Supplemental data

Table S1, related to Figure 3. (A) Gene list and normalized values for TRAP-Seq, 5hmC-Seq, and MeDIP-Seq. (B) Gene list of RNA-Seq from wt and Mecp2 KO cerebellum. (C) 5hmC-Seq values of all expressed genes from Mecp2 KO and WTGCs.

(Table is provided as a separate Microsoft Excel file titled “Mellen et al. Table S1.xls”).

Table S2, related to Figure 6. (A) List of up-regulated genes in each cell type. (B) Gene list and FPKM values of up-regulated genes in MeCP2 KO cerebellum (C) Gene list and FPKM values of down-regulated genes in MeCP2 KO cerebellum. (D) Gene list and FPKM values of GC enriched, dysregulated genes in MeCP2 KO cerebellum.

(Table is provided as a separate Microsoft Excel file titled “Mellen et al. Table S2.xls”).

Table S3, related to experimental procedures. A-D of this table list the primers used in this study as referenced in the extended supplemental procedures. (Table is provided as a separate Microsoft Excel file titled “Mellen et al. Table S3.xls”).

Figure S1, related to Figure 1. TRAP-Seq quality control

(A) Table showing procedure, number of reads processed and aligned per individual sample indicating sample, origin and number of replica per cell type used in this study.

(B) Scatter plot representation of expressed genes in 2 of the 4 TRAP-Seq replicas per cell type. PC scatter is shown in blue, GC in orange and BG in green. In each panel, Pearson correlation coefficients of TRAP-Seq replicas based on normalized values are shown. F indicates female and M, male. 1 or 2 indicate the number of replica.

(C) Hierachical cluster and heatmap representation of gene expression values per replica used in this study. F female, M male. 1 or 2 indicate the number of replica.

(D) Principal components analysis (PCA) showing the variance between the samples in the present study. PC TRAP-Seq samples in blue, GC orange and BG green.

(E) Left panel: GO analysis of the 86 genes specifically expressed in GC when compare to PC or BG, as show in Figure 1 D. p-value cut off 0.01. GO terms that explain the main biological features of GCs are highlighted. Right panel: Same as left panel for the 134 uniquely expressed genes in BG as shown in Figure 1

Figure S2, related to Figure 2. Genome-wide distribution of 5hmC and 5mC in PCs, GCs, and BG

(A) Saturation analysis of MeDIP-Seq data. 33 million reads per condition show a correlation of 0.90 between two half's of the sample, estimated correlation of 0.95 per sample.

(B) Pie charts showing the percentage of 5hmC peaks corresponding to each entity over all called peaks in each cell type. First chart, PC; second, GC and third, BG.

(C) Chromosomal read densities for 5hmC in PCs (left panel, blue), GCs (middle panel, orange) and BG (right panel, green), input DNA and values expected by chance are also indicated.

(D) Heatmaps where 5hmC FPKM values are represented in the expressed (first column) and non-expressed (NE, second column) entities. From top to bottom row, left panel: TSS ± 2Kb and Gene body. Right panel: TES ±2Kb, CpG islands (CGI) and intergenic region of all genome. First column in each heatmap shows PC values, second GC and third BG. Log_2 values of FPKM are represented.

(E) Funnel plots of FPKM values representing input DNA, 5hmC and 5mC respectively over the gene body throughout the genome ranked by gene expression. Darker color correspond to higher levels of the modification. Left panel shows PC (light blue 5hmC and dark blue 5mC), center panel GCs (light orange 5hmC and dark orange 5mC) and right panel BG (light green 5hmC and dark green 5mC).

Figure S3, related to Figure 3. Bisulfite sequencing and 5hmC, 5mC and TRAP-Seq data for individual PCs, GCs and BG expressed genes

(A) Results of bisulfite sequencing of selected regions in the *Diras2* and *Foxp4* gene loci. Regions *a* and *d* in *Diras2* gene, and regions *a* and *e* *Foxp4* gene are outside differentially modified region. Regions *b* and *c* in *Diras2* and region *e* in *Foxp4* are inside differentially modified region identified by MeDIP. Increase in converted bases indicated the presence of unmodified CpGs (open circles) in regions, which lack mC signal from MeDIP-seq results

(B) Representative examples of individual genes and their values of 5hmC, 5mC and gene expression from the three cell types. Left panel shows a gene enriched in PCs (*Foxp4*), middle panel shows a gene enriched in GCs (*Diras2*) and right panel - one enriched in BG (*Ppap2b*). In each panel, values from PCs are in tones of blue, GCs in orange and BG in green. First line of each group in light colors indicates 5hmC normalized values and second in dark colors, 5mC. Below, TRAP-Seq normalized values are shown per each cell type. The last line in black shows gene location in this particular region of the genome.

Figure S4, related to Figure 4. MeCP2 is the most abundant 5mC and 5hmC binding protein in neurons and glia

(A) The peptide coverage of MeCP2 in mass spectrometry (MS) analysis. The table compares the MS results from the band excised from 5mC affinity pulldown to that from 5hmC affinity pulldown. Orange letters represent the peptide hits in both 5mC and 5hmC pulldown. Green letters represent the peptide that was a hit in 5mC pulldown only.

(B) Expression levelsof the 5mC binding proteins MBD1, MBD2, MBD3, MBD4 and MeCP2 in PCs (blue), GCs (orange) and BG (green). (n=4, mean + SD represented, asterisks indicate p-val < 0.01 in T-test analysis)

(C) EMSA of C, 5mC (M) and 5hmC (H) containing probes made from different sequences in the mouse genome (probe1 [p1], probe 2 [p2] and probe 3[p3]). 1 pmol of MeCP2 (aa 1-205) was used for the experiment. The probes were either reacted with or without β-glucosyltransferase (βGT+) prior to labeling.

Figure S5, related to Figure 5. MeCP2 and other MBD proteins binding kinetics to 5hmC and 5mC containing DNA measured by SPR.

(A) The reference-subtracted dynamic SPR response of flow cells coupled with C, 5mC or 5hmC upon injection of serial dilutions of MeCP2 and MBD2 . The SPR response at the end of injection, shown by the dotted red line, was recorded as steady-state response to the corresponding protein concentration.

(B) Steady-state reference-subtracted SPR response was recorded at the end of 90 s injection o each flow cell immobilized with C, 5mC and 5hmC probes upon application of serial dilutions of MBD1, MBD3 and MBD4;. Values were averaged from at least 4 independent experiments performed in duplicates and plotted as a function of protein concentration. Error bars represent standard error.

(C) The Bmax values from Fig 5C were plotted to visualize the difference in WT MeCP2 and R133C MeCP2.

(D) Table of Bmax values for each protein binding to C, 5mC and 5hmC is shown with standard error and p-values of those calculated by F test to compare variances (right panel).

Figure S6, related to Figure 6. MeCP2 KO GC sorting and 5hmC levels comparison to WT.

(A) Scatterplots from an example of nuclei sorting in MeCP2 KO that allow us to enrich our sample in GC subpopulation of cells. Top panels, ND1 bacTRAP animals where GC were sorted by size that shows 92.4% of GFP positive, GC cells. In the bottom panel, selection of the same population is shown in MeCP2 KO.

(B) Genes were ranked by expression levels and grouped in deciles (from 1, higher expression to 10, lower expression) as shown in figure 6 C. FPKM values of 5hmC are shown in GC WT (orange) and GC KO (yellow) . * p-val <0.05 in T-test. Also, p-values and FC value per decile are shown in the table on the left.

(C) Number of counts per each value of 5hmC FPKM over the gene bodies in GC WT (left panel), GC KO (middle panel) and both merged (right panel).

(D) Scatter plot representation of expressed genes in 2 replicas per genotype (left panel WT and middle panel KO) and between genotypes (right panel). In each panel, Pearson correlation coefficients of replicas based on normalized values are shown. 1 or 2 indicate the number of replica.

Figure S7, related to Figure 7. 5hmC levels and chromatin accessibility over time of digestion

(A) Representative agarose gel photograph showing the MNase digestion pattern over time (from 0 to 60 min) of the DNA analyzed below (300 μg each).

(B) Amount of DNA released by MNase digestion of isolated nuclei was quantified qPCR and percentage of accessibility calculated. We show the correlation between the chromatin accessibility of individual genes and their expression (first row), gene body 5hmC values (second row), gene body 5mC (third row) and 5hmC/5mC values (fourth row) over time (from 0 min, left panel to 30 min, right panel). Pearson correlation values are shown. (n=3)

Extended experimental procedures

Animals. All animal protocols were approved by the Rockefeller University Institute Animal Care and Use Committee in accordance with the National Institute of Health guidelines. Pcp2, Neurod1 and Sept4 bacTRAP animals were generated as described (Doyle et al., 2008), bred to C57BL/6J mice and maintained as trans-heterozygotes. *Mecp2tm1.1Bird* hemizygous males were purchased from Jackson Labs at 4 weeks of age. Animals were sacrifice between 7 to 11 weeks of age for experimentation. All mice were maintained on a 12 h light/dark cycle and given *ad libitum* access to food and water.

