Supplementary Materials

Table 1. Methylation status of orthologous CGIs in human and mouse brain. Methods to select orthologous CGIs and to determine methylation status are described in Supplementary Methods.

	Promoter CGIs (n=1245)		**Intragenic CGIs (n=502)**	
	Unmethylated in human	Methylated in human	Unmethylated in human	Methylated in human
Unmethylated in mouse	1234	3	293	66
Methylated in mouse	6	2	36	107
P-value	~0		~0	

Table 2. Methylation, H3K4me3 and TSS status for human CGIs as determined by MeDIP-seq (brain), MRE-seq (brain), H4K4me3 ChIP-seq (brain), RNA-seq-inferred TSS (brain) and CAGE tags (multiple tissues). See Supplementary Methods for description of cutoffs.

type of CGI	Total number	1 Number (%) fully methylated by MeDIP	2 Number (%) partially methylated by MeDIP	3 Number (%) unmethylated by MeDIP	4 Number (%) unmethylated by MRE	5 Number (%) H3K4me3 positive	6 Number (%) smart tag positive	7 Number (%) CAGE tag positive	Number (%) with 3,4,5,7	number (%) with 3,4,5,6,7
promoter	11967	237 (2.0)	227 (1.9)	11503 (96.1)	10622 (88.8)	10341 (86.4)	8206 (68.6)	11500 (96.1)	9500 (79.4)	6523 (54.5)
intragenic	8092	3600 (44.5)	1519 (18.8)	2973 (36.7)	2454 (30.3)	1983 (24.5)	2397 (29.6)	5239 (64.7)	1448 (17.9)	794 (9.8)
3' of tx	512	152 (29.7)	105 (20.5)	255 (49.8)	228 (44.5)	132 (25.8)	142 (27.7)	400 (78.1)	110 (21.5)	45 (8.8)
intergenic	7068	1887 (26.7)	940 (13.3)	4241 (60.0)	3688 (52.2)	2615 (37.0)	3386 (48.0)	3866 (54.7)	1761 (24.9)	1146 (16.2)

Supplementary Figures

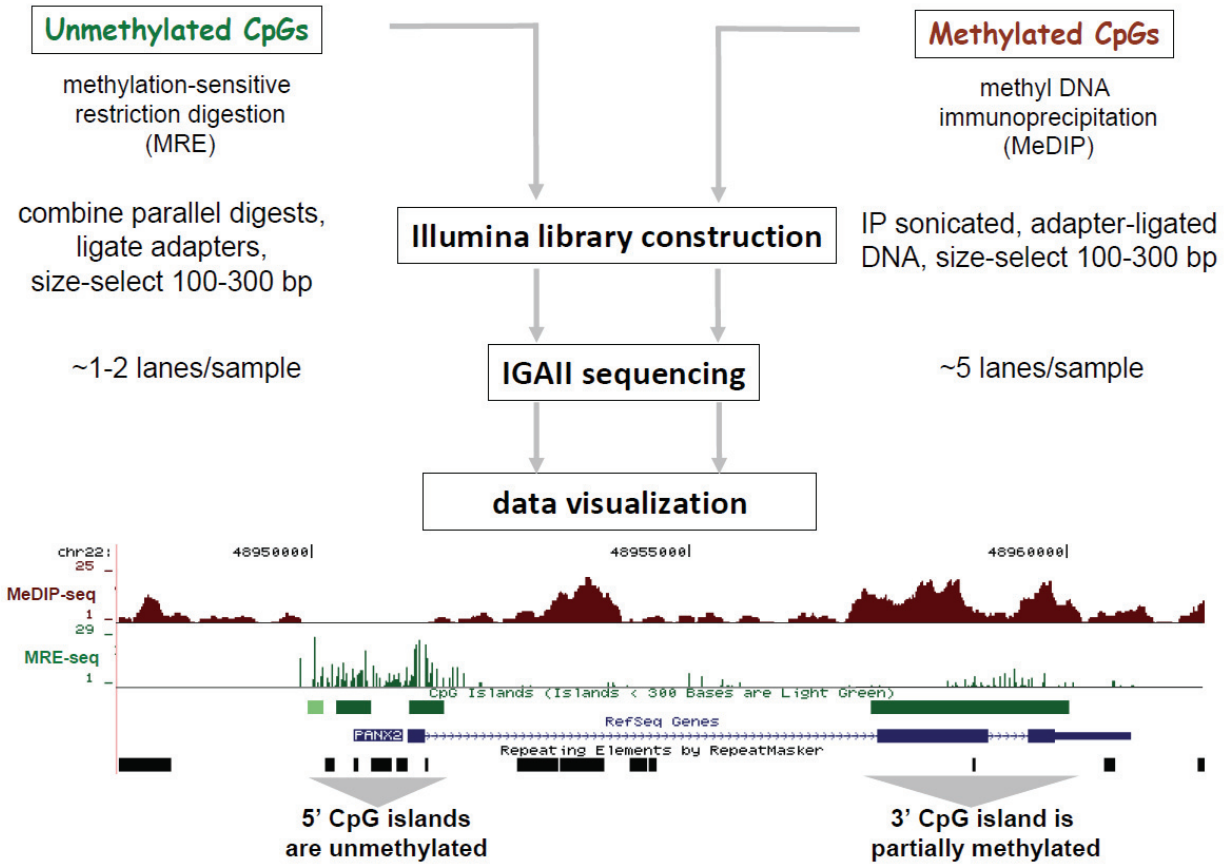


Figure S1. **Workflow for MRE- and MeDIP-seq**. For MRE-seq, unmethylated CpG sites are cut with methyl-sensitive restriction enzymes, adapter-ligated, and size selected for 100-300 bp, representing DNA fragments unmethylated at 2 CpG sites within close proximity. For MeDIP-seq, sonicated, adapter-ligated DNA is immunoprecipitated with anti-methylcytosine antibody and size-selected for 100-300 bp. Both libraries are subjected to Illumina Genome Analyzer II (IGAII) sequencing and reads are aligned to the reference genome. For visualization, the starts of MRE-seq reads are graphed as vertical lines, which identify single unmethylated CpG sites. The height of the line corresponding to an individual CpG site is determined from the number of reads originating at that site in either orientation. For visualization of MeDIP-seq, reads are computationally extended to 200 bp (based on average fragment size in library) and stacked to display regions of coverage.

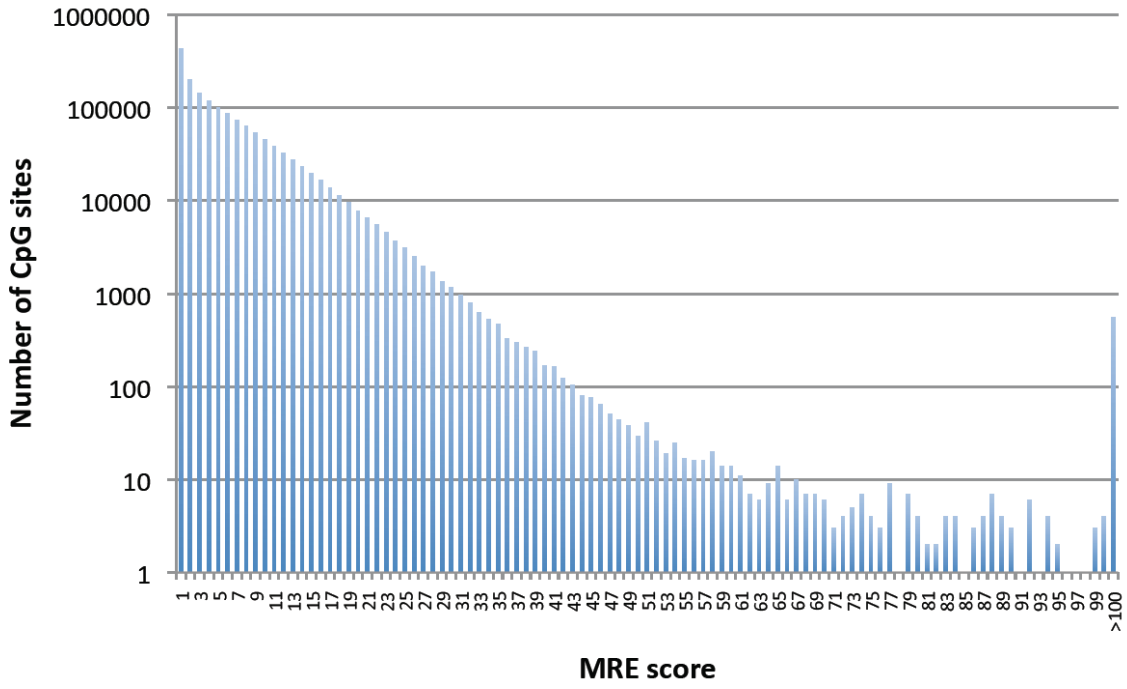


Figure S2. MRE-seq score distribution on CpG sites. Histogram of CpG sites with specific MRE-score is shown. The x-axis indicates MRE-scores. Y-axis indicates number of CpG sites with a specific MRE-score.

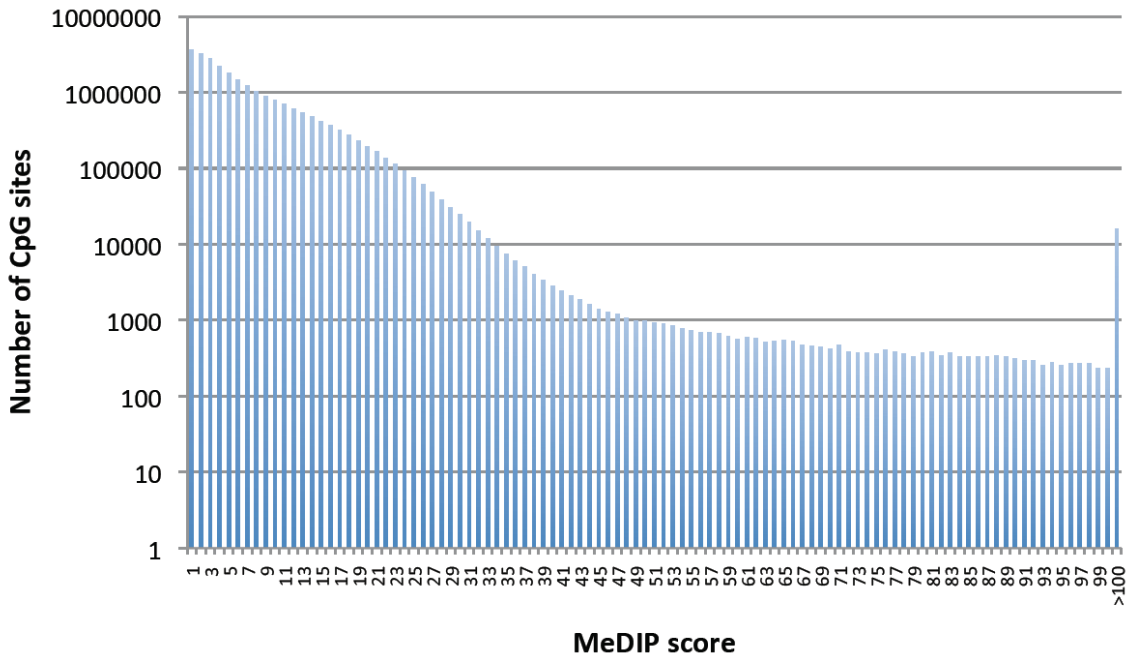


Figure S3. MeDIP-seq score distribution on CpG sites. Histogram of CpG sites with specific MeDIP-score is shown. The x-axis indicates MeDIP-scores. Y-axis indicates number of CpG sites with a specific MeDIP-score.

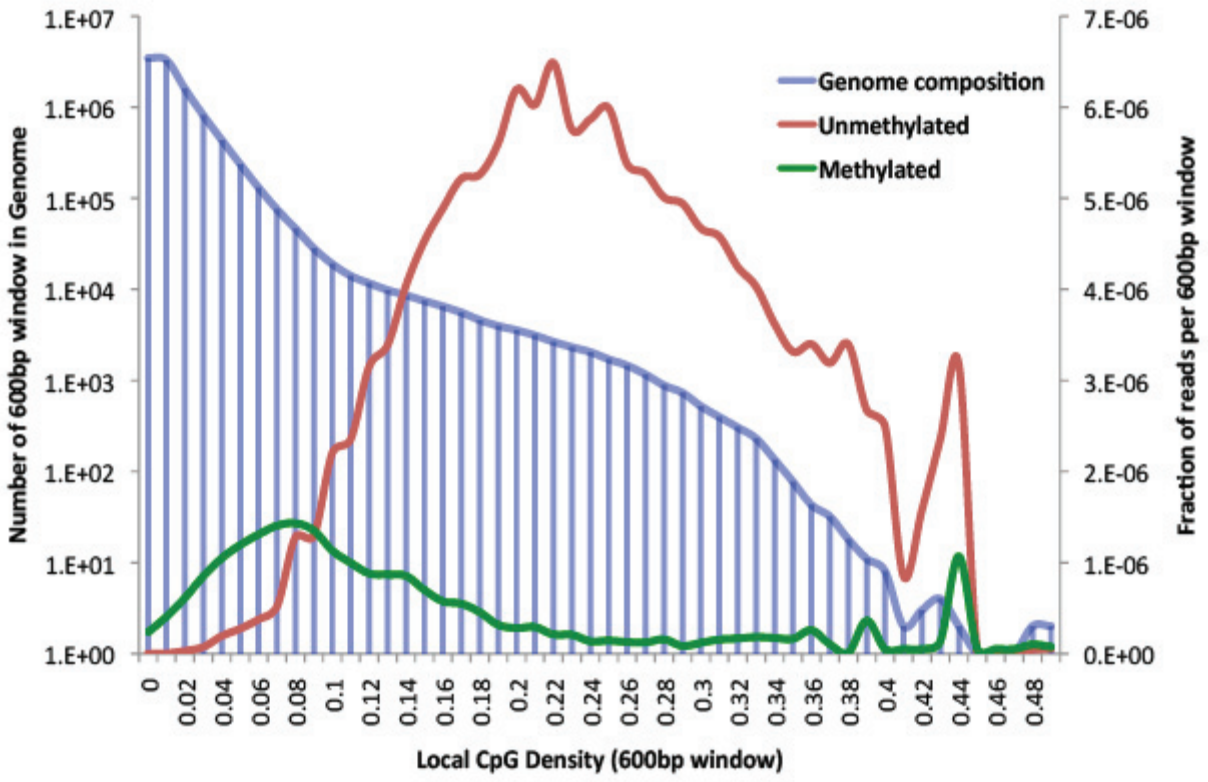


Figure S4. Genome-wide analysis illustrating the inverse correlation between MeDIP-seq and MRE-seq. The left Y-axis and the blue line show the distribution of CpG density in 600 bp windows across the human genome. The right Y-axis and the red and green lines show the fraction of MRE-seq and MeDIP-seq reads in each local bin of CpG density. Each method preferentially detects a different genomic fraction, and the occurrence of MeDIP-seq and MRE-seq reads is related to methylation status and CpG density. Within the CpG-poor fraction, the frequency of MeDIP-seq reads increased with CpG density. In contrast, MRE-seq reads were preferentially in the CpG-rich fraction of the genome, reflecting the unmethylated status of the majority of CGIs, particularly in 5' promoter regions.

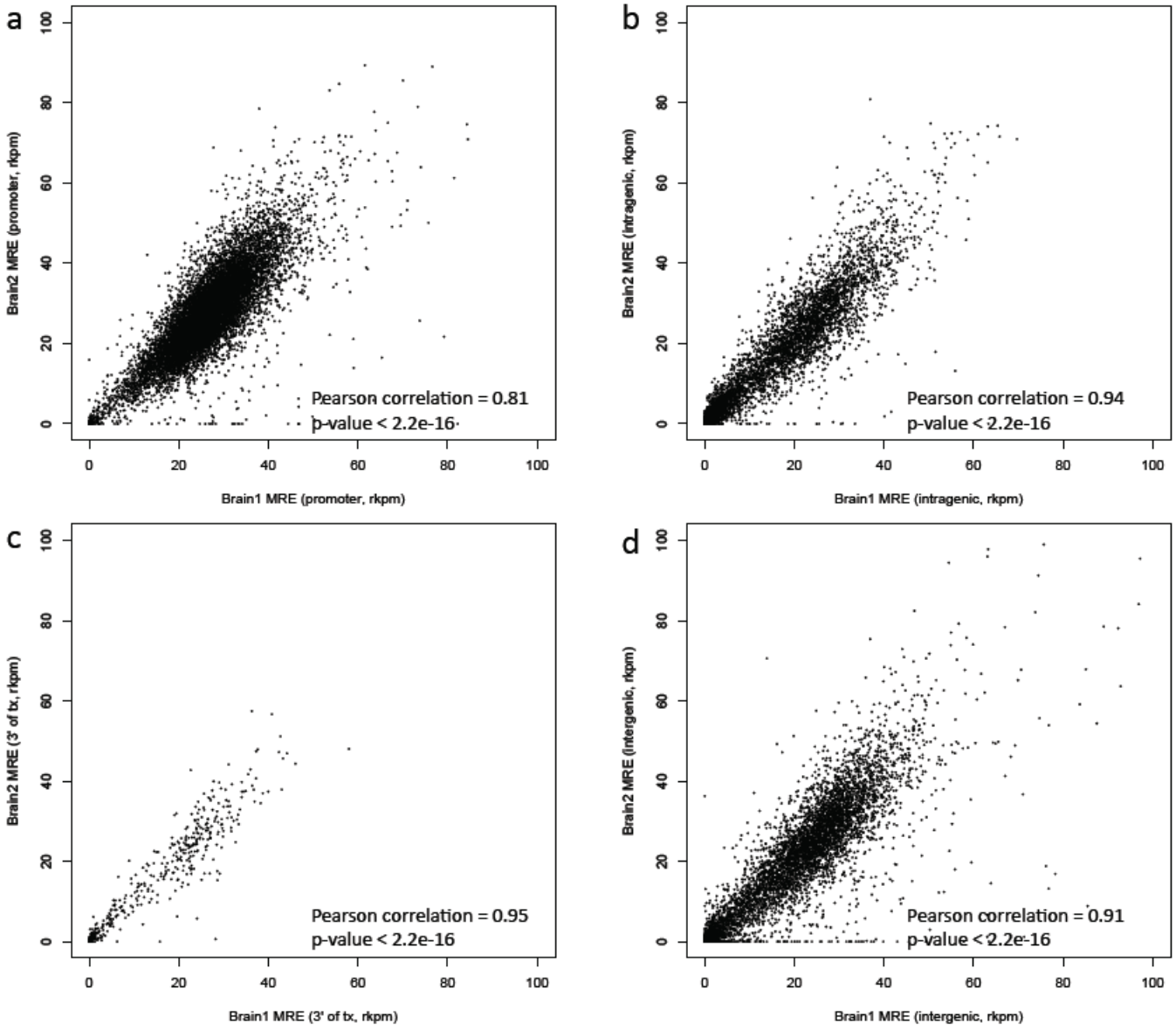


Figure S5. MRE-seq from two different human frontal cortex samples shows high reproducibility. a, Scatter plot of Reads per Kilobase per Million Reads (RPKM) for promoter CGIs, with Pearson correlation coefficient and p value shown. Similar plots are shown for intragenic (**b**), 3' (**c**), and intergenic (**d**) CGIs.

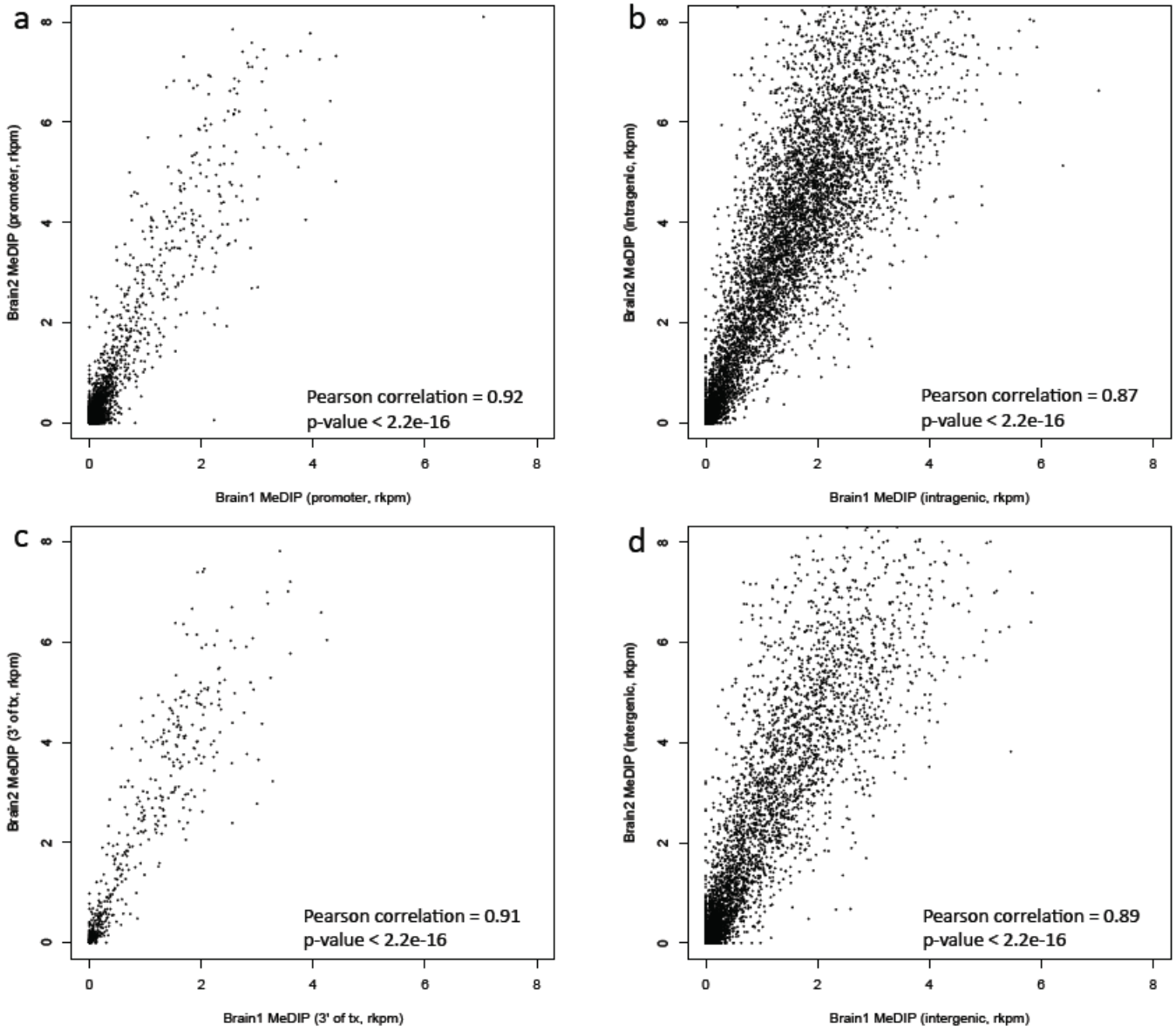


Figure S6. MeDIP-seq from two different human frontal cortex samples shows high reproducibility. a, RPKM scatter plot for promoter CGIs, with Pearson correlation coefficient and p value shown. Similar plots are shown for intragenic (**b**), 3' (**c**), and intergenic (**d**) CGIs.

