5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes

Marian Mellén[a], Pinar Ayata[a,1], and Nathaniel Heintz[a,b,2]

[a]Laboratory of Molecular Biology, The Rockefeller University, New York, NY 10065; and [b]Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065

Contributed by Nathaniel Heintz, July 3, 2017 (sent for review May 15, 2017; reviewed by Steven Henikoff and Rob Martienssen)

5-hydroxymethylcytosine (5hmC) occurs at maximal levels in postmitotic neurons, where its accumulation is cell-specific and correlated with gene expression. Here we demonstrate that the distribution of 5hmC in CG and non-CG dinucleotides is distinct and that it reflects the binding specificity and genome occupancy of methylcytosine binding protein 2 (MeCP2). In expressed gene bodies, accumulation of 5hmCG acts in opposition to 5mCG, resulting in "functional" demethylation and diminished MeCP2 binding, thus facilitating transcription. Non-CG hydroxymethylation occurs predominantly in CA dinucleotides (5hmCA) and it accumulates in regions flanking active enhancers. In these domains, oxidation of 5mCA to 5hmCA does not alter MeCP2 binding or expression of adjacent genes. We conclude that the role of 5-hydroxymethylcytosine in postmitotic neurons is to functionally demethylate expressed gene bodies while retaining the role of MeCP2 in chromatin organization.

epigenetics | 5-hydroxymethylcytosine | neuron | MeCP2

Proper maintenance of genomic cytosine methylation is essential for the normal functions of mammalian cells (1). The discovery of 5-hydroxymethylcytosine (5hmC) and the 10–11 translocation (TET) proteins (2, 3) have led to important insights into 5-methylcytosine (5mC) oxidation and its roles in DNA demethylation (4). Passive demethylation occurs following oxidation of 5mC to 5hmC in dividing cells because 5hmC prevents the maintenance DNA methyltransferase complex (DNMT1/UHRF1) from acting on newly synthesized hemimethylated DNA to restore 5mC on the nascent strand. This results in dilution of 5hmC in genomic DNA and, in successive rounds of division, demethylation (4). Oxidation of 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) can also occur (5). These derivatives are removed by thymine-DNA glycosylase and replaced with cytosine by the mammalian base excision repair mechanism, thus resulting in active DNA demethylation (6). Taken together, these mechanisms provide for both genome-wide and local cytosine demethylation in dividing cell populations.

In postmitotic neurons, 5hmC accumulates to approximately 10 times the levels present in peripheral cell types (2, 7, 8) and it serves as a stable epigenetic mark. Maximal accumulation of 5hmC occurs in expressed gene bodies (9–13). Although the distribution of 5hmC is cell-specific, the ratio of 5hmC/5mC correlates well with gene expression in all characterized neuronal cell types (11). The discovery that the most abundant 5mC binding protein in neurons, the methyl CpG binding protein 2 (MeCP2), can bind 5mC and 5hmC with similar affinities led to the hypothesis that MeCP2 is an important "reader" of this new epigenetic mark, and that together they constitute a cell-specific mechanism for the regulation of chromatin structure and gene expression (11). Recent data indicate, however, that the binding of recombinant MeCP2 to 5hmC containing DNA depends on the context in which it occurs (14–16). Thus, MeCP2 binds with similar high affinities to 5mCG, 5mCA, and 5hmCA containing DNA, but its binding to 5hmCG occurs at a low affinity similar to that measured for unmethylated DNA. Given the importance of MeCP2 for normal neuronal function demonstrated through studies of Rett syndrome mouse models (17, 18) and the differences in MeCP2 binding to 5hmCG and 5hmCA containing DNA (14–16), we have employed 5hmC mapping using bisulfite-sequencing data (MethylC-Seq) (10) and oxidative bisulfite sequencing (oxBS-Seq) (19), high-resolution mapping of native MeCP2 binding sites using the Occupied Regions of Genomes from Affinity-purified Naturally Isolated Chromatin (ORGANIC) approach (20), and ChIP mapping of informative histone modifications to understand the biochemical consequences of 5hmC accumulation in a single neuronal cell type in vivo.

Results

5hmCG and 5hmCH Occur Predominantly in Euchromatin. To investigate further the distributions and functions of 5hmC, we have chosen to focus on cerebellar granule cells. Granule cells are an abundant, postmitotic neuron type that provide an opportunity to examine and distinguish between the highly condensed, inactive domains characteristic of heterochromatin (HC), and the less compact, functionally active domains of euchromatin (EC) (21). Since standard MethylC-Seq does not distinguish between methylated and hydroxymethylated sites (10, 19) we used in parallel genome-wide oxBS-Seq because comparative analysis of these

Significance

The main insight from this study is that the role of 5-hydroxymethylcytosine (5hmC) in postmitotic neurons is to sculpt the genome occupancy of the very abundant 5-methylcyctosine binding protein 2 (MeCP2). Accumulation of 5hmCG in transcribed genes replaces high-affinity 5mCG binding sites with low-affinity sites, decreasing MeCP2 occupancy over the transcription unit and removing its repressive effect. We refer to this role for 5hmCG as "functional demethylation" because its biochemical effect with respect to MeCP2 is equivalent to chemical demethylation: Loss of high-affinity sites for interaction in the genome. This concept reinforces the roles of 5hmC in demethylation in dividing cells by a mechanism that achieves the same goal without requiring cell division or DNA damage.

Author contributions: M.M., P.A., and N.H. designed research; M.M. and P.A. performed research; M.M. analyzed data; and N.H. wrote the paper.

Reviewers: S.H., Fred Hutchinson Cancer Research Center; and R.M., Cold Spring Harbor Laboratory.

The authors declare no conflict of interest.

Data deposition: The sequence data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database [accession nos. GSE95628 and GSE42880 (cerebellum MeCP2 KO RNA-Seq data)].

[1]Present address: Laboratory of Brain Epigenetics, Icahn School of Medicine at Mount Sinai, New York, NY 10029.

[2]To whom correspondence should be addressed. Email: heintz@rockefeller.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1708044114/-/DCSupplemental.

