Supplementary Information

Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma

Jian Huang[1,2,3,4,7], Qing Deng[1,2,7], Qun Wang[1,2,7], Kun-Yu Li[2], Ji-Hong Dai[2], Niu Li[1,2], Zhi-Dong Zhu[3], Bo Zhou[1,2], Xiao-Yan Liu[1,2], Rui-Fang Liu[1,2], Qian-Lan Fei[2], Hui Chen[2], Bing Cai[5], Boping Zhou[4], Hua-Sheng Xiao[3], Lun-Xiu Qin[6], Ze-Guang Han[1,2,*]

[1]Human Genome Center of Rui-Jin Hospital, Shanghai Jiaotong University School of Medicine, 197 Rui-Jin II Road, Shanghai 200025, China;

[2]Shanghai-MOST Key Laboratory for Disease and Health Genomics, Chinese National Human Genome Center at Shanghai, 351 Guo Shou-Jing Road, Shanghai 201203, China;

[3]National Engineering Center for Biochip at Shanghai, Shanghai 201203, China;

[4]The Third People's Hospital of Shenzhen, Shenzhen 518020, China.

[5]Department of Hepatobiliary Surgery, Wuxi People's Hospital of Nanjing Medical University, Wuxi 214023, China;

[6]Liver Cancer Institute & Zhongshan Hospital, Fudan University, 180 Feng Lin Road, Shanghai 200032, China;

[7]These authors contributed equally to this work.

[*]**Correspondence should be addressed to** Ze-Guang Han. Chinese National Human Genome Center at Shanghai, 351 Guo Shou-Jing Road, Shanghai 201203, China; Tel: +86-21-50801325; Fax: +86-21-50800402; Email: hanzg@chgc.sh.cn.

Nature Genetics doi:10/1038/ng.2391

Supplementary Note

Illumina/Solexa sequencing. The extracted genomic DNA was randomly sonicated to produce fragment of 300-800 bp, which were then polished with T4 DNA polymerase and T4 polynucleotide kinase. NimbleGen linkers were added to the polished DNA fragments using T4 DNA ligase. The ligated products were then hybridized to the capture array according to the manufacturer's protocol, and the enriched DNA fragments were eluted and amplified with ligation-mediated PCR through the linkers added to the exonic DNA fragments. Before the second run of library construction, qPCR was performed to estimate the degree of enrichment of the exonic sequences. A minimum requirement of an 80-fold enrichment was achieved for all of the libraries prepared for the next procedure. The enriched exonic DNA was randomly blunt-end ligated with DNA ligase to fragments ranging in size from 2-5 kb. The resultant DNA products were sheared to 200 bp on average and were subjected to standard Illumina Genome Analyzer (GA) library preparation according to the Illumina protocol. The exome-enriched shotgun libraries were sequenced on the Illumina GA II platform, and single-end reads with an average size of 75 bp were generated. Image analysis and base calling were performed with the Genome Analyzer Pipeline version 1.3, using the default parameters. After removing reads containing sequencing adaptors and low-quality reads with more than five unknown bases, high

Nature Genetics doi:10/1038/ng.2391

quality reads were aligned to the NCBI human reference genome (hg19) using MAQ[1] with the default parameters. To identify potential mutations, we performed local realignments of the BWA-aligned reads using the Genome Analysis Tool Kit (GATK) [2]. The raw lists of potential somatic substitutions were called by using VarScan (v2.2) based on the MAQ alignments[3]. After data analysis, SNVs were considered to be of interest if they met the following criteria: (1) variants should have Phred-like scaled consensus scores or SNP quality scores > 20; (2) variants should have mapping qualities > 30; (3) both the tumors and the matched normal samples should be covered more than 10-fold at each genomic position being compared; (4) the average base quality for a given genomic position should be at least 15 in both the tumors and the normal samples; (5) variants should be supported by at least 10% of the total reads in the tumors if no high-quality, variant-supporting reads are allowed in the normal controls; (6) variants should be supported by at least five reads in the tumors; (7) after cross-referencing of the potential somatic mutations against the dbSNP and SNP datasets of the 1000 Genomes Project (http://www.1000genomes.org), any mutations present in these datasets must be filtered out; and (8) to manage the false positives associated with pseudo-genes and repeat sequences, simulated reads (80 bp in length) containing the potential mutations must be generated and aligned to the reference genome. If more than 10% of the simulated variant-containing

Nature Genetics doi:10/1038/ng.2391

reads of a given variant cannot be uniquely mapped to the reference genome, this variant must be discarded.

ABI SOLiD sequencing. DNA was randomly fragmented by sonication into fragments in the target range of 150-180 bp. Then, the DNA fragments were end-repaired to produce blunt-ended fragments, and subsequently ligated to SOLiD-specific universal sequencing adaptors and size-selected. The correctly sized ligation products were nick-translated and PCR-amplified to generate the Prepped Library. The library hybridization was performed at 65℃ for 24 hours, and the captured hybrids were selected using magnetic beads. An emulsion PCR reaction was then performed, and the templates were enriched and modified for deposition onto a glass slide. The slides were the site of massively parallel sequencing performed by ligation to generate a monoclonal 50-bp sequence read from each template bead. On the basis of the size of the target region, the sequence coverage, and the number of libraries, an eight-partition deposition chamber for each single-flow cell, or octet, was used for each sample. SOLiD™ BioScope™ software was used for multiple rounds of data analysis, including mapping or aligning the reads to a reference genome and generating a *.bam file after the completion of the mapping and pairing. BioScope™ software uses the diBayes tool to identify single nucleotide polymorphisms (SNPs). This

Nature Genetics doi:10/1038/ng.2391

tool utilizes the color-space reads, the quality values, the reference sequence, and the error information from each SOLiD™ slide as the input to call the SNPs. The diBayes package performs an independent SNP analysis at each position in the reference, using either a Bayesian or frequentist algorithm. The frequentist algorithm is employed when the coverage is for a given position is high (for example, 60×). Given the assumption that the errors will follow a Poisson distribution, the frequentist algorithm calculates the probability of the null hypothesis that the observed valid dicolor mismatches are errors. If the probability of the null hypothesis is too low, the hypothesis is rejected and the position is deemed a SNP. The Bayesian algorithm is used when the coverage of the position is not particularly high. The Bayesian algorithm evaluates the posterior probability of the existence of a heterozygote or a non-reference homozygote at the position in question. On the basis of the probability of the SNP evidence being a mis-color call, a position error, or a probe error, the Bayesian algorithm uses the prior probability of the position being a heterozygote and the probability of the observation being correct. The probabilities are calculated from the quality values of the color calls, from the frequencies of dicolor read mismatches as a function of the positions in a read, or from the frequencies of mismatches as the function of the 6-mer probe prefix, respectively. Th detection of indel variants using a split-read technique is achieved using BioScope™ software's small indel

Nature Genetics doi:10/1038/ng.2391

(insertion and deletion) caller with the BAM files produced from the fragment library types. The small indel pipeline determines the high-quality calls for insertions and deletions in two stages. In the first stage, the pipeline evaluates the gapped alignments on a bead-by-bead basis. In the second stage, the indel caller takes these gap alignments, forms pileups, filters the pileups based on specific heuristics, determines the zygosity, and annotates the indel sequences. This procedure results in concisely annotated and highly accurate indel calls. For fragment libraries, indel sizes up to 11 for deletions and 3 for insertions are considered valid.

Identification of candidate somatic mutations. To identify candidate somatic mutations using two massively parallel sequencing platforms, as described in **Supplementary Figure 2**, we first found potential SNP loci in the exomes of sequenced tumors (primary tumors and/or PVTTs) using Illumina/Solexa and ABI SOLiD™ sequencing data and the SOAPsnp and BioScope™ software tools respectively, with default parameters. The raw data produced by the two massively parallel sequencing platforms were first mapped to the reference genome (hg19); 122,574 and 124,208 potential base substitutions were identified in the ten tumor genomes with the SOAPsnp and BioScope programs respectively, based on Illumina/Solexa and ABI SOLiD sequencing platforms. Next, candidate variation sites were identified that met the following prerequisites: 1)

Nature Genetics doi:10/1038/ng.2391

homozygous genotypes in the matched, non-tumorous liver samples; 2) potential base substitution mutations in the primary tumor and PVTT samples compared with the genotypes of the matched, non-tumorous liver samples; 3) loci containing base substitutions that were not deposited in the dbSNP or 1000Genomes databases; 4) loci located in exons and splice sites; After being subjected to these four filters, 550 and 579 candidate somatic mutations that had been identified with the Illumina/Solexa and ABI SOLiD sequencing platforms, respectively, were combined with additional sequencing data for further analyses and comparisons. The following filters were applied to classify the candidate somatic mutations: 1) the number of sequence reads with base substitutions was ≤ 2 in the non-tumorous liver samples and 2) the number of sequence reads with base substitutions was ≥ 2 in the primary tumor and PVTT samples. On the basis of these criteria, the somatic mutations were classified into three groups: 475 somatic mutations that were simultaneously identified with Solexa and SOLiD sequencing, 256 with Solexa sequencing alone and 301 with SOLiD sequencing alone. These somatic mutations were annotated to determine their mutation types.

Bioinformatic tools. The SeattleSeq Annotation server provides annotations for both known and novel SNPs. These annotations include dbSNP rs IDs, gene names and accession numbers, SNP functions (*e.g.*,

Nature Genetics doi:10/1038/ng.2391

missense), protein positions and amino acid changes, conservation scores, HapMap frequencies, PolyPhen predictions, and clinical associations. (http://gvs.gs.washington.edu/SeattleSeqAnnotation/). The somatic mutations identified in the HCC samples were further filtered using the annotation information provided by SeattleSeq Annotation. The functional effects of the mutations were predicted by PolyPhen[4-6] (http://genetics.bwh.harvard.edu/pph2/).

Loss-of-function by RNA interference screen for cell viability. Each sequence was filtered using NCBI BLAST to validate its specificity and to reduce off-target effects. The siRNA target sequences are listed in **Supplementary Table 12**. The double-stranded siRNA oligonucletotides were chemically synthesized by Shanghai GenePharma Co. Ltd. and were dissolved overnight in a 20-μM solution of double-distilled H_2O treated with diethylpyrocarbonate (DEPC, Sigma). The cell-based siRNA screen was performed as follows: 3×10^3 – 5×10^3 cells per well were seeded in 96-well culture plates. Likewise, the siRNA duplexes were arranged in 96-well culture plates and delivered into the HCC cells with Lipofectamine™2000 (Invitrogen), according to the manufacturer's instructions. The siRNA concentration was optimized to minimize off-target effects. The cells were incubated at 37℃ for 72 hours prior to the cell viability assay. The cell viability was measured using the Cell

Nature Genetics doi:10/1038/ng.2391

Counting Kit-8 (Dojindo Laboratories), a sensitive colorimetric assay for the determination of the number of viable cells in cell proliferation and cytotoxicity studies, according to the manufacturer's instructions. A microplate spectrophotometer (Bio-Tek, MQX200) was used to collect the raw data from 96-well plates twice with a 450-nm filter. A FAM-labeled siRNA was used as a transfection control, two irrelevant siRNAs served as negative controls, and two siRNAs against AURKA were positive controls in each plate. The functional assays of these mutated genes employed the following 8 cell lines: Huh-7, PLC/PRF/5, Focus, WRL68, MHCC-97L, MHCC-97H, HCC-LM3 and HCC-LM6.

The raw absorbance value from each plate were normalized and analyzed according to a previously described method[7]. First, the output profile of cell viability of each cell line was evaluated and illustrated in **Supplementary Figure 13b**. The one-sample t-test was used to examine the mean difference between each sample and the known population mean. The results indicate that the knockdown of the majority of the genes had no significant effect on cell viability in 4 of the 8 HCC cell lines. (one sample t-test, WRL68, P = 0.0801; PLC/PRF/5, P = 0.3028; HCC-LM3, P = 0.1617; HCC-LM6, P = 0.5181). However, the loss of function of these genes may contribute to the cell viability of the other 4 HCC lines (one-sample t-test, Focus, P = 0.0111; Huh-7, P = 0.0001; MHCC-97L, P = 0.0013; MHCC-97H, P = 0.0436). Second, to identify

Nature Genetics doi:10/1038/ng.2391

hits from the screen data, phenotypes were scored for their statistical significance. As a measure of the evidence for loss-of-function phenotype, a score was calculated according to formula $\mathbf{Z} = \frac{y-M}{s}$ **,** where y and s are the mean and standard deviation, respectively, of the triplicate normalized value of each siRNA, and M is the median of the distribution of the y values for all of the siRNAs. Meanwhile, the *P* value was calculated using one-tailed t-test to assess the statistical significance of the difference. As shown in **Supplementary Fig. 13c**, positive results in the loss-of-function assay in each cell line were determined to promote viability when $\mathbf{Z} < -5$ and $\boldsymbol{P} < 0.01$ and to inhibit cell viability when $\mathbf{Z} > 5$ and $\boldsymbol{P} < 0.01$. Third, we selected the genes that produced at least one siRNA hit in two or more cell lines as candidates for further research. The total Z score is the sum of the z scores of the siRNAs hits against an effector gene, and the average *P* value is the average of the *P* values of the siRNA hits against the effector gene. To validate the effect on cell viability of RNA interference against the mutated genes, the gene expression levels and siRNA-mediated knockdown were evaluated in the corresponding cell lines using quantitative PCR.

Nature Genetics doi:10/1038/ng.2391

Supplementary discussion

Mutational signatures are known to carry the specific imprints of previous mutagenic exposures or DNA repair defects; hence, these signatures provide insights into cancer etiology. Chronic HBV infection is the most common etiology of HCC in Asian countries. The integration of the HBV viral DNA into the human genome can induce chromosomal instability, resulting in rearrangements and deletions. However, viral DNA integration could not be directly responsible for the elevated number of transversions noted above. As previously mentioned, dietary exposure to AFB1 increases the risk of HCC commonly found in Asia. AFB1-exposed individuals frequently have distinct TP53 mutations, such as G>T transversions in codon 249 (AGG>AGT) and a characteristic mutational spectrum dominated by C:G>A:T mutations[8-11]. Importantly, the C:G>A:T transversion was significantly enriched in half (5/10) of all of the HCC samples in this study, implying that AFB1 exposure may be a critical contributor in these HBV-associated HCCs, although the AGG>AGT mutation at codon 249 of *TP53* was not found in the ten HCC patients. However, whether AFB1 leads to the T:A>A:T transversion is uncertain. We suspect that other carcinogens are involved in the oncogenesis of HBV-associated HCC. Aristolochic acid, a known carcinogen that has been associated with the development of progressive renal fibrosis, human urothelial cancer, and rodent tumors, is a

Nature Genetics doi:10/1038/ng.2391

component of plant extracts, including herbal remedies[12]. Aristolochic acid has been implicated in the T:A>A:T transversion in some cancer genes, including *TP53* and *Ras*[13,14]. Vinyl chloride, a gas used in the plastics industry, is known as a carcinogen that causes liver cancers, including HCC[15]. This carcinogen may cause the T:A>A:T transversion.

Among the genes carrying two confirmed non-silent mutations, *ELMO1* encodes a Rac1 regulator that controls vascular morphogenesis[16]; DSE is considered a tumor rejection antigen[17]; and germline nonsynonymous variants of *CSMD3* have been associated with familial colorectal cancer[18]. Despite their low mutation frequencies, the prospect that these mutated genes may play significant roles in HCC oncogenesis should be further investigated.

Of 7 potential cancer-associated genes, as indicated by RNAi secreen, in addition to *VCAM1* and *CDK14*, *HOXA1* has been associated with oncogenic transformation and anchorage-independent survival[19,20]. Whether and how these genes might be involved in HCC oncogenesis should be examined in future research.

References

1. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851-1858 (2008).

2. McKenna, A. *et al*. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. **20**, 1297-1303 (2010).
3. Koboldt, D.C. *et al*. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283-2285 (2009).
4. Ramensky, V., Bork, P. & Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. **30**, 3894-3900 (2002).
5. Sunyaev, S., Ramensky, V. & Bork, P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet.* **16**, 198-200 (2000).
6. Sunyaev, S. *et al*. Prediction of deleterious human alleles. *Hum. Mol. Genet.* **10**,591-597 (2001).
7. Boutros, M., Brás, L.P. & Huber, W. Analysis of cell–based RNAi screens. *Genome Biol.* **7**, R66 (2006).
8. Hussain, S.P., Schwank, J., Staib, F., Wang, X.W. & Harris C.C. TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer. *Oncogene* **26**, 2166–2176 (2007).
9. Aguilar, F., Hussain, S.P. & Cerutti, P. Aflatoxin B1 induces the transversion of G—>T in codon 249 of the p53 tumor suppressor gene in human hepatocytes. *Proc. Natl. Acad. Sci. USA* **90**,

Nature Genetics doi:10/1038/ng.2391

8586–8590 (1993).

10. Hsu, I.C. *et al*. Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature* **350**, 427–428 (1991).
11. Besaratinia, A., Kim, S.I., Hainaut, P. & Pfeifer, G.P. In vitro recapitulating of TP53 mutagenesis in hepatocellular carcinoma associated with dietary aflatoxin B1 exposure. *Gastroenterology* **137**, 1127–1137 (2009).
12. Arlt, V.M., Stiborova, M. & Schmeiser, H.H. Aristolochic acid as a probable human cancer hazard in herbal remedies: a review. *Mutagenesis* **17**, 265–277 (2002).
13. Moriya, M. *et al*. TP53 Mutational signature for aristolochic acid: an environmental carcinogen. *Int. J. Cancer* **129**, 1532–1536 (2011).
14. Wang, Y. *et al*. Aristolochic acid–induced carcinogenesis examined by ACB–PCR quantification of H–Ras and K–Ras mutant fraction. *Mutagenesis* **26**, 619–628 (2011).
15. Barbin, A. Etheno–adduct–forming chemicals: from mutagenicity testing to tumor mutation spectra. *Mutat. Res.* **462**, 55–69 (2000).
16. Epting, D. *et al*. The Rac1 regulator ELMO1 controls vascular morphogenesis in zebrafish. *Circ. Res.* **107**, 45–55 (2010).
17. Sasatomi, T. *et al*. Expression of tumor rejection antigens in colorectal carcinomas. *Cancer* **94**, 1636–1641 (2002).
18. Gylfe, A.E. *et al*. Somatic mutations and germline sequence variants

Nature Genetics doi:10/1038/ng.2391

in patients with familial colorectal cancer. *Int. J. Cancer* **127**, 2974–2980 (2010).

19. Zhang, X. *et al*. Human growth hormone–regulated HOXA1 is a human mammary epithelial oncogene. *J. Biol. Chem.* **278**, 7580–7590 (2003).
20. Zhang, X. *et al*. HOXA1 is required for E–cadherin–dependent anchorage–independent survival of human mammary carcinoma cells. *J. Biol. Chem.* **281**, 6471–6481 (2006).

Nature Genetics doi:10/1038/ng.2391

Supplementary Figures

Supplementary Figure 1. The average depth of coverage of the 10 HCC patients whose exomes were sequenced on two types of massively parallel sequencing platforms.

Supplementary Figure 2. A flowchart for identifying somatic mutations based on the Illumina/Solexa and SOLiD sequencing data.

Supplementary Figure 3. The classification and evaluation of 682 potential somatic mutations.

Supplementary Figure 4. The identification of somatic insertion and deletion variants in 10 HCC samples.

Supplementary Figure 5. Twenty-five somatic indel variants were manually confirmed by using Integrative Genomics Viewer (IGV).

Supplementary Figure 6. The diversity of somatic mutations in advanced HCC.

Supplementary Figure 7. Functional categories of the mutated genes.

Nature Genetics doi:10/1038/ng.2391

Supplementary Figure 8. Mutated genes harboring two non-silent mutations confirmed by Sanger sequencing.

Supplementary Figure 9. Genotype comparison and analysis of some of the HCC cell lines with an Affymetrix Genome-Wide Human SNP Array 5.0 dataset.

Supplementary Figure 10. An ARID1A knockdown via RNA interference promoted cell migration and invasion in MHCC-97L and MHCC-97H cells.

Supplementary Figure 11. Functional analysis of ARID1A via RNAi knockdown in various HCC cell lines.

Supplementary Figure 12. The mutated genes were further examined via Sanger sequencing in additional HCC samples.

Supplementary Figure 13. Screen of knockdowns by RNA interference of the mutated genes in eight HCC cell lines.

Nature Genetics doi:10/1038/ng.2391

Supplementary Figure 14. The efficacies of RNA interference against VCAM1, TMEM2, CDK14, HOXA1, TMEM35, ELL and CSNK1G3 in the HCC cells were evaluated with real-time quantitative PCR.

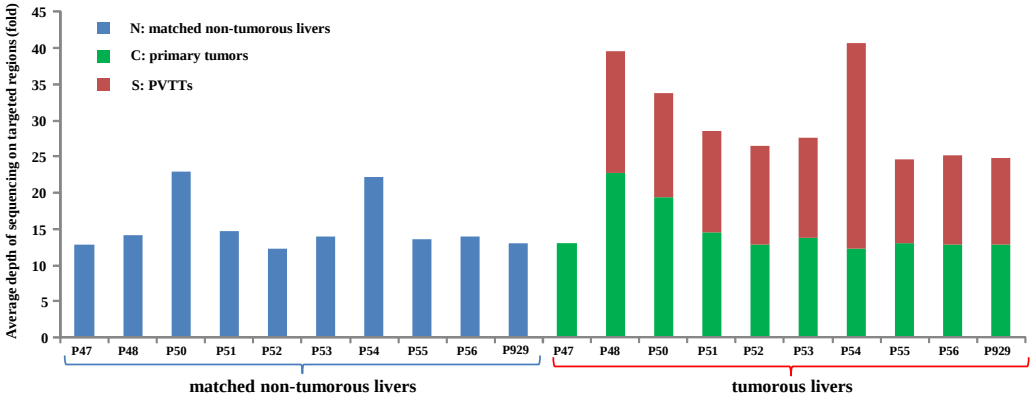
Supplementary Figure 15. The effects of the mutated genes on cell behaviors determined using cell-based RNA interference (RNAi).

Supplementary Figure 16. Functional assays of VCAM1 and CDK14 in various HCC cell lines.

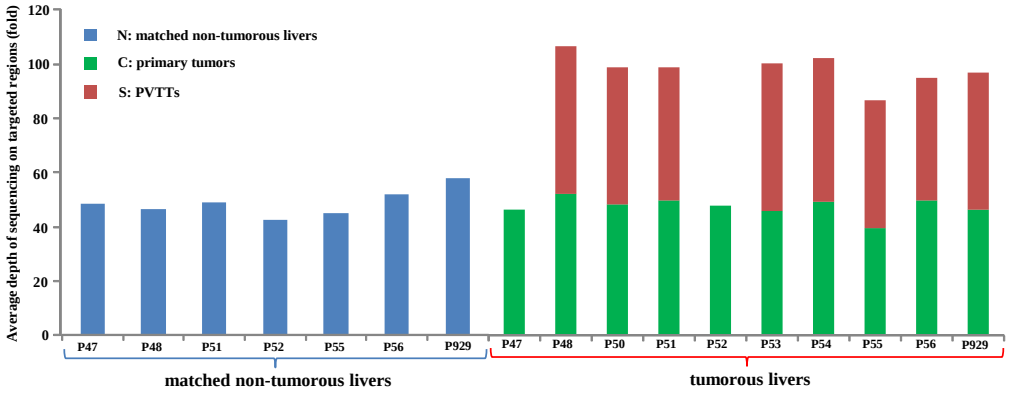
Supplementary Figure 17. The expression patterns in HCC samples of 7 cancer-associated genes from the RNAi screen.

Nature Genetics doi:10/1038/ng.2391

a

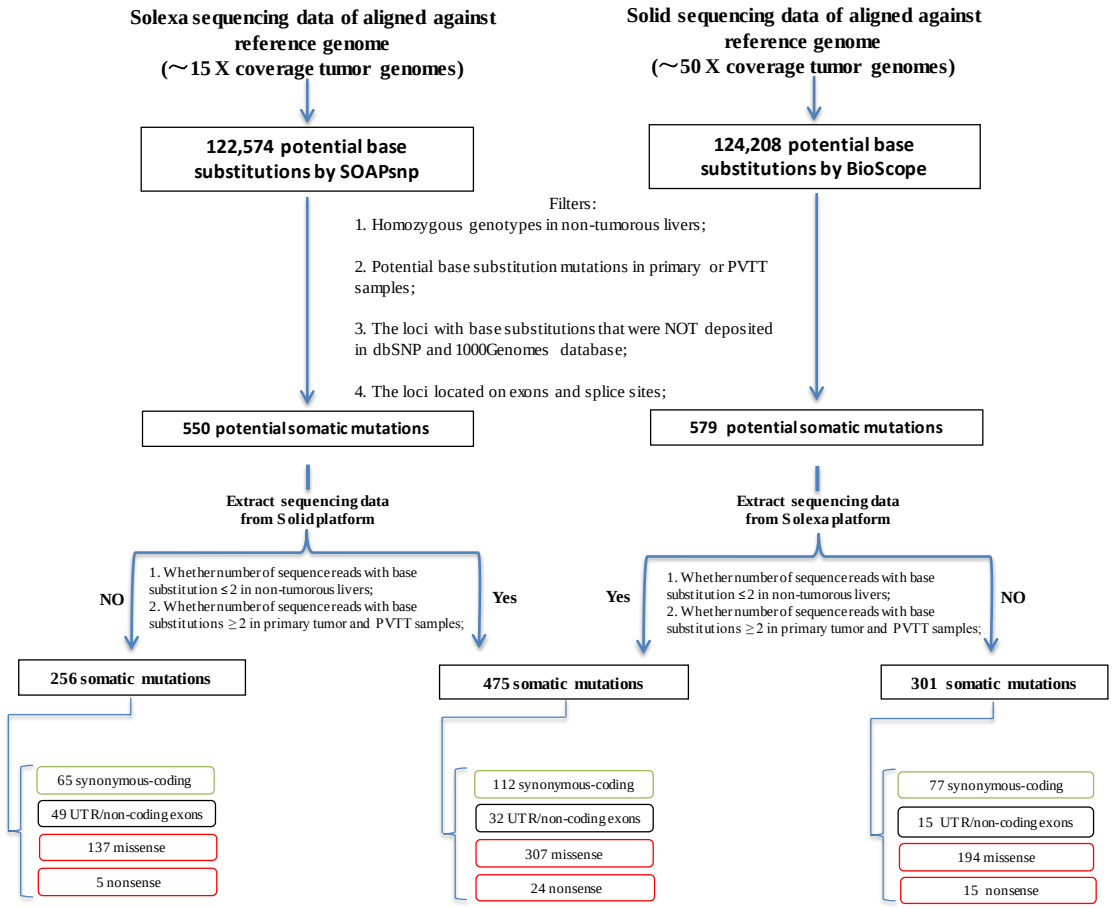


b



Supplementary Figure 1. The average depth of coverage of the 10 HCC patients whose exomes were sequenced on two types of massively parallel sequencing platforms. The average depth of coverage of the targeted bases of each sample is shown. The average depth of coverage of each tumor sample was calculated as the sum of the depths of coverage of the primary tumor and the PVTT samples, as evaluated with the Illumina/Solexa (**a**) and ABI SOLiD sequencing platforms (**b**).

Nature Genetics doi:10/1038/ng.2391



Supplementary Figure 2. A flowchart for identifying somatic mutations based on the Illumina/Solexa and SOLiD sequencing data. The raw data from the two massively parallel sequencing platforms were first mapped onto the reference genome (hg19). A total of 122,574 and 124,208 potential base substitutions were produced by the SOAPsnp and BioScope™ programs, respectively, based on data of the ten tumor exomes generated by the Illumina/Solexa and ABI SOLiD sequencing platforms. Next, the candidate single nucleotide variant mutations had to fulfill the following criteria: 1) homozygous genotypes detected in the matched, non-tumorous liver samples; 2) potential base substitution mutations in the primary tumor and PVTT samples compared with the

Nature Genetics doi:10/1038/ng.2391

genotypes of the non-tumorous liver samples; 3) loci containing base substitutions not deposited in the dbSNP or 1000 Genomes databases and 4) loci located in exons and splice sites. After filtering, 550 and 579 candidate somatic mutations were proposed based on the Illumina/Solexa and ABI SOLiD sequencing data, respectively. We then subjected these sequencing data to further filtering based on the following requirements: 1) the number of sequence reads containing base substitutions was ≤ 2 in the non-tumorous liver samples and 2) the number of sequence reads containing base substitutions was ≥ 2 in both the primary tumor and PVTT samples. Finally, the somatic mutations were classified into three categories: 475 somatic mutations were simultaneously detected by both Solexa and SOLiD sequencing platforms, 256 only by Solexa sequencing and 301 only by SOLiD sequencing. These somatic mutations were annotated to determine the mutation types.

Nature Genetics doi:10/1038/ng.2391

a

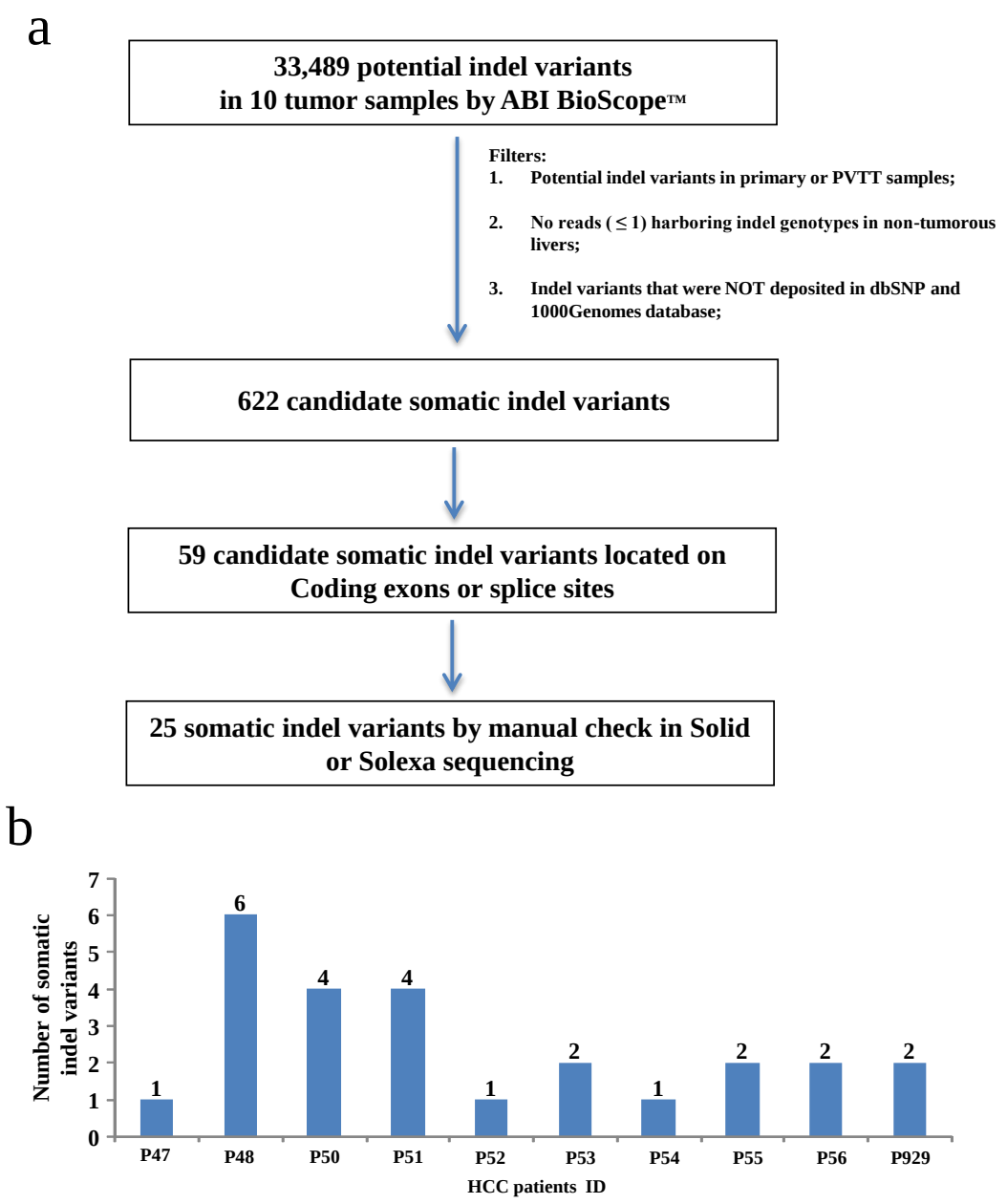
Non-silent mutations

b

Non-silent somatic mutations identified in primary tumors or PVTTs		Sanger sequencing		
		Number of performed PCR-Sanger sequence	Number of confimed mutations	Confirmatory proportion
Solexa & Solid	**331**	**224**	**193**	**86.2%**
Solid	**209**	**41**	**17**	**41.5%**
Solexa	**142**	**131**	**16**	**12.2%**

Supplementary Figure 3. The classification and evaluation of 682 potential somatic mutations. (**a**) The distribution of non-silent somatic mutations based on the Illumina/Solexa and SOLiD sequencing data. Of the total of 682 non-silent somatic mutations, 331 were supported by both the Illumina/Solexa and SOLiD sequencing data, 142 and 209 by only the Illumina/Solexa or SOLiD sequencing data, respectively. (**b**) The evaluation of the above three categories based on the different strategies was performed using conventional Sanger sequencing.

