### Supplementary Notes

#### Details for processing of human normal tissues

Ten human primary normal tissues were purchased from BioChain. Approximately 200 ng of genomic DNA from ten human primary tissues in the volume of 50 µL was fragmented into an average size of 400 bp in a Covaris micro TUBE with Covaris E210 ultrasonicator. Fragmented genomic DNA was converted into Illumina paired-end sequencing libraries using KAPA Library Preparation kit (KAPA Biosystems) following manufacturer’s instruction with modifications. After end-repair and dA-tailing, ligation with methylated adapters was performed at 20 ˚C for 15 min in the presence of 10-fold molar excess of Illumina methylated adapters (Illumina). The ligation mixture was purified with an equal volume of Agencourt AMPure XP beads (Beckman Coulter) and eluted with 23 µL of 10mM Tris-HCl, pH8.5. Next, 20 µL of adaptor ligated DNA was bisulfite converted using EZ DNA Methylation-Lightning kit (Zymo Research) following manufacturer’s protocol and eluted with 30 µL of 10mM Tris-HCl, pH8.5. Bisulfite converted DNAs were amplified using iQ SYBR Green Supermix (Bio-Rad) with 200 nM each of PCR primer PE1.0 and multiplexing PCR primer for 10 cycles in 100 µL total volume. PCR products were purified with 0.8X volume of Agencourt AMPure XP beads (Beckman Coulter) and eluted with 50 µL of 10mM Tris-HCl, pH8.5, pooled in equimolar ratios, and size selected using 6% TBE gels for 400-600 bp. The concentration of sequencing libraries was quantified by qPCR using KAPA Library Quantification kit (KAPA Biosystems). Libraries were sequenced on HiSeq2500 for PE 100 cycles.

#### Details for processing of patient tumor tissues.

Cancer tissue and plasma samples were collected from UCSD Moores Cancer Center. Clinical information, gender, age and TNM staging, on the patients was limited because the samples were de-identified. Informed consent was obtained from all subjects. All the samples are diagnosis to corresponding cancers according to the World Health Organization classification criteria[1](#_ENREF_1). 88.4% samples were derived from Caucasian population while 6.8% and 3.3% samples were from Asian and African population (detail see Supplementary Table 12). Genomic DNAs were extracted from 20-50 mg of primary tumor tissues from lung, colon and pancreatic cancer patients using DNeasy Blood and Tissue kit (QIAGEN) following the manufacturer’s instruction and eluted in 400 µL of AE buffer (QIAGEN). The concentration and quality of genomic DNA were assessed by Qubit dsDNA HS Assay kit (Life Technologies) and NanoDrop (Thermo Scientific), respectively. To generate RRBS sequencing libraries, 100 ng of gDNA were digested with 20 U of *Msp*I (Thermoscientific) in 1X Tango buffer (Thermoscientific) and 1 ng of unmethylated lambda DNA (Promega) in order to assess for bisulfite conversion rate in 30µL total volume for 3 h at 37 ˚C and heat inactivated at 65 ˚C for 20 min. Next, 5U of Klenow fragment, exo- (Thermoscientifc) and a mixture of dATP, dGTP, and dCTP (New England Biolabs) were added to *Msp*I-digested DNAs for a final concentration of 1 mM, 0.1 mM, and 0.1 mM for dATP, dGTP, and dCTP, respectively in 32 µL for end-repair and dA-tailing. The mixture was mixed and incubated at 30 ˚C for 20 min, 37 ˚C for 20 min, and heat inactivated at 75 ˚C for 10 min. dA-tailed DNA was purified with 2X volume of Agencourt AMPure XP beads (Beckman Coulter) and resuspended dA-tailed DNA with 20 µL nuclease-free water without discarding the magnetic beads. dA-tailed DNAs were then ligated to methylated adaptors in 30 µL total volume containing 30 U of T4 DNA ligase, HC (Thermoscientific), 1X Ligation buffer (Thermoscientific), and 500 nM individual TruSeq multiplexing methylated adaptors (Illumina). The ligation mixture was mixed well and incubated at 16 ˚C for 20 h, heat inactivated at 65 ˚C for 20 min, purified by adding 60 µL of PEG 8000/5M NaCl buffer (Teknova) to adaptor ligated DNA and bead mixture, and eluted in 20 µL of nuclease-free water. Next, the adaptor ligated DNA were bisulfite converted using the MethylCode Bisulfite Conversion kit (Life Technologies) following manufacturer’s protocol and eluted in 35 µL of Elution buffer (Life Technologies). Bisulfite treated DNAs were amplified using 5 U of PfuTurboCX (Agilent Technologies) and 300 nM each of TruS\_F and TruS\_R primers for 14 cycles in 100 µL total volume. PCR products were purified with an equal volume of Agencourt AMPure XP beads (Beckman Coulter) and eluted with 50 µL of 10mM Tris-HCl, pH8.5, pooled in equimolar ratios, and size selected using 6% TBE gels for 150-400 bp. The concentration of sequencing libraries was quantified by qPCR using KAPA Library Quantification kit (KAPA Biosystems). Libraries were sequenced on Illumina HiSeq2500 for PE 100 cycles.

