Supplementary Information

Whole Genome Sequencing of Liver Cancers Identifies Etiological Influences on Mutation Patterns and Recurrent Mutations in Chromatin Regulators

Akihiro Fujimoto*, Yasushi Totoki*, Tetsuo Abe, Keith A Boroevich, Fumie Hosoda, Ha Hai Nguyen, Masayuki Aoki, Naoya Hoshono, Michiaki Kubo, Fuyuki Miya, Yasuhito Arai, Hiroyuki Takahashi, Takuya Shirakihara, Masao Nagasaki, Tetsuo Shibuya, Kaoru Nakano, Kumiko Watanabe-Makino, Hiroko Tanaka, Hiromi Nakamura, Jun Kusuda, Hidenori Ojima, Kazuaki Shimada, Takuji Okusaka, Masaki Ueno, Yoshinobu Shigekawa, Yoshiiku Kawakami, Koji Arihiro, Hideki Ohdan, Kunihito Gotoh, Osamu Ishikawa, Shunichi Ariizumi, Masakazu Yamamoto, Terumasa Yamada, Kazuaki Chayama, Tomoo Kosuge, Hiroki Yamaue, Naoyuki Kamatani, Satoru Miyano, Hitoshi Nakagama, Yusuke Nakamura, Tatsuhiko Tsunoda, Tatsuhiro Shibata, and Hidewaki Nakagawa

*These authors contributed equally to this work.

Correspondence should be addressed to Hidewaki Nakagawa (hidewaki@ims.u-tokyo.ac.jp) or Tatsuhiro Shibata (tashibat@ncc.go.jp).

Nature Genetics: doi:10.1038/ng.2291

Supplementary Note

Somatic point mutation and short indel calls

Read sequences were mapped by BWA[1] to the human reference genome (GRCh37). Possible PCR duplicated reads were removed by SAMtools[2] and an in-house program. After filtering by pair mapping distance, mapping uniqueness and orientation between paired reads, the mapping result files were converted into pileup format by SAMtools.

We used three kinds of read filters: set1; both read pairs were uniquely mapped with consistent orientation and pair distance (within average ± 3 s.d.), set2; at least one read pair was uniquely mapped with consistent orientation and pair distance and, set3; all uniquely mapped paired reads and set2, as described elsewhere[3]. Mutation calls were done using all three sets of filtered reads, and mutations identified in the all three sets were considered as candidates.

To identify point mutations, we used the following criteria; (1) non-reference calls with a frequency ≥ 0.15 after removing bases calls with base quality < 10, and mapping quality < 20, (2) supported by at least two base calls including one base call with base quality ≥ 30, (3) a SAMtools consensus quality ≥ 20 and root mean square mapping quality ≥ 40, (4) if three or more single nucleotide variants (SNVs) were found within any 10bp windows, or distance from nearest indel was less than 5bp, we discarded all SNVs. (5) if candidate non-coding SNVs were in a tandem repeat region suggested by tandem repeat finder[4], we discarded the SNVs. (6) if candidate SNVs were in RepeatMasker repeat regions (http://www.repeatmasker.org) within 1Mb from the

Nature Genetics: doi:10.1038/ng.2291

centoromeric or telemeric gaps, we discarded the SNVs. (7) if a base with consensus quality lower than 20 occurs within 3bp on either side of the target SNV, we discarded the SNVs. After SNV calling in the tumor samples, candidate SNVs were filtered based on the lymphocyte sequence of the same patient; (1) candidate SNV alleles with a frequency ≥ 0.03 after removing reads with base quality < 15, and mapping quality < 20, (2) depth of coverage in lymphocyte ≤ 5, (3) depth of coverage in lymphocyte ≤ 10 and candidate SNV allele was represented in the dbSNP database v131 (http://www.ncbi.nlm.nih.gov/projects/SNP/).

Short indels were identified based on gaps in a read's alignment by BWA. We defined indels using the following criteria; (1) if indels were supported by a frequency ≥ 0.1 and ≥ 4 reads after removing reads with mapping quality < 20, and root mean square mapping quality ≥ 40, (2) if candidate non-coding indels were in repeat regions suggested by tandem repeat finder or RepeatMasker, we discarded the indels. After indel calling in tumor samples, the candidates were filtered based on the lymphocyte sequence of the same patient using the following criteria; (1) depth of coverage in lymphocyte ≤ 7, (2) for coding and non-coding region, if any indels were identified within 5bp or 10bp region in lymphocyte, respectively, the candidate indel was discarded. Gene information was based on UCSC[5] annotation.

The proportion of mutant alleles was estimated by the number of mutant and reference base calls. The mutation caller showed a false positive rate ≤ 0.05 in point mutations and ≤ 0.1 in short indels.

Estimation of false positive and false negative rates in somatic mutation calling

To examine accuracy of our analysis pipeline, we performed Sanger sequence validation for randomly selected point mutation and indel candidates. The false positive rate was 0/124 in coding point mutations, 2/44 (4.5%; 95% CI 0.6-15.5%) in non-coding point mutations, 1/16 (6.3%; 95% CI 0.16-30.2%) in coding indels and 3/32 (9.3%; 95% CI 0.20-25.0%) in non-coding indels. Note that false positive rates in repeat regions may be higher than these values because mutations in repeat regions were difficult to examine by PCR-Sanger sequencing method.

False negative rate (FNR) is very important, but difficult to estimate in the cancer genome sequence, because it is difficult to obtain sufficient number of already identified somatic mutations. Therefore, we used SNP array genotype for FNR estimation. We genotyped one cancer (HB6 tumor) and lymphocyte sample from another patient (HB5 lymphocyte) with Illumina OmniExpress array, and then ran our analysis pipeline by using HB6 tumor and HB5 lymphocyte data. We compared the results from our analysis pipeline with these from genotyping arrays. If our analysis pipeline did not call HB6 tumor-specific SNP alleles, i.e. heterozygous in HB6 tumor and homozygous in HB5 lymphocyte, we considered that SNP to be a false negative in our pipeline. By the comparison of HB6 tumor data and HB5 lymphocyte data, we identified 61,077 SNPs that were homozygous with reference allele on GRCh37 in HB5 lymphocyte and heterozygous in HB6 tumor by the genotyping array. Of these, 56,074 were identified by WGS data analysis, therefore FNR was estimated to be 8.2% (95% C.I. 8.0-8.4%). Four-fifth of false negatives (6.5%) were caused by insufficient depth of

coverage in HB5 lymphocyte.

Identification of common mutation in MCTs

We compared somatic mutation sites and mutation patterns of the two pairs of MC tumors (HC3 and HC7). In protein-coding regions, no common point somatic mutation was identified. Three genes, *ATM, FSIP2*, and *LRFN5*, were mutated in both the HC3-1 and HC3-2 tumors, but the mutations occurred at different positions. In non-coding regions, WGS identified 30 and 37 common somatic point mutations and indels in the HC3 and HC7 MCT HCC genomes, respectively. However, most of these occurred in repetitive regions and all candidates that could be sequenced by Sanger sequencing method (n=20) were found to be germline variants. This high false positive rate is due to the use of a common lymphocyte sample as a control. False negatives of germline variants from the lymphocyte sample could cause common false positive somatic mutations in both tumors. There was also no common structural alteration found in these MCT HCC genomes. These findings strongly suggest that these synchronous MCT tumors developed through an accumulation of a completely different set of genetic alterations.

Permutation test on the distance between MCTs

To evaluate the distance between the paired MCTs, we performed permutation tests based on the PCA (**Fig. 2c and 2d**). We randomly selected two sets of tumors from the 25 tumors and calculated the three-dimensional distance between tumors in a set. Then

the average distance of two sets was calculated. We repeated this process 100,000 times to obtain the null distribution. The average distance of the MC pair was tested under this null distribution.

Association between mutations in chromatin regulators and clinical factors

We examined the association of chromatin regulator mutations with several clinical factors. Tests of significance for tumor size, Edmondson grade, and liver fibrosis were performed by the Pearson's correlation test. Tests for portal vein invasion and hepatic vein invasion were carried out by the Fisher's exact test. We found that the stage of liver fibrosis in the background liver was associated with mutations and CNAs of the chromatin regulator genes (*P-value* = 0.026) or ARID family genes (*P-value* = 0.0086) (**Supplementary Table 11**). We also found that the presence of invasion to hepatic vein in tumors was associated with mutations and CNAs of *ARID1A* (*P-value* = 0.042) (**Supplementary Table 11**). Although we did not observe any significant association after the Bonferroni correction, mutations of the chromatin regulators in HCC may contribute to poor prognosis of HCC patients.

Copy number alternation (CNA)

To investigate copy number alterations, we analyzed the ratio of the depth of coverage across 5kb windows of the tumor sample to that of the matched lymphocytes. In total, we detected 294 deleted regions ($\log_2$R ratio $\leq$ -1) and 20 amplified regions ($\log_2$R ratio $\geq$ 2), of which 39 deletions and one amplification occurred recurrently. We also

observed 397 low-level losses (log_2R ratio ≤ -0.6) and one low-level gain (log_2R ratio ≥ 1) with a frequency greater than 15% (n ≥ 5), most of which have been reported previously[6,7] (**Supplementary Table 3**). Among the significantly mutated genes (**Table 1, Supplementary Table 7**), *TP53* harbored both a missense point mutation and low-level loss in five tumors, while HC6 had a 1bp-deletion in the coding region and low-level loss of *ARID1A*. Additionally, we identified several deletions or low-level losses of the chromatin regulators; *ARID1A* in 5 tumors, *MLL* in 4 tumors, *ARID1B* in one tumor and *MLL3* in one tumor (**Fig. 4a**).

Genomic rearrangement

Using inconsistently-mapped reads, we detected an average of 20.8 genomic rearrangements per tumor with a wide range among samples (0-59) (**Supplementary Table 2)**. Among the 561 PCR-validated rearrangements (**Supplementary Table 2, Supplementary Fig. 11**), breakpoints of 351 rearrangements were in intragenic regions and 30 rearrangements involved possible gene fusion events. Thirteen genes (*ACVR2A, AUTS2*, *CADPS2, EPAS1, FHIT, MBD5, NR3C1, PLAG1, PLEKHG3, RAD51L1, RBFOX3, STK39* and *XRCC4*) had rearrangement breakpoints in multiple HCCs. Seventy genes had both protein-alternating point mutations and a rearrangement. Fifteen genes found to have rearrangement breakpoints also contained recurrent protein-alternating point mutations (**Supplementary Tables 4 and 7**), including *CSMD1*, the chromatin regulator *ARID2* and two significantly mutated genes, *ALB* and *ATM*. We did not observe any recurrent genomic rearrangements that generated an

in-frame fusion gene in this cohort. There is no apparent difference of the observed genomic structure changes between HBV- and HCV-related HCCs **(Supplementary Fig. 2)**, and there is no common rearrangement in the two sets of the MCTs (HC3 pair and HC7 pair) (**Fig. 2a)**.

Recurrent somatic mutations in non-coding regions

We tested the number of mutations in the 5'UTR, 3'UTR, promoter regions (1kb region from transcription start sites) and non-coding RNA regions while considering the regional differences of the mutation rate and region lengths. As a result, 74 regions were identified with ≥ 3 point mutations or indels. Of these, 63 were considered significant after adjustment for multiple testing (q-value ≤ 0.05), which included regulatory regions of coding genes as well as multiple small and long intergenic non-coding RNAs (**Fig 2a, Supplementary Table 12**). Among them, *PHF1*, which is involved in DNA damage response and epigenetic modification and reported to be rearranged in rare uterine sarcomas[8], was significantly mutated in the promoter region as well as the protein-coding regions. Additionally, *MED1*, which is reported to play an important role in liver regeneration and liver carcinogenesis in animal models[9], was also significantly mutated in the 3'UTR as well as its protein-coding region.

References

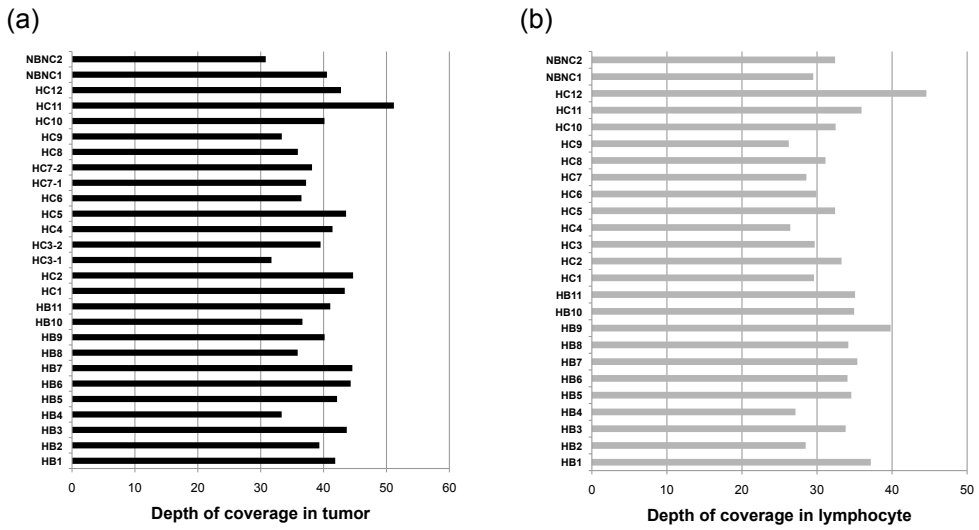
1. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
2. Li, H. et al. The Sequence Alignment/Map format and SAMtools.

Bioinformatics **25**, 2078-9 (2009).

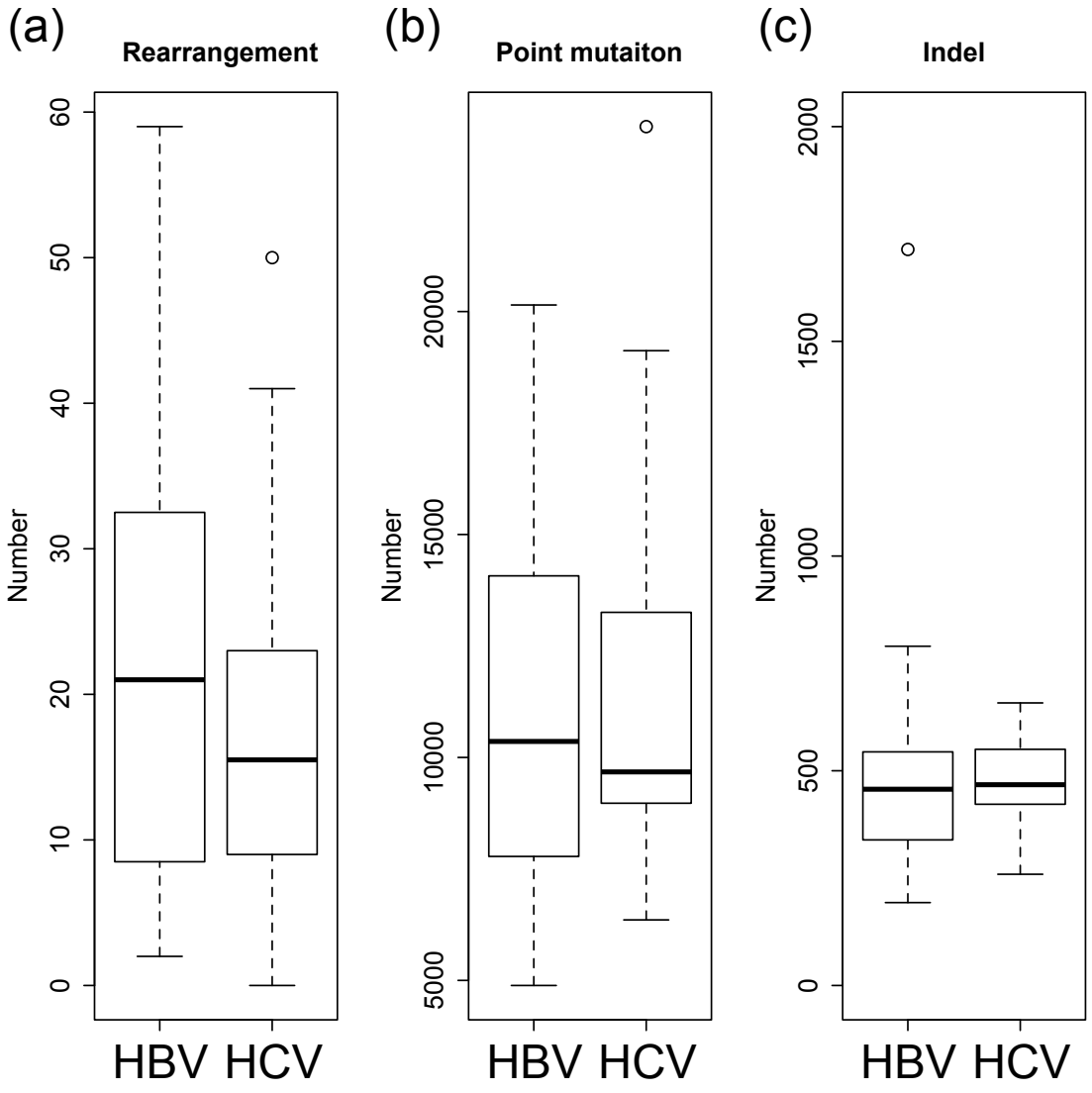
3. Totoki, Y. et al. High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* **43**, 464-9 (2011).
4. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573-80 (1999).
5. Fujita, P.A. et al. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* **39**, D876-82.
6. Laurent-Puig, P. & Zucman-Rossi, J. Genetics of hepatocellular tumors. *Oncogene* **25**, 3778-86 (2006).
7. Tsai, W.L. & Chung, R.T. Viral hepatocarcinogenesis. *Oncogene* **29**, 2309-24 (2010).
8. Micci, F., Panagopoulos, I., Bjerkehagen, B. & Heim, S. Consistent rearrangement of chromosomal band 6p21 with generation of fusion genes JAZF1/PHF1 and EPC1/PHF1 in endometrial stromal sarcoma. *Cancer Res* **66**, 107-12 (2006).
9. Matsumoto, K. et al. Critical role for transcription coactivator peroxisome proliferator-activated receptor (PPAR)-binding protein/TRAP220 in liver regeneration and PPARalpha ligand-induced liver tumor development. *J Biol Chem* **282**, 17053-60 (2007).