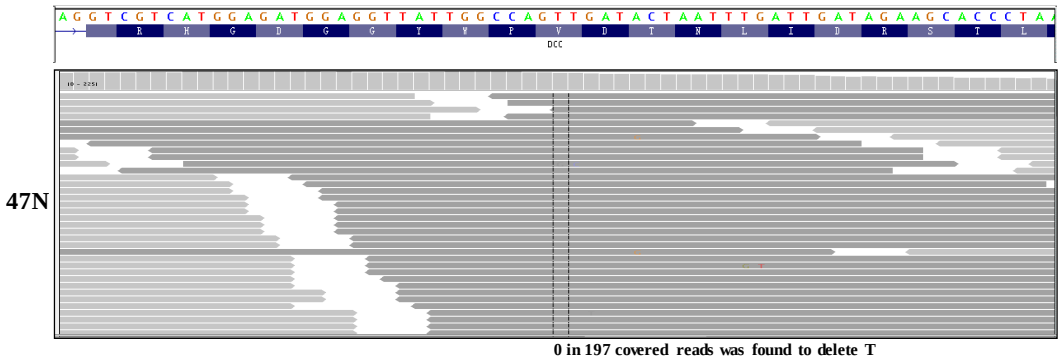
Nature Genetics doi:10/1038/ng.2391



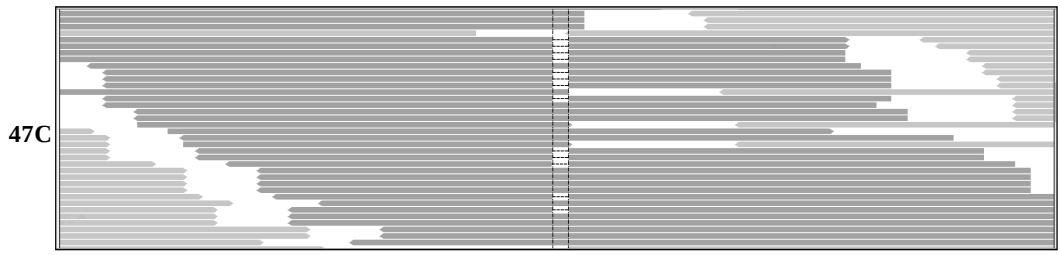
Supplementary Figure 4. The identification of somatic insertion and deletion variants in 10 HCC samples. (**a**) The flowchart used for identifying somatic insertion and deletion variants in the 10 HCC patients. (**b**) The distribution of the identified somatic indel mutations in the 10 HCC patients.

Nature Genetics doi:10/1038/ng.2391

Gene:DCC; Chromosome location: chr18:50961544-50961544; Variant type: DELETION, del T

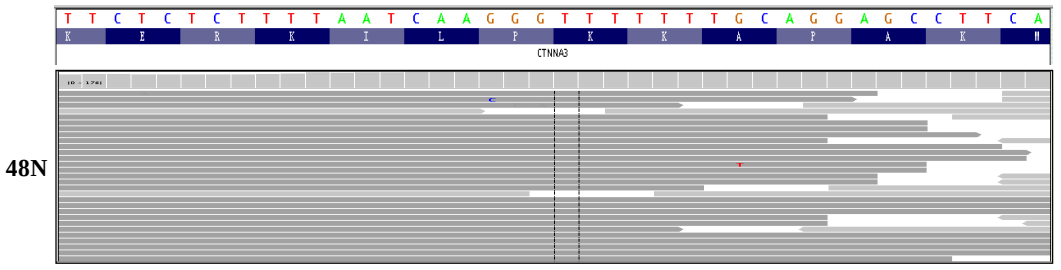


0 in 197 covered reads was found to delete T

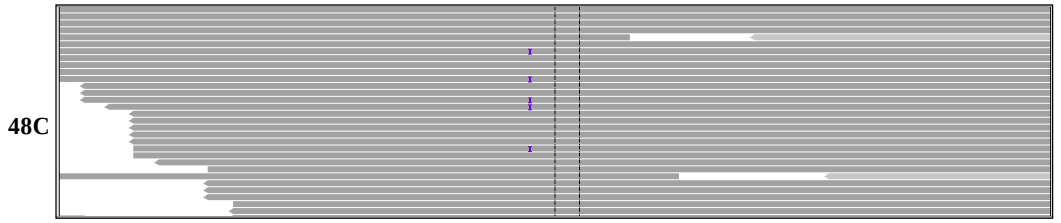


20 in 133 covered reads were found to delete T

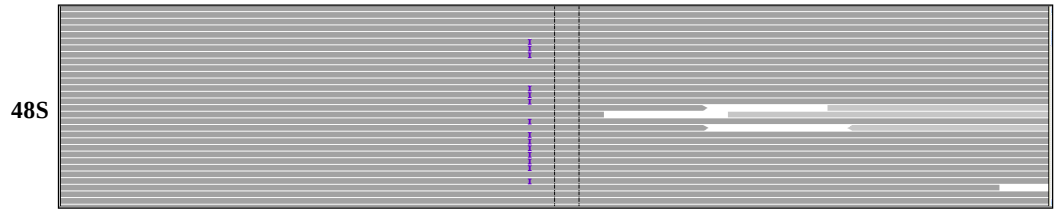
Gene:CTNNA3; Chromosome position: chr10:67680207-67680207; Varinat type: INSERTION, ins T



0 in 173 covered reads was found to insert T



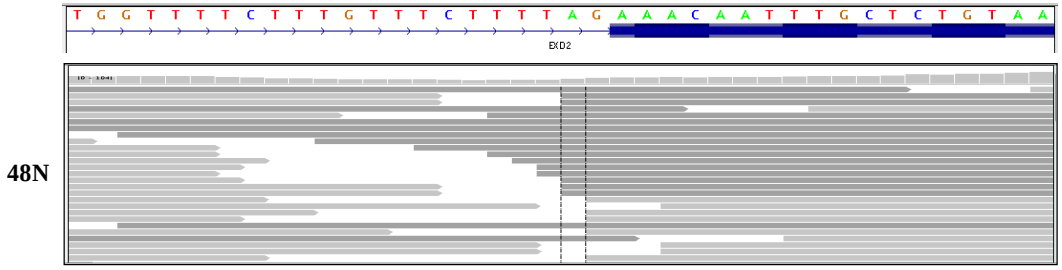
10 in 184 covered reads were found to insert T



52 in 248 covered reads were found to insert T

Nature Genetics doi:10/1038/ng.2391

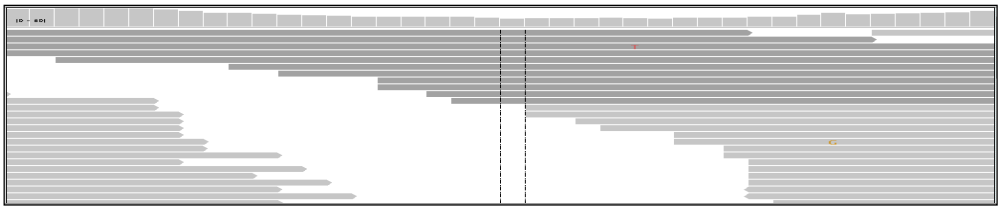
EXD2 chr14:69697187-69697187 DELETION del A



48N

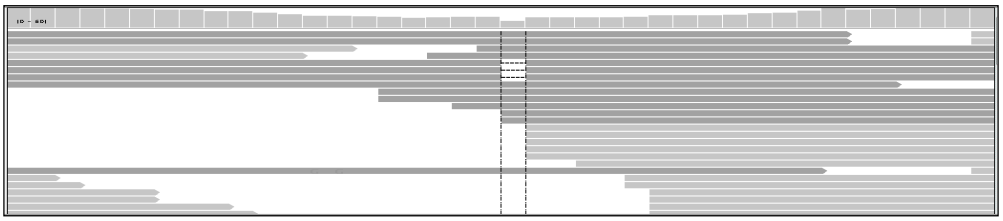
0 in 26 covered reads was found to delete A

48C



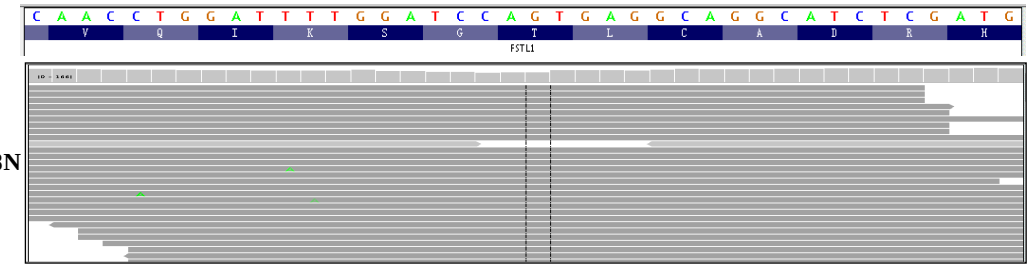
0 in 24 covered reads was found to delete A

48S



5 in 20 covered reads was found to delete A

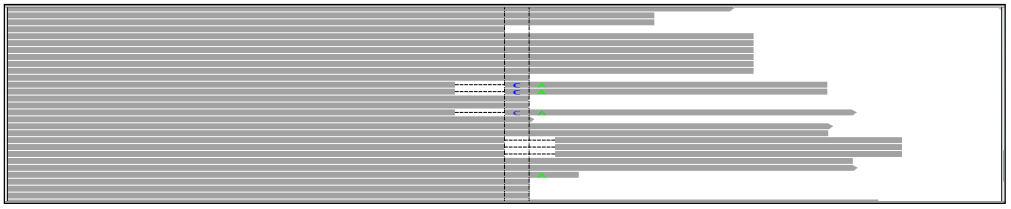
Gene: FSTL1; Chromosome position: chr3:120130742-120130743; Variant type: DELETION, del TG



48N

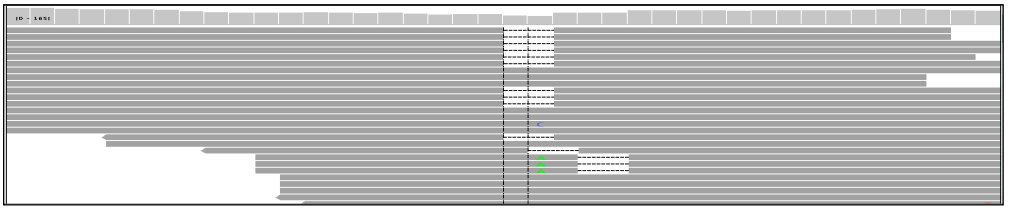
0 in 93～113 coverd reads was found to delete TG

48C



4 in 84～88 covered reads were found to delete TG

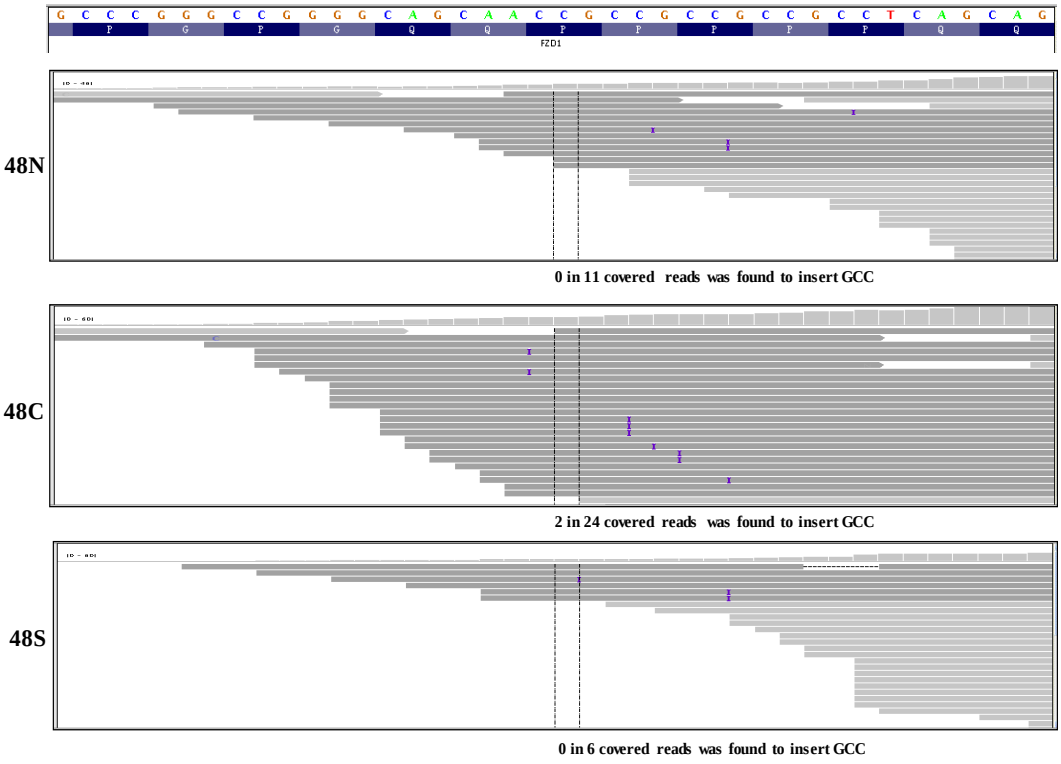
48S



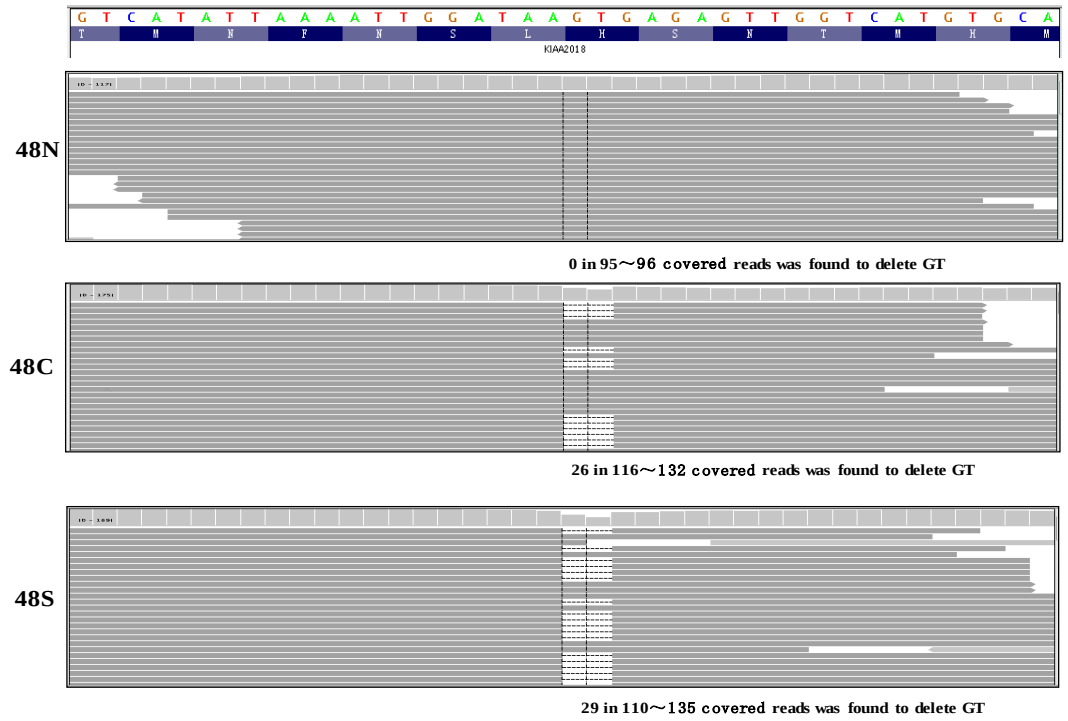
11 in 74～76 covered reads were found to delete TG

Nature Genetics doi:10/1038/ng.2391

Gene: FZD1; Chromosome position: chr7:90894459-90894459; Variant type: INSERTION, insGCC

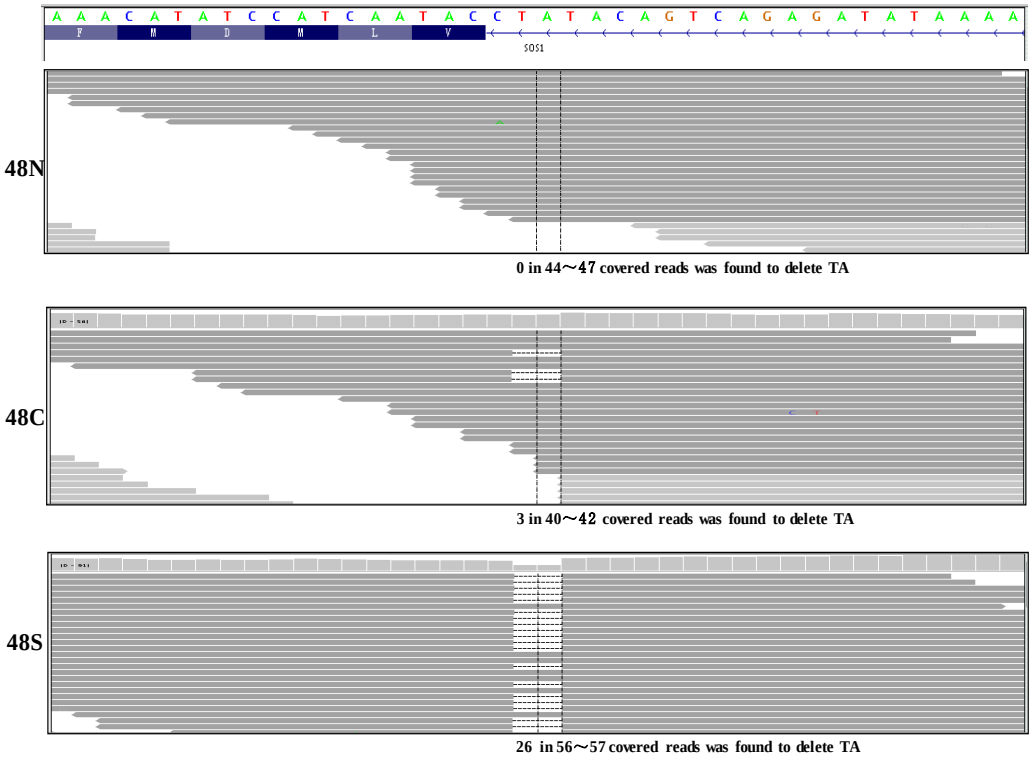


Gene: KIAA2018; Chromosome position: chr3:113373977-113373978; Variant type: DELETION, del GT



Nature Genetics doi:10/1038/ng.2391

Gene:SOS1;Chromosome position: chr2:39281966-39281967;Variant type: DELETION, del TA

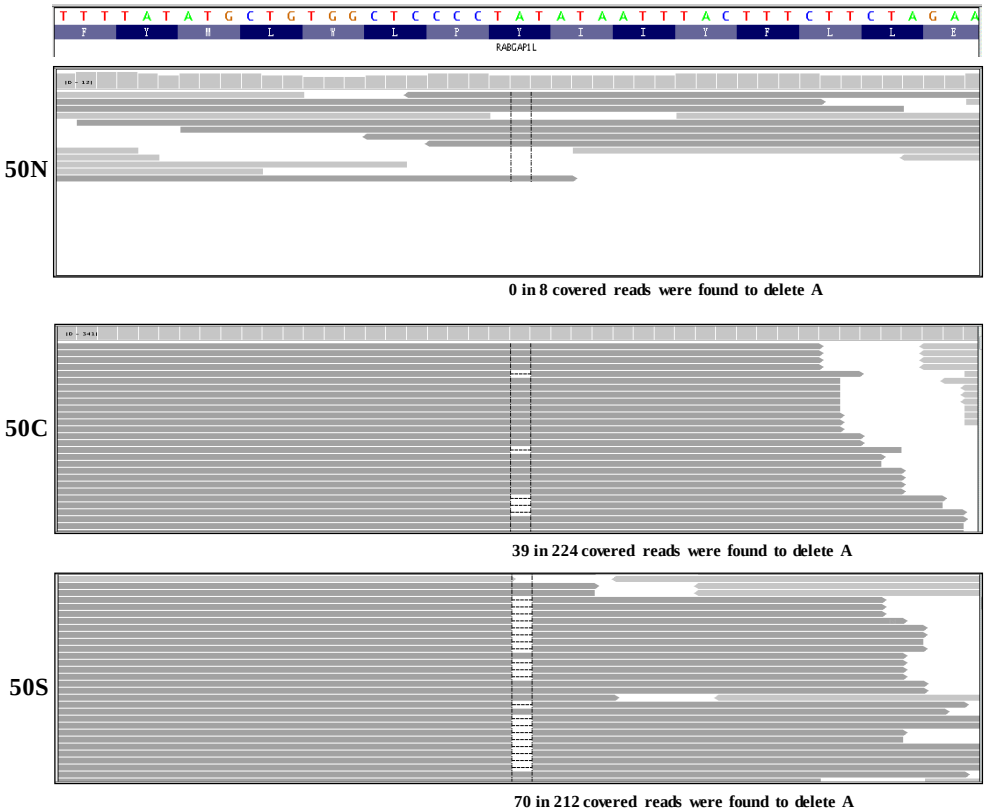


0 in 44～47 covered reads was found to delete TA

3 in 40～42 covered reads was found to delete TA

26 in 56～57 covered reads was found to delete TA

Gene： GPR52; Chromosome position: chr1:174418091-174418091; Variant type: DELETION,del A



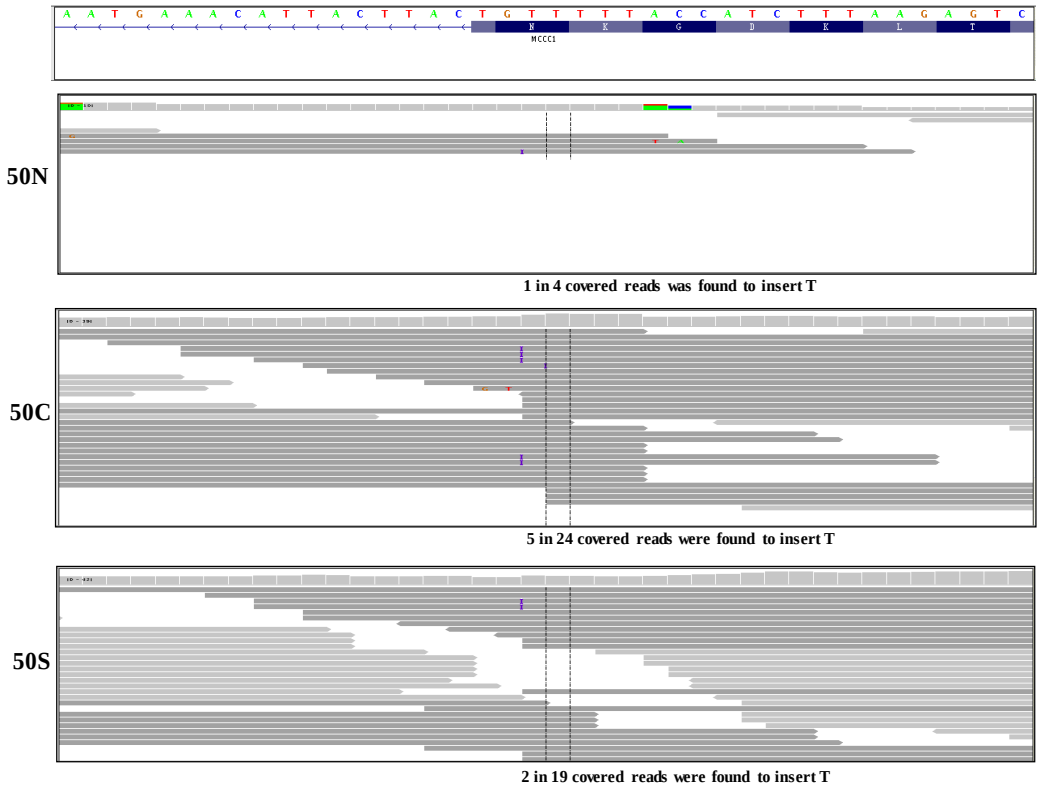
0 in 8 covered reads were found to delete A

39 in 224 covered reads were found to delete A

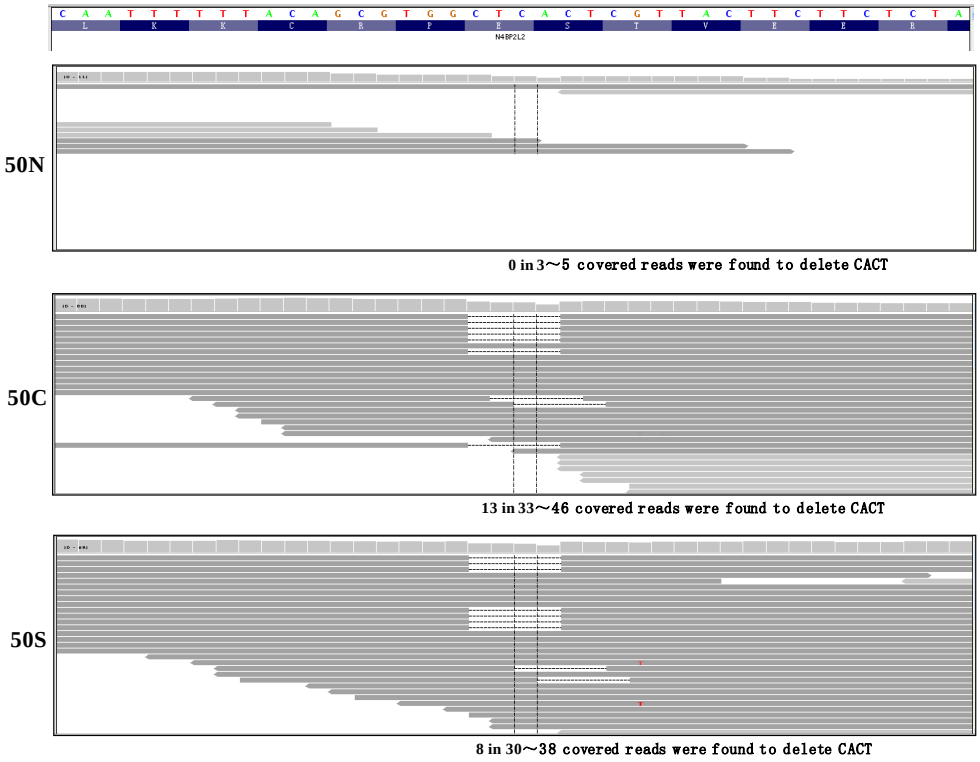
70 in 212 covered reads were found to delete A

Nature Genetics doi:10/1038/ng.2391

Gene: MCCC1; Chromosome position: chr3:182751780-182751780; Variant type: INSERTION, ins T

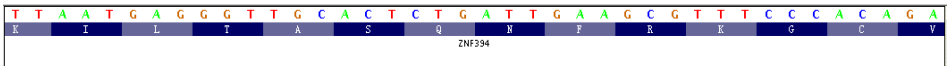


Gene: N4BP2L2; Chromosome position: chr13:33111105-33111108; Variant type: DELETION, del CACT

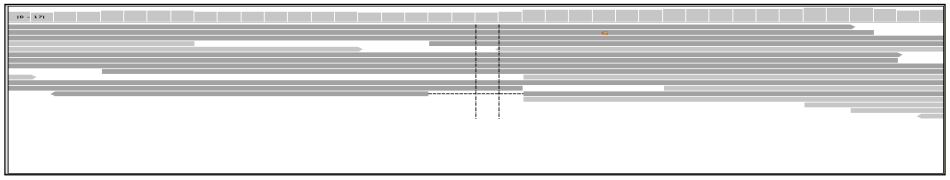


Nature Genetics doi:10/1038/ng.2391

Gene:ZNF394; Chromosome position: chr7:99091315-99091318; Variant type:DELETION, del ATTG

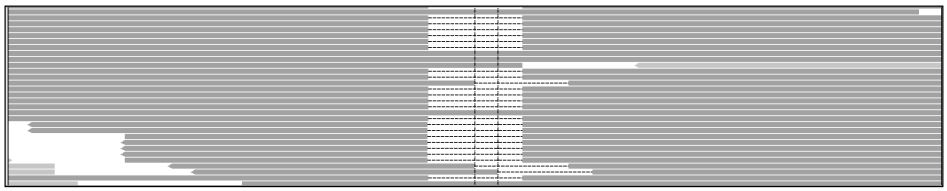


50N



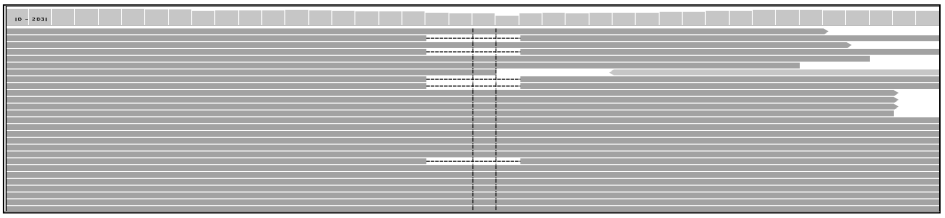
1 in 10～11 covered reads was found to delete ATTG

50C



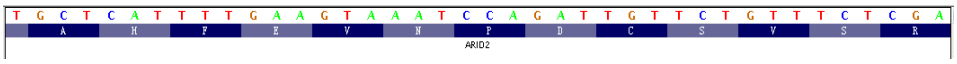
53 in 92～118 covered reads were found to delete ATTG

50S

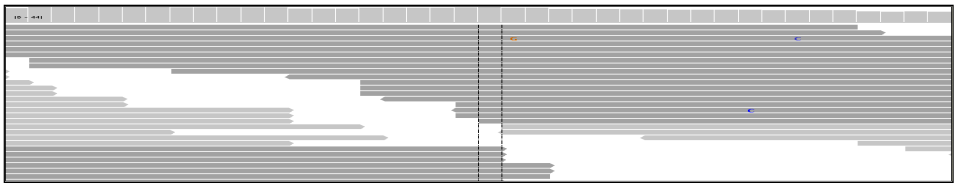


16 in 100～129 covered reads were found to delete ATTG

Gene:ARID2; Chromosome position: chr12:46242645-46242645; Variant type: DELETION, del CA

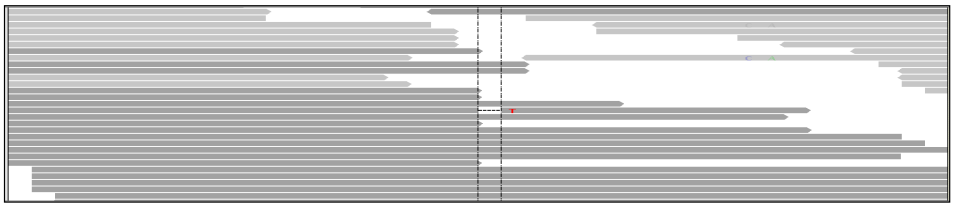


51N



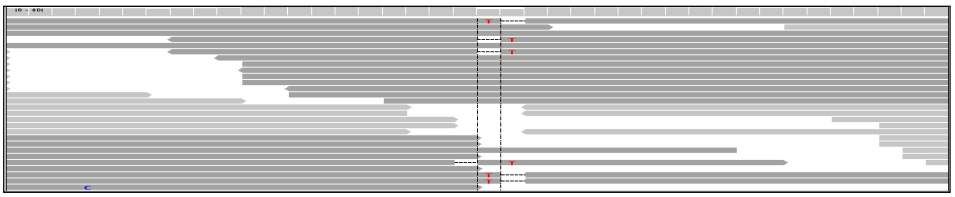
0 in 42 covered reads were found to delete CA