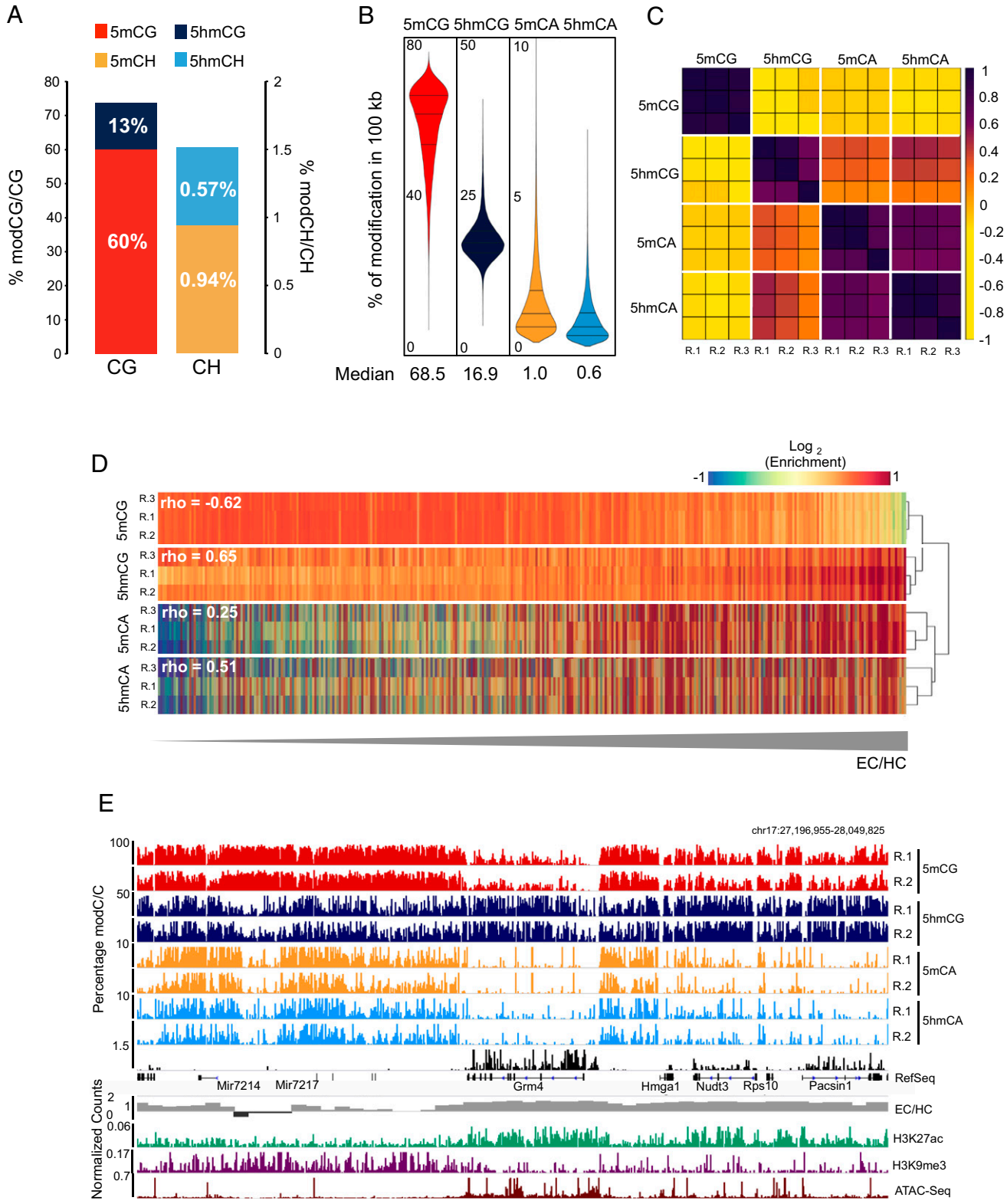

Fig. 1. 5hmCG and 5hmCH occur predominantly in EC. (*A*) Percentage of base calls with each modification in the CG and CH contexts in granule cell genome. (*B*) Genome-wide distributions of 5mC and 5hmC in CG and CA contexts in 10-kb windows. Median values of each distribution are indicated below. (*C*) Correlation matrix of Pearson's *r* of biological replicas (R.1, R.2, and R.3) of normalized cytosine modifications values in 100.kb windows. Samples are ordered by hierarchical clustering of modifications. (*D*) Heat-map representation of enrichment values of each cytosine modification genome-wide. Rows are ordered following EC/HC ratio. Higher values are more accessible (EC) and lower, less accessible (HC). Spearman correlation values between each modification and EC/HC are indicated. (*E*) Browser representation of a euchromatic region showing percentage of modification of individual sites for 5mCG, 5hmCG, 5mCA, and 5hmCA, in two independent biological replicas and nuclear RNA-Seq. EC/HC ratio in 10kb bins and the normalized counts for H3K9me3 and H3K27Ac ChIP-Seq and ATAC-Seq are also shown. Maximum enrichment levels or normalized counts are indicated on the top left corner. Genomic coordinates are shown on the top right corner of the plot. See also Fig. S2*A*.

datasets allows 5hmC levels to be determined with single-nucleotide resolution (19). MethylC-Seq data and oxBS-Seq data were collected from three independent biological replicates yielding average coverages of 13.5 and 25.4 per cytosine, respectively (cut off three reads, Dataset S1). The correlation coefficients for three oxBS-Seq technical replicates within the same biological replica (pairwise Pearson $r \geq 0.99$ for 5mCG and ≥ 0.92 for 5mCA in 1-kb bins) demonstrated that the oxidation and bisulfite conversion were highly reproducible. Analysis of the spike-in controls for the MethylC-Seq and oxBS-Seq biological replicates (Fig. S1*A* and Dataset S1) showed CG nonconversion rates of 0.59% for MethylC-Seq and 0.21% for oxBS-Seq and CH nonconversion rates of 0.46% and 0.21%, respectively. The oxBS-Seq 5hmC efficiency of oxidation (measured as the difference between 5hmC conversion in oxBS-Seq and MethylC-Seq) was 91.93% and 92.94% for CG and CH, respectively. Together these data demonstrate that oxBS-Seq can be used to assess accurately genome-wide CG and non-CG methylation and hydroxymethylation in granule cell DNA samples (Fig. S1*A* and Dataset S1).

To estimate the percentage of methylation and hydroxymethylation at each cytosine and to understand their distribution in the genome we used the maximum likelihood methylation levels (MLML) method described by Qu et al. (22). This approach combines information from MethylC-Seq and oxBS-Seq to arrive at maximum likelihood estimates for the 5mC and 5hmC levels per cytosine. A binomial test is performed for each methylation level calculated. If the estimated methylation level falls out of the confidence interval calculated from input coverage and methylation level, then this is counted as a conflict and removed from the analysis.

Our data (Fig. 1*A*) are generally consistent with previous reports of neuronal cytosine modification (9, 10, 12). Thus, of the modified CG dinucleotides, methylcytosine (~60%) is ~4.6 times more abundant than hydroxymethylcytosine CG (~13%) (Fig. 1*A*). The extent of cytosine modification in the CH context (defined as C followed by A, C, or T, ~1.5%) also agrees well with previous MethylC-Seq studies of total non-CG methylation in neurons (9, 10, 12) (Fig. 1*A*), as does our consensus nucleotide context indicating that non-CG methylation occurs predominantly as 5mCA (Fig. S1*C*). Our data demonstrate that 5hmCH (0.57%) occurs at reduced but substantial levels in granule cell genomes relative to 5mCH (0.94%). Although these values are higher than the levels of 5hmCA determined using Tet-assisted bisulfite sequencing to study 5hmC in neurons in the prefrontal cortex (10), the levels of 5hmC detection by oxBS-Seq are consistent with recent studies that used Pvu-Seal-Seq to determine that ~24% of 5hmC in mouse embryonic stem cells occurs in the non-CpG context (23, 24).

To determine the distribution of cytosine modifications in the granule cell genome we calculated the average percentage for each modification in 100-kb windows across the genome (Fig. 1*B*). As expected, 5mCG was the most abundant modification genome-wide, with 5mCG (median value 68.5%), and the majority of the 100-kb genomic intervals were heavily methylated. 5hmCG was significantly less abundant in each interval (median value 16.9%). The DNA fragment distribution for cytosine modifications in the CA context revealed that the majority of the genome contained low levels of both 5mCA (1.0%) and 5hmCA (0.65%) with few large intervals containing levels of modified CH that approach those seen in the CG context (Fig. 1*B*).

To determine whether the distribution of each form of cytosine modification in granule cell genomes was highly correlated in independent biological replicates, a Pearson's correlation coefficient matrix for the triplicate samples was generated (Fig. 1*C*). These data illustrate that the gross distributions of cytosine modifications mapped using MethylC-Seq and oxBS-Seq data are highly reproducible. They also reveal two features of these distributions that are of interest. First, although the distributions of 5mCH and 5hmCH are strongly correlated, the distributions of 5mCG and 5hmCG are inversely correlated; second, the average values per region of 5hmCG are more closely correlated with 5mCA and 5hmCA than they are with 5mCG. These data suggest that the accumulation of 5hmCG, 5mCA, and 5hmCA reflect aspects of granule cell development and nuclear structure that are not shared by 5mCG.

Given previous studies demonstrating enrichment of 5mCG in HC (25), we were interested in whether the distinct distributions of 5hmCG, 5mCA, and 5hmCA might reflect their accumulation in EC. To map heterochromatic and euchromatic domains in granule cell nuclei, ChIP-Seq data were collected for histone marks characteristic of EC (H3K27Ac) and HC (H3K9me3) (26). The ratio between these histone marks (H3K27Ac/H3K9me3) was used as a measurement of chromatin organization (EC/HC). DNA accessibility in euchromatic domains was confirmed using Assay for Transposase-Accessible Chromatin with high-throughput sequencing (27) (ATAC-Seq; Fig. 1*E*). We then rank-ordered the genomic segments analyzed above based on EC/HC and plotted cytosine methylation status for each replicate relative to this measure. As expected, 5mCG accumulates in heterochromatic domains in granule cell nuclei, and it is relatively depleted in EC (Spearman rho = –0.62) (Fig. 1*D*). In contrast, 5hmCG (rho = 0.65), 5mCA (rho = 0.25), and 5hmCA (rho = 0.51) are all enriched in euchromatic regions and depleted in heterochromatic segments.