Translating ribosome affinity purification and libraries preparation. Single animals per replica were euthanized with CO_2 and cerebella were dissected. 4 cerebella was immediately homogenized in ice-cold homogenization buffer (10 mM HEPES [pH 7.4], 150 mM KCl, 5 mM $MgCl_2$, 0.5 mM dithiothreitol (DTT), 100 µg/ml cycloheximide, Complete-EDTA-free protease inhibitors (Roche) and RNasin RNase inhibitors (Promega, Madison, WI) using a Teflon-glass homogenizer. Homogenates were centrifuged for 10 minutes at 2,000 × g, 4 °C, to pellet cell debris, and incubated with 1% IGEPAL CA-630 (NP-40, Sigma) and 30 mM DHPC (Avanti Polar Lipids, Alabaster, AL) for 5 min. Lysates were centrifuged for 15 minutes at 20,000 × g to pellet insoluble material. At this point, 30 µL of supernatant was saved as input sample. Two custom made mouse anti-GFP (clones 19C8 and 19F7; see Heiman et al., 2008) antibodies were captured on Dynal magnetic beads (Invitrogen Corporation, Carlsbad, CA) coupled with protein L (Pierce, Rockford, IL) . The homogenate was incubated with antibody coupled beads at 4°C with end-over-end rotation for approximately 16 hours. Beads were subsequently collected on a magnetic rack, washed three times with high-salt wash buffer (10 mM HEPES [pH 7.4], 350 mM KCl, 5 mM $MgCl_2$, 1% NP-40, 0.5mM DTT, 100 µg/ml cycloheximide) and RNA was released and purified using Rneasy Micro Kit (Qiagen, Valencia, CA) with in-column DNase digestion. RNA quantity and quality were determined with a Nanodrop 1000 spectrophotometer (Wilmington, DE) and Agilent 2100 Bioanalyzer using RNA 6000 Pico Chip. 4 biological replicas was produced per each cell type. 10 ng of total RNA per IP or input sample were converted to cDNA using the NuGEN Ovation RNA-Seq

(NuGEN,San Carlos,CA,USA) following manufacturers' instructions. The Single Primer Isothermal Amplification (SPIA) method used in this protocol allows the amplification of RNA target in double stranded cDNA under standardized conditions that markedly deplete rRNA without preselecting mRNA. cDNA obtained was quality scored by RNA 6000 PicoChip for Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). 1 µg of cDNA per sample was sonicated using Covaris-S2 system (Covaris Inc., Woburn, MA, USA), cDNA fragments of 200 bp were end-repaired and adapters were ligated for HiSeq 2000 (Ilumina Inc., San Diego, CA, USA) technology using TruSeq DNA Sample kit (Illumina) and following manufacturer`s instructions. Quality of libraries was assesed using HT DNA High Sensitivity Chip (Agilent) for 2100 Bioanalyzer.

Scoring Allen Brain Atlas in situ *hybridizations.* Filtering conditions of the TRAP-Seq raw data was chosen empirically by analyzing the expression pattern using the Allen Brain Atlas (ABA) *in situ* hybridization (ISH) database for adult mouse brain (Lein et al., 2007). Qualitative analysis was performed in the top 200 enriched genes and one every ten to number 500 in each cell type. Two independent observers scored the data considering the following parameters. A gene was considered specific if it has a clear expression by in situ in the particular cell type and not expressed in the rest of the tissue. Genes were classified as not specific when they showed non-specific expression, and *specific in other cells* if they were expressed in other cell types in the brain but not in the one subject to study. Genes that could not be found, were not expressed in the brain or have a bad quality in the ABA database were eliminated from the analysis. The best qualitative scoring (data not shown) was obtained after filtering our data allowing at least 5 reads per gene before performing the differential expression analysis using DESeq package.

Inmunohistochemistry. Animals were transcardially perfused with phosphate buffered saline (PBS) followed by 4% paraformaldehide (PFA) (w/v) for 10 min. Cerebella were dissected and postfixed overnight in 4% PFA. Tissue was sequentially cryopreserved in sucrose 5-10-30% (w/v) in PBS for 12 h each 4ºC, OCT TissueTek embedded (Sakura, The Netherlands) and kept at -20 C. Tissue was cut into 35 µm sections in a cryostat Leica CM 3050 S (Leica Mycrosystems, Buffalo Grove, IL). Free floating immunohostochemistry was performed by the following protocol: Sections were first permeated and blocked in 0.5% Triton X (v/v) (Sigma)

and10% (v/v) normal goat serum (Sigma) in PBS. DNA was denatured for 15 min in 2N HCl followed by 0.1M Sodium Borate pH 9 for 15 min. Sections were incubated with rabbit anti-5hmC antibody (1:500, Active Motif), mouse anti-5mC antibody (1:500, Eurogentech) and chicken anti-EGFP (1:2000, Abcam, Cambridge, MA, USA) in blocking medium overnight at 4ºC. Sections were then washed with PBS and incubatedwith Alexa-fluor secondary antibodies (Invitrogen): chicken A488, rabbit A 568 and mouse A647 in blocking medium. The sections were stained with DAPI and mounted using Vectashield (Vector Labs, Burlingame, CA). Samples were imaged on a Zeiss LSM700 confocal microscope (Carl Zeiss, LLC, Thornwood).

Nuclei sorting and DNA extraction. Isolation and sorting was performed as described in (Kriaucionis and Heintz, 2009). Briefly, cerebella were dissected as described above and homogenized in homogenization medium (0.25 M sucrose, 150 mM KCl, 5 mM $MgCl_2$, 20 mM Tricine pH 7.8, 0.15 mM spermine, 0.5 mM spermidine, EDTA-free protease inhibitor cocktail (Roche) using loose (A) and tight (B) glass-glass dounce.Homogenate was supplemented with 50% iodixanol (OptiPrep, Axis-Shield, Scotland), 150 mM KCl, 5 mM MgCl2, 20 mM Tricine pH 7.8; and laid on a 29% iodixanol cushion. Cerebella from 12 mice (6 males and 6 females) were used for PCs, 6 (3 males and 3 females) for BGs and 4 (2 males and 2 females) for GCs. Nuclei were pelleted by centrifugation 30 min, 10,000 g, 4 °C in swinging bucket rotor (SW41) in a Beckman Coulter XL-70 ultracentrifuge. The nuclear pellet was resuspended in homogenization buffer and co-stained with DyeCycle Ruby (Invirogen) to 20 µM final concentration. Nuclei were sorted in a BD FASCAria (BD Biosiences, San Jose, CA, USA) cell sorter using 635 nm and 488 nm excitation lasers and by gating with two parameters: high GFP signal (compared to wt mice) indicating bacTRAP positive cells and lowest signal for DC Ruby, indicating singlets. Nuclei were fast frozen in liquid nitrogen after sorting and kept at -80 C until analysis.

Alternatively, GC nuclei from WT and *Mecp2*-null mice were sorted by selecting a specific FSC/SSC population identified from GC EGFP+ mice. We observed that this FSC/SSC gated population in GC EGFP+ animals resulted >92% of the sorted nuclei being EGFP positive (Fig S6A). Thus, the gate selected allows us to enrich the sample from 65-70% to 92-96% on GC cells (data not shown).

5hmC pull down: Sorted nuclei from the three cell types were processed in parallel during the whole procedure. Nuclei were digested with Proteinase K 100 μg/mL in the presence of SDS 1% for 2 hours at 50º C. Then, the lysate was treated with RNAse A/T1 mix (Fermentas, Glen Burnie, MA, USA) for 1h 37º C to and DNA was extracted with phenol: chloroform: isoamyl alcohol (25:24:1), precipitated in 70% ethanol and dissolved in Tris-EDTA buffer. 500 μg of purified double stranded DNA was fragmented into approximately 100 bp using Covaris-S2 system. These fragments were end-repaired and ligated with adapters as described in TRAPSeq methods. Libraries before final amplification were resuspended in 18 μL of water. At this point 1 μL was saved as input. Glucosylation reaction was performed at 37 ºC for 1 hour in 10 μL of reaction volume by adding 100 μM of UDP-6-N3-glucose, 2 μM of βGT (kind gift of Dr. Chuan He, University of Chicago) and βGT reaction buffer (50 mM HEPES pH 7.9, 250 mM $MgCl_2$) to 8 μL of DNA library. Negative control reaction was performed by substituting 100 μM UDP-6-N3-Glucose for 100 μM UDP-Glucose (New England Biolabs, Ipswich, MA, USA). . After DNA purification, biotinylation reaction was performed by incubating the DNA with 150 μM dibenzocyclooctyne and disulfide biotin linker (gift from Dr. He) for 2 h at 37 ºC. Biotinylated DNA was purified and pulled down using Dynabeads® MyOne™ Streptavidin C1 (Invitrogen) following manufacturers' instructions. 5hmC enriched DNA was released by incubating the beads in 50 μL of fresh prepared 100 mM DTT in H_2O for 2 h at room temperature using gentle rotation, and beads were removed using a DynalMag-2 magnet (Invitrogen). After DNA purification a final step of amplification was done as described in TRUSeq DNA Sample preparation kit (Illumina). All DNA purification steps followed in this protocol were done using Agencourt AMPpure XP (Beckman Coulter) following manufacturers' instructions.

