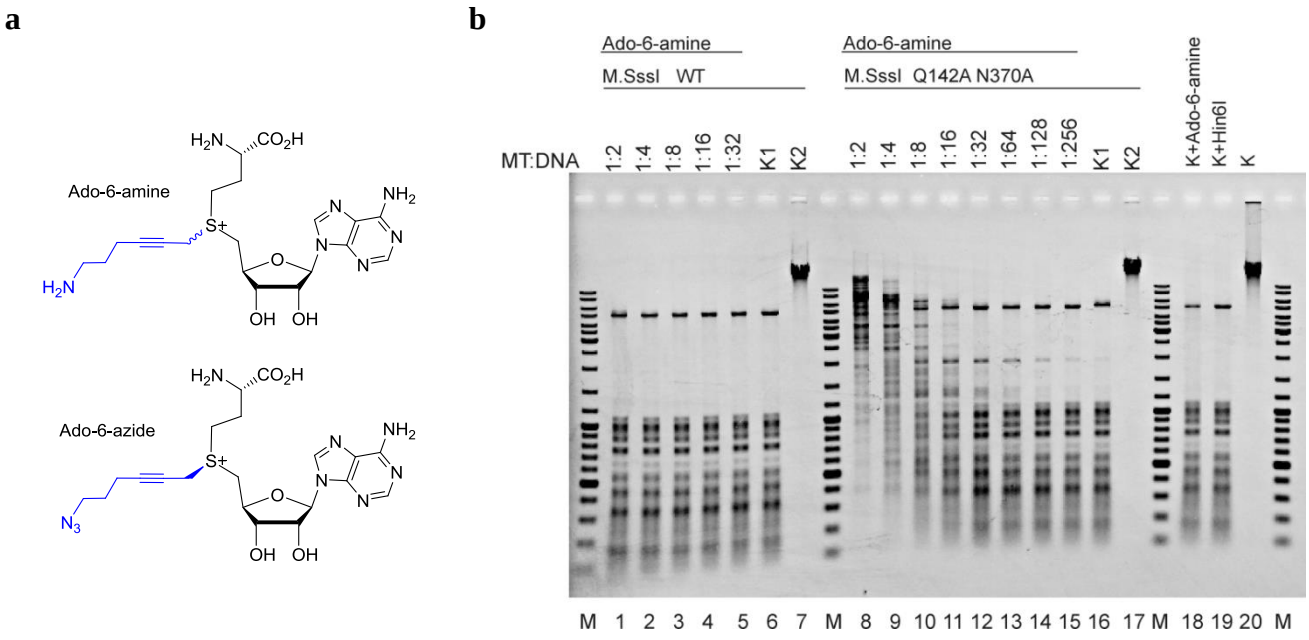
SUPPLEMENTARY INFORMATION

DNA unmethylome profiling by covalent capture of CpG sites

E. Kriukienė, V. Labrie, T. Khare, G. Urbanavičiūtė, A. Lapinaitė, K. Koncevičius, D. Li, T. Wang, S. Pai, C. Ptak, J. Gordevičius, S.-C. Wang, A. Petronis and S. Klimašauskas

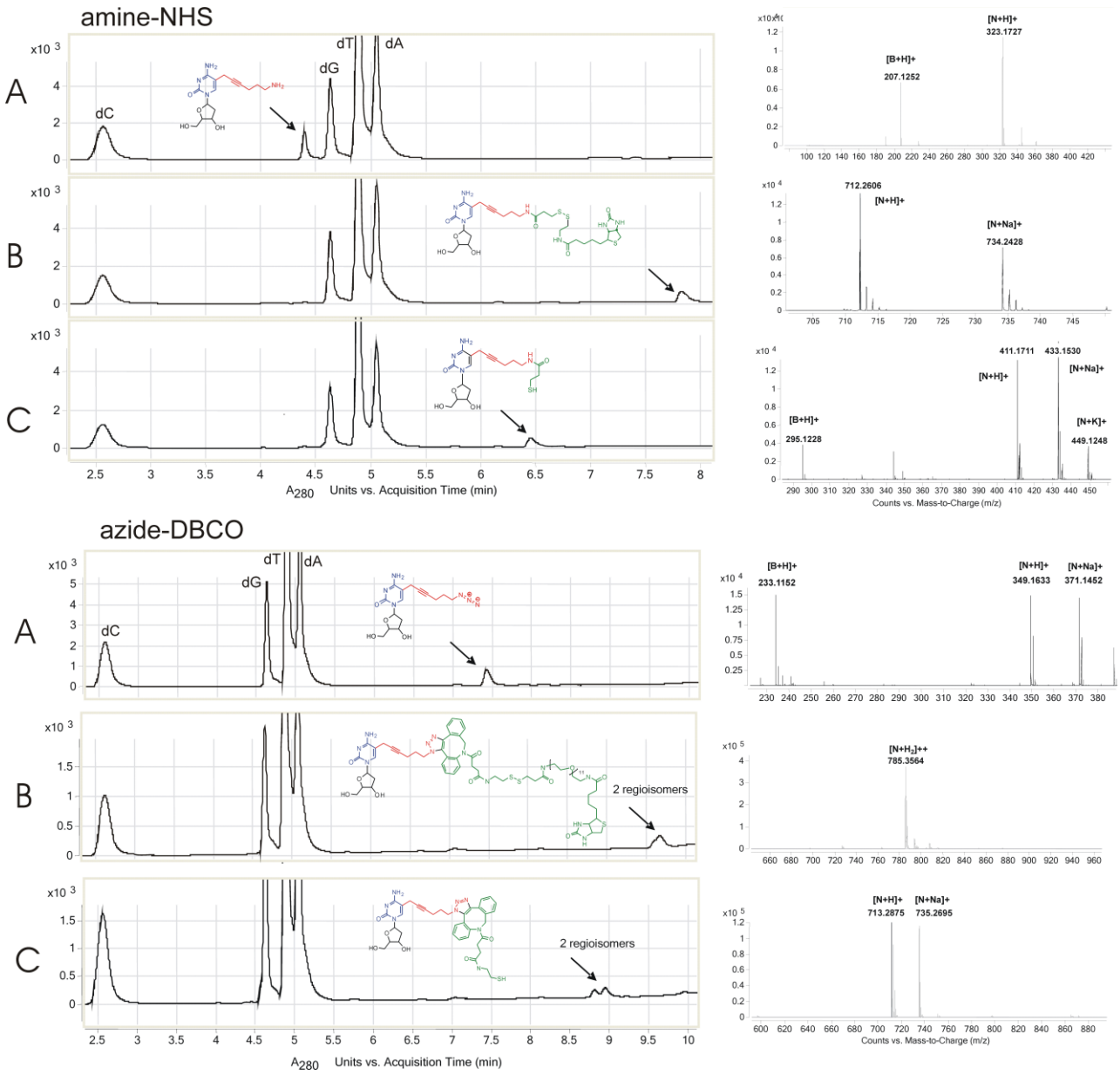
Correspondence to: SK (saulius.klimasauskas@bti.vu.lt) or AP (art.petronis@camh.ca)