The distributions presented above are consistent with the observation that modifications in CA dinucleotides are abundant in postmitotic neurons and depleted in regions with low accessibility (10, 12). However, inspection of cytosine modification data relative to chromatin organization and gene expression in the genome browser provides additional insights into the relationships between 5hmCG and 5hmCA (Fig. 1*E* and Fig. S2*A*). Three properties are evident. First, although reproducible distributions of each modification are clearly revealed in both of the biological replicates shown, the data do not correspond precisely at each genomic position. Rather, it is their enrichment or depletion over specific genomic features, such as gene bodies, promoters, or enhancers, that is conserved. Second, as expected, the accumulation of 5hmCG and depletion of 5mCG overexpressed gene bodies (e.g., Grm4) is clearly evident in the MethylC-Seq and oxBS-Seq combined data. Third, although 5hmCG, 5mCA, and 5hmCA are all depleted in heavily heterochromatic foci (Fig. S3*B*), the distributions of 5mCA and 5hmCA in EC are complex and differ from 5hmCG. For example, both 5mCA and 5hmCA are depleted from the highly expressed genes in Fig. 1*E*, indicating that the functions of 5hmC may vary depending on sequence context.

5hmCG Accumulation in Expressed Gene Bodies Results in "Functional Demethylation." The observation that 5hmCG accumulates in active gene bodies is interesting given recent data demonstrating that recombinant MeCP2 does not bind to 5hmCG (14–16). To determine whether this binding specificity is retained in vivo, we performed pull-down assays in cerebellar extracts using beads carrying control, methylated, or hydroxymethylated cytosine residues in a DNA fragment modeled on an endogenous sequence containing both CG and CA dinucleotides, or in precisely sequence-matched DNA probes containing only CA or CG dinucleotides (Fig. 2*A*). These data confirm the in vitro binding specificity of MeCP2 and demonstrate that brain-expressed MeCP2 binds with high affinity to 5mCA, 5hmCA, and 5mCG but that it does not bind to 5hmCG.

To understand the importance of sequence context, methylation, and hydroxymethylation on gene expression we next evaluated the accumulation of each cytosine modification in expressed euchromatic genes relative to those present in HC (Fig. 2*B* and Dataset S2). Highly expressed gene bodies in EC contain elevated levels of 5hmCG, which does not bind MeCP2, and they are depleted for the high-affinity MeCP2 binding residues 5mCG, 5mCA, and 5hmCA (Fig. 2*B*, *Upper*). Genes that are expressed poorly or not at all in euchromatic domains show the opposite properties: They accumulate high-affinity MeCP2 binding sites (5mCG, 5mCA, and

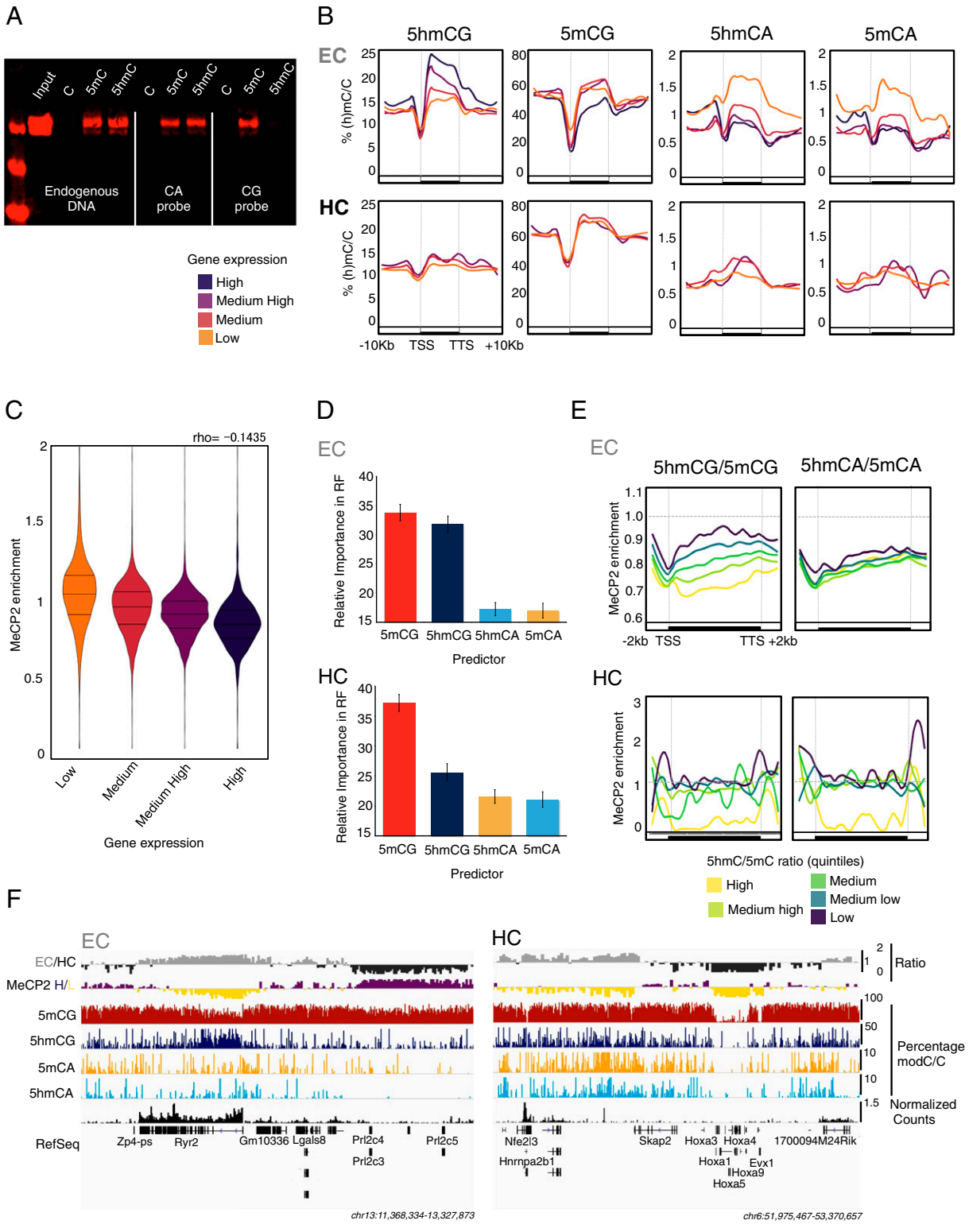

Fig. 2. 5hmCG accumulation in expressed gene bodies results in functional demethylation. (*A*) Western blot of MeCP2 from DNA pulldowns of mouse cerebellar nuclear protein extracts. Endogenous DNA probe contains Cs in both CG and CA contexts; CA and CG synthetic DNA probes were designed to carry Cs only in CA or CG, respectively. 5hmC, 5hmC in DNA probe; 5mC, 5mC in DNA probe; C, DNA probe is unmodified. (*B*) Average percentage of modifications in EC and HC transcripts grouped according to their expression levels by nuclear RNA-Seq. High > 3.5; Medium High = 3.5–1; Medium = 1–0.1; Low < 0.1 reads per kilobase per million (RPKM). TTS, transcription termination site. (*C*) Violin plots of MeCP2 enrichment distribution in transcripts grouped according to their expression levels as described previously. Each pairwise comparison was significant ($P < 10^{-16}$) by Wilcoxon–Mann–Whitney *U* test. (*D*) Relative importance of cytosine modifications in predicting MeCP2 binding in EC and HC transcripts using the random forest regressor algorithm. (*E*) Quantification of MeCP2 enrichment over EC and HC transcripts, grouped in quintiles according to their 5hmC/5mC ratio in CG context (*Left*) and CA context (*Right*). (*F*) Browser representation of EC (gray) and HC (black) regions, MeCP2 enrichment (purple ≥ 1, yellow < 1) in 10-kb bins, average percentage of 5mCG, 5hmCG, 5mCA, and 5hmCA in 100-bp bins, and nuclear RNA-Seq normalized counts. Maximum levels are represented on the top right corner of each track. Genomic coordinates are shown below.

5hmCA), and 5hmCG is not abundant in these transcription units. In HC, most genes are either not expressed or expressed at a very low level (Fig. S2*B*). As expected, these genes are highly methylated, and 5hmCG and 5hmCA do not accumulate (Fig. 2*B*, *Lower*). Granule cell promoters are demethylated relative to gene bodies in both EC and HC (Fig. 2*B*). As previously reported for cortical neurons (12), granule cells also contain a class of developmental regulatory genes that are hypomethylated, for example the HOX group of homeobox genes (Fig. 2*F*, *Right*). These are present in HC, they are not expressed, and they are nearly completely demethylated—thus, they have little or no modified cytosines that can serve as high-affinity binding sites for MeCP2.