Bioinformatics and statistical analysis. Principal Components Analysis of four principal components on single samples, hierarchical clustering (Euclidean distance metric), heatmap, scatter plots and Fold change/expression plots for TRAP-Seq in Figure 1 were done in Avadis NGS software version 1.3.0 (Strand Scientific Intelligence).

Saturation analysis and visualization was done using R package MEDIPS (Chavez L and Dietrich, 2010) version 1.4.0

Hierarchical clustering and FPKM heatmaps in Fig S2 E were made by calculating the FPKM for the gene body for each gene for 5hmC and 5mC in each cell type. FPKM values were median centered for each sample/condition and reported as FPKM deviation from the median. Data was clustered using k-means clustering for genes and hierarchical for sample/condition. Data was displayed using Java Treeview (Saldanha, 2004)

Gene and genome feature heatmaps in Fig S2 D were made by calculating the FPKM for the indicated feature. Intergenic regions were defined as regions larger than 100bp that were not within 5kb of any gene TSS or TTS. Values were reported as Log_2 FPKM. Data were displayed using Java Treeview.

Non expressed gene list was built by removing expressed genes from the Ensembl mouse genome annotation. Also Olfr, Vmn groups as well as and poorly annotated genes were also removed.

Funnel plots (heatmap representations of 5hmC and 5mC enrichment ordered by expression), 5kb regions upstream of the TSS and 5kb downstream from the TTS were binned into 50bp windows and enrichment calculated for each gene; each gene body (transcription unit) was segmented into 600 windows and coverage calculated and normalized for gene length. Each alignment was extended 200bp in the 3' direction to account for the average length of sequenced templates. Genes were ordered by expression values derived from TRAP-Seq data for each cell type. Data was displayed using Java Treeview.

Metagene analysis was done using functions from the R package ShortRead (Morgan M. et al 2009) and custom written scripts. Briefly, genomic coverage vector was created by extending all reads to 400 bp (average library size). ENSEMBL genomic annotation coordinates were used to sample the coverage vector for the coverage values in the enriched and input samples. The ratio of enriched/input was used for analysis and data representation.

Genomic views of read coverage were generated using Integrated Genomics Viewer tools and browser (IGV 2.0, http://www.broadinstitute.org/igv/).

Statistical analysis in Figure S4 B was performed in Microsoft Excel using Student's t-test.

Data shown in Fig 3B, 6C, 6E, 7A, S6C and S7 were analyzed by Pearson correlation in Microsoft Excel.

Data shown in Fig 7F and G were analyzed by Wilcoxon Signed-Rank test in VassarStats website for statistical computation (http://vassarstats.net) and box and wishker plots in Microsoft Excel.

Bisulfite sequencing. Bisulfite sequencing was done using Qiagen EpiTeck bisulfite conversion kit according manufacturers recommendations. Bisulfite converted DNA was amplified with the primers in Table S3a.

Preparation of Probes. C, 5mC and 5hmC probes were prepared using either unmodified dCTP (New England Biolabs), 5'-methylated dCTP (Trilink Biotechnologies, San Diego, CA, USA) or 5'-hydroxymethylated dCTP (Bioline, Taunton, MA, USA), along with dATP, dTTP and dGTP (New England Biolabs). A 120 bp region of mouse BDNF promoter was amplified with ChoiceTaq (Denville, South Plainfield, NJ, USA) according to manufacturer's recommendations using 5'-biotinylated or 5'-unmodified forward primers and 5'-unmodified reverse primers (Genelink, Hawthorne, NY, USA) listed in Table S3b. For CG-rich probe we amplified a 200 bp region of the pUC plasmid (Invitrogen). Probes were purified via Qiaquick PCR Purification Columns (Qiagen). 1 pmol of probe was end-labeled with 1 or 10 uCi of γ-P32 dATP (Perkin Elmer, Waltham, MA, USA) using T4 Polynucleotide Kinase (New England Biolabs) for 45 min at 37°C and purified using Illustra ProbeQuant G-50 Micro Columns (GE Healthcare, Little Chalfont, UK). ^{32}P was counted in scintillation fluid Ready Safe (Beckman Coulter) in 1209 Rackbeta scintillation counter (LKB-Wallac Pharmacia, Turku, Finland).

Nuclear protein extracts from cerebella. Nuclear extracts were prepared from 80 frozen rat cerebella (Pel-Freez, Rogers, AR, USA) as previously described (Klose and Bird, 2004). Briefly, rat brains were diluted 5 volumes to 1 in ice-cold Buffer A containing 10 mM Hepes pH 7.5, 25 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 1 mM EDTA, 2 M sucrose, 10% glycerol, and Halt Protease and Phosphatase Inhibitor Cocktail (Pierce Biotechnology, Rockford, IL, USA) followed by homogenization in a 60 mL Dounce (Bronwill Scientific Inc., Rochester, NY) on Glas-Col Tissue Homogenizer (Cole-Parmer, Vernon Hills, IL, USA) (5 strokes at 4,000 rpm). The homogenate was layered onto a 10 mL cushion of the Buffer A and centrifuged in pre-chilled SW2 (Beckman Coulter) rotor at 24,000 rpm in Beckman XL-70 ultracentrifuge for 1 hour at 2 °C. Recovered

nuclei were incubated in 5 volumes of Buffer B containing 20 mM Hepes (pH 7.9), 150 mM KCl, 1 mM EDTA, 0.5 mM DTT, Halt Protease and Phosphatase Inhibitor Cocktail on ice 2 min and then resuspended. The nuclei were counted by Neubauer haemocytometer (Hausser Scientific, Horsham, PA, USA). The nuclear proteins were extracted by stepwise addition of 3M KCl until a final concentration of 400 mM. The extraction was allowed to proceed for 20 min on ice, and then the nuclei were pelleted at 13,000 rpm for 10 min at 4°C. The supernatant was dialyzed in Slide-A-Lyzer cassettes with 10 kDa molecular weight cutoff (Pierce) overnight at 4°C against Buffer B. The protein concentration was measured by Quick Start Bradford Assay (BioRad) according to manufacturer's instructions. Alternatively, nuclear extracts were prepared from 10 wild type (WT) or *MECP2* KO mice using SW41 rotor with a 2 mL cushion of Buffer A.

5hmC affinity pull-down. 1 ug of 5'-biotinylated C, 5mC or 5hmC BDNF probe was immobilized on Dynabeads M-280 Streptavidin (Invitrogen) following manufacturer's recommendations and incubated with 2 mg of nuclear extract in presence of 1 mg of poly-dIdC competitor (Sigma Aldrich) in Buffer B supplemented with 1% Triton X-100 (Sigma Aldrich) 1 hour at room temperature. The isolated proteins were washed extensively and eluted in 1X LDS Sample Buffer (Invitrogen) under reducing conditions, separated by gel electrophoresis in 4-12% gradient BisTris Gels (Invitrogen) and stained by GelCode Blue (Pierce) or Silver Stain Kit (Pierce). Protein bands were excised from a GelCode stained gel and analyzed by mass spectrometry (MS). Alternatively, nuclear extracts prepared from WT or KO mice were incubated with magnetic beads coated with 5'-Biotinylated CG-rich probes.

Mass spectrometric identification of isolated proteins. MS analysis was conducted by the Proteomics Resource Center, Rockefeller University, New York, NY. Protein bands were reduced and alkylated and hereafter subjected to in-gel trypsinization following published protocol. (Shevchenko et al, 1996) Post digestion, peptides were extracted and separated by reversed phase based nano flow liquid chromatography (Dionex, Boston, MA, USA) connected to an Orbitrap XL mass spectrometer (Thermo, San Jose, CA, USA). Peptides were subjected to Tandem MS in CID mode. Tandem MS data were extracted using ProteomeDiscoverer v. 1.3 (Thermo, Bremen, Germany) and queried against the Rat International Protein Index (IPI rat) v 3.87 using MASCOT 2.3 (Matrixscience, London, UK). For each identified protein a rough measure of amount was

calculated based on the average area of the three most abundant peptides (Silva et al, 2006).