Supplementary Figures Contents

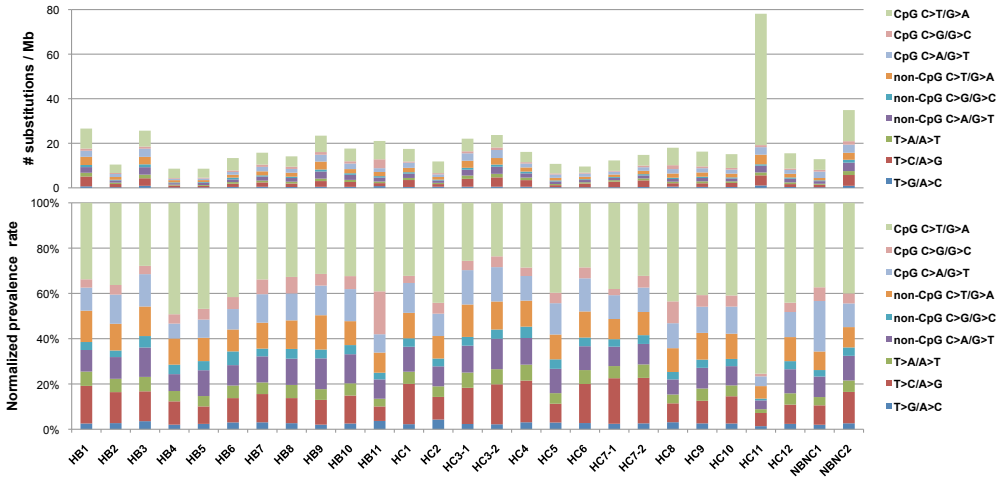
Nature Genetics: doi:10.1038/ng.2291



Supplementary Figure 1: Average depth of coverage with uniquely mapped reads in each sample. We calculated depth of coverage of each sample by using uniquely mapped reads after removing PCR-duplication.

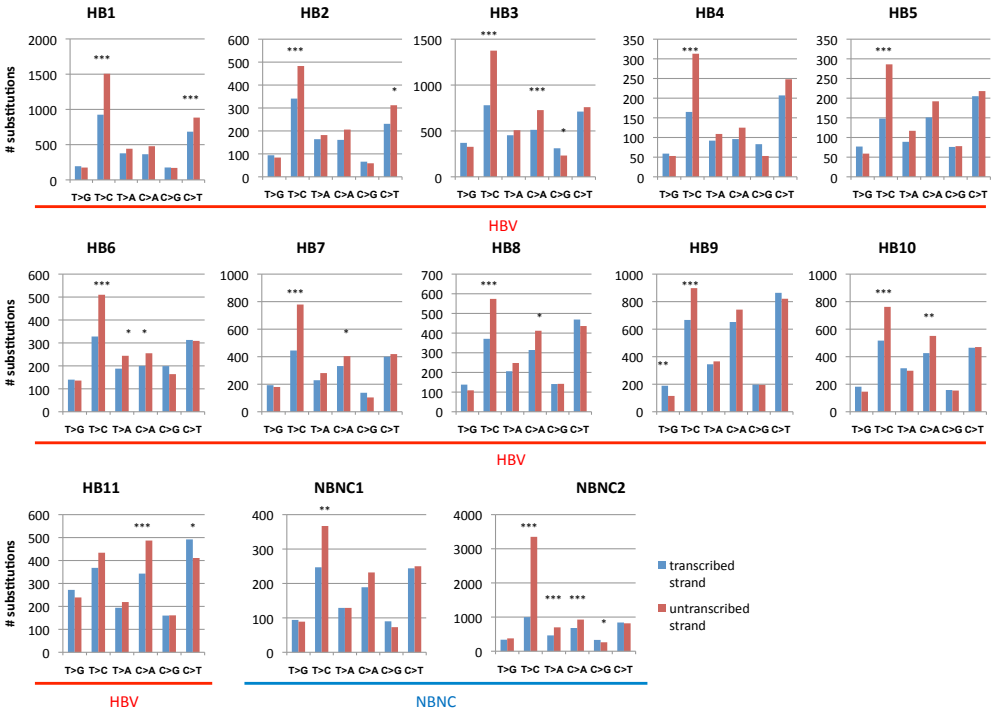


Supplementary Figure 2: Difference in mutation counts between HBV- and HCV-related tumors. (a) Number of rearrangements in HBV- and HCV-related tumors. (b) Number of point mutations in HBV- and HCV-related tumors. (c) Number of indels in HBV- and HCV-related tumors. No significant difference was observed in these comparisons.

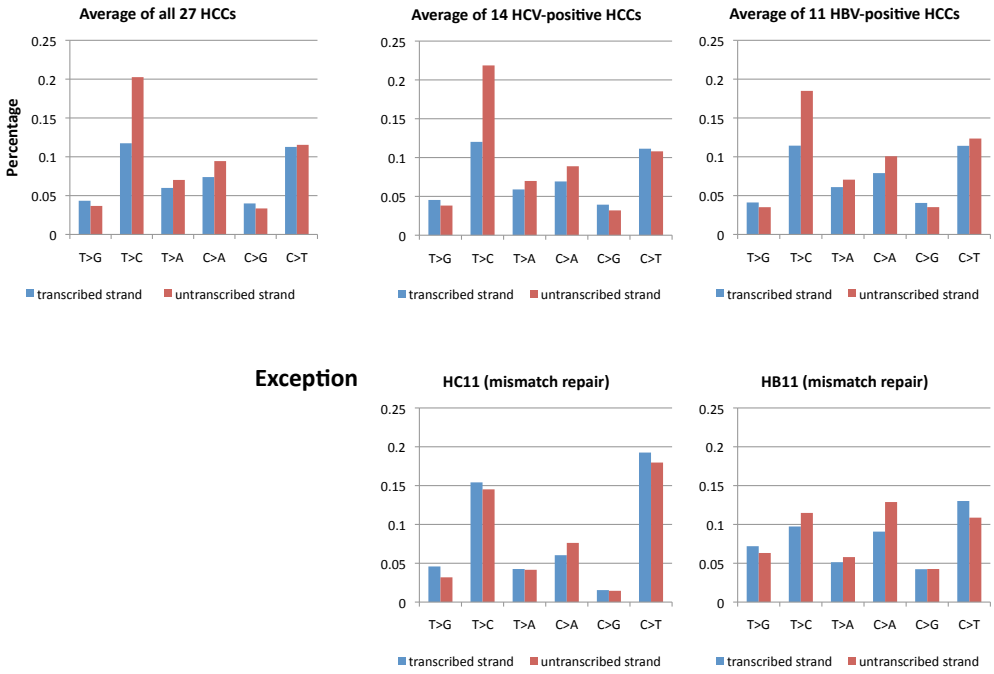


Supplementary Figure 3: Somatic substitution pattern of each tumor. (Top) Number of somatic substitutions in each of the nine classes per Mb in 27 HCC genomes. (Bottom) Frequency of the nine classes of somatic substitutions in 27 HCC genomes.

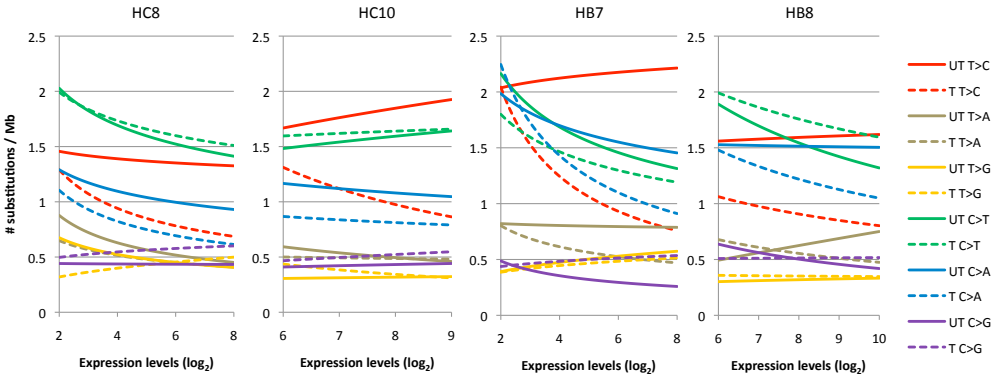
Supplementary Figure 4: The number of somatic substitutions on the transcribed strand and the untranscribed strand **(a)** in 14 HCV-related HCCs. ***; *P-value* < 10^{-6}, **; *P-value* < 0.0001, *; *P-value* < 0.01.



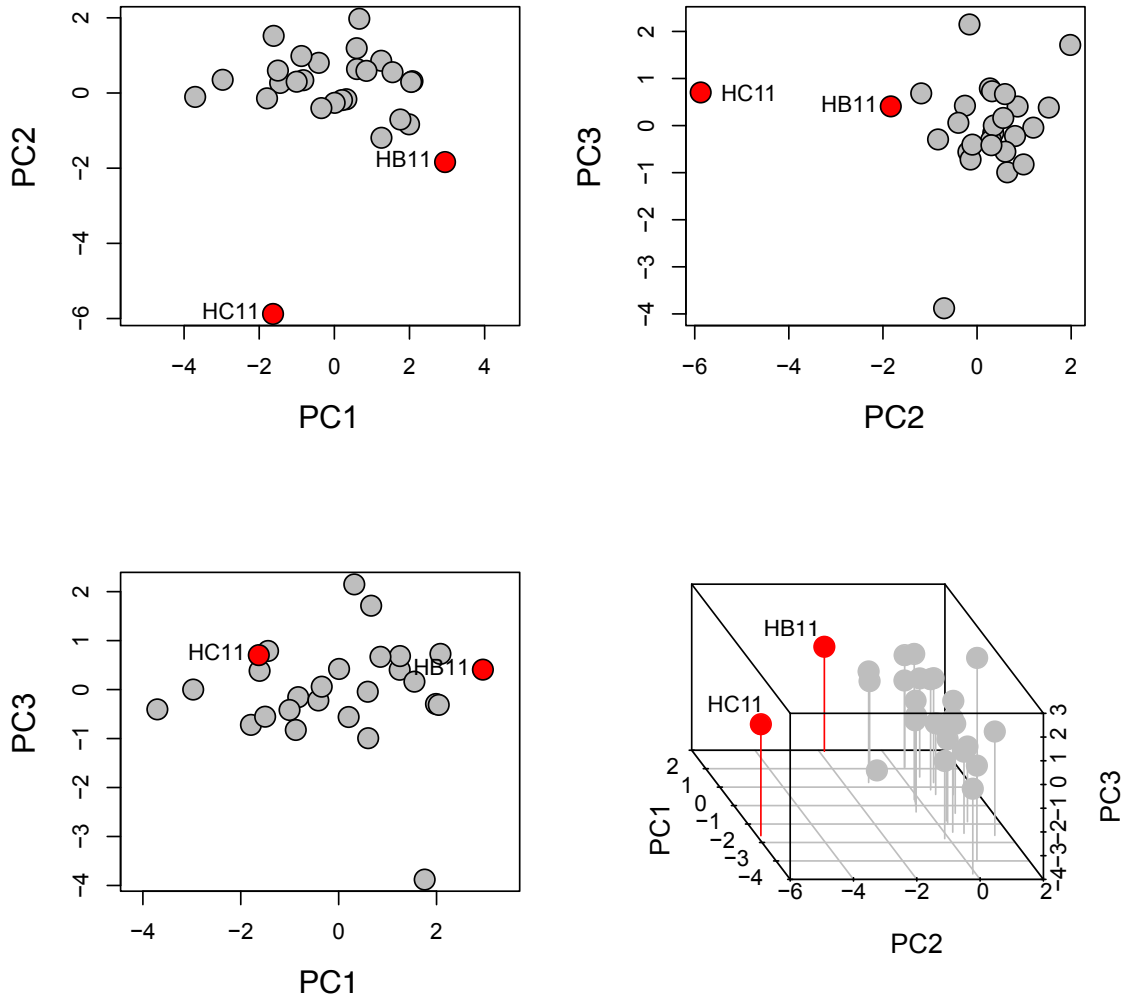
Supplementary Figure 4: The number of somatic substitutions on the transcribed strand and the untranscribed strand **(b)** in 11 HBV-related HCCs, and 2 HCCs without HBV and HCV infections (NBNC). ***; *P-value* < 10^{-6}, **; *P-value* < 0.0001, *; *P-value* < 0.01.