Taken together with the binding specificity of MeCP2, the contrasting relationships of 5hmCG and 5hmCA to gene expression suggest distinct roles for these two forms of hydroxymethylation. To determine whether these differences in gene body cytosine methylation predict MeCP2 genome occupancy in granule cells we used the ORGANIC profiling methodology (28). This is a modified native chromatin immunoprecipitation assay that does not use cross-linking or sonication, resulting in improved sensitivity and specificity for mapping of native protein–DNA interactions (Fig. S3 *B* and *C*). We then used the ORGANIC sequencing data to quantify MeCP2 enrichment in the gene body for each of the expressed transcript categories defined above (Dataset S2). As expected, we observed an inverse relationship between gene expression and the level of MeCP2 binding in the gene body (Fig. 2*C*). Thus, highly expressed genes bind low levels of MeCP2 and poorly expressed genes are enriched in MeCP2 binding, in agreement with a repressive role for MeCP2 in transcription (17, 18, 29). To reveal the relative contribution of each type of modified cytosine to MeCP2 binding in expressed genes we applied the random forest regressor algorithm (30). As shown in Fig. 2*D*, this analysis indicates that the most important contributions to MeCP2 gene occupancy come from 5mCG and 5hmCG, and that a significant but less important contribution comes from the levels of 5mCA and 5hmCA. Since 5mCG and 5hmCG act in opposition with respect to MeCP2 binding, we compared MeCP2 binding with the ratios of 5hmCG/5mCG and 5hmCA/5mCA. In each context, genes were binned into quintiles based on the ratio of 5hmC/5mC, and MeCP2 enrichment in each quintile plotted over the gene body. We find that that there is a very strong inverse correlation between 5hmCG/5mCG and binding: Euchromatic genes with a high ratio of 5hmCG/5mCG bind the lowest levels of MeCP2, and those with low ratios of 5hmCG/5mCG are enriched in MeCP2 binding (Fig. 2 *E* and *F*). In contrast, the relative levels of 5hmCA and 5mCA in gene bodies is not predictive. Although depletion of 5mCA and 5hmCA overexpressed gene bodies also occurs (Fig. 2*B*), these data demonstrate that the levels of 5hmCA and 5mCA in expressed genes must at most play a contributory role for MeCP2 binding in active genes. Taken together, our data suggest that the main role of 5hmCG in postmitotic neurons is to convert high-affinity 5mCG MeCP2 binding sites to 5hmCG low-affinity binding sites, thereby reducing the binding of MeCP2 to 5mCG and facilitating transcription. We refer to this process as "functional" demethylation because the consequence of 5mCG oxidation to 5hmCG is functionally equivalent to its demethylation with respect to MeCP2 binding.

In HC, 5mCG is the most important feature for the binding of MeCP2 (Fig. 2*D*, *Lower*). The levels of 5hmCG and 5hmCA are very low, and these genes generally bind high levels of MeCP2. However, there is a class of heterochromatic genes with very low binding of MeCP2 (Fig. 2*E*, *Lower* and Fig. 2*F*). These genes correspond to the hypo- or demethylated developmental regulatory genes previously mentioned (12). The fact that they are not expressed despite very low binding of MeCP2 indicates that the mechanisms responsible for repression of heterochromatic genes do not depend solely on the levels of MeCP2.

Functional Demethylation of Expressed Genes Contributes to the Modulation of Gene Expression by MeCP2. It has been established that loss of MeCP2 in specific neurons leads to dysregulation of a large number of genes (11, 15–18, 30, 31). If functional demethylation of expressed genes contributes to the modulation of gene expression by MeCP2, then both the magnitude and direction of change in $MeCP2_{null}$ granule cells should be related to methylation status and MeCP2 binding in the gene body. To determine if this is the case, we calculated the effect sizes for the change in expression observed in granule cells upon loss of MeCP2 in euchromatic (Fig. 3*A*) and heterochromatic (Fig. S4*A*) genes (Dataset S2). Positive effect sizes are associated with those genes whose expression is increased in the $MeCP2_{null}$. Negative effect sizes represent those genes whose expression decreases in $MeCP2_{null}$ granule cells. Comparison of effect size with features of granule cell genes revealed several aspects of MeCP2 function in neurons. As expected, MeCP2 enrichment in euchromatic gene bodies significantly correlates with both the magnitude and direction of the change in gene expression in $MeCP2_{null}$ granule cells. Thus, expression of genes enriched for MeCP2 increases as a consequence of its loss, whereas those depleted in MeCP2 relative to the genome average decrease expression in $MeCP2_{null}$ neurons. We conclude that MeCP2 binding has a repressive effect on transcription in euchromatic, expressed genes, and that those genes whose expression increases in $MeCP2_{null}$ granule cells are likely direct binding targets of MeCP2. Since the ratio of 5hmCG/5mCG is inversely correlated with MeCP2 occupancy, this relationship is also predictive for alterations in expression in response to loss of the MeCP2. Interestingly, neither the level of expression nor the ratio of 5hmCA to 5mCA determines the magnitude or the direction of change in the $MeCP2_{null}$. Taken together, our data support a model in which accumulation of 5hmCG over gene bodies facilitates transcription as a result of release from the repressive effects of MeCP2 binding.

A second influence on MeCP2 binding and neuronal gene expression, demethylation of promoters and adjacent domains, was also evident in our data (Fig. 4*B*). Given recent studies demonstrating that methylation of H3 lysine 4 can specifically block DNA methyltransferase activity (32, 33), and that genes marked by broad H3K4me3 domains around transcription start sites (TSS) have increased transcriptional consistency (34, 35), we were interested in the relative contributions of promoter demethylation and functional demethylation in expressed neuronal genes. To complement the oxBS data, H3K4me3 coverage was measured by ChIP-Seq and DNA accessibility was assessed by ATAC-Seq. We then calculated H3K4me3 gene coverage as the ratio H3K4me3 ChiP-Seq peak breadth to gene length (Fig. 3*B* and Fig. S3*B*) and confirmed that H3K4me3 coverage reflects gene body accessibility by ATAC-Seq (Fig. 3*D*). We observed that the presence of H3K4me3 within the gene body is associated with strong decreases in methylation, hydroxymethylation, and MeCP2 binding over the region of the gene covered by H3K4me3 (Fig. 3*C*, *Lower* and Fig. 3*E*). This phenomenon does not overcome the influence of 5hmCG for the majority of granule cell expressed genes because 5hmCG accumulation, as well as 5mCG, 5mCA, and 5hmCA depletion, occurs 3′ to the H3K4me3 marked domains in most genes (Fig. 3*C* and Fig. S4*B*). However, a small class of active genes that are covered substantially by H3K4me3 marks, that are demethylated, and bind little MeCP2 escape modulation by 5hmCG-mediated functional demethylation.

5mCA and 5hmCA Accumulate in Intergenic DNA Adjacent to Active Enhancers and Promoters. Non-CG methylation is a prominent feature of neuronal genomes that accumulates as cells become postmitotic and terminally differentiate (10). Recent studies have established that the distribution of modified CH dinucleotides plays an important role in gene expression (9, 12) and MeCP2 function (15). Given the relatively minor contributions