51C



1 in 29～30 covered reads was found to delete CA

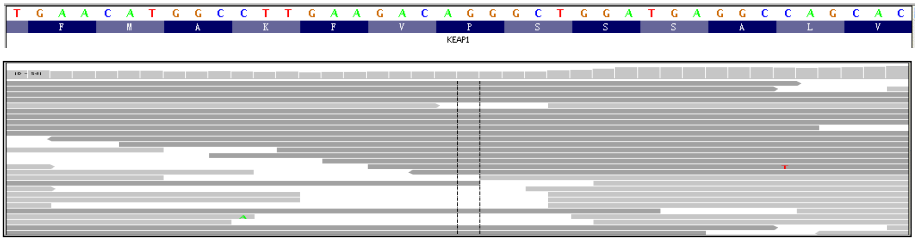
51S



5 in 21～22 covered reads was found to delete CA

Nature Genetics doi:10/1038/ng.2391

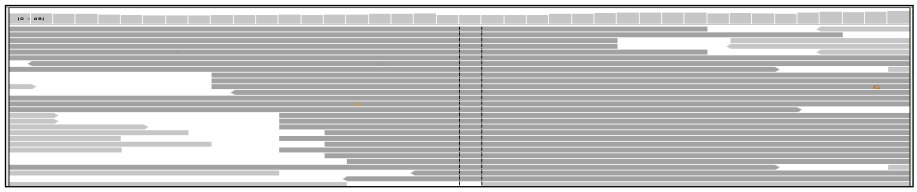
Gene: KEAP1;Chromosome position: chr19:10610396-10610396; Variant type: DELETION, del G



51N

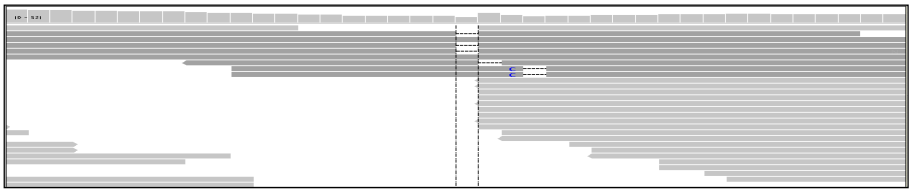
0 in 23 covered reads were found to delete G

51C



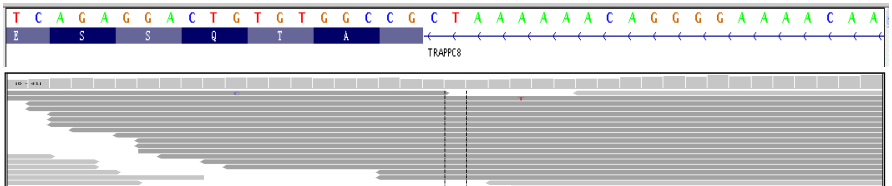
0 in 36 covered reads were found to delete G

51S



4 in 21 covered reads were found to delete G

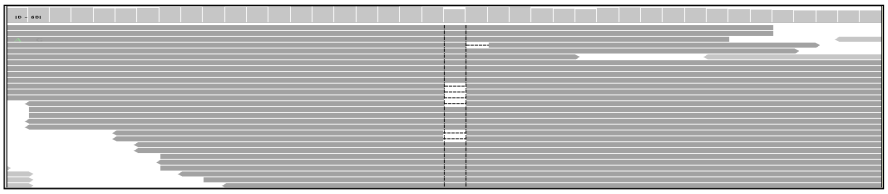
Gene: TRAPPC8; Chromosome position: chr18:29432469-29432469; Variant type: DELETION, del T



51N

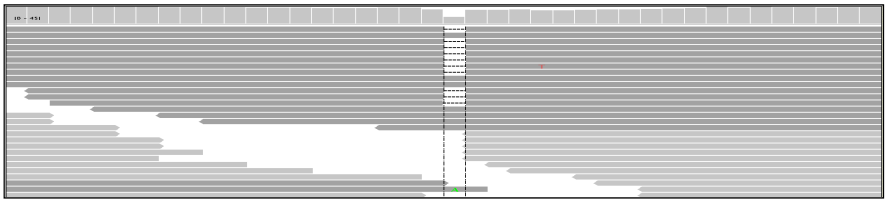
0 in 23 covered reads were found to delete T

51C



6 in 47 covered reads were found to delete T

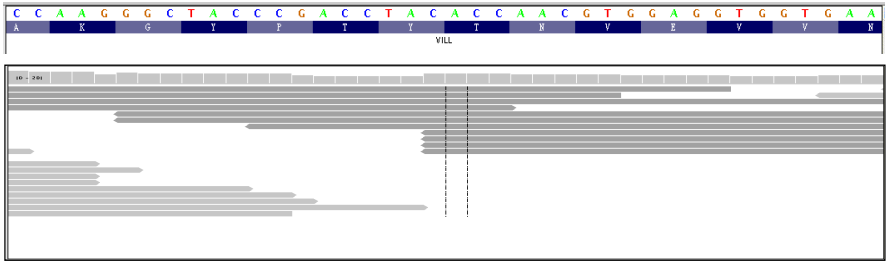
51S



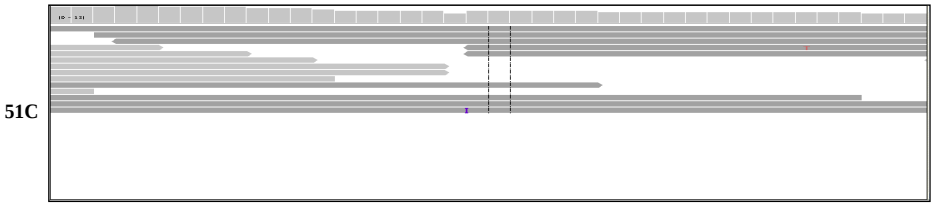
13 in 30 covered reads were found to delete T

Nature Genetics doi:10/1038/ng.2391

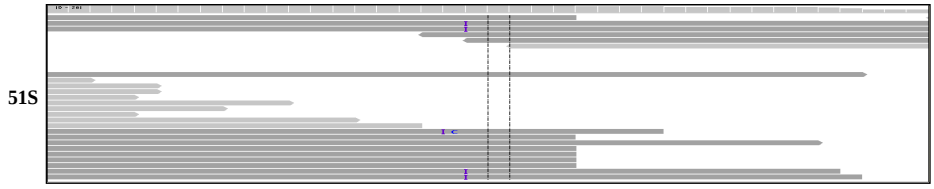
Gene:VILL; Chromosome position: chr3:38040434-38040434;Variant type: INSERTION, ins ACC



0 in 11 covered reads were found to insert ACC

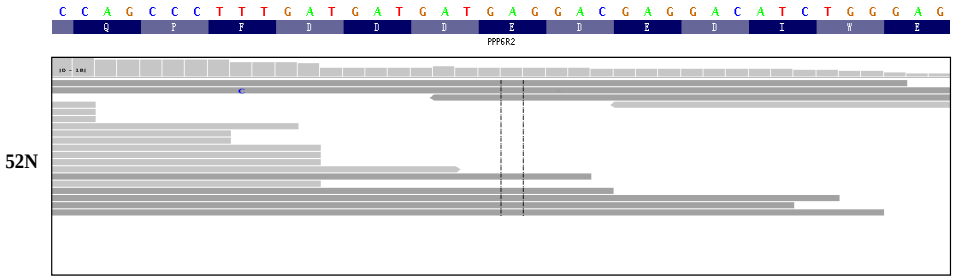


1 in 7 covered reads was found to insert ACC

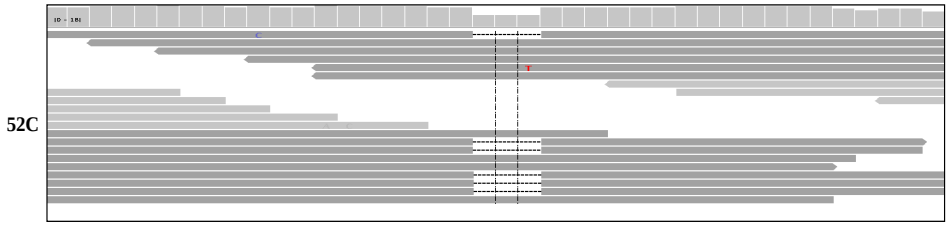


4 in 14 covered reads were found to insert ACC

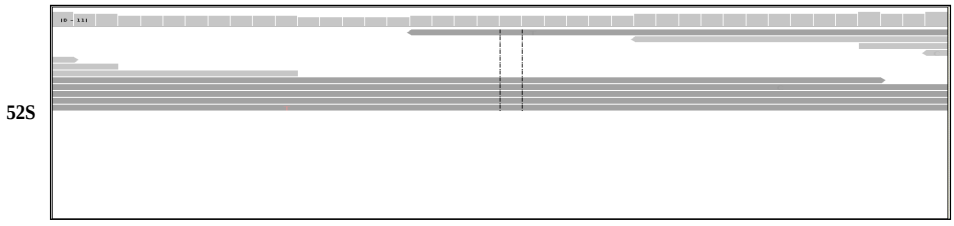
Gene: PPP6R2; Chromosome position: chr22:50876663-50876665; Variants type: DELETION, delAGG



0 in 8 covered reads were found to delete AGG



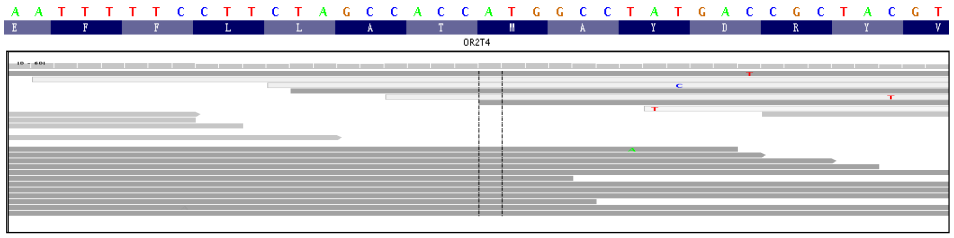
6 in 15 covered reads were found to delete AGG



0 in 6 covered reads were found to delete AGG

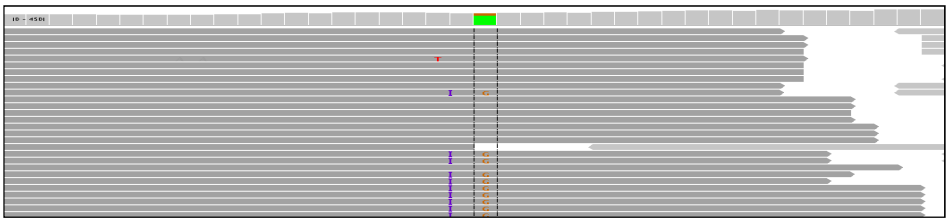
Nature Genetics doi:10/1038/ng.2391

Gene: OR2T4; Chromosome position: chr1:248525328-248525328; Variant type: INSERTION, ins TA



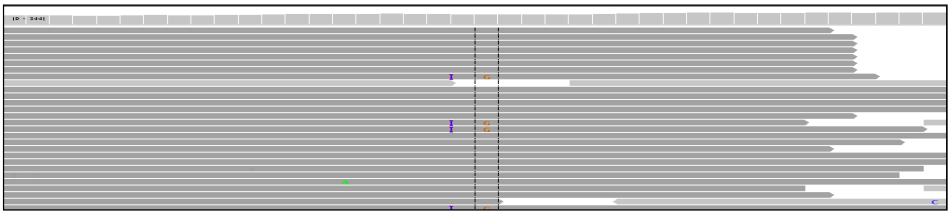
53N

0 in 17 covered reads were found to insert TA



53C

12 in 194 covered reads were found to insert TA



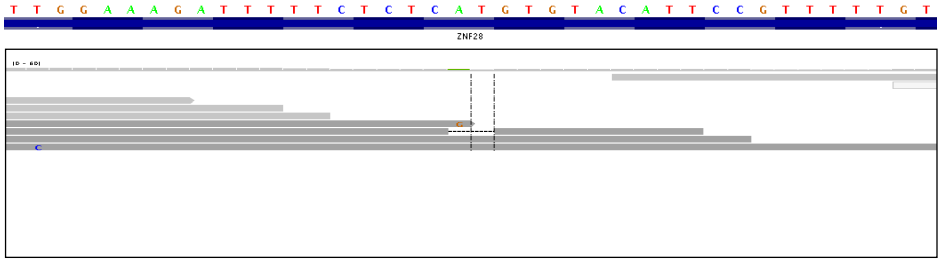
53S

60 in 287 covered reads were found to insert TA

Nature Genetics doi:10/1038/ng.2391

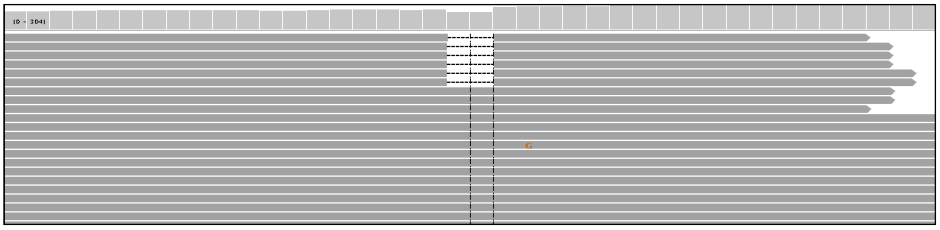
Gene:ZNF28; Chromosome position: chr19:53304469-53304470; Variant type: DELETION, del AT

53N



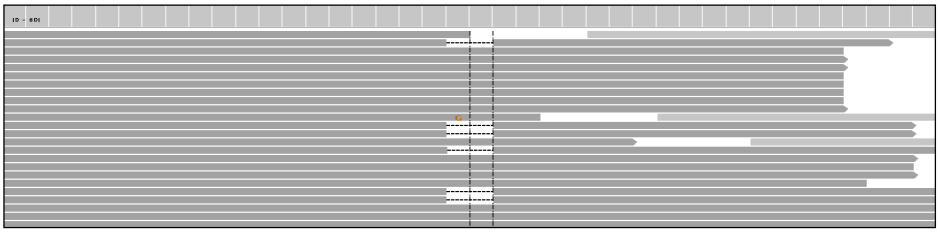
1 in 3～4 covered reads was found to delete AT

53C



24 in 232～233 covered reads was found to delete AT

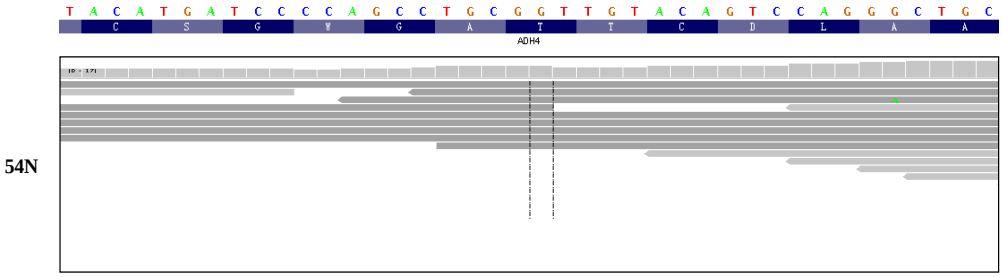
53S



37 in 152～153 covered reads was found to delete AT

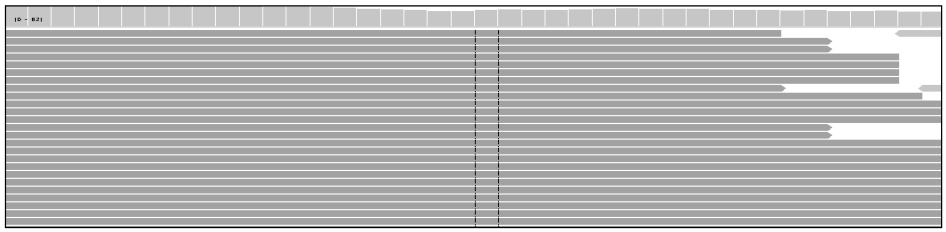
Nature Genetics doi:10/1038/ng.2391

Gene: ADH4; Chromosome position: chr4:100048473-100048473; Variant type: INSERTION, insT



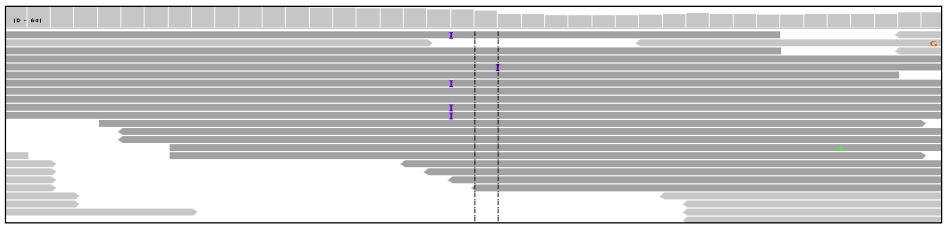
54N

0 in 9 covered reads were found to insert T



54C

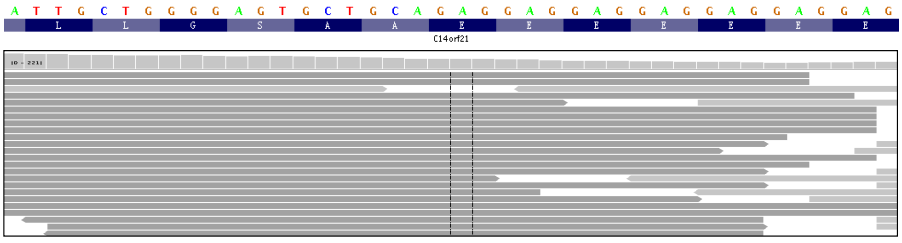
0 in 57 covered reads were found to insert T



54S

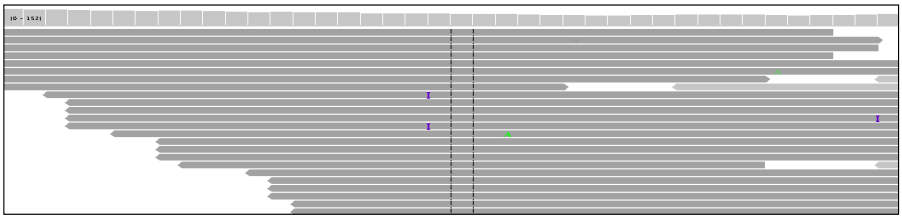
12 in 50 covered reads were found to insert T

Gene: C14orf21; Chromosome position: chr14:24769849-24769849; Variant type: INSERTION, ins GGA



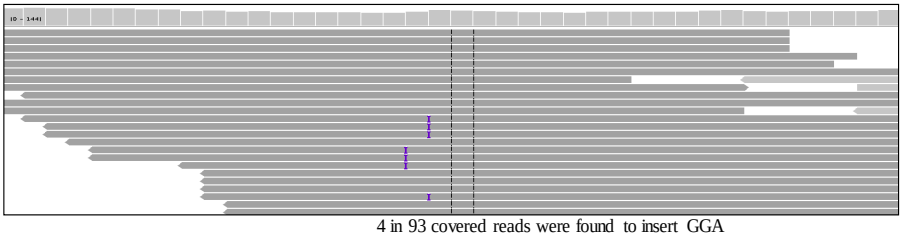
55N

0 in 116 covered reads were found to insert GGA



55C

2 in 81 covered reads were found to insert GGA

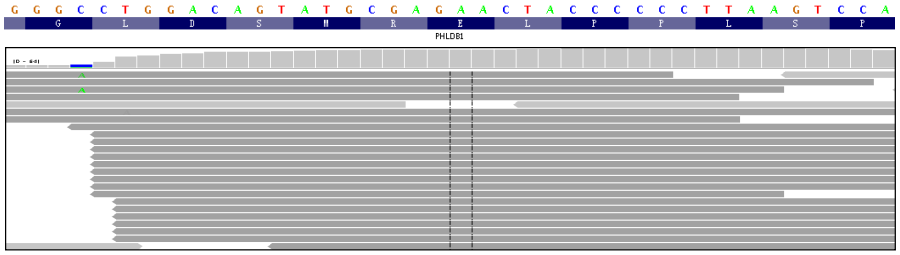


55S

4 in 93 covered reads were found to insert GGA

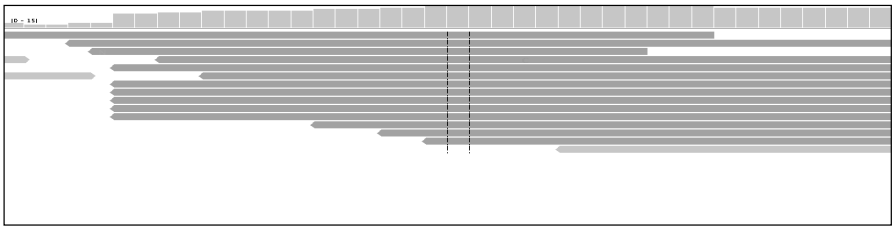
Nature Genetics doi:10/1038/ng.2391

Gene: PHLDB1; Chromosome position: chr11:118498906-118498906; Variant type: DELETION, delA



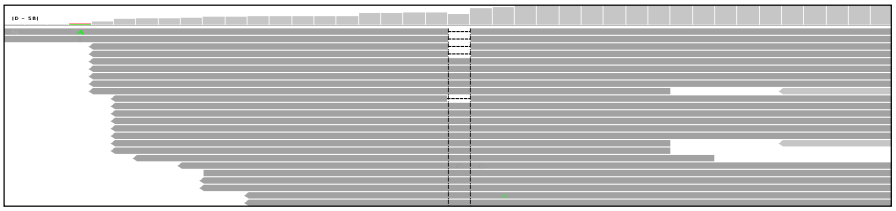
55N

0 in 55 covered reads were found to delete A



55C

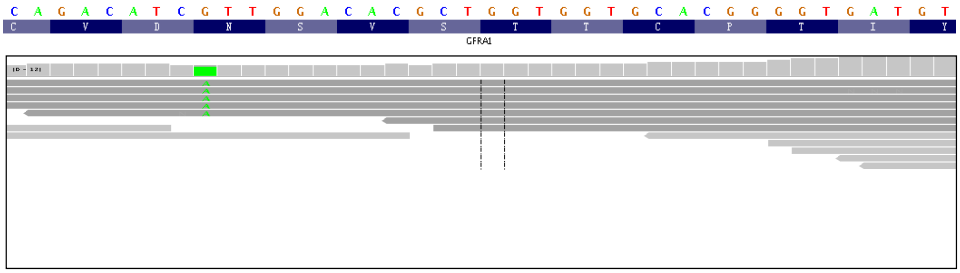
0 in 14 covered reads were found to delete A



55S

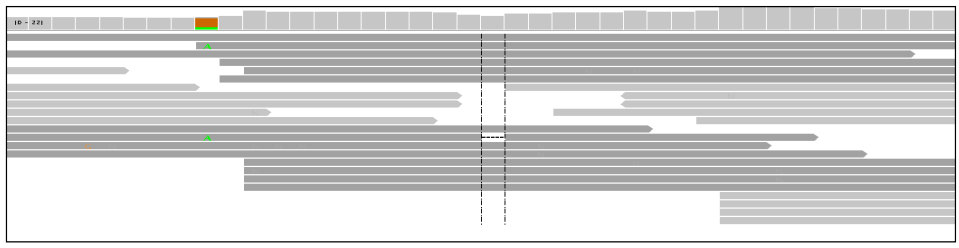
5 in 34 covered reads were found to delete A

Gene: GFRA1; Chromosome position: chr10:117884962-117884962; Variant type: DELETION, delG



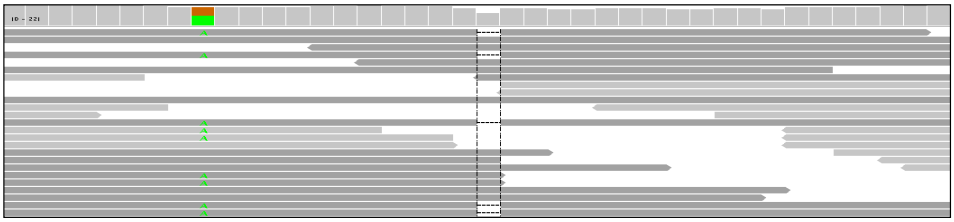
56N

0 in 7 covered reads were found to delete G



56C

1 in 14 covered reads were found to delete G

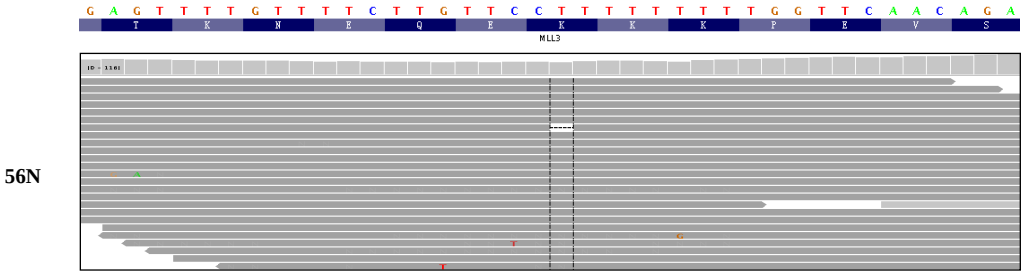


56S

6 in 19 covered reads were found to delete G

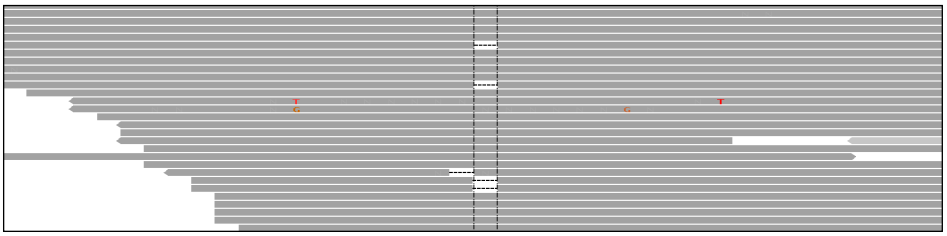
Nature Genetics doi:10/1038/ng.2391

Gene: MLL3; Chromosome position: chr7:151874148-151874148; Variant type: DELETION, delT



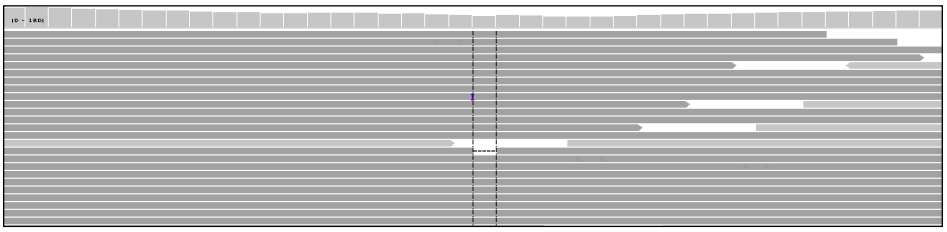
56N

1 in 65 covered reads was found to delete T



56C

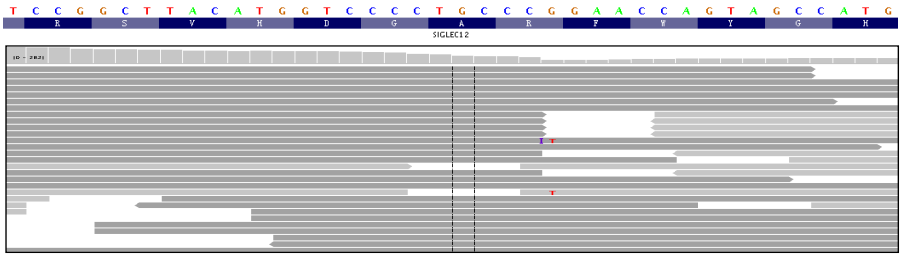
5 in 89 covered reads was found to delete T



56S

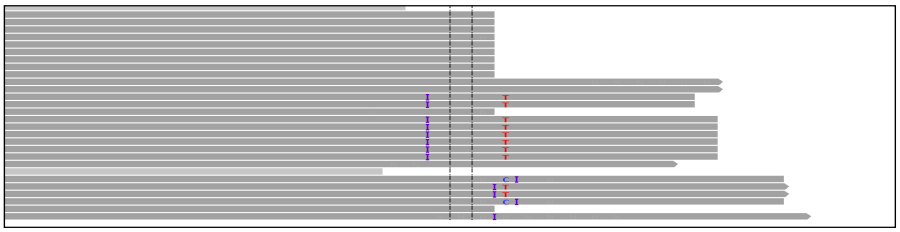
2 in 95 covered reads was found to delete T

Gene: SIGLEC12; Chromosome position: chr19:52004791-52004791; Variant type: INSERTION, ins C



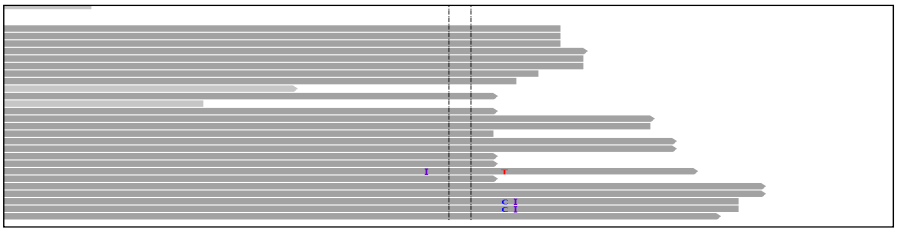
929N

0 in 136 covered reads were found to insert C



929C

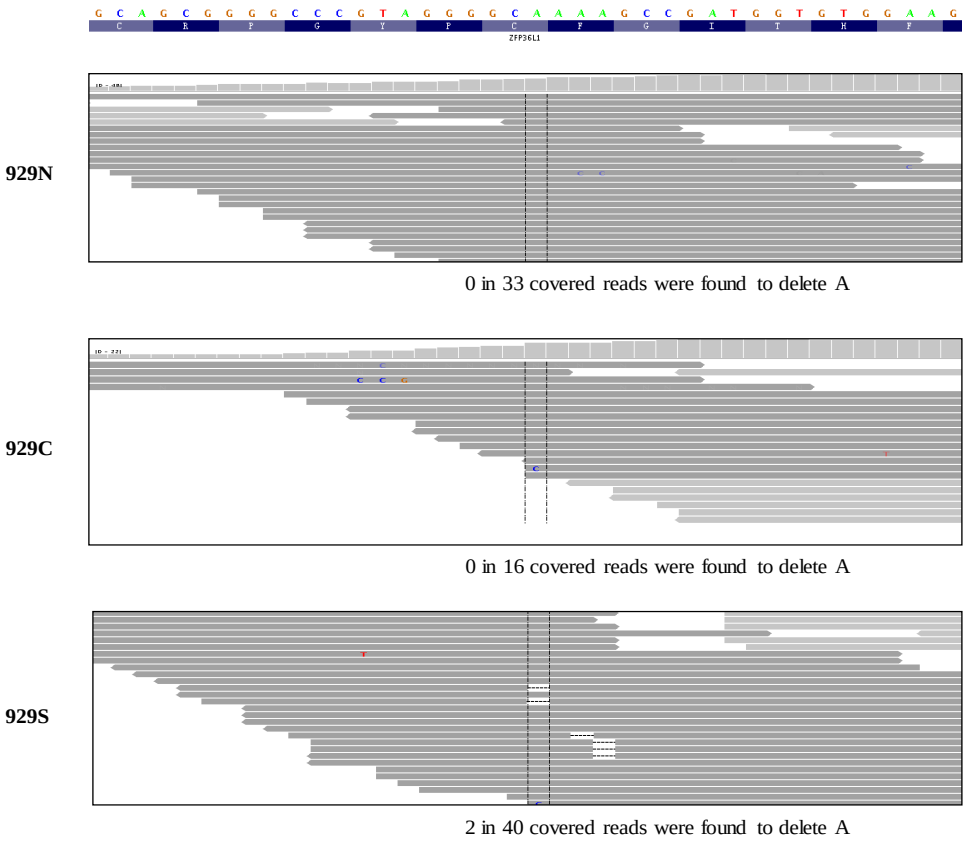
9 in 83 covered reads were found to insert C



929S

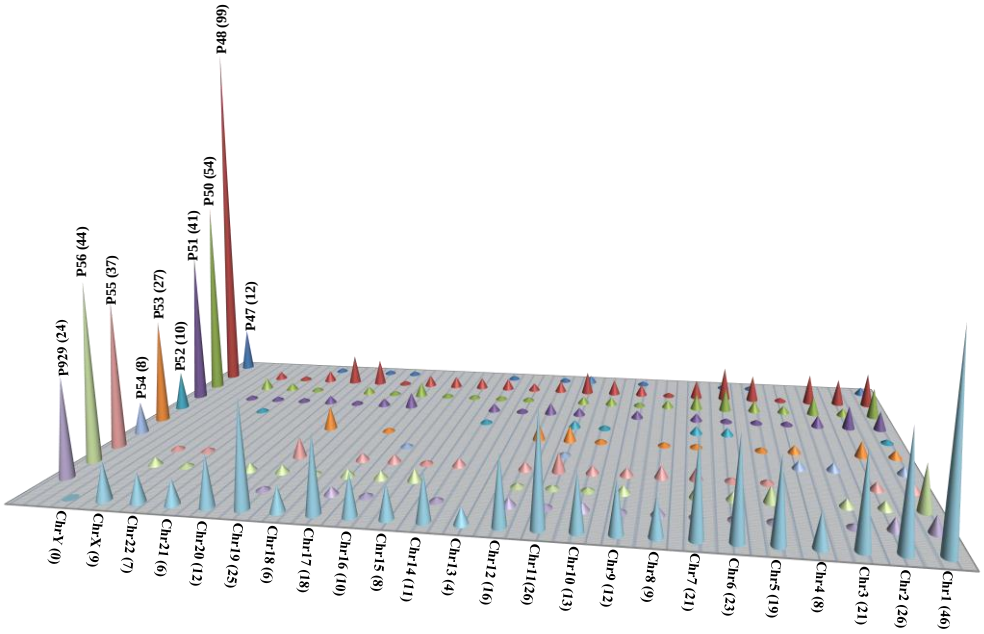
1 in 70 covered reads were found to insert C

Nature Genetics doi:10/1038/ng.2391



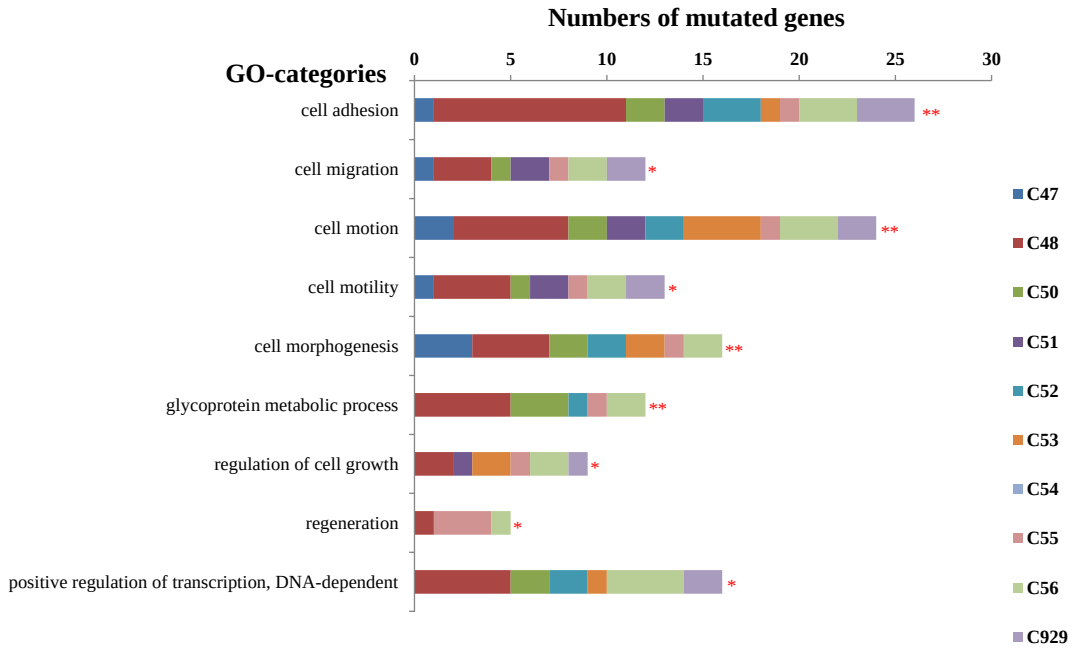
Supplementary Figure 5. Twenty-five somatic indel variants were manually confirmed by using Integrative Genomics Viewer (IGV).