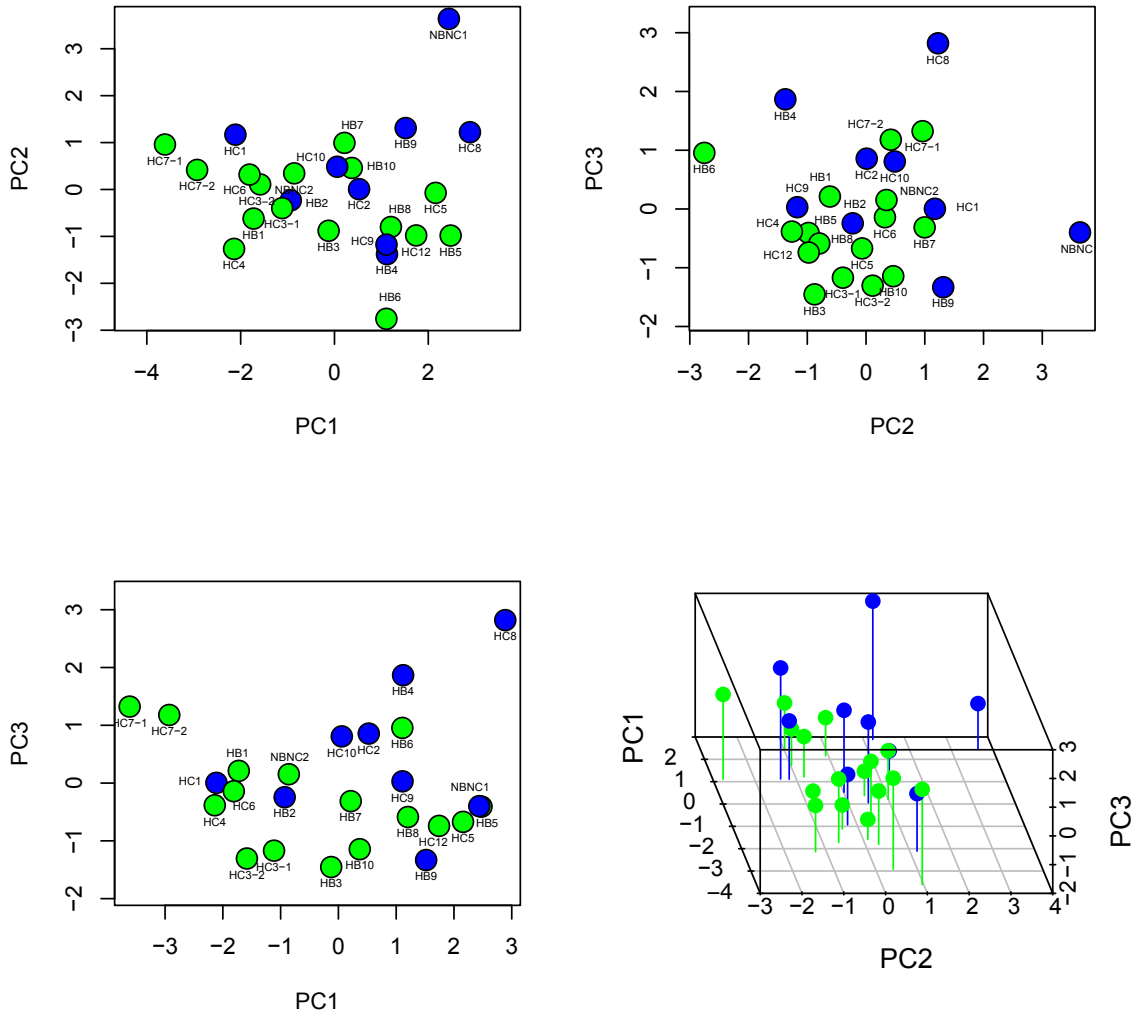
Supplementary Figure 4: The number of somatic substitutions on the transcribed strand and the untranscribed strand. **(c)** The average percentage of somatic substitutions in all 27 HCCs, 14 HCV-related HCCs and 11 HBV-related HCCs (top). Two exceptions (bottom). TCR did not occur in the mismatch-repair deficient tumor (HC11) with somatic nonsense mutation of *MLH1*. Another case (HB11) had a familial disposition to cancer and exhibited a distinct mutation signature (increased indels, suggesting mismatch-repair deficiency, less dominant T>C/A>G transitions, and a decreased effect of TCR at T>C transition).



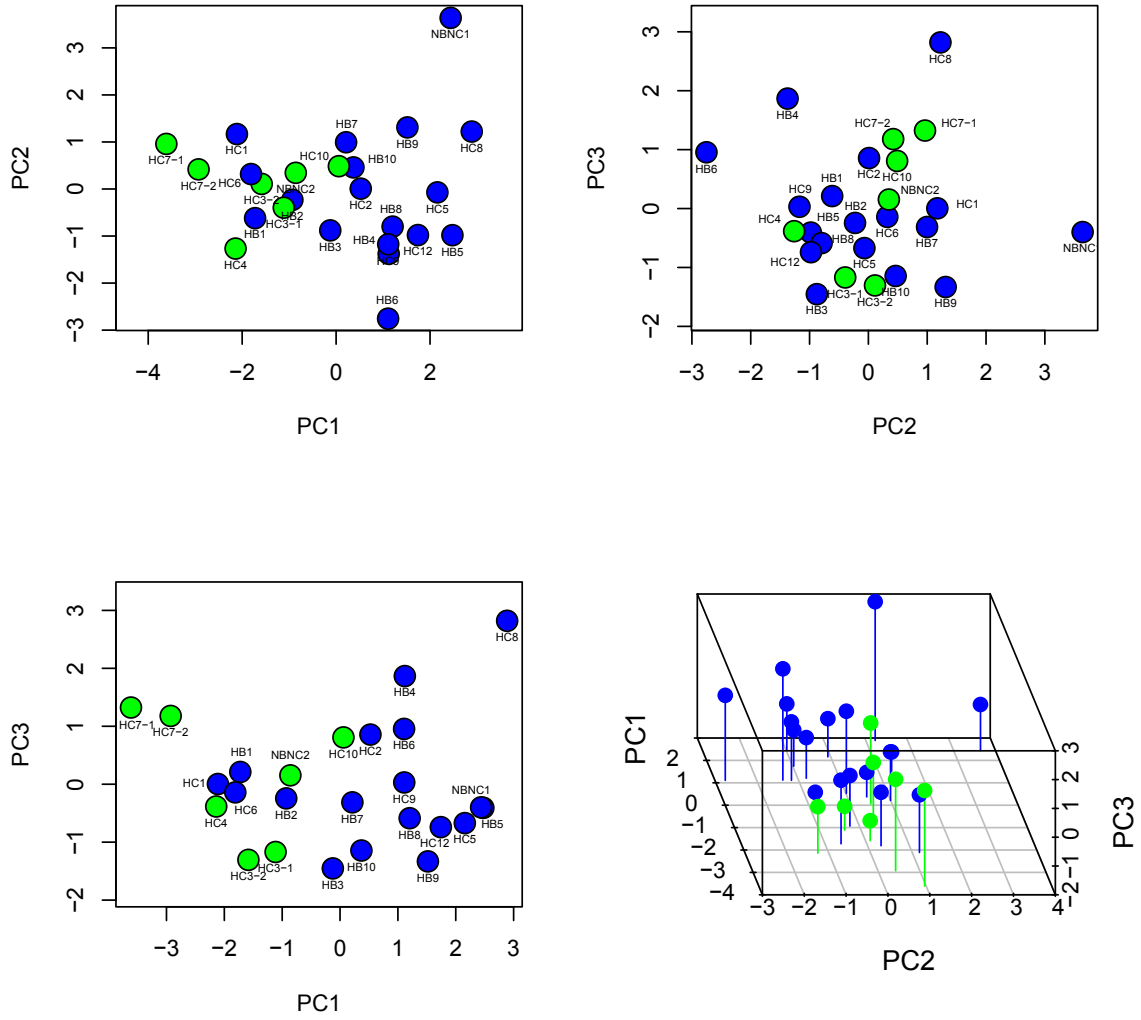
Supplementary Figure 5: Repair on the transcribed strand. Fitted curves showing the effect of gene expression and strand bias on substitution prevalence. We used Agilent microarray expression data (Whole Human Genome (8×60K) Oligonucleotide Microarray) in this TCR analysis and Expression level (log_2) indicate Agilent microarray intensity level units with log_2 scale. UT, untranscribed strands; T, transcribed strands.



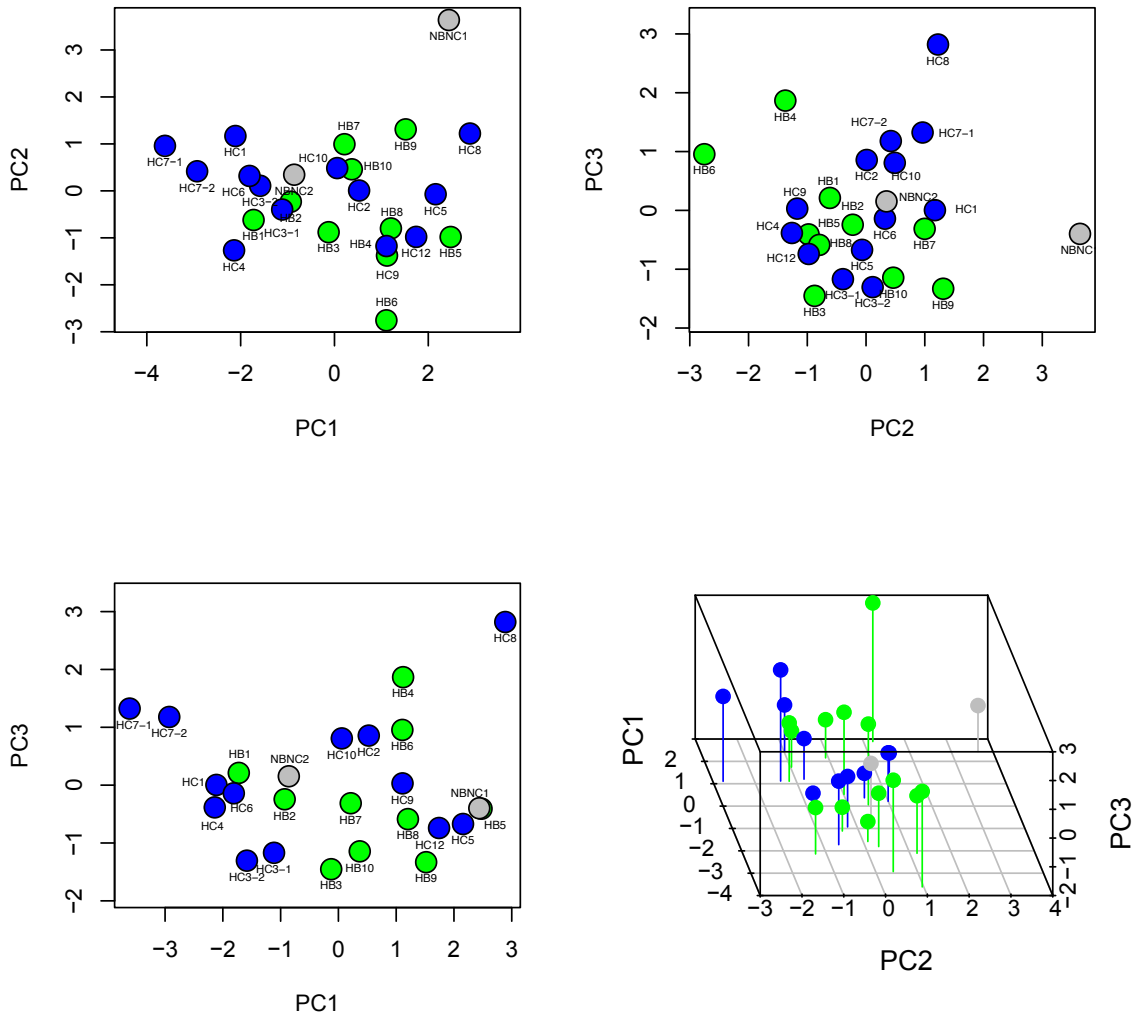
Supplementary Figure 6: PCA of somatic substitution patterns. **(a)** HB11 and HC11 are outliers, probably because they apparently show mismatch-repair deficient phenotype due to a somatic nonsense mutation (p.E234*) of *MLH1* for HC11 and unknown factors for HB11



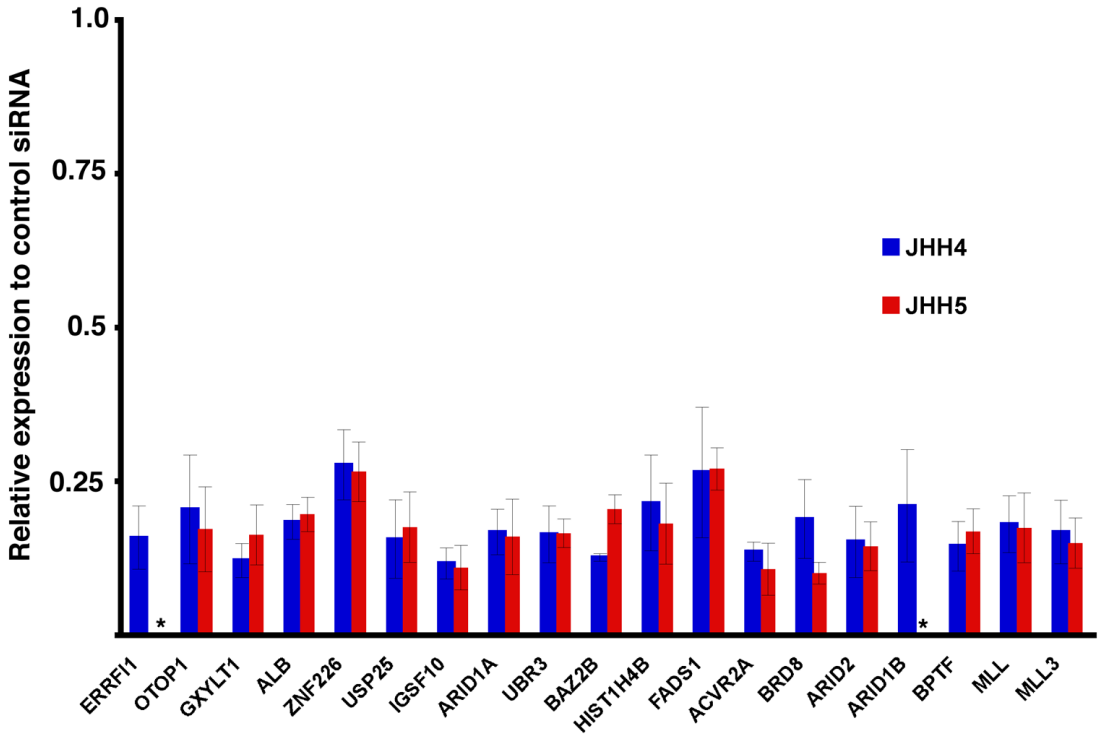
Supplementary Figure 6: PCA of somatic substitution patterns. **(b)** Comparison between tumors from habitual alcohol drinking (green) and non-drinking (blue) HCC patients. Habitual alcohol drinking was significantly correlated with somatic substitution patterns in HCCs (*P*-value = 0.028).



