Supporting Information

Chen et al. 10.1073/pnas.1505909112

SI Materials and Methods

ChIP. ChIP was performed as described previously (1). Briefly, mouse hypothalamus tissues of 7-wk-old mice were minced and fixed in in 1% formaldehyde in PBS for 10 min at room temperature. After quenching the fixation with 125 mM glycine in PBS, the samples were homogenized in hypotonic solutions to release the cytoplasmic fractions. Nuclei were collected at 11,200 × *g* in a table-top centrifuge. Chromatin was prepared using a method of sonication followed by a brief Micrococcal Nuclease (Worthington) digestion for 5 min at 37 °C. Reactions were stopped by addition of EGTA. The chromatin was then collected by spinning at 15,700 × *g* in a table-top centrifuge for 10 min.

Chromatin was diluted 1:10 in ChIP Dilution Buffer (Millipore) and precleared with Protein A Dynabeads (Invitrogen). An aliquot of 5% (vol/vol) of chromatin was stored as input and the remaining amount was incubated overnight at 4 °C with 2 μg of anti-GFP antibody (Abcam, ab6556) overnight at 4 °C. Next, 20 μL of Protein A Dynabeads were then added to each reaction and allowed to bind for 2 h at 4 °C. The beads were then washed in Low-Salt Immune Complex Wash Buffer (Millipore), High-Salt Immune Complex Wash Buffer (Millipore), LiCl Immune Complex Wash Buffer (Millipore), and 1× TE (10 mM Tris HCl, 1 mM EDTA). The beads were eluted twice in 1% SDS/100 mM $NaHCO_3$. Precipitated chromatin and input samples were then reverse cross-linked and treated with proteinase K. DNA was recovered and eluted in water using a Qiagen PCR Purification Kit.

ChIP-Seq Data Analysis. All raw reads were mapped to mouse genome mm9 by bowtie, allowing up to two mismatches. Reads that could be mapped to multiple locations with the same best mapping score were removed. Mapped reads from each biological replicate were subject to the function Dtriple in DANPOS v2.2.1 (2) for MeCP2 occupancy calculation. Each read was extended toward the 3′ end to be 200 base pairs. MeCP2 occupancy (or MeCP2 binding) was then calculated as the number of reads at each base pair. Three independent biological replicates of MeCP2 ChIP-Seq were normalized and then pooled for downstream analysis of high-resolution MeCP2 binding profiles.

Independent Datasets on MeCP2 Occupancy and DNA Methylation. The following previously published independent datasets were used. Raw reads from mouse whole brains were downloaded from the Gene Expression Omnibus (GEO) under the accession no. GSM494290. Reads from ES cells, differentiated neurons, and related inputs were downloaded under the accession nos. GSM972976, GSM972981, and GSM1161419, respectively. DNA methylation data were downloaded under the access no. GSE47966. Specifically, mCG and mCH data from adult mouse brain (6-wk-old) were downloaded under the accession no. GSM1173783. hmCG data from adult mouse brain (6-wk-old) were downloaded under the accession number GSM1173795. mCG and mCH data from fetal mouse brain were downloaded under the accession no. GSM1173779.

Scatter Plot Analyses of Correlation. For downstream scatter plot analyses of correlation, after MeCP2 occupancy was calculated by counting the number of reads in 1-Mb bins across the whole mouse genome, the number of total reads in each MeCP2 ChIP-Seq dataset was normalized to 75 M to correct for variance in ChIP-Seq depth from different experiments. mCG and mCH densities were calculated as the count of mCG and mCH observed in each 1-Mb bin across the whole mouse genome. The total number of cytosines (in mCG, mCH, and nonmethylated C) in all reads was normalized to be equal to the mouse genome size. The value at each dot was the normalized MeCP2 occupancy, mCG density, or mCH density at each 1-Mb bin region across the genome. Correlation coefficient was calculated based on the indicated methods in the figure legends and shown in each plot.

mRNA-Seq and Gene Ontology Data Analysis. For analysis of mRNA-Seq data, raw reads were mapped to mouse genome mm9 using Tophat based on the default parameters. Samtools was used in processing the output files of Tophat. University of California Santa Cruz (UCSC) knownGene was downloaded from the UCSC Table Browser (genome.ucsc.edu/cgi-bin/hgTables?command=start). Reads count for each gene was calculated by the htseq-count function in the tool HTSeq. Differential genes were defined using the negative binomial model provided in EdgeR.