Nature Genetics doi:10/1038/ng.2391



Supplementary Figure 6. The diversity of somatic mutations in advanced HCC. A three-dimensional map depicting the landscape of 356 non-synonymous somatic mutations, including 331 SNV mutations and 25 indels, distributed across 24 chromosomes in ten HCC patients. The total number of mutations in a given patient or chromosome is indicated within parentheses.

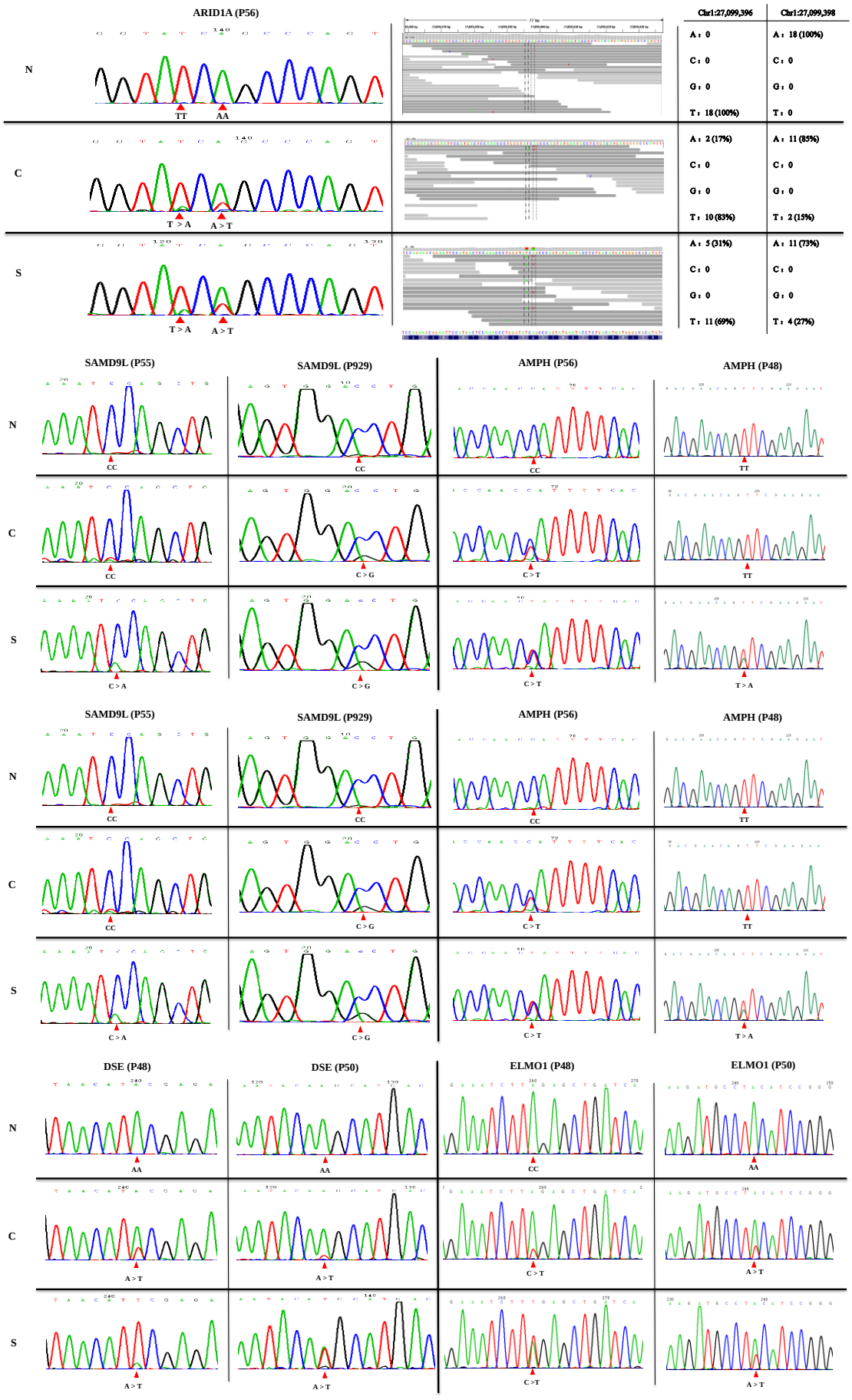
Nature Genetics doi:10/1038/ng.2391



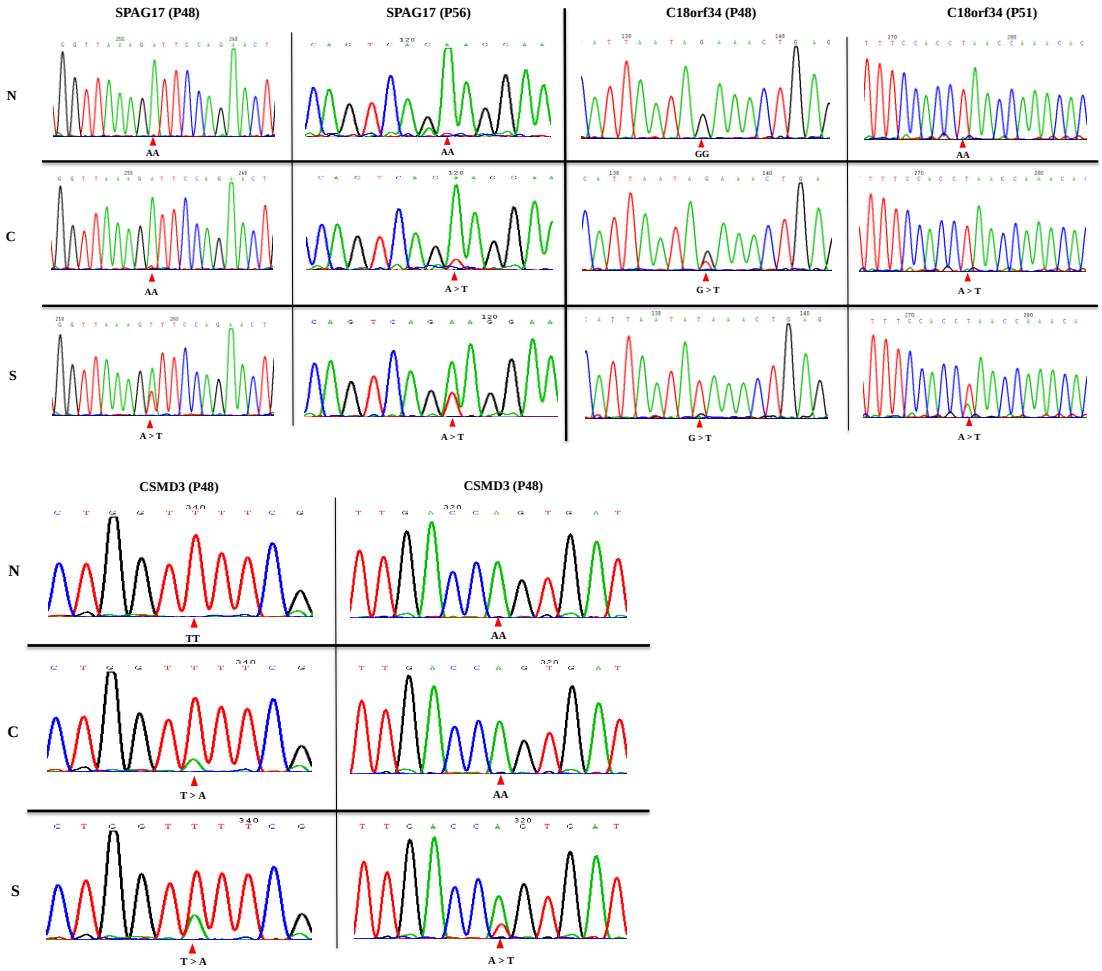
Supplementary Figure 7. Functional categories of the mutated genes.

The mutated genes were assigned to different Gene Ontology (GO) categories based on their biological processes and were statistically analyzed with all human genes as the background for comparison. *P < 0.05, **P < 0.01.

Nature Genetics doi:10/1038/ng.2391



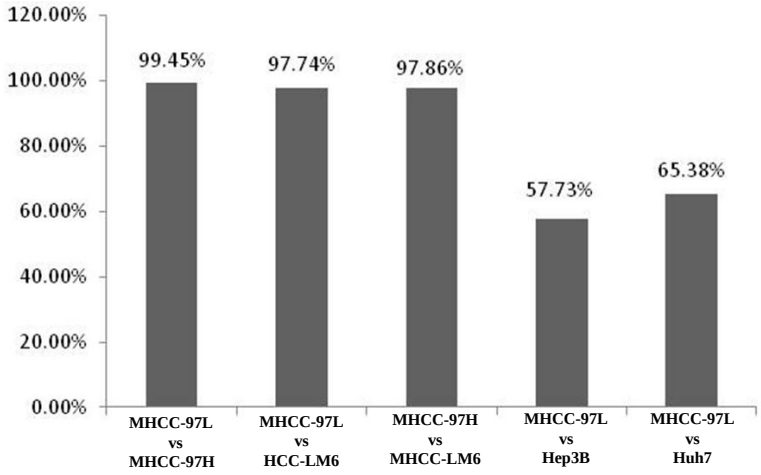
Nature Genetics doi:10/1038/ng.2391



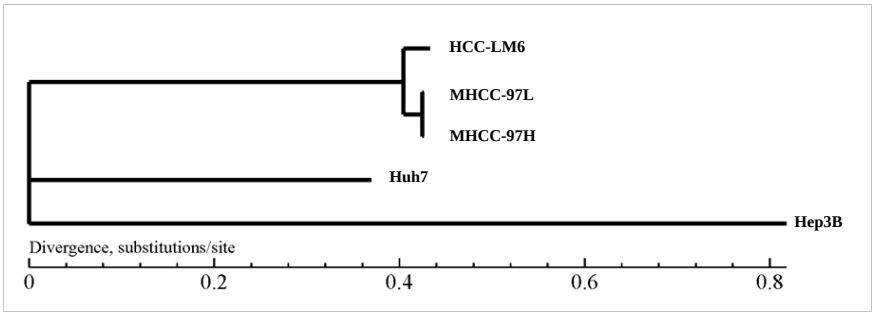
Supplementary Figure 8. Mutated genes harboring two non-silent mutations confirmed by Sanger sequencing. The red triangles indicate the mutation sites. N: adjacent liver, C: primary tumor, S: PVTT.

Nature Genetics doi:10/1038/ng.2391

a

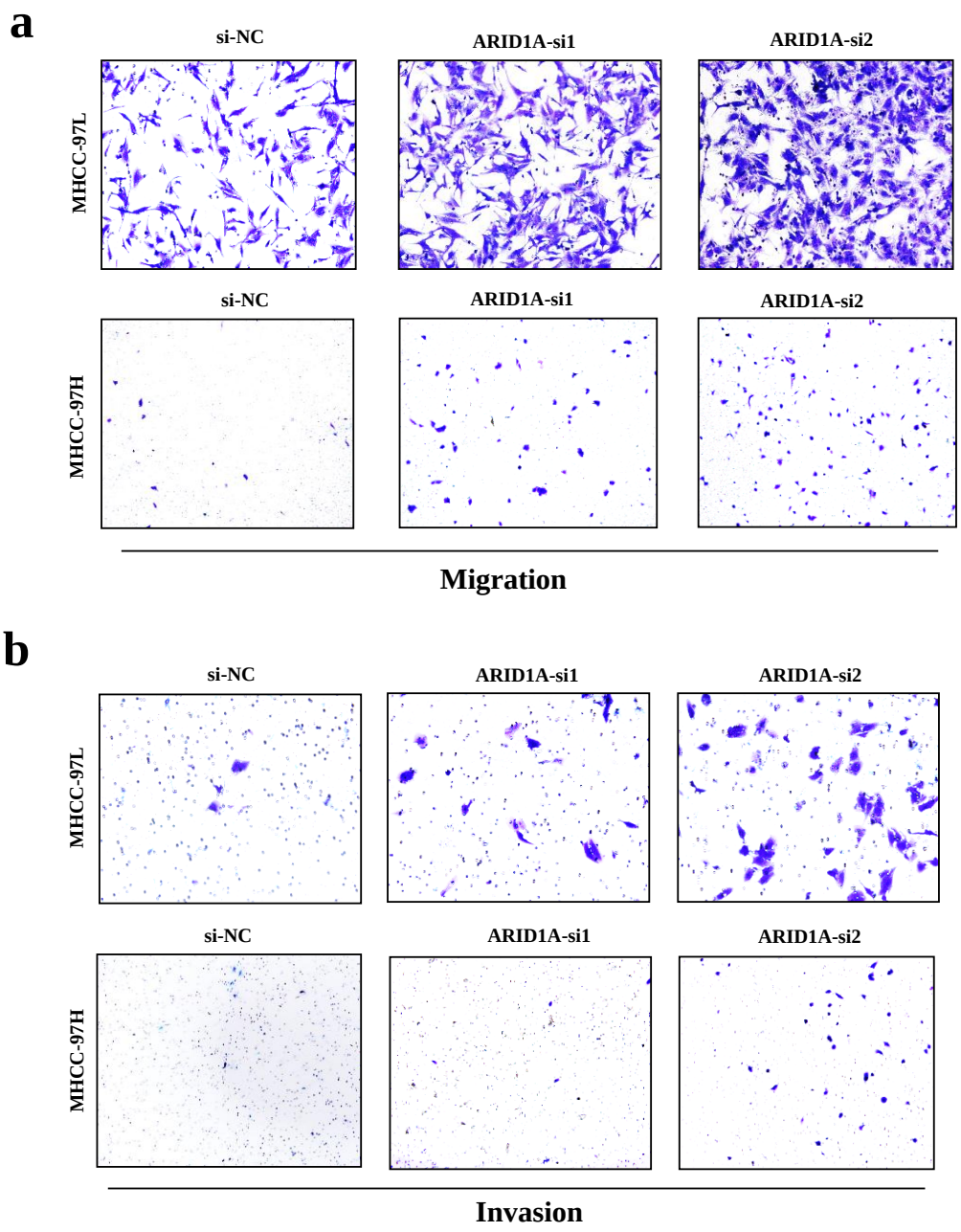


b

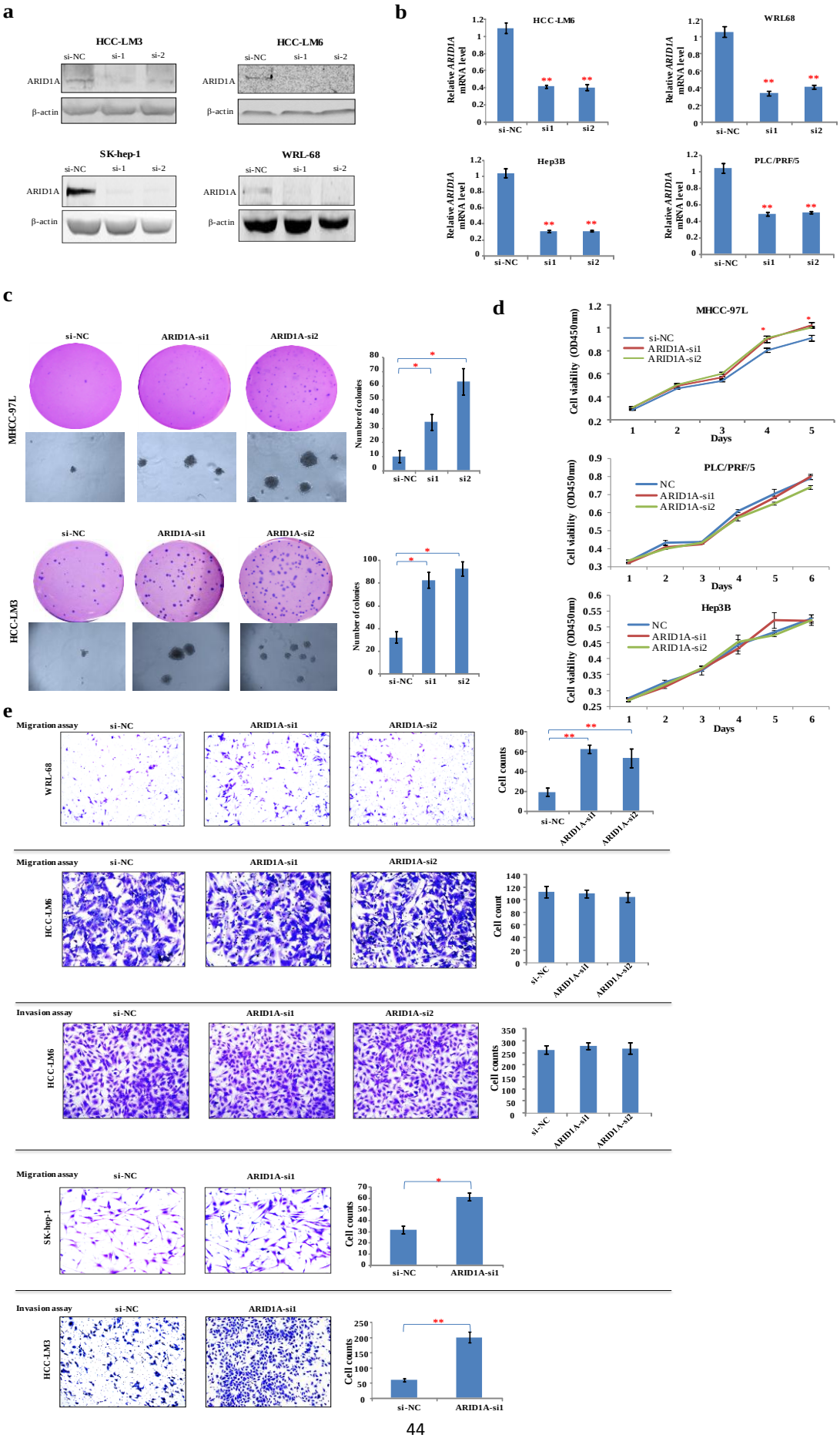


Supplementary Figure 9. Genotype comparison and analysis of some of the HCC cell lines with an Affymetrix Genome-Wide Human SNP Array 5.0 dataset.

Nature Genetics doi:10/1038/ng.2391



Supplementary Figure 10. An ARID1A knockdown via RNA interference promoted cell migration (a) and invasion (b) in MHCC-97L and MHCC-97H cells. The cell migration and invasion were evaluated using Matrigel transwell assays; the permeable cells were stained with crystal violet and counted. The cell migration and invasion assays were repeated three times.



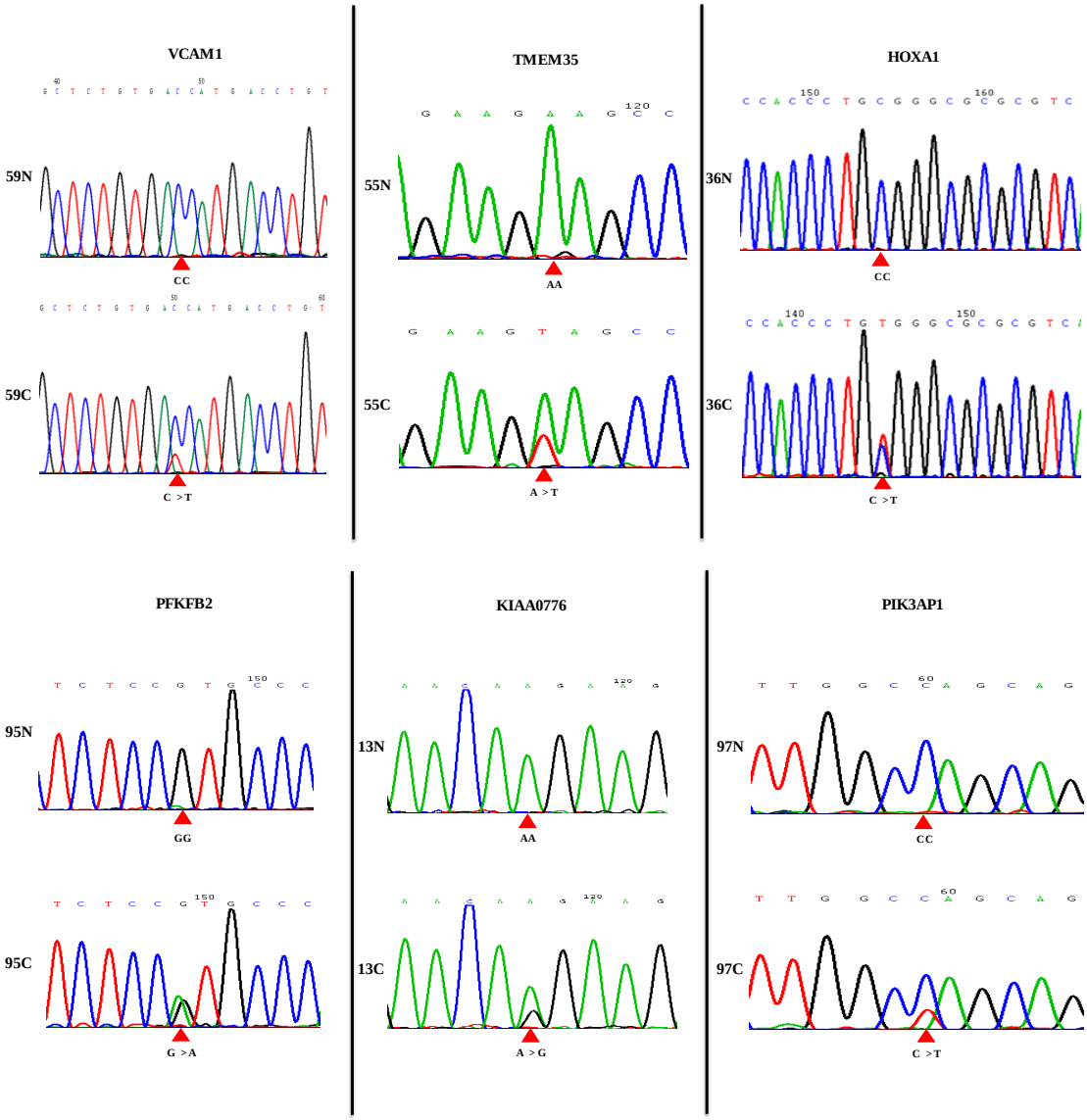
Nature Genetics doi:10/1038/ng.2391

Supplementary Figure 11. Functional analysis of ARID1A via RNAi knockdown in various HCC cell lines. (**a**) The efficacies of two siRNAs (si-1 and si-2) against ARID1A were evaluated with western blotting in HCC-LM3, HCC-LM6, SK-hep-1 and WRL68 cells, β-actin was used as a loading control. (**b**) The *ARID1A* mRNA expression levels were examined with real-time quantitative PCR to evaluate the efficacies of the two siRNAs (si-1 and si-2) in the HCC-LM6, WRL68, Hep3B and PLC/PRF/5 cell lines. The statistical significance of the *ARID1A* kncokdown by siRNAs was calculated with two-tailed t-tests. *, $P < 0.05$; **, $P < 0.01$. (**c**) The role of ARID1A in cell proliferation was assessed with an anchorage-independent colony formation assay after the ARID1A knockdown in MHCC-97L and HCC-LM3 cells. The colonies grown in soft-agar medium were stained with crystal violet, and photographs (upper) and microphotographs (lower) were taken to count colonies (right). The significance of the difference between colony numbers was calculated with a two-tailed t-test. *, $P < 0.05$; **, $P < 0.01$. (**d**) The growth curves of MHCC-97L, PLC/PRF/5 and Hep3B cells after the ARID1A knockdown were assessed using cell counting kit-8 (CCK-8). (**e**) Cell migration and invasion were evaluated with Matrigel transwell assays following ARID1A knockdown via RNA interference in various cell lines. HCC-LM6 cells with the *ARID1A* mutations were used as a negative control. The permeable cells were stained with crystal violet and

Nature Genetics doi:10/1038/ng.2391

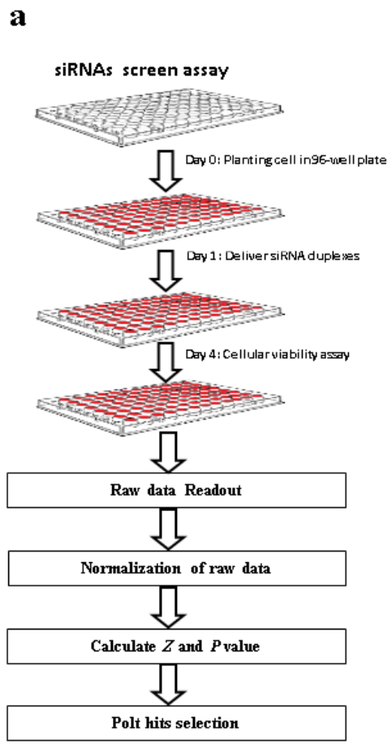
counted. The cell migration and invasion assays were repeated three times, and the results were statistically analyzed with a two-tailed t-test. * $P < 0.05$; ** $P < 0.01$.

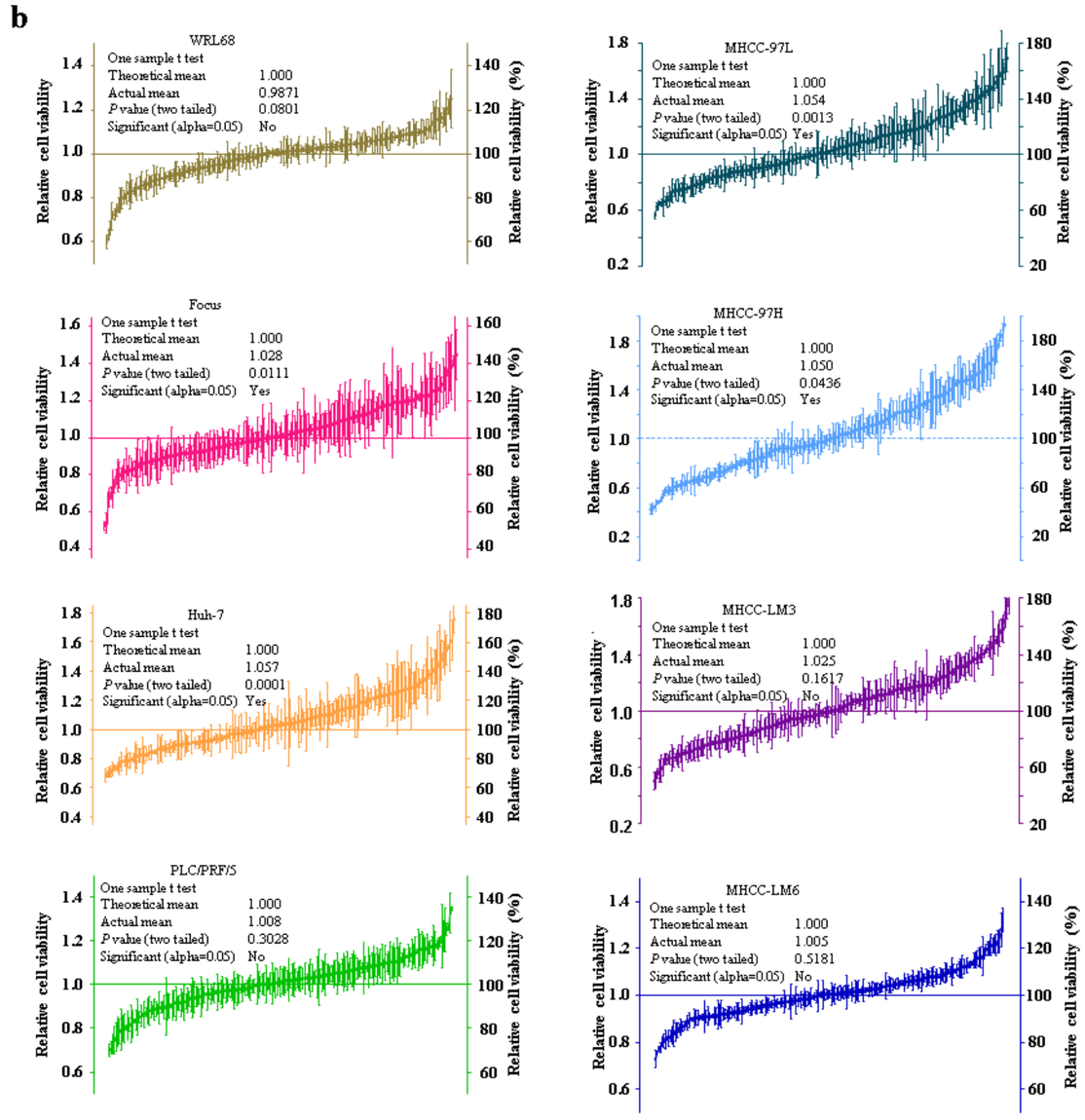
Nature Genetics doi:10/1038/ng.2391

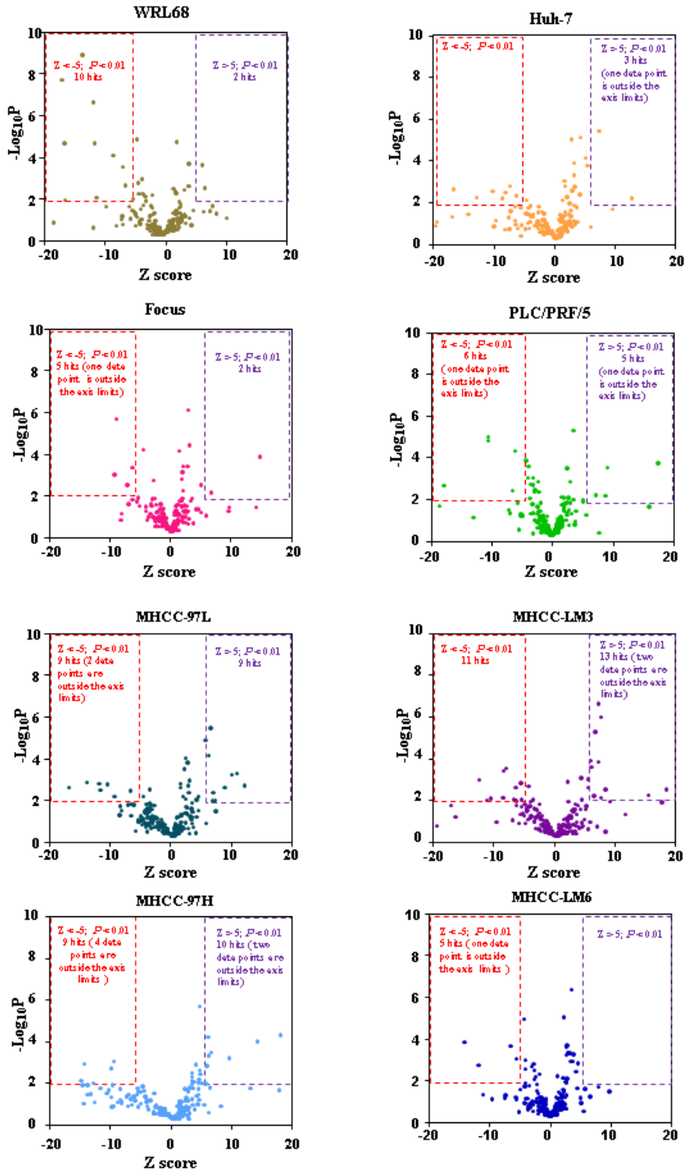


Supplementary Figure 12. The mutated genes were further examined via Sanger sequencing in additional HCC samples.