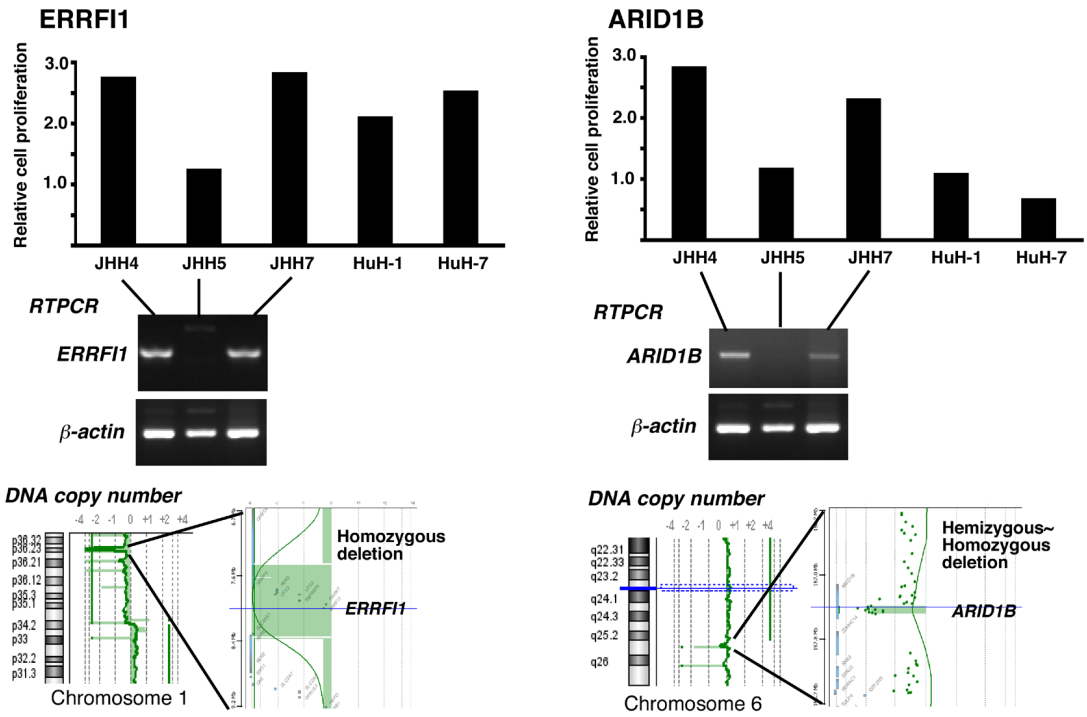
Supplementary Figure 6: PCA of somatic substitution patterns. **(c)** Comparison between synchronous/metachronous multiple liver nodules (green) and the other (blue) HCCs. The occurrence of synchronous or metachronous multiple HCCs, most of which are likely to be MC, was significantly correlated with somatic substitution patterns (*P*-value = 0.016)



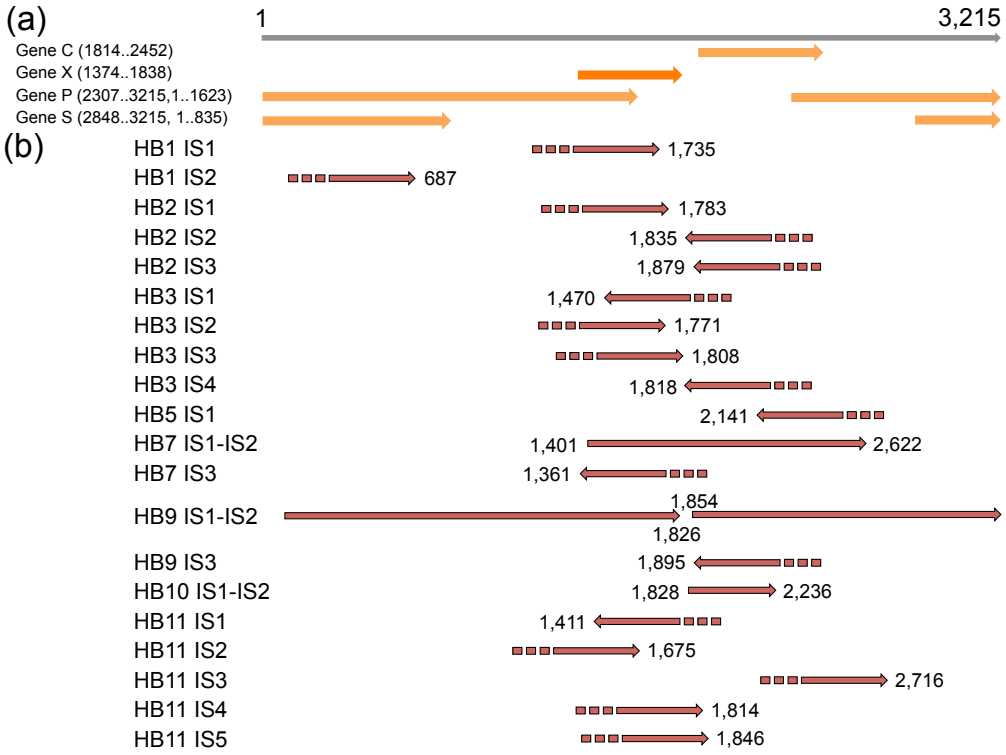
Supplementary Figure 6: PCA of somatic substitution patterns. **(d)** Comparison between HBV-related (green), HCV-related (blue), and NBNC (grey) HCCs. Different substitution patterns between the HBV-, HCV-related and NBNC HCCs were observed (*P*-value = 0.091, canonical correlation analysis, and *P*-value = 0.020, correlation analysis on the second component: PC2).



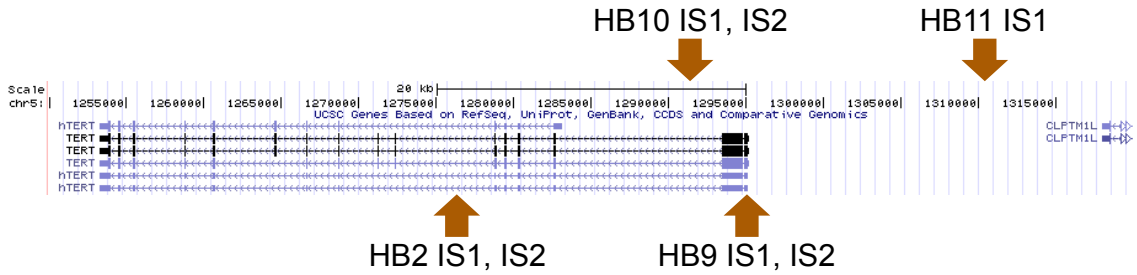
Supplementary Figure 7: Reduction of the target gene expression by siRNA. Relative expression of the target genes compared to the control siRNA treated cells was measured by quantitative RT-PCR. Asterisk shows the genes whose expression was not detected because of homozygous gene deletion (see **Supplementary Figure 8**). Data indicates mean +/- SD.



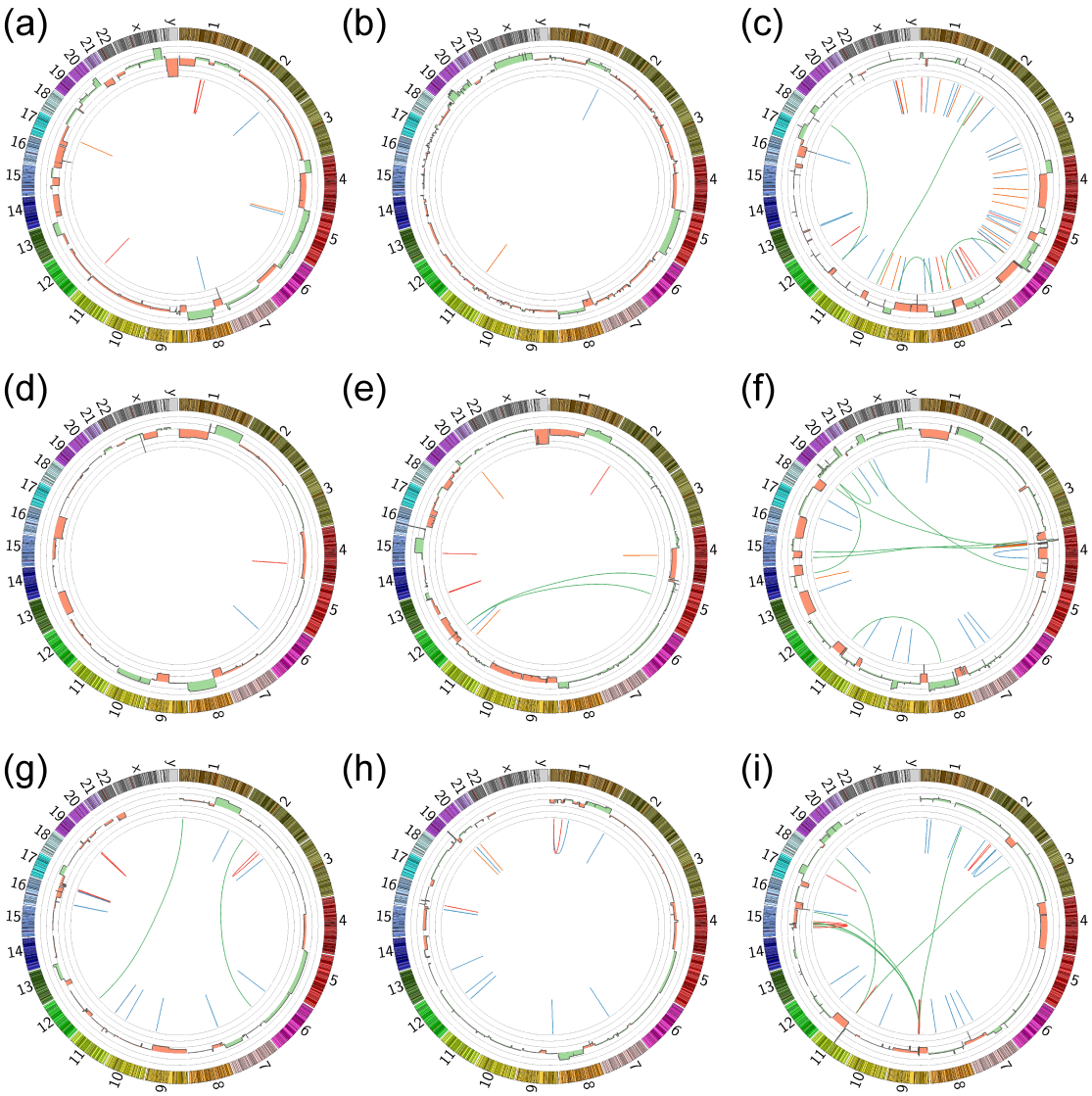
Supplementary Figure 8: Down-regulation of *ERRFI1* and *ARID1B* genes did not affect proliferation of JHH5 cells which lack expression of these two genes. (Top) Relative cell proliferation of five HCC cell lines treated with siRNAs against *ERRFI1* and *ARID1B* genes. Note that down-regulation of these genes did not affect proliferation of JHH5 cells. (Middle) RT-PCR analysis of *ERRFI1, ARID1B* and *β-Actin* (control) genes in three HCC cell lines. *ERRFI1* and *ARID1B* mRNAs were not detected in JHH5 cells. (Bottom) Array CGH analysis of *ERRFI1* and *ARID1B* loci in JHH5 cells. A homozygous deletion including the whole *ERRFI1* gene was detected on *1p36*. A hemizygous deletion with a minute homozygous deletion within the *ARID1B* gene was detected on *6q25*. A high-density oligonucleotide array containing 244,000 independent probes (Human Genomic Microarray 44B, Agilent Technologies) was used for array comparative genomic hybridization employing the manufacturer's protocol.

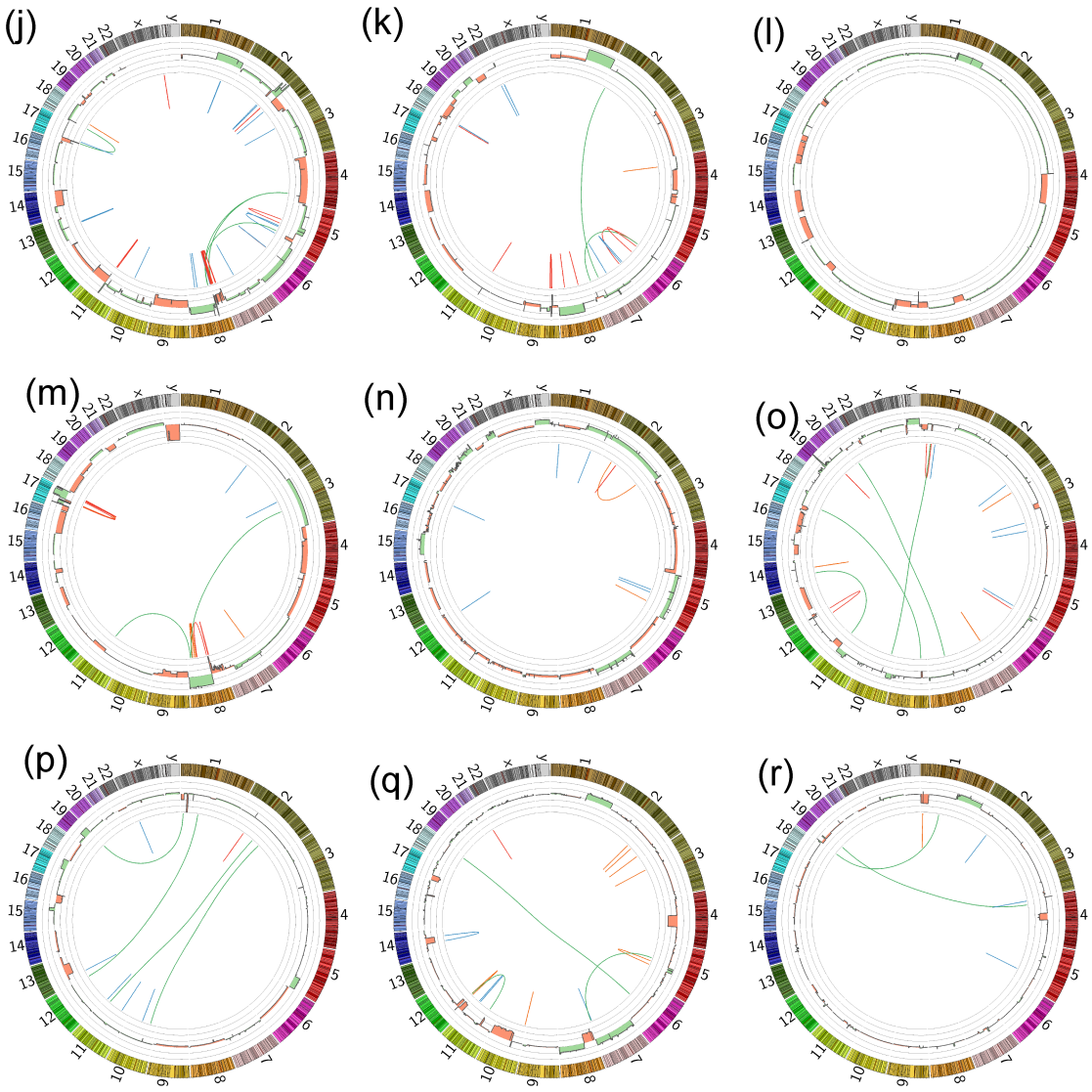


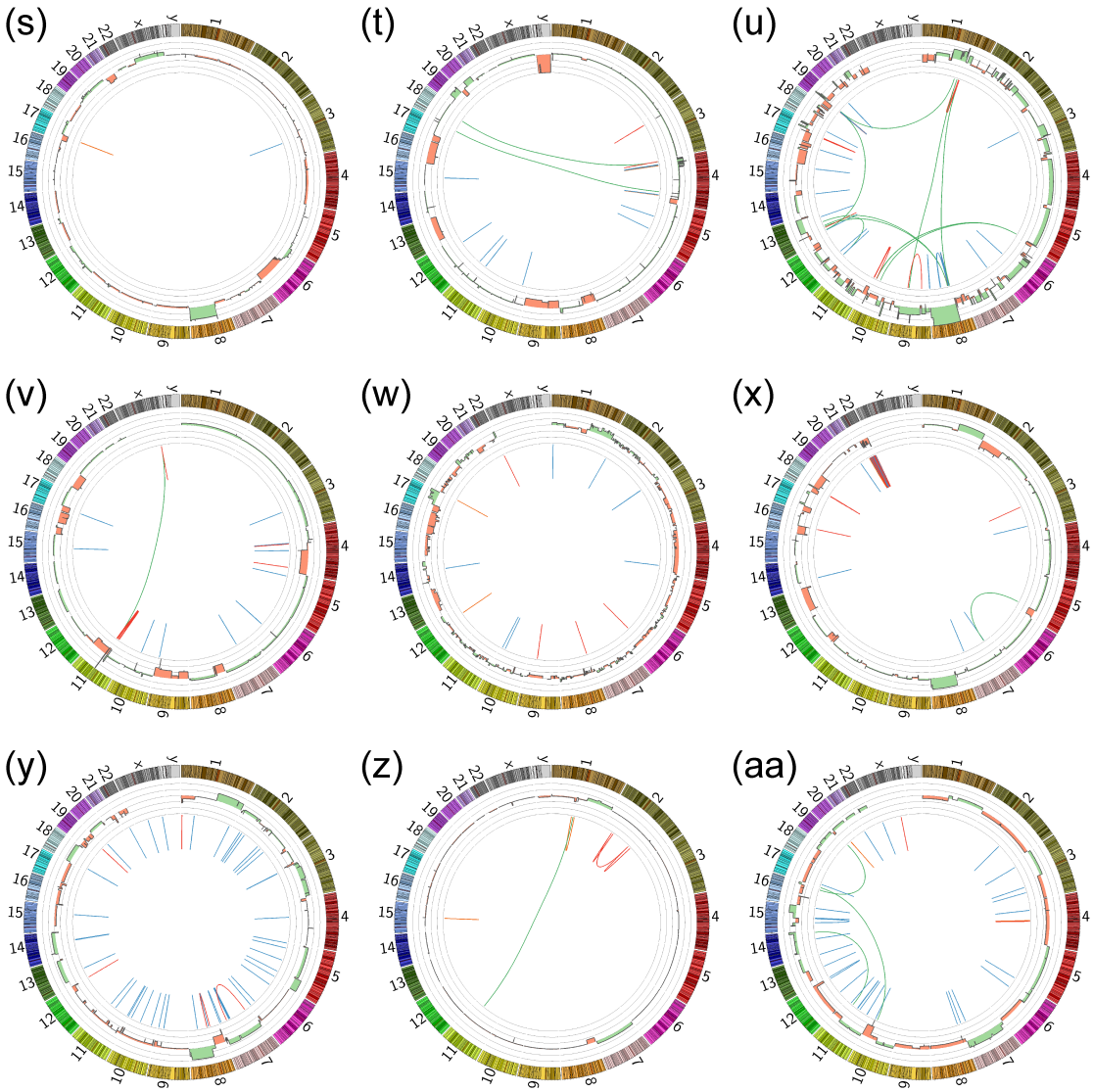
Supplementary Figure 9: Integration sites on HBV genome. **(a)** Structure of HBV genome. Arrows indicate gene structure. **(b)** Integration site on HBV genome. Arrows represents direction and location of integrated HBV genome. Integrated sequences were completely identified on HB9 IS1-IS2 and HB10 IS1-IS2. Integration site in host genome suggested that HB2 IS1 and IS2, and HB2 IS3 and IS4 are both ends of single integration events (**Supplementary Table 13**).