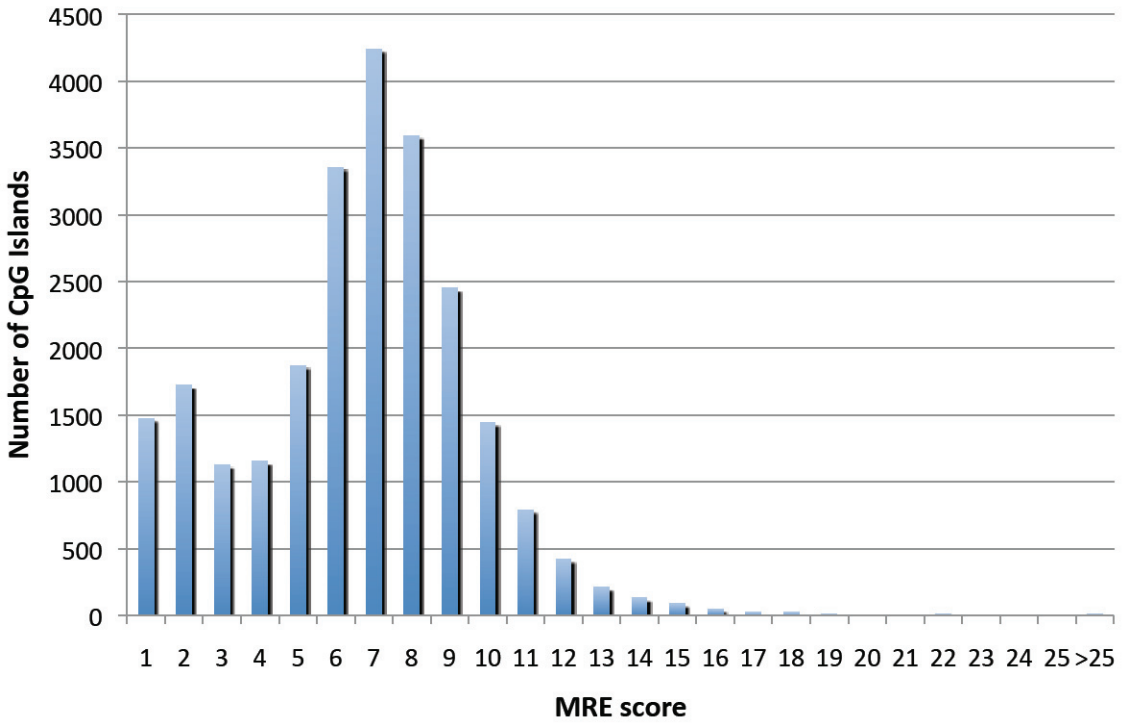


Figure S7. MRE-seq score distribution for CpG islands. Histogram of CpG islands with specific MRE-score is shown. The x-axis indicates MRE-scores. Y-axis indicates number of CpG islands with a specific MRE-score. The scores display bi-model distributions, suggesting two main populations of CpG islands, being either methylated or unmethylated.

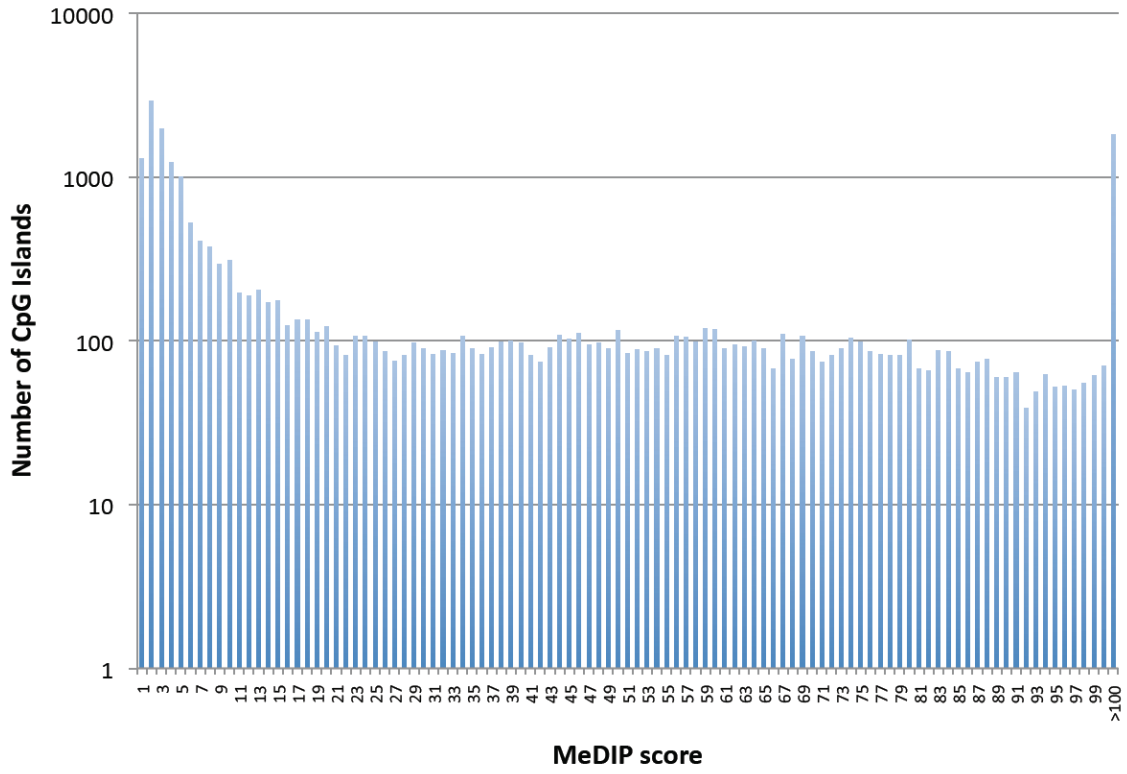
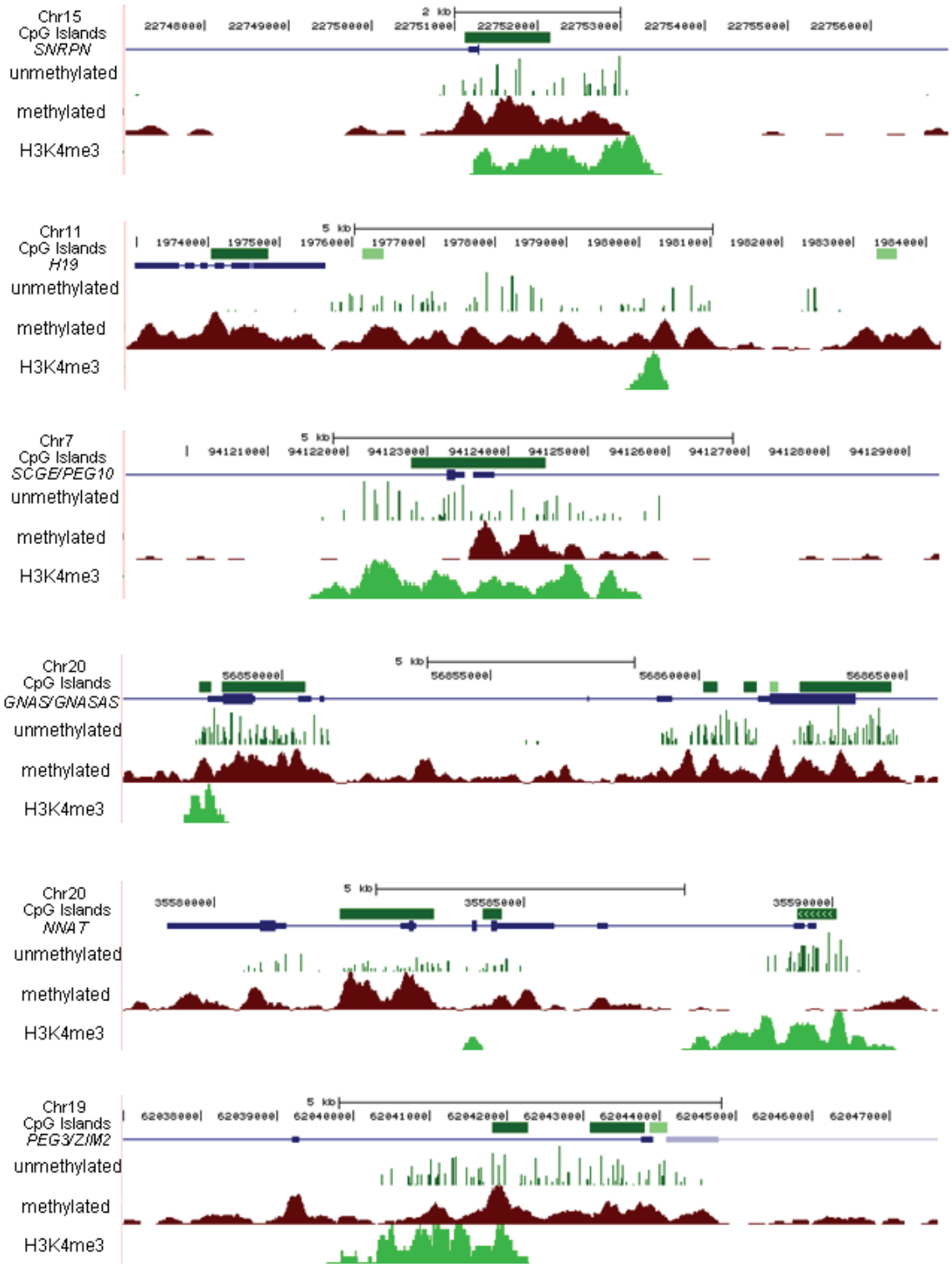
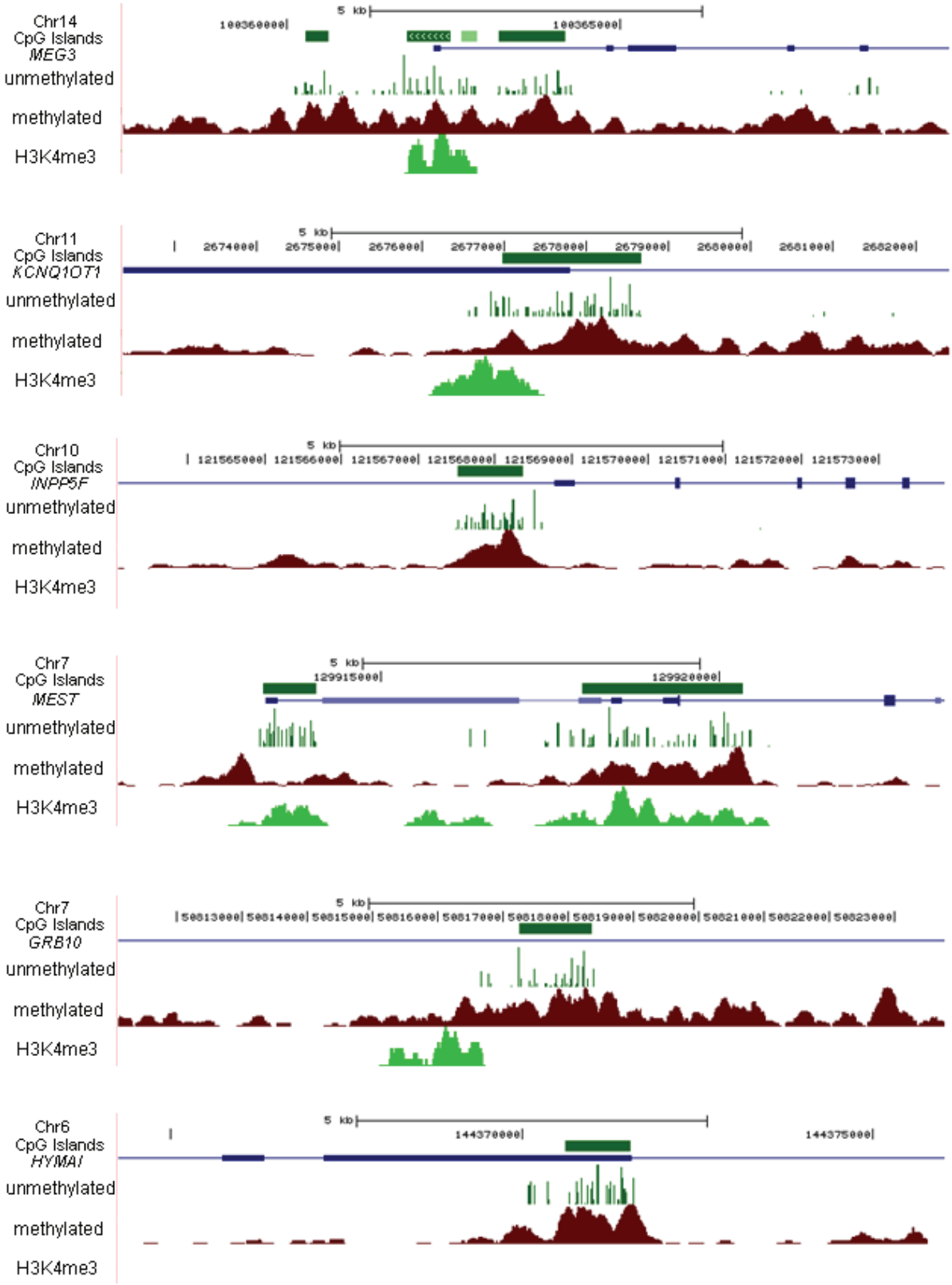


Figure S8. MeDIP-seq score distribution for CpG islands. Histogram of CpG islands with specific MRE-score is shown. The x-axis indicates MRE-scores. Y-axis indicates number of CpG islands with a specific MRE-score.

Figure S9


Chr15
CpG Islands
SNRPN
unmethylated
methylated
H3K4me3
Chr11
CpG Islands
H19
unmethylated
methylated
H3K4me3
Chr7
CpG Islands
SCGE/PEG10
unmethylated
methylated
H3K4me3
Chr20
CpG Islands
GNAS/GNASAS
unmethylated
methylated
H3K4me3
Chr20
CpG Islands
NNAT
unmethylated
methylated
H3K4me3
Chr19
CpG Islands
PEG3/ZIM2
unmethylated
methylated
H3K4me3


Chr14
CpG Islands
MEG3
unmethylated
methylated
H3K4me3
Chr11
CpG Islands
KCNQ1OT1
unmethylated
methylated
H3K4me3
Chr10
CpG Islands
INPP5F
unmethylated
methylated
H3K4me3
Chr7
CpG Islands
MEST
unmethylated
methylated
H3K4me3
Chr7
CpG Islands
GRB10
unmethylated
methylated
H3K4me3
Chr6
CpG Islands
HYMAI
unmethylated
methylated
H3K4me3

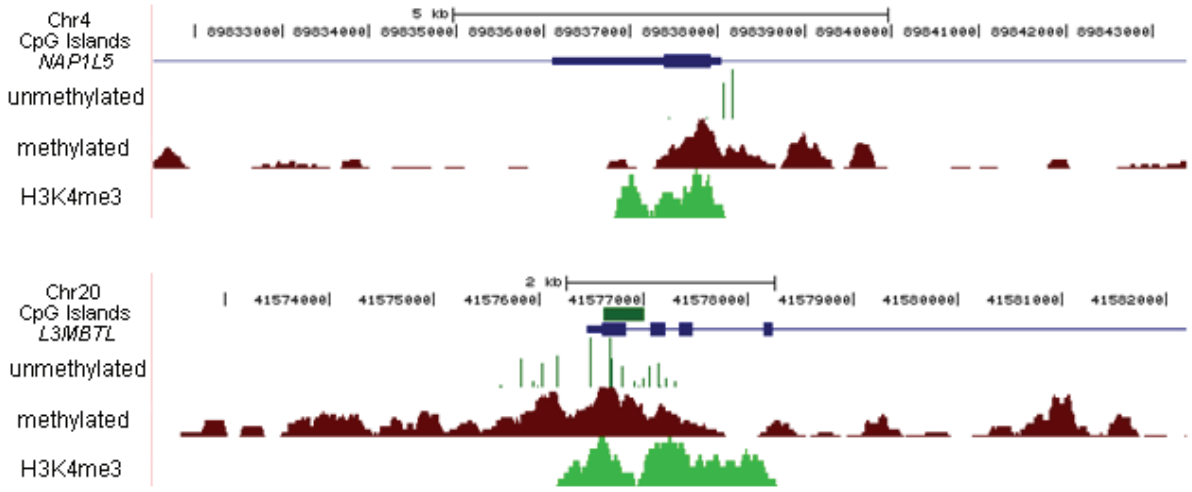


Figure S9. MeDIP- and MRE-seq consistently identify known differentially methylated regions (DMRs) of imprinted genes. Approximately 10 kb-wide UCSC genome browser windows are shown for each locus. 14/14 known imprinted DMRs showed overlapping MeDIP- and MRE-seq reads, consistent with allele-specific DNA methylation. Furthermore, H3K4me3 ChIP-seq of the same human brain demonstrated the presence of H3K4me3 at most known imprinted DMRs (12/14). 13/14 known DMRs overlapped with a CpG island. The one DMR without a corresponding CpG island, NAP1L5, did show overlapping MeDIP- and MRE-seq reads, although less dense MRE-seq reads were observed, likely due to the absence of dense MRE recognition sites (and thus fewer fragments) at this locus compared to most CpG islands.

Figure S10a

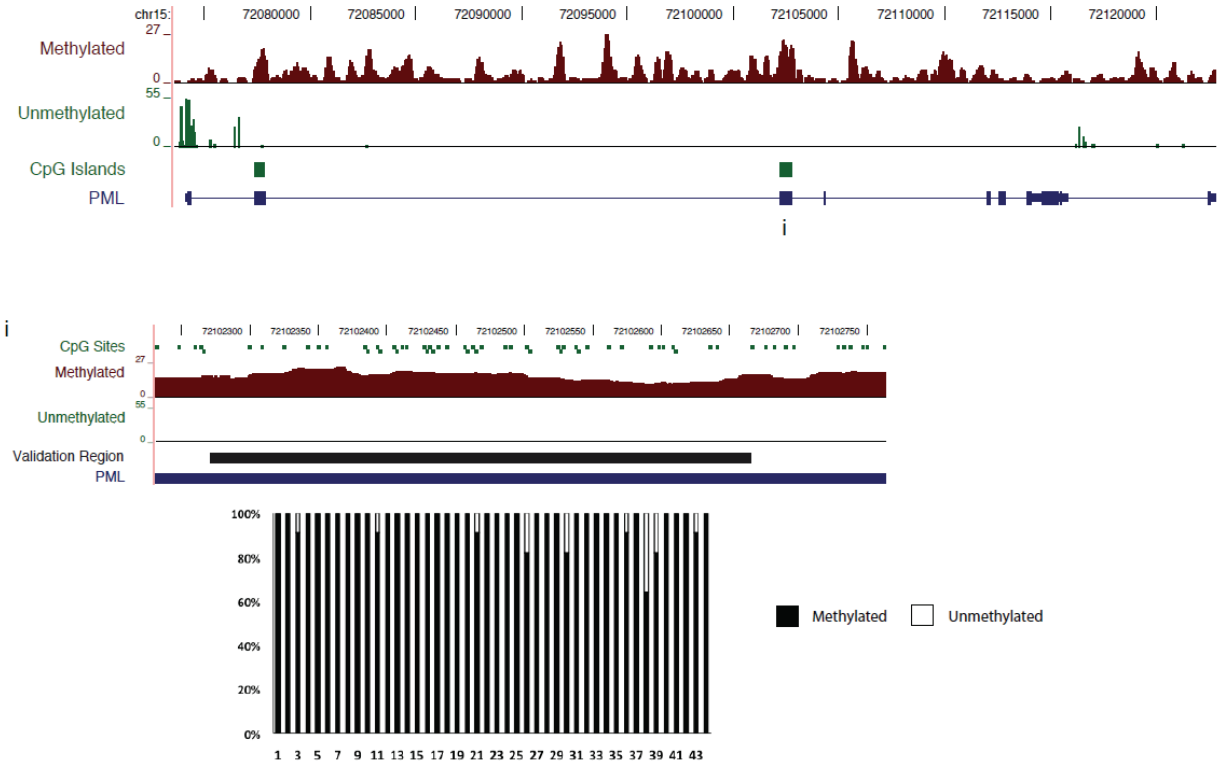

chr15:
72080000
72085000
72090000
72095000
72100000
72105000
72110000
72115000
72120000
Methylated
27
0
Unmethylated
55
0
CpG Islands
PML
i
i
72102300
72102350
72102400
72102450
72102500
72102550
72102600
72102650
72102700
72102750
CpG Sites
Methylated
27
0
Unmethylated
55
0
Validation Region
PML
100%
80%
60%
40%
20%
0%
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43
Methylated
Unmethylated

Figure S10b

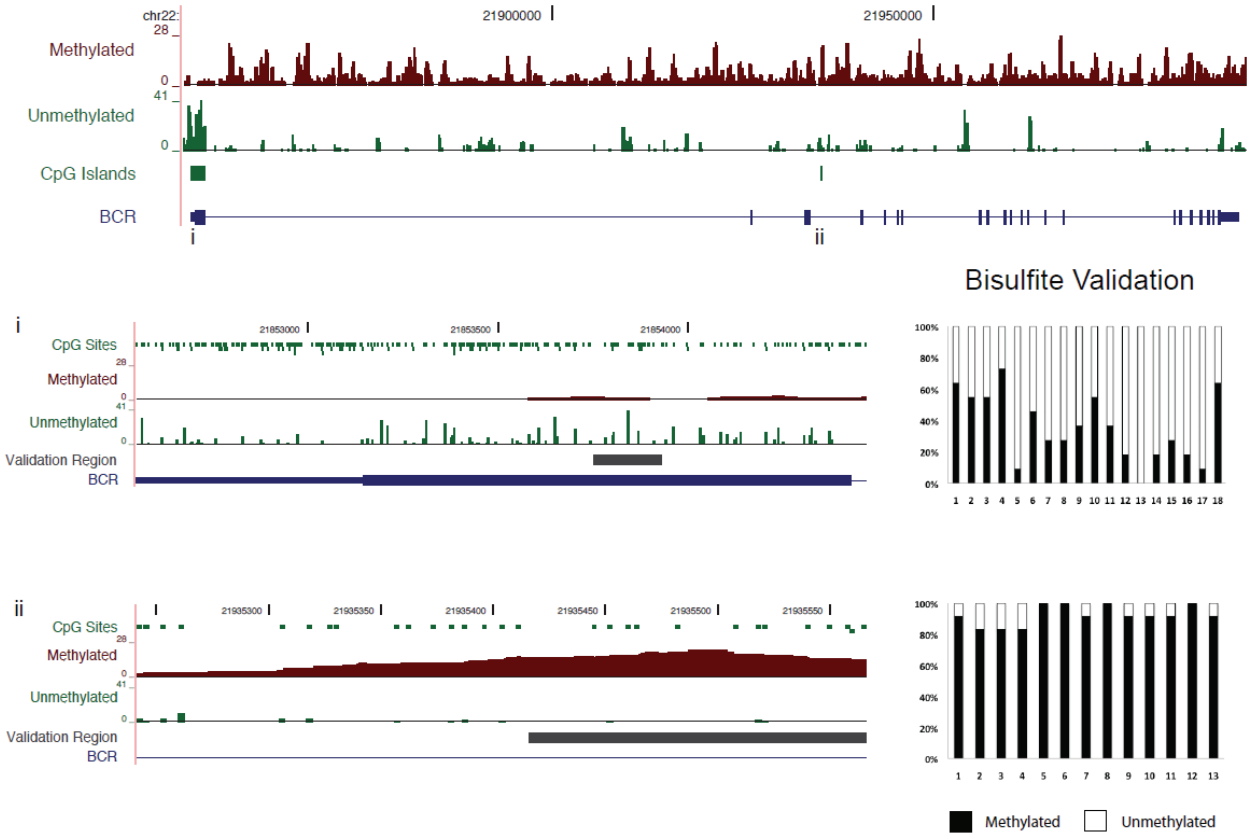

chr22:
21900000
21950000
Methylated
Unmethylated
CpG Islands
BCR
i
ii
Bisulfite Validation
CpG Sites
Validation Region
21853000
21853500
21854000
21935300
21935350
21935400
21935450
21935500
21935550
Methylated
Unmethylated