Supplementary Figures



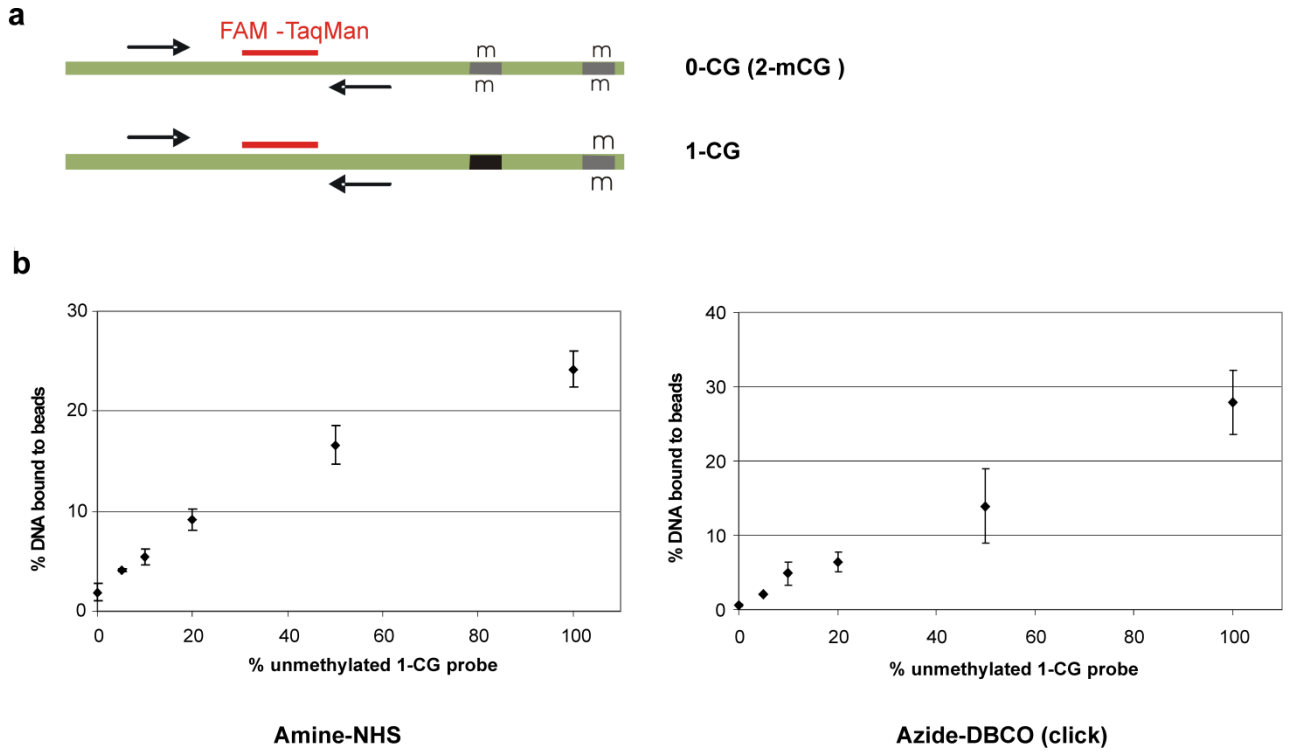
Supplementary Figure S1. Methyltransferase-directed transfer of activated groups catalyzed by the engineered M.SssI MTase

(**a**) Chemical structures of the cofactor analogs Ado-6-amine and Ado-6-azide; (**b**) Enhanced aminoalkylation activity of the engineered M.SssI MTase The DNA methyltransferase activity with Ado-6-amine cofactors was estimated using a DNA protection assay [22]. Modification reactions of lambda DNA were performed with sequential two-fold dilutions (shown as molar ratios of Mtase to DNA target sites) of WT and engineered (variant Q142A/N370A) M.SssI at 37°C for 2 h. Modified DNA was digested with R.Hin6I and analyzed by agarose gel electrophoresis. A DNA protection level observed with the highest concentration (1:2) of the WT M.SssI is weaker than that observed with the most diluted sample (1:256) of the engineered enzyme, indicating that the latter shows at least 100-fold higher transferase activity with the Ado-6-amine cofactor. M – GeneRuler DNA Ladder Mix (ThermoFisher); K1 and K2, controls with no added cofactor and no R.Hin6I (K2).



Supplementary Figure S2. HPLC-MS analysis of a modified cytosine residue formed in DNA upon action of the engineered M.SssI and cofactor Ado-6-amine.

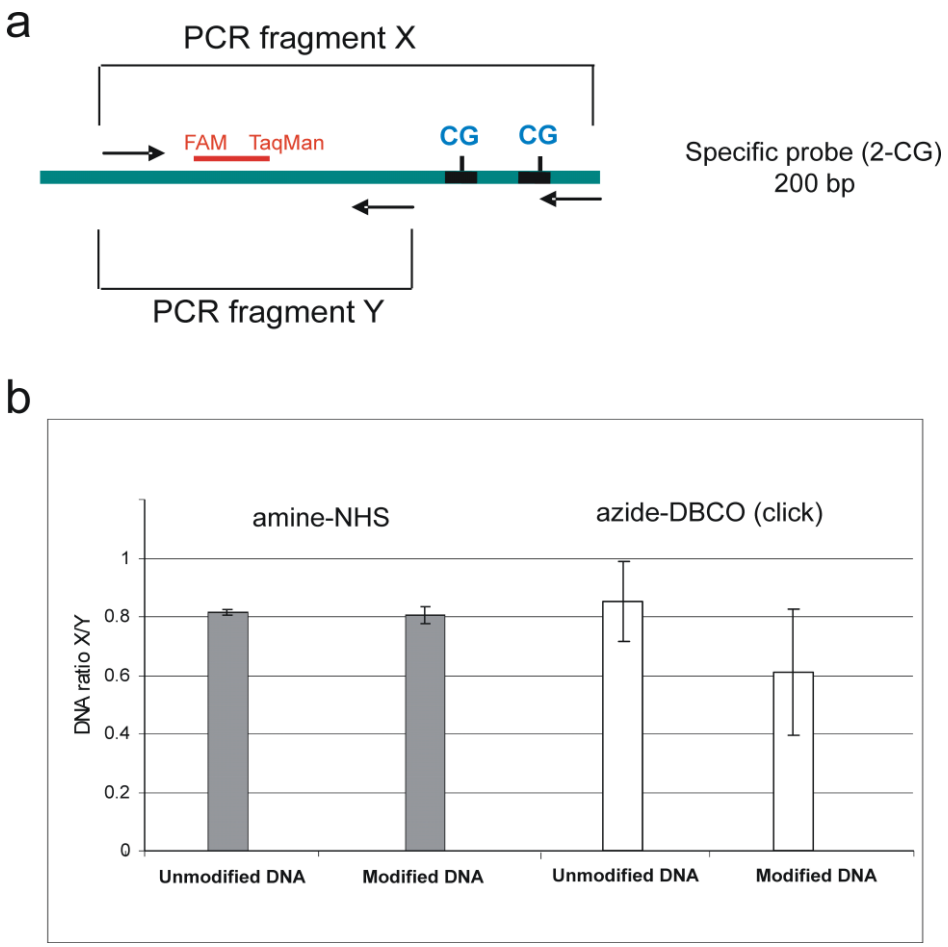
24-mer oligonucleotide duplex (5 μM) containing a single CG site (5'-taataatgcgctaataataataat/ 5'-attattattattagcgcattatta) was treated with 5 μM eM.SssI in the presence of 400 μM Ado-6-amine or 200 μM Ado-6-azide for 1 hour at 37 °C (sample A). The DNA was biotinylated with 8.6 mM biotin-SS-NHS or 5 mM of dibenzylcyclooctyne-S-S-PEG11-Biotin, respectively (sample B). Purified DNA was then incubated with 100 mM DTT at RT for 1 hour (sample C). All samples were digested to nucleosides by incubation with a mixture of nuclease P1 and FastAP alkaline phosphatase. Samples were analyzed on an integrated HPLC/ESI-MS system (Agilent 1290 Infinity) equipped with a Discovery C18 column (75 x 2.1 mm) by elution with a linear gradient of solvents A (20 mM ammonium formate, pH 3.5) and B (80% aqueous methanol) at a flow of 0.3 ml/min at 30 °C as follows: 0–2 min, 0% A; 2–10 min, 0–100% B. High-resolution mass spectra of modification products were acquired on an Agilent Q-TOF 6520 mass analyzer (100–2500 m/z range, positive ionization mode). UV chromatograms (left) show efficient conversions of mTAG-derivatized DNA at biotinylation (B) and SS-cleavage (C) steps. MS data (right panel; N, nucleoside; B, nucleobase) confirm the formation of the expected modified nucleoside products.