Nature Genetics doi:10/1038/ng.2391



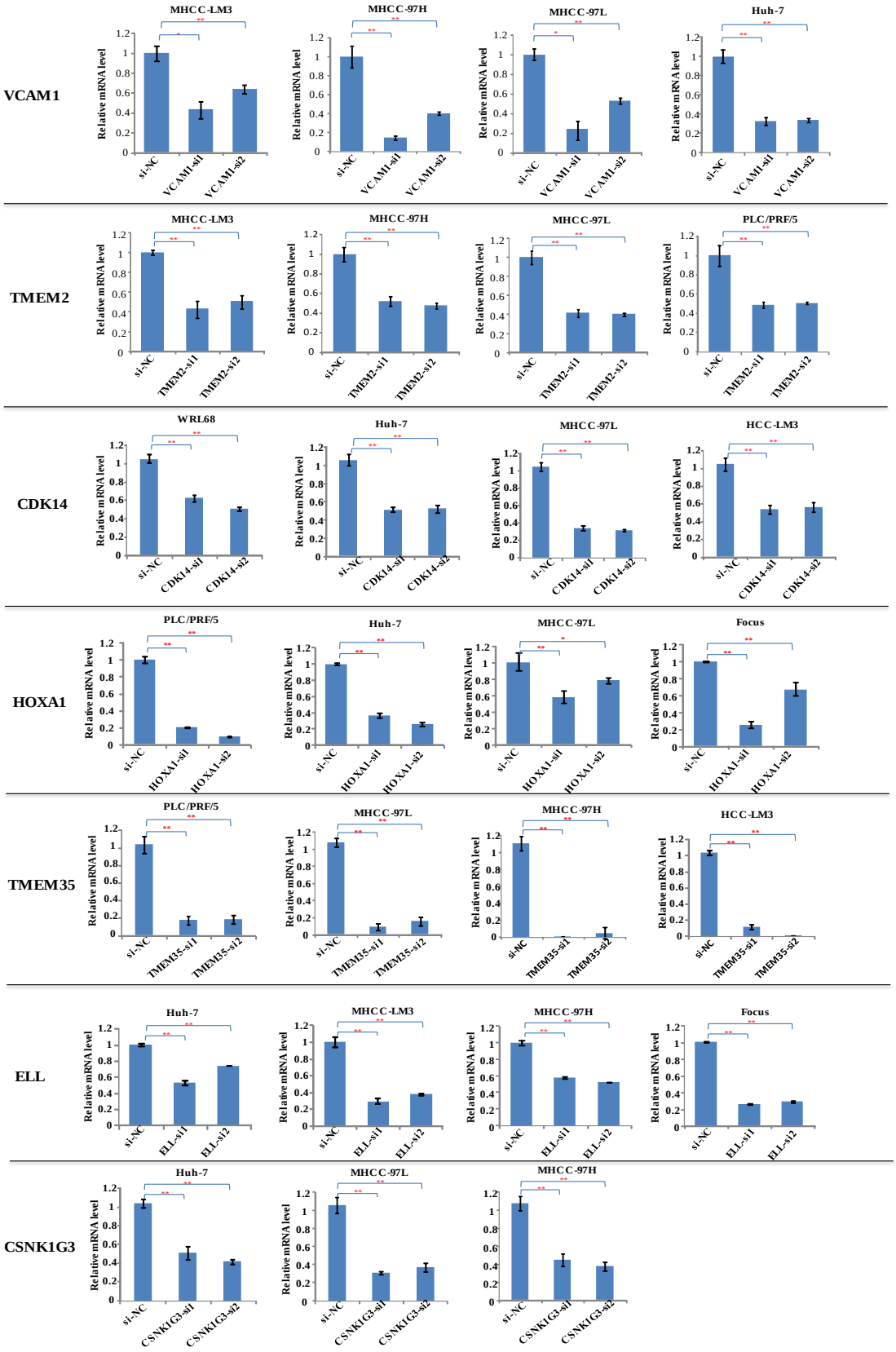




Nature Genetics doi:10/1038/ng.2391

Supplementary Figure 13. Screen of knockdowns by RNA interference of the mutated genes in eight HCC cell lines. (**a**) A flowchart of the cell viability assay screen. The cells were seeded in 96-well culture plates for 24 hours of culture, and the siRNAs were delivered to the adherent cells. After 72 hours, the cell viability was assayed with the Cell Counting Kit-8. The raw absorbance value readouts were normalized, and *Z* scores and *P* values were calculated to evaluate the siRNA effects. (**b**) The output profile created from the normalized readout data from the RNAi screen. Overall, the knockdown of the majority of the genes had no significant effect on cell viability in 4 of the eight HCC cell lines (one-sample t-test; WRL68: $P = 0.0801$; PLC/PRF/5: $P = 0.3028$; MHCC-LM3: $P = 0.1617$; MHCC-LM6, $P = 0.5181$); In contrast, the loss of function of these genes may have influenced the cell viability of the other four HCC lines (one-sample t-test; Focus, $P = 0.0111$; Huh-7, $P = 0.0001$; MHCC-97L, $P = 0.0013$; MHCC-97H, $P = 0.0436$). (**c**) The *Z* score was calculated to measure the generated phenotype. In addition, the *P* value was calculated, using one-tailed t-test to assess statistical significance of the difference. Hits in each cell line were determined to promote or inhibit cell viability using the cutoffs of, $Z < -5$, $p < 0.01$ and $Z > 5$, $p < 0.01$, respectively.

Nature Genetics doi:10/1038/ng.2391

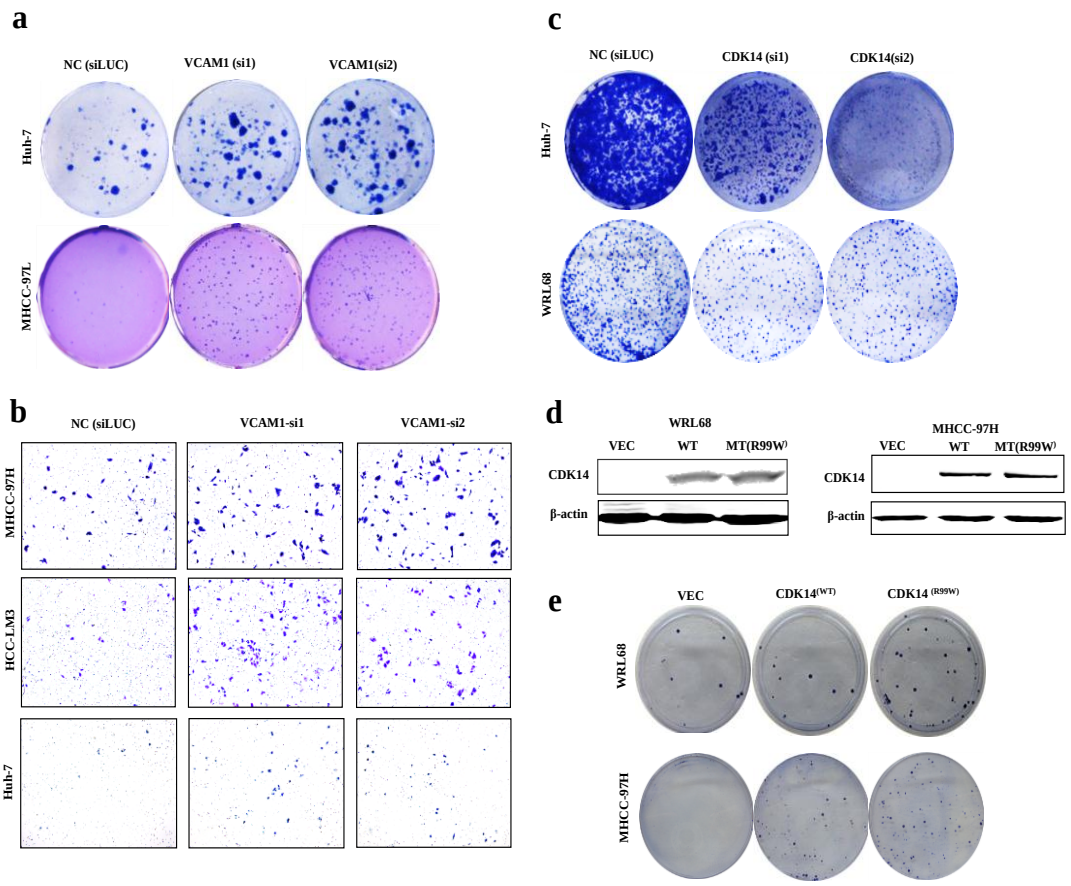


Supplementary Figure 14. The efficacies of RNA interference against VCAM1, TMEM2, CDK14, HOXA1, TMEM35, ELL and CSNK1G3

Nature Genetics doi:10/1038/ng.2391

in the HCC cells were evaluated with real-time quantitative PCR. The statistical significance of the gene knockdown results was determined usig a two-tailed t-test. * $P < 0.05$; ** $P < 0.01$.

Nature Genetics doi:10/1038/ng.2391

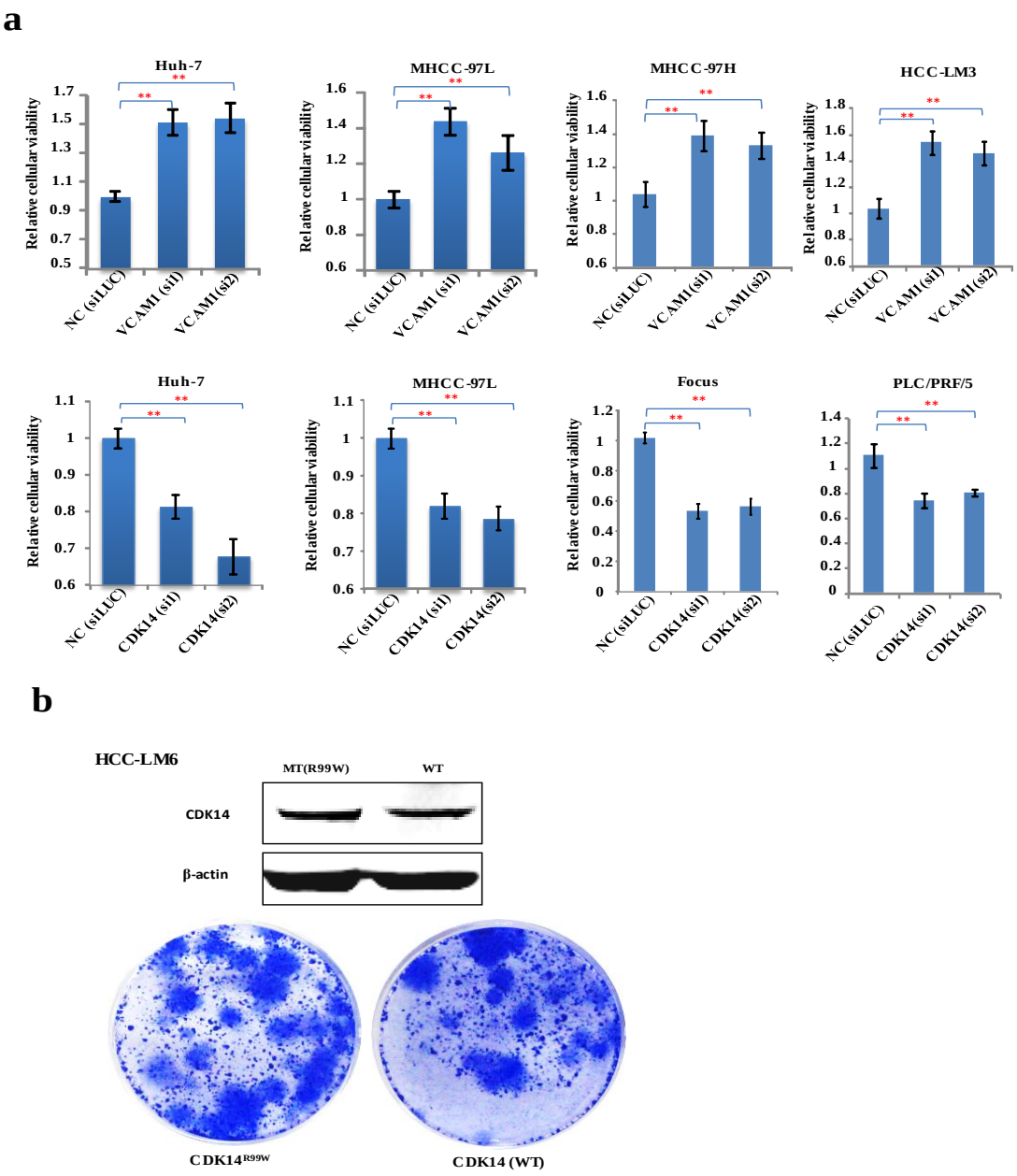


Supplementary Figure 15. The effects of the mutated genes on cell behaviors determined using cell-based RNA interference (RNAi). (a) Anchorage-dependent (upper) and -independent (lower) colony formation was assessed via RNAi against VCAM1 in the Huh-7 and MHCC-97L cell lines. (b) The effect of VCAM1 on cell migration was evaluated in MHCC-97H, HCC-LM3 and Huh-7 cells via RNAi. (c) The effect of CDK14 on colony formation was evaluated in Huh-7 and WRL68 cells via RNAi. (d) The expression of the vectors carrying wild type and mutant CDK14 ($CDK14^{R99W}$) in WRL68 and MHCC-97H cells was evaluated with Western blotting (lower); β-actin was used as a loading control. (e) The effect of mutant $CDK14^{R99W}$ on colony formation was

Nature Genetics doi:10/1038/ng.2391

evaluated in the WRL68 and MHCC-97H cell lines, with the wild-type CDK14 and the empty vector used as controls.

Nature Genetics doi:10/1038/ng.2391

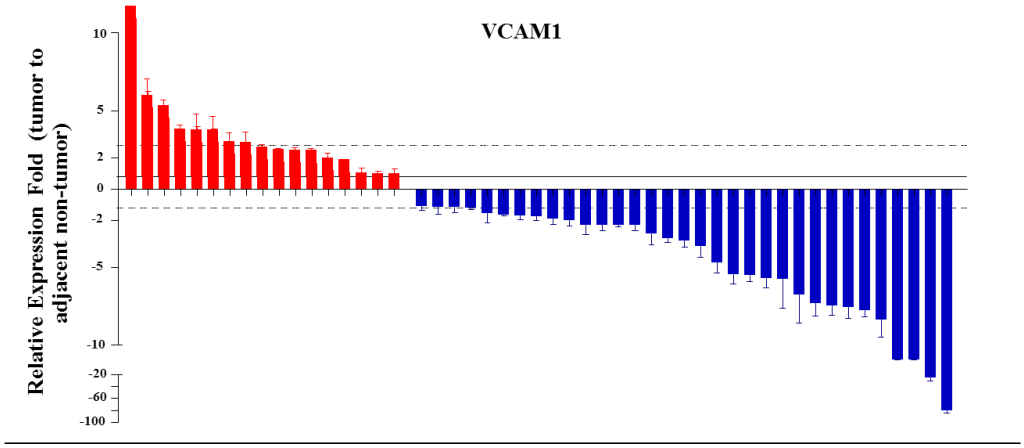


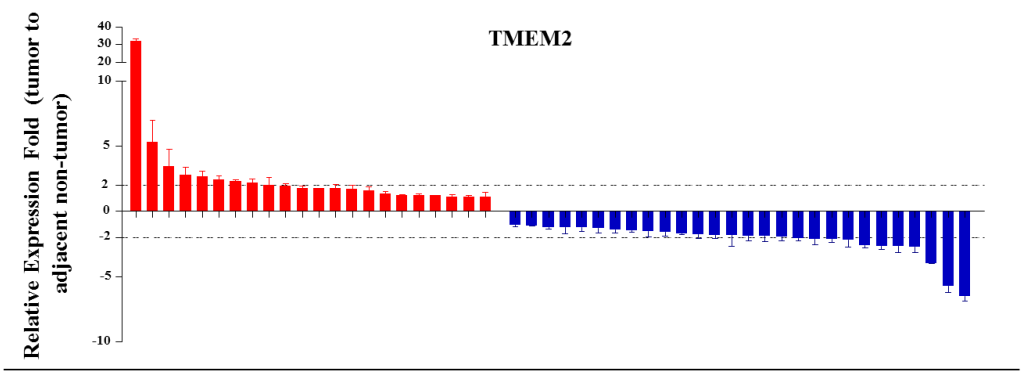
Supplementary Figure 16. Functional assays of VCAM1 and CDK14 in various HCC cell lines. (**a**) The cell viability results in response to the siRNA-mediated loss of gene function of VCAM1 (upper) and CDK14 (lower) are presented. The statistical significance was determined with a two-tailed t-test. *, $P < 0.05$, **, $P < 0.01$. (**b**) The effect of the $CDK14^{R99W}$ mutation on the HCC-LM6 cells was evaluated with a colony formation assay (lower), with wild type CDK14 used as a control.

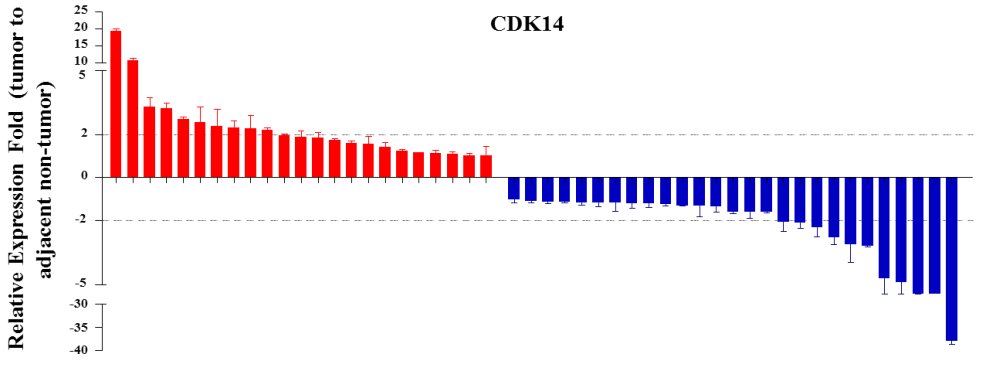
Nature Genetics doi:10/1038/ng.2391

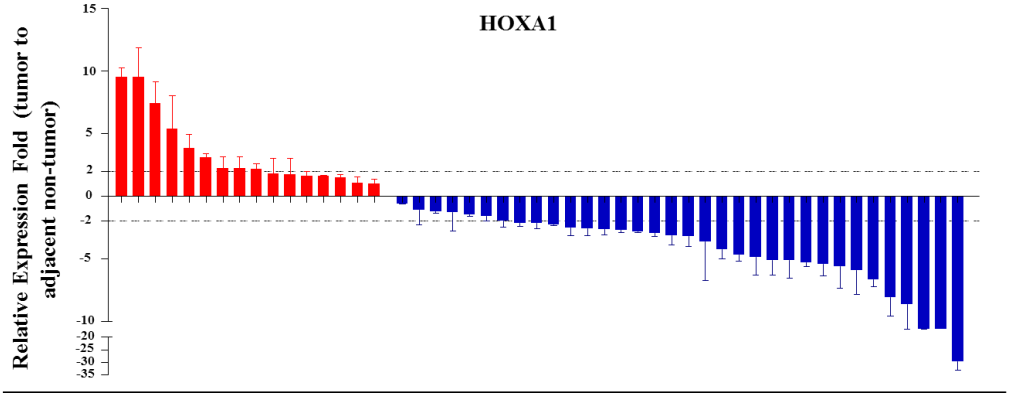
Western blotting (upper) revealed the expression of wild type and mutant CDK14 in the HCC-LM6 cells.

Nature Genetics doi:10/1038/ng.2391

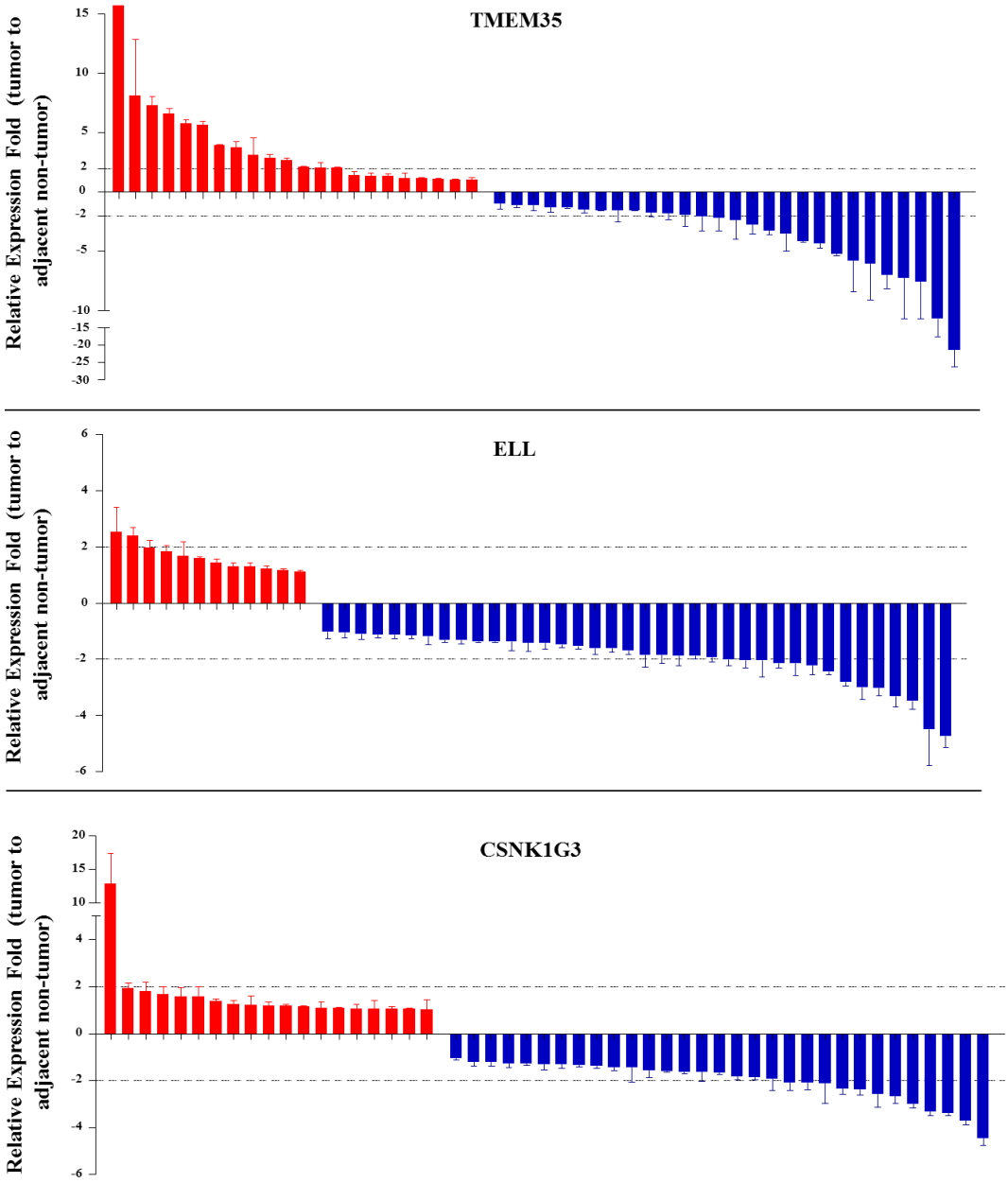








Nature Genetics doi:10/1038/ng.2391



Supplementary Figure 17. The expression patterns in HCC samples of 7 cancer-associated genes from the RNAi screen.

Nature Genetics doi:10/1038/ng.2391

Supplementary Table 1. The clinical and pathological features of the ten HCC patients whose samples were used for exomic sequencing.

Patient IDs	Gender	Age	Cell type	Differentiation stages	Size of primary tumors (cm)	HBV/HAV/HCV	AFP (ng/ml)	PVTT
P47	Male	31	HCC	Low	> 15	HbsAg(+), HAV(-), HCV(-)	1000	+
P48	Male	42	HCC	Low	> 15	HbsAg(+), HAV(-), HCV(-)	317	+
P50	Male	51	HCC	Middle	12 × 10	HbsAg(+), HAV(-), HCV(-)	49610	+
P51	Male	80	HCC	Low to middle	4 × 4	HbsAg(+), HAV(-), HCV(-)	654	+
P52	Female	69	HCC	Middle	11 × 9	HbsAg(-), HAV(-), HCV(-)	0.887	+
P53	Male	43	HCC	Middle	8 × 6	HbsAg(+), HAV(-), HCV(-)	4609	+
P54	Male	66	HCC	Low to middle	3.5 × 3.5	HbsAg(+), HAV(-), HCV(-)	717	+
P55	Male	54	HCC	Low to middle	12 × 10	HbsAg(-), HAV(-), HCV(-)	111	+
P56	Male	50	HCC	Low	3 × 4	HbsAg(+), HAV(-), HCV(-)	51.8	+
P929	Male	54	HCC	Low	7 × 8	HbsAg(+), HAV(-), HCV(-)	25410	+

Nature Genetics doi:10/1038/ng.2391

Supplementary Table 2. The exomic sequencing profiles of ten HCC patients with PVTTs.

HCC patients ID		Solexa						
		Total Target BP	Bases covered on targeted regions	Average depth of sequencing on targeted regions (fold)	Percent of target bp covered at ⩾ 1×	Percent of target bp covered at ⩾ 4×	Percent of target bp covered at ⩾ 10×	Percent of target bp covered at ⩾ 20×
P47	**P47_N**	34108810	32964609	12.77	96.60%	86.30%	56.10%	19.90%
	P47_C	34108810	32944866	13.05	96.60%	85.80%	55.80%	20.80%
P48	**P48_N**	34108810	32819440	14.1	96.20%	84.40%	58.30%	25.60%
	P48_C	34108810	33660337	22.84	98.70%	93.50%	77.10%	48.30%
	P48_S	34108810	33378855	16.9	97.90%	90.20%	67.40%	32.90%
P50	**P50_N**	34108810	33551881	22.97	98.40%	91.50%	73.00%	48.30%
	P50_C	34108810	32813641	19.41	96.20%	84.40%	63.10%	39.00%
	P50_S	34108810	32921996	14.51	96.50%	85.80%	58.40%	25.40%
P51	**P51_N**	34108810	32993262	14.64	96.70%	86.90%	60.40%	26.10%
	P51_C	34108810	32989954	14.59	96.70%	86.60%	60.10%	25.90%
	P51_S	34108810	32837181	14.01	96.30%	83.60%	55.40%	23.90%
P52	**P52_N**	34108810	32898333	12.22	96.70%	85.30%	54.40%	17.90%
	P52_C	34108810	32919005	12.88	96.70%	84.90%	55.90%	20.70%
	P52_S	34108810	33060985	13.62	97.10%	86.50%	58.80%	22.90%

Nature Genetics doi:10/1038/ng.2391

P53	**P53_N**	34108810	32821727	14	96.20%	85.20%	57.40%	24.00%
	P53_C	34108810	32837822	13.78	96.30%	85.10%	56.80%	23.40%
	P53_S	34108810	32998745	13.89	96.70%	86.30%	57.90%	23.60%
P54	**P54_N**	34108810	32868575	22.1	96.40%	83.50%	62.60%	41.10%
	P54_C	34108810	33166903	12.4	97.20%	86.40%	56.00%	18.50%
	P54_S	34108810	33151299	28.4	97.20%	88.20%	71.90%	50.40%
P55	**P55_N**	34108810	32774741	13.48	96.10%	83.70%	55.50%	22.90%
	P55_C	34108810	33119227	13.01	97.10%	86.20%	57.30%	20.70%
	P55_S	34108810	32645825	11.73	95.70%	80.10%	46.90%	16.90%
P56	**P56_N**	34108810	32495095	13.97	95.30%	78.40%	44.80%	15.20%
	P56_C	34108810	33157949	12.81	97.20%	86.80%	55.80%	19.10%
	P56_S	34108810	32827036	12.36	96.20%	82.10%	49.90%	18.30%
P929	**P929_N**	34108810	32886528	13	96.40%	87.90%	60.50%	20.00%
	P929_C	34108810	32430656	12.81	95.10%	81.80%	52.70%	21.90%
	P929_S	34108810	32752135	11.99	96.00%	85.50%	54.10%	16.80%

HCC patients ID		**SOLID**						
		Total Target BP	Bases covered on targeted regions	Average depth of sequencing on targeted regions (fold)	Percent of target bp covered at ⩾ 1×	Percent of target bp covered at ⩾ 5×	Percent of target bp covered at ⩾ 10×	Percent of target bp covered at ⩾ 20×
P47	**P47_N**	38754065	35758043	48.65	92.27%	86.18%	79.58%	67.25%

Nature Genetics doi:10/1038/ng.2391

	P47_C	38754065	35592721	46.5	91.84%	84.42%	76.97%	63.86%
P48	**P48_N**	38754065	36087108	46.72	93.12%	87.78%	81.38%	68.28%
	P48_C	38754065	36251967	52.4	93.54%	89.25%	83.69%	71.43%
	P48_S	38754065	35707181	54.37	92.14%	86.41%	80.35%	68.66%
P50	**P50_N**	—	—	—	—	—	—	—
	P50_C	38754065	36018601	48.28	92.94%	87.04%	80.41%	67.57%
	P50_S	38754065	35860025	50.71	92.53%	86.04%	79.21%	66.79%
P51	**P51_N**	38754065	35878593	48.77	92.58%	87.22%	81.10%	68.61%
	P51_C	38754065	35951236	49.64	92.77%	87.21%	81.02%	68.66%
	P51_S	38754065	36179307	49.44	93.36%	87.84%	81.25%	68.04%
P52	**P52_N**	38754065	36009202	42.76	92.92%	87.11%	79.84%	64.92%
	P52_C	38754065	35613708	47.8	91.90%	85.83%	79.46%	67.00%
	P52_S	—	—	—	—	—	—	—
P53	**P53_N**	—	—	—	—	—	—	—
	P53_C	38754065	33601921	46.15	86.71%	78.38%	71.90%	60.64%
	P53_S	38754065	35157077	54.24	90.72%	85.13%	79.74%	69.13%
P54	**P54_N**	—	—	—	—	—	—	—
	P54_C	38754065	33527625	49.34	86.51%	78.04%	72.17%	62.29%
	P54_S	38754065	32242094	52.94	83.20%	72.93%	66.88%	58.16%

Nature Genetics doi:10/1038/ng.2391

P55	**P55_N**	38754065	34972901	45.25	90.24%	82.29%	75.24%	63.29%
	P55_C	38754065	35838924	39.35	92.48%	85.71%	77.74%	61.96%
	P55_S	38754065	35844234	47.46	92.49%	86.50%	79.87%	66.77%
P56	**P56_N**	38754065	36042287	52.11	93.00%	87.91%	81.96%	70.24%
	P56_C	38754065	34772411	49.87	89.62%	82.48%	76.56%	65.82%
	P56_S	38754065	35749166	45.05	92.25%	86.24%	79.55%	66.15%
P929	**P929_N**	38754065	36156204	58.03	93.30%	88.71%	83.59%	73.37%
	P929_C	38754065	33482268	46.19	86.29%	77.23%	70.39%	59.24%
	P929_S	38754065	36304601	50.95	93.68%	89.14%	83.35%	70.85%

HCC patients ID		**Actual overlap between NimbleGen and Agilent**		
		Bases covered by Solexa and Solid sequencing reads	Percent of expected coverage in Nimblegen capture	Percent of expected coverage in Agilent capture
P47	**P47_N**	30402144	89.13%	78.45%
	P47_C	30260463	88.72%	78.08%
P48	**P48_N**	30555163	89.58%	78.84%
	P48_C	31490611	92.32%	81.26%
	P48_S	30767872	90.21%	79.39%
P50	**P50_N**	—	—	—
	P50_C	30496100	89.41%	78.69%
	P50_S	30456251	89.29%	78.59%

Nature Genetics doi:10/1038/ng.2391

P51	**P51_N**	30535864	89.52%	78.79%
	P51_C	30598532	89.71%	78.96%
	P51_S	30665757	89.91%	79.13%
P52	**P52_N**	30648007	89.85%	79.08%
	P52_C	30311578	88.87%	78.22%
	P52_S	—	—	—
P53	**P53_N**	—	—	—
	P53_C	28481446	83.50%	73.49%
	P53_S	29922376	87.73%	77.21%
P54	**P54_N**	—	—	—
	P54_C	28681320	84.09%	74.01%
	P54_S	27583931	80.87%	71.18%
P55	**P55_N**	29579378	86.72%	76.33%
	P55_C	30629056	89.80%	79.03%
	P55_S	30190707	88.51%	77.90%
P56	**P56_N**	30230297	88.63%	78.01%
	P56_C	29712403	87.11%	76.67%
	P56_S	30269693	88.74%	78.11%
P929	**P929_N**	30677873	89.94%	79.16%

Nature Genetics doi:10/1038/ng.2391

	P929_C	27990300	82.06%	72.23%
	P929_S	30675008	89.93%	79.15%

Note: N, matched non-tumorous livers; C, primary tumors; S, PVTTs

Nature Genetics doi:10/1038/ng.2391

Supplementary Table 6. Summary of the types and prevalence of the somatic mutations found in 10 HCC patients.

