irf2bp1	SNV	NS	7	19589817	G>AG	11W>XW	0.84	1.04
Jun	SNV	S	4	94717641	C>CT	424Q>QQ	6.15	2.60
Jun	SNV	S	4	94717644	C>CT	423E>EE	6.15	2.60
Jun	SNV	S	4	94717991	G>AG	308L>LL	6.15	2.60
Junb	SNV	S	8	87501936	C>CG	237G>GG	0.50	0.30
Junb	SNV	NS	8	87501814	C>CT	278W>WX	0.50	0.30
Junb	SNV	NS	8	87502211	C>CT	146G>GS	0.50	0.30
Jund	SNV	S	8	73223213	G>AG	86G>GG	1.52	1.11
Jund	SNV	NS	8	73223881	G>AG	309S>NS	1.52	1.11
Jund	SNV	NS	8	73223175	G>AG	74A>TA	1.52	1.11
Klhl13	SNV	NS	X	22797884	C>CT	508R>RQ	2.46	1.46

Krt8	SNV	S	15	101834373	C>CT	99E>EE	899.20	213.93
Krt8	SNV	S	15	101834421	C>CT	83K>KK	899.20	213.93
Ldlr	SNV	S	9	21528266	C>CT	16L>LL	12.26	2.57
Maob	SNV	S	X	16293534	G>AG	404P>PP	2.57	1.75
Mc3r	SNV	NS	2	172074866	G>AG	170V>IV	2.54	1.65
Mettl7b	SNV	S	10	128397538	G>AG	152S>SS	725.01	1.49
Mov10l1	SNV	NS	15	88835831	G>AG	514A>TA	2.35	1.45
Mtf1	SNV	S	4	124482127	G>AG	58L>LL	2.36	1.46
Nfkbia	SNV	S	12	56593306	C>CT	69Q>QQ	0.58	0.13
Nkapl	SNV	S	13	21560171	C>CG	62R>RP	2.46	2.04
Nr0b2	SNV	S	4	133109875	C>CT	179N>NN	25.83	1.43
Nr0b2	SNV	NS	4	133109616	G>AG	93C>YC	25.83	1.43
Nr0b2	SNV	NS	4	133109619	G>AG	94C>YC	25.83	1.43
Nt5e	SNV	S	9	88222612	G>AG	30E>EE	2.37	1.45
Olfr395	SNV	NS	11	73720101	C>CT	297M>MI	2.47	1.49
Olfr750	SNV	NS	14	51690384	C>AC	228V>FV	–	–
Onecut1	SNV	S	9	74710870	G>AG	256L>LL	65.68	20.36
P4hb	SNV	NS	11	120434054	C>CT	23E>EK	6.33	1.57
Paqr9	SNV	NS	9	95460521	C>CT	49P>PS	10.09	1.27
Pck1	SNV	NS	2	172979068	G>AG	60M>IM	–	–
Pck1	SNV	NS	2	172979479	G>AG	90S>NS	–	–
Pck1	SNV	NS	2	172980228	G>AG	145S>NS	–	–
Pcsk9	SNV	S	4	106136386	C>CT	43E>EE	87.60	104.05
Pdzk1	SNV	NS	3	96658479	A>AG	162N>ND	60.18	7.95
Plk2	SNV	S	13	111186500	G>AG	129E>EE	8.10	5.19
Plk3	SNV	S	4	116805814	G>AG	139F>FF	14.32	1.57
Plk3	SNV	S	4	116805856	G>AG	125R>RR	14.32	1.57
Plk3	SNV	S	4	116805898	C>CT	111E>EE	14.32	1.57
Ptma	SNV	S	1	88426050	G>AG	46E>EE	0.28	1.28
Qsox1	SNV	S	1	157630652	G>AG	343Y>YY	19.30	0.79
Rac3	SNV	S	11	120583903	C>CT	63D>DD	2.20	1.61
Rbp4	SNV	NS	19	38198851	G>AG	61A>VA	9360.58	3515.97
Rbp4	SNV	NS	19	38198896	G>AG	46A>VA	9360.58	3515.97
Rbp4	SNV	S	19	38198853	G>AG	60I>II	9360.58	3515.97
Rdh8	SNV	S	9	20627811	C>CT	110S>SS	2.35	1.43
Rrs1	SNV	NS	1	9536104	G>CG	167M>IM	1.08	2.80

Scd1	SNV	NS	19	44481883	G>GT	6L>IL	318.00	0.95
Sephs2	SNV	S	7	134416519	G>AG	352L>LL	20.37	1.89
Sephs2	SNV	NS	7	134417076	G>AG	166A>VA	20.37	1.89
Sgk1	SNV	NS	10	21714771	C>CT	26A>AV	2.47	1.47
Sik1	SNV	S	17	31991940	C>CT	18Q>QQ	–	–
Slc27a2	SNV	NS	2	126379128	C>CT	80T>TI	15.48	1.51
Slc7a5	SNV	S	8	124431107	C>CT	133K>KK	0.04	1.42
Smg1	SNV	NS	7	125309982	T>GT	1910K>KN	–	–
Socs3	SNV	S	11	117829254	C>AC	126V>VV	0.36	1.19
Sox9	SNV	NS	11	112644154	G>AG	86W>XW	17.49	1.93
St3gal6	SNV	NS	16	58473929	C>CT	152G>GD	0.05	0.56
Tk1	SNV	NS	11	117687088	C>CT	2S>SN	0.22	0.92
Tkt	SNV	S	14	31362521	C>CT	21R>RR	0.50	0.61
Tmem49	SNV	NS	11	86400004	G>AG	357T>IT	2.80	1.31
Tob2	SNV	S	15	81681801	C>CT	167Q>QQ	2.50	1.30
Ttr	SNV	NS	18	20825069	G>AG	61W>XW	291.13	37.23
Zfp36	SNV	S	7	29163230	C>CT	102R>RR	1.75	0.18
Zfp36	SNV	NS	7	29163204	C>CT	111R>RH	1.75	0.18
Zfp36	SNV	NS	7	29163274	C>CT	88A>AT	1.75	0.18
Zfp36l2	SNV	S	17	84586208	C>CT	113A>AA	1.21	0.94
Zp2	SNV	NS	7	127278692	G>AG	549A>VA	2.49	1.49

 # NS: nonsynonymous S: synonymous

***Values are expression ratio of adult liver (AL) relative to bone marrow (BM).**

****** Values are expression ratio of fetal liver (FL) relative to BM.**

The genes which have been reported to be mutated in human HCC tissues are marked with half-tone dot meshing.