Supplementary Figure S3. Streptavidin capture of unmodified CpG sites in model DNA fragments.

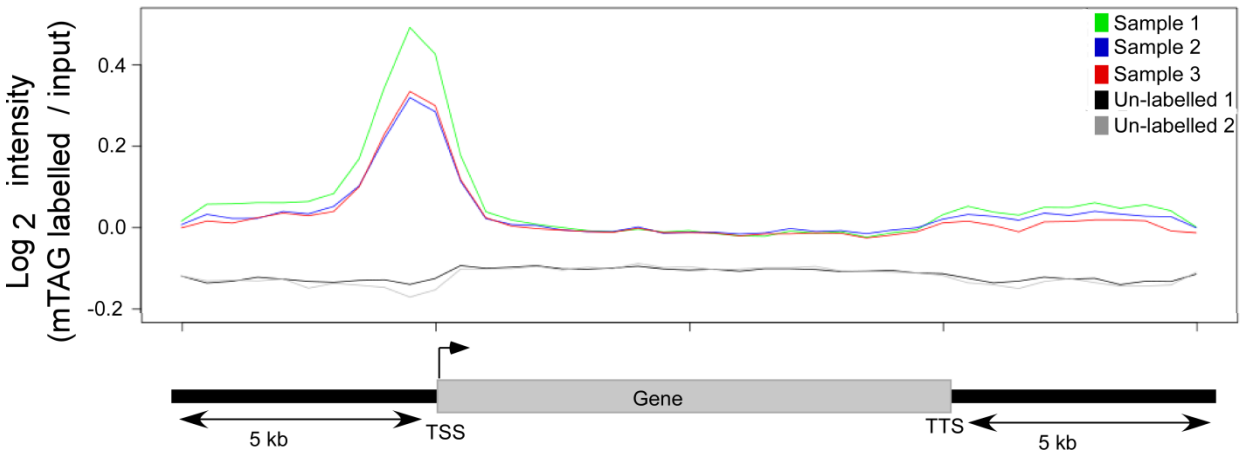
(**a**) Schematic view of a fully methylated (2-mCG) and specific (1-CG) DNA probes derived from mouse genome for quantization of DNA using TaqMan qPCR. CpG sites are shown as darker sections and qPCR primers are shown as arrows. "m" denotes a premethylated CpG site.

(**b**) Capture of unmethylated DNA probe in the presence of methylated DNA fragment on streptavidine beads. The specific DNA probe (1-CG) was combined with the pre-methylated probe (2-mCG) at specified ratios in the presence of 300 ng sheared genomic DNA. Samples were mTAG-labeled at Medium Intensity using the conventional (left) or click (right) chemistries and processed as described in Methods. The amount of captured DNA was determined by on-beads qPCR analysis. Error bars defined as ±s.d. from duplicate experiments.



Supplementary Figure S4. qPCR analysis of DNA amplification efficiency through mTAG modified CpG sites.

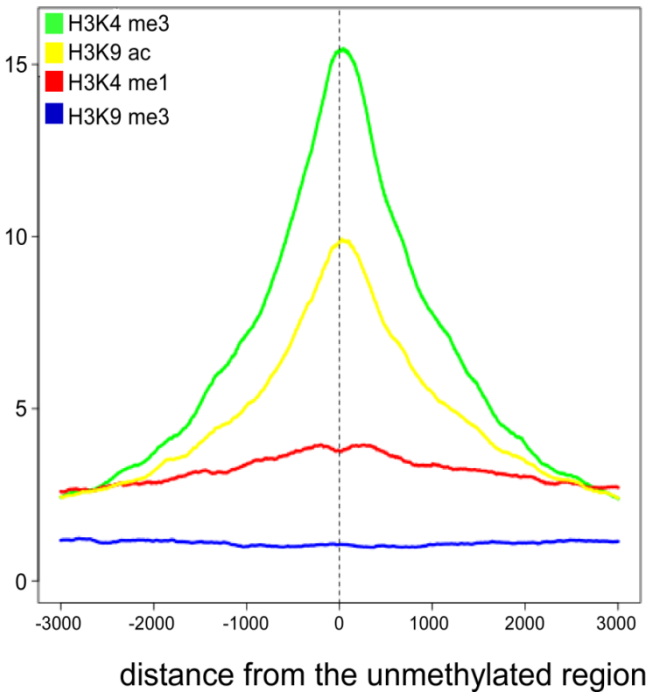
The specific 2-CG probe (**a**) was mTAG-labeled at CpG sites (High labeling intensity), captured on streptavidin beads and recovered after cleavage of the biotin linker with DTT. In the released DNA, CG sites retain a fraction of the original biotin linking chains (see Figure 1b). (**b**) Unmodified and modified DNA probes as shown were amplified (10 cycles, Taq polymerase in 40 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol) with 2 sets of primers to give fragments X and Y, and their ratio was determined using qPCR. Error bars defined as ±s.d. from triplicate experiments. .



Supplementary Figure S5. mTAG-chip DNA unmethylation profile of a composite gene.