Figure S10c

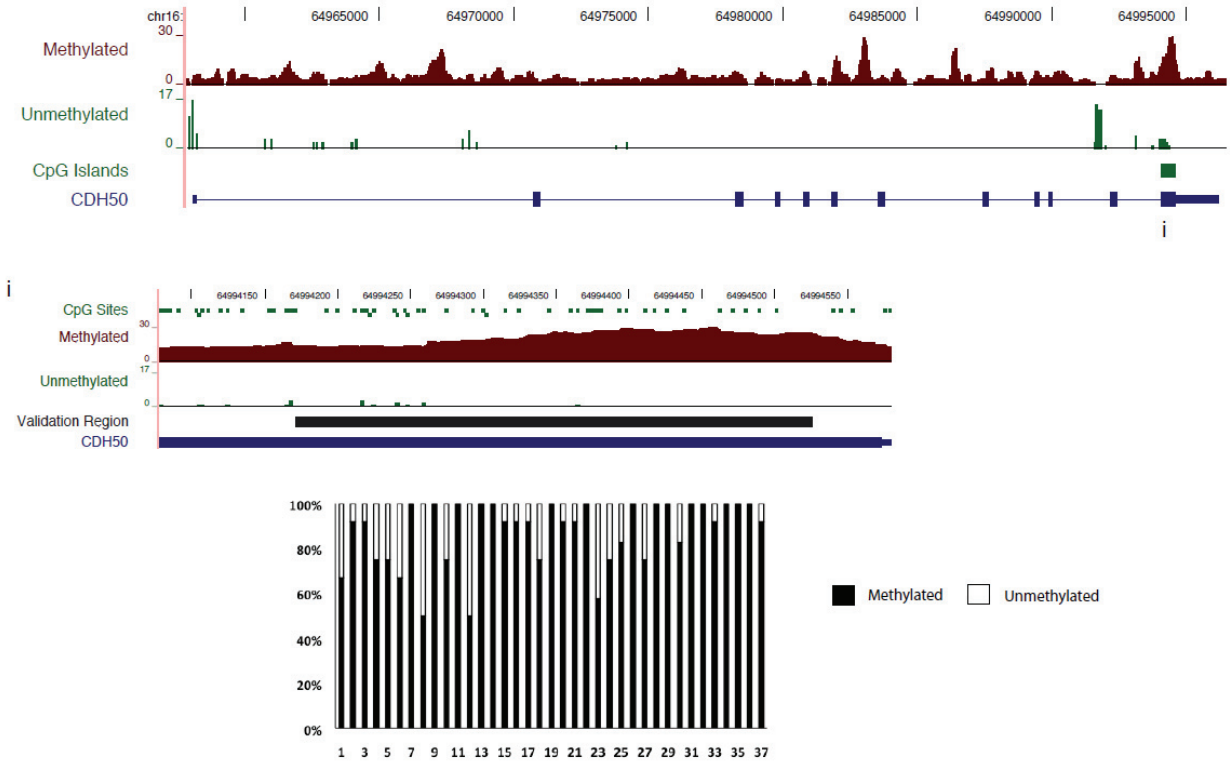

chr16:
64965000
64970000
64975000
64980000
64985000
64990000
64995000
Methylated
Unmethylated
CpG Islands
CDH50
i
i
64994150
64994200
64994250
64994300
64994350
64994400
64994450
64994500
64994550
CpG Sites
Methylated
Unmethylated
Validation Region
CDH50
100%
80%
60%
40%
20%
0%
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37
Methylated
Unmethylated

Figure S10d

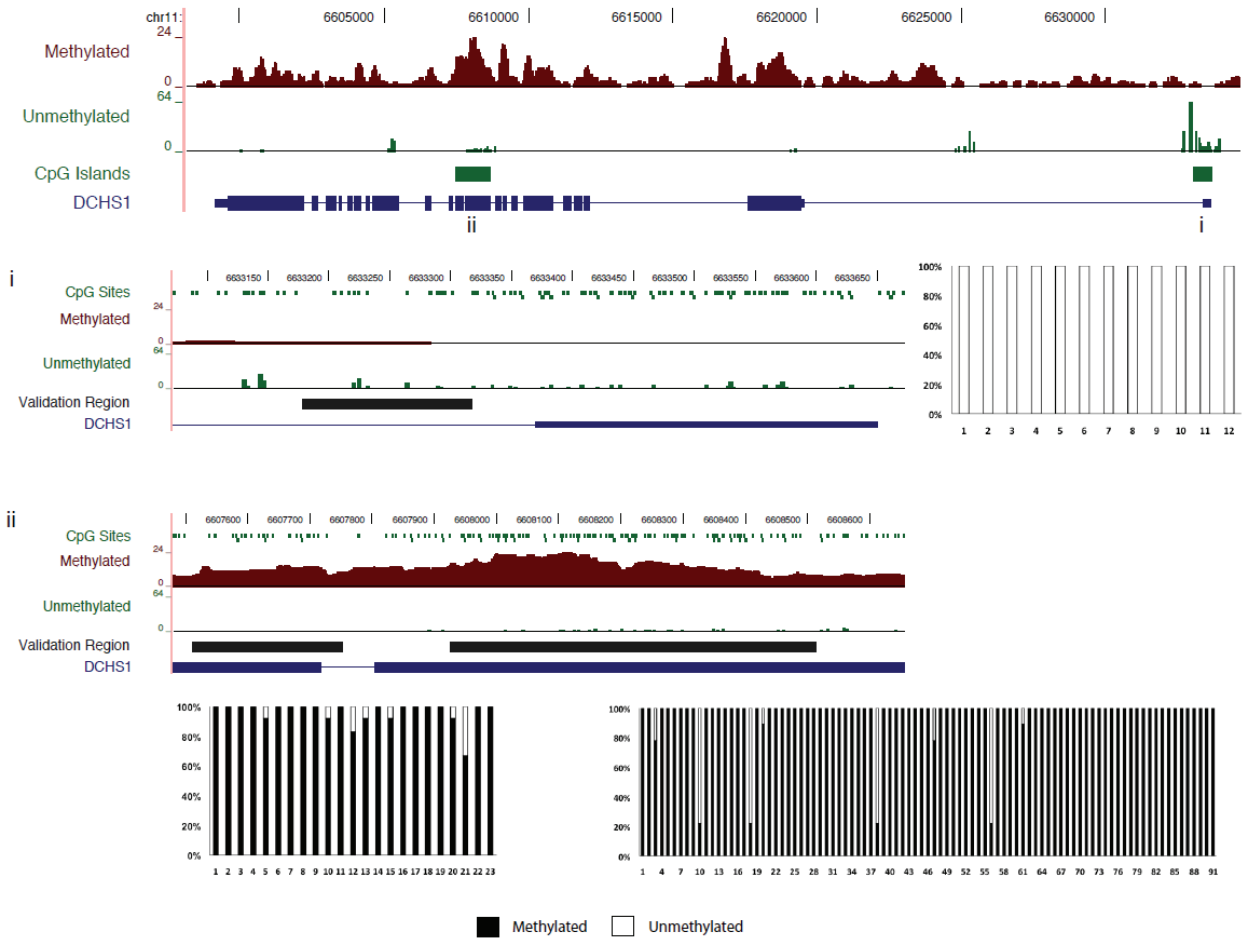
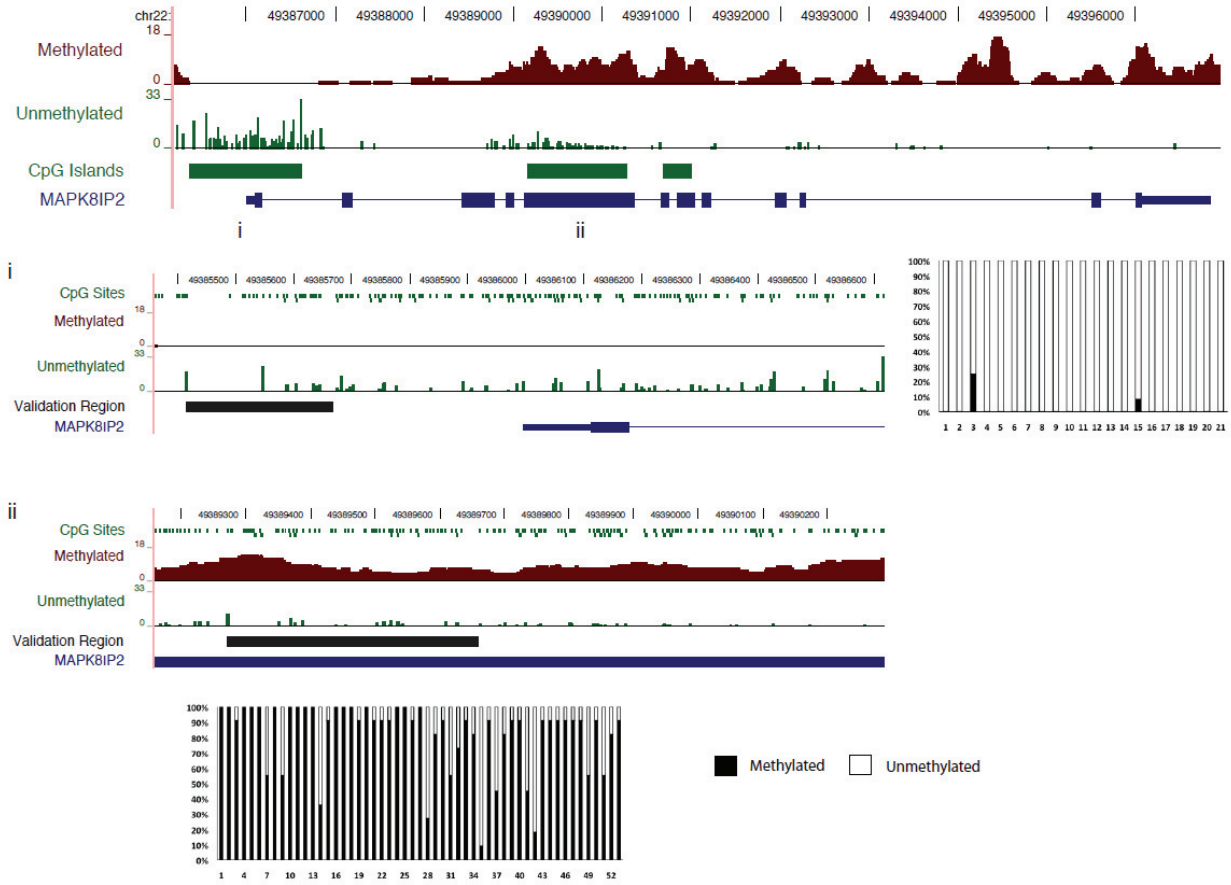

chr11:
6605000
6610000
6615000
6620000
6625000
6630000
Methylated
Unmethylated
CpG Islands
DCHS1
ii
i
i
CpG Sites
Methylated
Unmethylated
Validation Region
DCHS1
ii
Methylated
Unmethylated

Figure S10e


chr22:
49387000
49388000
49389000
49390000
49391000
49392000
49393000
49394000
49395000
49396000
Methylated
Unmethylated
CpG Islands
MAPK8IP2
i
ii
CpG Sites
Validation Region
Methylated
Unmethylated

Methylated
Unmethylated

Figure S10f

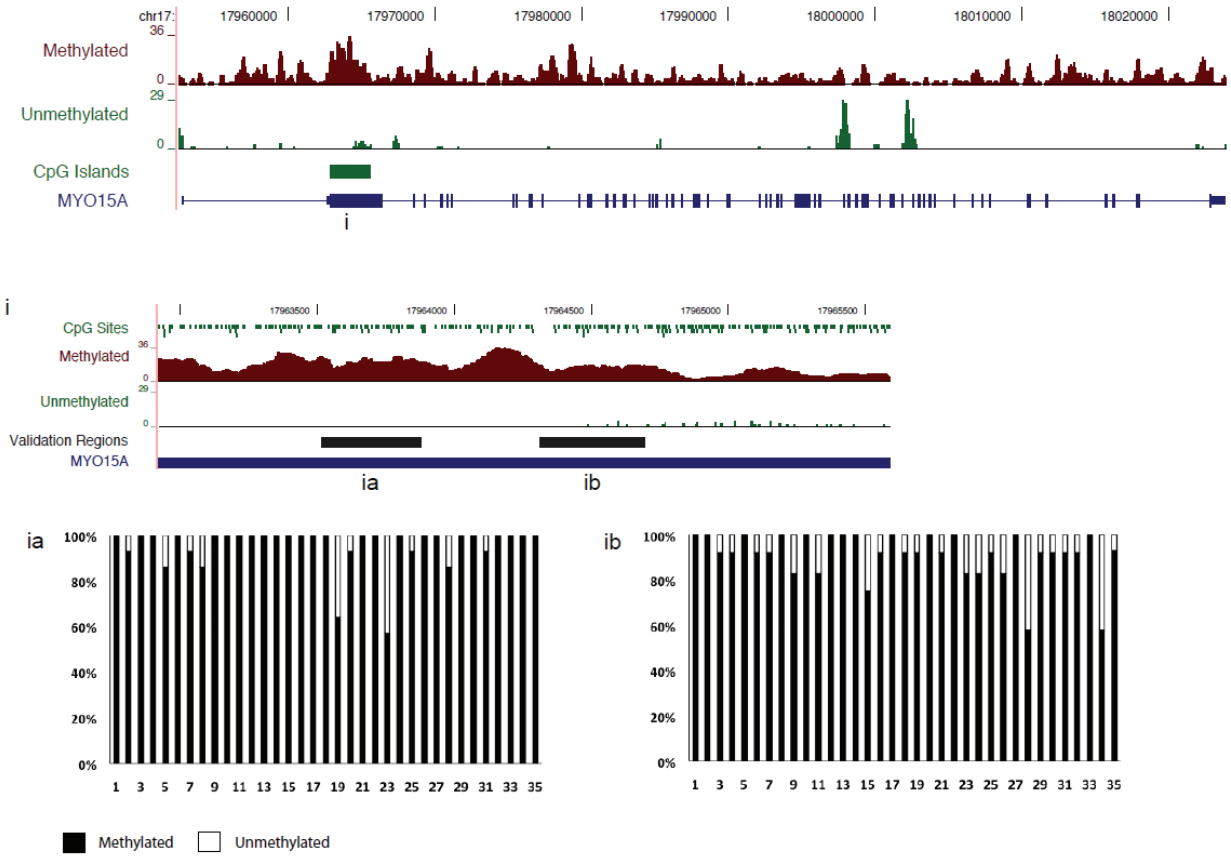

chr17:
17960000
17970000
17980000
17990000
18000000
18010000
18020000
Methylated
Unmethylated
CpG Islands
MYO15A
i
i
17963500
17964000
17964500
17965000
17965500
CpG Sites
Methylated
Unmethylated
Validation Regions
MYO15A
ia
ib
ia
ib
100%
80%
60%
40%
20%
0%
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
Methylated
Unmethylated

Figure S10g

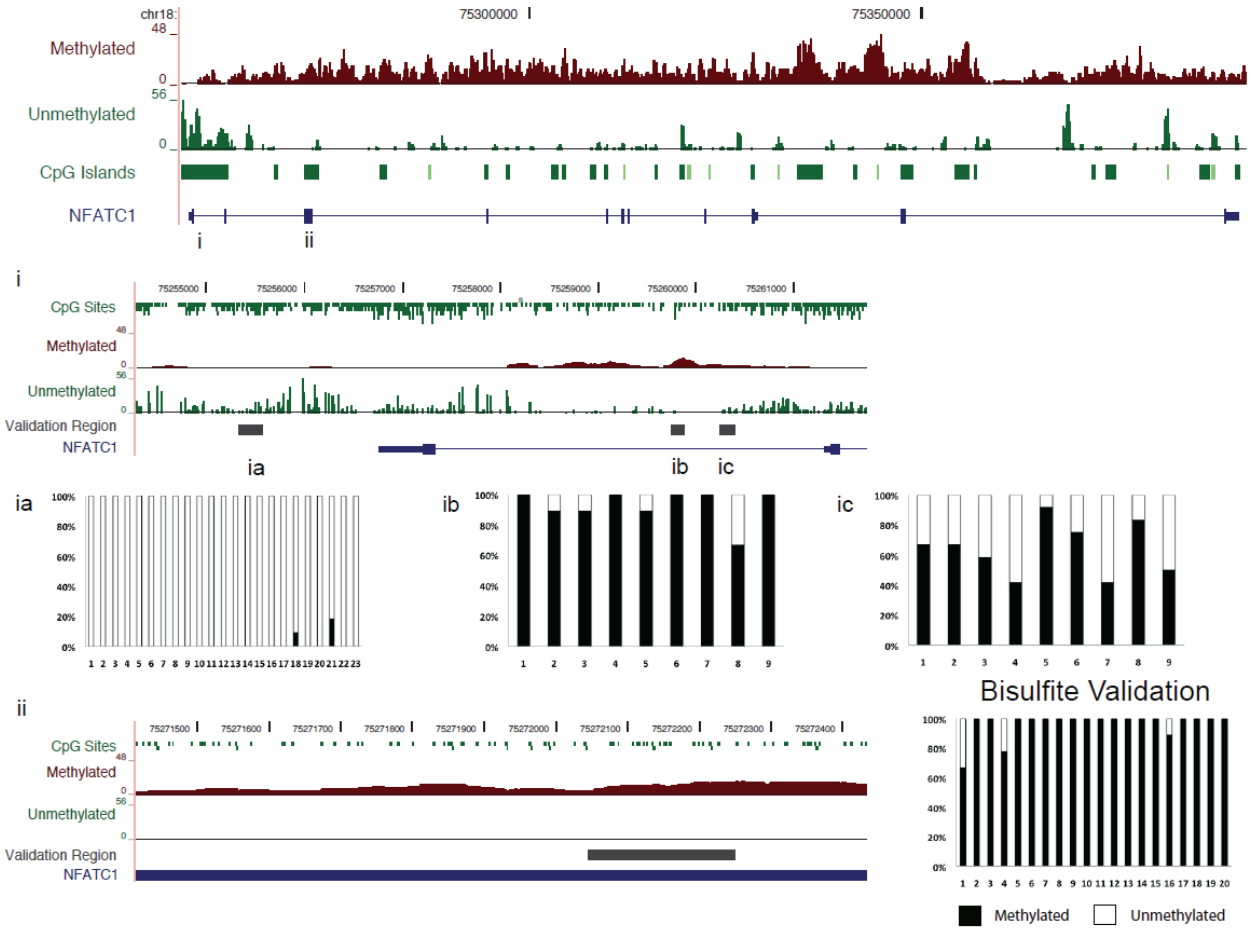

chr18:
75300000
75350000
Methylated
Unmethylated
CpG Islands
NFATC1
i
ii
CpG Sites
Validation Region
ia
ib
ic
Bisulfite Validation
Methylated
Unmethylated

Figure S10h

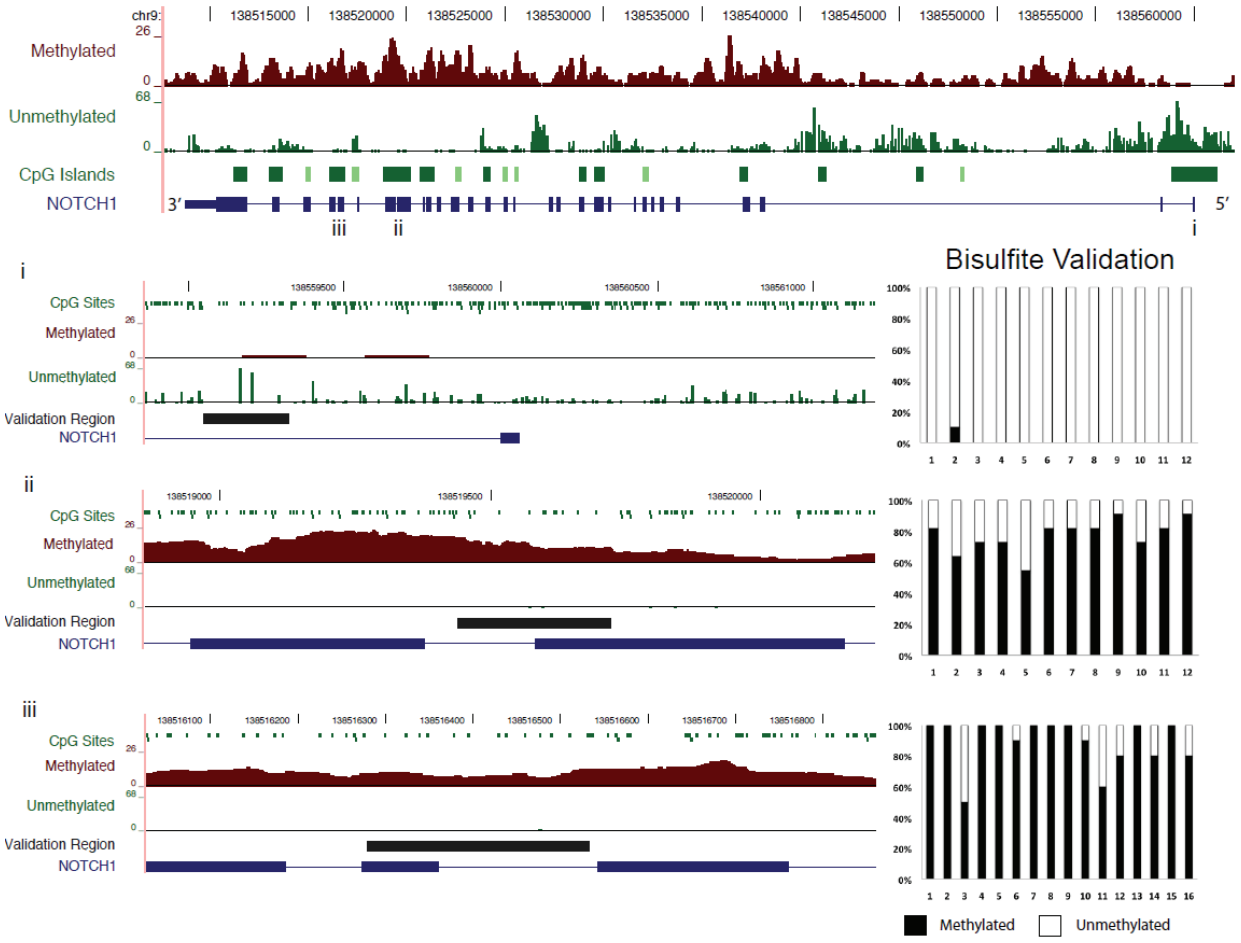

chr9:
138515000
138520000
138525000
138530000
138535000
138540000
138545000
138550000
138555000
138560000
Methylated
Unmethylated
CpG Islands
NOTCH1
3′
5′
iii
ii
i
Bisulfite Validation
CpG Sites
Validation Region
Methylated
Unmethylated

