**Supplementary Material and Methods**

**Alignment of high-throughput sequencing reads and somatic variant detection**

Analysis of sequencing data and somatic variant calling were performed as described previously ([1](#_ENREF_1), [2](#_ENREF_2)). Paired-end reads generated from high-throughput sequencing were aligned to the reference genome (Build hg19/GRCh37) using BWA-MEM (version: 0.7.10-r789, URL) in default mode. Duplicates were marked with Samblaster ([3](#_ENREF_3)), and the bam files were further processed according to GATK best practices using Genome Analysis Toolkit (version: 3.4, URL), by performing Indel Realignment and Base Quality Score Recalibration.

Somatic variants were detected using VarScan2 somatic command (version: 2.4.0) ([4](#_ENREF_4)). A locus covered by at least 10 reads in germline and 14 reads in tumor was considered for variant calling. Initial somatic variants were processed with VarScan processSomatic command to detect high confident variants with the requirements of VAF being more than 0.08 in the tumor while less than 0.02 in the germline, and P-value of the Fisher's exact test being less than 0.01. High confident variants were further filtered for potential false-positives using VarScann2 fpfilter command. Variants supported by a minimum 7 reads for non-reference allele were considered for downstream analyses. Candidate variants were annotated using annovar and filtered against dbSNP135 for commonly found SNPs in general population, while retaining clinically associated and COSMIC variants ([5](#_ENREF_5)). Finally, candidate variants were manually checked and curated.

**Identification of potential driver mutations in M-WES**

To identify potential driver mutations in our cohort, we first nominated potential cancer genes by comparing all mutated genes against: i) COSMIC cancer gene census (Aug 2015) ([6](#_ENREF_6)); ii) Pan-cancer analysis by Lawrence *et al*. ([7](#_ENREF_7)); and iii) recent large-scale HCC sequencing results ([8-12](#_ENREF_8)). Next, all of the nonsilent variants in these potential cancer genes were evaluated, and identified as driver mutations only when they met one of these 3 following criteria: 1) If the gene was annotated as being recessive by COSMIC, and the variant caused damage to the protein (Stopgain, splicing or frameshift); 2) If the gene was annotated as being recessive by COSMIC, and the variant was predicted as deleterious by either SIFT ([13](#_ENREF_13)) or Polyphen ([14](#_ENREF_14), [15](#_ENREF_15)); 3) In COSMIC either exact matched mutation or ≥ 3 mutations located within 15bp of the variant were found.

**Analysis of copy number variations, tumor cellularity and cancer cell fraction (CCF)**

For in-house data which had deep-sequencing results only:

Copy number alterations in tumor samples were detected using ReCapSeg, which is re-implemented as a part of Genome Analysis Tool Kit (GATK v4). Briefly, read counts for each of the exome targets were extracted from all samples and were divided by the total number of reads to generate proportional coverage. Each of the tumor samples was compared to a panel of normal controls, followed by tangent normalization. These normalized coverage profiles were then segmented using circular binary segmentation, which were fed into ABSOLUTE algorism to determine CCF ([16](#_ENREF_16)).

For TCGA HCC data which had matched deep-sequencing and Affymetrix SNP Array 6.0 results:

TCGA Mutation Annotation Format (MAF) files along with matched segmented copy number files (Generated by Affymetrix Genome-Wide Human SNP Array 6.0) for liver cancer samples were downloaded from TCGA Data Portal (URL) on Aug 2015. Similarly, tumor cellularity was estimated using ABSOLUTE algorism based on copy number results ([16](#_ENREF_16)). CCF of each mutation was estimated by integrating variant allele frequency, copy number data and tumor cellularity as outlined by McGranahan et al ([17](#_ENREF_17)). Clonal status of each mutation was estimated based on the confidence interval (CI) of the CCF: mutations were defined as subclonal if the 95% CI was less than 1, and otherwise clonal.

**Mutational signature analysis**

Somatic variants (both silent and non-silent) were classified as either truncal or branch based on either their presence or absence across all tumor sections of each HCC case. Truncal and branch variants were used separately to generate mutational signatures. We used a recently published multiple regression approach, deconstructSigs ([18](#_ENREF_18)), to extract the signatures based on COSMIC Mutational Signature Framework ([19](#_ENREF_19)), and to quantify the contribution of each signature for each HCC subject.