For gene ontology analysis, genes misregulated in MeCP2 knockout (KO) and MeCP2 transgenic (Tg) mice were analyzed grouped as outlined in Fig. S4*B*. For each group, official gene names were submitted to the DAVID database (david.abcc.ncifcrf.gov) and the Functional Annotation Chart function was used to retrieve a combined list of enriched functional terms (Dataset S1). We then further used Fold Enrichment threshold 2 and Q value (Benjamini-adjusted *P* value) threshold 0.05 to remove functional terms that were less significantly enriched.

MNase-Seq Data Analysis. Raw reads from each biological replicate were mapped to mouse genome mm9 by bowtie, allowing up to two mismatches. Reads that could be mapped to multiple locations with the same best mapping score were removed. Mapped reads from each replicate were subject to the Dpos function in DANPOS v2.2.1 (2) for occupancy calculation. The 5′ end of each read was shifted 74 base pairs toward the 3′ end direction, and then each read was extended to be 74 base pairs. Nucleosome occupancy was calculated as the number of reads at each base pair. Nucleosome occupancy in each replicate was normalized to have an average value of 3 across the whole genome. The occupancy data from replicates were then pooled for downstream analysis. Nucleosome positions were defined by an occupancy cut-off of 5.

Heatmap and Analyses of Functional Profiles. To relate MeCP2 binding with DNA methylation and nucleosome profiles, we plotted each heatmap and normalized profiles based on the Profile function in DANPOS v2.2.1 followed by normalization based on flanking region as described previously (3). Briefly, MeCP2 density, methylation ratios, and nucleosome occupancy in wiggle format files were subject to DANPOS v2.2.1, along with the gene coordinate file. We used the *Mus musculus* genomic annotation version mm9 for all our analyses. Gene annotations were obtained from UCSC (genome.ucsc.edu/). Values in the wiggle format files for each gene region from 100 kb upstream of 5′ end of each gene body to 100 kb downstream of 3′ end of each gene body were retrieved. The length of each gene body was normalized to 100 kb. The signal at each gene was normalized so that the mean value in the flank region of each gene body will be the same. These values were then subjected to MEV v4.5.0 to plot the heatmap figures. Normalized profiles were calculated for the average of each indicated gene set. In all profile figures, the *x* axis showed the genomic coordinates, and the *y* axis showed the raw values of the averaged signal.

RT-PCR Primer Sequences. For RT-PCR analysis of gene expression, the following primers are used: *Gapdh* forward 5′-ggagattgttgccatcaacga-3′, reverse 5′-tgaagacaccagtagactccacgac-3′; *Bdnf* forward 5′-cgccatgcaatttccactatcaataatttaac-3′, reverse 5′-ctttttcagtcactacttgtcaaagtaaac-3′. For RT-PCR analysis following ChIP experiments, the following primers were used: Mouse major satellite DNA, forward 5′-catccacttgacgacttgaaaa-3′, reverse, 5′-gaggtccttcagtgtgcattt-3′; *Sst* forward 5′- cattgacaggtacccaactga-3′, reverse 5′-cagccacataggagcacactt-3′; *Bdnf* forward 5′-ggattcccttcatcctcagat-3′, reverse 5′-ccaaagagtaagtgtgcccttc-3′; high-affinity site 1, forward 5′-atgttccagggatgatgaactt-3′, reverse 5′-gctacgcagtagagacatgctg-3′; high-affinity site 2, forward 5′-atgtagacgtacacgcccttg-3′, reverse 5′-agttctgactgtcctggaactg-3′; low-affinity site 1, forward 5′-ccagctactcgcggctttacgg-3′, reverse 5′-cctcccgccctgcttatccagt-3′; low-affinity site 2, forward 5′-ggaagtgaggacagaaatggac-3′, reverse 5′-attggctcataccaaaccagag-3′.

1. Baker SA, et al. (2013) An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. *Cell* 152(5):984–996.
2. Chen K, et al. (2013) DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res* 23(2):341–351.
3. Lister R, et al. (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* 341(6146):1237905.