Figure S10i

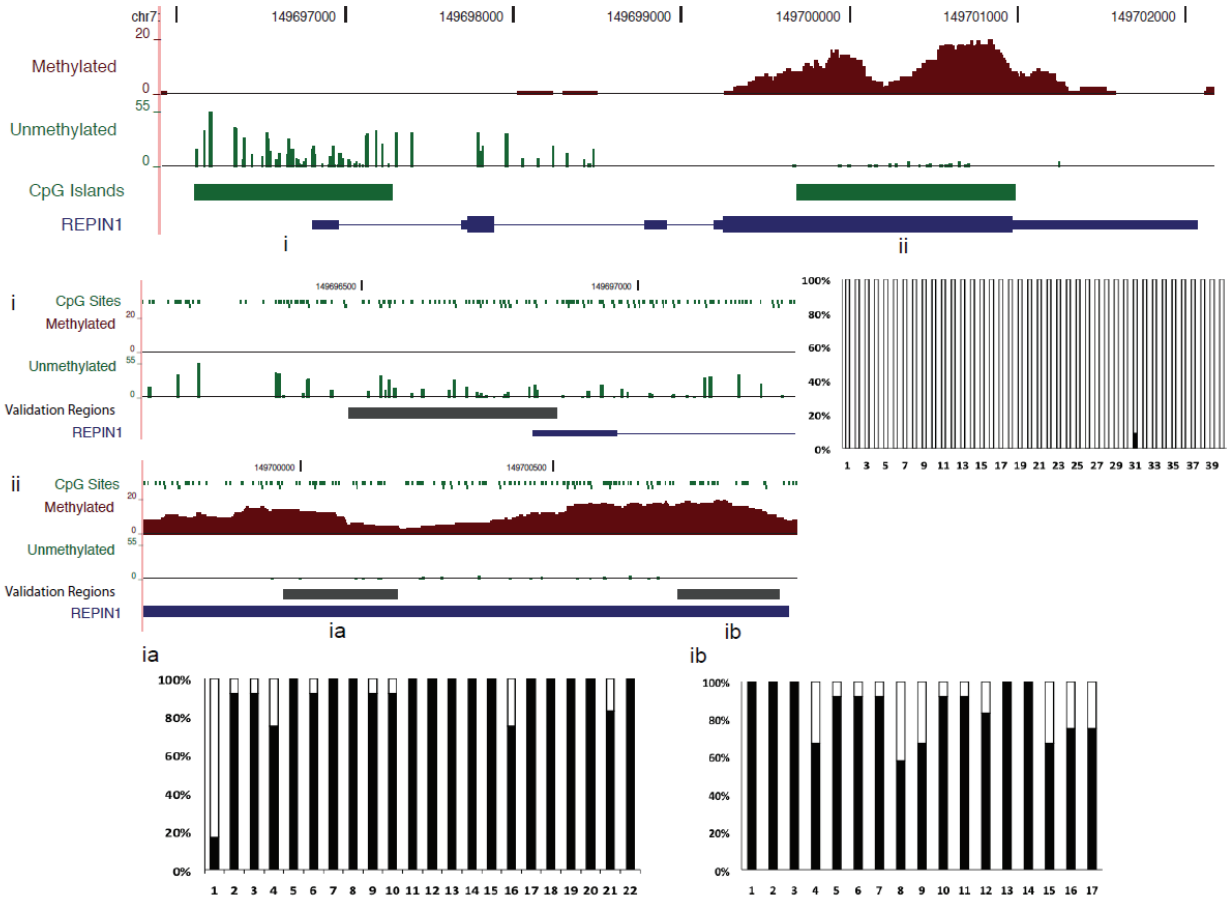

chr7:
149697000
149698000
149699000
149700000
149701000
149702000
Methylated
Unmethylated
CpG Islands
REPIN1
i
ii
CpG Sites
Validation Regions
149696500
149697000
149700000
149700500
ia
ib
100%
80%
60%
40%
20%
0%

Figure S10j

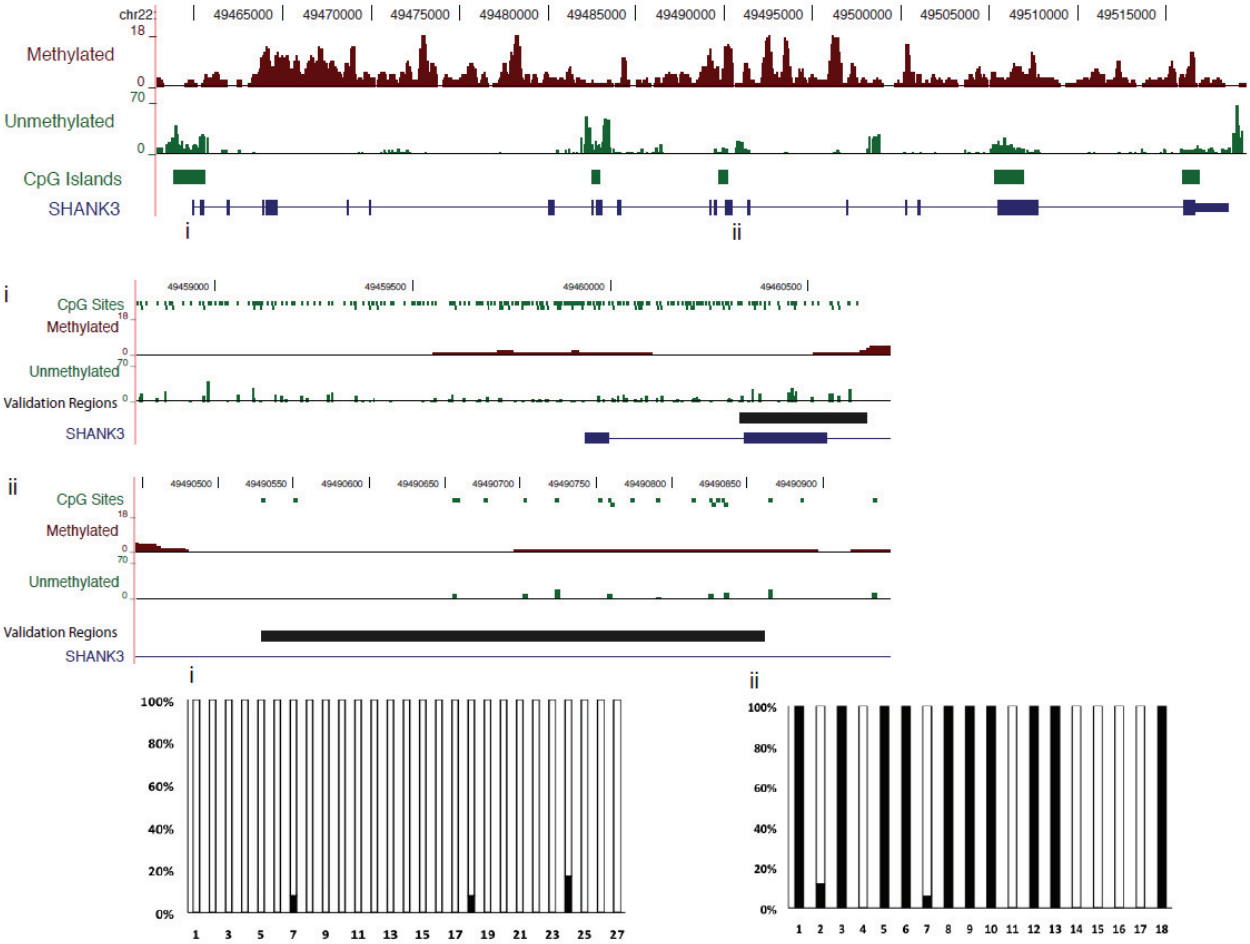

chr22
49465000
49470000
49475000
49480000
49485000
49490000
49495000
49500000
49505000
49510000
49515000
Methylated
Unmethylated
CpG Islands
SHANK3
i
ii
CpG Sites
Validation Regions
100%
80%
60%
40%
20%
0%

Figure S10k

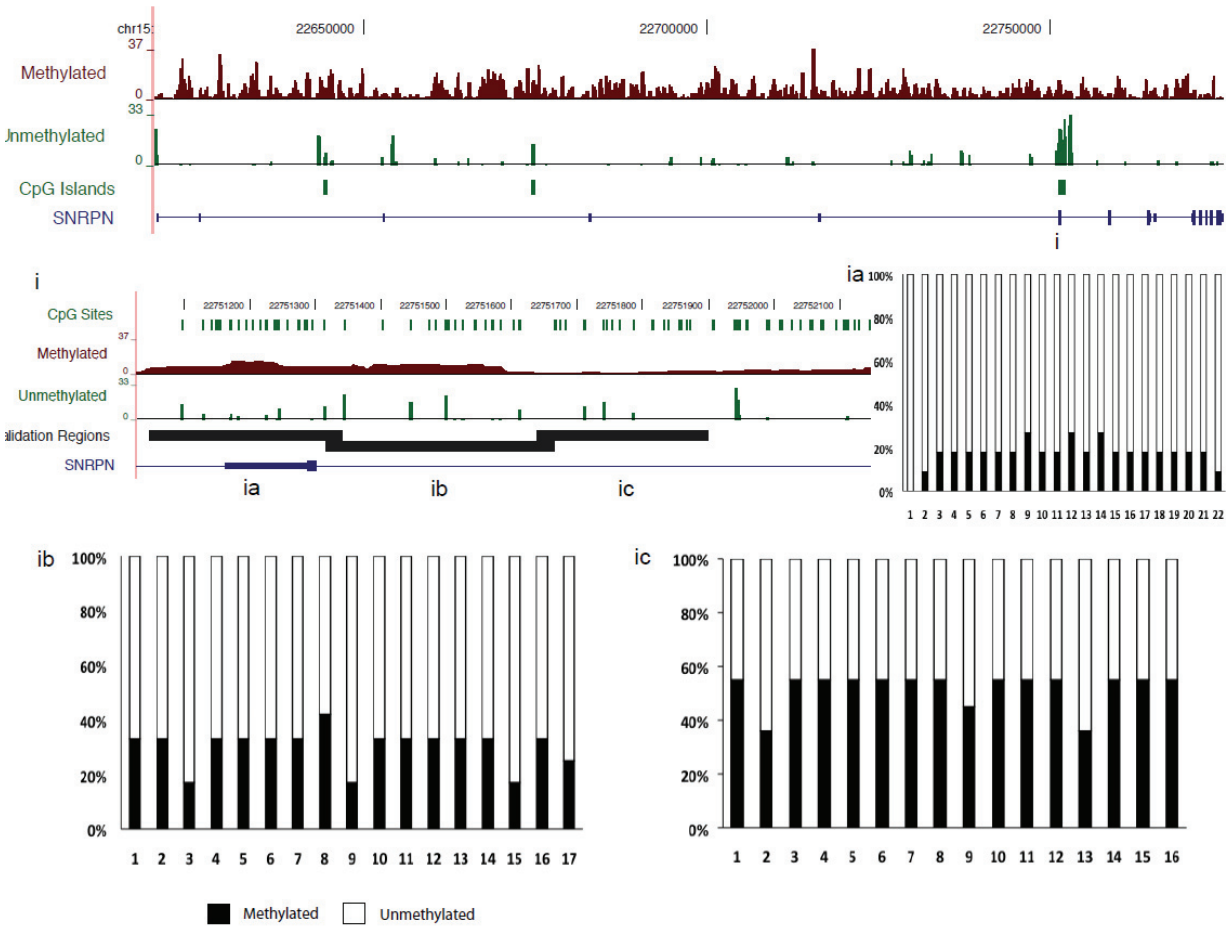

chr15
22650000
22700000
22750000
Methylated
Unmethylated
CpG Islands
SNRPN
i
ia
CpG Sites
Methylated
Unmethylated
SNRPN
ia
ib
ic
ib
ic
Methylated
Unmethylated

Figure S10l

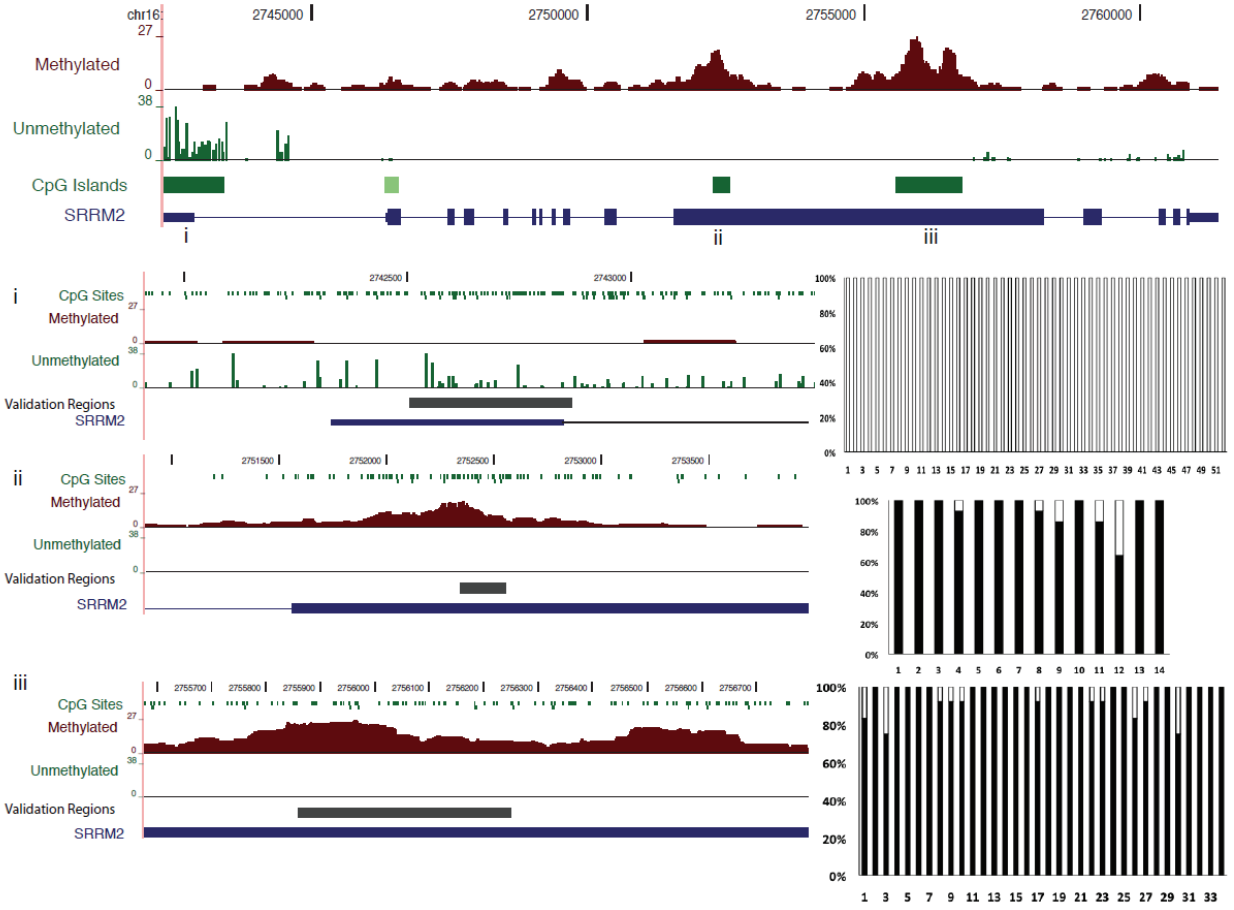

chr16
2745000
2750000
2755000
2760000
Methylated
Unmethylated
CpG Islands
SRRM2
i
ii
iii
CpG Sites
Validation Regions
2742500
2743000
2751500
2752000
2752500
2753000
2753500
2755700
2755800
2755900
2756000
2756100
2756200
2756300
2756400
2756500
2756600
2756700
100%
80%
60%
40%
20%
0%

Figure S10m

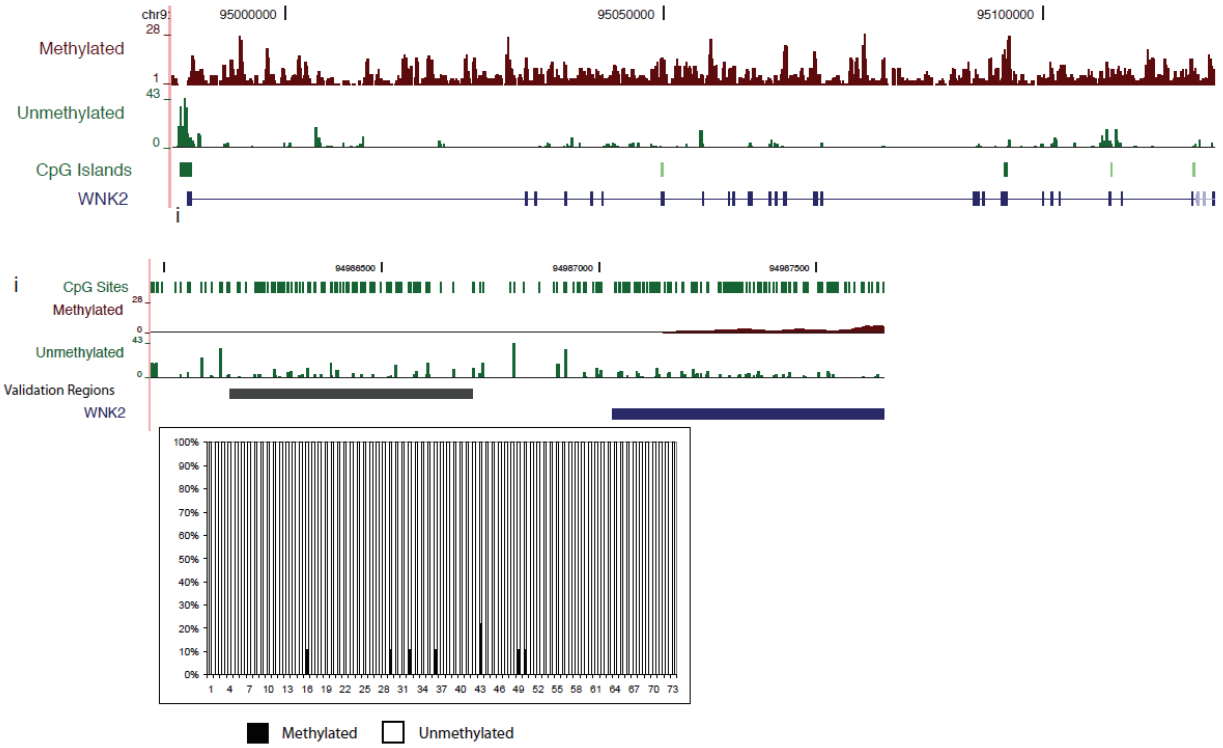


Figure S10a-m. Bisulfite sequencing and validation of human brain MeDIP- and MRE-seq data for 13 selected loci. We treated total genomic DNA with sodium bisulfite for 16 hours and carried out PCR using primers listed in Supplementary Methods, and cloned products into pCR2.1/TOPO (Invitrogen). We selected ≥ 10 individual colonies and sequenced inserts using the ABI 3700 automated DNA sequencer. Methylation is graphed for each assayed CpG site as percent methylation. Crude bisulfite data are shown in Supplementary Excel File 3.

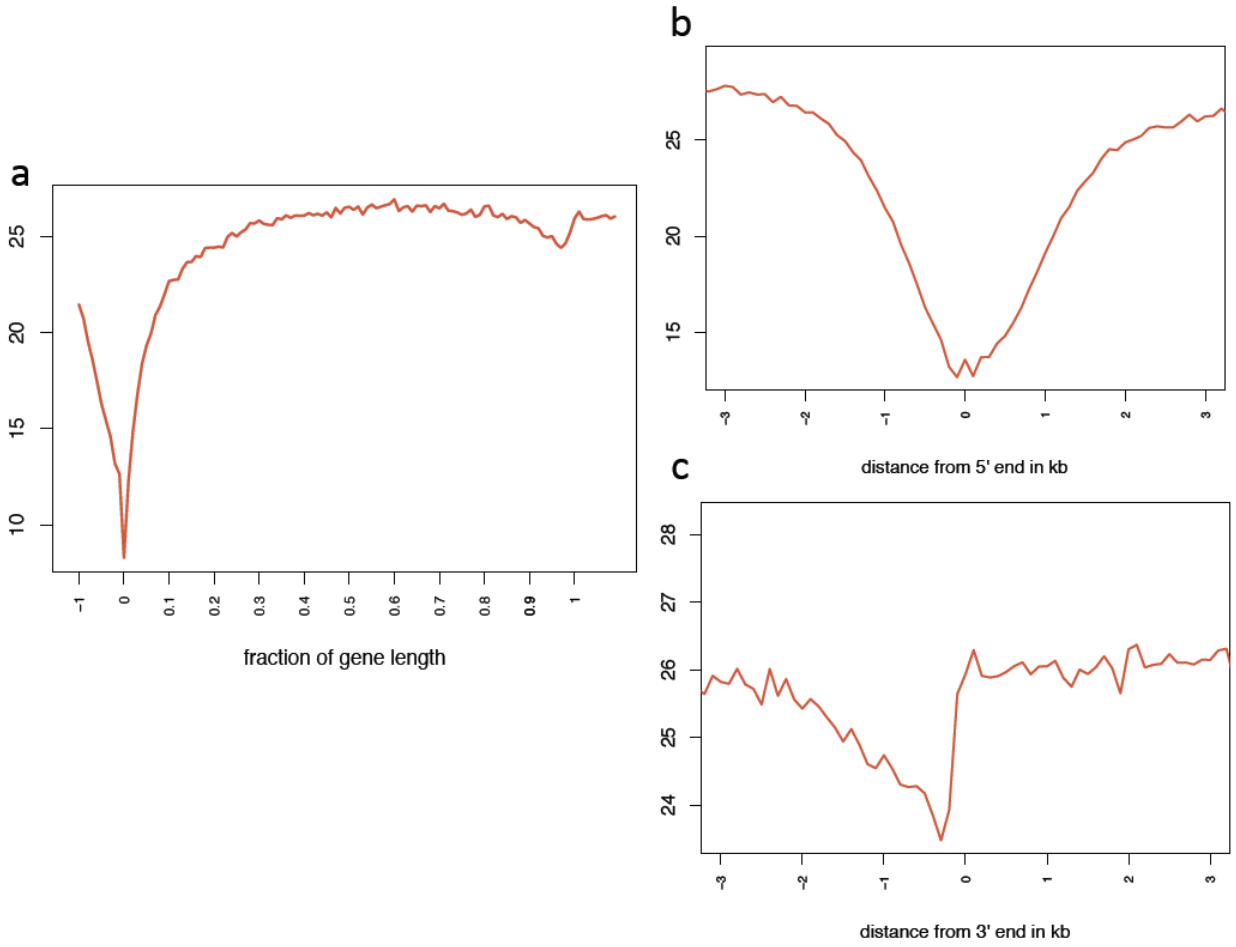


Figure S11. Gene-level plot of average DNA methylation levels. a, Average methylation level is plotted across all known genes, with data binned according to fractions of the total length of each gene. DNA methylation is high across the gene body, and decreases dramatically around transcription start sites, consistent with previous reports[1,2]. Interestingly, we also observe a decrease of methylation at 3' ends of genes, although the magnitude of this decrease is much smaller compared to 5' ends. Close up views of 5' (**b**) and 3' ends (**c**) reveal that decreased methylation is symmetric around 5'ends with a span of about 2-4kb, while at 3' ends, methylation levels start to drop 2kb upstream of 3'end and return to high levels immediately after the gene ends.