Supplementary Figure 10: HBV genome integration sites into the *TERT* locus. HBV genome integration was observed within or in the upstream region of the *TERT* gene in four HBV-related HCCs.


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)


(j)
(k)
(l)
(m)
(n)
(o)
(p)
(q)
(r)



Supplementary Figure 11: Graphical representation of 27 HCC genomes. (a) HB1, (b) HB2, (c) HB3, (d) HB4, (e) HB5, (f) HB6, (g) HB7, (h) HB8, (i) HB9, (j) HB10, (k) HB11, (l) HC1, (m) HC2, (n) HC3-1, (o) HC3-2, (p) HC4, (q) HC5, (r) HC6, (s) HC7-1, (t) HC7-2, (u) HC8, (v) HC9, (w) HC10, (x) HC11, (y) HC12, (z) NBNC1, (aa) NBNC2. Each circle plot represents validated rearrangements (inner arcs), and copy number alternations (inner rings). In rearrangements, green, blue, orange and red lines show translocation, deletion, inversion and tandem duplication. Copy number gain and loss region were shown in green and red.

Supplementary Table 1: Clinical and pathological features of 27 HCCs from 25 patiens analyzed by whole genome sequencing

ID	Age	Gender	Hepatitis	Alcohol drinking	Tumor size (mm)	Pretreatment	Edmondson grade[d]	Portal vein invasion
HB1	56	M	HBV	+	33	-	II	-
HB2	46	M	HBV	-	24	-	II	-
HB3	74	M	HBV	+	60	-	III	+
HB4	61	M	HBV	-	30	-	II-III	-
HB5	38	M	HBV	+	18	-	II-III	-
HB6	57	M	HBV	+	30	-	III	+
HB7	58	M	HBV	+	22	-	II	+
HB8	68	F	HBV	+	45	TACE[b]	II-III	+
HB9	60	M	HBV	-	80	-	II-III	+
HB10	56	M	HBV	+	63	-	III	+
HB11	41	F	HBV	+	30	-	II-III	-
HC1	62	M	HCV	-	87	-	II-III	+
HC2	71	M	HCV	-	45	-	III	-
HC3-1[a]	69	M	HCV	+	20	-	II	-
HC3-2[a]	69	M	HCV	+	20	-	II	-
HC4	61	M	HCV	+	18	-	II	-
HC5	58	M	HCV	+	24	-	II	-
HC6	61	M	HCV	+	35	TACE	II	-
HC7-1[a]	64	M	HCV	+	25	-	II	-
HC7-2[a]	64	M	HCV	+	25	-	II	-
HC8	73	F	HCV	-	21	-	II	-
HC9	66	M	HCV	-	40	-	II-III	-
HC10	62	F	HCV	-	20	-	III	-
HC11	71	F	HCV	-	35	-	III	+
HC12	57	M	HCV	+	43	-	I	-
NBNC1	81	F	NBNC	-	70	TACE+PEIT[C]	II	-
NBNC2	73	M	NBNC	+	60	TACE	II	-

Supplementary Table 1 (continued)

ID	Hepatic vein invasion	Liver fibrosis[e]	Platelet count ($x10^4$)/mm^2	Serum AFP level (ng/mL)	Other diseases	Family history of cancer	Multiple liver nodules
HB1	-	0	19.0	106.8	-	mother:HCC	-
HB2	-	4	11.3	15.6	pancreatic tumor	-	-
HB3	+	1	16.6	99.9	esophageal cancer	-	-
HB4	+	2	28.3	382.4	-	-	-
HB5	-	3	15.0	6.9	-	mother: breast cancer	-
HB6	-	4	8.9	2302.0	-	mother & brother: HCC	-
HB7	-	3	21.1	4.3	-	-	-
HB8	-	4	6.1	nd	-	-	-
HB9	-	2	14.4	1630.3	-	-	-
HB10	-	3	11.9	48.6	colon cancer	-	-
HB11	-	4	23.1	1678	-	mother:thyroid cancer, father: cancer of unknown origin	synchronous
HC1	+	4	12.4	1457.4	-	-	-
HC2	-	3	11.3	27.7	-	-	-
HC3-1[a]	-	4	14.4	7.0	-	-	synchronous
HC3-2[a]	-	4	14.4	7.0	-	-	synchronous
HC4	-	2	12.9	<5	HCC	father: stomach cancer	metachronous
HC5	-	3	15.3	<5	-	-	-
HC6	+	4	7.9	<5	-	mother: HCC, father: lung cancer	-
HC7-1[a]	-	4	22.6	3209.0	diabetes	-	synchronous
HC7-2[a]	-	4	22.6	3209.0	diabetes	-	synchronous
HC8	-	4	15.1	20	-	-	-
HC9	-	3	13.2	nd	-	-	-
HC10	-	4	11.3	26.8	lung cancer	-	metachronous
HC11	-	2	14.2	nd	-	-	-
HC12	-	2	22.7	5.2	diabetes	-	-
NBNC1	-	4	16.8	6.9	-	-	-
NBNC2	-	0	14.3	96.2	diabetes, colon cancer	-	metachronous

a; Multicentric occurance or intrahepatic metastasis was determined by considering their location, size, and histopathological features.

b; Trancecathter arterail chemoembolixation. c; Percutaneous ethanol injection.

d; Edmondson tumor grading. e; Fibrosis in non-cacerous liver tissue is determined accodirung the New Inuyama Classification.

Supplementary Table 2: Summary of somatic mutations in 27 HCC genomes

		Point mutation					Somatic indel		
Sample	Hepatitis	Total	Coding	Synonymous	Nonsynonymous	Splice site	Total	Coding indel	Splice site
HB1	HBV	20,147	133	38	95	5	616	5	0
HB2	HBV	7,083	49	12	37	2	472	6	0
HB3	HBV	19,586	120	30	90	4	790	4	0
HB4	HBV	4,974	26	11	15	1	315	0	0
HB5	HBV	4,886	35	9	26	0	193	2	0
HB6	HBV	8,477	62	14	48	3	470	5	0
HB7	HBV	10,619	84	26	58	3	435	6	0
HB8	HBV	9,583	81	22	59	1	252	1	0
HB9	HBV	16,186	136	38	98	2	363	2	0
HB10	HBV	11,959	103	26	77	2	457	3	0
HB11	HBV	10,359	91	18	73	1	1,714	10	1
HC1	HCV	13,055	69	15	54	4	502	2	0
HC2	HCV	7,120	45	13	32	1	466	5	0
HC3-1	HCV	17,294	115	30	85	0	461	6	0
HC3-2	HCV	19,122	112	26	86	2	550	9	0
HC4	HCV	13,254	68	12	56	3	469	4	1
HC5	HCV	6,357	36	7	29	0	336	2	0
HC6	HCV	7,248	55	13	42	0	259	2	0
HC7-1	HCV	9,006	49	11	38	1	391	6	0
HC7-2	HCV	11,370	63	14	49	0	485	2	1
HC8	HCV	9,426	57	7	50	2	561	2	1
HC9	HCV	9,925	91	25	66	2	658	4	0
HC10	HCV	9,293	66	17	49	1	426	2	0
HC11	HCV	24,147	376	109	267	2	8,950	58	0
HC12	HCV	8,973	74	18	56	3	422	3	0
NBNC1	nBnC	6,464	25	5	20	0	430	2	0
NBNC2	nBnC	22,964	242	62	180	3	496	8	0
Total		318,877	2,463	628	1,835	48	21,939	161	4
Max		24,147	376	109	267	5	8,950	58	1
Min		4,886	25	5	15	0	193	0	0
Average		11,810	91	23	68	2	813	6	0

Supplementary Table 2 (continued)

		Somatic rearrangement				
Sample	Hepatitis	Total	Deletion	Inversion	Tandem duplicaiton	Translocation
HB1	HBV	8	3	3	2	0
HB2	HBV	2	1	0	1	0
HB3	HBV	59	27	11	16	5
HB4	HBV	2	1	1	0	0
HB5	HBV	9	1	3	3	2
HB6	HBV	25	14	1	2	8
HB7	HBV	21	12	6	1	2
HB8	HBV	15	10	2	3	0
HB9	HBV	40	18	12	0	10
HB10	HBV	48	20	20	2	6
HB11	HBV	22	7	11	2	2
HC1	HCV	0	0	0	0	0
HC2	HCV	16	2	6	6	2
HC3-1	HCV	9	6	1	2	0
HC3-2	HCV	19	5	7	3	4
HC4	HCV	10	5	1	0	4
HC5	HCV	15	4	1	7	3
HC6	HCV	7	3	0	1	3
HC7-1	HCV	2	1	0	1	0
HC7-2	HCV	22	11	4	5	2
HC8	HCV	41	19	12	0	10
HC9	HCV	23	11	10	1	1
HC10	HCV	15	7	6	2	0
HC11	HCV	40	15	23	1	1
HC12	HCV	50	43	7	0	0
NBNC1	nBnC	6	0	3	2	1
NBNC2	nBnC	35	27	3	2	3
Total		561	273	154	65	69
Max		59	43	23	16	10
Min		0	0	0	0	0
Average		21	10	6	2	3

Supplementary Table 7: Genes recurrently mutated in 27 HCC genomes (n ≥ 2)

Gene	chr	start	last	CDS length (bp)	Coding indel	Nonsynonymous	Splice site SNV	Splice site indel	Total	*P-value*	*q-value*
TP53	17	7,572,927	7,579,912	1,218	0	11	3	0	14	0	0
ZIC3	X	136,648,851	136,652,229	1,412	0	3	0	0	3	0.0005	0.0207
ERRFI1	1	8,073,270	8,075,679	1,397	1	2	0	0	3	0.0002	0.0207
HIST1H4B	6	26,027,169	26,027,480	312	1	1	0	0	2	0.0004	0.0207
HLA-A	6	29,910,331	29,913,232	1,126	1	2	0	0	3	0.0004	0.0207
ZNF740	12	53,575,768	53,581,374	602	0	2	0	0	2	0.0008	0.0267
C14orf176	14	21,569,503	21,571,417	674	0	2	0	0	2	0.0012	0.0310
CTNNB1	3	41,265,560	41,280,833	2,398	0	3	0	0	3	0.0015	0.0310
GXYLT1	12	42,481,588	42,538,448	1,351	0	3	0	0	3	0.0013	0.0310
OTOP1	4	4,190,530	4,228,591	1,859	1	2	0	0	3	0.0015	0.0310
SLC25A5	X	118,602,479	118,605,021	909	0	2	0	0	2	0.0017	0.0324
FADS1	11	61,569,883	61,584,401	1,550	0	2	0	0	2	0.0047	0.0357
ATM	11	108,098,352	108,236,235	9,415	1	4	0	0	5	0.0037	0.0357
TNFRSF1B	1	12,227,149	12,267,077	1,422	0	2	0	0	2	0.0038	0.0357
TYW5	2	200,797,790	200,820,193	976	0	2	0	0	2	0.0038	0.0357
LRRC30	18	7,231,137	7,232,042	906	0	2	0	0	2	0.0047	0.0357
MMACHC	1	45,966,005	45,974,887	861	0	2	0	0	2	0.0030	0.0357
NCS1	9	132,934,943	132,988,766	597	0	2	0	0	2	0.0030	0.0357
CELA1	12	51,722,361	51,740,422	805	1	1	0	0	2	0.0023	0.0357
BRE	2	28,117,424	28,550,300	1,288	1	1	0	0	2	0.0047	0.0357
ZNF226	19	44,674,234	44,681,827	2,424	1	2	0	0	3	0.0043	0.0357
FFAR2	19	35,940,617	35,941,609	993	0	2	0	0	2	0.0038	0.0357
APOE	19	45,409,882	45,412,507	962	0	2	0	0	2	0.0047	0.0357
PEX14	1	10,535,024	10,690,044	1,166	2	0	0	0	2	0.0038	0.0357
CXADR	21	18,885,448	18,965,511	1,087	0	2	0	0	2	0.0047	0.0357
OR6C74	12	55,641,072	55,642,010	939	0	2	0	0	2	0.0030	0.0357
ALB	4	74,270,045	74,286,015	1,882	3	0	0	0	3	0.0022	0.0357
USP25	21	17,102,713	17,250,794	3,260	1	2	0	0	3	0.0051	0.0374
MTMR9	8	11,142,398	11,180,297	1,686	1	1	0	0	2	0.0056	0.0400
WWP1	8	87,386,280	87,479,122	2,857	2	1	0	0	3	0.0060	0.0409
EFNA5	5	106,716,956	107,006,314	703	0	2	0	0	2	0.0066	0.0415
BEST1	11	61,722,607	61,730,621	1,843	0	2	0	0	2	0.0066	0.0415
SLC16A1	1	113,456,513	113,471,930	1,515	1	1	0	0	2	0.0066	0.0415
SIRPA	20	1,876,094	1,918,214	1,543	0	2	0	0	2	0.0078	0.0420
PSPH	7	56,079,455	56,088,905	694	0	0	0	2	2	0.0078	0.0420
C18orf8	18	21,083,583	21,111,668	2,050	0	2	0	0	2	0.0078	0.0420
IL31RA	5	55,147,399	55,212,948	2,351	0	2	0	0	2	0.0078	0.0420
TMEM39A	3	119,150,828	119,180,921	1,495	1	1	0	0	2	0.0078	0.0420
DDC	7	50,530,929	50,611,783	1,491	0	2	0	0	2	0.0102	0.0428
SS18L1	20	60,718,877	60,754,264	1,231	0	2	0	0	2	0.0102	0.0428
RFX5	1	151,314,662	151,318,796	1,883	0	1	1	0	2	0.0102	0.0428
FUT11	10	75,532,092	75,535,443	1,487	0	2	0	0	2	0.0102	0.0428
IGSF10	3	151,154,477	151,176,497	7,892	0	4	0	0	4	0.0091	0.0428
DYRK1B	19	40,316,355	40,322,507	1,846	0	2	0	0	2	0.0089	0.0428
CREB3L2	7	137,565,222	137,686,451	1,607	0	2	0	0	2	0.0102	0.0428
ZNF770	15	35,273,560	35,275,635	2,076	0	2	0	0	2	0.0102	0.0428
OR10K1	1	158,435,352	158,436,293	942	0	2	0	0	2	0.0089	0.0428
COQ2	4	84,185,352	84,206,067	1,290	0	2	0	0	2	0.0089	0.0428
FBXO43	8	101,146,030	101,157,363	2,143	0	2	0	0	2	0.0102	0.0428
OAS2	12	113,416,414	113,448,289	2,200	0	2	0	0	2	0.0115	0.0439
ARID1A	1	27,022,895	27,107,247	6,934	2	1	0	0	3	0.0109	0.0439
ZNF670	1	247,200,751	247,241,898	1,182	0	2	0	0	2	0.0115	0.0439
TAS2R30	12	11,285,884	11,286,843	960	0	2	0	0	2	0.0115	0.0439
LIPI	21	15,481,314	15,579,244	1,482	0	2	0	0	2	0.0115	0.0439
LCTL	15	66,840,828	66,857,703	1,752	0	2	0	0	2	0.0129	0.0484
SORBS2	4	186,508,781	186,611,725	2,081	0	1	1	0	2	0.0144	0.0494
GLT25D2	1	183,907,895	184,006,491	1,925	0	2	0	0	2	0.0144	0.0494
ATP6V1A	3	113,497,601	113,528,274	1,906	0	2	0	0	2	0.0144	0.0494
C9	5	39,285,301	39,364,566	1,720	0	2	0	0	2	0.0144	0.0494
ACVR2A	2	148,602,722	148,684,843	1,582	1	1	0	0	2	0.0144	0.0494
SARM1	17	26,699,054	26,723,305	2,205	0	2	0	0	2	0.0159	0.0529
FERMT2	14	53,325,095	53,417,286	2,120	0	2	0	0	2	0.0159	0.0529
MBTPS1	16	84,088,054	84,135,388	3,243	0	2	0	0	2	0.0175	0.0539
PHF20	20	34,389,445	34,535,549	3,103	0	2	0	0	2	0.0175	0.0539
MGEA5	10	103,546,208	103,577,779	2,811	0	2	0	0	2	0.0175	0.0539
XPO5	6	43,491,606	43,543,601	3,739	0	2	0	0	2	0.0175	0.0539
LUZP2	11	24,518,790	25,100,204	1,085	0	2	0	0	2	0.0175	0.0539
UBR3	2	170,684,018	170,938,353	5,819	0	3	0	0	3	0.0185	0.0559
MUSK	9	113,431,185	113,563,268	2,662	0	2	0	0	2	0.0192	0.0565
XRCC5	2	216,974,160	217,069,925	2,279	0	2	0	0	2	0.0192	0.0565
ITIH3	3	52,828,820	52,842,697	2,757	0	2	0	0	2	0.0209	0.0583
RUSC1	1	155,291,565	155,300,362	2,741	0	2	0	0	2	0.0209	0.0583
SEMA4B	15	90,744,811	90,771,875	2,566	0	2	0	0	2	0.0209	0.0583
ENPP3	6	131,958,525	132,068,096	2,724	0	2	0	0	2	0.0209	0.0583
6.Mar	5	10,354,011	10,433,796	2,833	0	2	0	0	2	0.0246	0.0618
ZNF880	19	52,873,219	52,888,567	1,746	0	2	0	0	2	0.0246	0.0618
CTNND1	11	57,558,951	57,583,780	2,979	0	2	0	0	2	0.0246	0.0618
TBC1D15	12	72,233,562	72,316,984	2,144	0	2	0	0	2	0.0246	0.0618
FRG1	4	190,862,165	190,884,284	809	0	2	0	0	2	0.0227	0.0618
ZCCHC2	18	60,190,658	60,243,812	3,589	0	2	0	0	2	0.0246	0.0618
PLEKHG5	1	6,527,623	6,579,571	3,336	0	2	0	0	2	0.0246	0.0618
BAZ2B	2	160,176,776	160,335,230	6,643	0	3	0	0	3	0.0241	0.0618
PLXNA3	X	153,688,524	153,701,028	5,740	1	1	0	0	2	0.0285	0.0690