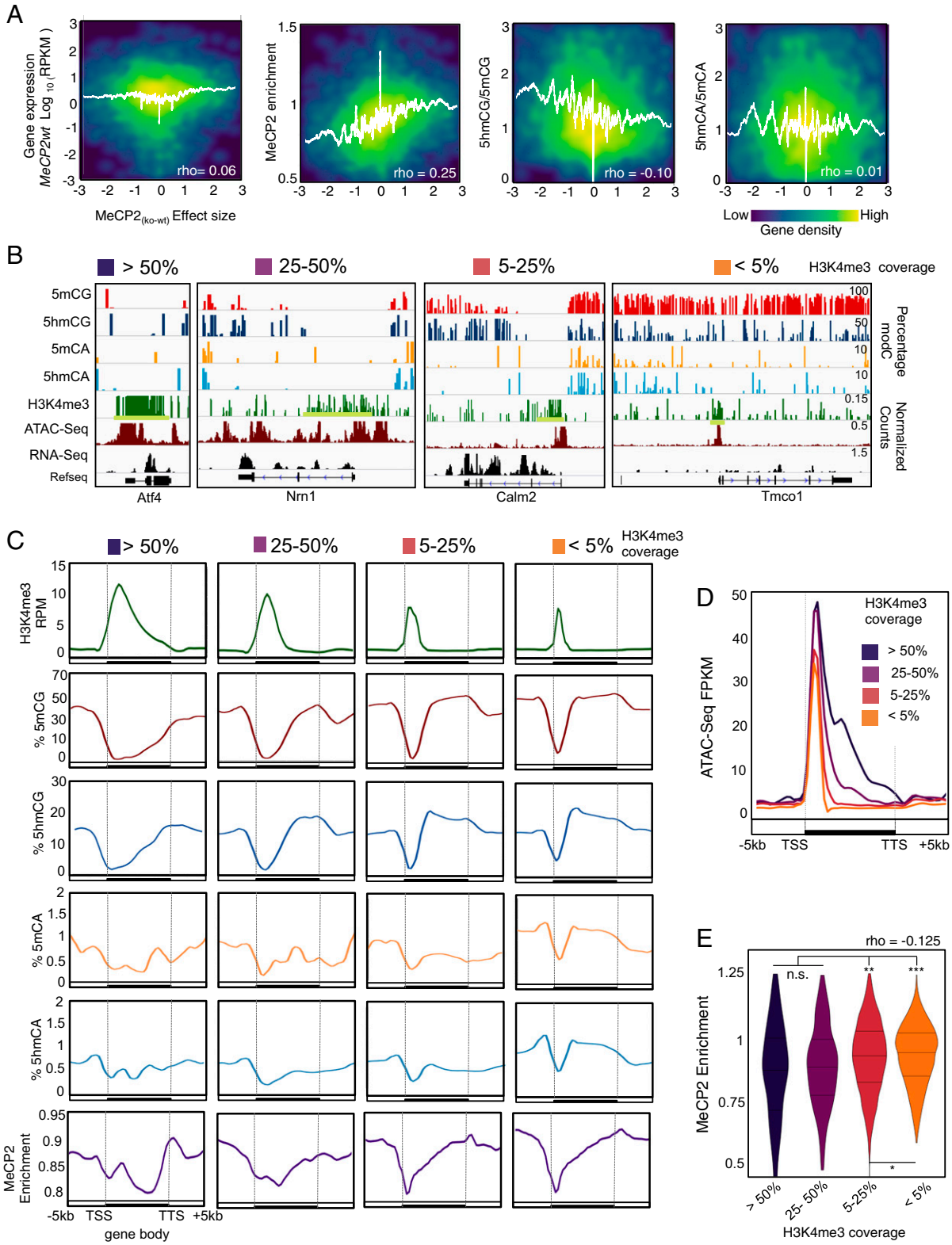

Fig. 3. 5hmCG accumulation in gene bodies influences MeCP2 binding and impacts MeCP2-dependent gene expression regulation. (*A*) Density plots of EC transcripts showing correlations between the effect size [($MeCP2_{null}$ – $MeCP2_{wt}$)/SD pooled] and expression in the WT, MeCP2 enrichment, 5hmCG/5mCG ratio, and 5hmCA/5mCA ratio, respectively. White lines represent moving mean values of *y* axis grouped according to effect size (200 transcripts bins, one transcript step). RPKM, reads per kilobase per million. (*B*) Browser representation of examples in each H3K4me3 coverage category (>50%, 25–50%, 5–25%, and <5%) showing 5mCG, 5hmCG, 5mCA, and 5hmCA normalized enrichment in 100-bp bins, and H3K4me3, ATAC-Seq, and nuclear RNA-Seq normalized counts. Maximum enrichment levels or normalized counts are indicated on the top of the panel. (*C*) Average H3K4me normalized values (*Upper*) percentage of modifications (*Middle*) and MeCP2 enrichment (*Lower*) grouped according to their H3K4me3 coverage as described in *B*. TTS, transcription termination site. (*D*) Average ATAC-Seq normalized values per each H3K4me3 coverage category. (*E*) Violin plots of MeCP2 enrichment distribution in transcripts grouped according to H3k4me3 coverage levels as described previously. Each pairwise comparison is shown (n.s., not significant, $*P < 10^{-8}$, $**P < 10^{-12}$, $***P < 10^{-16}$ by Wilcoxon–Mann–Whitney *U* test).

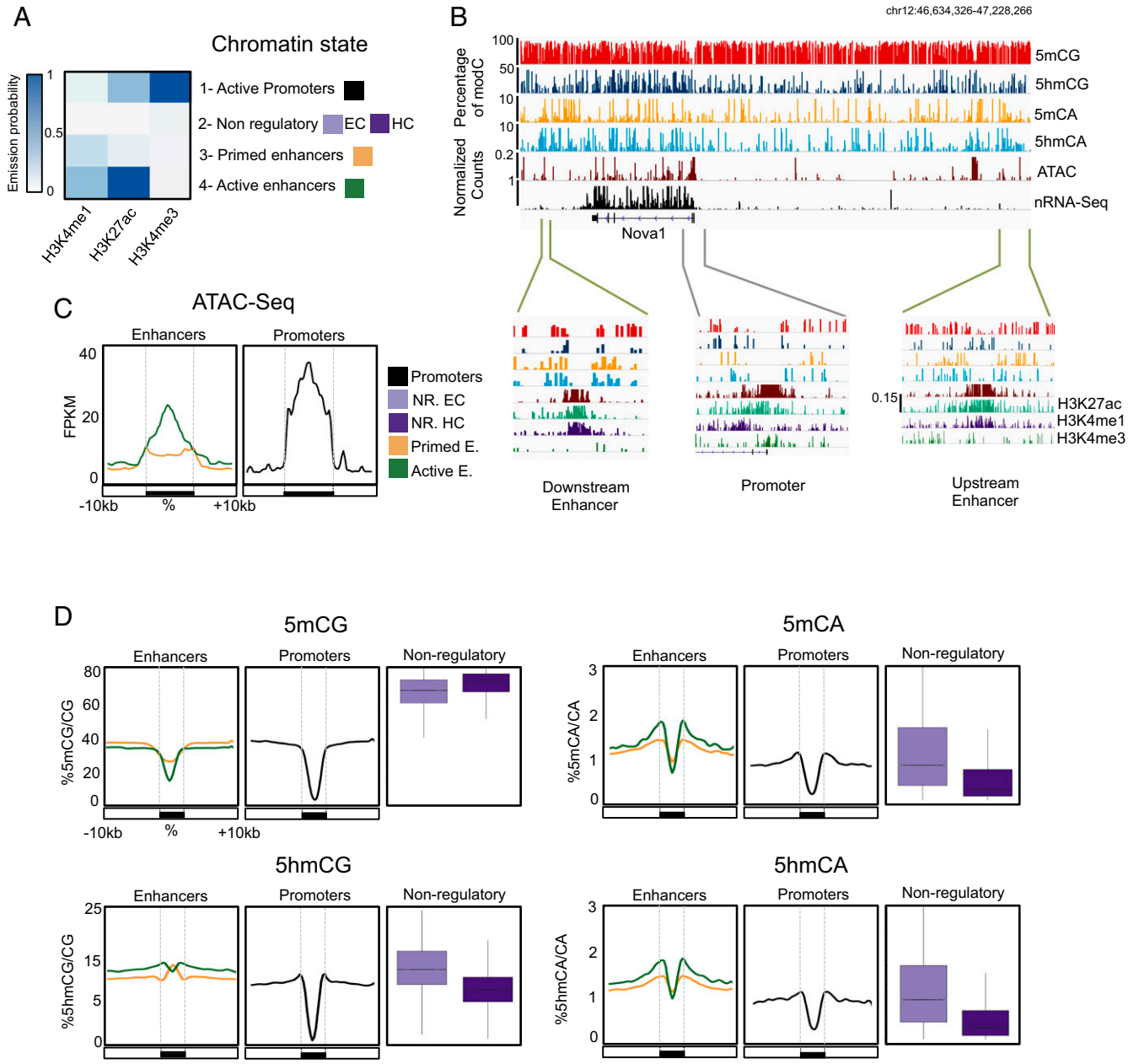

Fig. 4. 5hmCA is preferentially accumulated in active enhancer shores. (*A*) Four chromatin states were generated to define combinatorial patterns of H3K4me3, H3K27Ac, and H3K4me1. Color key reflects the frequencies of each chromatin mark in each state, as the emission probabilities of ChromHMM. (*B*) Browser representation of *Nova1* genomic region showing average percentage of 5mCG, 5hmCG, 5mCA, and 5hmCA in 100-bp bins, and H3K4me3, H3K27Ac, and H3K4me1 ChIP-Seq, ATAC-Seq, and nuclear RNA-Seq normalized counts. Two examples of enhancers (46,655,945-46,678,317 and 47,144,638–47,177,401, respectively) and a promoter (46,798,150-46,836,857) are shown in closer detail in lower panels. Maximum values are represented on the top left corner of each track. (*C* and *D*) Average ATAC-Seq read coverage (*C*) and percentage of cytosine modifications (*D*) within and surrounding promoters and enhancers. In *D*, distributions of *C* modifications per each 1 kb from nonregulatory regions are plotted in boxplots.

