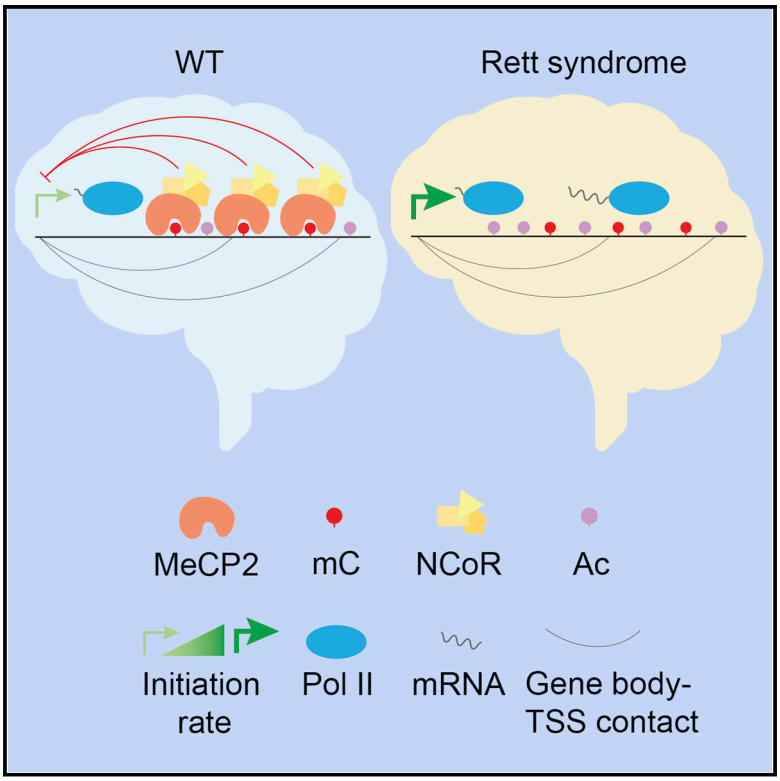
Article

Molecular Cell

MeCP2 Represses the Rate of Transcriptional Initiation of Highly Methylated Long Genes

Graphical Abstract



Authors

Lisa D. Boxer, William Renthal, Alexander W. Greben, ..., Eric C. Griffith, Boyan Bonev, Michael E. Greenberg

Correspondence

michael_greenberg@hms.harvard.edu

In Brief

Mutations in MeCP2 cause the neurodevelopmental disorder Rett syndrome. Although MeCP2 binding is enriched in methylated gene bodies, Boxer et al. find that MeCP2 decreases the rate of transcriptional initiation, but not elongation, of highly methylated long genes in the brain through its interaction with the NCoR co-repressor complex.

Highlights

- MeCP2 represses transcription of highly methylated long genes through NCoR
- Direct measurements of transcriptional initiation and elongation rates in the mouse brain
- MeCP2 reduces transcriptional initiation, not elongation, of highly methylated long genes
- Gene body-TSS contacts position distal MeCP2 molecules at the TSS

Boxer et al., 2020, Molecular Cell 77, 294–309
January 16, 2020
https://doi.org/10.1016/j.molcel.2019.10.032

CellPress

CellPress

Molecular Cell
Article

MeCP2 Represses the Rate of Transcriptional Initiation of Highly Methylated Long Genes

Lisa D. Boxer,[1,3] William Renthal,[1,3] Alexander W. Greben,[1] Tess Whitwam,[1] Andrew Silberfeld,[1] Hume Stroud,[1] Emmy Li,[1] Marty G. Yang,[1] Benyam Kinde,[1] Eric C. Griffith,[1] Boyan Bonev,[2] and Michael E. Greenberg[1,4,*]

[1]Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, MA 02115, USA
[2]Helmholtz Pioneer Campus, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[3]These authors contributed equally
[4]Lead Contact
*Correspondence: michael_greenberg@hms.harvard.edu
https://doi.org/10.1016/j.molcel.2019.10.032

SUMMARY

Mutations in the methyl-DNA-binding repressor protein MeCP2 cause the devastating neurodevelopmental disorder Rett syndrome. It has been challenging to understand how MeCP2 regulates transcription because MeCP2 binds broadly across the genome and MeCP2 mutations are associated with widespread small-magnitude changes in neuronal gene expression. We demonstrate here that MeCP2 represses nascent RNA transcription of highly methylated long genes in the brain through its interaction with the NCoR co-repressor complex. By measuring the rates of transcriptional initiation and elongation directly in the brain, we find that MeCP2 has no measurable effect on transcriptional elongation, but instead represses the rate at which Pol II initiates transcription of highly methylated long genes. These findings suggest a new model of MeCP2 function in which MeCP2 binds broadly across highly methylated regions of DNA, but acts at transcription start sites to attenuate transcriptional initiation.