Sample	Synonymous	missense	nonsense	Indel	Total	Mutations per Mb DNA	NS/S
P47	4	10	1	1	16	0.454	2.75
P48	28	85	8	6	127	3.607	3.32
P50	14	47	3	4	68	1.932	3.57
P51	9	34	3	4	50	1.420	4.11
P52	2	8	1	1	12	0.341	4.50
P53	19	23	2	2	46	1.307	1.32
P54	5	7	0	1	13	0.369	1.40
P55	8	34	1	2	45	1.278	4.38
P56	17	38	4	2	61	1.733	2.47
P929	6	21	1	2	30	0.852	3.67
Average	11.2	30.7	2.4	2.5	46.8	1.329	2.96

Nature Genetics doi:10/1038/ng.2391

Supplementary Table 8. The signaling pathways and biological processes of the mutated genes.

Category	Term	Count	P-Value	Fold Enrichment
GOTERM_BP_FAT	cell motion	23	1.70E-04	2.5
GOTERM_BP_FAT	neuron projection morphogenesis	14	3.00E-04	3.3
GOTERM_BP_FAT	axonogenesis	13	4.20E-04	3.4
GOTERM_BP_FAT	neuron projection development	15	5.40E-04	3
GOTERM_BP_FAT	cell morphogenesis involved in neuron differentiation	13	8.60E-04	3.2
GOTERM_BP_FAT	cell projection morphogenesis	14	1.10E-03	2.9
GOTERM_BP_FAT	cell part morphogenesis	14	1.70E-03	2.8
GOTERM_BP_FAT	cellular component morphogenesis	18	2.20E-03	2.3
GOTERM_BP_FAT	cell projection organization	17	2.50E-03	2.3
GOTERM_BP_FAT	cell adhesion	26	2.80E-03	1.9
GOTERM_BP_FAT	biological adhesion	26	2.80E-03	1.9
GOTERM_BP_FAT	cell morphogenesis involved in differentiation	13	3.20E-03	2.7
GOTERM_BP_FAT	cell morphogenesis	16	4.50E-03	2.3
GOTERM_BP_FAT	glycoprotein metabolic process	11	6.50E-03	2.8
GOTERM_BP_FAT	neuron development	15	7.10E-03	2.3
GOTERM_BP_FAT	cell-cell adhesion	13	8.30E-03	2.4
GOTERM_BP_FAT	neuron differentiation	17	1.30E-02	2
GOTERM_BP_FAT	glycoprotein biosynthetic process	9	1.30E-02	2.9
KEGG_PATHWAY	Adipocytokine signaling pathway	6	1.50E-02	4.1
GOTERM_BP_FAT	positive regulation of biosynthetic process	23	1.80E-02	1.7
GOTERM_BP_FAT	membrane organization	15	1.80E-02	2
GOTERM_BP_FAT	localization of cell	13	1.80E-02	2.2
GOTERM_BP_FAT	cell motility	13	1.80E-02	2.2
GOTERM_BP_FAT	regulation of dopamine secretion	3	1.90E-02	13.9

Nature Genetics doi:10/1038/ng.2391

GOTERM_BP_FAT	cell migration	12	2.00E-02	2.2
GOTERM_BP_FAT	intracellular signaling cascade	36	2.10E-02	1.5
GOTERM_BP_FAT	positive regulation of organelle organization	6	2.30E-02	3.7
GOTERM_BP_FAT	regulation of organelle organization	10	2.70E-02	2.3
GOTERM_BP_FAT	positive regulation of cellular biosynthetic process	22	2.90E-02	1.6
GOTERM_BP_FAT	forebrain development	8	3.00E-02	2.7
GOTERM_BP_FAT	synapse organization	5	3.10E-02	4.2
GOTERM_BP_FAT	positive regulation of transcription from RNA polymerase II promoter	14	3.10E-02	1.9
KEGG_PATHWAY	Olfactory transduction	15	3.50E-02	1.8
GOTERM_BP_FAT	vesicle-mediated transport	19	3.50E-02	1.7
GOTERM_BP_FAT	regulation of cell growth	9	3.70E-02	2.4
GOTERM_BP_FAT	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	20	3.80E-02	1.6
GOTERM_BP_FAT	biopolymer glycosylation	7	4.00E-02	2.8
GOTERM_BP_FAT	protein amino acid glycosylation	7	4.00E-02	2.8
GOTERM_BP_FAT	glycosylation	7	4.00E-02	2.8
GOTERM_BP_FAT	extracellular structure organization	8	4.20E-02	2.5
GOTERM_BP_FAT	homophilic cell adhesion	7	4.40E-02	2.7
GOTERM_BP_FAT	regeneration	5	4.60E-02	3.7
GOTERM_BP_FAT	regulation of secretion	9	4.60E-02	2.3
GOTERM_BP_FAT	neurological system process	33	4.90E-02	1.4
GOTERM_BP_FAT	positive regulation of transcription, DNA-dependent	16	4.90E-02	1.7
GOTERM_BP_FAT	positive regulation of nitrogen compound metabolic process	20	5.00E-02	1.6
KEGG_PATHWAY	Fatty acid metabolism	4	5.60E-02	4.5
KEGG_PATHWAY	Cell adhesion molecules (CAMs)	7	6.80E-02	2.4
KEGG_PATHWAY	ABC transporters	4	7.00E-02	4.1
KEGG_PATHWAY	Neuroactive ligand-receptor interaction	10	1.10E-01	1.8
KEGG_PATHWAY	Colorectal cancer	5	1.10E-01	2.7

Nature Genetics doi:10/1038/ng.2391

KEGG_PATHWAY	Hedgehog signaling pathway	4	1.20E-01	3.2
KEGG_PATHWAY	O-Glycan biosynthesis	3	1.40E-01	4.5
KEGG_PATHWAY	Tight junction	6	1.70E-01	2
KEGG_PATHWAY	Fructose and mannose metabolism	3	1.70E-01	4
KEGG_PATHWAY	Wnt signaling pathway	6	2.30E-01	1.8
KEGG_PATHWAY	Endometrial cancer	3	3.10E-01	2.6
KEGG_PATHWAY	Basal cell carcinoma	3	3.40E-01	2.5
KEGG_PATHWAY	Insulin signaling pathway	5	3.40E-01	1.7
KEGG_PATHWAY	Huntington's disease	6	3.60E-01	1.5
KEGG_PATHWAY	GnRH signaling pathway	4	3.60E-01	1.9
KEGG_PATHWAY	T cell receptor signaling pathway	4	4.20E-01	1.7
KEGG_PATHWAY	Vascular smooth muscle contraction	4	4.40E-01	1.6
KEGG_PATHWAY	Focal adhesion	6	4.50E-01	1.4
KEGG_PATHWAY	Alzheimer's disease	5	4.80E-01	1.4
KEGG_PATHWAY	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	3	5.00E-01	1.8
KEGG_PATHWAY	Adherens junction	3	5.00E-01	1.8
KEGG_PATHWAY	Fc epsilon RI signaling pathway	3	5.10E-01	1.7
KEGG_PATHWAY	Axon guidance	4	5.40E-01	1.4
KEGG_PATHWAY	ECM-receptor interaction	3	5.50E-01	1.6
KEGG_PATHWAY	Small cell lung cancer	3	5.50E-01	1.6
KEGG_PATHWAY	Hypertrophic cardiomyopathy (HCM)	3	5.60E-01	1.6
KEGG_PATHWAY	Apoptosis	3	5.70E-01	1.6
KEGG_PATHWAY	ErbB signaling pathway	3	5.70E-01	1.6
KEGG_PATHWAY	Pathways in cancer	8	5.80E-01	1.1
KEGG_PATHWAY	Dilated cardiomyopathy	3	6.00E-01	1.5
KEGG_PATHWAY	Fc gamma R-mediated phagocytosis	3	6.20E-01	1.4
KEGG_PATHWAY	Valine, leucine and isoleucine degradation	2	6.20E-01	2.1
KEGG_PATHWAY	Arginine and proline metabolism	2	6.90E-01	1.7

Nature Genetics doi:10/1038/ng.2391

KEGG_PATHWAY	Amyotrophic lateral sclerosis (ALS)	2	6.90E-01	1.7
KEGG_PATHWAY	Regulation of actin cytoskeleton	5	7.00E-01	1.1
KEGG_PATHWAY	Non-small cell lung cancer	2	7.00E-01	1.7
KEGG_PATHWAY	Glycolysis / Gluconeogenesis	2	7.40E-01	1.5
KEGG_PATHWAY	Calcium signaling pathway	4	7.50E-01	1
KEGG_PATHWAY	Glioma	2	7.50E-01	1.4
KEGG_PATHWAY	Neurotrophin signaling pathway	3	7.60E-01	1.1
KEGG_PATHWAY	Parkinson's disease	3	7.80E-01	1.1
KEGG_PATHWAY	p53 signaling pathway	2	7.80E-01	1.3
KEGG_PATHWAY	Epithelial cell signaling in Helicobacter pylori infection	2	7.80E-01	1.3
KEGG_PATHWAY	Chemokine signaling pathway	4	7.80E-01	1
KEGG_PATHWAY	Long-term depression	2	7.80E-01	1.3
KEGG_PATHWAY	PPAR signaling pathway	2	7.80E-01	1.3
KEGG_PATHWAY	Renal cell carcinoma	2	7.90E-01	1.3
KEGG_PATHWAY	Pancreatic cancer	2	8.00E-01	1.3
KEGG_PATHWAY	B cell receptor signaling pathway	2	8.10E-01	1.2
KEGG_PATHWAY	Chronic myeloid leukemia	2	8.10E-01	1.2
KEGG_PATHWAY	Cardiac muscle contraction	2	8.20E-01	1.2
KEGG_PATHWAY	MAPK signaling pathway	5	8.40E-01	0.9
KEGG_PATHWAY	TGF-beta signaling pathway	2	8.60E-01	1
KEGG_PATHWAY	Prostate cancer	2	8.60E-01	1
KEGG_PATHWAY	Melanogenesis	2	8.90E-01	0.9
KEGG_PATHWAY	Toll-like receptor signaling pathway	2	8.90E-01	0.9
KEGG_PATHWAY	Leukocyte transendothelial migration	2	9.30E-01	0.8
KEGG_PATHWAY	Cell cycle	2	9.40E-01	0.7
KEGG_PATHWAY	Oxidative phosphorylation	2	9.50E-01	0.7
KEGG_PATHWAY	Ubiquitin mediated proteolysis	2	9.50E-01	0.7
KEGG_PATHWAY	Purine metabolism	2	9.70E-01	0.6

Nature Genetics doi:10/1038/ng.2391

KEGG_PATHWAY	Jak-STAT signaling pathway	2	9.70E-01	0.6
KEGG_PATHWAY	Endocytosis	2	9.80E-01	0.5
KEGG_PATHWAY	Non-homologous end-joining	1	1.00E+00	3.5
KEGG_PATHWAY	Riboflavin metabolism	1	1.00E+00	2.8
KEGG_PATHWAY	Thiamine metabolism	1	1.00E+00	5.7
KEGG_PATHWAY	Gap junction	1	1.00E+00	0.5
KEGG_PATHWAY	DNA replication	1	1.00E+00	1.3
KEGG_PATHWAY	Butanoate metabolism	1	1.00E+00	1.3
KEGG_PATHWAY	Homologous recombination	1	1.00E+00	1.6
KEGG_PATHWAY	Glutathione metabolism	1	1.00E+00	0.9
KEGG_PATHWAY	Chondroitin sulfate biosynthesis	1	1.00E+00	2.1
KEGG_PATHWAY	Thyroid cancer	1	1.00E+00	1.6
KEGG_PATHWAY	Notch signaling pathway	1	1.00E+00	1
KEGG_PATHWAY	Retinol metabolism	1	1.00E+00	0.8
KEGG_PATHWAY	Limonene and pinene degradation	1	1.00E+00	3.2
KEGG_PATHWAY	SNARE interactions in vesicular transport	1	1.00E+00	1.2
KEGG_PATHWAY	Drug metabolism	1	1.00E+00	0.7
KEGG_PATHWAY	Histidine metabolism	1	1.00E+00	1.6
KEGG_PATHWAY	Ether lipid metabolism	1	1.00E+00	1.3
KEGG_PATHWAY	Alanine, aspartate and glutamate metabolism	1	1.00E+00	1.5
KEGG_PATHWAY	Amino sugar and nucleotide sugar metabolism	1	1.00E+00	1
KEGG_PATHWAY	Lysosome	1	1.00E+00	0.4
KEGG_PATHWAY	Glycerolipid metabolism	1	1.00E+00	1
KEGG_PATHWAY	Hematopoietic cell lineage	1	1.00E+00	0.5
KEGG_PATHWAY	RNA degradation	1	1.00E+00	0.8
KEGG_PATHWAY	Systemic lupus erythematosus	1	1.00E+00	0.5
KEGG_PATHWAY	Cytokine-cytokine receptor interaction	1	1.00E+00	0.2
KEGG_PATHWAY	Arachidonic acid metabolism	1	1.00E+00	0.8

Nature Genetics doi:10/1038/ng.2391

KEGG_PATHWAY	Tyrosine metabolism	1	1.00E+00	1
KEGG_PATHWAY	Natural killer cell mediated cytotoxicity	1	1.00E+00	0.3
KEGG_PATHWAY	RIG-I-like receptor signaling pathway	1	1.00E+00	0.6
KEGG_PATHWAY	Progesterone-mediated oocyte maturation	1	1.00E+00	0.5
KEGG_PATHWAY	Ascorbate and aldarate metabolism	1	1.00E+00	2.7
KEGG_PATHWAY	NOD-like receptor signaling pathway	1	1.00E+00	0.7
KEGG_PATHWAY	Tryptophan metabolism	1	1.00E+00	1.1
KEGG_PATHWAY	Glycerophospholipid metabolism	1	1.00E+00	0.7
KEGG_PATHWAY	Metabolism of xenobiotics by cytochrome P450	1	1.00E+00	0.8
KEGG_PATHWAY	Type II diabetes mellitus	1	1.00E+00	1
KEGG_PATHWAY	Complement and coagulation cascades	1	1.00E+00	0.7
KEGG_PATHWAY	Taste transduction	1	1.00E+00	0.9
KEGG_PATHWAY	VEGF signaling pathway	1	1.00E+00	0.6
KEGG_PATHWAY	Bladder cancer	1	1.00E+00	1.1
KEGG_PATHWAY	O-Mannosyl glycan biosynthesis	1	1.00E+00	15.1
KEGG_PATHWAY	Acute myeloid leukemia	1	1.00E+00	0.8
KEGG_PATHWAY	Nucleotide excision repair	1	1.00E+00	1
KEGG_PATHWAY	Basal transcription factors	1	1.00E+00	1.3
KEGG_PATHWAY	Viral myocarditis	1	1.00E+00	0.6
KEGG_PATHWAY	Dorso-ventral axis formation	1	1.00E+00	1.8
KEGG_PATHWAY	Melanoma	1	1.00E+00	0.6
KEGG_PATHWAY	Aldosterone-regulated sodium reabsorption	1	1.00E+00	1.1
KEGG_PATHWAY	Prion diseases	1	1.00E+00	1.3
KEGG_PATHWAY	Primary immunodeficiency	1	1.00E+00	1.3
KEGG_PATHWAY	Propanoate metabolism	1	1.00E+00	1.4
KEGG_PATHWAY	alpha-Linolenic acid metabolism	1	1.00E+00	2.5
KEGG_PATHWAY	Spliceosome	1	1.00E+00	0.4
KEGG_PATHWAY	Oocyte meiosis	1	1.00E+00	0.4

Nature Genetics doi:10/1038/ng.2391

KEGG_PATHWAY	Linoleic acid metabolism	1	1.00E+00	1.6
KEGG_PATHWAY	Lysine degradation	1	1.00E+00	1
KEGG_PATHWAY	D-Glutamine and D-glutamate metabolism	1	1.00E+00	11.4
KEGG_PATHWAY	Glycosphingolipid biosynthesis	1	1.00E+00	1.8
KEGG_PATHWAY	beta-Alanine metabolism	1	1.00E+00	2.1
KEGG_PATHWAY	Nitrogen metabolism	1	1.00E+00	2
KEGG_PATHWAY	Pyruvate metabolism	1	1.00E+00	1.1
KEGG_PATHWAY	Glycosphingolipid biosynthesis	1	1.00E+00	3.2

Nature Genetics doi:10/1038/ng.2391

Supplementary Table 9. The signaling pathways and biological processes separated by HCC patient.

Symbol	Gene ID	Pathways and biological processes	Database	Number of mutations (Patient ID)
BMP8B	656	Hedgehog signaling pathway	KEGG_PATHWAY	1 (P51)
CSNK1G3	1456	Hedgehog signaling pathway	KEGG_PATHWAY	1 (P50)
LRP2	4036	Hedgehog signaling pathway	KEGG_PATHWAY	1 (P55)
WNT2	7472	Hedgehog signaling pathway	KEGG_PATHWAY	1 (P55)
FZD1	8321	Wnt signaling pathway	KEGG_PATHWAY	1 (P48)
MAPK8	5599	Wnt signaling pathway	KEGG_PATHWAY	1 (P48)
PPP2R1B	5519	Wnt signaling pathway	KEGG_PATHWAY	1 (P929)
TP53	7157	Wnt signaling pathway	KEGG_PATHWAY	1 (P56)
VANGL1	81839	Wnt signaling pathway	KEGG_PATHWAY	1 (P54)
WNT2	7472	Wnt signaling pathway	KEGG_PATHWAY	1 (P55)
MAPK8	5599	ErbB signaling pathway	KEGG_PATHWAY	1 (P48)
PAK6	56924	ErbB signaling pathway	KEGG_PATHWAY	1 (P56)
SOS1	6654	ErbB signaling pathway	KEGG_PATHWAY	1 (P48)
APAF1	317	Apoptosis	KEGG_PATHWAY	1 (P56)
PRKAR1B	5575	Apoptosis	KEGG_PATHWAY	1 (P48)
TP53	7157	Apoptosis	KEGG_PATHWAY	1 (P56)
CACNA2D1	781	MAPK signaling pathway	KEGG_PATHWAY	1 (P51)
MAPK8	5599	MAPK signaling pathway	KEGG_PATHWAY	1 (P48)
PLA2G6	8398	MAPK signaling pathway	KEGG_PATHWAY	1 (P50)
SOS1	6654	MAPK signaling pathway	KEGG_PATHWAY	1 (P48)
TP53	7157	MAPK signaling pathway	KEGG_PATHWAY	1 (P56)
BMP8B	656	TGF-beta signaling pathway	KEGG_PATHWAY	1 (P51)
PPP2R1B	5519	TGF-beta signaling pathway	KEGG_PATHWAY	1 (P929)
CCNH	902	Cell cycle	KEGG_PATHWAY	1 (P48)

Nature Genetics doi:10/1038/ng.2391

TP53	7157	Cell cycle	KEGG_PATHWAY	1 (P56)
PLA2G6	8398	VEGF signaling pathway	KEGG_PATHWAY	1 (P50)
PSENEN	55851	Notch signaling pathway	KEGG_PATHWAY	1 (P56)
LIFR	3977	Jak-STAT signaling pathway	KEGG_PATHWAY	1 (P55)
SOS1	6654	Jak-STAT signaling pathway	KEGG_PATHWAY	1 (P48)
ITGA6	3655	Cell adhesion molecules (CAMs)	KEGG_PATHWAY	1 (P51)
NRXN2	9379	Cell adhesion molecules (CAMs)	KEGG_PATHWAY	1 (P52)
NFASC	23114	Cell adhesion molecules (CAMs)	KEGG_PATHWAY	1 (P52)
NRCAM	4897	Cell adhesion molecules (CAMs)	KEGG_PATHWAY	1 (P56)
PVRL1	5818	Cell adhesion molecules (CAMs)	KEGG_PATHWAY	1 (P47)
SIGLEC1	6614	Cell adhesion molecules (CAMs)	KEGG_PATHWAY	1 (P56)
VCAM1	7412	Cell adhesion molecules (CAMs)	KEGG_PATHWAY	1 (P929)
CTCF	10664	chromatin binding	GOTERM_MF_FAT	1 (P929)
ATRX	546	chromatin binding	GOTERM_MF_FAT	1 (P50)
CHAF1B	8208	chromatin binding	GOTERM_MF_FAT	1 (P48)
MPO	4353	chromatin binding	GOTERM_MF_FAT	1 (P55)
PBRM1	55193	chromatin binding	GOTERM_MF_FAT	1 (P48)
RUNX2	860	chromatin binding	GOTERM_MF_FAT	1 (P52)
TP53	7157	chromatin binding	GOTERM_MF_FAT	1 (P56)
AGAP6	414189	small GTPase regulator activity	GOTERM_MF_FAT	1 (P50)
FGD6	55785	small GTPase regulator activity	GOTERM_MF_FAT	1 (P51)
IQGAP1	8826	small GTPase regulator activity	GOTERM_MF_FAT	1 (P56)
ARHGEF5	7984	small GTPase regulator activity	GOTERM_MF_FAT	1 (P50)
CIT	11113	small GTPase regulator activity	GOTERM_MF_FAT	1 (P55)
DOCK2	1794	small GTPase regulator activity	GOTERM_MF_FAT	1 (P56)
PLEKHG2	64857	small GTPase regulator activity	GOTERM_MF_FAT	1 (P48)
SOS1	6654	small GTPase regulator activity	GOTERM_MF_FAT	1 (P48)
TTN	7273	small GTPase regulator activity	GOTERM_MF_FAT	1 (P48)

Nature Genetics doi:10/1038/ng.2391

ARID1A	8289	transcription regulator activity	GOTERM_MF_FAT	2 (P56)
CTCF	10664	transcription regulator activity	GOTERM_MF_FAT	1 (P929)
ELK3	2004	transcription regulator activity	GOTERM_MF_FAT	1 (P929)
GLIS1	148979	transcription regulator activity	GOTERM_MF_FAT	1 (P50)
KIAA2018	205717	transcription regulator activity	GOTERM_MF_FAT	1 (P48)
LHX9	56956	transcription regulator activity	GOTERM_MF_FAT	1 (P55)
POU1F1	5449	transcription regulator activity	GOTERM_MF_FAT	1 (P50)
POU3F3	5455	transcription regulator activity	GOTERM_MF_FAT	1 (P48)
TBX18	9096	transcription regulator activity	GOTERM_MF_FAT	1 (P50)
TBX19	9095	transcription regulator activity	GOTERM_MF_FAT	1 (P48)
TAF1L	138474	transcription regulator activity	GOTERM_MF_FAT	1 (P48)
BRWD1	54014	transcription regulator activity	GOTERM_MF_FAT	1 (P47)
DCC	1630	transcription regulator activity	GOTERM_MF_FAT	1 (P47)
ELL	8178	transcription regulator activity	GOTERM_MF_FAT	1 (P48)
GTF3C1	2975	transcription regulator activity	GOTERM_MF_FAT	1 (P54)
GCM1	8521	transcription regulator activity	GOTERM_MF_FAT	1 (P48)
HOXA1	3198	transcription regulator activity	GOTERM_MF_FAT	1 (P52)
KEAP1	9817	transcription regulator activity	GOTERM_MF_FAT	1 (P51)
MTF1	4520	transcription regulator activity	GOTERM_MF_FAT	1 (P56)
MYT1	4661	transcription regulator activity	GOTERM_MF_FAT	1 (P56)
NR2E1	7101	transcription regulator activity	GOTERM_MF_FAT	1 (P48)
PPARGC1A	10891	transcription regulator activity	GOTERM_MF_FAT	1 (P53)
PLAGL2	5326	transcription regulator activity	GOTERM_MF_FAT	1 (P48)
PFDN5	5204	transcription regulator activity	GOTERM_MF_FAT	1 (P53)
RUNX2	860	transcription regulator activity	GOTERM_MF_FAT	1 (P52)
TSHZ1	10194	transcription regulator activity	GOTERM_MF_FAT	1 (P48)
TRIM32	22954	transcription regulator activity	GOTERM_MF_FAT	1 (P51)
TP53	7157	transcription regulator activity	GOTERM_MF_FAT	1 (P56)

Nature Genetics doi:10/1038/ng.2391

ZBTB7B	51043	transcription regulator activity	GOTERM_MF_FAT	1 (P56)
ZSCAN16	80345	transcription regulator activity	GOTERM_MF_FAT	1 (P50)
ZFP36L1	677	transcription regulator activity	GOTERM_MF_FAT	1 (P929)
ZNF394	84124	transcription regulator activity	GOTERM_MF_FAT	1 (P50)
ARID1A	8289	DNA binding	GOTERM_MF_FAT	2 (P56)
ARID2	196528	DNA binding	GOTERM_MF_FAT	2 (P48;P51)
AKNA	80709	DNA binding	GOTERM_MF_FAT	1 (P48)
BRIP1	83990	DNA binding	GOTERM_MF_FAT	1 (P56)
CTCF	10664	DNA binding	GOTERM_MF_FAT	1 (P929)
ELK3	2004	DNA binding	GOTERM_MF_FAT	1 (P929)
GLIS1	148979	DNA binding	GOTERM_MF_FAT	1 (P50)
HMGXB4	10042	DNA binding	GOTERM_MF_FAT	1 (P55)
KIAA2018	205717	DNA binding	GOTERM_MF_FAT	1 (P48)
LHX9	56956	DNA binding	GOTERM_MF_FAT	1 (P55)
POU1F1	5449	DNA binding	GOTERM_MF_FAT	1 (P50)
POU3F3	5455	DNA binding	GOTERM_MF_FAT	1 (P48)
RAD50	10111	DNA binding	GOTERM_MF_FAT	1 (P48)
SMARCAL1	50485	DNA binding	GOTERM_MF_FAT	1 (P929)
TBX18	9096	DNA binding	GOTERM_MF_FAT	1 (P50)
TBX19	9095	DNA binding	GOTERM_MF_FAT	1 (P48)
TAF1L	138474	DNA binding	GOTERM_MF_FAT	1 (P48)
ATRX	546	DNA binding	GOTERM_MF_FAT	1 (P50)
GTF3C1	2975	DNA binding	GOTERM_MF_FAT	1 (P54)
GCM1	8521	DNA binding	GOTERM_MF_FAT	1 (P48)
HOXA1	3198	DNA binding	GOTERM_MF_FAT	1 (P52)
MTF1	4520	DNA binding	GOTERM_MF_FAT	1 (P56)
MYT1	4661	DNA binding	GOTERM_MF_FAT	1 (P56)
MLL3	58508	DNA binding	GOTERM_MF_FAT	1 (P56)

Nature Genetics doi:10/1038/ng.2391

NR2E1	7101	DNA binding	GOTERM_MF_FAT	1 (P48)
PPARGC1A	10891	DNA binding	GOTERM_MF_FAT	1 (P53)
PLAGL2	5326	DNA binding	GOTERM_MF_FAT	1 (P48)
POGZ	23126	DNA binding	GOTERM_MF_FAT	1 (P56)
PBRM1	55193	DNA binding	GOTERM_MF_FAT	1 (P48)
RAG1	5896	DNA binding	GOTERM_MF_FAT	1 (P48)
RUNX2	860	DNA binding	GOTERM_MF_FAT	1 (P52)
SOS1	6654	DNA binding	GOTERM_MF_FAT	1 (P48)
TSHZ1	10194	DNA binding	GOTERM_MF_FAT	1 (P48)
TSN	7247	DNA binding	GOTERM_MF_FAT	1 (P51)
TP53	7157	DNA binding	GOTERM_MF_FAT	1 (P56)
ZBTB7B	51043	DNA binding	GOTERM_MF_FAT	1 (P56)
ZSCAN16	80345	DNA binding	GOTERM_MF_FAT	1 (P50)
ZNF28	7576	DNA binding	GOTERM_MF_FAT	1 (P53)
ZNF280B	140883	DNA binding	GOTERM_MF_FAT	1 (P51)
ZNF32	7580	DNA binding	GOTERM_MF_FAT	1 (P48)
ZNF324	25799	DNA binding	GOTERM_MF_FAT	1 (P47)
ZFP36L1	677	DNA binding	GOTERM_MF_FAT	1 (P929)
ZNF394	84124	DNA binding	GOTERM_MF_FAT	1 (P50)
ZNF671	79891	DNA binding	GOTERM_MF_FAT	1 (P53)
ZNF84	7637	DNA binding	GOTERM_MF_FAT	1 (P48)
ZMAT2	153527	DNA binding	GOTERM_MF_FAT	1 (P56)
ACSL1	2180	Fatty acid metabolism	KEGG_PATHWAY	1 (P50)
ADH4	127	Fatty acid metabolism	KEGG_PATHWAY	1 (P54)
ALDH2	217	Fatty acid metabolism	KEGG_PATHWAY	1 (P929)
CPT1A	1374	Fatty acid metabolism	KEGG_PATHWAY	1 (P54)
ACSL1	2180	Adipocytokine signaling pathway	KEGG_PATHWAY	1 (P50)
CAMKK2	10645	Adipocytokine signaling pathway	KEGG_PATHWAY	1 (P51)