Southwestern Blotting. Nuclear extract (isolated as described above) was separated on duplicate gels and transferred to PVDF membranes (BioRad) in Mini Trans-Blot Electrophoretic Transfer Cell (BioRad) at 100 V for 2 hours at room temperature and assayed for DNA binding as previously described (Campoy et al., 1995). The blotted proteins were re-natured by immersing the blot in Buffer B with 6 M guanidine HCl, which was then serially diluted to 3, 1.5, 0.75, 0.375, 0.188, and 0.094 M using binding Buffer B with incubation at 4 °C for 5 min each time. The blot was blocked at room temperature for 30 min in Buffer B containing 5% non-fat milk, then incubated with 10^5 cpm/mL CG-rich 5mC or 5hmC probes with 10 ug/mL poly-dIdC for 2 hours at room temperature, extensively washed with Buffer B and air-dried. The autoradiography was measured by exposing a storage phosphor screen (Molecular Dynamics, Sunnyvale, CA) that was scanned by Typhoon Imager (GE Healthcare) and analyzed by ImageJ software (http://rsbweb.nih.gov/ij/).

Expression and purification of recombinant proteins. cDNA clones of human MeCP2, and MBD family of proteins were purchased from Open Biosystems (Lafayette, CO) and cloned into pet28a vector (Novagen, Madison, WI, USA) using primers listed in Table S3c with NdeI and XhoI (NEB) restriction sites.

The C-terminally 6-His tagged recombinant proteins were expressed in BL21 strain of *E. coli* (Invitrogen) in Overnight Express Autoinduction System (EMD Millipore, Billerica, MA, USA), purified using Ni-NTA Spin columns (Qiagen) according to manufacturer's recommendations and dialyzed against Buffer B supplemented with 10% glycerol. The protein concentration was measured by Quick Start Bradford Assay (BioRad).

Electrophoretic mobility shift assay (EMSA). Unless stated otherwise, 1 pmol of MeCP2 (aa 1-205), 1.2 pmol of MBD1, 0.5 pmol of MBD2, 50 pmol of MBD3 or 2.5 pmol of MBD4 was incubated with 10 fmol of ^{32}P-end-labeled C, 5mC or 5hmC BDNF probe in presence of 1 μg of pdIdC in Buffer B supplemented with 5% Ficoll for 30 min at room temperature. 10 μL of the binding reaction was electrophoresed in native 6% 29:1 acrylamide:bis-acrylamide gel in 0.5 % TBE buffer at 4°C for 2-4 h. The gels were then dried in vacuum (Fisher Scientific, Hampton, NH, USA) for 1 h at 70°C, exposed to a storage phosphor screen (Molecular Dynamics) and scanned by Typhoon Imager (GE Healthcare).

Surface Plasmon Resonance (SPR). Four flow cells (Fc) of Streptavidin (SA)-Sensor Chips (GE Healthcare) were cleaned with 3, 1-min pulses of 1M NaCl in 50mM NaOH in Biacore 3000 (GE Healthcare). 5'-biotinylated C, 5mC and 5hmC BDNF probes were injected onto Fc 2, 3 and 4, respectively, at 2 μL/min in Buffer B until SPR response increased by 500(+/-25) Resonance Units (RU). Fc 1 was immobilized with 2 mM Biotin to serve as reference cell. 8-12 serial dilutions of protein samples within 0.01 Kd to 10 Kd range in Buffer B supplemented with 0.02% Tween and 50 ng/μL poly-dIdC were injected onto each Fc in parallel at 30 μL/min for 3 min. The surfaces were regenerated by 1 min injection of Buffer B with 1.5 M KCl. The change in SPR response with respect to reference cell 5 seconds before the end of injection was recorded for steady state analysis. Each run was performed in duplicates and each experiment was repeated with at least 4 samples of each protein preparation on two independently prepared surfaces. Bmax values and their standard error were automatically extracted in GraphPad Prism Software from steady state values and they were divided by the corresponding protein size. The F test to compare variances was performed using GraphPad Prism software.

Single gene chromatin accessibility. Nuclei were digested using MNase (EpiQ enzyme, BioRad). Single gene chromatin analysis was performed by qPCR in a Roche LightCycler 480 System (Roche) using SYBR green method (BioRad). The accessibility data obtained were compared with GC data since this cell type is the most abundant (around 75%) of the tissue. For single gene chromatin analysis by qPCR, 5 million cerebellar nuclei from a single animal were incubated in the presence of 0.5 μL of MNase in digestion buffer (EpiQ analysis, BioRad) and stopped with stopping buffer after 1, 3, 5, 10, 30 and 60 min of incubation at 37 ºC. Control condition was incubated for 30 min without enzyme. Nuclei were then digested with 40 μg of proteinase K (Roche) at 50 ºC for 1 h, DNA extracted using Qiagen Qiaquick columns (Qiagen), and quantified by Nanodrop. DNA digestion patterns were then visualized by loading 300 ng of DNA in a 2% (w/v) agarose gel in TBE buffer. Single gene chromatin analysis was performed by qPCR in a Roche LightCycler 480 System (Roche) using SYBR green method (BioRad) following this cycling protocol: 96 ºC for 5 minutes; 40 cycles of: 96 C for 15 s, 66 C for 1 min (plate read), 80 ºC 30 s; Melt curve, 70-96 ºC. 1 μg of DNA per sample was amplified for three biological and experimental replicas. Cq values of each digestion were used to calculate % of DNA accessibility using Pfaffl method (Pfaffl 2001).

Accesibility value per gene per condition was calculated following this formula:

$$\% \text{ Accessibility} = [(E_{\text{target}})^{\Delta Cq \text{ target}} / (E_{Ccnb1ip1})^{\Delta Cq\ Ccnb1ip1}] \times 100$$

where E is the percentage of efficiency of the primers and ΔCq the difference between undigested and digested Cq of a given target. E of each pair of primers were close to 100%. Ccnb1ip1 probe was used as an undigested reference, with a ΔCq always less than 1.5, demonstrating inaccessibility, as suggested in EpiQ Chromatin Analysis (BioRad). Calm1 probe was chosen as a control of digestion since the ΔCq was the highest in all conditions. In all the biological triplicates and digestion times ΔCq between undigested reference gene and undigested control gene were less than 1.5.

Targets were chosen randomly from the TRAP-Seq gene expression list covering the maximum range of expression values. Probes were designed to amplify 300 bp in a region that covers the first section of the transcript after TSS, since this is the region were differences between 5hmC and 5mC values were maximum

(see metagenes in Fig 3 A). Table S3d describes the probes used for amplification of the targets, on order of 5hmC/5mC ratio value for GCs.

Southeastern Blot. 18 million nuclei from cerebella of WT or KO mice were resuspended in buffer containing 10 mM Tris pH 8.0 ans 150 mM NaCl supplemented with 5mM $CaCl_2$, 10 mM Na-butyrate, 0.15 mM spermine, 0.5 mM spermidine and Halt Protease and Phosphatase Inhibitor Cocktail (Pierce). The nuclei were partially digested by the addition of 100 Kunitz units of Micrococcal Nuclease (NEB) and aliquots of the reaction were stopped by the addition of excess amount of EGTA at various time points. The nuclei were then incubated with RNase A/T1 Mix (Fermentas) lysed in 1% SDS and incubated with Proteinase K for at 37°C for 30 min and 65°C for 30 min. The DNA was extracted with phenol: chloroform: isoamyl alcohol (25:24:1), precipitated in 70% ethanol and dissolved in Tris-EDTA buffer. 0.5-1 μg of DNA was resolved in 2% TAE agarose gel and transferred to nitrocellulose membrane (Whatman plc, Maidstone, Kent, UK) overnight by capillary action in 20X SSC buffer (3M NaCl, 300mM Na-Citrate) following denaturation in 0.4 M NaOH and 0.6 M NaCl and neutralization in 0.5 M Tris (pH 7.5) and 1.5 M NaCl of the agarose gel. The membrane was equilibrated in 6X SSC buffer, air-dried, UV-crosslinked, blocked in 5% fat-free milk in TBS buffer with 0.2% Tween and incubated with rabbit anti 5hmC antibody (1:500, Active Motif) or mouse anti 5mC antibody (1:250, Diagenode, Denville NJ, USA) for 2 hours at room temperature. The membranes were washed for 3 X 5 minutes and incubated with a secondary antibody conjugated with horse radish peroxidase (1:4000, Abcam) for 30 min at room temperature. The blots were washed for 3 X 10 minutes, rinsed with water and then reacted with ECL substrate (Pierce). The signal was detected by exposing reacted blots to chemiluminescence films (Sigma Aldrich) that were developed in film processor (Konica Minolta, Osaka, Japan). The gel images from five gels from three independent experiments were analyzed in ImageJ software. The average values, standard error of the mean and p values by unpaired t-test were calculated via GraphPad Prism Software.

We refer to this method as Southeastern blot, since it combines the transfer of an electrophoretically separated DNA sample onto a membrane as in classical Southern blot technique with the assaying of chemical modifications as in Eastern blot technique instead of hybridizing with a sequence-specific probe.