GCFC1	21	34,107,261	34,143,979	2,822	1	1	0	0	2	0.0285	0.0690
TUT1	11	62,342,566	62,359,071	2,771	0	2	0	0	2	0.0285	0.0690
OSBPL8	12	76,749,669	76,881,331	2,758	0	1	1	0	2	0.0305	0.0731
KIAA1107	1	92,633,414	92,649,887	4,093	0	2	0	0	2	0.0326	0.0746
ELTD1	1	79,356,839	79,472,339	2,129	0	2	0	0	2	0.0326	0.0746
TNPO1	5	72,112,565	72,201,237	2,789	1	1	0	0	2	0.0326	0.0746
SWT1	1	185,129,974	185,259,935	2,771	0	2	0	0	2	0.0326	0.0746
NLRP7	19	55,435,108	55,453,079	3,066	0	2	0	0	2	0.0347	0.0786
ZNF827	4	146,686,136	146,859,559	3,282	0	2	0	0	2	0.0369	0.0809
MAP3K13	3	185,146,370	185,200,244	2,949	0	2	0	0	2	0.0369	0.0809
AGTPBP1	9	88,162,024	88,327,448	3,657	0	2	0	0	2	0.0369	0.0809
MAST4	5	65,892,484	66,462,879	7,984	0	3	0	0	3	0.0379	0.0822
SBF1	22	50,885,571	50,913,269	5,842	0	2	0	0	2	0.0392	0.0824
USP53	4	120,160,915	120,214,366	3,278	0	2	0	0	2	0.0392	0.0824
ERAP1	5	96,098,904	96,139,629	2,919	0	2	0	0	2	0.0392	0.0824
BRD8	5	137,475,763	137,514,303	3,812	0	2	0	0	2	0.0415	0.0846
PIK3CA	3	178,916,614	178,952,152	3,283	0	2	0	0	2	0.0415	0.0846
SUPT6H	17	27,000,420	27,028,643	5,321	0	1	1	0	2	0.0415	0.0846
PCDHA11	5	140,248,689	140,389,522	2,862	0	2	0	0	2	0.0438	0.0860
PCDHA13	5	140,261,854	140,389,522	2,865	0	2	0	0	2	0.0438	0.0860
PCDHA7	5	140,213,969	140,389,522	2,826	0	2	0	0	2	0.0438	0.0860
XPOT	12	64,803,816	64,841,911	2,981	0	2	0	0	2	0.0438	0.0860
COL11A1	1	103,343,575	103,573,734	5,685	0	3	0	0	3	0.0460	0.0866
VAV3	1	108,115,953	108,507,491	2,648	0	2	0	0	2	0.0462	0.0866
NLRP8	19	56,459,269	56,499,279	3,183	0	2	0	0	2	0.0462	0.0866
ZNF804B	7	88,389,291	88,966,346	4,062	0	3	0	0	3	0.0460	0.0866
THBS1	15	39,874,059	39,887,570	3,593	1	1	0	0	2	0.0462	0.0866
MET	7	116,339,139	116,436,178	4,249	0	2	0	0	2	0.0487	0.0903
SYNJ2	6	158,402,994	158,517,396	4,595	0	2	0	0	2	0.0512	0.0924
PAM	5	102,201,900	102,364,769	3,021	0	2	0	0	2	0.0512	0.0924
OSMR	5	38,869,147	38,933,546	3,004	0	2	0	0	2	0.0512	0.0924
MDN1	6	90,353,724	90,529,326	17,195	1	3	0	0	4	0.0526	0.0942
ARID2	12	46,123,620	46,298,861	5,588	1	1	0	0	2	0.0537	0.0954
WDR96	10	105,889,897	105,992,004	5,146	0	2	0	0	2	0.0563	0.0991
COBL	7	51,085,248	51,384,330	3,834	0	2	0	0	2	0.0589	0.1028
MYH10	17	8,379,122	8,526,564	6,087	0	2	0	0	2	0.0643	0.1112
PIWIL1	12	130,827,137	130,856,143	2,662	0	2	0	0	2	0.0698	0.1188
XIRP2	2	167,759,993	168,115,873	2,952	0	2	0	0	2	0.0698	0.1188
EVC	4	5,713,108	5,812,764	3,059	0	2	0	0	2	0.0726	0.1226
MIA3	1	222,791,453	222,838,961	5,832	0	2	0	0	2	0.0754	0.1243
NPAT	11	108,029,669	108,093,263	4,352	1	0	0	1	2	0.0754	0.1243
MMRN1	4	90,816,123	90,874,569	3,715	0	2	0	0	2	0.0754	0.1243
ABCA5	17	67,243,678	67,310,556	5,077	0	2	0	0	2	0.0783	0.1261
AKD1	6	109,814,572	109,996,948	5,892	0	2	0	0	2	0.0783	0.1261
COL19A1	6	70,589,460	70,916,978	3,625	0	2	0	0	2	0.0783	0.1261
LRFN5	14	42,355,829	42,373,378	2,172	0	2	0	0	2	0.0812	0.1278
IGF1R	15	99,192,811	99,500,671	4,184	1	1	0	0	2	0.0812	0.1278
ARID1B	6	157,099,064	157,529,025	6,826	0	2	0	0	2	0.0812	0.1278
TNRC6B	22	40,574,140	40,719,245	5,590	0	2	0	0	2	0.0842	0.1294
RGNEF	5	72,980,677	73,236,838	5,336	0	2	0	0	2	0.0842	0.1294
NID2	14	52,472,204	52,535,712	4,212	0	2	0	0	2	0.0842	0.1294
ITSN1	21	35,091,134	35,260,604	5,318	0	2	0	0	2	0.0872	0.1313
CUBN	10	16,866,974	17,171,764	11,136	0	4	0	0	4	0.0873	0.1313
PCDH17	13	58,206,681	58,299,428	3,492	0	2	0	0	2	0.0872	0.1313
CNTNAP4	16	76,311,561	76,592,580	4,018	0	2	0	0	2	0.0902	0.1318
ZFYVE26	14	68,215,153	68,282,680	7,780	0	2	0	0	2	0.0902	0.1318
CDH18	5	19,483,458	19,839,095	1,761	0	2	0	0	2	0.0902	0.1318
CNTRL	9	123,850,605	123,939,464	7,142	0	2	0	0	2	0.0902	0.1318
LAMC1	1	182,992,852	183,111,925	4,938	0	2	0	0	2	0.0963	0.1397
WDR52	3	113,010,404	113,152,511	5,697	0	2	0	0	2	0.0994	0.1432
CDH7	18	63,430,079	63,548,130	2,398	0	2	0	0	2	0.1026	0.1457
CEP192	18	12,999,424	13,124,769	7,786	0	2	0	0	2	0.1026	0.1457
CNTN1	12	41,302,235	41,463,837	3,145	0	2	0	0	2	0.1057	0.1482
TRPV5	7	142,605,680	142,630,556	2,246	0	2	0	0	2	0.1057	0.1482
COL24A1	1	86,196,229	86,622,079	5,381	0	2	0	0	2	0.1089	0.1516
FREM3	4	144,498,764	144,621,828	6,448	0	2	0	0	2	0.1154	0.1595
ZNF536	19	30,934,470	31,048,071	3,915	0	2	0	0	2	0.1186	0.1618
MKI67	10	129,897,454	129,923,931	9,823	0	3	0	0	3	0.1184	0.1618
FSIP2	2	186,603,355	186,697,953	21,079	1	4	0	0	5	0.1240	0.1680
EPHA6	3	96,533,468	97,467,545	3,461	0	2	0	0	2	0.1319	0.1776
F5	1	169,483,551	169,555,624	6,771	0	2	0	0	2	0.1386	0.1842
SCN3A	2	165,946,660	166,032,904	6,103	1	1	0	0	2	0.1386	0.1842
VPS13B	8	100,026,017	100,887,894	12,309	0	3	0	0	3	0.1406	0.1857
COL22A1	8	139,601,496	139,895,415	5,133	0	2	0	0	2	0.1489	0.1929
TRPS1	8	116,426,251	116,635,864	3,905	0	2	0	0	2	0.1489	0.1929
MYH9	22	36,678,714	36,745,281	6,039	0	2	0	0	2	0.1489	0.1929
MYOM2	8	1,998,881	2,092,905	4,538	0	2	0	0	2	0.1523	0.1937
SYCP2	20	58,439,366	58,497,468	4,761	0	2	0	0	2	0.1523	0.1937
ITPR2	12	26,492,330	26,985,714	8,330	0	2	0	0	2	0.1523	0.1937
ATR	3	142,168,271	142,297,546	8,119	0	2	0	0	2	0.1558	0.1957
ZFC3H1	12	72,004,208	72,057,390	6,106	0	2	0	0	2	0.1558	0.1957
NOTCH3	19	15,271,473	15,311,716	7,094	0	2	0	0	2	0.1593	0.1989
HMCN1	1	185,703,912	186,159,010	17,332	0	4	0	0	4	0.1614	0.2002
SCN7A	2	167,262,090	167,334,206	5,141	0	2	0	0	2	0.1628	0.2008
ATP10B	5	159,992,460	160,115,081	4,470	0	2	0	0	2	0.1698	0.2082

DCC	18	49,867,158	51,057,023	4,456	0	2	0	0	2	0.1734	0.2088
TRANK1	3	36,869,766	36,986,300	8,866	0	2	0	0	2	0.1734	0.2088
BPTF	17	65,821,841	65,978,404	8,828	1	1	0	0	2	0.1734	0.2088
MAP1B	5	71,403,359	71,501,066	7,431	0	2	0	0	2	0.1948	0.2320
ITPR1	3	4,558,176	4,887,909	8,464	0	2	0	0	2	0.1948	0.2320
NCKAP5	2	133,430,862	134,275,097	5,798	0	2	0	0	2	0.2166	0.2564
AHNAK	11	62,284,216	62,303,570	17,681	0	3	0	0	3	0.2192	0.2577
PREX2	8	68,864,630	69,143,613	4,977	0	2	0	0	2	0.2202	0.2577
ODZ3	4	183,245,174	183,721,504	8,204	0	2	0	0	2	0.2239	0.2605
PRUNE2	9	79,229,486	79,520,879	9,339	1	1	0	0	2	0.2495	0.2888
C12orf51	12	112,600,191	112,757,291	13,033	1	1	0	0	2	0.2532	0.2914
DCHS1	11	6,643,010	6,662,844	9,973	0	2	0	0	2	0.2569	0.2940
HERC1	15	63,901,280	64,067,822	14,890	0	1	1	0	2	0.2606	0.2949
PCDH7	4	30,723,045	31,144,471	3,752	0	2	0	0	2	0.2606	0.2949
RELN	7	103,113,259	103,629,803	10,639	0	2	0	0	2	0.2716	0.3041
MLL	11	118,307,228	118,392,887	12,050	1	1	0	0	2	0.2716	0.3041
TRRAP	7	98,478,774	98,609,978	11,769	0	2	0	0	2	0.2936	0.3270
RNF213	17	78,237,481	78,367,298	16,039	0	2	0	0	2	0.2973	0.3293
FAT4	4	126,237,567	126,412,923	15,010	1	2	0	0	3	0.2990	0.3294
ABCA13	7	48,211,081	48,685,108	15,421	0	3	0	0	3	0.3071	0.3365
LAMA2	6	129,204,391	129,837,492	9,625	0	2	0	0	2	0.3338	0.3638
SPTA1	1	158,581,054	158,656,307	7,464	0	1	1	0	2	0.3374	0.3658
DNAH6	2	84,744,951	85,046,532	12,777	0	2	0	0	2	0.3589	0.3871
LRP2	2	169,985,173	170,218,909	14,280	1	1	0	0	2	0.4012	0.4305
ANK3	10	61,802,449	62,149,296	13,302	0	2	0	0	2	0.4422	0.4707
LRP1B	2	140,990,755	142,888,298	14,160	0	3	0	0	3	0.4433	0.4707
CSMD1	8	2,796,107	4,851,938	10,971	0	2	0	0	2	0.4488	0.4742
EYS	6	64,430,492	66,205,303	9,591	1	1	0	0	2	0.4718	0.4959
FBN2	5	127,595,147	127,873,296	8,995	0	2	0	0	2	0.4783	0.5001
MLL3	7	151,833,917	152,132,871	14,968	0	2	0	0	2	0.5160	0.5369
USH2A	1	215,799,123	216,595,678	15,889	0	2	0	0	2	0.5460	0.5652
DNAH7	2	196,602,645	196,933,435	12,331	0	2	0	0	2	0.5747	0.5920
DYNC2H1	11	102,980,304	103,349,981	13,301	0	2	0	0	2	0.5858	0.6004
CSMD3	8	113,237,000	114,449,083	11,404	0	2	0	1	3	0.5967	0.6085
MUC16	19	8,959,608	9,091,814	43,856	1	2	0	0	3	0.6641	0.6739
PCLO	7	82,387,891	82,791,908	15,525	0	3	0	0	3	0.6773	0.6839
GPR98	5	89,854,713	90,459,717	19,277	0	2	0	0	2	0.7307	0.7342
TTN	2	179,391,739	179,669,369	101,512	0	4	1	0	5	0.9935	0.9935