of 5mCA and 5hmCA to MeCP2 binding and function within gene bodies, we were interested in analyzing in detail the levels and distribution of 5hmCA relative to 5mCA in intergenic regions (Fig. S5*A*) and its potential impact on MeCP2 regulatory function. We mapped H3K4me1, H3K27Ac, and H3K4me3 in granule cells and then used a multivariate hidden Markov model (ChromHMM) (36) to characterize chromatin states based on these markers. This allowed us to parse the genome into four categories based on previous characterization of the relationships of these marks to regulatory domains: active promoters, active enhancers, primed enhancers, and nonregulatory domains (Fig. 4 *A* and *B*). To confirm that the categories assigned properly distinguish active from inactive regulatory domains, ATAC-Seq was used to identify accessible regions (Fig. 4*C*). As predicted, active enhancers and promoters were highly accessible relative to primed enhancers and nonregulatory regions in the ATAC-Seq data (Fig. 4 *B* and *C*). To understand whether 5mCA and 5hmCA play distinct roles in enhancer or promoter activity, we next studied the distribution of modified cytosines relative to intergenic active enhancers (Fig. 4 *B* and *D* and Dataset S3). As previously reported (12), the core domains of both active enhancers and promoters are demethylated (Fig. 4*D*). The core of

active enhancers, but not primed enhancers, displays also a decrease in 5hmC levels that correspond with an increase of DNA accessibility measured by ATAC-Seq (Fig. 4 *C* and *D*). Consistent with their location in EC (Fig. S5*B*), regions flanking active enhancers in granule cells contain reduced levels of 5mCG relative to the genome average. While there is an enrichment of 5hmCG in these domains, its accumulation near active enhancers is minor relative to its accumulation in expressed gene bodies. In contrast, 5mCA and 5hmCA are highly enriched in 5′ and 3′ flanking DNA in close proximity to active enhancers, similar to the accumulation levels found in silent genes (Fig. 4*D*).

To assess whether 5hmCA accumulation in enhancer flanking DNA is associated with the level of CG methylation, intergenic enhancers situated at different distances upstream and downstream a TSS (Fig. S5 *C* and *D*) were rank-ordered based on the level of 5mCA or 5hmCA occurring 2 kb upstream and downstream of the enhancer core (Fig. 5 *A* and *B* and Dataset S3). Using these rank orders to plot the occurrence in the remaining cytosine modifications, it is evident that 5hmCA and 5mCA accumulation is correlated, and that neither 5mCG nor 5hmCG levels reflect the level of 5hmCA or 5mCA. To understand the significance of 5hmCA and 5mCA accumulation flanking active enhancers we binned them into quartiles based on the levels of 5mCA or 5hmCA (Fig. 5 *B* and *C* and Fig. S5*A*). For each group, the levels of enhancer RNA (eRNA) (37) and TSS RNA for the genes closest to each enhancer were calculated and plotted (Fig. 5*C*). These data demonstrate that genes with high levels of 5mCA or 5hmCA in their enhancers are associated with low levels of eRNA (Spearman rho = −0.33 and −0.28, respectively) and TSS RNA (Spearman rho = −0.13 and −0.10, respectively), whereas those genes with high levels of transcription do not accumulate 5mCA or 5hmCA in nearby enhancers. Although the 5mCG enrichment in these regions is subtle compared with CA modifications, higher levels of 5mCG also correlate with low levels of eRNA and TSS RNA (Spearman rho = −0.3 and −0.15, respectively). Interestingly, 5hmCG accumulation is not significantly associated with either a decrease or an increase of eRNA or TSS RNA (Spearman rho = 0.01 and 0.008, respectively).

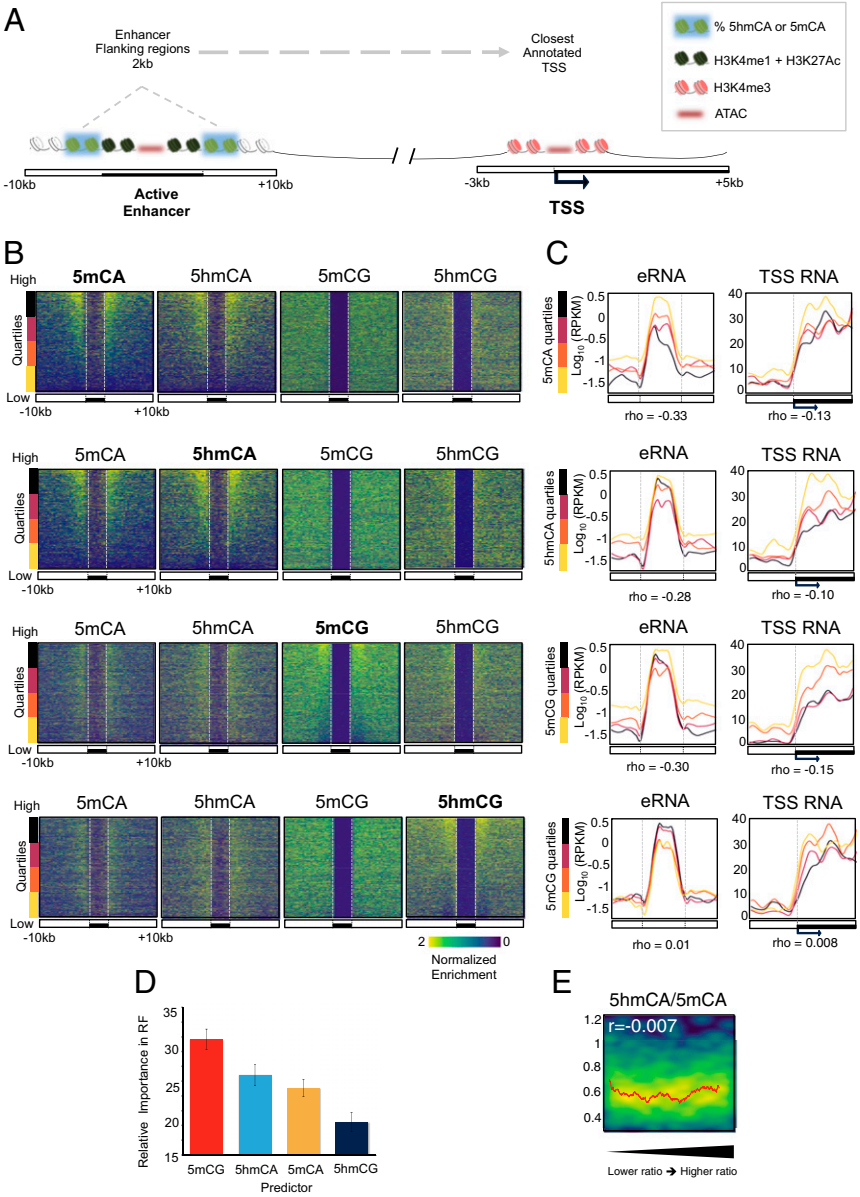

Fig. 5. Oxidation of 5mCA to 5hmCA does not influence MeCP2 binding in active enhancer flanking regions. (*A*) Schematic representation of an active enhancer region and the closest TSS; 2-kb enhancer surrounding regions are highlighted. These regions are considered for detection and functional study of methylation and hydroxymethylation in CA and compared with local and TSS transcription. (*B*) Intergenic enhancers and 10-kb surrounding regions were ranked based on the enrichment values of 5mCA (first row), 5hmCA (second row), 5mCG (third row), and 5hmCG (fourth row) in 2-kb shores. Enrichment of the other cytosine modifications were plotted following this rank (*Upper*). (*C*) Enhancer (*Left*) and associated TSS (*Right*) nuclear RNA-Seq mean coverage. Regions were grouped based on the 5mCA, 5hmCA, 5mCG, and 5hmCG enrichment ranking in 2 Kb surrounding enhancers (from higher, black to lower, yellow), as shown in previous figure. RPKM, reads per kilobase per million. (*D*) Relative importance of cytosine modifications in predicting MeCP2 binding in enhancer flanking regions using random decision forest regressor algorithm. (*E*) Density plot of MeCP2 enrichment in enhancer flanking region ordered from lower to higher 5hmCA/5mCA ratio. Red line represents moving mean values of MeCP2 enrichment grouped according to the ratio (20 region bins, one region step). Pearson's *r* value is indicated.