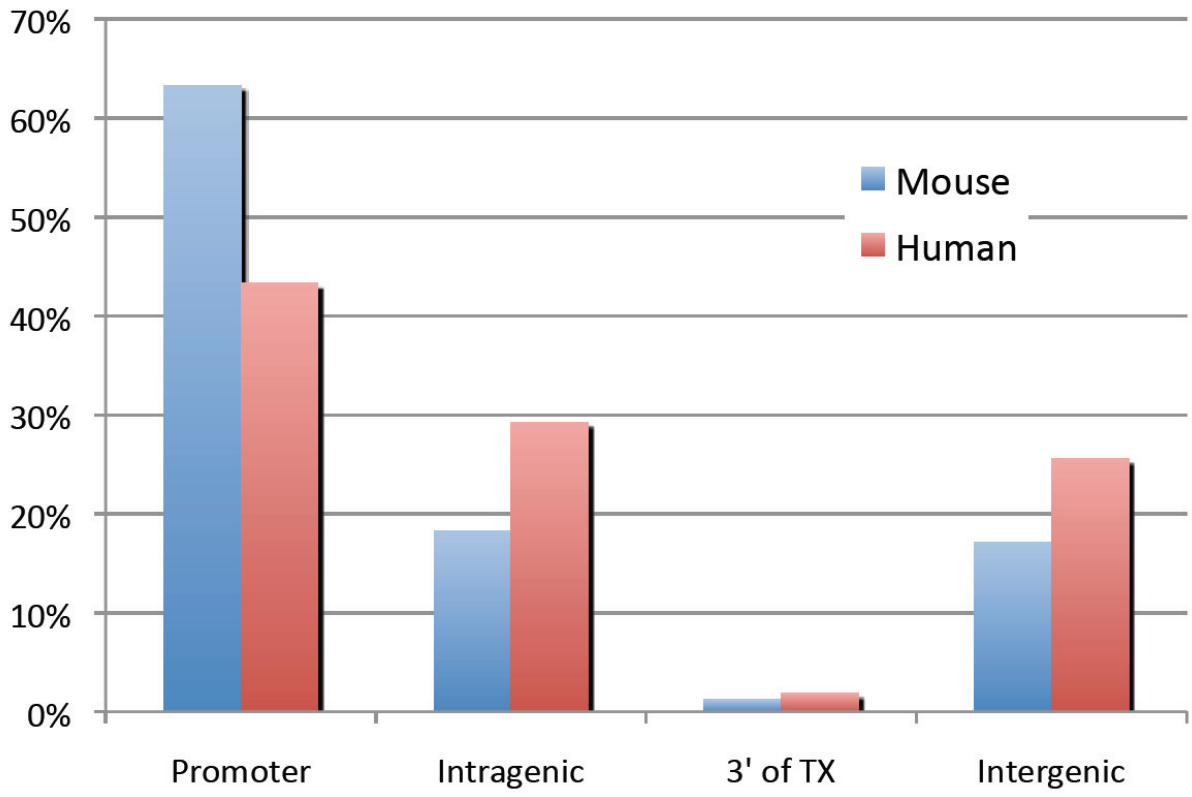


Figure S12. Categorization of mouse and human CpG islands (CGIs).
28226 human and 15948 mouse CGIs were classified into 4 types based on distance to genes: Promoter: -1000 ~ 300 of start of gene; Intragenic: 300 ~ -300 (within body of a gene); 3': -300 ~ 100 (around 3'end of gene); and Intergenic: 1000 upstream, or 300 downstream from a gene. The RefSeq geneset was used for all analyses, and overlapping RefSeq records were merged for simplicity. One gene corresponds to one genomic locus, specified by a start site and an end site.

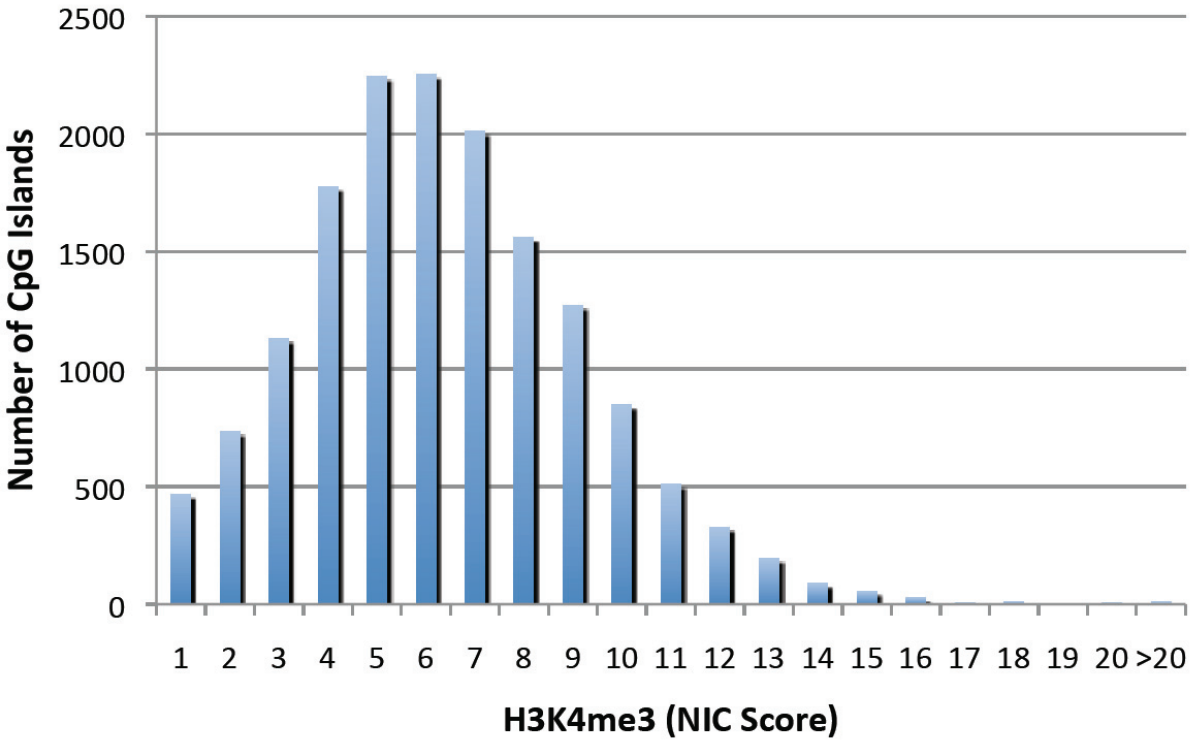


Figure S13. Distribution of H3K4me3 ChIP-seq NIC scores. This is a histogram of CpG islands with specific NIC score for H3K4me ChIP-seq. The x-axis indicates NIC scores. Y-axis indicates number of CpG islands with a specific NIC score. Method for calculating NIC score is described in Supplemental Methods.

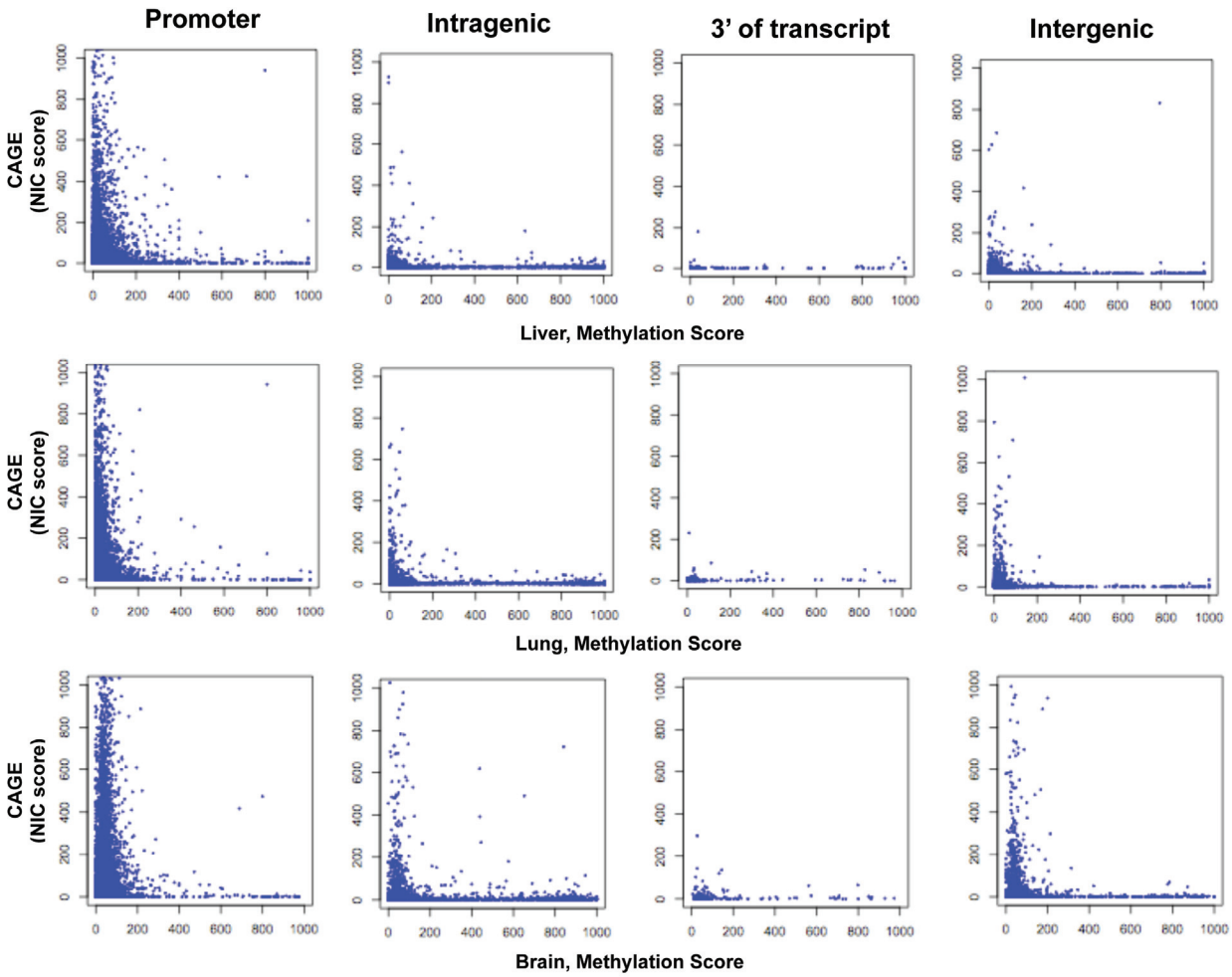


Figure S14. Inverse relationship between DNA methylation and CAGE tags. DNA methylation was measured by RRBS[3] in three cell types. Methylation scores for CpG islands were calculated according to Supplemental Methods. CAGE tags were extracted from three matching cell types based on Carninci et al[4]. An NIC score for CAGE signals over each CpG island was calculated according to Supplemental Methods. A scatter plot between methylation scores and CAGE scores indicates an inverse relationship.

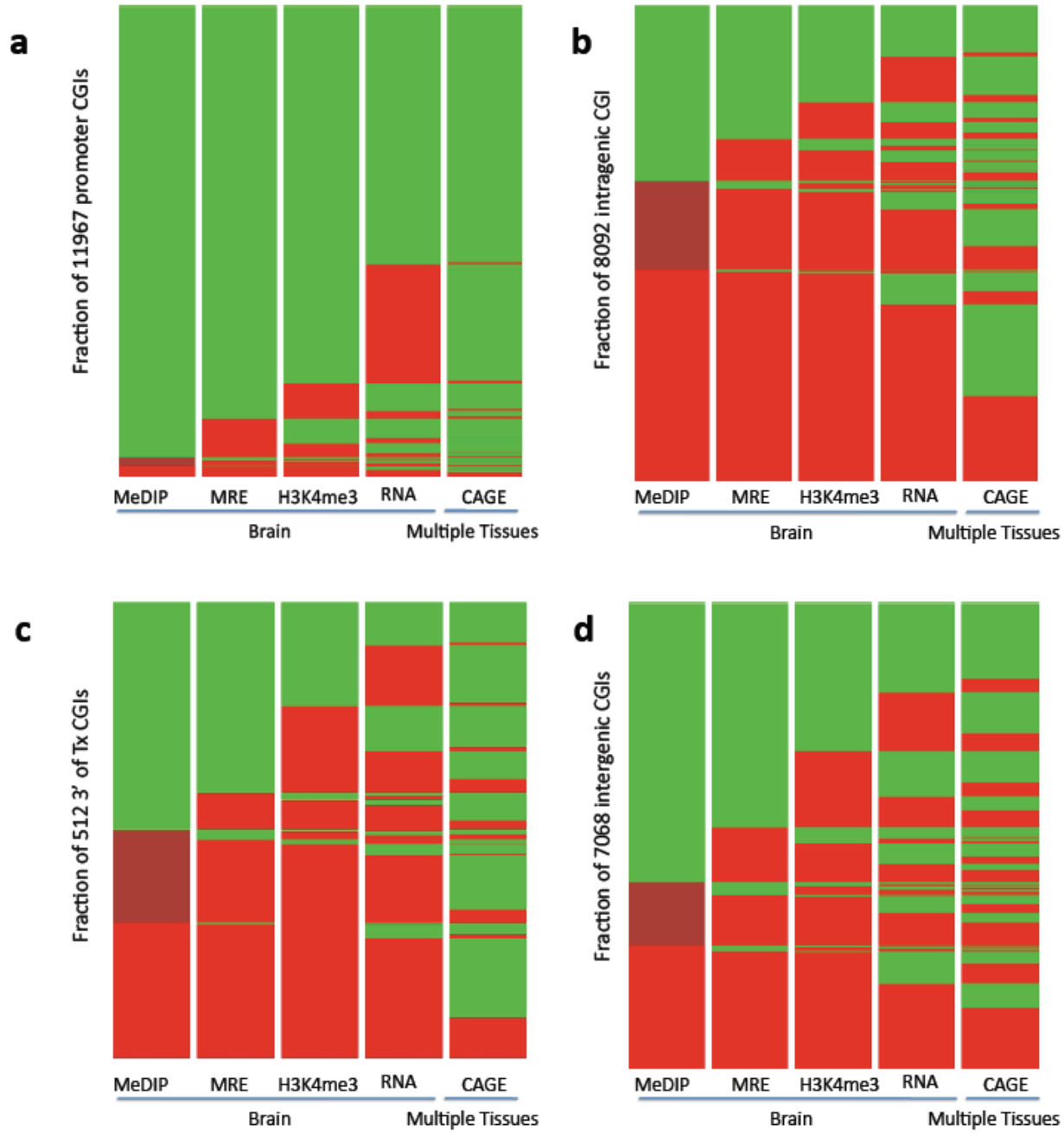


Figure S15. Heatmaps of all human CGIs sorted iteratively with five independent datasets. **a,** Each 5' CGI is coloured according to its status and CGIs are sorted from top to bottom in the order of increasing signal in MeDIP-seq, then within the three MeDIP-defined CGI subgroups by MRE-seq. This process is performed iteratively based on H3K4me3, RNA-seq TSS and CAGE status. For MeDIP-seq, green indicates unmethylated (0-20 reads/kb), maroon indicates partially methylated (20-50 reads/kb), and red indicates methylated (>50 reads/kb); For MRE-seq, green indicates unmethylated (MRE score 0-5), red indicates methylated (MRE score >5); For H3K4me3 ChIP-seq, green indicates active/with signal, red indicates inactive/without signal. For RNA-seq TSS, green indicates evidence for TSS, red indicates lack of evidence for TSS

(Supplemental Methods). For CAGE, green indicates CAGE tags from one or more tissues that overlap the CGI; red indicates lack of overlapping CAGE tags. **b,c,d,** Similar heatmaps for intragenic, 3' and intergenic CGIs, respectively.

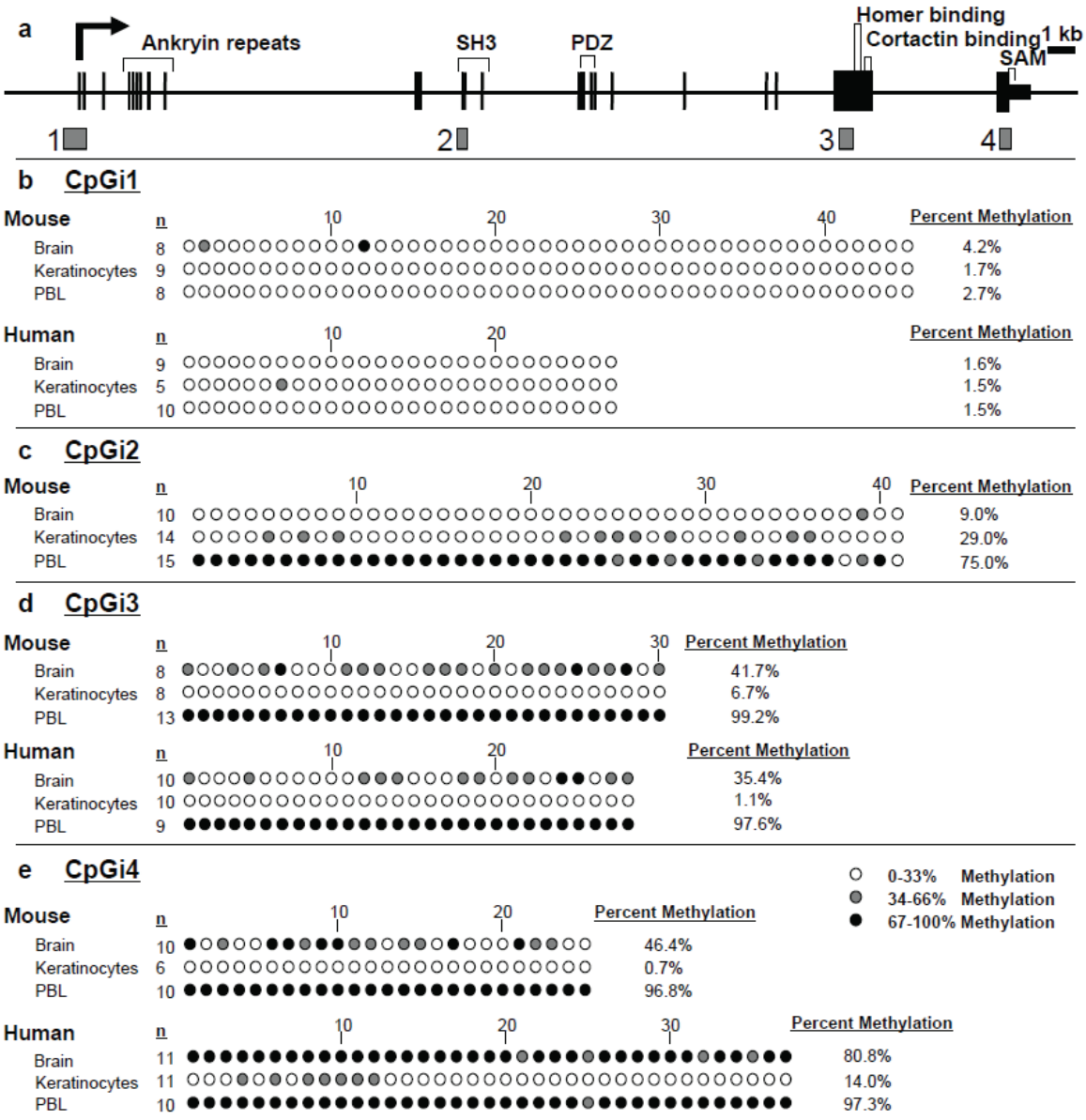


Figure S16. Bisulfite sequencing of matched tissues from mice and humans shows conserved patterns of DNA methylation at *SHANK3/Shank3* CGIs. The tissue-specific DNA methylation patterns of *SHANK3* CpG islands are evolutionarily conserved. **a,** Diagram of the ~60 kb *SHANK3* gene composed of 22 exons (black bars) and four conserved CpG islands (grey bars) as defined by a length≥200bp, an ObsCpG/ExpCpG≥0.65, and a %GC≥50%[5]. Three of the CpG islands are intragenic, while one overlaps with the 5'-most translational start site (arrow)[6]. The known SHANK3 protein domains encoded by each exon are indicated. **b-e,** Comparison of the DNA methylation patterns of the *SHANK3* CpG islands between normal mouse and human tissues. DNA methylation was determined by bisulfite-sequencing of the CpG island regions and

analysis of the indicated number of PCR clones (n). Each circle represents the overall methylation status of an individual CpG cytosine within the analyzed sequence, the color of which corresponds to the extent of methylation in all clones (white: 0-36% methylation; grey: 37-66% methylation; black: 67-100% methylation). The level of DNA methylation from each region is presented as the percentage of methylated CpG cytosines divided by the total number of CpG cytosines within the sequence analyzed from all clones. Sequence similarity was determined by BLAT alignment (UCSC Genome Browser) of the mouse and human genomes to identify orthologous CpG island regions for direct comparison of DNA methylation patterns. These mouse and human CpG island sequences are >80% identical. The presentation of the methylation data is based on 5'-3' orientation from left to right, as in *A*.