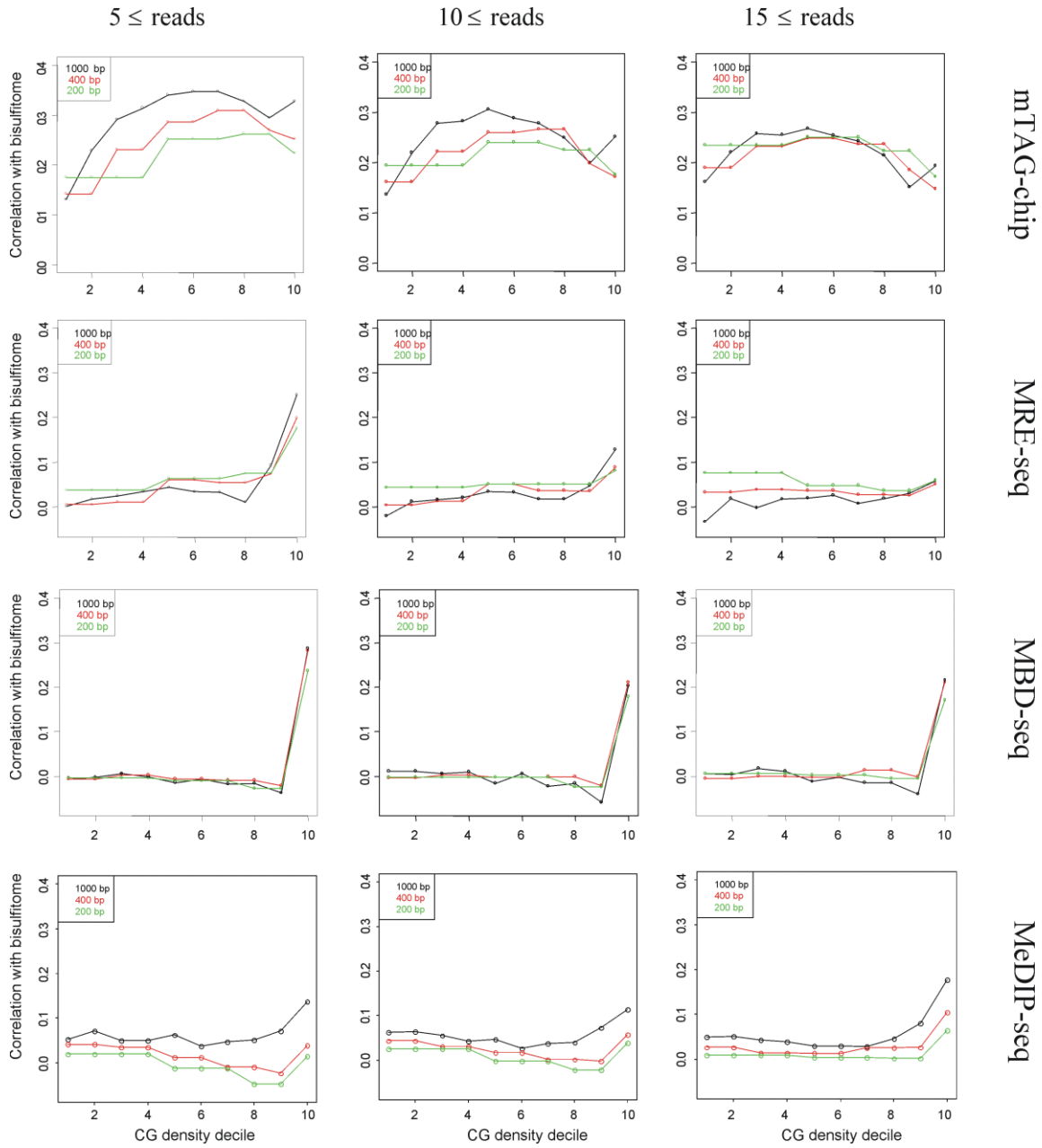
The mTAG-chip mean intensity ratios were plotted across all known genes (*Refgene*; UCSC *hg18*) on human chromosome 1, with data binned in 20 parts of each gene. In addition, 5 kb region upstream and downstream from the gene were each binned in 10 parts. Shown are the plots of mTAG labeled samples (n=3) and control unlabelled samples (n=2). The unmethylated peaks were observed at the promoter, around transcriptional start site (TSS) and to a lower scale at the downstream region after transcriptional termination site (TTS). Controlled unlabelled samples showed no enrichment at any fraction of the composite gene.

Pilot DNA unmethylome studies were performed with a human post mortem brain (prefrontal cortex) genomic DNA of individuals not diagnosed with any mental illness (Stanley collection, CAMH, Toronto). The mTAG protocol using amine-NHS chemistry at Medium labeling intensity was performed as described in Methods. The unmethylated DNA fraction was interrogated on Affymetrix GeneChip Human Tiling 2.0R (B arrays, chromosome 2, 9 and 19). Tiling array preprocessing and analysis was done as described in Methods, except that instead of quantile normalization the arrays were log2 transformed, sliding-window smoothed and log2 ratio was taken. Log ratios of the labeled to non-labeled were obtained for each pair of labeled and non-labeled chips. The measurement was duplicated and log ratios were averaged to represent the enrichment of labeled relative to non-labeled.



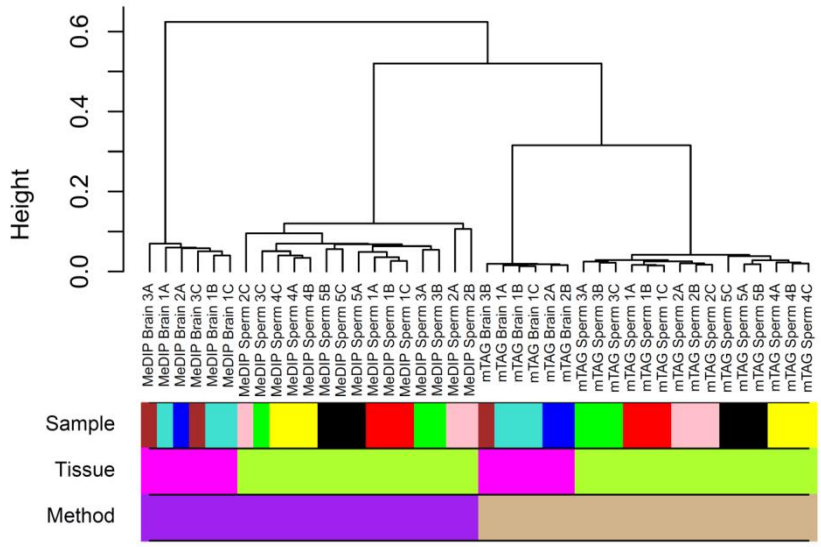
Supplementary Figure S6. Histone lysine modifications in mTAG-chip derived unmethylated genomic regions of the human brain.

CMARRT algorithm was used to identify unmethylated peaks on the normalized ratios, and these peaks were then extended to ± 3kb from the midpoint. Histone modification track measures (UCSC *hg19*; human Brain, Mid Frontal Lobe; Epigenome Roadmap: http://www.roadmapepigenomics.org/) were plotted on these regions; shown are H3 lysine 9 acetylation (H3K9ac), H3 lysine 4 mono (H3K4me1) and tri (H3K4me3) methylation and H3 lysine 9 trimethylation (H3K9 me3) profiles. Histone modifications associated to active chromatin either at the promoter of a gene (H3K9ac and H3K4 me3) or at the enhancers (H3K4me1 and H3K9ac) showed prominent peaks on these unmethylated regions, while modification associated to inactive chromatin (H3K9me3) was not different from the extended neighboring 3 kb region. Histone track measures were from GSM669965 (H3K9 me3); GSM670015 (H3K4 me1), GSM670016 (H3K4 me3) and GSM670021 (H3K9 ac). Prior to plotting histone modification measures, the unmethylated extended regions were genome lift-over to *hg19*.