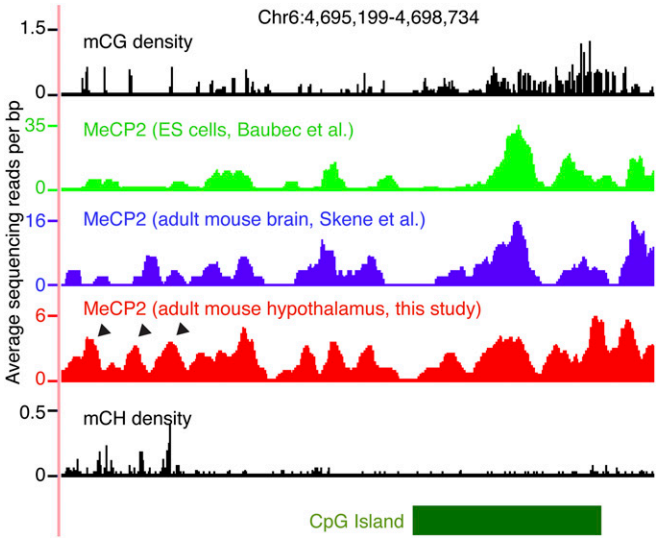

Fig. S1. Brower representation of MeCP2 binding at an example chromosomal region. Shown are mCG density in the adult mouse brain (black), MeCP2 binding in the ES cells (green), adult mouse brains (blue), and adult mouse hypothalamus (red) spanning an example chromosomal region (*Top*, representing the *x* axis). Arrowhead denotes additional MeCP2 binding sites in the adult mouse hypothalamus comparing to ES cells that coincide with mCH density peaks in the adult mouse brain (black). This region contains a methylated CpG island (green box at the bottom).

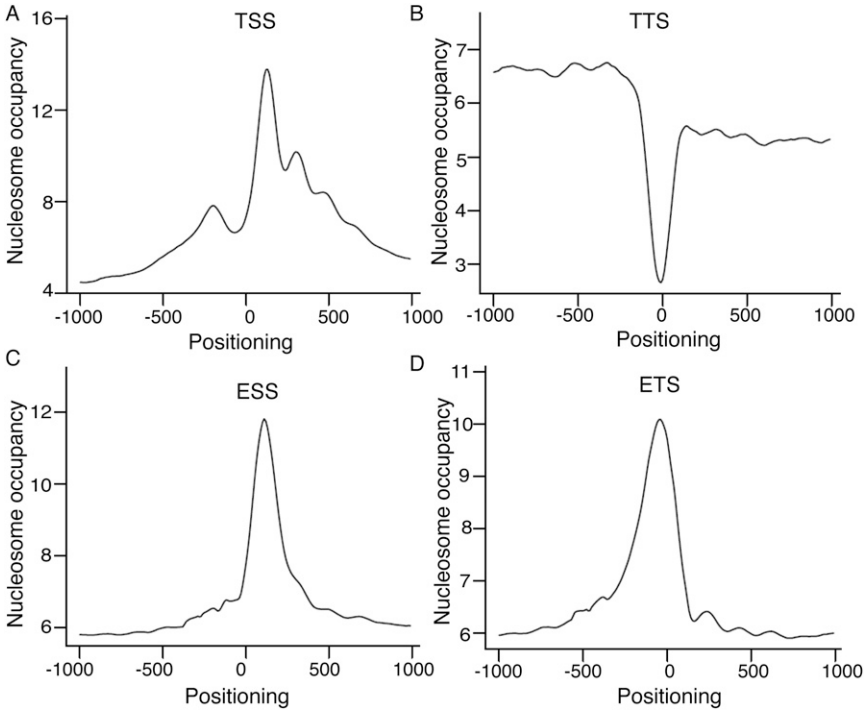

Fig. S2. Global features of nucleosome positioning in the adult mouse brain. (*A*) Normalized nucleosome occupancy at the transcriptional start site (TSS), showing nucleosome-depleted region at TSS followed by a +1 nucleosome. (*B*) Normalized nucleosome occupancy at the transcriptional termination site (TTS), showing nucleosome depletion at TTS. (*C* and *D*) Normalized nucleosome occupancy at the exon start site (ESS) and exon termination site (ETS) showing increased nucleosome occupancy at these two sites.

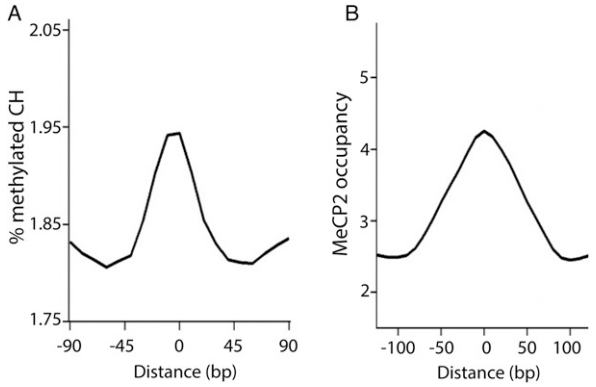

Fig. S3. Relationship between MeCP2, mCH, and nucleosome occupancy in the adult mouse brain. (*A*) Profile of mCH levels averaged and plotted around the nucleosome. The center indicates the highest nucleosome score. (*B*) Normalized MeCP2 binding surrounding the nucleosome, showing the overlap between MeCP2 and nucleosome occupancy. The center indicates the highest nucleosome score.