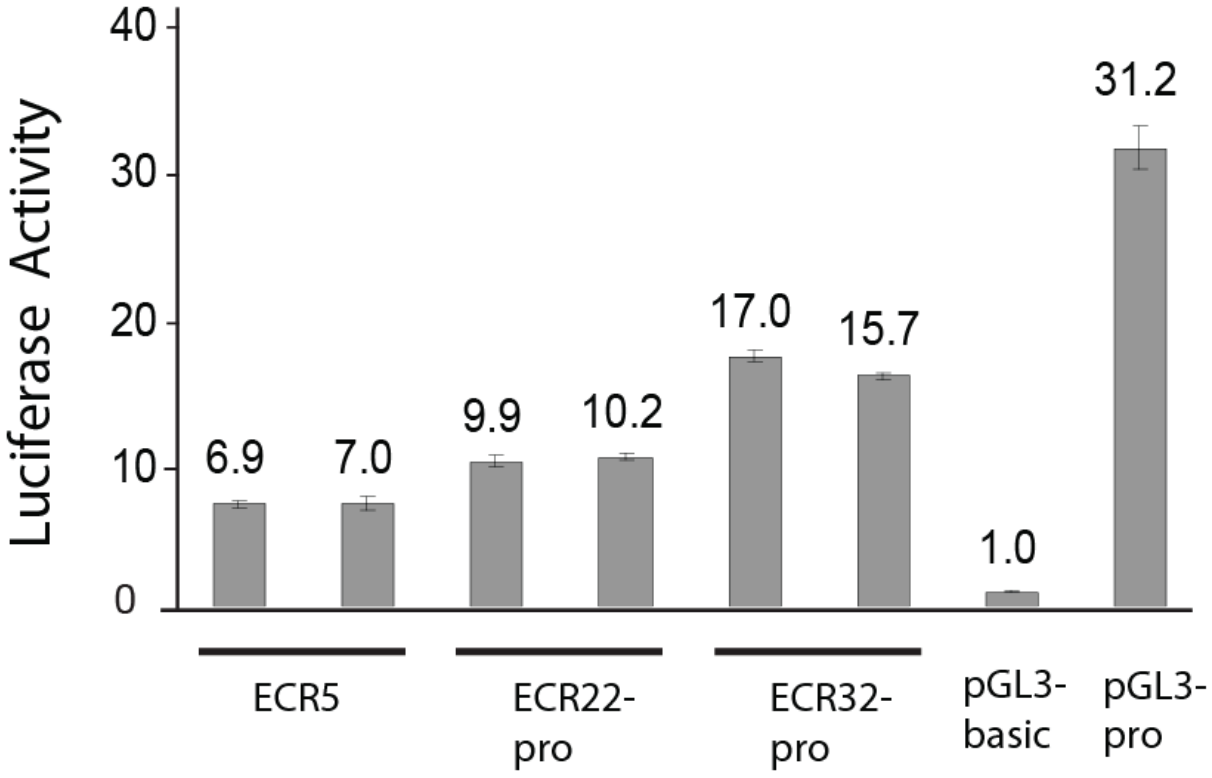


Figure S17. Mouse *Shank3* 5' and intragenic ECRs exhibit promoter activity assessed by dual luciferase promoter-reporter assays (fragments indicated in Fig. 3b). The values above each bar are the normalized ratio of luciferase activity for each construct.

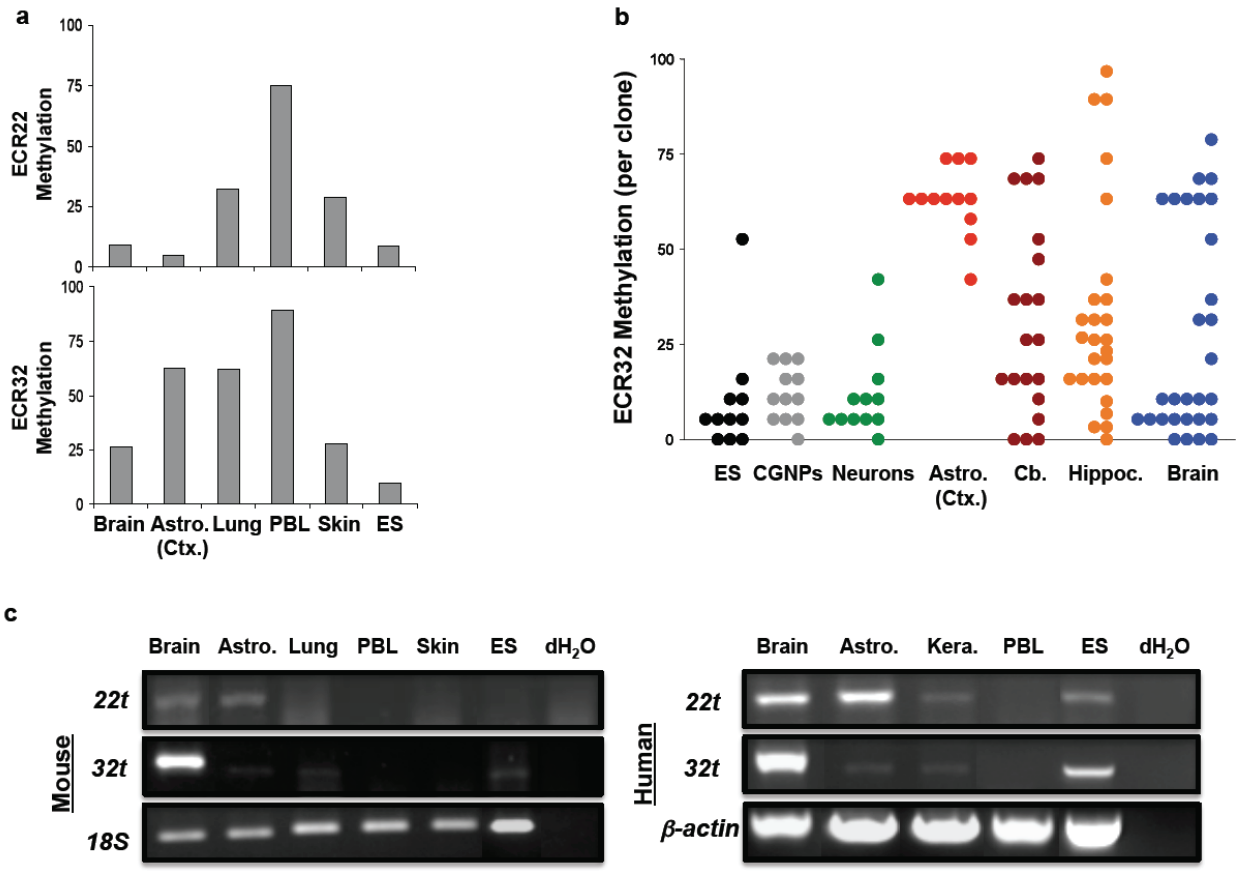


Figure S18. Internal *SHANK3* promoter regions are differentially methylated and their respective transcripts are differentially expressed. a, Overall levels of DNA methylation of the ECR22 (*top*) and ECR32 (*bottom*) promoter regions in tissues or cells. The DNA methylation levels were measured as described in Supplementary Fig. S10. **b,** Comparison of DNA methylation levels of individual alleles in brain tissue and purified cell types. DNA methylation levels are presented for each of the individual clones sequenced across the ECR32 promoter region in mouse tissue/cell types. The cell type-specific methylation patterns are inversely correlated with 32t expression levels (compare with 32t expression in Fig. 3g.). Note that data from cortical (Ctx.) astrocytes are displayed here. **c,** The expression patterns of the novel *SHANK3* transcripts are tissue-specific and conserved between mouse and human. The expression levels of *22t* and *32t* were determined by conventional reverse-transcription (RT)-PCR. Forward primers were designed to specifically recognize and amplify the internal transcripts based on the 5'-RACE sequences, thus amplification of the transcripts confirms the 5'-RACE results. *18S* expression levels were examined in each tissue/cell type as an internal control. Water was used as a negative control for the PCR. Astrocytes are cortical, as in b. Brain, whole brain tissue; Astro. Ctx., astrocytes from cerebral cortex; PBL, peripheral blood lymphocytes; ES, embryonic stem cells; Cb., cerebellar tissue, CGNPs, cerebellar granule neural progenitors; Hippoc., hippocampal tissue; Kera., keratinocytes.

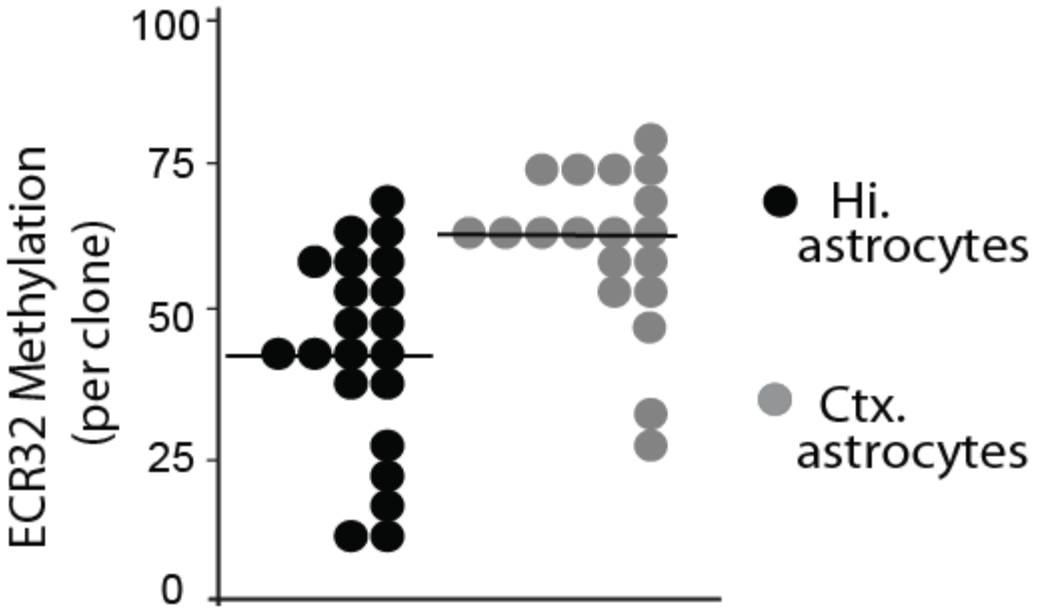
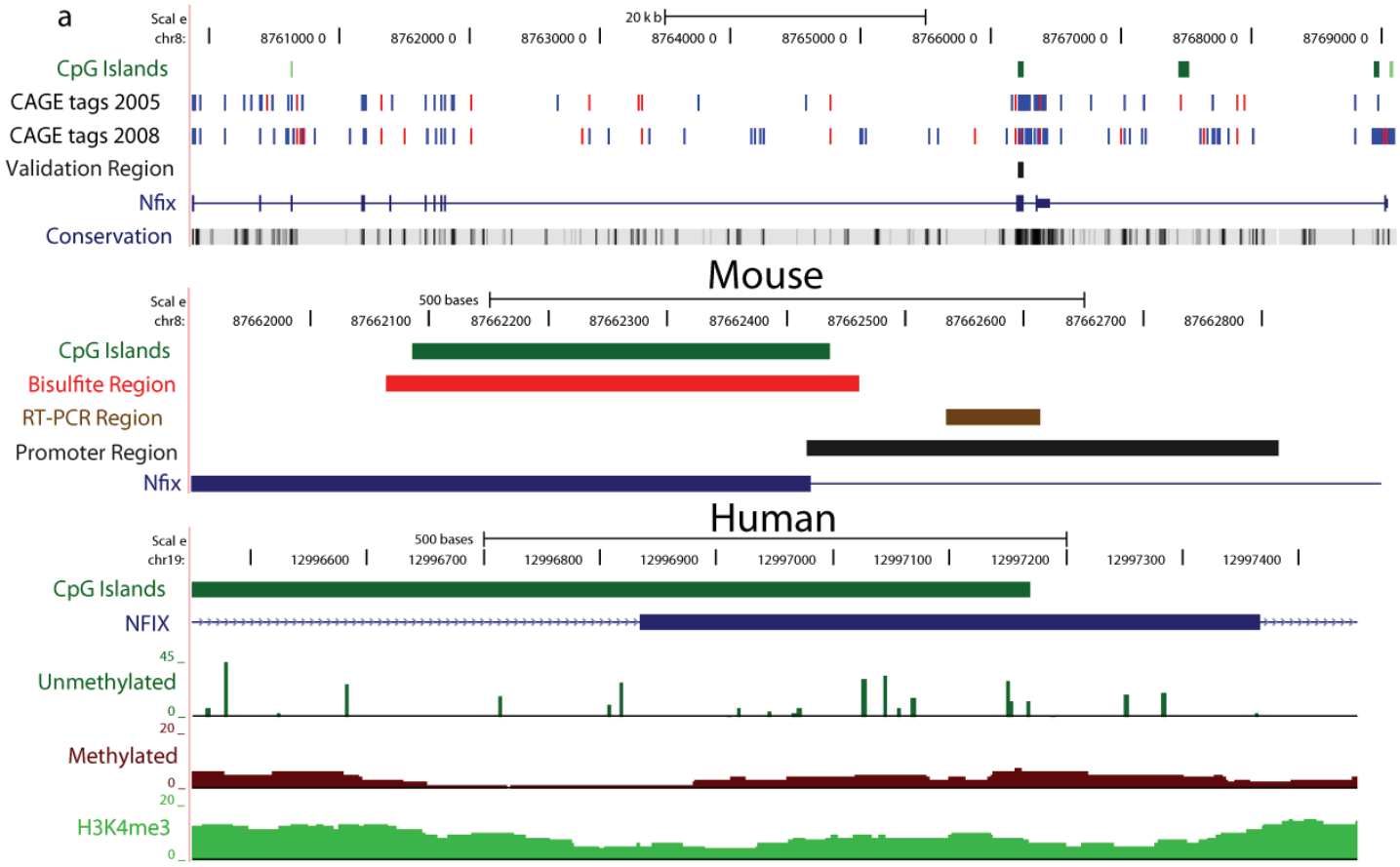
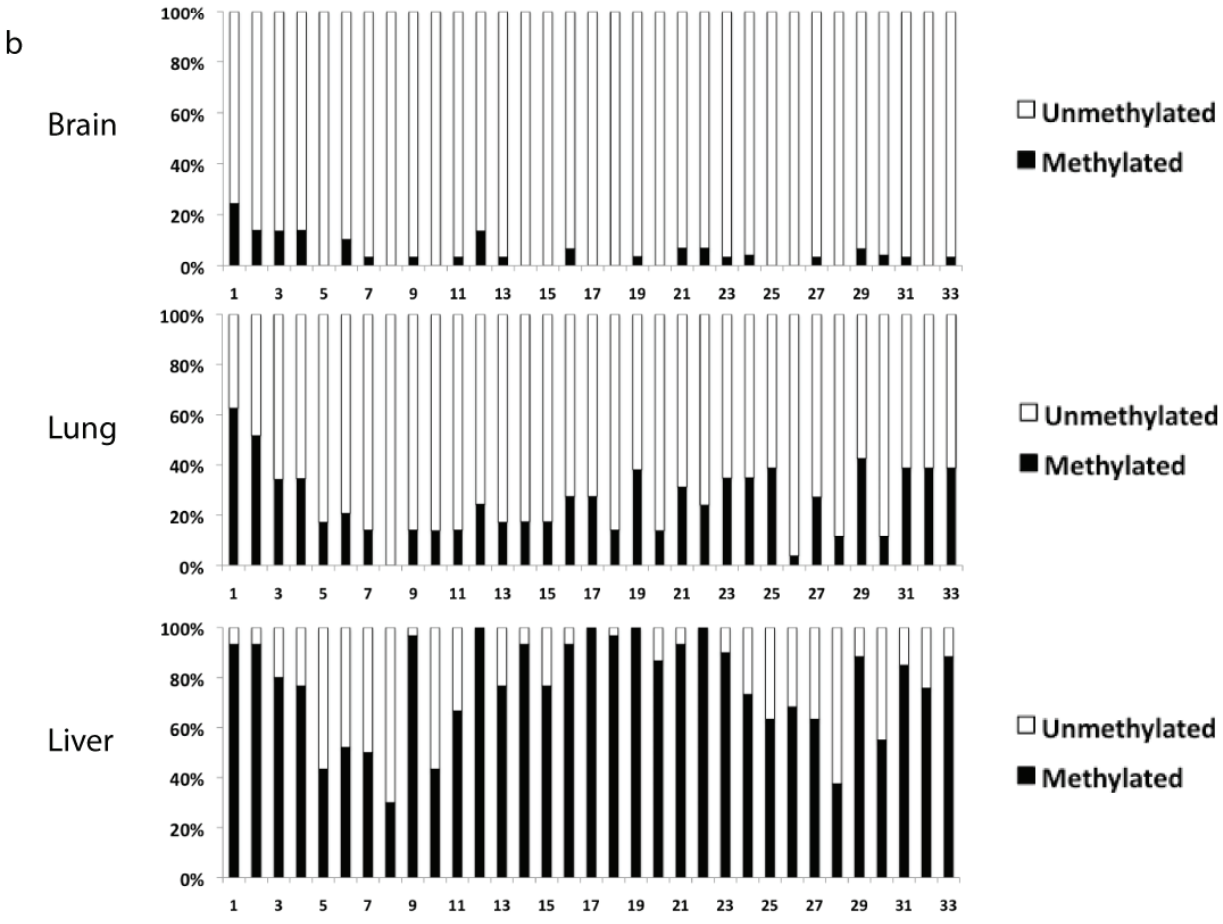


Figure S19. ECR32 DNA methylation differs significantly between morphologically similar astrocytes derived from two distinct brain regions. The percent methylation of each clone from bisulfite-PCR products from the CpG-dense promoter region of ECR32 is plotted from untreated primary cultures of astrocytes derived from normal P7 mouse hippocampal (Hi.) or cortical (Ctx.) brain regions. The overall DNA methylation levels (Hi., 42.6%; Ctx., 60.4% methylation; solid lines) between these two groups are significantly different ($p<0.001$, Student's t-test).


a
Scale
chr8:
20 kb
8761000 0
8762000 0
8763000 0
8764000 0
8765000 0
8766000 0
8767000 0
8768000 0
8769000 0
CpG Islands
CAGE tags 2005
CAGE tags 2008
Validation Region
Nfix
Conservation
Mouse
Scale
chr8:
500 bases
87662000
87662100
87662200
87662300
87662400
87662500
87662600
87662700
87662800
CpG Islands
Bisulfite Region
RT-PCR Region
Promoter Region
Nfix
Human
Scale
chr19:
500 bases
12996600
12996700
12996800
12996900
12997000
12997100
12997200
12997300
12997400
CpG Islands
NFIX
45
Unmethylated
0
20
Methylated
0
20
H3K4me3
0


b
Brain
Lung
Liver
100%
80%
60%
40%
20%
0%
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
33
Unmethylated
Methylated

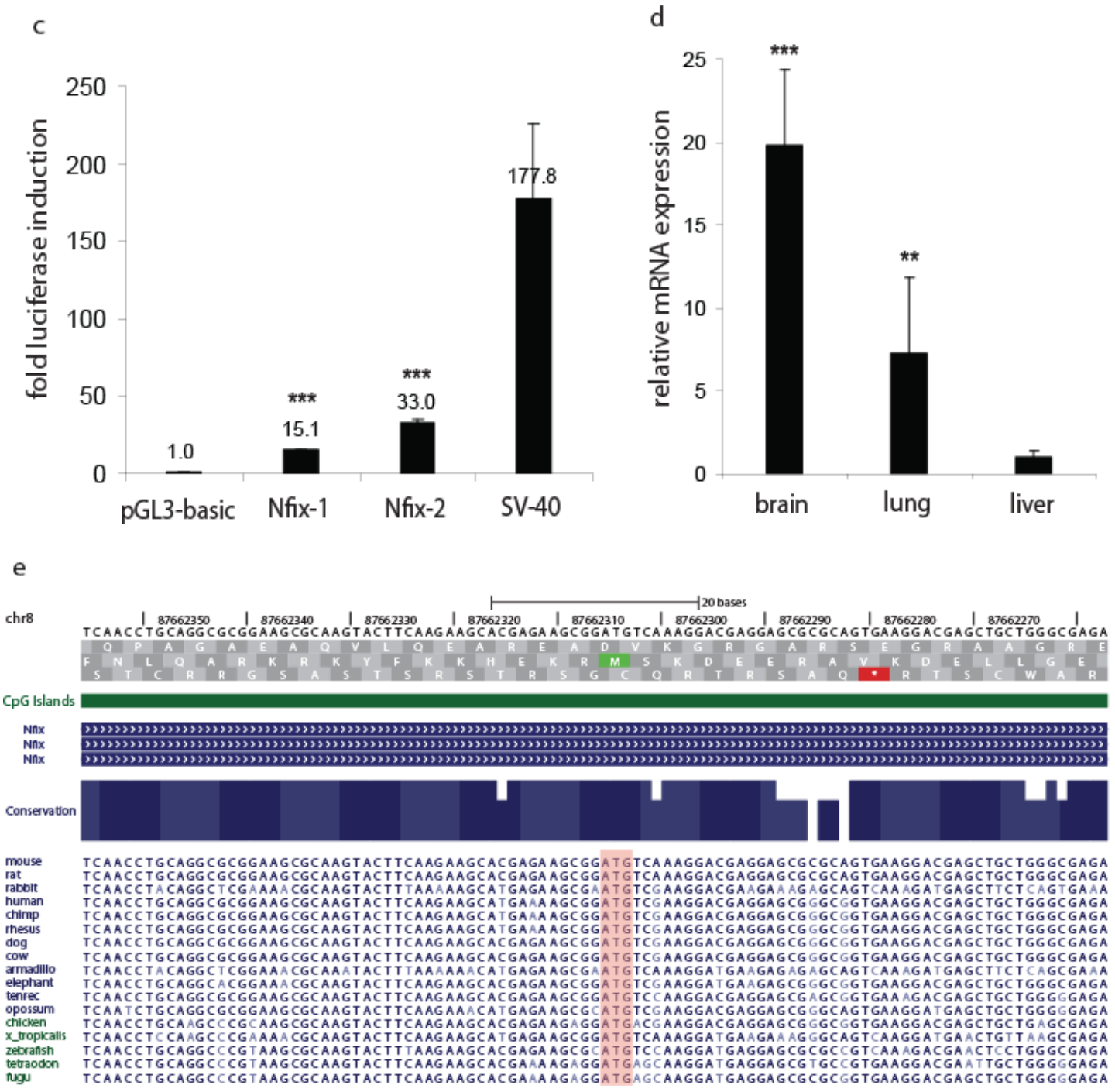


Figure S20. Tissue-specific DNA methylation of a novel intragenic promoter within *Nfix* is associated with expression of a novel transcript isoform. a, Top: UCSC genome browser windows showing the mouse gene *Nfix*, which contains an evolutionarily conserved intragenic CpG island overlapping a cluster of mouse CAGE tags, suggesting transcription initiation. **Middle:** UCSC browser window zoomed in on the mouse *Nfix* intragenic CpG island, with

horizontal bars showing the genomic locations of validation regions for bisulphite sequencing, promoter assay, and qRT-PCR. **Bottom:** MRE-seq, MeDIP-seq and H3K4me3 human brain data for the orthologous human region. **b,** Percent DNA methylation determined by bisulphite cloning and sequencing for individual CpG sites in mouse brain, lung and liver. Two wild-type C57Bl6 mice per tissue were analyzed and the average percent methylation is graphed. **c,** Promoter activity was confirmed in the genomic region partially overlapping the differentially methylated intragenic CpG island by luciferase assay. Fold induction of luciferase activity +/- s.d. relative to pGL3 basic vector is graphed. SV40, positive control. *** $P \leq 0.001$ (compared to empty vector control). **d,** qRT-PCR demonstrates high expression in mouse brain compared to lung and liver of a novel, tissue-specific transcript isoform that initiates within an intron immediately upstream of the differentially methylated intragenic *Nfix* CpG island. Two wild-type C57Bl6 mice per tissue were analyzed and each sample was run in triplicate. Mean fold change +/- s.d. is graphed for each tissue type relative to liver. ** $P \leq 0.01$, *** $P \leq 0.001$. **e,** Evolutionary conservation of a putative intragenic translation start site within *Nfix*. An ATG start codon in-frame with the full-length Nfix protein was found to have a strong predicted Kozak sequence by TIS Miner (http://dnafsminer.bic.nus.edu.sg/)[7]. The ATG start codon shows 100% conservation from humans to Fugu (highlighted in figure).

