Supplementary Methods:

**Experimental Procedure:**

**Tissue collection**

Normal liver tissue was obtained from deceased donor at the time of organ procurement at Zhongshan Hospital, Fudan University (Shanghai, China) after research consent from family was obtained. Hepatoma sample was obtained from a patient with hepatic carcinoma in Zhongshan Hospital (Shanghai, China). Informed consent was obtained from all subjects. Samples were flash frozen with liquid nitrogen.

**Hepatocellular carcinoma (HCC) cell lines**

97L and LM3 are hepatoma cell lines provided by the Liver Cancer Institute of Fudan University (Shanghai, China). 97L and LM3 are both derived from the parent metastatic human HCC cell line MHCC97, but have different metastatic potentials. 97L (full name MHCC97-L) has less metastatic potential while MHCC97-H has higher metastatic potential to lung. For LM3, MHCC97-H were inoculated into BALB/c nude mice, and the pulmonary metastatic lesions were harvested and re-implanted into nude mice for the second round of in vivo selection. The same procedure was repeated twice. LM3 was established from the third round with high metastatic potential to lung. Cell lines were cultured at 37˚C in 5% CO2 in high-glucose DMEM (Hyclone, USA) supplemented with 10% fetal bovine serum (Hyclone, USA), and collected under standard tissue culture protocols.

**Chromatin Immunoprecipitation Sequencing (ChIP-seq)**

ChIP was carried out in 97L and LM3 cell lines according to the standard protocol. Briefly, approximate 300μg fixed chromatin and 5μg antibody, including H3K4me3 (Abcam, ab8580), H3K27ac (Abcam, ab4729) and H3K36me3 (Abcam, ab9050), were used for each ChIP pull-down. ChIP library preparation and sequencing procedures were carried out according to Illumina protocols (Illumina, San Diego, CA). Liver ChIP-seq data were downloaded from GEO database, including H3K4me3 (GSM537697, GSM537709 and GSM621675), H3K27ac (GSM1112809) and H3K36me3 (GSM537699 and GSM537708).

**RNA-seq**

We extracted approximate 6 micrograms of total RNA and removed ribosomal RNA (rRNA) according to the manufacturer’s instructions with minor modifications (Invitrogen). The remaining RNA was processed for sequencing library preparation with RNA-seq Library Preparation Kit for Whole Transcriptome Discovery (Huijiebio, China). Briefly, the RNA was incubated with fragmentation buffer for 5 min at 95 ˚C, then immediately cooled down on ice and ligated a 5’ single-strand linker, the cDNA synthesis was performed followed by PCR purification and size selection (400-500bp). The obtained library was sequenced with pair-end model (2×100bp) by Hi Seq 2500 (Illumina, San Diego, CA).

**Bisulfite pyrosequencing**

The nested-PCR was performed to produce the single band PCR products for bisulfite pyrosequencing. The first round PCR was generated in a 10μL reaction volume with 5μL 2×Taq PCR Master Mix (LifeFeng, Cat No.PT102), 100 nM of forward primer and 100nM reverse primer which tagged by a sequence recognized by the universal primer, and 1μL Template. The second round PCR was generated in a 30μL reaction volume with 15μL 2×Taq PCR Master Mix, 200 nM of forward primer and 200 uM universal biotin reverse primer, and 2μL product from first round PCR as template. The two rounds PCR conditions were 98°C for 30 seconds, followed by 30 cycles of denaturation at 98°C for 10 seconds, annealing at 58°C for 30 seconds, and elongation at 72°C for 30 seconds, and finished with 1 cycle of final elongation at 72°C for 5 minutes.

PCR products (20μL) were added to a mix consisting of 3μL Streptavidin Sepharose HP (GE Healthcare, Germany, Cat No. 17-5113-01), 40μL binding buffer (Qiagen, Hilden, Germany, 979006) and 17μL ddH2O, then mixed at 1400 rpm for 10 minutes at room temperature. Using the Vacuum Prep Tool (Qiagen, Hilden, Germany, 972804), single-stranded PCR products were prepared following the manufacturer’s instructions. The sepharose beads with the single stranded templates attached were released into a PSQ 96 Plate Low (Qiagen, Hilden, Germany) containing a mix of 40μL annealing buffer (Qiagen, Hilden, Germany) with 400 nM of the corresponding sequencing primer. Pyrosequencing reactions were performed in a PyroMark ID System (Qiagen, Hilden, Germany) according to the manufacturer’s instructions using the PyroMark Gold 96 Reagent Kit (Qiagen, Hilden, Germany). CpG site quantification was performed using the methylation Software Pyro Q-CpG.

**Bioinformatic analysis:**

**ChIP-seq analysis**

We firstly filtered out low quality reads with NGS QC Toolkit v2.3.3 (Patel and Jain 2012). The remaining reads were aligned to human reference genome (hg19) with Bowtie (version 1.0.0) (Langmead and Salzberg 2012). The human reference genome (hg19) was downloaded from the UCSC website (http://www.genome.ucsc.edu). The alignment was unique mapping and allowed up to 2 mismatches. Peaks were identified using the MACS (version1.4) (Zhang et al. 2008) with “-nomodel” and other default parameters. For ChIP-seq comparison across samples, ChIP-seq signal was normalized according to the number of mapped reads of samples. Histone modification density in promoters or genebodies was calculated as ChIP-seq signal summation divided by promoter/genebody length. The promoter was defined as ±1kb around transcript start site. In correlation analysis between ChIP-seq signal and expression level, the curve in the scatter plot was based on a generalized additive model in R and fitted by the method of cubic smoothing.