Supplemental bibliography

Campoy, F.J., Meehan, R.R., McKay, S., Nixon, J., and Bird, A. (1995). Binding of histone H1 to DNA is indifferent to methylation at CpG sequences. J Biol Chem *270*, 26473-26481.

Chavez, L., Jozefczuk, J., Grimm, C., Dietrich, J., Timmermann, B., Lehrach, H., Herwig, R., and Adjaye, J. Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. Genome Res *20*, 1441-1450.

Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., Suarez-Farinas, M., Schwarz, C., Stephan, D.A., Surmeier, D.J.*, et al.* (2008). A translational profiling approach for the molecular characterization of CNS cell types. Cell *135*, 738-748.

Klose, R.J., and Bird, A.P. (2004). MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. J Biol Chem *279*, 46490-46496.

Kriaucionis, S., and Heintz, N. (2009). The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. Science *324*, 929-930.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J.*, et al.* (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168-176.

Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pages, H., and Gentleman, R. (2009). ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics *25*, 2607-2608.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res *29*, e45.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*,3246-3248.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silverstained polyacrylamide gels. Anal Chem *68*, 850-858.

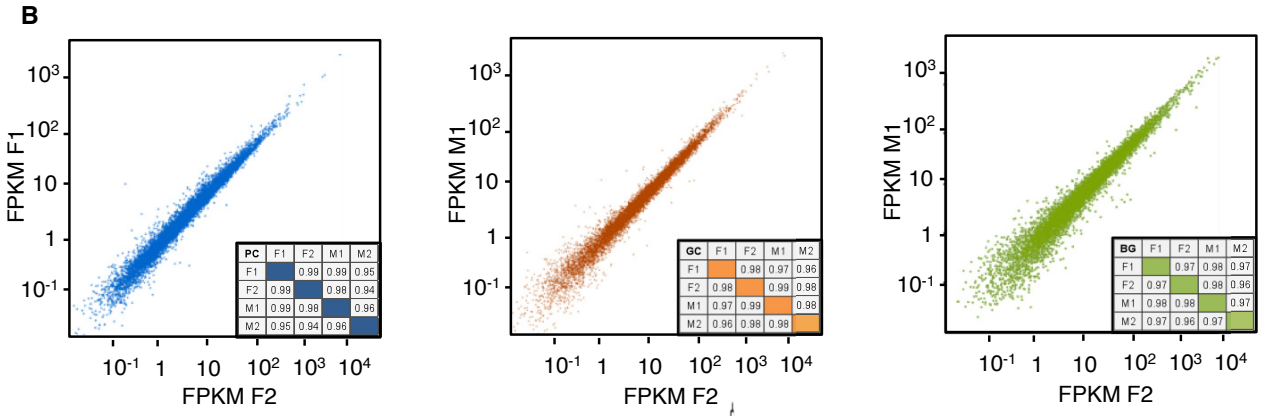
Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P., and Geromanos, S.J. (2006). Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics *5*, 144-156.

Figure S1

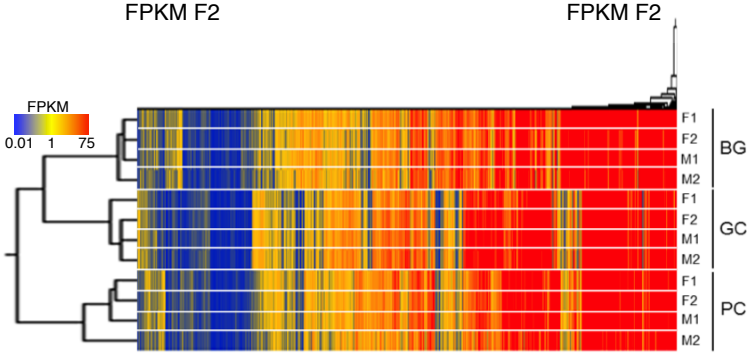
A

	Cell line	Sample type	Replicate	Sex	Sequencing Conditions	N of reads processed	Reads with one alignment or more	Reads Failing Alignment
TRAPSeq	PCP2	Input	1	F	50bp single end HiSeq200 Illumina	93122095	64477752	27587214
		IP				117952218	77408959	31345469
		Input	2			82891533	58917279	23015925
		IP				102102779	70211153	26819472
		Input	1	M		92600050	64680808	26815351
		IP				79197506	52477034	21191190
		Input	2			103496324	75146931	27031543
		IP				91648549	56416106	23815284
	NeuroD1	Input	1	F	50 bp single end HiSeq200 Illumina	65432818	51081294	13537901
		IP				121738952	91077615	29716126
		Input	2			100728986	81130221	18129756
		IP				93483961	70949224	21606702
		Input	1	M		86803595	70780971	14950888
		IP				104645740	84021093	19485400
		Input	2			107930363	88122422	18404184
		IP				70595179	55235833	14593509
	Septin4	Input	1	F	50 bp single end HiSeq200 Illumina	16539673	12648004	3636232
		IP				53760035	42533348	10470506
		Input	2			39581527	28111879	10993066
		IP				48208179	36515427	10780620
		Input	1	M		30304048	23112327	6602158
		IP				38091947	30687733	6922703
		Input	2			47206686	39066506	7527766
		IP				47633248	37417900	9267798