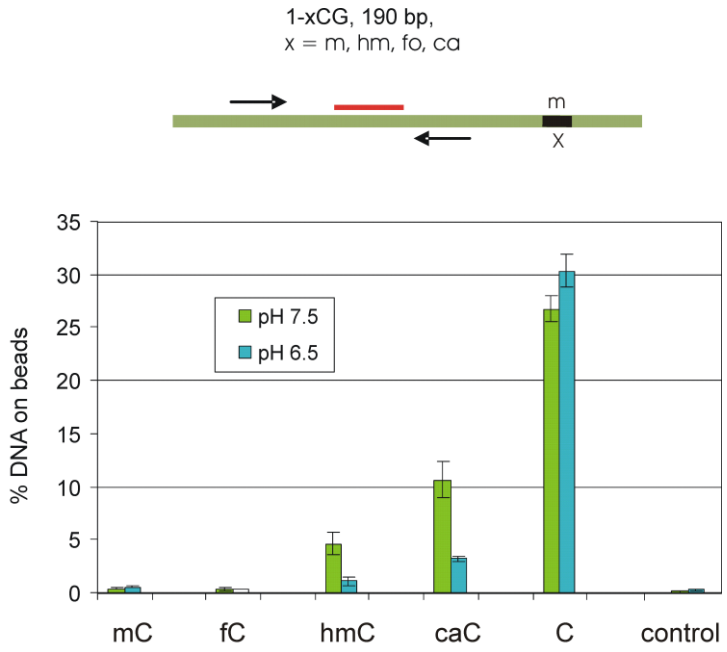
Supplementary Figure S7. Correlations of methylome profiling methods with MethylC-seq stratified according to sequencing depth and local CpG density.

Chromosome 4 was divided in tiles of 1 kb, 400 bp or 200 bp and mean log-ratios of the probes in the tiles are calculated. Pearson correlations with previously reported MethylC-seq data[3] (minimum 5, 10 or 15 reads) were determined. Results are stratified according to CpG density deciles. MethylC-seq data obtained with IMR90 cells were used as the reference methylome dataset for experimental mTAG-chip and MeDIP-chip, whereas H1 MethylC-seq data[3] were used for correlation analysis of published mDIP-seq, MRE-seq and MBD-seq data[8]. The genomic coordinates and Affymetrix probe annotations were relative to the human reference genome NCBI36 (hg18). Relevant genomic regions were divided into tiles as indicated, and tiles were stratified according to their CG densities. Mean log-ratios of probes for chip-data or mean numbers of reads for seq-data in the tiles were calculated and correlated to mean methylation scores of the MethylC-seq data. Missing-value tiles were excluded, and non-CG methylation sites in the H1 MethylC-seq data were removed prior to averaging and correlation with the MRE-seq and MBD-seq data.



Supplementary Figure S8. Comparison of mTAG-chip and MeDIP-chip datasets from human brain and sperm.

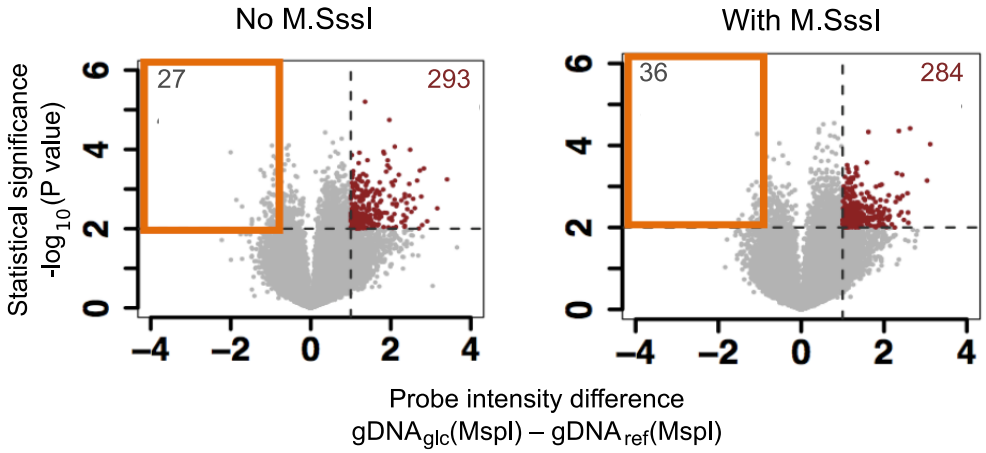
Clustering of technical replicates in mTAG-chip and MeDIP-chip datasets using gDNA from 3 brain and 5 sperm samples were enriched using mTAG and MeDIP protocol (3 replicates per sample and protocol) and analyzed on DNA microarrays (chr 10, 13, 14 and 17). Clustering of technical replicates in mTAG-chip and MeDIP-chip datasets was performed using 300,000 most variable probes.



Supplementary Figure S9. mTAG labeling of modified CpG sites in model DNA fragments.

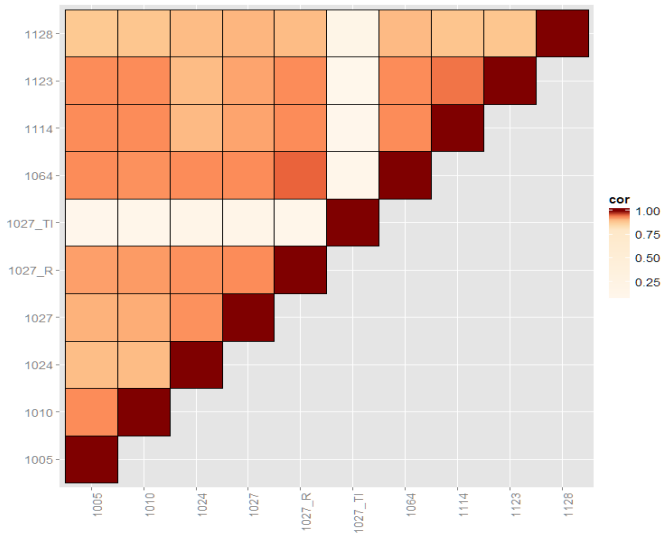
Top, schematic structure of a control DNA fragment (1-xCG) containing a single modified CpG site (x= C, mC, hmC, caC or fC; m= mC). DNA fragments were produced by PCR of mouse genomic DNA with Sp-PCR-dir primer (see Methods section) and corresponding reverse primer Mod-PCR-rev 5'-gcatcctggagattgtgggcaacatcXgg-3' (X=C, hmC, caC or fC). The upper strand of the modified target site or both strands in the unmodified substrate was methylated with M.HpaII and 300 µM SAM at 37°C for 1 hour and purified with QIAquick Nucleotide Removal kit. Control sample contained no M.SssI. qPCR primers are shown as arrows and a TaqMan probe is shown in red.