**RNA-seq analysis**

We firstly filtered out low quality reads with Trim Galore (v0.4.0, http://www.bioinformatics.babraham.ac.uk/projects/trim\_galore/). Then mRNA-seq reads were aligned to hg19 using TopHat2 (v2.0.12) (Kim et al. 2013) with default parameters. The mapped reads were further used for identifying gene FPKM (Fragments Per Kilobase of exon per Million fragments mapped) using Cufflinks (v2.2.1) (Trapnell et al. 2012) with default parameters. The RefSeq genes (hg19) were used as gene annotation and obtained from the UCSC website. The gene FPKMs in the four samples, including liver tissue, primary liver cancer, 97L and LM3, were normalized with normalize.quantiles and preprocessCore packages in the Bioconductor website (<https://www.bioconductor.org/)>. As for up- or down-regulated genes, a minimal of two-fold change in expression level (FPKM) was required and genes with low expression (FPKM<1) in both samples were removed.

**Simulated data generation**

Simulated pair-end reads were randomly selected from hg19 genome. Read1 and Read2 were of 100bp length. The length of the interval between a read pair was normally modeled with mean of 400bp and standard deviation of 50bp. The probability of methylated C in CpG was set at 70% while the probability of methylated C in CH was 1%. The location information was remained and used for verification of mapping accuracy.

**Comparison of GPS and WGBS**

Mapped percentage of GPS and WGBS datasets was defined as percentage of paired mapped reads out of total reads.

To compare GPS with WGBS, we normalized the GPS-detected CpG number to 100%, and compared the WGBS with the normalized GPS in the same regions. CpG sites were covered by at least 1 read except for repetitive elements whose CpG sites required at least 5x coverage for reliable detection due to high frequency of repetitive elements.

Pie graphs were used to show the distributions of CpG sites detected by WGBS or GPS over whole-genome according to the annotation of RefSeq genes (hg19). CpG sites were covered by at least five reads.

**Bisulfite pyrosequencing verification**

Firstly, CpG sites (≥5x coverage) with consistency (methylation difference ≤10%) between two replicates were selected in GPS and WGBS datasets, respectively. Secondly, among these selected CpG sites, we further picked up those detected by both GPS and WGBS commonly, ranked the common CpG sites according to the difference of methylation signal between GPS and WGBS from high to low, and excluded those sites where PCR primers cannot be designed. Finally, the top 10 sites represent those CpGs where methylation was the most different detected by WGBS and GPS. We used bisulfite pyrosequencing to verify methylation levels in these 10 sites with around regions.

**Methylation smoothing**

We carried out CpG methylation smoothing by BSmooth (Hansen et al. 2012) of R package bsseq (v1.10.0) from the source of Bioconductor with parameter alpha = 0.95. After smoothing, a sliding window was built with step length 50bp to generate methylation level for every fragment in genome by averaged the methylation level of neighboring CpGs. If the distance between two neighboring CpGs was exceeding 10kb, the region would not be assigned the methylation level.

**Variation calling**

Adjusted Read2 were used for variation calling. We firstly trimmed low quality reads and adapters using the Trim Galore (v0.4.0) and aligned onto reference genome (hg19) by Bowtie 2 with default parameters. The commands of samtools rmdup and mpileup were used for removing duplicated reads and generating pileupped files. We identified variations from the pileupped file using VarScan (v2.3.9) (Koboldt et al. 2009) with default parameters except “mpileup2snp --output-vcf 1 --min-coverage 10 --min-reads 25 --p-value 0.01”.

**Allele-specific methylation**

In order to analyze allele-specific methylation, the fragments coming from different allele need to be separated. We firstly selected the heterozygous variations with the cutoff of allele frequency within 40%-60% and ADP (Average per-sample depth of bases with Phred score ≥ 15) ≥30, and then removed those variations that located in repetitive elements. Secondly, for each heterozygous variation, Read2 and corresponding Read1 were separated into two groups, for example, allele A and allele B. Finally, the methylation difference between allele A and allele B was figured up by GPS analysis. We defined allele-specific methylation as alleles whose methylation difference exceeded the threshold of 50% in at least three CpG loci.

**Methylation Boundary Shift (MBS) analysis**

For each gene with length more than 5kb, the smoothed methylation value was used for defining the hypomethylated region boundary of every gene within ±5kb around Transcription Start Site (TSS), 10kb totally. TSS and genebody location were obtained from RefSeq gene (hg19). At first, the 10kb fragment was segmented into 100 bins and averaged methylation values were calculated for each bin, respectively. Then we compared each bin value with the averaged methylation level of the whole 10kb, from TSS to upstream direction and to downstream direction respectively. Finally, Methylation Boundary was defined as the distance to TSS where the methylation level of corresponding part exceeded averaged methylation level.

**Bar and box plot**

Bar plot showed the mean with standard error of the mean (SEM) as error bars, while box plot showed the median, 25th and 75th percentiles and whiskers showing min to max.

References and Notes:

Hansen KD, Langmead B, Irizarry RA. 2012. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* **13**: R83.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.

Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**: 2283-2285.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* **7**: e30619.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**: 562-578.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.