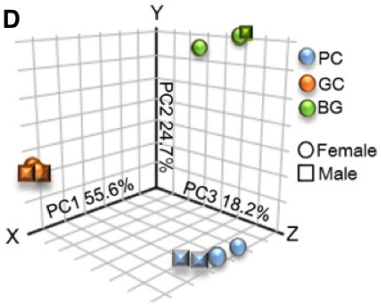
B



C



D



E

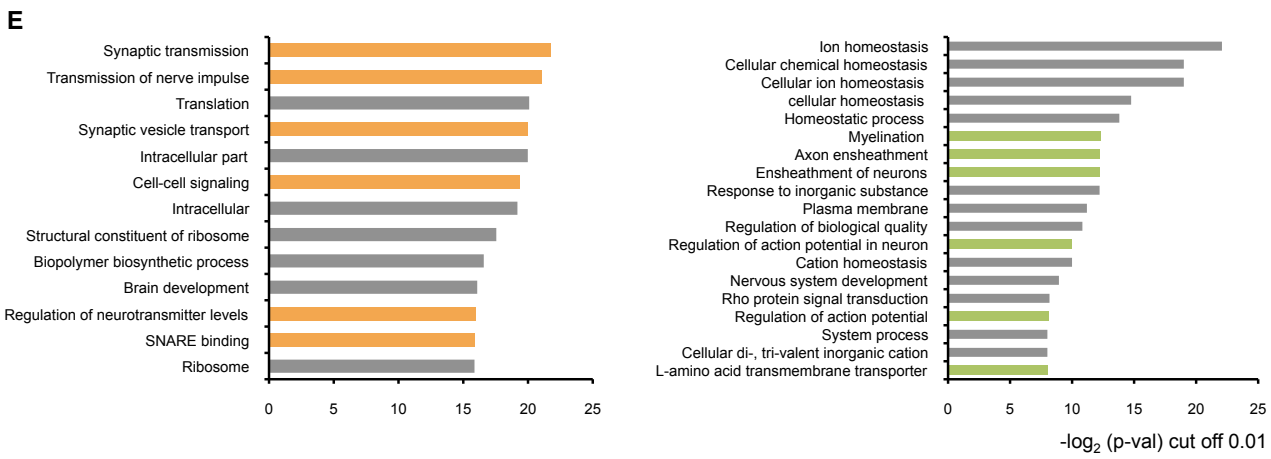


Figure S2

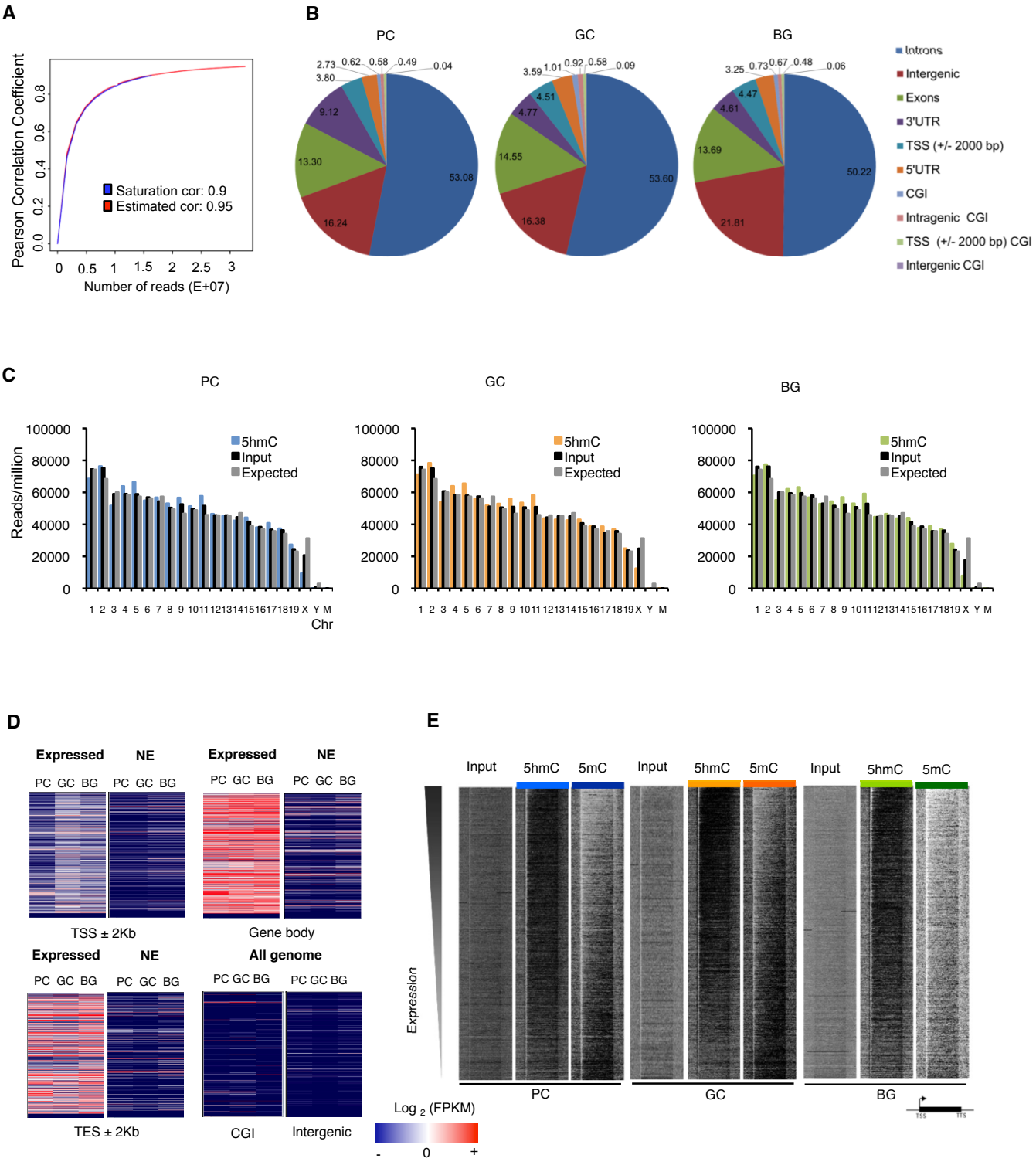

A
Pearson Correlation Coefficient
0.0
0.2
0.4
0.6
0.8
Saturation cor: 0.9
Estimated cor: 0.95
0
0.5
1
1.5
2
2.5
3
Number of reads (E+07)
B
PC
GC
BG
2.73 0.62 0.58 0.49 0.04
3.80
9.12
13.30
16.24
53.08
3.59 1.01 0.92 0.58 0.09
4.51
4.77
14.55
16.38
53.60
3.25 0.73 0.67 0.48 0.06
4.47
4.61
13.69
21.81
50.22
Introns
Intergenic
Exons
3'UTR
TSS (+/- 2000 bp)
5'UTR
CGI
Intragenic CGI
TSS (+/- 2000 bp) CGI
Intergenic CGI
C
PC
GC
BG
Reads/million
100000
80000
60000
40000
20000
0
5hmC
Input
Expected
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X Y M
Chr
D
Expressed
NE
All genome
PC GC BG
TSS ± 2Kb
Gene body
TES ± 2Kb
CGI
Intergenic
E
Input
5hmC
5mC
Expression
PC
GC
BG
TSS
TES
Log 2 (FPKM)
-
0
+

Figure S3

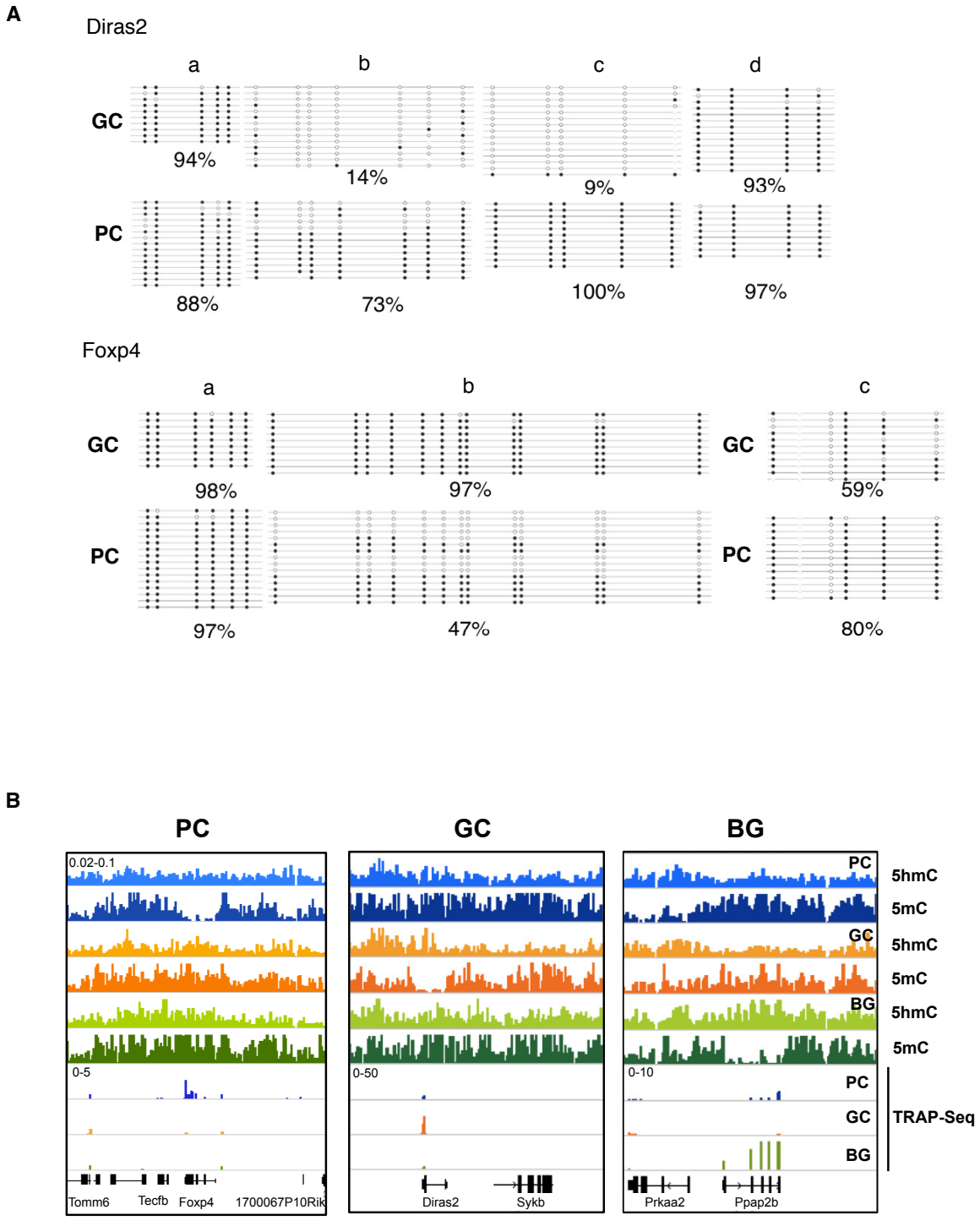

A
Diras2
a
b
c
d
GC
PC
94%
14%
9%
93%
88%
73%
100%
97%
Foxp4
a
b
c
GC
PC
98%
97%
59%
97%
47%
80%
B
PC
GC
BG
0.02-0.1
0-5
0-50
0-10
Tomm6
Tecfb
Foxp4
1700067P10Rik
Diras2
Sykb
Prkaa2
Ppap2b
PC
GC
BG
5hmC
5mC
5hmC
5mC
5hmC
5mC
PC
GC
BG
TRAP-Seq