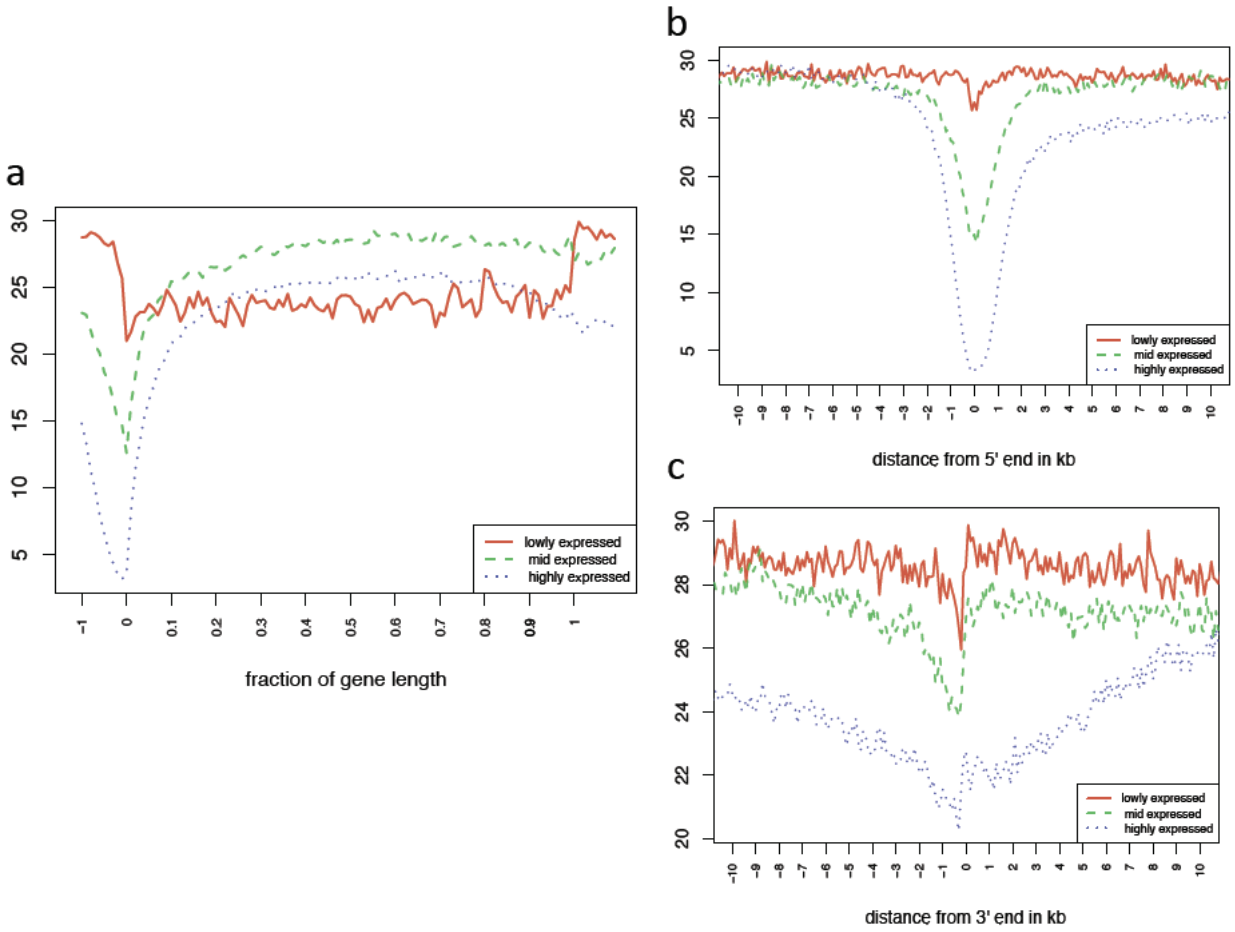


Figure S21. High gene body methylation is associated with moderately expressed genes. Average methylation level is plotted across all known genes, with data binned according to fractions of the total length of each gene. Average gene expression levels were generated from RNA-seq data and genes were divided in high, mid, and low expression. Close up views of 5' and 3' ends are shown in **b** and **c.**

Methods

DNA isolation

Cells were lysed in DNA extraction buffer (50 mM Tris pH 8.0, 0.5% sodium dodecyl sulfate, 0.5 mM EDTA pH 8.0, and 1 mg/ml proteinase K) overnight at 55° C. RNA was removed with RNase treatment (40 µg/ml, Roche DNase-free RNase) for 1 hr at 37° C. DNA was purified with 2 phenol/chloroform/isoamyl alcohol extractions followed by 2 chloroform extractions using phase lock gels. DNA was precipitated with sodium acetate and ethanol, washed with 70% ethanol, and resuspended in TE buffer.

MRE-seq

Three parallel digests were performed (*Hpa*II, *Aci*I, and *Hin*6I; Fermentas), each with 1-5 µg of DNA. Five units of enzyme per microgram DNA were added and incubated at 37° C in Fermentas "Tango" buffer for 3 hrs. A second dose of enzyme was added (5 units of enzyme per microgram DNA) and the DNA was incubated for an additional 3 hrs. Digested DNA was precipitated with sodium acetate and ethanol, and 500 ng of each digest were combined into one tube. Combined DNA was size-selected by electrophoresis on a 1% agarose TBE gel. A 100 – 300 bp gel slice was excised using a sterile scalpel and gel-purified using Qiagen Qiaquick columns, eluting in 30 µl of Qiagen EB buffer. Library construction was performed using the Illumina Genomic DNA Sample Kit (Illumina Inc., USA) with single end adapters, following the manufacturer's instructions with the following changes. For the end repair reaction, T4 DNA polymerase and T4 polynucleotide kinase were excluded and the klenow DNA polymerase was diluted 1:5 in water and 1 µl used per reaction. For single end oligo adapter ligation, adapters

were diluted 1:10 in water and 1 µl used per reaction. After the second size selection, DNA was eluted in 36 µl EB buffer using Qiagen Qiaquick columns, and 13 µl used as template for PCR, using Illumina reagents and cycling conditions with 18 cycles. After cleanup with Qiagen MinElute columns, each library is examined by spectrophotometry (Nanodrop, Thermo Scientific, USA) and Agilent DNA Bioanalyzer (Agilent, USA).

MeDIP-seq

For MeDIP, 5-15 µg DNA isolated as described above was sonicated to ~100-500 bp with a Bioruptor sonicator (Diagenode). Sonicated DNA was end-repaired, A-tailed, and ligated to single-end adapters following the standard Illumina protocol. After agarose size-selection to remove unligated adapters, 2-5 µg of adapter-ligated DNA was used for each immunoprecipitation using a mouse monoclonal anti-methylcytidine antibody (1 mg/ml, Eurogentec, catalog # BI-MECY-0100). For this, DNA was heat denatured at 95 ° C for 10 minutes, rapidly cooled on ice, and immunoprecipitated with 1 µl primary antibody per microgram of DNA overnight at 4° C with rocking agitation in 500 µl IP buffer (10 mM sodium phosphate buffer, pH 7.0, 140 mM NaCl, 0.05% Triton X-100). To recover the immunoabsorbed DNA fragments, 4 µl of rabbit anti-mouse IgG secondary antibody (2.5 mg/ml, Jackson Immunoresearch) and 100 µl Protein A/G beads (Pierce Biotechnology) were added and incubated for an additional 2 hr at 4° C with agitation. After immunoprecipitation a total of 6 IP washes were performed with ice cold IP buffer. A nonspecific mouse IgG IP (Jackson Immunoresearch) was performed in parallel to methyl DNA IP as a negative control. Washed beads were resuspended in TE with 0.25% SDS and 0.25 mg/ml proteinase K for 2 hrs at 55° C and then allowed to cool to room temperature. MeDIP and supernatant DNA were purified using

Qiagen MinElute columns and eluted in 16 µl EB (Qiagen, USA). Fifteen cycles of PCR were performed on 5 µl of the immunoprecipitated DNA using the single end Illumina PCR primers. The resulting reactions are purified over Qiagen MinElute columns, after which a final size selection (192 -392 bp) was performed by electrophoresis in 2% agarose. Libraries were QC'd by spectrophotometry and Agilent DNA Bioanalyzer analysis. An aliquot of each library was diluted in EB to 5 ng/µl and 1 µl used as template in 4 independent PCR reactions to confirm enrichment for methylated and de-enrichment for unmethylated sequences, compared to 5 ng of input (sonicated DNA). Two positive controls (*SNRPN* and *MAGEA1* promoters) and 2 negative controls (a CpG-less sequence on Chr15 and *GAPDH* promoter) were amplified (see Supplementary Materials for primer sequences). Cycling was 95° C for 30 s, 58° C for 30 s, 72° C for 30 s with 30 cycles. PCR products were visualized by 1.8% agarose gel electrophoresis.

ChIP-seq of H3K4me3

Human left hemisphere frontal cortex (Brodmann Area 10) was obtained from the Québec Suicide Brain Bank (QSBB, Montreal, Québec; http://www.douglasrecherche.qc.ca/brain-banks/suicide-bank.asp). All tissue was collected with written informed consent from next of kin. Experimentation with human brain tissue at the Genome Sciences Centre was carried out with approval from the University of British Columbia - British Columbia Cancer Agency Research Ethics Board (REB# H07-01589). For immunoprecipitation of H3K4me3-modified chromatin, human frontal cortex tissue (200-500mg each) from a 57 year old male suspended in chilled douncing buffer (250 µl; 10mM Tris-Cl pH7.5, 4mM $MgCl_2$, 1mM $CaCl_2$), and homogenized by repeated pipetting followed by passing through a 1 ml 26 gauge-syringe 6 times. The homogenate was then incubated with 5U/ml of micrococcal nuclease (Sigma, USA) for 7 min at

37°C (~90% was mononucleosomes after digestion). The reaction was terminated by addition of EDTA (10mM; ~5 µl). To this, 1 ml hypotonic lysis buffer (0.2mM EDTA (pH8.0), 0.1mM benzamidine, 0.1mM PMSF, 1.5mM DTT) with protease inhibitor cocktail was added. The homogenate was incubated on ice for 60 min, with brief vortexing at 10 min intervals. The homogenate was centrifuged at 3000g for 5 min, and the supernatant was transferred to a 1.5 ml non-stick tube. The micrococcal nuclease-digested chromatin fraction was pre-cleared with 100 µl of blocked Protein A/G sepharose beads (Amersham, USA) at 4°C for 2 hrs, and following centrifugation and the supernatant was transferred to fresh tubes. Chromatin immunoprecipitation was carried out either with anti-histone H3 trimethyl K4 (H3K4me3) antibody (ab8580, Abcam), or normal rabbit IgG antibody (12-370, Upstate Biotechnology) to assess fold enrichment. Antibodies were added in manufacturer recommended amounts, and the mixtures incubated at 4°C for 1 hr. To each reaction mixture, 20 µl of Protein A/G beads were added and incubated by rotating at 4°C overnight. Beads were recovered by centrifugation and washed twice with ChIP wash buffer (20 mM Tris-HCl [pH 8.0], 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl) and once with ChIP final wash buffer (20 mM Tris-HCl [pH 8.0], 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl). DNA-antibody complexes were eluted using 100 µl elution buffer (100 mM $NaHCO_3$, 1% SDS), and incubated with 5 µg of DNAse-free RNAse (Roche, Canada) at 68°C for 2 hrs. The beads were pelleted by centrifugation and the supernatant was collected. Elution was repeated with addition of 100 µl of elution buffer and incubation at 68°C for 5 min. After pooling the two eluates, DNA was recovered using the QIAquick PCR Purification kit (Qiagen, Germany). A ChIP-seq library were constructed as previously described[8] using 11-35 ng of immunoprecipitated DNA.

Categorization of CpG islands

We obtained genomic locations of CpG islands from the UCSC Genome Browser for human (hg18, 27639 islands) and mouse (mm8, 15948 islands). We obtained RefSeq gene definition from the UCSC Genome Browser for human (hg18, 29996 genes) and mouse (mm8, 22307 genes). We grouped CpG islands into four classes based on their distance to RefSeq genes. They are:

(1) promoter islands (if an island ends after 1000bp upstream of a RefGene transcription start site, and starts before 300bp downstream of a RefGene transcription start site);

(2) intragenic islands (if an island starts after 300bp downstream of a RefGene transcription start site and ends before 300bp upstream of a RefGene transcription end site);

(3) 3' transcript islands (if an island ends after 300bp upstream of a RefGene transcription end site and starts before 300bp downstream of a RefGene transcription end site);

(4) intergenic islands (if an island starts after 300bp downstream of a RefGene transcription end site and ends before 1000bp upstream of a RefGene transcription start site.

See Supplemental Fig. S12 for number of different classes of CpG islands in the human and mouse genome.

Definition of Islands with no CpG

We identified 94,239 CpG free regions in the human genome assembly (hg18) that span between 1kb to 3kb. We defined the middle 600bp of these regions to be islands with no CpG.

DNA methylation score for the mouse

We obtained reduced representation bisulfite sequencing data from Meissner et al. 2008[3]. We included data on the following cell types in this analysis: Astro_primary_p2, B cell, Brain, ES cell, Liver, Lung, Spleen, T cell CD4, and T cell CD8. Methylation score for individual CpG site is defined as number of CG/(CG+TG) from bisulfite sequencing reads. A CpG site will have a defined methylation score only when CG+TG is equal or greater than 5; otherwise, the score is undefined. Methylation score for individual CpG island is defined as the average score of all CpG sites with a defined methylation score within this island. The score is multiplied by 1000.

A CpG island is defined as completely methylated if its methylation score is equal or greater than 500; as partially methylated if its methylation score is between 100 and 500; and as unmethylated if its methylation score is less than 100.

MeDIP-seq and methylation score for the human

We sequenced the same sample on Illumina GAI and GAII with a total number of reads about 106 million. Redundant reads were removed, and 47 million reads were mapped to the current human genome assembly (hg18) with MAQ. We extended each mapped reads to 200bp in length. Overall, 24 million CpG sites are covered by at least one extended read. We define a methylation score for any region in the genome as number of extended reads per kb. A CpG island is defined as unmethylated if its methylation score is less than 20 reads/kb, as partially methylated if its methylation score is between 20 and 50 reads/kb, and as completely methylated if its methylation score is greater than 50 reads/kb. See Supplementary Fig. S3 for distribution of MeDIP-score across CpG sites and Fig S8 for MeDIP-score across CpG islands.

MRE-seq and MRE-score for the human

We sequenced the same sample with Illumina GAI and GAII with a total number of reads about 20 million. We mapped these reads to the human genome assembly (hg18) with MAQ with an additional constraint that the 5' end of a read must map to the CpG site within a MRE site. This resulted in about 11 million mapped MRE-reads. About 1.5 million CpG sites have at least one mapped MRE-read. We define MRE-score for each CpG site as the number of MRE-reads that map to the site, regardless of the orientation. We define MRE-score for each CpG island as the average MRE-score for all CpG sites that have a score within the island. See Supplementary Fig. S2 for a distribution of MRE-score across CpG sites and Fig S7 for MRE-score across CpG islands.

NIC (Normalized Internal Coverage) score

For any genome-wide data presented in wiggle format, NIC for any given region is defined as the total area of the data profile within the region normalized by the length of the region. See Supplementary Fig. S13 for distribution of NIC scores of CpG islands with respect to H3K4me3.

CAGE association

We used published CAGE data from mouse and human[4,9,10]. Tissue-specific CAGE data is available as "wiggle" tracks. For each CpG island, we extend the island boundary by 200bp in both upstream and downstream directions. If the extended island overlaps with any wiggle signal from the CAGE dataset, we calculate NIC score for the island.

Identifying conserved CpG islands between human and mouse

We first syntenically mapped all human CpG islands to the mouse genome assembly (mm8) and filtered out those that don't map. We further filtered out ones that when mapped to the mouse, they do not overlap annotated CpG islands. Next, we compared classification of these islands (promoter, intragenic, 3' of transcript or intergenic) and filtered out those pairs whose classifications do not match. This results in 2400 pairs of conserved CpG islands between human and mouse, 500 of which are intragenic.

RNA-seq; Identification of putative transcription start sites; Gene expression measurements

100 ng of total RNA was used to synthesize full-length single-stranded cDNAs using the SMART PCR cDNA Synthesis Kit (Clontech, Mountain View, CA, USA) following the protocol as described by Morin et al[11]. The resulting double-stranded cDNAs was assessed using an Agilent DNA 1000 series II assay (Agilent, Mississauga ON, Canada) and Nanodrop 7500 spectrophotometer (Nanodrop, Wilmington, DE, USA). Sonication was performed for a total of 50 minutes using Bioruptor UCD-200 (Diagenode Inc. Sparta, NJ, USA). The sheared cDNA was size separated by 8% PAGE and the 200-250bp DNA fraction excised and eluted from the gel slice overnight at 4 °C in 300 µl of elution buffer (5:1, LoTE buffer (3 mM Tris-HCl, pH 7.5, 0.2 mM EDTA)-7.5 M ammonium acetate), and purified using a QIAquick purification kit (Qiagen, Mississauga, ON, Canada). The library was constructed following the Illumina genomic DNA paired end library protocol with 10 cycles of PCR (Illumina Inc., Hayward CA, USA). The resulting PCR product was purified using 8% PAGE to remove small products including adapter dimers, and the DNA quality was assessed using an Agilent DNA 1000 series II assay

and quantified by Qubit fluorometer (Invitrogen, Burlington, ON, Canada) and then diluted to 10nM. The final concentration was double checked and determined by Quant-iT dsDNA HS Assay Kit using Qubit fluorometer (Invitrogen). Cluster generation and paired-end sequencing was performed on the Illumina cluster station and Genome Analyzer following manufacturer's instructions (Illumina Inc., Hayward CA, USA).

In total, 93 million paired-end reads (186 million reads) were generated for the frontal cortex WTSS-lite library. Custom scripts were used to identify 56.4 million reads that contained the SMART oligo sequence and a variable G stretch (added by the RT terminal transferase activity) on the 5' end. Putative TSS were found by identifying WTSS reads containing sequence corresponding to SMART oligo tags, clipping these tags informatically, and aligning the resulting sequence tag (representing the 5' end of a full-length mRNA) using Maq[12]. In detail: paired end reads were split into forward (read1) and reverse (read2) reads. Read 1s were parsed for those which contained reads starting with the SMART tag followed by a variable number of Gs and clipped after the terminal G. These variable length sequence strings were written to the SMART file (56.4 million reads). All Read2s and those Read1s that did not contain the SMART sequence tag were written to a NOSMART file (129.6 million reads). The SMART file was split into 14 subfiles based on read length and Maq (0.7.1) alignments were run and the resulting .map files merged. The NOSMART file was split into 2 subfiles (for the 75 and 50bp read lengths), and Maq aligned and the resulting .map files merged. The .map files were used to generate SMART and NOSMART wig tracks using FindPeaks 2[13] (xset5; no threshold). For gene expression analysis, the clipped and non-clipped reads were pooled (SMART and NOSMART .map files merged), and read counts generated at the exon and gene level using custom scripts.

To assess promoter activities of individual CpG islands, we first extended each island boundary by 200bp in both upstream and downstream directions and looked for evidence of TSS based on RNA-seq data in these regions. We tallied number of SMART and NOSMART RNA-seq reads overlapping with each island, and defined TSS activity as (1) having at least 5 SMART tagged reads, and (2) at least 70% of total RNA-seq reads are SMART tagged reads.

Normal tissues and cultured primary cells

For the *SHANK3* experiments normal human brain samples were provided from the Neurosurgery Tissue Bank at the University of California San Francisco (UCSF) and we collected adult peripheral blood lymphocytes (PBL) from healthy volunteers. All samples were obtained with informed consent, and their use was approved by the Committee on Human Research at UCSF. Normal human primary adult keratinocytes and normal human fetal astrocytes were purchased from Cambrex and were cultured for fewer than three passages. Normal human ES cells (HSF6) were kindly provided by Mary Firpo while at UCSF. Mouse whole brain, cerebella, hippocampi, lung, pancreas, heart, PBL, and sperm were isolated from normal 8-week old C57BL/6J mice. Keratinocytes from the skin of normal newborn NIH/Ola pups were isolated by physical separation of the epidermal layer from whole skin. In addition to adult stages, brain and lung tissues were derived from mice at pre- and post-natal developmental time points where indicated in the text. Astrocyte monolayers were derived from the postmortem cerebral cortex and hippocampus of postnatal day 7 C57BL/6J mice. The cerebral cortex dissection was preformed in such a way as to exclude all cells of the ventricular or subependymal region. Primary cultures were generated by mincing the tissue and incubating it with papain enzyme, after which cells were filtered through a 70 μm cell strainer. The resulting cell

suspensions were seeded on laminin coated plates in DMEM/F12 medium containing 10% (vol/vol) FCS supplemented with 2 mM glutamine and allowed to grow to confluence. The cells were confirmed to be astrocytes based on morphology and expression of the astrocyte-specific glial fibrillary acidic protein. Mouse ES cells (from C57BL/6J blastocysts) were kindly provided by Miguel Ramalho-Santos (UCSF). All tissue samples were homogenized for isolation of nucleic acids. All cultured cells were collected by trypsinization using 0.25% trypsin-EDTA and washed before cell lysis.

Demethylation and deacetylation experiments

Primary mouse astrocytes were seeded at 1×10^5 cells per well of a six-well plate, incubated for 24 hours in Dulbecco's Modified Eagle's Medium (DMEM) high glucose with 10% serum, and then supplemented with fresh media containing 5-aza-2'-deoxycytidine (5azadC) (1 or 5 μM; Sigma-Aldrich) for 72 hours or trichostatin A (TSA) (100 ng/ml; Sigma-Aldrich) for 12 hours. For the combination treatment, 1 or 5 μM 5azadC was present for 72 hours and TSA was added for the last 12 hours. The media containing drugs were changed every 24 hours.

Bisulfite treatment, PCR and sequencing

We treated total genomic DNA with sodium bisulfite for 16 hours and carried out PCR using primers listed in Supplementary Table 2, and cloned products into pCR2.1/TOPO (Invitrogen). We selected a specified number of individual colonies and sequenced inserts using the ABI 3700 automated DNA sequencer. DNA methylation patterns and levels were determined only from highly (>95%) converted sequences.