Supplementary Table 8: Mutation list of the significantly mutated genes

Gene	Sample	Chr	Genomic position	Genotype	RNA_id	Exon	aa position	Type	Reference aa	Mutant aa
TP53	HB7	17	7,578,555	CT	NM_000546	4	-	splice site	-	-
TP53	HB8	17	7,577,521	CT	NM_000546	6	254	nonsynonymous	I	V
TP53	HB8	17	7,577,523	GT	NM_000546	6	253	nonsynonymous	T	N
TP53	HB9	17	7,578,393	AT	NM_000546	4	179	nonsynonymous	H	N
TP53	HC11	17	7,578,479	CG	NM_000546	4	151	nonsynonymous	P	A
TP53	HB10	17	7,578,443	AT	NM_000546	4	163	nonsynonymous	Y	N
TP53	HB11	17	7,578,403	AC	NM_000546	4	176	nonsynonymous	C	F
TP53	HC9	17	7,578,260	AC	NM_000546	5	197	nonsynonymous	V	L
TP53	HC10	17	7,579,699	AC	NM_000546	2	-	splice site	-	-
TP53	HB1	17	7,578,208	CT	NM_000546	5	214	nonsynonymous	H	R
TP53	HC2	17	7,577,117	AG	NM_000546	7	274	nonsynonymous	V	N
TP53	HB5	17	7,579,355	AC	NM_000546	3	111	nonsynonymous	L	N
TP53	HC7-1	17	7,578,515	CT	NM_000546	4	139	nonsynonymous	K	E
TP53	HC1	17	7,576,927	AC	NM_000546	8	-	splice site	-	-
ERRFI1	HB6	1	8,074,109	AC	NM_018948	3	184	nonsynonymous	E	Stop
ERRFI1	HC7-1	1	8,074,313	AT	NM_018948	3	116	nonsynonymous	K	Stop
ERRFI1	HC7-1	1	8,073,922	+A	NM_018948	3	246	coding indel	-	-
ERRFI1	validation study	1	8,074,296	AT	NM_018948	3	121	nonsynonymous	N	K
ERRFI1	validation study	1	8,074,111	AT	NM_018948	3	183	nonsynonymous	L	Stop
ZIC3	HC11	X	136,649,473	AG	NM_003413	1	208	nonsynonymous	D	N
ZIC3	NBNC2	X	136,649,094	AG	NM_003413	1	82	nonsynonymous	A	T
ZIC3	NBNC2	X	136,649,095	AC	NM_003413	1	82	nonsynonymous	A	D
ZIC3	validation study	X	136,649,223	CT	NM_003413	1	125	nonsynonymous	S	P
ZIC3	validation study	X	136,649,386	GT	NM_003413	1	179	nonsynonymous	V	G
ZIC3	validation study	X	136,649,473	AG	NM_003413	1	208	nonsynonymous	D	N
ZIC3	validation study	X	136,649,735	AC	NM_003413	1	295	nonsynonymous	H	Q
CTNNB1	HB9	3	41,266,113	CT	NM_001904	2	37	nonsynonymous	S	F
CTNNB1	HC11	3	41,266,124	AG	NM_001904	2	41	nonsynonymous	T	A
CTNNB1	NBNC1	3	41,266,124	AG	NM_001904	2	41	nonsynonymous	T	A
GXYLT1	HB7	12	42,481,592	CT	NM_173601	8	440	nonsynonymous	K	R
GXYLT1	HB7	12	42,512,938	AT	NM_173601	3	117	nonsynonymous	H	L
GXYLT1	HB9	12	42,481,688	AT	NM_173601	8	408	nonsynonymous	K	I
GXYLT1	validation study	12	42,503,486	CT	NM_173601	3	165	nonsynonymous	N	S
OTOP1	HB11	4	4,204,211	AG	NM_177998	4	232	nonsynonymous	R	W
OTOP1	HB11	4	4,204,225	-TTGAG	NM_177998	4	227	coding indel	-	-
OTOP1	HB1	4	4,199,681	AG	NM_177998	5	294	nonsynonymous	R	C
OTOP1	validation study	4	4,228,490	-GA	NM_177998	1	34	coding indel	-	-
ALB	HC3-1	4	74,283,362	-AAG	NM_000477	11	468	coding indel	-	-
ALB	HB2	4	74,282,022	-AA	NM_000477	10	414	coding indel	-	-
ALB	HB2	4	74,285,285	-ATGG	NM_000477	13	572	coding indel	-	-
ALB	validation study	4	74,277,844	AT	NM_000477	14	-	splice site	-	-
ALB	validation study	4	74,274,520	-AAGT	NM_000477	4	160	coding indel	-	-
ALB	validation study	4	74,276,079	-A	NM_000477	6	222	coding indel	-	-
ALB	validation study	4	74,283,257	-C	NM_000477	11	433	coding indel	-	-
ATM	HC1	11	108,142,043	AT	NM_000051	19	996	nonsynonymous	H	L
ATM	HC3-1	11	108,218,045	AC	NM_000051	58	2,875	nonsynonymous	N	T
ATM	HC3-1	11	108,235,819	AG	NM_000051	61	2,954	nonsynonymous	Y	C
ATM	HC5	11	108,200,958	AC	NM_000051	49	2,442	nonsynonymous	Q	P
ATM	HC3-2	11	108,122,587	-TG	NM_000051	10	544	coding indel	-	-
ATM	validation study	11	108,098,352	AG	NM_000051	1	1	nonsynonymous	M	V
ATM	validation study	11	108,153,482	CG	NM_000051	24	1,208	nonsynonymous	D	H
ATM	validation study	11	108,235,810	AT	NM_000051	61	2,951	nonsynonymous	V	D
ATM	validation study	11	108,165,652	AT	NM_000051	62	-	splice site	-	-
ATM	validation study	11	108,114,682	-TG	NM_000051	5	167	coding indel	-	-
ATM	validation study	11	108,151,790	-AT	NM_000051	23	1,157	coding indel	-	-
ZNF226	HC11	19	44,680,766	GT	NM_001032372	4	451	nonsynonymous	E	Stop
ZNF226	HC11	19	44,681,808	-A	NM_001032372	4	798	coding indel	-	-
ZNF226	HC7-1	19	44,680,755	AG	NM_001032372	4	447	nonsynonymous	Y	C
ZNF226	validation study	19	44,680,551	AT	NM_001032373	4	379	nonsynonymous	Q	L
ZNF226	validation study	19	44,680,766	GT	NM_001032373	4	451	nonsynonymous	E	Stop
ZNF226	validation study	19	44,681,310	AC	NM_001032373	4	632	nonsynonymous	A	D
ZNF226	validation study	19	44,681,399	-A	NM_001032373	4	662	coding indel	-	-
USP25	HC9	21	17,197,356	CT	NM_013396	12	427	nonsynonymous	T	M
USP25	HC9	21	17,203,755	AC	NM_013396	16	600	nonsynonymous	E	D
USP25	HC3-2	21	17,250,777	-C	NM_013396	24	1,051	coding indel	-	-
WWP1	HC7-1	8	87,450,862	GT	NM_007013	15	600	nonsynonymous	F	V
WWP1	HB1	8	87,439,942	-C	NM_007013	9	410	coding indel	-	-
WWP1	HC2	8	87,393,800	+T	NM_007013	3	92	coding indel	-	-
WWP1	validation study	8	87,410,649	AT	NM_007013	6	138	nonsynonymous	V	E
WWP1	validation study	8	87,424,082	CA	NM_007013	9	347	nonsynonymous	N	T
WWP1	validation study	8	87,454,897	AT	NM_007013	18	630	nonsynonymous	L	I
WWP1	validation study	8	87,473,551	AT	NM_007013	23	-	splice site	-	-
WWP1	validation study	8	87,443,885	AG	NM_007013	14	505	nonsynonymous	R	K
IGSF10	HC11	3	151,166,940	AG	NM_178822	4	277	nonsynonymous	F	L
IGSF10	HC11	3	151,166,974	AT	NM_178822	4	265	nonsynonymous	K	N
IGSF10	HB11	3	151,166,144	AT	NM_178822	4	542	nonsynonymous	D	V
IGSF10	validation study	3	151,162,983	AC	NM_178822	4	1,596	nonsynonymous	A	S
IGSF10	validation study	3	151,171,523	CT	NM_178822	3	122	nonsynonymous	K	E
IGSF10	validation study	3	151,174,890	AT	NM_178822	2	83	nonsynonymous	L	Q
IGSF10	validation study	3	151,174,891	GT	NM_178822	2	83	nonsynonymous	L	M
IGSF10	validation study	3	151,176,344	AG	NM_178822	1	52	nonsynonymous	P	S
ARID1A	HC3-1	1	27,094,429	AG	NM_006015	11	1,046	nonsynonymous	R	K
ARID1A	HB1	1	27,059,259	-TCAG	NM_006015	4	632	coding indel	-	-
ARID1A	HC6	1	27,087,445	-G	NM_006015	5	673	coding indel	-	-
ARID1A	validation study	1	27,099,051	AG	NM_006015	13	1,156	nonsynonymous	E	G
ARID1A	validation study	1	27,100,114	GT	NM_006015	16	1,304	nonsynonymous	A	S
ARID1A	validation study	1	27,100,202	AGT	NM_006015	16	1,333	nonsynonymous	Q	L, R
ARID1A	validation study	1	27,106,985	CT	NM_006015	20	2,199	nonsynonymous	L	P
ARID1A	validation study	1	27,107,125	GT	NM_006015	20	2,246	nonsynonymous	E	Stop
ARID1A	validation study	1	27,107,151	CG	NM_006015	20	2,254	nonsynonymous	Y	Stop
ARID1A	validation study	1	27,089,777	CG	NM_006015	19	-	splice site	-	-
ARID1A	validation study	1	27,057,726	-CAG	NM_006015	3	478	coding indel	-	-
ARID1A	validation study	1	27,101,701	-A	NM_006015	18	1,661	coding indel	-	-
ARID1A	validation study	1	27,102,168	-A	NM_006015	19	1,698	coding indel	-	-
ARID1A	validation study	1	27,107,123	-C	NM_006015	20	2,245	coding indel	-	-
ARID1A	validation study	1	27,101,407	+C	NM_006015	18	1,563	coding indel	-	-
UBR3	HC9	2	170,802,827	AC	NM_172070	21	945	nonsynonymous	E	D
UBR3	HC2	2	170,753,095	AG	NM_172070	8	945	nonsynonymous	E	D
UBR3	HC3-2	2	170,863,649	AC	NM_172070	28	439	nonsynonymous	M	V
UBR3	validation study	2	170,806,424	AG	NM_172070	23	1,393	nonsynonymous	Q	H
BAZ2B	HB9	2	160,294,830	CT	NM_013450	6	1,132	nonsynonymous	I	V
BAZ2B	NBNC2	2	160,303,439	CT	NM_013450	4	426	nonsynonymous	E	G
BAZ2B	HB1	2	160,189,119	AG	NM_013450	32	184	nonsynonymous	I	V
BAZ2B	validation study	2	160,205,256	-T	NM_013450	28	1,959	nonsynonymous	Y	H
BAZ2B	validation study	2	160,295,136	-T	NM_013450	6	324	coding indel	-	-

Supplementary Table 9: **Gene set enrichment (GSE) analysis for genes with nonsense, coding indel and splice-site mutations**

Category	Term	Count	%	*P-value*	List Total	Pop Hits	Pop Total	Fold Enrichment	*q-value*
SP_PIR_KEYWORDS	phosphoprotein	162	55.1	6.92E-10	292	7263	19235	1.469292662	0.00000023
INTERPRO	IPR013032:EGF-like region, conserved site	19	6.5	8.02E-07	260	293	16659	4.154909425	0.00049
INTERPRO	IPR000742:EGF-like, type 3	15	5.1	2.14E-06	260	194	16659	4.954103886	0.0006
INTERPRO	IPR006210:EGF-like	15	5.1	3.25E-06	260	201	16659	4.781572905	0.0007
SP_PIR_KEYWORDS	egf-like domain	15	5.1	1.15E-05	292	230	19235	4.296083979	0.0019
UP_SEQ_FEATURE	domain:EGF-like 1	12	4.1	2.25E-06	292	120	19113	6.545547945	0.0039
SMART	SM00181:EGF	15	5.1	3.00E-05	175	201	9079	3.871641791	0.0043
SP_PIR_KEYWORDS	polymorphism	208	70.7	4.75E-05	292	11550	19235	1.18628951	0.0052
SP_PIR_KEYWORDS	calcium	28	9.5	9.42E-05	292	803	19235	2.2969515	0.0078
UP_SEQ_FEATURE	sequence variant	218	74.1	1.06E-05	292	11992	19113	1.189901144	0.0092
SP_PIR_KEYWORDS	**bromodomain**	6	2.0	2.89E-04	292	39	19235	10.13435195	0.019
SP_PIR_KEYWORDS	**chromatin regulator**	12	4.1	4.17E-04	292	213	19235	3.711171136	0.023
INTERPRO	IPR006209:EGF	10	3.4	1.66E-04	260	127	16659	5.045124167	0.025
SP_PIR_KEYWORDS	disease mutation	42	14.3	5.45E-04	292	1591	19235	1.738955426	0.026
INTERPRO	**IPR001487:Bromodomain**	6	2.0	3.69E-04	260	40	16659	9.610961538	0.044
SP_PIR_KEYWORDS	tumor suppressor	9	3.1	0.001157126	292	137	19235	4.327442256	0.047