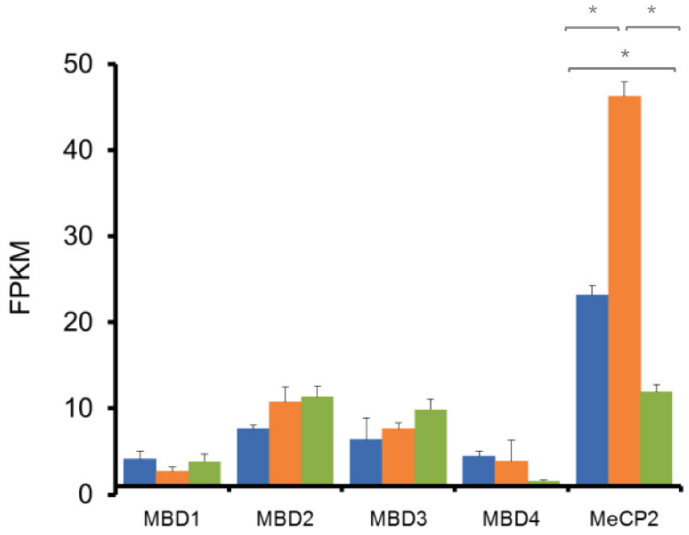
Figure S4

A

```
  1 MAAAAAAAAA AAAAAAAAAA AAAAAPSGGG GGEEERLEEK SEDQDLQGLK
 51 EKPLKFKKVK KDKKEDKEGK HEPLQPSAHH SAEPAEAGKA ETSESSGSAP
101 AVPEASASPK QRRSIIRDRG PMYDDPTLPE GWTRKLKQRK SGRSAGKYDV
151 YLINPQGKAF RSKVELIAYF EKVGDTSLDP NDFDFTVTGR GSPSRREQKP
201 PKKPKSPKAP GTGRGRGRPK GSGTGRPKAA ASEGVQVKRV LEKSPGKLLV
251 KMPFQASPGG KGEGGGATTS AQVMVIKRPG RKRKAEADPQ AIPKKRGRKP
301 GSVVAAAAAE AKKKAVKESS IRSVQETVLP IKKRKTRETV SIEVKEVVKP
351 LLVSTLGEKS GKGLKTCKSP GRKSKESSPK GRSSSASSPP KKEHHHHHHH
401 AESPKAPMPL LPPPPPPEPQ SSEDPISPPE PQDLSSSICK EEKMPRAGSL
451 ESDGCPKEPA KTQPMVAAAA TTTTTTTTTV AEKYKHRGEG ERKDIVSSSM
501 PRPNREEPVD SRTPVTERVS
```

Affinity Pulldown	Accession	Score	Coverage	# Unique Peptides	# PSMs	Area	# AAs	MW [kDa]	calc. pI
5mC	114225270	1659.82	55.58	27	62	47574554	520	55.0	9.89
5hmC	114225270	1210.32	49.62	21	50	60711257	520	55.0	9.89

B



C

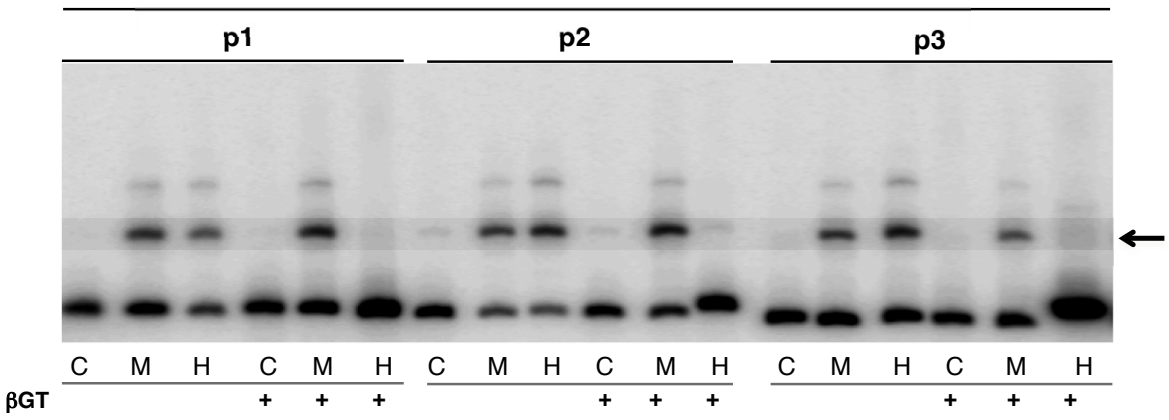
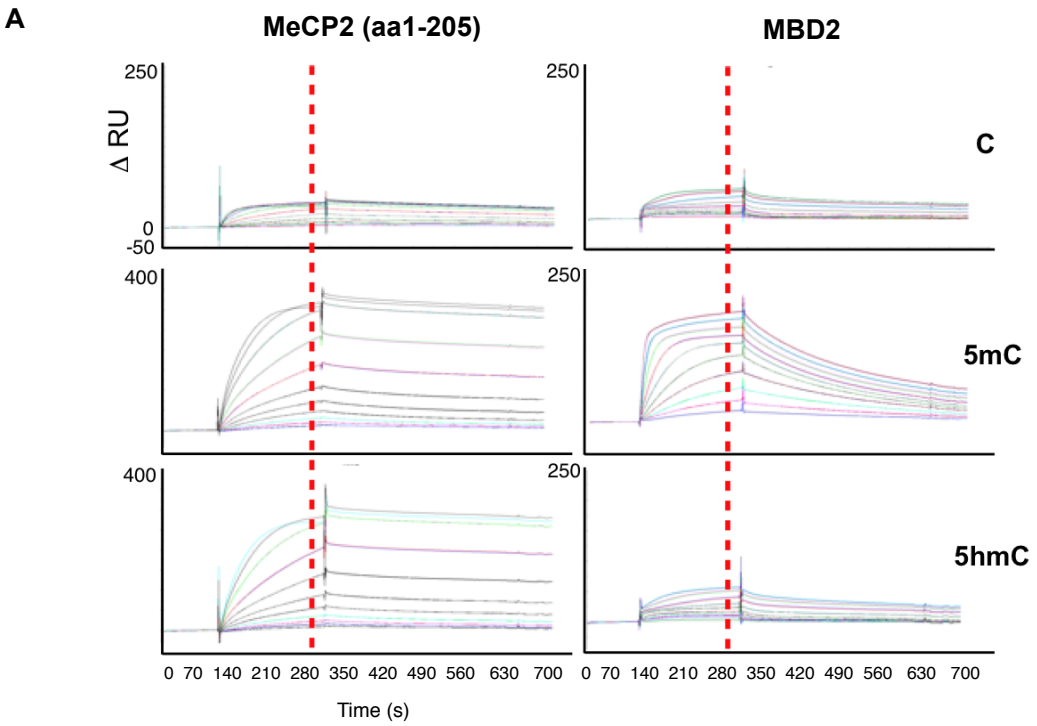
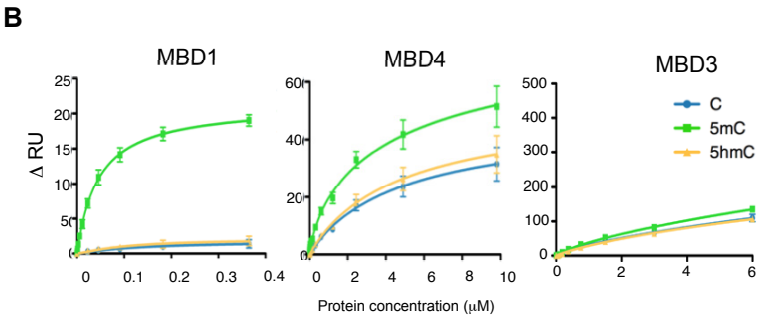
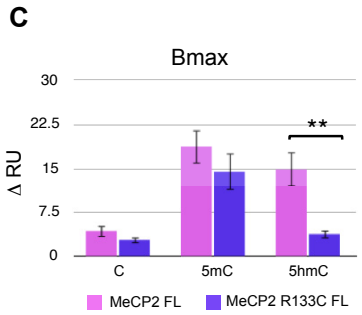