Oxidation of 5mCA to 5hmCA in Enhancer Shores Does Not Alter MeCP2 Binding. Given the accumulation of 5hmCA and 5mCA adjacent to active enhancers, two important questions arise: Which cytosine modifications contribute to MeCP2 binding in these regulatory domains, and does conversion of 5mCA to 5hmCA in these domains disturb MeCP2 binding? To understand the relative importance of each cytosine modification to MeCP2 binding in regions flanking active enhancers, we used again the random forest classifier algorithm. In contrast to the minor contributions of non-CG modification to MeCP2 binding in expressed gene bodies (Fig. 2*D*), 5hmCA and 5mCA were important determinants of MeCP2 binding in active enhancer flanking regions (Fig. 5*D*). Furthermore, the very strong influence of 5hmCG on MeCP2 binding in gene bodies was not apparent in enhancers. Rather, the high-affinity binding sites for MeCP2 5mCG, 5hmCA, and 5mCA ranked first, second, and third, respectively, in predicting MeCP2 occupancy of these regulatory domains. Given these data, and the fact that 5hmCG acts in opposition to 5mCG in expressed gene bodies, we were interested in the effect of 5hmCA on MeCP2 binding in these enriched domains. If 5hmCA is a "neutral" modification with respect to MeCP2 binding, then the ratio of 5hmCA to 5mCA should have no impact on binding. As shown in Fig. 5*F*, the level of MeCP2 binding does not vary significantly as the ratio of 5hmCA to 5mCA increases in these enhancer domains. This is consistent with the biochemical data demonstrating the MeCP2 binds with high affinity to both 5mCA and 5hmCA. We conclude that oxidation of 5mCA to 5hmCA in intergenic regulatory domains is neutral for MeCP2 binding, thus maintaining functional DNA methylation.

Discussion

The discovery of 5hmC in the mammalian genome (2) has led to important insights into its role as an intermediate in passive and active cytosine demethylation in dividing cell populations (4, 38). Despite the very high levels of 5hmC found in neurons, its role has been difficult to decipher. Here we report two functions for 5hmC in postmitotic neurons. Oxidation of 5mCG to 5hmCG occurs predominantly in expressed gene bodies. This conversion of high-affinity MeCP2 binding sites (5mCG) to low-affinity sites (5hmCG) results in diminished MeCP2 binding, release of its repressive effect, and increased transcription. We refer to this role for 5hmC as functional demethylation because 5hmCG is stably present in these domains, yet it inhibits binding and function of the most abundant neuronal 5mC binding protein, MeCP2. 5hmCA, however, is present in active enhancers and promoters that are also enriched in 5mCA. Since both 5mCA and 5hmCA are high-affinity binding sites for MeCP2, oxidation of 5mCA to 5hmCA maintains MeCP2 binding and its repressive effects on gene expression. We refer to this role for 5hmC as neutral with respect to MeCP2 function because 5hmCA is a stable mark in postmitotic neurons that allows MeCP2 to retain its repressive functions at these important regulatory sites. We propose that 5hmCG-mediated functional demethylation of expressed genes has arisen both to facilitate and stabilize gene expression in postmitotic neurons and that retention of high-affinity binding of 5hmCA by MeCP2 allows functional demethylation to occur without disturbing MeCP2 function in other genomic domains.

The functions of 5hmC identified here are consistent with several recent studies that have established an important relationship between cytosine methylation and MeCP2 function (14–16). Thus, our data confirm the influence of both CG and non-CG methylation on MeCP2 genome occupancy (14–16), and they support the general conclusion that MeCP2 binding within genes and regulatory domains has a repressive effect on gene expression (18, 29). However, the identification of distinct functions for 5hmCG and 5hmCA enhances considerably our understanding of the important roles of hydroxymethylation in sculpting the final landscape of functional cytosine methylation in the brain. Furthermore, our data highlight the fact that an understanding of cytosine modification at CG dinucleotides must incorporate analysis of both 5mCG and 5hmCG because they have opposing impacts on MeCP2 binding, genome occupancy, and function. The demonstration that 5mCG and 5hmCG are the most important determinants of MeCP2 function in expressed gene bodies is consistent with previous studies reporting a positive correlation of 5hmC accumulation and gene expression (9–13, 15). Given the present demonstration that functional demethylation is the mechanism regulating the impact of MeCP2 on expressed gene bodies, our data also provide support for the inverse relationship between non-CG gene body methylation and expression (9, 10, 12, 15), since depletion of these high-affinity binding sites for MeCP2 can also contribute to functional demethylation.

Our understanding of relative roles of 5mCH and 5hmCH in neuronal genomes is less well developed. The present findings confirm reports that 5mCH occurs at maximal levels in CA dinucleotides (9, 10, 12), and they extend these results to identify active enhancers as highly enriched in both 5mCA and 5hmCA. They demonstrate the functional equivalence of 5mCA and 5hmCA for MeCP2 binding and gene expression. Based on these findings, we have proposed that 5hmCA function is neutral with respect to MeCP2 binding and function in vivo. While this model may explain retention of high-affinity binding of MeCP2 to 5hmCA, our data do not exclude other roles for 5hmCA in neurons. It will be interesting to assess, for example, whether enhancers targeted by the MeCP2/NCoR complex (18) are enriched in 5mCA and 5hmCA, whether their relative abundance in these targets influences enhancer activity, and whether the functions of other transcription factors are altered by accumulation of non-CG methylation and hydroxymethylation (39).

The level of expression of any gene reflects the combined action of a large number of complex and interacting genetic and epigenetic factors. We have identified two roles for 5hmC in neurons that help sculpt MeCP2 genome occupancy and determine its consequences for gene expression. Although our data highlight the differential effects 5hmCG and 5hmCA on the binding of this critical neuronal protein, we do not yet understand why the additional regulatory functions of 5hmC and MeCP2 are required principally in vertebrate neurons. Insight into this important issue will require elucidation of the mechanisms targeting Tet oxidases to specific genomic domains and definition of precise roles of MeCP2 in the organization of EC.

Materials and Methods

Mice. All procedures were in accordance with protocols approved by the Rockefeller University Institute Animal Care and Use Committee in accordance with the National Institute of Health guidelines.

Ox-BS-Seq and MethylC-Seq. Ox-BS-Seq and MethylC-Seq experiments were performed using TrueMethyl-Seq reagents (CEGX) and workflow, following the manufacturer's instructions. A total of eight cerebella (four males and four females) were pooled before sorting. Nuclei were digested with 100 μg/mL Proteinase K in the presence of SDS 1% for 2 h at 50 °C. Then, the lysate was treated with RNase A/T1 mix (Thermo Fisher) for 1 h at 37 °C and DNA was extracted with phenol:chloroform:isoamyl alcohol (25:24:1), precipitated in 70% ethanol and dissolved in Tris-EDTA buffer. At least 1 μg of DNA per sample was sonicated using a Covaris-S2 system, and DNA fragments of ∼200 bp were end-repaired using TruSeq DNA Sample kit (Illumina) as per the manufacturer's instructions. Four nanograms of TrueMethyl sequencing spike-in controls were added to the DNA sample before adapter ligation. After TruSeq DNA adapters ligation, libraries were repurified to eliminate potential contaminating compounds using 80% acetonitrile (Sigma-Aldrich) and TrueMethyl-Seq magnetic beads. DNA libraries in bead pellets were then incubated 5 min at 37 °C in 20 μL of denaturing solution. ssDNA libraries were then transferred into two independent tubes: One microliter of oxidation solution (19) was added to the oxBS sample and 1 μL of ultrapure water was added to MethylC sample. After the oxidation step, bisulfite conversion, desulfonation, and purification were repurified in both the oxBS library and MethylC library as per the manufacturer's instructions. PCR amplification suitable for HiSeq2000 sequencing was performed using

TruSeq DNA primers and TrueMethyl-Seq reagents. Quality of libraries was assessed using High Sensitivity D1000 ScreenTape (Agilent) for the 2200 TapeStation system.