q-value was obtained by Benjamini and Hochberg's FDR method

Supplementary Table 10: Mutation list of the chromatin regulator genes

Gene	Sample	Chr	Genomic position	Genotype	RNA_id	Exon	aa position	Type	Reference aa	Mutant aa
ARID1B	HB9	6	157,528,540	CG	NM_020732	20	2,089	nonsynonymous	L	V
ARID1B	HC5	6	157,454,194	AG	NM_020732	8	802	nonsynonymous	M	V
ARID1B	validation study	6	157,099,508	CG	NM_020732	1	149	nonsynonymous	Q	E
ARID1B	validation study	6	157,100,285	CT	NM_020732	1	408	nonsynonymous	Q	Stop
ARID1B	validation study	6	157,099,981	-GGC	NM_020732	1	306	coding indel	-	-
ARID1B	validation study	6	157,100,046	-AGC	NM_020732	1	328	coding indel	-	-
ARID1B	validation study	6	157,100,137	-G	NM_020732	1	358	coding indel	-	-
ARID1B	validation study	6	157,100,222	+C	NM_020732	1	387	coding indel	-	-
ARID1B	validation study	6	157,100,309	-C	NM_020732	1	416	coding indel	-	-
ARID1B	validation study	6	157,505,455	-T	NM_020732	13	1,146	coding indel	-	-
ARID2	HC9	12	46,244,449	+T	NM_152641	15	848	coding indel	-	-
ARID2	HB3	12	46,243,834	AC	NM_152641	15	643	nonsynonymous	S	Stop
ARID2	validation study	12	46,123,699	AT	NM_152641	1	27	nonsynonymous	H	L
ARID2	validation study	12	46,231,200	CG	NM_152641	9	374	nonsynonymous	G	R
ARID2	validation study	12	46,244,897	GT	NM_152641	15	997	nonsynonymous	Q	H
ARID2	validation study	12	46,285,641	GT	NM_152641	17	1,667	nonsynonymous	C	W
ARID2	validation study	12	46,123,699	-C	NM_152641	1	27	coding indel	-	-
ARID2	validation study	12	46,285,661	-T	NM_152641	17	1,674	coding indel	-	-
ARID2	validation study	12	46,298,726	-A	NM_152641	21	1,791	coding indel	-	-
MLL	HC11	11	118,344,186	-C	NM_005933	3	771	coding indel	-	-
MLL	HC7-2	11	118,342,410	CG	NM_005933	3	179	nonsynonymous	P	R
MLL	validation study	11	118,307,445	AG	NM_005933	1	73	nonsynonymous	G	E
MLL	validation study	11	118,352,504	AG	NM_005933	7	1,237	nonsynonymous	V	M
MLL3	HC3-1	7	151,853,128	CT	NM_170606	46	3,943	nonsynonymous	G	R
MLL3	HC4	7	151,859,936	CT	NM_170606	43	3,576	nonsynonymous	I	V
MLL3	validation study	7	151,921,595	CT	NM_170606	19	1,028	nonsynonymous	C	Y
MLL3	validation study	7	151,932,906	CG	NM_170606	16	922	nonsynonymous	P	R
MLL3	validation study	7	151,932,946	CT	NM_170606	16	909	nonsynonymous	R	G
MLL3	validation study	7	151,962,189	AGT	NM_170606	8	373	nonsynonymous	I	T, K
MLL3	validation study	7	151,879,611	-T	NM_170606	36	1,778	coding indel	-	-
BRD8	HB1	5	137,496,591	AC	NM_139199	18	806	nonsynonymous	E	Stop
BRD8	HB6	5	137,496,741	CT	NM_139199	18	756	nonsynonymous	T	A
BRD8	validation study	5	137,497,513	CT	NM_139199	17	741	nonsynonymous	A	T
BPTF	HB10	17	65,907,062	AG	NM_004459	13	1,147	nonsynonymous	D	G
BPTF	HC3-1	17	65,871,775	-C	NM_004459	5	656	coding indel	-	-
BPTF	validation study	17	65,907,062	AG	NM_004459	13	1,147	nonsynonymous	D	G
BPTF	validation study	17	65,919,058	CT	NM_004459	16	2,013	nonsynonymous	V	A
HIST1H4B	HB11	6	26,027,251	+A	NM_003544	1	77	coding indel	-	-
HIST1H4B	HC4	6	26,027,170	AT	NM_003544	1	104	nonsynonymous	Stop	L
HIST1H4B	validation study	6	26,027,170	AT	NM_003544	1	104	nonsynonymous	Stop	K
BRE	HC11	2	28,248,273	-A	NM_004899	4	161	coding indel	-	-
BRE	HC4	2	28,464,244	CG	NM_004899	8	279	nonsynonymous	L	V

Supplementary Table 11: Association between mutations in chromatin regulators and clinical factors

data	Tumor size	Edmondson grade	Portal vein invasion
Point mutation and indel in the chromatin regulator genes	0.67	0.75	1.00
Point mutation, indel and CNA in the chromatin regulator genes	0.78	0.62	1.00
Point mutation and indel in the *ARID* genes (*ARID1A*, *ARID2*, and *ARID1B*)	0.56	0.94	0.36
Point mutation, indel and CNA in the *ARID* genes (*ARID1A*, *ARID2*, and *ARID1B*)	0.80	0.83	0.38
Point mutation and indel in *ARID1A* gene	0.44	0.20	0.59
Point mutation, indel and CNA in *ARID1A* gene	0.39	0.67	0.36

Supplementary Table 11 (continued)

data	Hepatic vein invasion	Liver fibrosis
Point mutation and indel in the chromatin regulator genes	1.00	0.061
Point mutation, indel and CNA in the chromatin regulator genes	0.61	0.026
Point mutation and indel in the *ARID* genes (*ARID1A*, *ARID2*, and *ARID1B*)	0.27	0.16
Point mutation, indel and CNA in the *ARID* genes (*ARID1A*, *ARID2*, and *ARID1B*)	0.27	0.0086
Point mutation and indel in *ARID1A* gene	0.39	0.63
Point mutation, indel and CNA in *ARID1A* gene	0.042	0.16

P-values for Tumor size, Edmondson grade, and Liver fibrosis were calculated by the Pearson's correlation test.

P-values for Portal vein invasion and Hepatic vein invasion were obtained by the Fisher's exact test.

Supplementary Table 12: Recurrent somatic mutations of non-coding regions (n ≥ 3)

Symbol	Type	Chr	Start	Last	SNV	indel	Length (bp)	*q*-value
ANKRD30BL	ncRNA	2	132,905,164	133,015,542	8	6	110,378	0
AF146191.4	lincRNA	4	190,742,266	190,838,477	12	0	96,211	8.10463E-15
U2	snRNA	10	103,124,602	103,124,792	4	0	190	1.59648E-07
RP5-875O13.1	lincRNA	1	16,860,381	16,862,144	3	3	1,763	1.69297E-07
AD000090.1	miRNA	19	36,066,627	36,066,721	3	0	94	2.41516E-06
AC020926.1	lincRNA	16	20,685,853	20,693,496	6	0	7,643	7.84436E-06
HMHA1	UTR5	19	1,067,174	1,086,627	3	0	19,453	1.19856E-05
RNU2-2	snRNA	11	62,609,091	62,609,281	2	1	190	1.19856E-05
LOC100130581	PRO	17	41,447,213	41,466,266	2	2	19,053	3.12021E-05
CTBP2	UTR3	10	126,676,418	126,849,103	2	2	172,685	5.73097E-05
NEAT1	lincRNA	11	65,190,269	65,213,011	9	0	22,742	9.66691E-05
ANKRD36BP2	lincRNA	2	89,065,385	89,106,126	7	0	40,741	9.91213E-05
PRR15	UTR3	7	29,603,427	29,606,911	3	0	3,484	0.000193272
DKFZp434L192	PRO	7	56,563,916	56,564,977	4	0	1,061	0.000357403
PHF1	PRO	6	33,378,773	33,384,230	3	0	5,457	0.000391151
HLA-DRB5	PRO	6	32,485,154	32,498,006	4	0	12,852	0.000664692
MED1	UTR3	17	37,560,538	37,607,527	4	0	46,989	0.000664692
AC079305.1	lincRNA	2	178,107,103	178,111,509	2	2	4,406	0.000724295
PRSS3	PRO	9	33,795,559	33,799,229	3	0	3,670	0.000745744
CTD-2384A14.1	lincRNA	14	29,400,776	29,437,598	3	0	36,822	0.000745744
LOC100130581	ncRNA	17	41,453,296	41,466,266	3	1	12,970	0.000813248
CTB-113P19.1	lincRNA	5	151,056,506	151,067,471	3	1	10,965	0.001123863
NIP7	UTR3	16	69,373,415	69,377,013	3	0	3,598	0.001307338
SMG1	UTR3	16	18,816,175	18,937,726	3	0	121,551	0.001768832
RIBC1	UTR5	X	53,449,839	53,456,776	3	0	6,937	0.001768832
IL22RA2	UTR3	6	137,464,957	137,494,785	3	0	29,828	0.002785282
URB1	UTR3	21	33,683,330	33,765,312	3	1	81,982	0.003785807
SRSF3	UTR3	6	36,562,090	36,572,244	1	2	10,154	0.003791246
LOC150622	PRO	2	6,072,819	6,120,350	3	0	47,531	0.003791246
TRY6	PRO	7	142,478,757	142,482,399	3	0	3,642	0.006317702
LOC121952	ncRNA	13	103,532,449	103,548,383	3	0	15,934	0.006317702
HCG22	lincRNA	6	31,021,984	31,027,653	2	2	5,669	0.006388021
PIK3R3	UTR3	1	46,505,812	46,598,380	3	0	92,568	0.007245836
KIAA1430	UTR3	4	186,080,819	186,125,182	3	0	44,363	0.007507514
PTAR1	UTR3	9	72,324,438	72,374,876	5	0	50,438	0.007507514
LOC220980	lincRNA	10	45,306,472	45,455,137	3	0	148,665	0.007994232
SCD	UTR3	10	102,106,772	102,124,588	3	0	17,816	0.009002124
SMAD6	ncRNA	15	66,994,674	67,074,337	3	0	79,663	0.009002124
SRSF3	ncRNA	6	36,562,090	36,572,244	1	2	10,154	0.009522649
PRPF40A	UTR3	2	153,508,107	153,573,975	4	0	65,868	0.009808589
MAGI3	UTR3	1	113,933,475	114,228,545	2	1	295,070	0.009808589
RP11-622O11.2	lincRNA	8	126,953,373	126,963,433	3	0	10,060	0.011986707
FLJ39653	lincRNA	4	16,228,286	16,259,810	3	0	31,524	0.012564096
NFIA	UTR3	1	61,547,980	61,928,460	4	0	380,480	0.017052264
EDNRB	UTR3	13	78,469,616	78,492,966	3	0	23,350	0.017648691
AP000459.4	lincRNA	21	24,733,426	24,757,182	3	0	23,756	0.018307238
HCG2P7	ncRNA	6	29,866,808	29,870,431	3	0	3,623	0.01897249
AC011477.2	lincRNA	19	19,945,686	20,008,579	3	0	62,893	0.019644199
CSGALNACT1	ncRNA	8	19,261,672	19,540,261	3	0	278,589	0.021434704
ATXN7L3B	UTR3	12	74,931,551	74,935,232	3	0	3,681	0.024476649
PTGER3	UTR3	1	71,471,537	71,513,491	4	0	41,954	0.028065538
AFF4	UTR3	5	132,211,071	132,299,354	3	0	88,283	0.029118099
RORA	UTR3	15	60,780,483	60,884,707	3	1	104,224	0.029118099
MYEF2	UTR3	15	48,431,629	48,470,558	3	0	38,929	0.029118099
FAM126B	UTR3	2	201,838,441	201,936,392	3	0	97,951	0.029384429
RLIM	UTR3	X	73,802,811	73,834,461	3	0	31,650	0.032953863
ADCY2	UTR3	5	7,396,343	7,830,194	3	0	433,851	0.032953863
KLF6	ncRNA	10	3,818,188	3,827,473	3	0	9,285	0.032953863
ZC3H6	UTR3	2	113,033,178	113,097,640	2	1	64,462	0.03371726

LOC400940	ncRNA	2	6,122,110	6,128,364	3	0	6,254	0.045347117
AP000476.1	lincRNA	21	25,801,054	25,862,621	3	0	61,567	0.045347117
PAM	ncRNA	5	102,201,527	102,366,808	3	0	165,281	0.046140706

UTR3; 3'UTR, UTR5; 5'UTR, PRO; promoter

Supplementary Table 13: HBV integration sites (IS) identified in HBV-related HCC genomes

Sample	Candidate name	Chr1	Pos1	Chr2	Pos2	Number of support read-pair	Affected gene
HB1	IS1	HBV	1,735	5	88,432,442	14	
	IS2	HBV	687	5	88,754,909	18	
HB2	IS1	HBV	1,783	5	1,275,381	3	***TERT***
	IS2	HBV	1,835	5	1,275,390	8	***TERT***
	IS3	HBV	1,879	5	173,125,603	9	
HB3	IS1	HBV	1,470	6	68,664,024	11	
	IS2	HBV	1,771	8	39,200,450	14	*ADAM5P*
	IS3	HBV	1,808	4	79,048,012	19	*FRAS1*
	IS4	HBV	1,818	4	79,048,008	15	*FRAS1*
HB4	-	-	-	-	-	-	
HB5	IS1	HBV	2,141	3	24,499,331	11	*THRB*
HB6	-	-	-	-	-	-	
HB7	IS1	HBV	1,401	7	52,802,419	11	
	IS2	HBV	2,622	7	52,802,387	5	
	IS3	HBV	1,361	3	121.439,129	3	*HNF1A*
HB8	-	-	-	-	-	-	
HB9	IS1	HBV	1,854	5	1,295,172	26	***TERT***
	IS2	HBV	1,726	5	1,295,200	13	***TERT***
	IS3	HBV	1895	7	70,267,023	4	
HB10	IS1	HBV	1,828	5	1,293,404	37	***TERT***
	IS2	HBV	2,236	5	1,293,416	25	***TERT***
HB11	IS1	HBV	1,411	5	1,310,429	74	***TERT-CLPTM1L***[a]
	IS2	HBV	1,675	17	15,412,626	34	*FAM18B2*
	IS3	HBV	2,716	20	30,502,092	18	*TTLL9*
	IS4	HBV	1,814	9	24,396,870	14	
	IS5	HBV	1,846	4	151,954,261	4	

Integration site was not identified in HB4, HB6, and HB8.

a; Intergenic region between *TERT* and *CLPTM1L*

Supplementary Table 14: Association between etiological factors and mutation pattern

Data set	Disctance between MC tumors[j]	HBV vs. HCV vs. NBNC	Liver fibrosis	Alcohol drinking	Edmondson grade[k]	Age
Point mutation (6 category)[a]	0.0073	0.3812	0.9328	0.0051	0.5866	0.7804
Point mutation[b]	0.0005	0.0909	0.5716	0.0283	0.5324	0.2423
Point mutation+indel[c]	0.0008	0.0685	0.5252	0.0408	0.4137	0.2375
Point mutation+indel+Rearrangement[d]	0.0001	0.0682	0.3835	0.0128	0.4149	0.2716
Point mutation+indel length[e]	0.0003	0.1147	0.4427	0.0049	0.6310	0.4401
Point mutation+indel length+Rearrangement[f]	0.0002	0.1891	0.5016	0.0055	0.8440	0.6076
Point mutation+indel length+ns_rate[g]	0.0006	0.0216	0.2497	0.0256	0.5051	0.2610
Point mutation+indel+Rearrangement+ns_rate[h]	0.0003	0.0146	0.2151	0.0892	0.3218	0.2058
Point mutation+total_num[i]	0.0006	0.0877	0.1529	0.0398	0.5388	0.2494

Supplementary Table 14 (continued)

Data set	Tumor size	Platelet count ($x10^4$)/mm^2	Portal vein invasion	Hepatic vein invasion	Serum AFP level[l]	Multiple liver nodules
Point mutation (6 category)[a]	0.0411	0.3601	0.5337	0.8128	0.0681	0.0118
Point mutation[b]	0.0784	0.5838	0.2971	0.8416	0.1420	0.0161
Point mutation+indel[c]	0.1066	0.7748	0.4683	0.8277	0.2155	0.0168
Point mutation+indel+Rearrangement[d]	0.0939	0.7317	0.3926	0.8394	0.2707	0.0135
Point mutation+indel length[e]	0.1029	0.6230	0.3990	0.7856	0.2601	0.0173
Point mutation+indel length+Rearrangement[f]	0.0810	0.6289	0.4068	0.7854	0.2805	0.0204
Point mutation+indel length+ns_rate[g]	0.2538	0.4773	0.4026	0.6012	0.1704	0.0186
Point mutation+indel+Rearrangement+ns_rate[h]	0.1576	0.8231	0.4465	0.7129	0.2054	0.0108
Point mutation+total_num[i]	0.0475	0.7064	0.1603	0.8449	0.4671	0.0119

a; number of substitutions without considering CpG. b; number of substitutions with CpG. c; number of substitutions and indels. d; number of substitutions, indels and rearrangements.

e; number of substitutions and length distribution of indels (number of 1bp, 2bp, 3bp, 5bp and >=5bp indels). f; number of substitutions, length distribution of indels and rearrangements.

g; number of substitution, length distribution of indels and the ratio of number of nonsynonymous to that of coding mutations.

h; number of substitution, indel, rearrangement and the ratio of number of nonsynonymous to that of coding mutations.

i; number of substitutions with CpG and total number of point mutations. j; Distance between MC tumors are tested by permutation test.

k; Edmondson tumor grading, l; Samples were classfied into two groups with AFP level (>10ng/mL or not).