Nature Genetics doi:10/1038/ng.2391

CPT1A	1374	Adipocytokine signaling pathway	KEGG_PATHWAY	1 (P54)
MAPK8	5599	Adipocytokine signaling pathway	KEGG_PATHWAY	1 (P48)
PPARGC1A	10891	Adipocytokine signaling pathway	KEGG_PATHWAY	1 (P53)
PRKCQ	5588	Adipocytokine signaling pathway	KEGG_PATHWAY	1 (P55)
GUCA1C	9626	Olfactory transduction	KEGG_PATHWAY	1 (P51)
OR1S1	219959	Olfactory transduction	KEGG_PATHWAY	1 (P51)
OR10H5	284433	Olfactory transduction	KEGG_PATHWAY	1 (P48)
OR11H6	122748	Olfactory transduction	KEGG_PATHWAY	1 (P51)
OR2C1	4993	Olfactory transduction	KEGG_PATHWAY	1 (P56)
OR2G3	81469	Olfactory transduction	KEGG_PATHWAY	1 (P50)
OR2M5	127059	Olfactory transduction	KEGG_PATHWAY	1 (P929)
OR2T4	127074	Olfactory transduction	KEGG_PATHWAY	1 (P53)
OR5A2	219981	Olfactory transduction	KEGG_PATHWAY	1 (P50)
OR5B12	390191	Olfactory transduction	KEGG_PATHWAY	1 (P53)
OR5D13	390142	Olfactory transduction	KEGG_PATHWAY	1 (P55)
OR5F1	338674	Olfactory transduction	KEGG_PATHWAY	1 (P48)
OR52N2	390077	Olfactory transduction	KEGG_PATHWAY	1 (P56)
OR9G4	283189	Olfactory transduction	KEGG_PATHWAY	1 (P53)
OR9Q1	219956	Olfactory transduction	KEGG_PATHWAY	1 (P55)
AKAP3	10566	cell motion	GOTERM_BP_FAT	1 (P53)
CD97	976	cell motion	GOTERM_BP_FAT	1 (P53)
POU3F3	5455	cell motion	GOTERM_BP_FAT	1 (P48)
ANK3	288	cell motion	GOTERM_BP_FAT	1 (P53)
ASTN1	460	cell motion	GOTERM_BP_FAT	1 (P48)
CATSPER1	117144	cell motion	GOTERM_BP_FAT	1 (P48)
CCKAR	886	cell motion	GOTERM_BP_FAT	1 (P50)
DOCK2	1794	cell motion	GOTERM_BP_FAT	1 (P56)
DCC	1630	cell motion	GOTERM_BP_FAT	1 (P47)

Nature Genetics doi:10/1038/ng.2391

EDN3	1908	cell motion	GOTERM_BP_FAT	1 (P51)
ELMO1	9844	cell motion	GOTERM_BP_FAT	2 (P48;P50)
GDNF	2668	cell motion	GOTERM_BP_FAT	1 (P929)
HOXA1	3198	cell motion	GOTERM_BP_FAT	1 (P52)
ITGA6	3655	cell motion	GOTERM_BP_FAT	1 (P51)
MAPK8	5599	cell motion	GOTERM_BP_FAT	1 (P48)
NFASC	23114	cell motion	GOTERM_BP_FAT	1 (P52)
NRCAM	4897	cell motion	GOTERM_BP_FAT	1 (P56)
NR2E1	7101	cell motion	GOTERM_BP_FAT	1 (P48)
PVRL1	5818	cell motion	GOTERM_BP_FAT	1 (P47)
SEMA3A	10371	cell motion	GOTERM_BP_FAT	1 (P53)
SPOCK1	6695	cell motion	GOTERM_BP_FAT	1 (P56)
VCAM1	7412	cell motion	GOTERM_BP_FAT	1 (P929)
WNT2	7472	cell motion	GOTERM_BP_FAT	1 (P55)
CD72	971	biological adhesion	GOTERM_BP_FAT	1 (P48)
CD97	976	biological adhesion	GOTERM_BP_FAT	1 (P53)
ASTN1	460	biological adhesion	GOTERM_BP_FAT	1 (P48)
CLSTN1	22883	biological adhesion	GOTERM_BP_FAT	1 (P48)
CLSTN2	64084	biological adhesion	GOTERM_BP_FAT	1 (P48)
CTNNA3	29119	biological adhesion	GOTERM_BP_FAT	1 (P48)
COL5A3	50509	biological adhesion	GOTERM_BP_FAT	2 (P48;P56)
DDR1	780	biological adhesion	GOTERM_BP_FAT	1 (P48)
DLG1	1739	biological adhesion	GOTERM_BP_FAT	1 (P50)
ITGA6	3655	biological adhesion	GOTERM_BP_FAT	1 (P51)
NEDD9	4739	biological adhesion	GOTERM_BP_FAT	1 (P52)
NRXN2	9379	biological adhesion	GOTERM_BP_FAT	1 (P52)
NFASC	23114	biological adhesion	GOTERM_BP_FAT	1 (P52)
NRCAM	4897	biological adhesion	GOTERM_BP_FAT	1 (P56)

Nature Genetics doi:10/1038/ng.2391

PVRL1	5818	biological adhesion	GOTERM_BP_FAT	1 (P47)
PTPRS	5802	biological adhesion	GOTERM_BP_FAT	1 (P48)
PCDH21	92211	biological adhesion	GOTERM_BP_FAT	1 (P929)
CDHR2	54825	biological adhesion	GOTERM_BP_FAT	1 (P51)
PCDHB1	29930	biological adhesion	GOTERM_BP_FAT	1 (P48)
PCDHB6	56130	biological adhesion	GOTERM_BP_FAT	1 (P48)
SIGLEC1	6614	biological adhesion	GOTERM_BP_FAT	1 (P56)
SIGLEC12	89858	biological adhesion	GOTERM_BP_FAT	1 (P929)
SIGLEC8	27181	biological adhesion	GOTERM_BP_FAT	1 (P55)
SPOCK1	6695	biological adhesion	GOTERM_BP_FAT	1 (P56)
STAB1	23166	biological adhesion	GOTERM_BP_FAT	1 (P50)
VCAM1	7412	biological adhesion	GOTERM_BP_FAT	1 (P929)
CD72	971	cell adhesion	GOTERM_BP_FAT	1 (P48)
CD97	976	cell adhesion	GOTERM_BP_FAT	1 (P53)
ASTN1	460	cell adhesion	GOTERM_BP_FAT	1 (P48)
CLSTN1	22883	cell adhesion	GOTERM_BP_FAT	1 (P48)
CLSTN2	64084	cell adhesion	GOTERM_BP_FAT	1 (P48)
CTNNA3	29119	cell adhesion	GOTERM_BP_FAT	1 (P48)
COL5A3	50509	cell adhesion	GOTERM_BP_FAT	2 (P48;P56)
DDR1	780	cell adhesion	GOTERM_BP_FAT	1 (P48)
DLG1	1739	cell adhesion	GOTERM_BP_FAT	1 (P50)
ITGA6	3655	cell adhesion	GOTERM_BP_FAT	1 (P51)
NEDD9	4739	cell adhesion	GOTERM_BP_FAT	1 (P52)
NRXN2	9379	cell adhesion	GOTERM_BP_FAT	1 (P52)
NFASC	23114	cell adhesion	GOTERM_BP_FAT	1 (P52)
NRCAM	4897	cell adhesion	GOTERM_BP_FAT	1 (P56)
PVRL1	5818	cell adhesion	GOTERM_BP_FAT	1 (P47)
PTPRS	5802	cell adhesion	GOTERM_BP_FAT	1 (P48)

Nature Genetics doi:10/1038/ng.2391

PCDH21	92211	cell adhesion	GOTERM_BP_FAT	1 (P929)
CDHR2	54825	cell adhesion	GOTERM_BP_FAT	1 (P51)
PCDHB1	29930	cell adhesion	GOTERM_BP_FAT	1 (P48)
PCDHB6	56130	cell adhesion	GOTERM_BP_FAT	1 (P48)
SIGLEC1	6614	cell adhesion	GOTERM_BP_FAT	1 (P56)
SIGLEC12	89858	cell adhesion	GOTERM_BP_FAT	1 (P929)
SIGLEC8	27181	cell adhesion	GOTERM_BP_FAT	1 (P55)
SPOCK1	6695	cell adhesion	GOTERM_BP_FAT	1 (P56)
STAB1	23166	cell adhesion	GOTERM_BP_FAT	1 (P50)
VCAM1	7412	cell adhesion	GOTERM_BP_FAT	1 (P929)
ANK3	288	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P53)
ATL1	51062	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P47)
CCKAR	886	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P50)
DCC	1630	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P47)
HOXA1	3198	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P52)
MAP1B	4131	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P48)
NTNG1	22854	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P48)
NFASC	23114	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P52)
NRCAM	4897	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P56)
NR2E1	7101	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P48)
PVRL1	5818	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P47)
PTPRZ1	5803	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P48)
SEMA3A	10371	cell morphogenesis involved in differentiation	GOTERM_BP_FAT	1 (P53)
ANK3	288	cell morphogenesis	GOTERM_BP_FAT	1 (P53)
ATL1	51062	cell morphogenesis	GOTERM_BP_FAT	1 (P47)
CCKAR	886	cell morphogenesis	GOTERM_BP_FAT	1 (P50)
DOCK2	1794	cell morphogenesis	GOTERM_BP_FAT	1 (P56)
DCC	1630	cell morphogenesis	GOTERM_BP_FAT	1 (P47)

Nature Genetics doi:10/1038/ng.2391

DLG1	1739	cell morphogenesis	GOTERM_BP_FAT	1 (P50)
HOXA1	3198	cell morphogenesis	GOTERM_BP_FAT	1 (P52)
LIFR	3977	cell morphogenesis	GOTERM_BP_FAT	1 (P55)
MAP1B	4131	cell morphogenesis	GOTERM_BP_FAT	1 (P48)
NTNG1	22854	cell morphogenesis	GOTERM_BP_FAT	1 (P48)
NFASC	23114	cell morphogenesis	GOTERM_BP_FAT	1 (P52)
NRCAM	4897	cell morphogenesis	GOTERM_BP_FAT	1 (P56)
NR2E1	7101	cell morphogenesis	GOTERM_BP_FAT	1 (P48)
PVRL1	5818	cell morphogenesis	GOTERM_BP_FAT	1 (P47)
PTPRZ1	5803	cell morphogenesis	GOTERM_BP_FAT	1 (P48)
SEMA3A	10371	cell morphogenesis	GOTERM_BP_FAT	1 (P53)
TNIP1	10318	glycoprotein metabolic process	GOTERM_BP_FAT	1 (P50)
GALNT7	51809	glycoprotein metabolic process	GOTERM_BP_FAT	1 (P48)
COG7	91949	glycoprotein metabolic process	GOTERM_BP_FAT	1 (P48)
DSE	29940	glycoprotein metabolic process	GOTERM_BP_FAT	2 (P48;P50)
FUT10	84750	glycoprotein metabolic process	GOTERM_BP_FAT	1 (P50)
FUT9	10690	glycoprotein metabolic process	GOTERM_BP_FAT	1 (P48)
GYPC	2995	glycoprotein metabolic process	GOTERM_BP_FAT	1 (P56)
IDE	3416	glycoprotein metabolic process	GOTERM_BP_FAT	1 (P52)
LRP2	4036	glycoprotein metabolic process	GOTERM_BP_FAT	1 (P55)
PSENEN	55851	glycoprotein metabolic process	GOTERM_BP_FAT	1 (P56)
POMT2	29954	glycoprotein metabolic process	GOTERM_BP_FAT	1 (P48)
ASTN1	460	cell-cell adhesion	GOTERM_BP_FAT	1 (P48)
CLSTN1	22883	cell-cell adhesion	GOTERM_BP_FAT	1 (P48)
CLSTN2	64084	cell-cell adhesion	GOTERM_BP_FAT	1 (P48)
CTNNA3	29119	cell-cell adhesion	GOTERM_BP_FAT	1 (P48)
DLG1	1739	cell-cell adhesion	GOTERM_BP_FAT	1 (P50)
NRCAM	4897	cell-cell adhesion	GOTERM_BP_FAT	1 (P56)

Nature Genetics doi:10/1038/ng.2391

PVRL1	5818	cell-cell adhesion	GOTERM_BP_FAT	1 (P47)
PCDH21	92211	cell-cell adhesion	GOTERM_BP_FAT	1 (P929)
CDHR2	54825	cell-cell adhesion	GOTERM_BP_FAT	1 (P51)
PCDHB1	29930	cell-cell adhesion	GOTERM_BP_FAT	1 (P48)
PCDHB6	56130	cell-cell adhesion	GOTERM_BP_FAT	1 (P48)
SIGLEC1	6614	cell-cell adhesion	GOTERM_BP_FAT	1 (P56)
VCAM1	7412	cell-cell adhesion	GOTERM_BP_FAT	1 (P929)
POU3F3	5455	cell motility	GOTERM_BP_FAT	1 (P48)
ASTN1	460	cell motility	GOTERM_BP_FAT	1 (P48)
CATSPER1	117144	cell motility	GOTERM_BP_FAT	1 (P48)
CCKAR	886	cell motility	GOTERM_BP_FAT	1 (P50)
DOCK2	1794	cell motility	GOTERM_BP_FAT	1 (P56)
DCC	1630	cell motility	GOTERM_BP_FAT	1 (P47)
EDN3	1908	cell motility	GOTERM_BP_FAT	1 (P51)
GDNF	2668	cell motility	GOTERM_BP_FAT	1 (P929)
ITGA6	3655	cell motility	GOTERM_BP_FAT	1 (P51)
NRCAM	4897	cell motility	GOTERM_BP_FAT	1 (P56)
NR2E1	7101	cell motility	GOTERM_BP_FAT	1 (P48)
VCAM1	7412	cell motility	GOTERM_BP_FAT	1 (P929)
WNT2	7472	cell motility	GOTERM_BP_FAT	1 (P55)
POU3F3	5455	cell migration	GOTERM_BP_FAT	1 (P48)
ASTN1	460	cell migration	GOTERM_BP_FAT	1 (P48)
CCKAR	886	cell migration	GOTERM_BP_FAT	1 (P50)
DOCK2	1794	cell migration	GOTERM_BP_FAT	1 (P56)
DCC	1630	cell migration	GOTERM_BP_FAT	1 (P47)
EDN3	1908	cell migration	GOTERM_BP_FAT	1 (P51)
GDNF	2668	cell migration	GOTERM_BP_FAT	1 (P929)
ITGA6	3655	cell migration	GOTERM_BP_FAT	1 (P51)

Nature Genetics doi:10/1038/ng.2391

NRCAM	4897	cell migration	GOTERM_BP_FAT	1 (P56)
NR2E1	7101	cell migration	GOTERM_BP_FAT	1 (P48)
VCAM1	7412	cell migration	GOTERM_BP_FAT	1 (P929)
WNT2	7472	cell migration	GOTERM_BP_FAT	1 (P55)
HTRA3	94031	regulation of cell growth	GOTERM_BP_FAT	1 (P53)
DDR1	780	regulation of cell growth	GOTERM_BP_FAT	1 (P48)
MAP1B	4131	regulation of cell growth	GOTERM_BP_FAT	1 (P48)
NPR1	4881	regulation of cell growth	GOTERM_BP_FAT	1 (P929)
NRCAM	4897	regulation of cell growth	GOTERM_BP_FAT	1 (P56)
PRKCQ	5588	regulation of cell growth	GOTERM_BP_FAT	1 (P55)
CDHR2	54825	regulation of cell growth	GOTERM_BP_FAT	1 (P51)
SEMA3A	10371	regulation of cell growth	GOTERM_BP_FAT	1 (P53)
TP53	7157	regulation of cell growth	GOTERM_BP_FAT	1 (P56)
IGSF10	285313	regeneration	GOTERM_BP_FAT	1 (P56)
LIFR	3977	regeneration	GOTERM_BP_FAT	1 (P55)
LRP2	4036	regeneration	GOTERM_BP_FAT	1 (P55)
MAP1B	4131	regeneration	GOTERM_BP_FAT	1 (P48)
PRKCQ	5588	regeneration	GOTERM_BP_FAT	1 (P55)
ARID1A	8289	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	2 (P56)
CTCF	10664	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P929)
GLIS1	148979	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P50)
POU1F1	5449	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P50)
POU3F3	5455	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P48)

Nature Genetics doi:10/1038/ng.2391

TBX19	9095	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P48)
CCNH	902	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P48)
DRD3	1814	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P48)
GDNF	2668	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P929)
HOXA1	3198	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P52)
MTF1	4520	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P56)
PPARGC1A	10891	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P53)
PLAGL2	5326	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P48)
RUNX2	860	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P52)
TP53	7157	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P56)
ZBTB7B	51043	positive regulation of transcription, DNA-dependent	GOTERM_BP_FAT	1 (P56)
POU1F1	5449	cell fate commitment	GOTERM_BP_FAT	1 (P50)
TBX19	9095	cell fate commitment	GOTERM_BP_FAT	1 (P48)
GDNF	2668	cell fate commitment	GOTERM_BP_FAT	1 (P929)
GCM1	8521	cell fate commitment	GOTERM_BP_FAT	1 (P48)
NR2E1	7101	cell fate commitment	GOTERM_BP_FAT	1 (P48)
RUNX2	860	cell fate commitment	GOTERM_BP_FAT	1 (P52)

Nature Genetics doi:10/1038/ng.2391

TP53	7157	cell fate commitment	GOTERM_BP_FAT	1 (P56)

Nature Genetics doi:10/1038/ng.2391

Supplementary Table 10. The ARID1A mutations were found in the HCCs with different etiologies and cell lines.

ARID1A mutations profile

Etiologies	Total examined cases	Cases with mutations	*P*-value[a]	Non-silent mutations	Synonymous mutations	dN/dS	*P*-value[b]
HBV	110	14	—	22	1	22:1	4.88×10^{-3}
HCV	26	2	0.74	2	0	2:0	0.60
HBV plus HCV	1	1	0.14	1	0	1:0	0.75
Nonviral	10	1	1.00	1	0	1:0	0.75
Total	147	18	—	26	1	26:1	5.07×10^{-3}

[a]The statistical significance of ARID1A mutations in HCC cases with different etiologies was calculated with the chi-squared test, as compared to ARID1A mutations occurred in HBV-infected HCCs. [b]Statistical-analyses were performed with a Fisher's exact test to distinguish between the observed results and a null model (dN/dS = 2).

Sample Names	Tissue and cell types	mRNA positions	Nucleotides (cDNA)	Amino acids (protein)	Mutation types
HCC-LM6	Cell line	433	c.60_62delGCC	p.del21Pro	indel
PLC/PRF/5	Cell line	501	c.128_129dupGGC	p.ins46Ala	indel
HCC-171#	HCV-infected HCC	564	c.191_204delGGCCGCCGCAGCCGCT	fs	indel
HCC-157#	Non-viral-infected HCC	691	c.317_409delCGCGGGCCCTAGGCCCGCCCTG AACAATAACCTCACGGAGCCGCCCGGCGGCGGCGG TGGCGGCAGCAGCGATGGGGTGGGGGCGCCTCCTC	fs	Indel

Nature Genetics doi:10/1038/ng.2391

HCC-165#	HBV plus HCV-infected HCC	1059	c.686A>G	p.Tyr229Cys	missense
*HCC-55#	HBV-infected HCC	1069	c.696C>G	p.Ser232Arg	missense
Hep3B	Cell line	1206	c.833G>C	p.Gly278Ala	missense
YY-8103	Cell line	1206	c.833G>C	p.Gly278Ala	missense
HCC-61#	HBV-infected HCC	1710	c.1337_1338delCT	fs	Indel
*HCC-72#	HBV-infected HCC	2132	c.1759T>A	p.Ser587Thr	missense
HCC-7#	HBV-infected HCC	2796	c.2423G>A	p.Gly808Asp	missense
HCC-40#	HBV-infected HCC	3149	c.2776A>G	p.Thr926Ala	missense
HCC-63#	HBV-infected HCC	3149	c.2776A>G	p.Thr926Ala	missense
HCC-34#	HBV-infected HCC	3477	c.3104C>A	p.Ala1035Asp	missense
HCC-61#	HBV-infected HCC	3512	c.3139delA	fs	Indel
HCC-140#	HCV-infected HCC	3593	c.3220C>T	p.Arg1074Trp	missense
*HCC-56	HBV-infected HCC	4006	c.3633T>A	p.Tyr1211*	nonsense
*HCC-56	HBV-infected HCC	4008	c.3635A>T	p.Gln1212Leu	missense
HCC-81#	HBV-infected HCC	4649	c.4276G>A	p.Asp1426Asn	missense
HCC-18#	HBV-infected HCC	4754	c.4381C>T	p.Arg1461*	nonsense
HCC-82#	HBV-infected HCC	5771	c.5398A>G	p.Asn1800Asp	missense
HCC-82#	HBV-infected HCC	5779	c.5406_5413delGGAGAAGCTGA	fs	Indel
HCC-82#	HBV-infected HCC	5796	c.5423A>C	p.Lys1808Thr	missense
HCC-81#	HBV-infected HCC	6346	c.5973_5974delGT	fs	Indel
HCC-LM6	Cell line	6503	c.6130A>G	p.Asn2044Asp	missense
HCC-47#	HBV-infected HCC	6543	c.6170G>A	p.Arg2057Gln	missense
HCC-47#	HBV-infected HCC	6544	c.6171G>A	p.Arg2057Arg	synonymous
HCC-47#	HBV-infected HCC	6557	c.6184G>A	p.Val2062Ile	missense
HCC-LM6	Cell line	6613	c.6240T>G	p.Ile2080Met	missense
HCC-31#	HBV-infected HCC	7084	c.6711_6712delGC	fs	Indel
*HCC-68#	HBV-infected HCC	7174	c.6801G>A	p.Met2267Ile	missense

Nature Genetics doi:10/1038/ng.2391

HCC-LM6	Cell line	463	c.90G>A	P.(=)	synonymous
HCC-LM6	Cell line	6244	c.5871C>T	P.(=)	synonymous
HCC-LM6	Cell line	6253	c.5880C>T	P.(=)	synonymous
HCC-LM6	Cell line	6301	c.5928T>C	P.(=)	synonymous
HCC-LM6	Cell line	6307	c.5934C>T	P.(=)	synonymous
HCC-LM6	Cell line	6316	c.5943C>T	P.(=)	synonymous
HCC-LM6	Cell line	6325	c.5952G>C	P.(=)	synonymous
HCC-LM6	Cell line	6337	c.5964T>C	P.(=)	synonymous
HCC-LM6	Cell line	6340	c.5967A>G	P.(=)	synonymous
HCC-LM6	Cell line	6349	c.5976A>G	P.(=)	synonymous
HCC-LM6	Cell line	6364	c.5991T>C	P.(=)	synonymous
HCC-LM6	Cell line	6400	c.6027C>T	P.(=)	synonymous
HCC-LM6	Cell line	6442	c.6069A>G	P.(=)	synonymous
HCC-LM6	Cell line	6469	c.6096A>G	P.(=)	synonymous
HCC-LM6	Cell line	6502	c.6129C>T	P.(=)	synonymous
HCC-LM6	Cell line	6544	c.6171G>A	P.(=)	synonymous
HCC-LM6	Cell line	6553	c.6180C>G	P.(=)	synonymous
HCC-LM6	Cell line	6554	c.6181T>C	P.(=)	synonymous
HCC-LM6	Cell line	6559	c.6186T>C	P.(=)	synonymous
HCC-LM6	Cell line	6562	c.6189A>C	P.(=)	synonymous
HCC-LM6	Cell line	6583	c.6210G>A	P.(=)	synonymous
HCC-LM6	Cell line	6595	c.6222T>C	P.(=)	synonymous
HCC-LM6	Cell line	6601	c.6228C>T	P.(=)	synonymous
HCC-LM6	Cell line	6604	c.6231C>T	P.(=)	synonymous
HCC-LM6	Cell line	6682	c.6309T>C	P.(=)	synonymous
HCC-LM6	Cell line	6685	c.6312C>A	P.(=)	synonymous
HCC-LM6	Cell line	6691	c.6318G>A	P.(=)	synonymous
HCC-LM6	Cell line	6709	c.6336T>C	P.(=)	synonymous

Nature Genetics doi:10/1038/ng.2391

HCC-LM6	Cell line	6715	c.6342G>C	P.(=)	synonymous
HCC-LM6	Cell line	6722	c.6349C>T	P.(=)	synonymous
HCC-LM6	Cell line	6778	c.6405T>C	P.(=)	synonymous
HCC-LM6	Cell line	6787	c.6414A>T	P.(=)	synonymous
HCC-LM6	Cell line	6793	c.6420C>T	P.(=)	synonymous
HCC-LM6	Cell line	6796	c.6423C>T	P.(=)	synonymous
HCC-LM6	Cell line	6820	c.6447C>T	P.(=)	synonymous
HCC-LM6	Cell line	6823	c.6450T>C	P.(=)	synonymous
HCC-LM6	Cell line	6820	c.6447C>T	P.(=)	synonymous
HCC-LM6	Cell line	6823	c.6450T>C	P.(=)	synonymous
HCC-LM6	Cell line	6889	c.6516C>A	P.(=)	synonymous
HCC-LM6	Cell line	6892	c.6519C>T	P.(=)	synonymous
HCC-LM6	Cell line	6898	c.6525C>T	P.(=)	synonymous
HCC-LM6	Cell line	6919	c.6546T>C	P.(=)	synonymous
HCC-LM6	Cell line	6922	c.6549T>G	P.(=)	synonymous
HCC-LM6	Cell line	6946	c.6573T>C	P.(=)	synonymous
HCC-LM6	Cell line	6964	c.6591C>T	P.(=)	synonymous
HCC-LM6	Cell line	6970	c.6597A>G	P.(=)	synonymous
HCC-LM6	Cell line	6985	c.6612C>T	P.(=)	synonymous
HCC-LM6	Cell line	7012	c.6639C>A	P.(=)	synonymous
HCC-LM6	Cell line	7021	c.6648C>G	P.(=)	synonymous
HCC-LM6	Cell line	7024	c.6651C>T	P.(=)	synonymous
HCC-LM6	Cell line	7033	c.6660C>T	P.(=)	synonymous
HCC-LM6	Cell line	7045	c.6672G>A	P.(=)	synonymous
HCC-LM6	Cell line	7081	c.6708C>A	P.(=)	synonymous
HCC-LM6	Cell line	7084	c.6711G>A	P.(=)	synonymous
HCC-LM6	Cell line	7094	c.6721T>C	P.(=)	synonymous
HCC-LM6	Cell line	7108	c.6735C>T	P.(=)	synonymous

Nature Genetics doi:10/1038/ng.2391

HCC-LM6	Cell line	7125	c.6752T>C	P.(=)	synonymous
HCC-LM6	Cell line	7135	c.6762C>T	P.(=)	synonymous
HCC-LM6	Cell line	7138	c.6765A>G	P.(=)	synonymous
HCC-LM6	Cell line	7159	c.6786G>C	P.(=)	synonymous
HCC-LM6	Cell line	7162	c.6789A>G	P.(=)	synonymous
HCC-LM6	Cell line	7168	c.6795G>A	P.(=)	synonymous
HCC-LM6	Cell line	7169	c.6796T>C	P.(=)	synonymous

TP53 mutaion profile in HBV-infected HCC cases

*HCC-68#	HBV-infected HCC	325	c.128T>A	p.Leu43Stop	nonsense
HCC-30#	HBV-infected HCC	399	c.202G>T	p.Glu68Stop	nonsense
HCC-71#	HBV-infected HCC	465	c.268_269insC	fs	indel
HCC-52#	HBV-infected HCC	603	c.406C>T	p.Gln136Stop	nonsense
HCC-56#	HBV-infected HCC	625	c.428_429insG	fs	indel
HCC-56#	HBV-infected HCC	627	c.430_435delTGCAGC	fs	indel
HCC-87#	HBV-infected HCC	649	c.452C>G	p.Pro151Arg	missense
HCC-29#	HBV-infected HCC	669	c.472_473insT	fs	indel
HCC-32#	HBV-infected HCC	670	c.473G>A	p.Arg 158 His	missense
HCC-67#	HBV-infected HCC	672	c.475G>C	p.Ala159Pro	missense
HCC-17#	HBV-infected HCC	721	c.524G>A	p.Arg175His	missense
HCC-95#	HBV-infected HCC	724	c.527G>T	p.Cys176Phe	missense
HCC-65#	HBV-infected HCC	724	c.527G>T	p.Cys176Phe	missense
*HCC-72#	HBV-infected HCC	756	c.559G>A	p.Gly187Ser	missense
HCC-24#	HBV-infected HCC	774	c.577C>T	p.His193Tyr	missense
HCC-88#	HBV-infected HCC	811	c.614A>G	p.Tyr205Cys	missense
HCC-43#	HBV-infected HCC	811	c.614A>G	p.Tyr205Cys	missense

Nature Genetics doi:10/1038/ng.2391

HCC-45#	HBV-infected HCC	812	c.615T>G	p.Tyr205Stop	nonsense
HCC-42#	HBV-infected HCC	841	c.644G>T	p.Ser215Ile	missense
*HCC-56	HBV-infected HCC	856	c.659T>C	p.Tyr220Cys	missense
HCC-15#	HBV-infected HCC	898	c.701A>G	p.Tyr234Cys	missense
HCC-93#	HBV-infected HCC	922	c.725G>T	p.Cys242Phe	missense
HCC-11#	HBV-infected HCC	933	c.736A>G	p.Met246Val	missense
HCC-31#	HBV-infected HCC	943	c.746G>C	p.Arg249Thr	missense
HCC-15#	HBV-infected HCC	944	c.747G>T	p.Arg249Ser	missense
HCC-93#	HBV-infected HCC	944	c.747G>T	p.Arg249Ser	missense
HCC-48	HBV-infected HCC	948	c.751T>A	p.Ile251Phe	missense
HCC-58#	HBV-infected HCC	951	c.754_762delCTCACCATC	fs	indel
HCC-11#	HBV-infected HCC	972	c.775G>T	p.Asp259Tyr	missense
HCC-96#	HBV-infected HCC	997	c.800G>C	p.Arg267Pro	missense
HCC-53#	HBV-infected HCC	1041	c.844C>T	p.Arg282Trp	missense
HCC-38#	HBV-infected HCC	1117	c.920C>T	p.Ala307Val	missense
HCC-77#	HBV-infected HCC		splicesite		
*HCC-55#	HBV-infected HCC		splicesite		
HCC-38#	HBV-infected HCC		splicesite		

Note: * indicated the cases harboring *TP53* and *ARID1A* mutant.

Nature Genetics doi:10/1038/ng.2391

Supplementary Table 11. Summary of the mutated genes examined using PCR-based Sanger sequencing in additional HCC samples.