Nuclear RNA-Seq. One cerebellum was immediately homogenized in ice-cold homogenization buffer as described above, in the presence of 0.5 mM DTT and RNasin RNase inhibitors (Promega). RNeasy Micro kit (Qiagen) with in-column DNase digestion was used. Ten nanograms of total RNA per sample were converted to cDNA using the NuGEN Ovation RNA-Seq kit (NuGEN) following the manufacturers' instructions. cDNA obtained was quality-scored by RNA 6000 PicoChip (Agilent) for Agilent 2100 Bioanalyzer. One milligram of cDNA per sample was sonicated using Covaris-S2, cDNA fragments of 200 bp were end-repaired and adapters were ligated for HiSEq. 2000 technology using TruSeq DNA Sample kit and following the manufacturer's instructions.

MeCP2 ORGANIC ChIP-Seq. ORGANIC profiling protocol in granule cells was adapted from ref. 20. After sorting, 3×10^6 nuclei were centrifuged (1,200 × *g* 15 min at 4 °C) and resuspended in 500 μL of prewarmed MNase digestion buffer: 15 mM Tris·HCl, pH 8, 1 mM CaCl2, 15 mM NaCl, 60 mM KCl, and 0.5 mM spermidine. Nuclei were digested with 2.5 U of micrococcal nuclease (Worthington) at 37 °C for 6 min. Digestion was stopped on ice by adding 5 mM of EDTA. After spinning, the pellet was resuspended in 500 μL of extraction buffer: 10 mM Tris·HCl, pH 8, 150 mM of NaCl, 0.1% Triton-X 1.5 mM EDTA, pH 8, and 0.5 spermidine. Supernatant (S1) was kept at 4 °C after NaCl and Triton-X were adjusted to 150 mM and 0.1%, respectively. DNA was mechanically extracted using a 26-gauge needle 10 times on ice before 4 °C end-over-end incubation for 2 h. The extract was then centrifuged at 9,500 × *g* and 4 °C for 10 min and supernatant was kept as fraction S2. S1 and S2 were combined and 25 μL was reserved as Input. Immunoprecipitation was performed at 4 °C overnight in an end-to-end rotator by combining the extract (S1+S2) with anti MeCP2 antibody (AB1)-coated beads [50 μL of protein G and 50 μL of protein A Dynabeads (Thermo Fisher) per sample]. After incubation, beads were washed in extraction buffer, DNA was extracted, and a library was prepared for Illumina sequencing [for both immunoprecipitation (IP) and input] as described in *SI Materials and Methods* for ChIP-Seq.

ACKNOWLEDGMENTS. We thank I. Ibañez for comments on the manuscript; B. López for technical assistance with animal breeding; A. Mousa for bioinformatics support; and E. Stoyanova, X. Xu, and F. Piccolo for discussion. We also thank C. Zhao, C. Lai, and N. Nnatubeugo from the Rockefeller University Genomics Resource Center; S. Mazel, S. Han, S. Semova, and S. Tadesse from the Rockefeller University Flow Cytometry Resource Center; and Y. Zhang and G. E. Zentner for technical discussions. This work was supported by the Howard Hughes Medical Institute (N.H.).

1. Suzuki MM, Bird A (2008) DNA methylation landscapes: Provocative insights from epigenomics. *Nat Rev Genet* 9:465–476.
2. Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324:929–930.
3. Tahiliani M, et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324:930–935.
4. Wu H, Zhang Y (2014) Reversing DNA methylation: Mechanisms, genomics, and biological functions. *Cell* 156:45–68.
5. Ito S, et al. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333:1300–1303.
6. He YF, et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333:1303–1307.
7. Globisch D, et al. (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 5:e15367.
8. Münzel M, et al. (2010) Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed Engl* 49:5375–5377.
9. Guo JU, et al. (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci* 17:215–222.
10. Lister R, et al. (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* 341:1237905.
11. Mellén M, Ayata P, Dewell S, Kriaucionis S, Heintz N (2012) MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151:1417–1430.
12. Mo A, et al. (2015) Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* 86:1369–1384.
13. Song CX, et al. (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* 29:68–72.
14. Ayata P (2013) Decoding 5hmC as an active chromatin mark in the brain and its link to Rett Syndrome. PhD thesis (The Rockefeller Univ, New York).
15. Chen L, et al. (2015) MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc Natl Acad Sci USA* 112:5509–5514.
16. Gabel HW, et al. (2015) Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* 522:89–93.
17. Lombardi LM, Baker SA, Zoghbi HY (2015) MECP2 disorders: From the clinic to mice and back. *J Clin Invest* 125:2914–2923.
18. Lyst MJ, et al. (2013) Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat Neurosci* 16:898–902.
19. Booth MJ, et al. (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336:934–937.
20. Kasinathan S, Orsi GA, Zentner GE, Ahmad K, Henikoff S (2014) High-resolution mapping of transcription factor binding sites on native chromatin. *Nat Methods* 11:203–209.
21. Palay SL, Chan-Palay V (1974) *Cerebellar Cortex: Cytology and Organization* (Springer, New York).
22. Qu J, Zhou M, Song Q, Hong EE, Smith AD (2013) MLML: Consistent simultaneous estimates of DNA methylation and hydroxymethylation. *Bioinformatics* 29:2645–2646.
23. Wu H, Zhang Y (2015) Charting oxidized methylcytosines at base resolution. *Nat Struct Mol Biol* 22:656–661.
24. Sun Z, et al. (2015) A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution. *Mol Cell* 57:750–761.
25. Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: Patterns and paradigms. *Nat Rev Genet* 10:295–304.
26. Barth TK, Imhof A (2010) Fast signals and slow marks: The dynamics of histone modifications. *Trends Biochem Sci* 35:618–626.
27. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10:1213–1218.
28. Orsi GA, Kasinathan S, Zentner GE, Henikoff S, Ahmad K (2015) Mapping regulatory factors by immunoprecipitation from native chromatin. *Curr Protoc Mol Biol* 110:21.31.1–25.
29. Lyst MJ, Bird A (2015) Rett syndrome: A complex disorder with simple roots. *Nat Rev Genet* 16:261–275.
30. Rube HT, et al. (2016) Sequence features accurately predict genome-wide MeCP2 binding in vivo. *Nat Commun* 7:11025.
31. Ben-Shachar S, Chahrour M, Thaller C, Shaw CA, Zoghbi HY (2009) Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Hum Mol Genet* 18:2431–2442.
32. Ooi SK, et al. (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448:714–717.
33. Rose NR, Klose RJ (2014) Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim Biophys Acta* 1839:1362–1372.
34. Benayoun BA, et al. (2014) H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* 158:673–688.
35. Chen K, et al. (2015) Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat Genet* 47:1149–1157.
36. Ernst J, Kellis M (2012) ChromHMM: Automating chromatin-state discovery and characterization. *Nat Methods* 9:215–216.
37. Kim TK, et al. (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465:182–187.
38. Lee HJ, Hore TA, Reik W (2014) Reprogramming the methylome: Erasing memory and creating diversity. *Cell Stem Cell* 14:710–719.
39. Sayeed SK, Zhao J, Sathyanarayana BK, Golla JP, Vinson C (2015) C/EBPβ (CEBPB) protein binding to the C/EBP|CRE DNA 8-mer TTGC|GTCA is inhibited by 5hmC and enhanced by 5mC, 5fC, and 5caC in the CG dinucleotide. *Biochim Biophys Acta* 1849:583–589.
40. Doyle JP, et al. (2008) Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* 135:749–762.
41. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14:178–192.
42. Xi Y, Li W (2009) BSMAP: Whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* 10:232.
43. Sun D, et al. (2014) MOABS: Model based analysis of bisulfite sequencing data. *Genome Biol* 15:R38.
44. Song Q, et al. (2013) A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One* 8:e81148.
45. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
46. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
47. Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589.
48. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14:1188–1190.
49. Dobin A, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
50. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.
51. Ghanbarian AT, Hurst LD (2015) Neighboring genes show correlated evolution in gene expression. *Mol Biol Evol* 32:1748–1766.
52. Sullivan GM, Feinn R (2012) Using effect size-or why the P value is not enough. *J Grad Med Educ* 4:279–282.