Bottom, streptavidin capture of modified CpG sites. mTAG labeling reactions were performed under standard conditions (pH 7.5) containing Ado-6-azide cofactor (25 µM), eM.SssI (180 nM) in buffer (10 mM Tris-HCl pH 7.6, 50 mM NaCl, 0.1 mg/ml) at 37°C for 30, followed by heating at 80°C for 10 min or under optimized mTAG conditions (pH 6.5) 100 µM of Ado-6-azide and 0.36 µM of eM.SssI in buffer (25 mM Tris-Bis-HCl pH 6.5, 50 mM NaCl, 0.5 mM EDTA, 0.1 mg/ml) at 37°C for 30 min. Biotinylation (DBCO click), enrichment and qPCR analysis were performed as described in Methods. Error bars defined as ±s.d. from duplicate experiments.



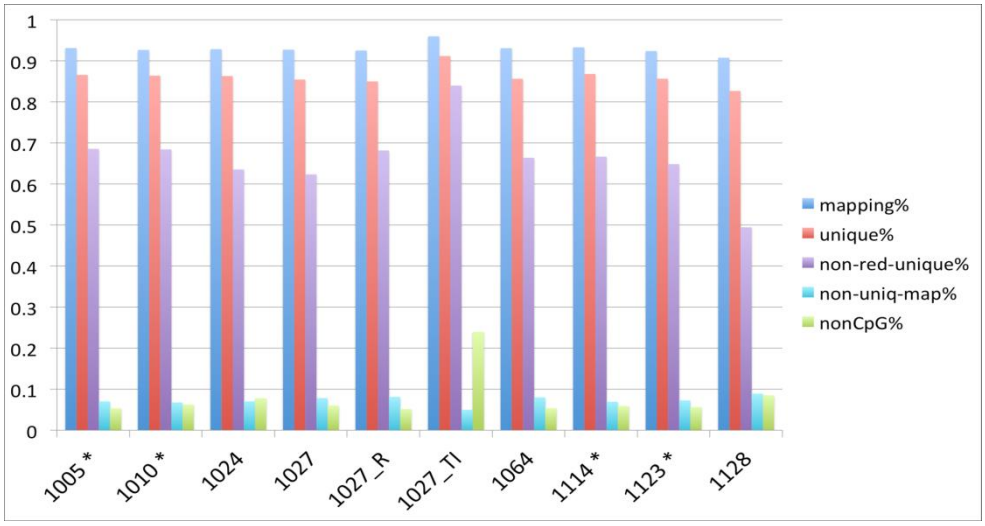
Supplementary Figure S10. Effect of M.SssI on detectable 5-hydroxymethylcytosine levels in human brain genomic DNA.

Genomic DNA from a pool of 5 human brain samples was treated with M.SssI under conditions equivalent to High intensity mTAG labeling (Standard Buffer:10 mM Tris pH 7.6, 50 mM NaCl, 0.1 mg/ml BSA and 400 nM M.SssI) in the absence of cofactor to maximize any possible effects to levels of hmC (4 technical replicates). hmC sites were identified using a enzymatic glucosylation coupled with R.MspI restriction endonuclease digestion, followed by microarray analysis4. In control reactions, the brain DNA was exposed to the same labeling conditions, only in the absence of M.SssI (4 technical replicates). Shown are chr1 and chr6 probes containing CCGG target sites (75,585 probes). *x* axis, log difference in glucosylated DNA relative to unglucosylated DNA; *y* axis, statistical significance of change (two-tailed *t* test). Probes with statistical significance $P < 0.05$ and log-difference >=1, indicating probe-level hmC detection (red dots), lie in the top-right quadrant; number beside rectangle indicates probes that fall in the region. The number of the orange rectangle bounds probes with equal but negative magnitude; as expected, these number fewer than those in the right quadrant; the results are similar when the p-value threshold is removed (data not shown). Volcano plots show that equivalent amount of hmC is detected in untreated (left; n=4 samples) and M.SssI-treated (right; n=4) samples.

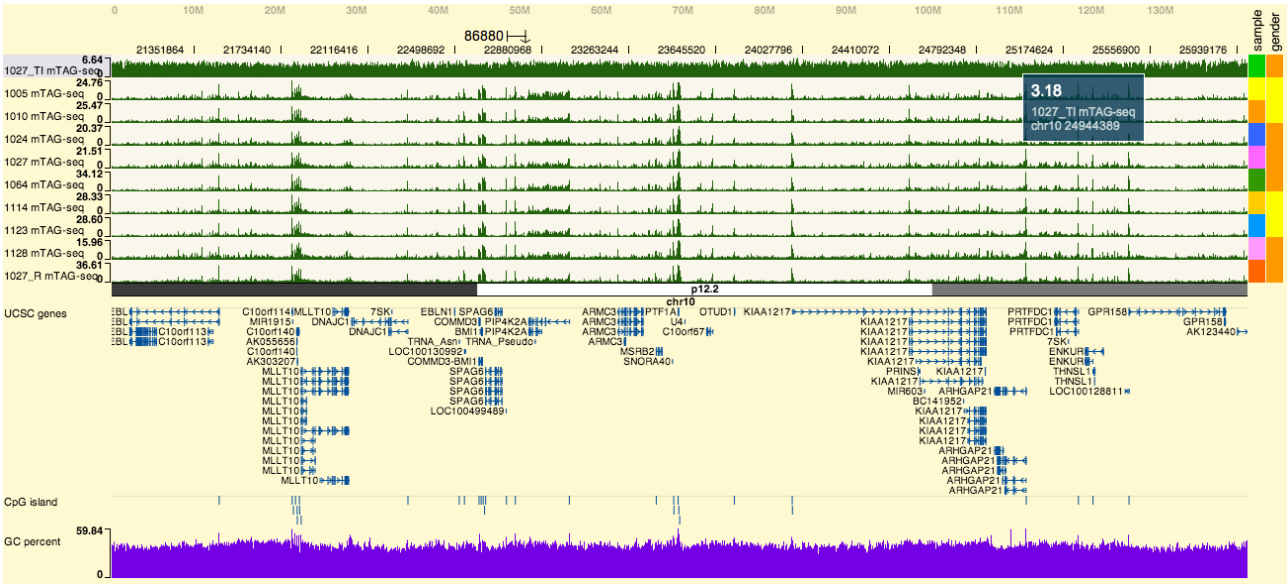


Supplementary Figure S11. Quality assessment of pilot mTAG-seq data.

Genome-wide pairwise correlation (500 bp window, rkpm values) of mTAG-enriched datasets between biological samples (human post mortem brain, encoded with a four-digit number, _R, a replicate sample) and a non-enriched (TI, a total input) control.

