Supplementary Methods

1. MSCC library construction

The two custom adaptors (A and B) that contained a 5’ CG overhang and 3’ NN overhang, respectively, were created by Sangon Biotech (Shanghai, China). These two adaptors also contained end sequences required for Illumina library construction and sequencing. Standard DNA was combined with 2 μg of genomic DNA and digested with HpaII (New England Biolabs [NEB]) for the HpaII library. After that, Adaptor A was ligated to the resulting fragments. The reaction products were incubated with Bst DNA polymerase (NEB) and then digested with MmeI (NEB). After incubation, the reaction mixture with Adaptor B and Agencount AMPure XP Beads (Beckman) were used to purify the products. A 2% E-Gel® EX Gel (Invitrogen) was used to isolate the target band at 140 bp from the DNA mixture. The target DNA was then amplified by an eight-cycle polymerase chain reaction (PCR) reaction. For the inverse library, antarctic phosphatase was used to deactivate the fragmented ends (NEB). The products were digested with MspI and then treated with the same procedure as the HpaII library.

2. MSCC data analysis

The output sequencing data were then analyzed by the following procedures: The .bcl files generated from the sequencing system were first transformed into a fastq file using CASAVA software. Maximum oligonucleotide mapping (MOM) software was then used for mapping the sequencing reads to the dataset of all possible 18-bp tags in the human genome (hg19)(1). After normalizing the number of reads in both libraries by the counts of standard DNA, the methylation level of each site was calculated.

3. Sequenom EpiTYPER assay

Genomic DNA was first incubated with bisulfite (unmethylated cytosine was converted to uracil), and then PCR amplification was performed using a 5 μl reaction mixture with the following conditions: 95°C for 4 min, 35 cycles of 95°C for 25 s, 58°C for 25 s and 72°C for 70 s, and 72°C for 7 min followed by a 4°C

hold. Next, the reactions were performed followed by Shrimp Alkaline Phosphatase (SAP) purification and T cleavage. Clean Resin (6 mg; Sequenom, San Diego, CA) and H2O (20 μl) were used to remove bilvalent cation adducts from the T cleavage transcription products. The samples were then sequenced on a MassARRAY analyzer (Sequenom, San Diego, CA). Primers of the two genes were as follows: *GABBR2*-F (aggaagagagTAGTTTGTGATAGGAAAGGATTTGG), *GABBR2*-R (cagtaatacgactcactatagggagaaggctACAACCAACTTCATACACCTACCAT), *CBFA2T3*-F (aggaagagagGGGGGTAGTTTTTGTTTTTAATTTG) and *CBFA2T3*-R (cagtaatacgactcactatagggagaaggctCTAACCCTCCTAATTTCCTCATTTA).

References

1. Eaves HL, Gao Y. MOM: maximum oligonucleotide mapping. Bioinformatics **2009**;25:969-70.

Figure legends

Supplementary Figure S1. The results proved that the whole-genome DNA methylation sequencing results were accurate and reliable from different aspects.

(a) The genomic distribution (according to the relative positions compared to genes) of CCGG sites with high sequencing depth of 40+ reads was very similar to that of all the CCGG sites in paired lung cancer tissues, showing that the CCGG data with high sequencing depth of 40+ reads could well reflect the data of all the CCGG sites.

(b) When mapping of MSCC sites to their relative location to associated genes, the moving average methylation levels of each region of genes indicate that the transcription starting site (TSS) region was hypomethylated compared to other relative locations of genes, including promotor region, 50% gene body, transcription end site (TES), and terminator region.

(c) The CCGG sites with 40+ reads were mapped to CGIs and CGI shores. The methylation level of CCGG sites inside CpG islands were significantly lower than those of outside CpG islands.