5'-Rapid amplification of cDNA ends

Total RNA from brain and lung of normal 8-week old C57BL/6J mice were used to amplify the 5' end of *SHANK3* mRNA with the Gene Racer kit (Invitrogen) based on the protocol supplied by the manufacturer. The mRNA was ligated to the Gene Racer oligo, reverse-transcribed, and amplified using *SHANK3*-specific reverse primers R1 or R2 (Supplementary Table 2) with PfuUltra high-fidelity DNA polymerase (Stratagene) under the following 3-step 'touch-down' cycling parameters: (1) 5 cycles of 94°C for 30 sec, 72°C for 1 min, (2) 5 cycles of 94°C for 30 sec, 70°C for 1 min, (3) 30 cycles of 94°C for 30 sec, 62°C for 30 sec, and 72°C for 1 min, followed by 72°C for 10 min. The amplification products were gel purified, cloned into pCR4-TOPO (Invitrogen), and inserts were sequenced. The sequence data for the novel *SHANK3* transcripts, *22t* and *32t*, have been deposited into the dbEST database and correspond to accession numbers GD253656 and GD253657, respectively. The unique first exon sequences of *22t* and *32t* correspond to chr15:89,354,730-89,355,012 and chr15:89,363,250-89,363,804, respectively (Mouse July 2007 assembly; http://genome.ucsc.edu). Another transcript with a transcription start site downstream of *32t* and lacking the full-length *SHANK3* exon 18 was also identified by 5'-RACE (accession number: GD253658).

Reverse transcription, standard and real-time reverse transcription-PCR

Reverse transcription reactions were performed essentially as previously described[14]. From mouse samples, we measured the expression of full-length *SHANK3* and an internal control *GusB* with probe/primer assays Mm00498775_m1 and Mm00446953_m1 (Applied Biosystems), respectively, by real-time RT-PCR using the Opticon2 Continuous Fluorescence Detector (MJ Research) and calculated relative expression levels using the deltaC_t-method. Expression levels

of *22t* and *32t* were measured by RT-PCR using *18S* and *β-actin* as internal controls for mouse and human samples, respectively. Primers and their corresponding PCR conditions are listed in Supplementary Table 1.

Integration of promoter-associated features at SHANK3

For the *SHANK3* locus (chr15:89,328,288-89,388,754; Mouse July 2007 assembly), we combined three distinct 'features' associated with promoters described in the text. We identified ECRs throughout *SHANK3* using 'ECR Browser': http://ecrbrowser.dcode.org. CAGE tag sequences along *SHANK3* were obtained from: http://fantom31p.gsc.riken.jp/cage_analysis. ECRs with 4 or more CAGE tags are shown with arrows in Fig 3a. ChIP-Seq data of H3K4me3 and H3K27me3 marks across *SHANK3* in ES cells were obtained from: http://www.broad.mit.edu/seq_platform/chip. Because all of these features are sequence-based, we were able to precisely align them in relationship to the corresponding *SHANK3* genomic sequence.

Cloning of ECRs, transfection, and promoter-reporter assays

From mouse or human genomic DNA, selected ECR sequences were PCR amplified with PfuUltra high-fidelity DNA polymerase (Stratagene) using primers designed to contain specific restriction sites (Supplementary Table 2). We subcloned each PCR product into the TOPO-TA cloning vector, selected and sequenced positive colonies, and isolated plasmid DNA containing correct insert sequences. We digested the plasmids, gel-purified the inserts, and re-ligated them into a similarly digested pGL3-Basic vector (Promega). We screened for and confirmed positive

colonies by restriction digestion and sequencing, respectively, and isolated plasmid DNA. Using the FuGENE6 reagent (Roche) and according to the manufacturer's instructions, 1 μg of each construct and 10 ng of an internal control vector (pRL-hTK; Promega) were co-transfected into HEK-293 cells that were cultured in six-well plates containing DMEM media with 10% serum. The pGL3-Basic vector without insert and the pGL3 vector containing an SV40 promoter served as negative and positive controls, respectively. Firefly luciferase and Renilla luciferase activities were each measured 48 hours after transfection by the Dual-Luciferase Reporter Assay System (Promega). As a measure of 'promoter' strength, luciferase activities were calculated from the intensity of light produced as a consequence of beetle luciferin oxidation by Firefly luciferase expressed from each ECR construct relative to that of the promoter-less pGL3-basic vector after normalizing for transfection efficiency as measured by the intensity of light produced as a consequence of coelenterazine oxidation by Renilla luciferase expressed from a co-transfected plasmid. Sequences containing promoter activity within ECR5, ECR22, and ECR32 have been deposited into the GenBank database and correspond to accession numbers FJ215690, FJ215689, FJ215688, respectively.

In vitro DNA methylation assay

Each pGL3-ECR promoter construct was treated with 2 mM S-adenosylmethionine (New England Biolabs) in the presence (methylated) or absence ('mock'-methylated) of 6 units of *M.Sss*I (CpG) methylase per μg of DNA for 4 hours at 37°C. Aliquots of purified constructs were digested with *Hpa*II to confirm the methylation status (data not shown).

Supplementary Methods Table 1: List of Primers used in this study

Species	Primer	Application (template)	Primer Sequence
Human	SNRPN promoter R	MeDIP verification	CGCTCAACACCCCCTAAATA
Human	SNRPN promoter F	(IP DNA)	GGTGGAGGTGGGTACATCAG
Human	MAGEA1 promoter F	MeDIP verification	GTT CCC GCC AGG AAA CAT C
Human	MAGEA1 promoter R	(IP DNA)	GAA CTC TAC GCC GTC CCT CAG
Human	GABRB3 intron CpG-less F	MeDIP verification	CCTGCAACTTTACTGAATTTAGC
Human	GABRB3 intron CpG-less R	(IP DNA)	GGAATCTCACTTTCACCACTGG
Human	GAPDH promoter F	MeDIP verification	CGTAGCTCAGGCCTCAAGAC
Human	GAPDH promoter R	(IP DNA)	GCTGCGGGCTCAATTTATAG
Human	Notch1_BS1_F	bisulfite PCR	AGTGTTTGAGGTTAGTAAGAAGAAG
Human	Notch1_BS1_R	(bisulfite-gDNA)	TCAAAAACCTAAAAAATAAAAAATC
Human	Notch1_BS2_F	bisulfite PCR	GGGGGTTTTTATATTTTTTTT
Human	Notch1_BS2_R	(bisulfite-gDNA)	CTCCAAAACAACCCCCATAC
Human	Notch1_BS3_F	bisulfite PCR	TTTAATGATTTTTGGAAGAATTGTA
Human	Notch1_BS3_R	(bisulfite-gDNA)	TACCTCCAACCCACTAACCAA
Human	BCR_BS1_F	bisulfite PCR	TGTTGTTGTTGTATATGTGTTTTTA
Human	BCR_BS1_R	(bisulfite-gDNA)	AAACACTAATCATCAAATTAATACC
Human	BCR_BS2_F	bisulfite PCR	TTTTTGGGTAGTTAGGTTATGTAGATG
Human	BCR_BS2_R	(bisulfite-gDNA)	ATCAAATTATCCTTCAAAAAAC
Human	MAPK4_BS1_F	bisulfite PCR	TTGTTGATTTTTAATTTTTGGGTTT
Human	MAPK4_BS1_R	(bisulfite-gDNA)	CTCACCCTACAAATCAATACCCTTA
Human	MAPK4_BS2_F	bisulfite PCR	AGGGTTTTTGAGTGATTGGAGT
Human	MAPK4_BS2_R	(bisulfite-gDNA)	CCACTAAACCCCCACCTACTAC
Human	MAPK4_BS3_F	bisulfite PCR	GTTTTTTTTAAAGAAAGGTGGTGAG
Human	MAPK4_BS3_R	(bisulfite-gDNA)	TATTCTACCAACCCCAAAACAAC
Human	NFATC1_BS1_F	bisulfite PCR	TTTTTTGTAATAAGAGGAAGTATAGTTTTA
Human	NFATC1_BS1_R	(bisulfite-gDNA)	ATCTCCCAAATCCAAACTACTATC
Human	NFATC1_BS2_F	bisulfite PCR	GTTAAATTTTGGTTTTAGGTTTGG
Human	NFATC1_BS2_R	(bisulfite-gDNA)	ACCATTTCCAAAAATCTCTACCTAC
Human	NFATC1_BS3_F	bisulfite PCR	TTGGAGTTTTAAAAATAATGTATAGGT
Human	NFATC1_BS3_R	(bisulfite-gDNA)	TCACCATATTAACCAAAATAATCTC
Human	NFATC1_BS4_F	bisulfite PCR	TTGGAGAAAATTAGTTAGTGAAAGG
Human	NFATC1_BS4_R	(bisulfite-gDNA)	TATCAAAACTAACAACCCCAAAACT
Human	MAPK8IP2-i#1BisF	bisulfite PCR	GGTTGTGTAGTTTTTATTGAGTGTTTA
Human	MAPK8IP2-i#1BisR	(bisulfite-gDNA)	AATCCCCCCAAAAACCCTAAC
Human	MAPK8IP2-i#2BisF	bisulfite PCR	GGTATYGAGGTTGATTTGAGAAGT
Human	MAPK8IP2-i#2BisR	(bisulfite-gDNA)	TCACCAAATTACTAAACAACCAAAAC
Human	SHANK3-i#1BisF	bisulfite PCR	GTTTYGTTGTAGAAGTGTTTG
Human	SHANK3-i#1BisR	(bisulfite-gDNA)	TCCCACTTATATTTCCCCACTAC
Human	SHANK3-ECR32OutBisF	bisulfite PCR	ATGTTGAGTTTTGAGTTTTGTGGT
Human	SHANK3-ECR32OutBisR	(bisulfite-gDNA)	TCATCTATCCTAAACACCCCACTAT
Human	SHANK3-ECR32InnerBisF	bisulfite PCR	TTGTGATAAGGAAGTTAGAAGAGGA
Human	SHANK3-ECR32InnerBisR	(bisulfite-gDNA)	AAAACCTCCTAAACCCAAACTATAC
Human	WNK2NIIBisF2	bisulfite PCR	CCTGCCAGTCAGCCAGCCCG

Human	WNK2NIIBisR3	(bisulfite-gDNA)	TCCTGGGCCAGGGCTGGGCAGG
Human	SNRPN_BisMore1F	bisulfite PCR	GTGTTGTGGGGTTTTAGGGGTTTAG
Human	SNRPN_BisMore1R:	(bisulfite-gDNA)	CAAACTATCTCTTAAAAAAAACCACC
Human	SNRPN_BisMore2F	bisulfite PCR	TGGTTTTTTTTAAGAGATAGTTTGG
Human	SNRPN_BisMore2R:	(bisulfite-gDNA)	AAACTACAATCACCCTAATATACCCAC
Human	SNRPN_BisFewF:	bisulfite PCR	TGGGTATATTAGGGTGATTGTAGTTT
Human	SNRPN_BisFewR:	(bisulfite-gDNA)	CCTAATCCACTACCATAACCTCCTC
Human	REPIN1_PromtBisF	bisulfite PCR	TTTAGTTTTTTATAAGGTAGTTAAGGT
Human	REPIN1_PromtBisR	(bisulfite-gDNA)	AATTCACAACTACCAAACCC
Human	REPIN1-intra129-2_BisF	bisulfite PCR	TTTAGTTAGAAGTTTAATTTGGTGT
Human	REPIN1-intra129-2_BisR	(bisulfite-gDNA)	CTAATAAACCAAAAATCTCTCCTC
Human	REPIN1-intra129-1_BisF	bisulfite PCR	TTTTTAGTGTGTTTGTTGTGGTAAG
Human	REPIN1-intra129-1_BisR	(bisulfite-gDNA)	ATAAAACAACAAATTAAATTTATACC
Human	Myo15A_1BisF	bisulfite PCR	TTTGAGGGTTTTTAGGATTTG
Human	Myo15A_1BisR	(bisulfite-gDNA)	CATAAAAATCCAAATAATAACCATAA
Human	Myo15A_2BisF	bisulfite PCR	GAAGTTGGAGGTGTTTTTGTTATTT
Human	Myo15A_2BisR	(bisulfite-gDNA)	AATCCATACCAACCAACTTATAACC
Human	CDH5_is50BisF	bisulfite PCR	GAGATGGATATTATTAGTTA
Human	CDH5_is50BisR	(bisulfite-gDNA)	ACATCTTAAACCTAAATCCCCAATC
Human	SRRM2_PromIs177_BisF	bisulfite PCR	GAAATTTTTTTTGTGTTTGAAAAA
Human	SRRM2_PromIs177_BisR	(bisulfite-gDNA)	TAACTCCCTCACCCTCTCAC
Human	SRRM2_Intr1_Is24_BisF	bisulfite PCR	GTTTAGAATATTTGTTAGG
Human	SRRM2_Intr1_Is24_BisR	(bisulfite-gDNA)	TCTAACTAAAATTCTAAAAC
Human	SRRM2_Intr2_Is105_BisF	bisulfite PCR	AAGGAGAGAGAAAATAAGAATAATT
Human	SRRM2_Intr2_Is105_BisR	(bisulfite-gDNA)	ATCACTAAAATCCTAAACCTTAACC
Human	DCHS1_intra_1BisF	bisulfite PCR	AGATGTGTTAGGAGGGTTGTTTTTA
Human	DCHS1_intra_1BisR	(bisulfite-gDNA)	AAAATCCCCACCAACCTACCTTACC
Human	DCHS1_intra_2BisF	bisulfite PCR	GTTGGTTTTTTGGGAGG
Human	DCHS1_intra_2BisR	(bisulfite-gDNA)	TACACTCACCTACACACTAATAAAC
Human	DCHS1_PromotBisF	bisulfite PCR	TAAAGATTTTTAGGTTTTGTTTTTA
Human	DCHS1_PromotBisR	(bisulfite-gDNA)	ACCTCCCCTAAACTA
Human	PML_intra2_is60BisF	bisulfite PCR	TTGATAGTAGTTATAGTGAGTTTAAGTG
Human	PML_intra2_is60BisR	(bisulfite-gDNA)	TACATATCCAACACCTCCTAATCC
Human	Shank3 CpGi1 F	Touchdown PCR	GTT TYG TTG TAG AAG TGT TTG
Human	Shank3 CpGi1 R	(bisulfite-gDNA)	TCC CAC TTA TAT TTC CCC ACT AC
Human	Shank3 CpGi3 F	Touchdown PCR	GTT TTA TGA TTA TTT TGT AGG ATT
Human	Shank3 CpGi3 R	(bisulfite-gDNA)	CTA CTT CAC CAA AAA ACT CTT AC
Human	Shank3 CpGi4 F	Touchdown PCR	TGG TTG GAG AGT ATT TAT TTA GG
Human	Shank3 CpGi4 R	(bisulfite-gDNA)	ATA AAA ACA CTA AAC AAT ATA CAA AA
Human	Shank3 ECR32 F	Outer PCR	ATG TTG AGT TTT GAG TTT TGT GGT
Human	Shank3 ECR32 R	(bisulfite-gDNA)	TCA TCT ATC CTA AAC ACC CCA CTA T
Human	Shank3 ECR32 F	Inner PCR	TTG TGA TAA GGA AGT TAG AAG AGG A
Human	Shank3 ECR32 R	(outer product)	AAA ACC TCC TAA ACC CAA ACT ATA C
Human	Shank3 *22t* F	RT-PCR	GCT TCC TTG CAT TGC GGG GTT CC
Human	Shank3 *22t* R	(cDNA)	CCG CGT TTC AGG CCG TGT GTC A
Human	Shank3 *32t* F	RT-PCR	CCG CCA GCA CCC TTG ATC GTC TC
Human	Shank3 *32t* R	(cDNA)	GCG GCT CTG GAG TCT GCG TCT GC
Human	β-actin F	RT-PCR	GGA TCT TCA TGA GGT AGT CAG TC
Human	β-actin R	(cDNA)	CCT CGC CTT TGC CGA TCC
Mouse	Shank3 ECR5 F	PCR	(KpnI: GG GGTACC) GCG GGA GAG TGC AGT GGA AGG

Mouse	Shank3 ECR5 R	(gDNA)	(XhoI: CCG CTCGAG) GGC CGA CTG AGA GGG AGA GAC AA
Mouse	Shank3 ECR22 F	PCR	(KpnI: GG GGTACC) ACC ACA GGA GTC AGG CCC AGA G
Mouse	Shank3 ECR22 R	(gDNA)	(XhoI: CCG CTCGAG) CGC CTC CTC CGC CAG CTC
Mouse	Shank3 ECR32 F	PCR	(XhoI: CCG CTCGAG) GCC CTC CTT TCC GAC GCC AGC AG
Mouse	Shank3 ECR32 R	(gDNA)	(HindIII: CCC AAGCTT) CCG CCT AAA GCA CCC AGA CCC A
Mouse	Shank3 CpGi1 F	Touchdown PCR	GTT TTG AAG GTT ATG ATT GAG GG
Mouse	Shank3 CpGi1 R	(bisulfite-gDNA)	CCT ACA ACA AAC RAT AAA CCT TTA C
Mouse	Shank3 CpGi2 F	Outer PCR	TTG TTG GAA TGA TTT AGG GAT ATT T
Mouse	Shank3 CpGi2 R	(bisulfite-gDNA)	AAA ACA ACA AAC AAA AAC ATA AAC C
Mouse	Shank3 CpGi2 F	Inner PCR	TTT TAA ATG GTA ATG AAA TAG GT
Mouse	Shank3 CpGi2 R	(outer product)	TCT ACA ACC CCA ATA AAC CC
Mouse	Shank3 CpGi3/ECR22 F	Outer PCR	GGT ATT TAT AGA GAG GAA TAT AGT GGA A
Mouse	Shank3 CpGi3/ECR22 R	(bisulfite-gDNA)	AAC TTC AAA CCA AAC CTA AAA ACT C
Mouse	Shank3 CpGi3/ECR22 F	Inner PCR	TTG TAG TTA AGG TTG TTT GGT
Mouse	Shank3 CpGi3/ECR22 R	(outer product)	CCA TAA TCA TAC TAC AAA ACT CTA C
Mouse	Shank3 CpGi4 F	Touchdown PCR	GGT TGG AGA GTA TTT ATT TAG G
Mouse	Shank3 CpGi4 R	(bisulfite-gDNA)	CAA ACA ACA ACA AAC AAA AC
Mouse	Shank3 nCGi F	Outer PCR	TTT TTA GTT GGT TTT TTT GTG TGT G
Mouse	Shank3 nCGi R	(bisulfite-gDNA)	TTA AAA CAC CTT CCC ATC AAA TAA C
Mouse	Shank3 nCGi F	Inner PCR	GAG TTG GGA TGA GTT AAG GTT ATG T
Mouse	Shank3 nCGi R	(outer product)	AAA CCT CTC TCT CTC CAC AAT AAA A
Mouse	Shank3 ECR32 F	Outer PCR	TAT TTG AAG AGG GGA TAG TTG GTT
Mouse	Shank3 ECR32 R	(bisulfite-gDNA)	ATT CTA CTC TAA AAA TCC ACC CTA C
Mouse	Shank3 ECR32 F	Inner PCR	TGT TTG ATT TTT GAT TAG GTT TTT G
Mouse	Shank3 ECR32 R	(outer product)	AAT CTC CTA CCA AAC TAC CTC CTA C
Mouse	Shank3 F	5'-RACE (cDNA)	GCA GGG AAG GCA GGT GTG GGT GTA A
Mouse	Shank3 R	5'-RACE (cDNA)	AGG CCC TGG CGC TCA AAC AAT GAG
Mouse	Shank3 *22t* F	RT-PCR	CGC CTT GTG GGG TTC CTG TGT CA
Mouse	Shank3 *22t* R	(cDNA)	CCC CGG AGA ACA AAG CCA AAC CC
Mouse	Shank3 32t F	RT-PCR	GGC AGT GGA GCG CAA GTG GCA TT
Mouse	Shank3 32t R	(cDNA)	GGC CGC CCT CGA GTC AGC ATC T
Mouse	18S F	RT-PCR	AGT GCG GGT CAT AAG CTT GC
Mouse	18S R	(cDNA)	CGC AGG TTC ACC TAC GGA AA
Mouse	Nfix intragenic	bisulfite PCR	TTTTTATATTTGGTTTAATTTGTAGG
Mouse	Nfix intragenic	(bisulfite-gDNA)	CCCATCAATACTTTCCAAAAAAATC
Mouse	Nfix intragenic	bisulfite PCR	AATATTGTGTGTATGTGTGTGTGTAA
Mouse	Nfix intragenic	(bisulfite-gDNA)	CCTACAAATTAAACCAAATATAAAAA
Mouse	Nfix intragenic small prom	PCR	AGCAACATTGTGTGCATGTG
Mouse	Nfix intragenic small prom	(gDNA)	AACGGGTGGAACTCATCCT
Mouse	Nfix intragenic large prom	PCR	AGCAACATTGTGTGCATGTG
Mouse	Nfix intragenic large prom	(gDNA)	CGCTCCCCATCAGTACTTTC
Mouse	Nfix intragenic	qRT-PCR	GTGGCTCCAAACCACACTTC
Mouse	Nfix intragenic	(cDNA)	ACCGCCTTCTTCTCCTTCTC
Mouse	Nfix intragenic	RT-PCR	GAAGGAGAAGAAGGCGGTTT
Mouse	Nfix intragenic	(cDNA)	CGCTCCCCATCAGTACTTTC

References

1. Ball, M.P. *et al.*, Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* **27** (4), 361-368 (2009).
2. Zhang, X. *et al.*, Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126** (6), 1189-1201 (2006).
3. Meissner, A. *et al.*, Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454** (7205), 766-770 (2008).
4. Carninci, P. *et al.*, The transcriptional landscape of the mammalian genome. *Science* **309** (5740), 1559-1563 (2005).
5. Gardiner-Garden, M. & Frommer, M., CpG islands in vertebrate genomes. *J Mol Biol* **196** (2), 261-282 (1987).
6. Beri, S. *et al.*, DNA methylation regulates tissue-specific expression of Shank3. *J Neurochem* **101** (5), 1380-1391 (2007).
7. Liu, H. & Wong, L., Data mining tools for biological sequences. *J Bioinform Comput Biol* **1** (1), 139-167 (2003).
8. Robertson, G. *et al.*, Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **4** (8), 651-657 (2007).
9. Carninci, P. *et al.*, Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38** (6), 626-635 (2006).
10. Valen, E. *et al.*, Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Res* **19** (2), 255-265 (2009).
11. Morin, R. *et al.*, Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* **45** (1), 81-94 (2008).
12. Li, H., Ruan, J., & Durbin, R., Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18** (11), 1851-1858 (2008).
13. Fejes, A.P. *et al.*, FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics (Oxford, England)* **24** (15), 1729-1730 (2008).
14. Hong, C., Bollen, A.W., & Costello, J.F., The contribution of genetic and epigenetic mechanisms to gene silencing in oligodendrogliomas. *Cancer Res* **63** (22), 7600-7605 (2003).