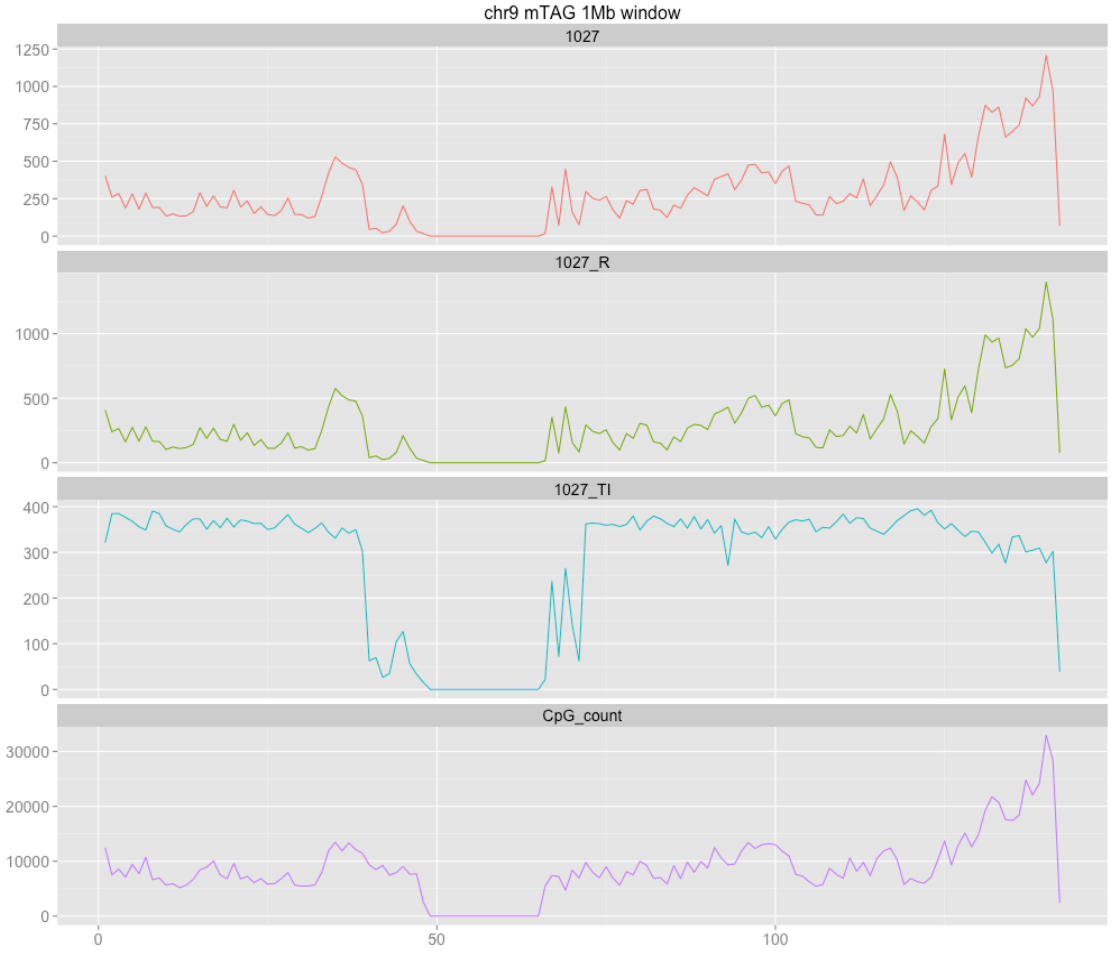


Supplementary Figure S12. Mapping statistics of mTAG-seq data obtained from human brain gDNA.
Biological samples (post mortem brain) are encoded with four-digit numbers, _R designates a replicate, _TI - a total input control; * indicates a female sample.
mapping%: percentage of mapped read pairs (in hg19);
unique%: percentage of uniquely mapped read pairs;
non-red-unique%: percentage of non-redundant uniquely mapped read pairs;
non-uniq-map%: percentage of non-uniquely mapped reads pair;
nonCpG%: percentage of read pairs that map to regions with no CpG.



Supplementary Figure S13. A typical WUSTL Epigenome Browser view of mTAG-seq data.

http://epigenomegateway.wustl.edu/browser/**.** gDNA samples are as above.



Supplementary Figure S14. A bird-eye view of mTAG-seq at the whole chromosome level.

mTAG-seq profile was highly reproducible (1027 and 1027_R) across chromosome 9. Non-enriched input signal (_TI) was uniform across the chromosome 9 (except for centromere regions, due to incomplete genome assembly). An overall similarity between the mTAG signals and the CpG density at the chromosome level is observed.

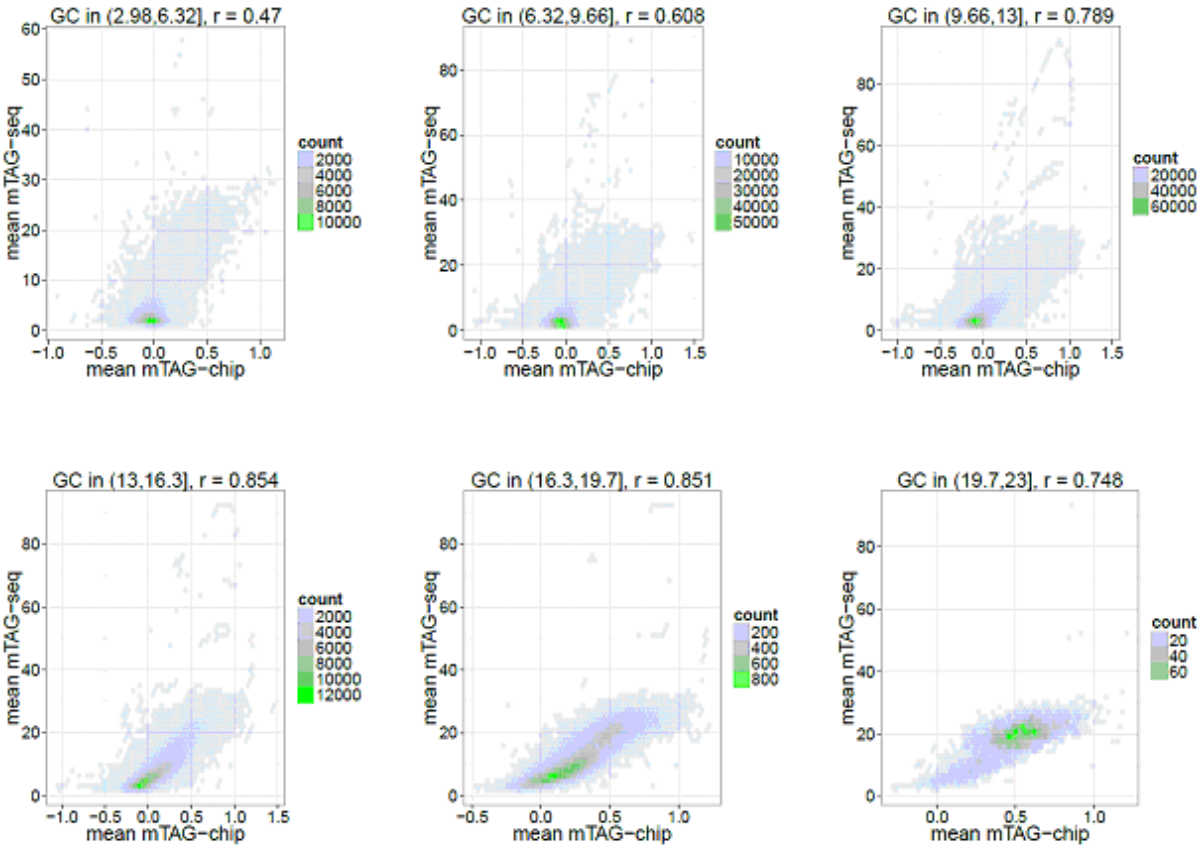


Figure S15. Pearson correlation between mTAG-seq and mTAG-chip data stratified according to local GC density.