#### Details for processing of plasma samples

Normal plasma samples were obtained from UCSD Shirley Eye center. Information such as gender and age was limited because the samples were de-identified. Informed consent was obtained from all subjects. Plasma samples from patients were processed using the QIAamp Circulating Nucleic Acid Kit (Qiagen) to extract circulating DNA. The DNA extracted from plasma were then concentrated using ethanol precipitation and eluted in 15 uL nuclease-free water. Next, 1-10 ng of DNA were digested with 10 U of *Msp*I (Thermoscientific), 1X Tango buffer (Thermoscientific), and 10 pg of unmethylated lambda DNA (New England Biolabs) as control for ~13 h at 37 ˚C, then heat inactivated at 65 ˚C for 20 min. Next, 5 U of Klenow fragment, exo- (Thermoscientifc) and a mixture of dATP, dGTP, and dCTP (New England Biolabs) were added for a final concentration of 1 mM, 0.1 mM, and 0.1 mM for dATP, dGTP, and dCTP respectively. The mixture was gently vortexed, and incubated at 30 ˚C for 20 min, 37 ˚C for 20 min, and finally 75 ˚C for 10 min. To perform adaptor ligation, the dA-tailed DNA were added to a 5 uL mixture of 1X Tango buffer, 30 U of T4 DNA Ligase, HC (Thermoscientific), 2.5 mM ATP, and 500 nM individual TruSeq multiplexing methylated adaptors. The combined mixture was gently vortexed, incubated at 16 ˚C for ~20 h, then heat inactivated at 65 ˚C for 20 min. The ligation mixture was purified using Agencourt AMPure XP beads (Beckman Coulter), and eluted in 20 uL of nuclease-free water. The ligated products were then bisulfite converted using the MethylCode Bisulfite Conversion kit (Life Technologies). Two rounds of amplification were performed after bisulfite conversion. The first round was using PfuTurboCX (Agilent Technologies) for 12 cycles in 50 uL total volume, then the second round was performed using Phusion HotStart Flex (New England Biolabs) master mix for 9 cycles in 50 uL total volume. Final PCR products were purified, pooled in equimolar ratios, and size selected using polyacrylamide gels for 150-400 bp. Libraries were sequenced on both Illumina MiSeq and HiSeq2500 for PE 100 cycles.

#### Methylation haplotype analyses.

1. We first partitioned the human genome into non-overlapping “sequenceable and mappable” segments using a set of in-house generated WGBS data from 10 tissues from a 25-yr adult male individual (5x mappable genome coverage per tissue, 50x for 10 tissues combined). A total of 1,072,789 autosomal segments (minimal size: 80bp; average size: 2.35Kb; total size: 2.52Gb) that have a minimal read depth of 10x were identified.

bedtools genomecov -bg -split -ibam N37\_10\_tissue\_pool\_chrXX.bam > N37\_10\_tissue\_pooled.chrXX.genomecov.bed

awk '$4>9 {print $1"\t"$2"\t"$3} N37\_10\_tissue\_pooled.chrXX.genomecov.bed | bedtools merge -d 10 -i - > N37\_10\_tissue\_pooled.chrXX.RD10.genomecov.bed

awk '$3-$2>80 {print $1"\t"$2"\t"$3"\t"$3-$2+1}' N37\_10\_tissue\_pooled.chrXX.RD10.genomecov.bed > N37\_10\_tissue\_pooled.chrXX.RD10\_80up.genomecov.bed

1. Mapped reads from WGBS data sets were converted into methylation haplotypes in each segment. Calculation of methylation linkage disequilibrium (the r2 statistics) was performed on the combined methylation haplotypes from all the five data sets. A binary partitioning strategy was used to split each segment into methylation haplotype blocks (MHBs). We define a methylation haplotype block as a genomic region in which the r2 value of two adjacent CpG sites is no less than a threshold (r2 >= 0.5). At this threshold, 50% of the variance of a CpG methylation status can be predicted by the status of an adjacent site.

mergedBam2hapInfo.pl N37\_10\_tissue\_pooled.chrXX.RD10\_80up.genomecov.bed N37\_10\_tissue\_pool\_chrXX.bam > N37\_10\_tissue\_pool\_chr1.RD10\_80up\_bin.hapInfo.txt

cat \*.chrXX.RD10\_80up\_bin.hapInfo.txt | /mergeHapInfo.pl > WGBS\_pooled\_mappable\_bins.chrXX.hapInfo.txt

hapInfo2mld\_block.pl WGBS\_pooled\_mappable\_bins.chrXX.hapInfo.txt 0.5 > WGBS\_pooled\_mappable\_bins.mld\_blocks\_r2-0.5.bed

1. After MHBs were defined, methylation haplotypes for each MHB were extracted from the bam file, and the methylation haplotype load (MHL) for each MHB was calculated.

#Iterate through all the sample and chromosome combinations

mergedBam2hapInfo.pl WGBS\_pooled\_mappable\_bins.mld\_blocks\_r2-0.5.bed SampleID\_chrXX.bam > SampleID\_chrXX.hapInfo.txt

#Place all hapInfo.txt files for one data set in one folder, calculate MHL and report the values for all samples at all MHBs in one matrix

get\_methHapLoad\_matrix.pl hapInfo\_data\_set\_folder > Data\_set\_name\_mhl\_matrix.txt

merge\_mhl\_matrix.pl Data\_set\_A\_mhl\_matrix.txt Data\_set\_B\_mhl\_matrix.txt Data\_set\_C\_mhl\_matrix.txt > All\_data\_sets\_matrix.txt