Gene symbol	Number of examined samples	Number of mutated samples	Nucleotide variations (cDNA)	Amino acid variations (protein)
PFKFB2	117	1	c.G1487A	p.Arg496His
PIK3AP1	116	1	c.C1581T	p.(=)
KIAA0776	112	1	c.A2283G	p.(=)
TMEM35	117	1	c.A427T	p.Lys143Trp
IGSF10	120	1	c.T4752A	p.(=)
HOXA1	118	1	c.G568A	p.Arg289His
ATL1	120	1	c.A84G	p.(=)
VCAM1	89	1	c.C728T	p.Thr243Ile
ABCA1	120	0	—	—
ABCB5	120	0	—	—
ALDH2	120	0	—	—
AMPH	116	0	—	—
ANKRD53	117	0	—	—
APAF1	120	0	—	—
ARID2	120	0	—	—
ARMC4	113	0	—	—

Nature Genetics doi:10/1038/ng.2391

ASB13	114	0	—	—
ATP5S	120	0	—	—
BRIP1	117	0	—	—
BRMS1	119	0	—	—
BRWD1	116	0	—	—
C10orf92	119	0	—	—
C1orf14	116	0	—	—
C1orf77	114	0	—	—
C6orf168	120	0	—	—
C9orf125	119	0	—	—
CAMKV	120	0	—	—
CATSPER1	118	0	—	—
CCKAR	118	0	—	—
CDKL4	119	0	—	—
CHRDL1	119	0	—	—
CHRNA6	115	0	—	—
CHRND	120	0	—	—
CIT	120	0	—	—
CLTC	119	0	—	—
COTL1	120	0	—	—
CRYBA4	118	0	—	—
CTCF	120	0	—	—
DCHS1	116	0	—	—
DDX59	94	0	—	—
DHPS	118	0	—	—

Nature Genetics doi:10/1038/ng.2391

DHX37	119	0	—	—
DSE	92	0	—	—
ETHE1	117	0	—	—
FAM129B	120	0	—	—
FAM38B	114	0	—	—
FAM71A	120	0	—	—
FNTA	120	0	—	—
FSHR	118	0	—	—
FUT10	120	0	—	—
GABRA1	120	0	—	—
GALNT14	116	0	—	—
GCLC	120	0	—	—
GCM1	88	0	—	—
GDNF	116	0	—	—
GHRHR	115	0	—	—
GLB1L2	120	0	—	—
GLUD1	120	0	—	—
GNB5	120	0	—	—
GOSR2	115	0	—	—
GPATCH8	120	0	—	—
GPS1	120	0	—	—
GTDC1	119	0	—	—
HCN1	119	0	—	—
HMCN1	120	0	—	—
HSPB7	117	0	—	—

Nature Genetics doi:10/1038/ng.2391

HTRA3	119	0	—	—
IFNA5	118	0	—	—
INADL	119	0	—	—
IPO8	94	0	—	—
IQGAP1	118	0	—	—
IRAK1BP1	116	0	—	—
IREB2	119	0	—	—
ITGA6	120	0	—	—
KCNJ9	119	0	—	—
KIAA0907	119	0	—	—
LACTB2	118	0	—	—
LHX9	116	0	—	—
LIFR	120	0	—	—
LPHN2	120	0	—	—
LRP2	120	0	—	—
LRRIQ1	120	0	—	—
MANEAL	119	0	—	—
MAP1B	120	0	—	—
MCTP2	118	0	—	—
MDN1	118	0	—	—
MFSD9	118	0	—	—
MMP9	118	0	—	—
MTF1	117	0	—	—
MYT1	120	0	—	—
NEDD9	113	0	—	—

Nature Genetics doi:10/1038/ng.2391

NFASC	94	0	—	—
NLK	94	0	—	—
NRCAM	117	0	—	—
NRXN2	111	0	—	—
NWD1	119	0	—	—
OR2G3	120	0	—	—
OR2M5	110	0	—	—
OR5A2	117	0	—	—
OR5B12	120	0	—	—
OR9Q1	120	0	—	—
PAK6	105	0	—	—
PAPOLG	119	0	—	—
PCDH21	119	0	—	—
PDE10A	117	0	—	—
PFKFB4	109	0	—	—
PFTK1/CDK14	117	0	—	—
PLA2G6	115	0	—	—
PMPCB	115	0	—	—
POGZ	111	0	—	—
POU6F2	117	0	—	—
PPARGC1A	106	0	—	—
PPM2C	116	0	—	—
PPP2R1B	117	0	—	—
PPP6C	116	0	—	—
PRCC	113	0	—	—

Nature Genetics doi:10/1038/ng.2391

RAB3D	104	0	—	—
RAG1	116	0	—	—
RASGRF1	112	0	—	—
RGL3	117	0	—	—
RNF17	112	0	—	—
RUNX2	120	0	—	—
SCN3A	118	0	—	—
SGK493	106	0	—	—
SIGLEC8	99	0	—	—
SLC39A10	118	0	—	—
SMARCAL1	116	0	—	—
SORL1	115	0	—	—
SPAG17	104	0	—	—
SPOCK1	119	0	—	—
STAMBP	85	0	—	—
TAF1L	110	0	—	—
TBX18	119	0	—	—
TBX19	85	0	—	—
TMEM2	118	0	—	—
TNIP1	119	0	—	—
TRIM68	97	0	—	—
TRIM9	110	0	—	—
TSHZ1	115	0	—	—
TSN	118	0	—	—
TTC3	112	0	—	—

Nature Genetics doi:10/1038/ng.2391

VCAM1	119	0	—	—
XKR7	114	0	—	—
ZBTB7B	116	0	—	—
ZMAT2	119	0	—	—
ZNF671	119	0	—	—
ZSCAN16	116	0	—	—

Nature Genetics doi:10/1038/ng.2391

Supplementary Table 12. The sequences of siRNAs.

siRNA ID	Targeting mutated genes	Targeted sequences (sense strand)	Providers
1	ARID1A	GCAGGCACCACUAACUUAU	GenePharma
2	ARID1A	GAUGAGACCUCAGCCAUAU	GenePharma
3	C1orf14	GGGACAAGAUACUGAUAUACA	GenePharma
4	C1orf14	GGACCUUUCUUUCCAAGAAUU	GenePharma
5	C1orf77	GCCUGGGUAAGAGUAACAUCC	GenePharma
6	C1orf77	CCACCAAGAUGUCUCUAAAUG	GenePharma
7	DDX59	GGCCCAUUAUUGACUUUGAAC	GenePharma
8	DDX59	GGCGACUUCUGGAUAUAAUAA	GenePharma
9	FAM71A	CAAGUAUGCACCGAUAUUUGA	GenePharma
10	FAM71A	GACGGUUGGUUCUAUGACACC	GenePharma
11	HMCN1	GGGCGAUCCUAAUGUUGAACU	GenePharma
12	HMCN1	GACCCAUAUUGGAUAUAAAGU	GenePharma
13	KCNJ9	UGCCAGCUUUCACGAGACUUU	GenePharma
14	KCNJ9	CUUCGAGAUCGUCGUUAUCCU	GenePharma
15	LHX9	GGACCAUGAAAUCCUACUUUG	GenePharma
16	LHX9	GGAGUGGACAUCGUCAAUUAC	GenePharma
17	MANEAL	CGGAGCACUUCAUCAAAGAGA	GenePharma
18	MANEAL	GCCUGUAUUUGGACUACCUGC	GenePharma
19	NFASC	CCCUGCGUAUCACAAAUGUCU	GenePharma
20	NFASC	CCGCUCUAUAUUGGAAACAGG	GenePharma
23	OR2G3	GUGGGAAACUUCACCAUAAUC	GenePharma
24	OR2G3	GUGUGGCGCAACUCUAUAUUU	GenePharma
25	OR2M5	GCUCUAUGGAUGCAAUCAUUG	GenePharma
26	OR2M5	CAGGUUUGUUCAUGUACAUAC	GenePharma
27	PRCC	GGAGAGAAGAAAUCAACUUUG	GenePharma
28	PRCC	CAGCAAUGGAUGACUAAGUCA	GenePharma
29	SPAG17	GCUCACCUAACUGAUUUAUUC	GenePharma
30	SPAG17	GGAGAUGGAACAACUAUUAUU	GenePharma
31	TBX19	GCCCAUCGAAUGGUAACAAAC	GenePharma
32	TBX19	CUCCCUCAGUGAAUUUGAUAG	GenePharma
33	VCAM1	CACCCAAGAAUACAGUUAUUU	GenePharma
34	VCAM1	GGAGCCUCAAAUAUACUUUGG	GenePharma
35	ZBTB7B	CUCCUUGAAUUUGCCUAUACA	GenePharma
36	ZBTB7B	GUAGCCACUACUUCAAGAAGC	GenePharma
37	ASB13	AUGUGUGAGGCUUCUUAUUGA	GenePharma
38	ASB13	AGUGCUUCGAGUACUACGAAA	GenePharma
39	PCDH21	CAGCCUUACUCAAGAUUGACA	GenePharma
40	PCDH21	CCAGGUACAACUUCUAUGUGA	GenePharma
41	PIK3AP1	GCUCAAGAGUCACAUUAAAGA	GenePharma
42	PIK3AP1	GAGGACGAGAAGGUUGUUUCC	GenePharma
43	NRXN2	CGGCAGGAAACUUUGAUAACG	GenePharma

Nature Genetics doi:10/1038/ng.2391

44	NRXN2	CUGGAGGAGAGUUAAUAUUGC	GenePharma
45	NRXN2	GCUCCAUGUACAUGAAGAUCA	GenePharma
46	OR5A2	CCCAUACACUUUGUUUAAAGA	GenePharma
47	OR5A2	GUGGGUGGAUUCCUUAGUUCU	GenePharma
48	OR5B12	CUGGUGGACUUUGGUUAUUCC	GenePharma
49	OR5B12	GGCCUCUAUAGGAUUCAUAUU	GenePharma
50	OR9Q1	CCAGUGAGAUUGACUUUAUUU	GenePharma
51	OR9Q1	GUCUGCUACUCAUCUAUCACU	GenePharma
52	PPP2R1B	GGGCAUCAAAUGCUGUUAAAG	GenePharma
53	PPP2R1B	CUGAGCAGCUGGGAAAUUUCA	GenePharma
54	PPP2R1B	GUCCUGACGUUCGUUUGAAUA	GenePharma
55	SORL1	GGACCUCACUACUACACAUGG	GenePharma
56	SORL1	GCUCGAACAAAGAGAAUGUCC	GenePharma
57	SORL1	GGCUGCUAGUAACUUUACAGA	GenePharma
58	ARID2	CUCUCGACUUGCAGUAAAUUA	GenePharma
59	ARID2	GGGUAGGGUUUCAGAACAUUG	GenePharma
60	CIT	GGGACAGGAAGUACAUUGUCC	GenePharma
61	CIT	CCUCAUACCAGGAUAAAUUAA	GenePharma
62	DHX37	GGCCUGUUGAAGACUUAAUCC	GenePharma
63	DHX37	GUGUGCUGCUUAAAGAAAUCC	GenePharma
64	ATL1	CCGAGAGCCUAGAUAUUAAAG	GenePharma
65	ATL1	GCUCAAUACAGGUAUAUAACU	GenePharma
66	KIAA1409	GCUGAGGAAUCCGAAUUUAAG	GenePharma
67	KIAA1409	GCUGCGUAUAUCGCACAAAGA	GenePharma
68	TRIM9	CUGGACAAGAUGAGCCUAUAC	GenePharma
69	TRIM9	CACGGACUUUGACUUGAGUCU	GenePharma
70	IQGAP1	CCCAGUCCUACCAGAAAUUCA	GenePharma
71	IQGAP1	GCUCACGCAAGCUGAAAUUCA	GenePharma
72	PAK6	GAAGUGAUCUCCAGGUCUUUG	GenePharma
73	PAK6	CACAGGUGUUGUGACACAUGA	GenePharma
74	RAB27A	GGACCAGAGAGUAGUGAAAGA	GenePharma
75	RAB27A	GGUGCGAUCAAAUGGUCAUGC	GenePharma
76	COTL1	AGGAGGUCGUACAGAAUUUCG	GenePharma
77	COTL1	UGAGCAAGAGGUCCAAGUUUG	GenePharma
78	FAHD1	GAUCAUAACCUUGGAAGAAGG	GenePharma
79	FAHD1	GGGAGUUGGACCGGUUAAAGA	GenePharma
80	BRIP1	GGAGGCUAAUCAUAUCAUUAA	GenePharma
81	BRIP1	GUCCCUACAACUAUCUUCUAG	GenePharma
82	GPATCH8	GAGAGAGUUUGCUCGAAAUGU	GenePharma
83	GPATCH8	GGCCAUCAGAACUAUUAAUUU	GenePharma
84	GPS1	GUGCCCUCAUCCAGUAUUUCA	GenePharma
85	GPS1	GAGACAUCAUCUUCAAAUUCU	GenePharma
86	NLK	CCGACAGGUUAAAGAAAUUAU	GenePharma
87	NLK	GGUGUUGUCUGGUCAGUAACA	GenePharma
88	CCBE1	GCCAUACUGUCUGGAUAUUGA	GenePharma

Nature Genetics doi:10/1038/ng.2391

89	CCBE1	CCUGAUCUGUCCCACAUUAAG	GenePharma
90	TSHZ1	CAGCUUAGCUCUGGAUUUAAA	GenePharma
91	TSHZ1	GCCACUCAAUCAGAUUCAAGA	GenePharma
92	DHPS	GGGAUCAAUAGGAUCGGAAAC	GenePharma
93	DHPS	GCGACAUGAUCUUCUUCCAUU	GenePharma
94	ELL	GAGGCCAUCUAUCCGAUUUCA	GenePharma
95	ELL	GCAGCAUACAGGACAAGAUCA	GenePharma
96	ETHE1	GGCUGACUUACACAUUGAGGA	GenePharma
97	ETHE1	UCUACCCUGCUCACGAUUACC	GenePharma
98	NWD1	GGAUCGCUCUAUACUUGUUUG	GenePharma
99	NWD1	GAAGAUCGGUGCGGAUAUUCU	GenePharma
100	ZNF671	GGCCCAAUAUUGAAAGAUACC	GenePharma
101	ZNF671	GUCGGAAACACACACUUGUUC	GenePharma
102	CHRND	UCUCCCUGAAAGAAGUUGAGG	GenePharma
103	CHRND	GUGGAUCAUCAUUGAUCCUGA	GenePharma
104	FSHR	GCUGAAUCUAAGCGAUAAUAA	GenePharma
105	FSHR	GUGCUAGUGAUCCUAACUACC	GenePharma
106	MFSD9	GUGAUGGGAAUAGUAAAUUAA	GenePharma
107	MFSD9	GGCUACCUCAUCAGUUACAGC	GenePharma
108	PAPOLG	GGAACGGAAUGAAUUUAUUAC	GenePharma
109	PAPOLG	ACCGCAUCUGAAUGGAAUGUC	GenePharma
110	SMARCAL1	GAGGCAGACAUCAGUUAUUCA	GenePharma
111	SMARCAL1	GCGCAAGAUAGUGGUGAUUGC	GenePharma
112	EMILIN3	GCUACGUAAAGGCUGAAUACC	GenePharma
113	EMILIN3	CAGUACUCAGACCCAAAUACA	GenePharma
116	BRWD1	CGGAGCUGUACUUCCUUAUCG	GenePharma
117	BRWD1	GGCGUAAAUAUCCAAAUUAUG	GenePharma
118	BRWD1	GGCCGCUUGUUAUCUACAUUA	GenePharma
123	PFKFB4	GUGGUCAAGACCUACAAAUCU	GenePharma
124	PFKFB4	CGAGGAAAUUCAGGAUAAUUA	GenePharma
125	ACSL1	GUGGGUGAUUAUUGAACAAGG	GenePharma
126	ACSL1	GAGCUGCGGAACUAUUUCAGG	GenePharma
127	HTRA3	GGUUCUGGCUUCAUCAUGUCA	GenePharma
128	HTRA3	CGGAUGCCAUCAUCAACUACG	GenePharma
129	PPARGC1A	CUGGUACACAAGGCAAUAACU	GenePharma
130	PPARGC1A	GUCGCAGUCACAACACUUACA	GenePharma
131	DOCK2	GCCAGAUUAUGCAAUGUAUUC	GenePharma
132	DOCK2	GACCAUCAUAGGCUACUUUGA	GenePharma
133	GABRA1	CCGCUGCUAUUUGGAAUCUUU	GenePharma
134	GABRA1	GCCGUCAUUACAAGAUGAACU	GenePharma
135	GDNF	GCUGAGCAGUGACUCAAAUAU	GenePharma
136	GDNF	AGGCUGGUGAGUGACAAAGUA	GenePharma
137	GDNF	AUGGCAGUGCUUCCUAGAAGA	GenePharma
138	HCN1	GAGGUGUCCACUCUGAUUUCC	GenePharma
139	HCN1	CCAUCCCAGUGGAUUAUAUCU	GenePharma

Nature Genetics doi:10/1038/ng.2391

140	LIFR	CUUGCGACUACGUCAUUAAGU	GenePharma
141	LIFR	GGAAGUUCUACAAGUAAAUUC	GenePharma
142	MAP1B	GACGCCCAAUGAGAUUAAAGU	GenePharma
143	MAP1B	CUCCCGAUUUCCUACUUAACU	GenePharma
144	SPOCK1	UGCCAUCUACCUGGAUAAGUA	GenePharma
145	SPOCK1	CCGCUUUCGAGACGAUGAUUA	GenePharma
146	BTN3A3	GGAGAGAUCUCGUUCUAUAAU	GenePharma
147	BTN3A3	CUGCAACAGAGCAAGAAAUAA	GenePharma
148	DSE	GGCAGGCAAACGCUAUAAAUU	GenePharma
149	DSE	GGUCUAAAUACAAGCAUGACC	GenePharma
150	GCLC	GGCAGACAAUGAGAUUUAAGC	GenePharma
151	GCLC	GGAGGCUACUUCUAUAUUAGA	GenePharma
152	GCM1	GAGGAAGAAAUGACAUACUUG	GenePharma
153	GCM1	GGACAUUUAAUAGCUAACACU	GenePharma
154	KIAA0776	GGACUAUUCAGUGCUAUUACC	GenePharma
155	KIAA0776	GUGGUCGAGUAAACAUUGUUG	GenePharma
156	RUNX2	CACGCUAUUAAAUCCAAAUUU	GenePharma
157	RUNX2	GCAGUUCCCAAGCAUUUCAUC	GenePharma
158	ZSCAN16	CCCAGGCAAUUCAGAAAGACG	GenePharma
159	ZSCAN16	CCUGAGCAUUCUUCCUAAAGA	GenePharma
160	AMPH	GGUGGAAACACUGCAUGAUUU	GenePharma
161	AMPH	GUCGCUGCAUGAAGUCUAUGA	GenePharma
162	GHRHR	GAGGAGGAAUCUUACUUCUCC	GenePharma
163	GHRHR	CACCCAGAAUGUGACUUCAUC	GenePharma
164	HOXA1	GACCAUAGGAUUACAACUUUC	GenePharma
165	HOXA1	GCUAUGGCUCACAGAACUUCA	GenePharma
166	NRCAM	GCUCCUUUAUUGGACAAUACA	GenePharma
167	NRCAM	CAGGAACGCUCAUAAUUAACA	GenePharma
168	NRCAM	GCACCACCAAACUGAAGUUUC	GenePharma
169	CDK14	GGUUCUUGGAACACCAAAUUU	GenePharma
170	CDK14	GGACACCUUUCACAGCUAUUU	GenePharma
171	CHRNA6	GGUGGACAGAGUAUUUCUUUG	GenePharma
172	CHRNA6	GCCACAAGCAAGAGAAGAUUA	GenePharma
173	FNTA	GCAUCGACAAUGGGUUAUUCA	GenePharma
174	FNTA	CCGGGAUGCUAUUGAGUUAAA	GenePharma
175	FUT10	GGGCUAAUAUCAGGCUUCAGG	GenePharma
176	FUT10	GUGGAGCAGAUGCUUGUUUCU	GenePharma
177	MTMR7	CUCCCUCCAUGAGUGAUUUCC	GenePharma
178	MTMR7	CUCCAGUUAUUGACCAGUUCA	GenePharma
181	ABCA1	CUCCGAGUCAAGAAGUUAAUG	GenePharma
182	ABCA1	GCACUAGGAUGGCAAUCAUGG	GenePharma
185	TAF1L	GUGGAUUUCAGUAGUUACUCU	GenePharma
186	TAF1L	CCCAGUUAAUAAGAAGUUUGU	GenePharma
187	TMEM2	GGUGCUCCGAGGUAUUAAUCA	GenePharma
188	TMEM2	GGAGACCCGUCUGUUAUUUCU	GenePharma

Nature Genetics doi:10/1038/ng.2391

189	TMEM35	GGAGCAACCCUCCUUAUAUGA	GenePharma
190	TMEM35	CGUGCUUACAAGAGCUAUGUU	GenePharma
191	LRRIQ1	GACACCCGCUUUGGAUAAACU	GenePharma
192	LRRIQ1	GGCUAGUUGAGGAAUCAAAUA	GenePharma
197	DISC1	GAGGUAAUAUCCUUAAGAUUA	GenePharma
198	DISC1	GGAAGAAAGUUAACGAUAUUG	GenePharma
199	DISC1	GCGUGACAUGCAUUCUUUACC	GenePharma
200	DISC1	GUGCCUUUACCUCAAGCUUUA	GenePharma
201	CDKL4	CCUCUGGACAAGUAGUAGCUGUUAA	Invitrogen
202	CDKL4	GCACUAAGAGAAAUACGUAUGUUGA	Invitrogen
203	CDKL4	CCCAAGACAUCAAUCAAUCUUUAAA	Invitrogen
204	CSNK1G3	GGCUGACACAUUAAAGGAGAGGUAU	Invitrogen
205	CSNK1G3	UCUUCGUUAUGUAAGAAGGCUAGAU	Invitrogen
206	CSNK1G3	GACUUGUUUGACUUGUGUGACAGAA	Invitrogen
207	NLK	GGGAUGUAUCUUUGCAGAACUACUA	Invitrogen
208	NLK	GGAAGCAAUGAGGACAGCUUGUGAA	Invitrogen
209	NLK	GGCAGCCGUCAUUACAGCAAUGCUA	Invitrogen
210	CAMKV	GCCAUCGGAGGUGACUGACAGAUAU	Invitrogen
211	CAMKV	GGGAUGAUAUUUCGCAGGCAGCCAA	Invitrogen
212	CAMKV	GCUUUCAGGCAAUCCACCUUUCUAU	Invitrogen
213	CIT	GAGCAGUGACUUUCUUGAUCUGAUU	Invitrogen
214	CIT	CCCACUUCAACUCACUCGAAGUAAU	Invitrogen
215	CIT	ACGAGGAGCAGAAACUGGAGCUCAA	Invitrogen
216	CDK14	GCAUUCUGAGAACAAUGCUUGCAUU	Invitrogen
217	CDK14	CCUCCACUGGCAAAGAGUCACCUAA	Invitrogen
218	CDK14	AGGCUGCCUUGAGCCACGAGUAUUU	Invitrogen
219	PAK6	ACUUCAACGUGGUGGAGAUGUACAA	Invitrogen
220	PAK6	GCUCAGAUCAGCAAAGACGUCCCUA	Invitrogen
221	PAK6	ACAUCGUCUCCCAAGUCAGGCUGAA	Invitrogen
222	PKDCC	CAGUAUCUGCAGAACUCCACGGCAA	Invitrogen
223	PKDCC	GAGUACCAGUGUAUCCCAGACAGCA	Invitrogen
224	PKDCC	CCAACGUGCUGCAGCUCUAUGGCUA	Invitrogen
225	Negative Control	UUCUCCGAACGUGUCACGU	GenePharma

Nature Genetics doi:10/1038/ng.2391

Supplementary Table 13. RNAi screen of the mutated genes via statistical evaluation and scoring.

Genes	Description	No. of cell line	Cell line	Z score	-$Log_{10}P$
Loss of function promotes cell viability					
VCAM1	Vascular cell adhesion molecule 1	**4**	MHCC- LM3	5.54	3.02
			Huh-7	5.4	3.76
			MHCC-97H	5.1	4.13
			MHCC-97L	7.02	2.36
TMEM2	Transmembrane protein 2	**4**	MHCC- LM3	21.87	2.41
			PLC/PRF/5	17.34	3.74
			MHCC-97L	5.74	4.9
			MHCC-97H	18.02	4.33
LHX9	LIM homeobox 9	**3**	Focus	6.78	2.16
			MHCC-97L	6.22	4.17
			MHCC-97H	10.62	10.61
BTN3A3	Butyrophilin, subfamily 3, member A3	**2**	MHCC- LM3	8.43	2.53
			MHCC-97H	6.64	3.5
ACSL1	Acyl-CoA synthetase long-chain family member 1	**2**	Huh-7	40.1	4.44
			MHCC-97H	14.18	4.02
DSE	Dermatan sulfate epimerase	**2**	MHCC- LM3	6.53	2.23
			MHCC-97L	8.94	2.62
FAHD1	Fumarylacetoacetate hydrolase domain containing 1	**2**	MHCC- LM3	5.51	2.61
			Focus	14.76	3.87
GPATCH8	G patch domain containing 8	**2**	MHCC- LM3	15.47	2.24
			MHCC-97H	5.87	2.85
LRRIQ1	Leucine-rich repeats and IQ motif containing 1	**2**	MHCC- LM3	20.71	3.07

Nature Genetics doi:10/1038/ng.2391

NRXN2	Neurexin 2	**2**	MHCC-97L	6.6	5.48
			MHCC-LM3	7.65	2.14
			MHCC-97L	10.98	3.3
PAK6	P21(CDKN1A)-activated kinase 6	**2**	Huh-7	12.77	2.23
			MHCC-97L	7.29	2.03
ATL1	Atlastin GTPase 1	**1**	MHCC-97H	6.2	3.34
COTL1	Coactosin-like 1 (Dictyostelium)	**1**	MHCC- LM3	6.01	3.89
DHPS	Deoxyhypusine synthase	**1**	MHCC-97L	12.18	2.7
DISC1	Disrupted in schizophrenia 1	**1**	MHCC- LM3	6.71	5.28
FNTA	Farnesyltransferase, CAAX box, alpha	**1**	MHCC-97H	5.05	2.01
GCM1	Glial cells missing homolog 1 (Drosophila)	**1**	MHCC-97L	10.06	3.22
GDNF	Glial cell derived neurotrophic factor	**1**	MHCC- LM3	6.05	3.61
GPS1	G protein pathway suppressor 1	**1**	PLC/PRF/5	8.73	2.18
HCN1	Hyperpolarization activated cyclic nucleotide-gated potassium channel 1	**1**	WRL68	6.48	2.55
KCNJ9	Potassium inwardly-rectifying channel, subfamily J, member 9	**1**	PLC/PRF/5	9.07	3.54
KIAA0776	KIAA0776	**1**	MHCC- LM3	7.62	6
NFASC	Neurofascin homolog (chicken)	**1**	PLC/PRF/5	36.36	6.02
OR5A2	Olfactory receptor, family 5, subfamily A, member 2	**1**	MHCC- LM3	7.21	3.84
PIK3AP1	Phosphoinositide-3-kinase adaptor protein 1	**1**	PLC/PRF/5	7.13	2.21
PPARGC1A	Peroxisome proliferator-activated receptor gamma, coactivator 1 alpha	**1**	MHCC-97H	9.58	3.24
RUNX2	Runt-related transcription factor 2	**1**	MHCC-97H	53.19	3.72
TAF1L	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 210kDa-like	**1**	MHCC- LM3	18.5	2.52

Nature Genetics doi:10/1038/ng.2391

ZSCAN16	Zinc finger and SCAN domain containing 16	**1**	WRL68	5.98	3.66

Loss of function suppresses cell viability

CDK14	Cyclin-dependent kinase 14	**4**	WRL68	-8.67	4.11
			Huh-7	-7.42	2.78
			MHCC-97L	-16.73	2.61
			HCC-LM3	-8.4	2.13
HOXA1	Homeobox A1	**4**	Huh-7	-8.13	2.5
			Focus	-7.1	2.53
			PLC/PRF/5	-6.24	4.34
TMEM35	Transmembrane protein 35	**4**	MHCC-97L	-6.67	2.48
			PLC/PRF/5	-10.67	4.81
			MHCC-97L	-11.56	2.5
			MHCC-97H	-14.33	2.93
ELL	Elongation factor RNA polymerase II	**4**	HCC-LM3	-12.41	3.01
			Huh-7	-10.25	2.54
			Focus	-6.22	3.34
			MHCC-97H	-9.89	2.72
CSNK1G3	Casein kinase 1, gamma 3	**3**	HCC-LM3	-8.39	3.43
			Huh-7	-6.35	2.55
			MHCC-97L	-11.84	2.8
			MHCC-97H	-9.49	3.18
ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1	**3**	PLC/PRF/5	-27.5	2.13
			MHCC-97L	-20.89	3.54
			MHCC-97H	-57.91	2.8
CHRNA6	Cholinergic receptor, nicotinic, alpha 6	**2**	Huh-7	-16.68	2.65

Nature Genetics doi:10/1038/ng.2391

ETHE1	Ethylmalonic encephalopathy 1	**2**	MHCC-LM3	-5.53	2.83
			Huh-7	-6.35	2.33
			PLC/PRF/5	-17.9	2.67
GHRHR	Growth hormone releasing hormone receptor	**2**	Focus	-22.31	6
			PLC/PRF/5	-10.65	5.01
IQGAP1	IQ motif containing GTPase activating protein 1	**2**	MHCC-97L	-13.81	2.86
			MHCC-97H	-12.84	2.03
NLK	Nemo-like kinase	**2**	Focus	-9.22	3.03
			MHCC-97L	-9.28	2.19
NWD1	NACHT and WD repeat domain containing 1	**2**	WRL68	-11.38	2.1
			MHCC-97H	-43.35	2.51
OR9Q1	Olfactory receptor, family 9, subfamily Q, member 1	**2**	Huh-7	-12.87	2.24
			MHCC-LM6	-11.84	2.74
SMARCAL1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a-like 1	**2**	WRL68	-13.75	8.92
			MHCC-97H	-29.17	2.55
TBX19	T-box 19	**2**	MHCC-LM3	-6.69	2.04
			MHCC-LM6	-6.52	3.69
ARID2	AT rich interactive domain 2 (ARID, RFX-like)	**1**	MHCC-LM3	-7.91	3.54
C1orf77	Chromosome 1 open reading frame 77	**1**	MHCC-LM3	-6.37	2.6
CCBE1	Collagen and calcium binding EGF domains 1	**1**	WRL68	-11.74	4.68
CDHR1	Cadherin-related family member 1	**1**	Huh-7	-20.24	2.22
CIT	Citron (rho-interacting, serine/threonine kinase 21)	**1**	MHCC-97L	-10.42	2.77
EMILIN3	Elastin microfibril interfacer 3	**1**	WRL68	-16.63	4.69
FAHD1	Fumarylacetoacetate hydrolase domain containing 1	**1**	MHCC-97L	-24.27	2.67
FUT10	Fucosyltransferase 10 (alpha (1,3) fucosyltransferase)	**1**	WRL68	-17.09	7.71
GABRA1	Gamma-aminobutyric acid (GABA) A receptor, alpha	**1**	MHCC-LM6	-5.68	3.07

Nature Genetics doi:10/1038/ng.2391

	1				
GCM1	Glial cells missing homolog 1 (Drosophila)	**1**	MHCC-LM6	-20.19	4.24
GPS1	G protein pathway suppressor 1	**1**	WRL68	-7.02	3.55
HMCN1	Hemicentin 1	**1**	MHCC-LM6	-14.18	3.85
KIAA1409	KIAA1409	**1**	MHCC-97H	-14.88	2.13
LHX9	LIM homeobox 9	**1**	WRL68	-6.61	2.67
MAP1B	Microtubule-associated protein 1B	**1**	PLC/PRF/5	-6.67	2.43
NRCAM	Neuronal cell adhesion molecule	**1**	MHCC-LM3	-8.51	2.12
OR2G3	Olfactory receptor, family 2, subfamily G, member 3	**1**	MHCC-LM3	-10.54	2.1
OR2M5	Olfactory receptor, family 2, subfamily M, member 5	**1**	MHCC-97H	-24.65	2.4
RUNX2	Runt-related transcription factor 2	**1**	WRL68	-11.94	6.65
TAF1L	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 210kDa-like	**1**	MHCC-LM3	-5.39	2.04
TASP1	Taspase, threonine aspartase, 1	**1**	MHCC-LM3	-11.16	2.01
TRIM9	Tripartite motif-containing 9	**1**	WRL68	-5.36	2.01
ZNF671	Zinc finger protein 671	**1**	Focus	-8.92	5.68

Nature Genetics doi:10/1038/ng.2391

Supplementary Table 14. Some genes identified in RNAi screen were further evaluated in HCC cell lines to determine their genotypes and knockdown effects.

Genes	Genotype of HCC cell lines		Whether target gene was significantly knockdowned by siRNAs in corresponding cell line
VCAM1	MHCC- LM3	Wildtype	Yes
	Huh-7	Wildtype	Yes
	MHCC-97H	Wildtype	Yes
	MHCC-97L	Wildtype	Yes
TMEM2	MHCC- LM3	Wildtype	Yes
	PLC/PRF/5	Wildtype	Yes
	MHCC-97L	Wildtype	Yes
	MHCC-97H	Wildtype	Yes
LHX9	Focus	Wildtype	No detected expression
	MHCC-97L	Wildtype	
	MHCC-97H	Wildtype	
CDK14	WRL68	Wildtype	Yes
	Huh-7	Wildtype	Yes
	MHCC-97L	Wildtype	Yes
	MHCC-LM3	Wildtype	Yes
HOXA1	Huh-7	Wildtype	Yes
	Focus	Wildtype	Yes
	PLC/PRF/5	Wildtype	Yes
	MHCC-97L	Wildtype	Yes
TMEM35	PLC/PRF/5	Wildtype	Yes
	MHCC-97L	Wildtype	Yes
	MHCC-97H	Wildtype	Yes
	MHCC-LM3	Wildtype	Yes
ELL	Huh-7	Wildtype	Yes
	Focus	Wildtype	Yes
	MHCC-97H	Wildtype	Yes
	MHCC-LM3	Wildtype	Yes
CSNK1G3	Huh-7	Wildtype	Yes
	MHCC-97L	Wildtype	Yes
	MHCC-97H	Wildtype	Yes
ABCA1	PLC/PRF/5	Wildtype	No significant knockdown
	MHCC-97L	Wildtype	
	MHCC-97H	Wildtype	

Nature Genetics doi:10/1038/ng.2391