**Enrichment analysis of variably- and invariably methylated probes**

Genomic contexts defined as followings were used: Promoters were defined as 1.5kbp up/down stream of RefSeq TSS, CpG islands (HMM method) ([20](#_ENREF_20)), shores and enhancers were defined using standard Illumina 450K annotation. Partially Methylated Domains (PMD) were called using Roadmap ([21](#_ENREF_21)) liver sample using MethPipe method ([22](#_ENREF_22)). We also included 10 Asian and HBV+ samples as well as 50 nonmalignant liver samples from TCGA for Supplementary Figure 6. Enrichment/Depletion p-value for the presence of variable and invariable probes within each genomic contexts were computed from hypergeometric test based on the background as the probes remaining on the array after the filtering step described above.

**Analysis of the intratumoral heterogeneity of TCGA HCC samples**

We used TCGA exome data to estimate the intratumoral genomic heterogeneity based on Mutant-Allele Tumor Heterogeneity (MATH), a mathematical algorithm developed recently by Mroz et al ([23](#_ENREF_23), [24](#_ENREF_24)). Briefly, MATH is the ratio of the standard deviation to the mean of the distribution of Variant-Allele Fractions (VAF) among all mutated loci. Because both the subclonality and copy number variations of the mutation affect VAF value, a heterogeneous tumor would tend to have a wider distribution of VAF among all mutation loci, centered at a lower fraction, than a homogeneous tumor. Following this method, we first calculated the VAF value for each mutant locus based on the publicly available WES data, and we then determined both the median and the median absolute deviation (MAD) of its VAF values. In the last step, the MATH value for each HCC sample was calculated as the percentage of MAD to the median of the distribution of VAFs among the mutant loci.

In the analysis of Methylation Intratumoral Heterogeneity (mITH), we selected tumor-specific hypermethylated CpG loci by requiring that they had low methylation ( values < 0.3) in all normal liver samples and immune cells (9), and were hypermethylated ( values > 0.3) in tumor samples. We computed mITH score as the median absolute deviation of  values of hypermethylated probes, divided by the median (essentially a similar formula as the genetic-based MATH score). The mITH score thus captures the extent of the intratumoral heterogeneity at the DNA methylation level.

**Assessment of the degree of intratumoral methylation heterogeneity by Shannon entropy**

We employed an adjusted Shannon entropy measurement specifically designed for analysis of DNA methylation data ([25](#_ENREF_25)). This entropy value measures how variable the relative methylation values are across different regions; ranging from zero (probes differentially methylated in a single region) to a maximal score based on the number of regions (probes uniformly methylated across all regions). In other words, lower entropy score corresponds to larger methylation heterogeneity between regions of the same tumor.

**Immunohistochemistry (IHC)**

Preparation of formalin-fixed, paraffin-embedded sections (5 µm thickness) and IHC staining were performed by standard protocols as described before ([26](#_ENREF_26)). Briefly, tissue sections were de-waxed, followed by antigen retrieval. Endogenous peroxidase activity was blocked by 1% H2O2 incubation. Primary antibodies against p53 (Cell Signaling Technology, #2527), FAT4 (Origene, TA322181) and CD31 (Golden Bridge Biotechnology, ZA-0568) were diluted in the blocking buffer and incubated at 4℃ overnight. The slides were washed in PBST and incubated with secondary antibody (Fuzhou Maixin Biotech., Kit-9706-C) for 10 min at room temperature. DAB (3,3'-Diaminobenzidine) substrate was used to detect the bound peroxidase on the slides, and the counter-staining was performed with haematoxylin. Slides were mounted by neutral balsam (Sinopharm Chemical Reagent, 10004160).

**Supplementary Figure Legend**

**Supplementary Figure 1. Unmodified phylogenetic trees associated with Figure 1**

Phylogenetic trees were constructed from all somatic variants by Wagner parsimony method using PHYLIP (See Method).

**Supplementary Figure 2. Number of clonal and subclonal variants of HCC cases sequenced by TCGA**

Each column represents one individual HCC case, and is ordered by increased ratio of subclonal/clonal variants.

**Supplementary Figure 3. Intratumoral heterogeneity of SCNAs in HCC**

Heatmap exhibiting significant SCNAs for each tumor region. Chromosome segment with a log2 ratio greater than 0.5 was classified as gain (Red), and those with less than -0.5 were categorized as loss (Blue).

**Supplementary Figure 4.** **Six mutation classes in HCC**

Fraction analysis of both truncal and branch variants on the basis of the six mutation classes in each HCC case. Numbers of variants are indicated on top of the columns.

**Supplementary Figure 5.** **Permutation test of deconstructSigs**

Results of permutation test calculating the contributions of each of 30 signatures using all possible combinations given the number of samples from 1 to 11 with the method deconstructSigs.

**Supplementary Figure 6. Overall intratumoral heterogeneity of DNA methylation in HCC**

Unsupervised hierarchical clustering of the methylation profiling of five HCC. Rows of the heat maps denote the methylation levels of the variable CpG sites across all samples. Columns represent samples.