Chromosome 1 intensities and read counts of 3 human brain samples were used for correlation analysis. mTAG-chip signals were log2 transformed and quantile normalized and the average total input was subtracted from each sample to correct for probe affinity bias. For each chip probe, its value was estimated as weighted average of the nearby probes. Similarly, for sequencing data, a read-count was estimated at chromosome coordinates matching the centers of chip probes using neighbouring read-count values. Pearson correlations were computed for probes representing 6 levels of GC density (shown are numbers of G+C nucleotides in a probe) using Gaussian kernel smoothing (bandwidth of 1900 bp for mTAG-seq and 3100 bp for mTAG-chip data).

Supplementary Tables

Supplementary Table S1. DNA primers for qPCR probes.

Fragment	Primer name	Primer sequence (5'→3')
Specific: 0-CGm, 1-CG 2-CG, 4-CG	Specific-dir	atgtgttggagtgtgcctga
	Specific-rev	gtggctctgattgatggctc
	TaqMan probe	FAM-tccctgtgtgatcacccctatgcttg-BHQ1
Nonspecific: 0-CGx	Nonspecific-dir	caggcctcttcaagggtca
	Nonspecific-rev	aagagatgagggcctggg
	TaqMan probe	JOE-tggcccatacctcttcaagggca-BHQ1

Supplementary Table S2. Distribution of CpG sites in fragments of different CpG content on human chromosomes 4, 15 and 18.

CG density decile	1	2	3	4	5	6	7	8	9	10
# of CGs/ kbp	1-2	3	4	5	6	7	8-9	10-11	12-16	17-177
% of total CGs	.6	.9	.5	.8	.3	.4	1.8	0.0	8.1	2.5

Supplementary Methods

Construction and purification of the mutant M.SssI methyltransferase

His_6-C-tagged M.SssI was obtained by cloning the synthetic *sssIM* gene (kind gift of A. Lubys) into the pBAD expression system. Further site-directed mutagenesis (Q142A/N370A) of M.SssI was performed using megaprimer mutagenesis protocol [52]. *E. coli* ER2267 cells harboring the resulting pBAD-sssIQN plasmid were grown at 37 °C in LB containing 0.1 mg/ml ampicillin to OD_{600}=0.6–0.7. Protein expression was induced with 0.2% arabinose at 16 °C. The supernatant of sonicated cells applied onto a 5 ml HiTrap IMAC column (GE Healthcare). Protein was eluted with 5–500 mM imidazole. Fractions containing the methyltransferase were applied onto a 5 ml HiTrap SP sepharose column (GE Healthcare), eluted with a 0.1–1 M NaCl and dialyzed to remove endogenous AdoMet.

mTAG labeling and enrichment for high-throughput analysis (96-well plate format)

The procedure can be scaled-up for a high throughput, cost-effective analysis of numerous samples. This protocol follows the same principles as described previously, but contains several modifications.

Sonication of genomic DNA was performed using a Covaris S220 instrument (200 cycles/burst, 10% duty factor) in 10 mM Tris-HCl (pH 8.5) for 3 min/sample to yield fragments with a peak size of 200 bp. Sheared DNA was end-filled using the Fast DNA End Repair Kit (Fermentas). Sheared DNA was mixed with 1X End Repair Buffer and 4 µl of End Repair Enzyme Mix (140 µl final volume) and then incubated at 20 °C for 10 min, followed by an enzyme inactivation step of 75 °C for 20 min and 55 cycles of 75 °C 10 sec with -1 °C per cycle. DNA was purified using the ZR-96 DNA Clean & Concentrator-5 kit (Zymo Research).

mTAG-click labeling was performed at Medium Intensity conditions. Biotinylation of labeled sites was completed (in a 96-well PCR plate) immediately after labeling. Initially, 30 µl of 10 mM Tris-HCl pH 8.5 was added to each sample. Each sample then received a 5 µl aliquot containing 4.4 µl of 10 mM Tris-HCl pH 8.5 and 0.6 µl of freshly prepared 20 mM dibenzylcyclooctyne-S-S-PEG3-biotin (Click Chemistry Tools) dissolved in dimethylformamide and was incubated for 2 h at 37 ˚C. Samples were purified using the ZR-96 DNA Clean & Concentrator-5 kit (Zymo Research) and eluted to a final volume of 32 µl of 10 mM Tris-HCl pH 8.5.

To capture biotinylated DNA fragments, Dynabeads MyOne Streptavidin T1 (Invitrogen) were concentrated to 6.25 mg/ml with 25 mM Tris-HCl (pH 7.5), 2.5 mM EDTA, and 5 M NaCl. Samples placed in a round bottom 96-well plate were each given 8 µl of beads (total volume 40 µl) and were mixed at 900 rpm for 2 h at room temperature. The beads were collected and the supernatant (unbound fraction) was removed. The beads were washed three times with 15 mM Tris-HCl (pH 7.5), 1.5 mM EDTA, 3 M NaCl and twice with 5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 1 M NaCl.

DNA was recovered by resuspending beads in 20 µl of 50 mM DTT, 60 mM Tris (pH 7.8), followed by a 1 h incubation with mixing at 900 rpm in a round bottom 96-well plate. Samples were placed on a magnetic plate to separate the beads from the released DNA. The recovered DNA was ligated to adaptors in a reaction containing 0.5 mM ATP, 10 mM $MgCl_2$, 5% PEG, 3.3 µM adaptors and 10 a.u. T4 Ligase (ThermoFisher) (30 µl total volume). Samples were ligated overnight at 22 °C followed by an enzyme inactivation step of 65 ˚C 15 min and 40 cycles of 65 °C 10 sec -1°C/cycle. Immediately afterwards, samples were treated with 32 mM 2-mercaptoethanol for 10 min at room temperature and PCR amplified in two rounds. The first PCR reaction performed by adding Buffer (63 mM Tris-HCl (pH 8.8), 17 mM $(NH_4)_2SO_4$, 0.0084% (v/v) Tween 20) 590 µM dNTPs, 1.5 µM A-25 Primer and 6.3 a.u. Taq. PCR cycling conditions: 1 min 50 °C, 2 min 72 °C, 1 min 94 °C, 4 min 94 °C, 1 min 65 °C, 5 min 72 °C, 15 cycles of 1 min 94 °C, 1 min 65 °C, 1 min 72°C, and the final extension step is at 72 °C for 2 min. The second round of PCR amplification and microarray analysis were carried out as described previously.

Supplementary Reference

52. Barik, S. Site-directed mutagenesis by double polymerase chain reaction. *Mol Biotechnol* **3**, 1–7 (1995).