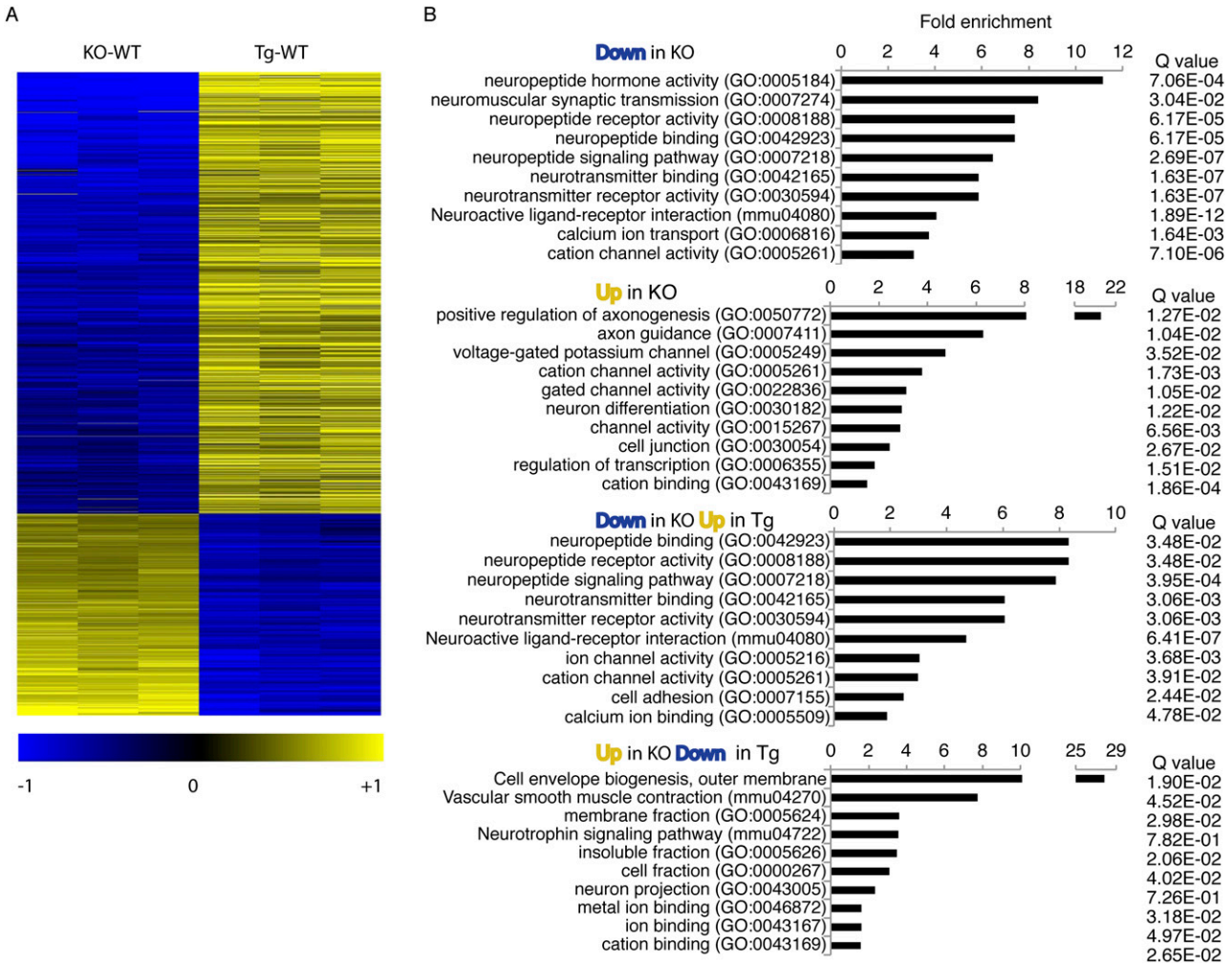

Fig. S4. Characterization of transcriptional regulation by MeCP2. (*A*) Heatmap showing fold changes in gene expression by mRNA-Seq in MeCP2 knock (KO) mice and MeCP2 transgenic (Tg) mice. Yellow and blue colors indicate increased and decreased expression, respectively, relative to their corresponding wild-type (WT). Each column represents the difference between one indicated genotype comparing to its corresponding one wild-type control, and each row represents one gene. Expression levels are depicted according to the color scale at the bottom. False-discovery rate-adjusted $P < 10^{-5}$. (*B*) Function analysis of misregulated genes in KO (top two panels), as well as genes that are inversely mis-regulated between KO and Tg mice (bottom two panels). We observe enrichment for neuronal functions in all groups. Ten representative functional terms were selected by requiring Q value (indicate at the right side) smaller than 0.05 or fold enrichment (*x* axis) larger than 2. Full list of functional terms reported by the DAVID database as enriched for each group have been provided as Dataset S1.

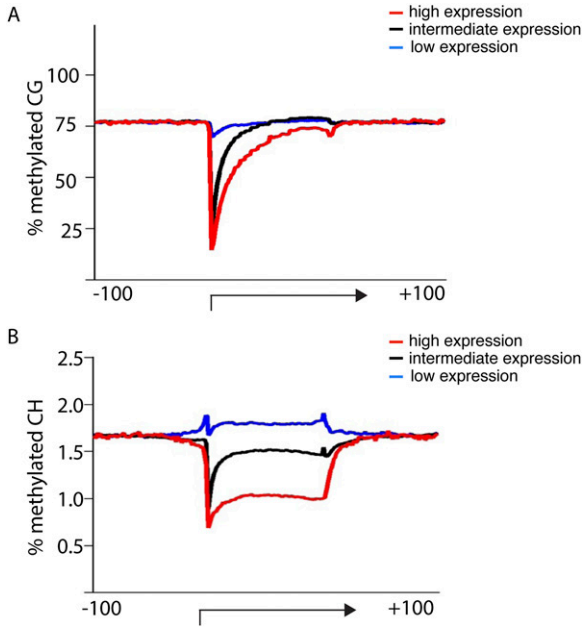

Fig. S5. Relationship between DNA methylation and transcriptional activity in wild-type mouse brain. Averaged mCG levels (*A*) and mCH levels (*B*) were plotted for genes with indicated expression levels in the adult mouse brain, showing negative correlation with expression of associated genes. High-expression genes (red) and low-expression genes (blue) were defined to have fragments per kilo bases of exons for per million mapped reads (FPKM) >10 and < 0.01, respectively. Intermediate expression genes (black) were defined to have FPKM between 0.01 and 10. Signal values are normalized so that the mean signal value in flanking regions of each gene will be the same.

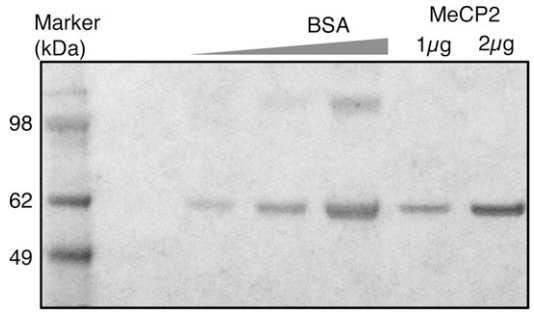

Fig. S6. Purification of recombinant full-length MeCP2 to homogeneity. Purified MeCP2 proteins were dialyzed into 1× PBS overnight at 4 °C. Next, 1 μg and 2 μg of MeCP2 proteins were loaded onto SDS/PAGE and stained with Coomassie brilliant blue to check purity.

Dataset S1. Gene Ontology of MeCP2 misregulated genes

Dataset S1

A full list of functional terms reported by the DAVID database as enriched for each group highlighted in Fig. S4. Datasets for mRNA-Seq were conducted from 7-wk-old hypothalamus from MeCP2 knockout mice and their age and genetic background matched wild-type control mice, and also from 7-wk-old hypothalamus from MeCP2 transgenic mice and their age and genetic background matched wild-type control mice. Datasets for MeCP2 ChIP-Seq were conducted from ~7-wk-old hypothalamus tissues from Mecp2$^{-/y}$; MECP2-EGFP mice. Datasets for Mnase-Seq were conducted from 7-wk-old hypothalamus from MeCP2 knockout mice and their age and genetic background matched wild-type control mice. All sequencing data generated in this work have been submitted to GEO under accession no. GSE66871.