Figure S5







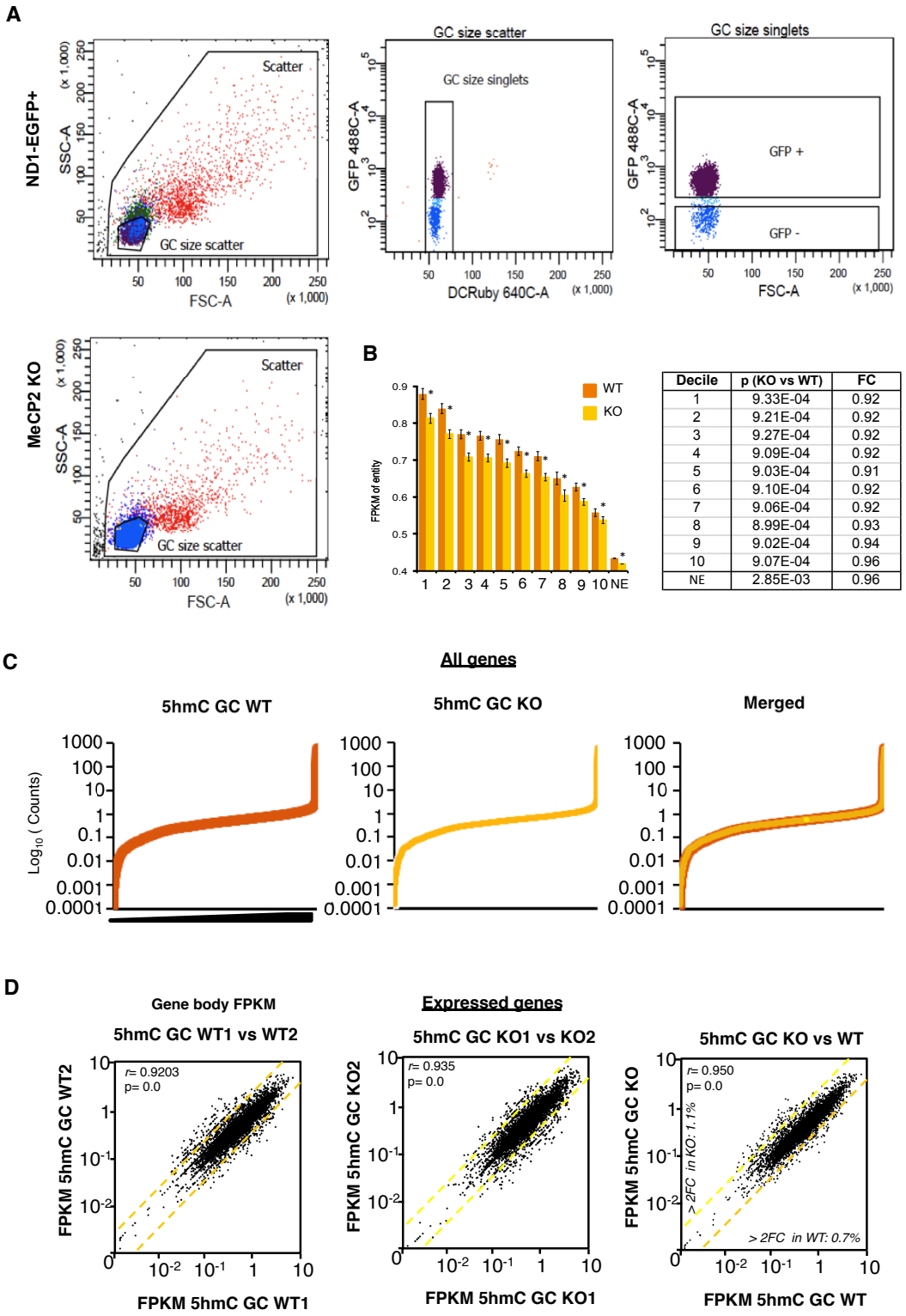
D

	Bmax C	Bmax M	Bmax H	Bmax std C	Bmax std M	Bmax std H
MBD1	1.50	17.80	1.77	1.14	1.18	0.97
MBD2	3.50	20.80	3.80	2.40	1.00	1.83
MBD3	3.44	4.155	3.50	3.54	2.48	3.22
MBD4	5.52	12.48	6.10	1.12	3.14	1.04
MeCP2 aa 1-205	4.20	21.25	19.25	0.50	1.92	1.90
MeCP2 FL	4.35	18.81	15.00	1.05	2.91	2.93
R133C FL	2.81	14.28	3.80	0.54	1.40	0.7

p values (unpaired t-test)	c/m	m/h	c/h
MBD1	<0.0001 ****	<0.0001 ****	0.0843
MBD2	<0.0001 ****	<0.0001 ****	0.5831
MBD3	0.9583	0.7259	0.7654
MBD4	0.0181 *	0.0446 *	0.7045
MeCP2 aa 1-205	0.0001 ***	0.7543	0.0003 ***
MeCP2 FL	<0.0001 ****	0.3523	<0.0001 ****
R133C FL	<0.0001 ****	0.0006 ***	0.0725

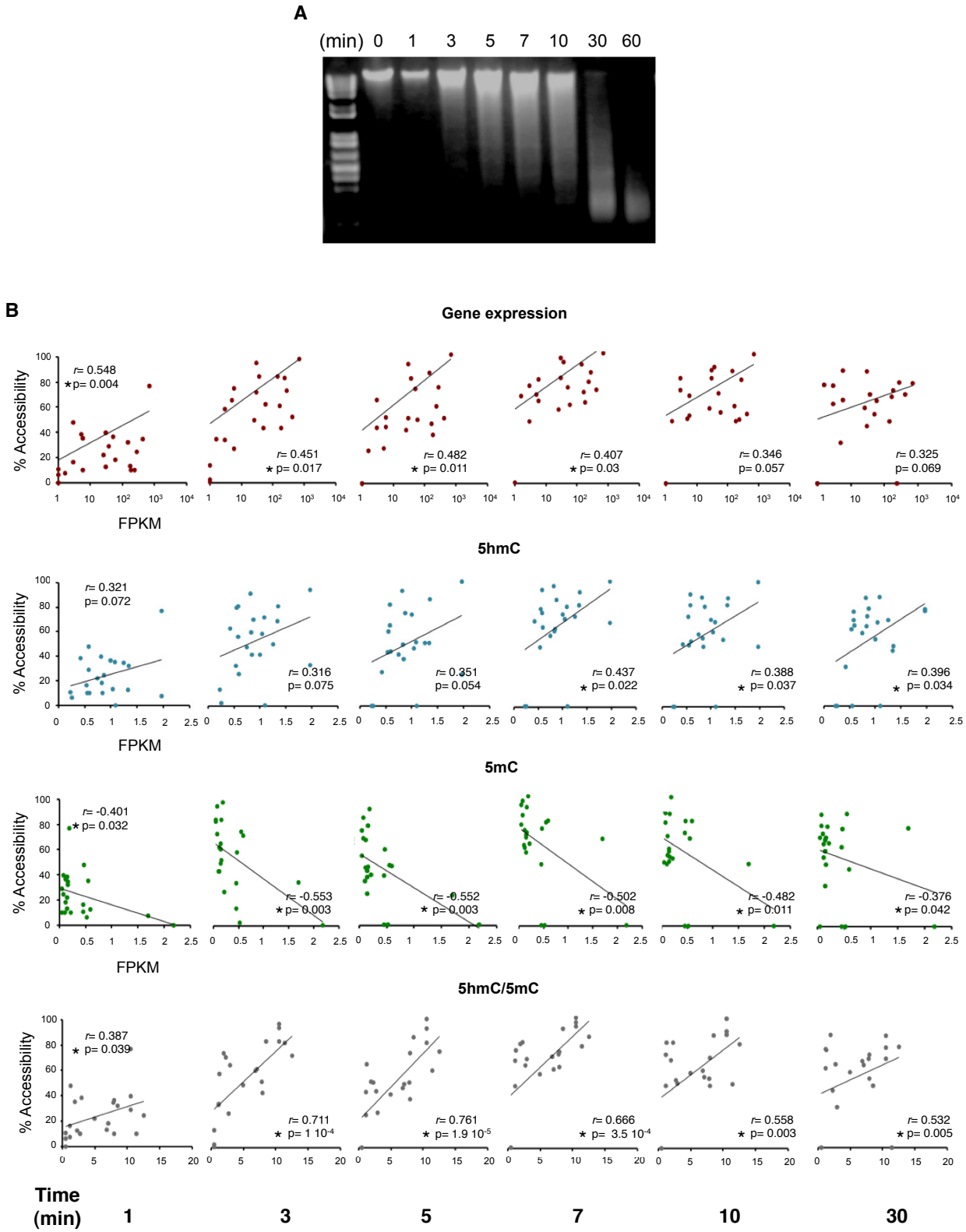
p values	WT vs R133C
C	0.1856
M	0.7690
H	0.0029 **

Figure S6



Decile	p (KO vs WT)	FC
1	9.33E-04	0.92
2	9.21E-04	0.92
3	9.27E-04	0.92
4	9.09E-04	0.92
5	9.03E-04	0.91
6	9.10E-04	0.92
7	9.06E-04	0.92
8	8.99E-04	0.93
9	9.02E-04	0.94
10	9.07E-04	0.96
NE	2.85E-03	0.96

Figure S7


A
(min) 0 1 3 5 7 10 30 60
B
Gene expression
r= 0.548
* p= 0.004
r= 0.451
* p= 0.017
r= 0.482
* p= 0.011
r= 0.407
* p= 0.03
r= 0.346
p= 0.057
r= 0.325
p= 0.069
% Accessibility
FPKM
5hmC
r= 0.321
p= 0.072
r= 0.316
p= 0.075
r= 0.351
p= 0.054
r= 0.437
* p= 0.022
r= 0.388
* p= 0.037
r= 0.396
* p= 0.034
5mC
r= -0.401
* p= 0.032
r= -0.553
* p= 0.003
r= -0.552
* p= 0.003
r= -0.502
* p= 0.008
r= -0.482
* p= 0.011
r= -0.376
* p= 0.042
5hmC/5mC
r= 0.387
* p= 0.039
r= 0.711
* p= 1 10-4
r= 0.761
* p= 1.9 10-5
r= 0.666
* p= 3.5 10-4
r= 0.558
* p= 0.003
r= 0.532
* p= 0.005
Time (min) 1 3 5 7 10 30