**Supplementary Figure 7. Inter-individual homogeneity of normal liver tissues**

Unsupervised hierarchical clustering of the methylation profiling of 22 in-house HCC tumors, 4 in-house normal liver tissues, 10 TCGA HCC tumors and 50 TCGA normal liver samples. Rows of the heat maps denote the methylation levels of the variable CpG sites across all samples. Columns represent samples. Considering the possible impact of ethic and viral factors, we only included HBV+ Asian HCC cases from TCGA.

**Supplementary Figure 8. Invariably methylated DNA in HCC**

The methylation levels of invariably hyper- or hypo-methylated CpG sites across different tumor regions were indicated by the heat maps. Columns represent samples. Number of probes (n) is provided for each case.

**Supplementary Figure 9. Estimation of the degree of intratumoral methylation heterogeneity**

(A) The density distribution of the entropy values of all the probes in each case, which was determined with the analytic method developed by Zhang et al ([25](#_ENREF_25)). Note that we performed bootstrapping to down-sample cases with 5 intratumoral regions (HCC8010 and HCC8257) to all possible 4-region combinations (shown as lines having the same color), in order to directly compare with the other three cases, which had only 4 intratumoral regions. (B) Correlation between the mean of entropy values with the number of variable probes across different intratumoral regions in each case. me-ITH, methylation intratumoral heterogeneity. (C) Correlation between the mean of entropy values with Euclidean pair-wise distance matrix, which measured the global methylation difference across different intratumoral regions in each case.

**Supplementary Figure 10. Prognostic values of** **genomic and epigenomic intratumoral heterogeneity.**

Kaplan-Meier survival curves of patients in TCGA HCC cohort with either low or high intratumoral heterogeneity, on the basis of either MATH scores (A, B) or mITH scores (C, D). Different cutoffs and *P* values are indicated in each plot.

**Supplementary Figure 11. Histological examination of the present HCC samples**

(A) Representative H&E photos of HCC samples. (B) Representative photos of IHC staining of CD31 in HCC samples. N.A., not available.

**References**

1. Lin DC*, et al.* (2014) The genomic landscape of nasopharyngeal carcinoma. *Nature genetics* 46(8):866-871.

2. Lin DC*, et al.* (2014) Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nature genetics* 46(5):467-473.

3. Faust GG & Hall IM (2014) SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 30(17):2503-2505.

4. Koboldt DC*, et al.* (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research* 22(3):568-576.

5. Wang K, Li M, & Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* 38(16):e164.

6. Futreal PA*, et al.* (2004) A census of human cancer genes. *Nature reviews. Cancer* 4(3):177-183.

7. Lawrence MS*, et al.* (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505(7484):495-501.

8. Fujimoto A*, et al.* (2012) Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nature genetics* 44(7):760-764.

9. Guichard C*, et al.* (2012) Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nature genetics* 44(6):694-698.

10. Kan Z*, et al.* (2013) Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome research* 23(9):1422-1433.

11. Schulze K*, et al.* (2015) Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nature genetics* 47(5):505-511.

12. Totoki Y*, et al.* (2014) Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nature genetics* 46(12):1267-1273.

13. Kumar P, Henikoff S, & Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4(7):1073-1081.

14. Adzhubei IA*, et al.* (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7(4):248-249.

15. Adzhubei I, Jordan DM, & Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* Chapter 7:Unit7 20.

16. Carter SL*, et al.* (2012) Absolute quantification of somatic DNA alterations in human cancer. *Nature biotechnology* 30(5):413-421.

17. McGranahan N*, et al.* (2015) Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med* 7(283):283ra254.

18. Rosenthal R, McGranahan N, Herrero J, Taylor BS, & Swanton C (2016) DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome biology* 17:31.

19. Alexandrov LB*, et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463):415-421.

20. Wu H, Caffo B, Jaffee HA, Irizarry RA, & Feinberg AP (2010) Redefining CpG islands using hidden Markov models. *Biostatistics* 11(3):499-514.

21. Kundaje A*, et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539):317-330.

22. Song Q*, et al.* (2013) A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PloS one* 8(12):e81148.

23. Mroz EA*, et al.* (2013) High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma. *Cancer* 119(16):3034-3042.

24. Mroz EA, Tward AD, Hammon RJ, Ren Y, & Rocco JW (2015) Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. *PLoS medicine* 12(2):e1001786.

25. Zhang Y*, et al.* (2011) QDMR: a quantitative method for identification of differentially methylated regions by entropy. *Nucleic acids research* 39(9):e58.

26. Lin DC*, et al.* (2015) Genomic and Functional Analysis of the E3 Ligase PARK2 in Glioma. *Cancer research* 75(9):1815-1827.