INTRODUCTION

Mutations in *MECP2* cause Rett syndrome (RTT), a severe neurodevelopmental disorder characterized by progressive loss of speech and social engagement, seizures, and motor disabilities (Amir et al., 1999). *MECP2* encodes a protein that is expressed at ~10-fold higher levels in neurons than other cell types (Skene et al., 2010), in which it preferentially binds methylated DNA and interacts with transcriptional co-repressors (Lewis et al., 1992; Nan et al., 1998). Missense mutations in either the methyl-binding domain (MBD) or transcriptional repressor domain (TRD) of *MECP2* are sufficient to cause RTT (Tillotson et al., 2017; Yusufzai and Wolffe, 2000), suggesting that at least one function of MeCP2 is to repress gene expression in a DNA methylation-dependent manner.

Nevertheless, it has been challenging to understand precisely how MeCP2 regulates transcription because there are numerous small-magnitude up- and downward changes in gene expression when MeCP2 function is disrupted, and it is unclear which of these changes are direct transcriptional consequences of MeCP2 loss and which are secondary to associated cellular or developmental abnormalities (Chahrour et al., 2008; Tudor et al., 2002). Moreover, unlike classic sequence-specific transcriptional repressors that bind to defined sites within gene regulatory regions, MeCP2 binds broadly throughout the genome to methylated DNA at CG and CA dinucleotides (Chen et al., 2015; Cohen et al., 2011; Gabel et al., 2015; Lagger et al., 2017; Skene et al., 2010), complicating attempts to link MeCP2 binding to the repression of specific genes.

Several recent studies searched for characteristic features of the genes that are misregulated with loss of MeCP2 and found that genes that are long (>100 kb) and have high levels of gene body mCA are preferentially upregulated when MeCP2 is mutated and downregulated when MeCP2 is overexpressed (Chen et al., 2015; Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017; Renthal et al., 2018; Stroud et al., 2017; Sugino et al., 2014). The repression of highly methylated long genes by MeCP2 in neurons may have particular relevance to the pathophysiology of RTT because the levels of gene body mCA and MeCP2 protein are higher in neurons compared with other cell types. In addition, the levels of mCA and MeCP2 protein increase significantly as RTT ensues in the postnatal period (Gabel et al., 2015; Lister et al., 2013; Skene et al., 2010). The preferential repression of long genes with high levels of gene body mCA and MeCP2 binding has led to the hypothesis that MeCP2 might function as a transcriptional speed bump, in which it acts as a weak repressor of elongating RNA polymerase (Pol II) (Cholewa-Waclaw et al., 2019; Kinde et al., 2016).

However, the role of MeCP2 in repressing highly methylated long genes remains controversial, as a recent study suggested that MeCP2-dependent repression of long genes is not observed when nascent RNA transcripts are analyzed, attributing the increased expression of mature long mRNAs to a yet-to-be-defined post-transcriptional mechanism (Johnson et al., 2017). It has also been suggested that the observed upregulation of long mRNAs when MeCP2 function is disrupted is an artifact of amplification bias during library preparation for RNA sequencing

(RNA-seq) or microarray experiments (Raman et al., 2018). Last, it has been suggested that MeCP2 might function directly as an activator of transcription or to regulate RNA splicing (Chahrour et al., 2008; Young et al., 2005), so there remains significant controversy about if and how MeCP2 affects the production of nascent RNA transcripts, whether MeCP2 actually functions as a transcriptional repressor, and if so, at what step in the process of mRNA synthesis and processing MeCP2 acts.

In this study, we generate a resource of high-confidence MeCP2-dependent gene expression changes across dozens of biological replicates of wild-type (WT) versus MeCP2-mutant brains to demonstrate that when MeCP2 function is disrupted, the expression of both nascent and mature RNA transcripts derived from highly methylated long genes is significantly increased. By characterizing Pol II binding at base pair resolution in MeCP2-mutant mice, we made the surprising observation that although MeCP2 is distributed throughout the body of highly methylated long genes, MeCP2 acts at the transcription initiation site to limit the synthesis of mRNAs, rather than functioning as a brake on transcriptional elongation. This suggests a model of MeCP2 function in which MeCP2 inhibits the rate of transcriptional initiation via its interaction with transcription start sites (TSSs), gene bodies, and Mb-scale regions up- and downstream of genes—regions that we show are in direct three-dimensional contact with the TSS. The finding that MeCP2 represses the rate of transcriptional initiation of highly methylated long genes suggests that the development of MeCP2 mimetics that selectively inhibit transcriptional initiation might be a useful therapeutic approach for treating RTT.

RESULTS

Upregulation of Highly Methylated Long Genes with Loss of MeCP2 Is Reproducible and Not Due to Amplification Bias

To provide sufficient statistical power to reproducibly detect small-magnitude MeCP2-dependent changes in the expression of nascent and mature mRNA transcripts, we sequenced whole-cell, nuclear, and chromatin-associated RNA from the forebrains of ten 8-week-old male MeCP2 knockout (KO) mice and ten WT littermate control mice (Figures 1A, S1A, and S1B). To avoid batch effects, 20 separate libraries (10 KO and 10 WT) from each RNA fraction were prepared at the same time. We initially focused on the whole-cell RNA and identified 2,870 differentially expressed genes between MeCP2-KO and WT samples (1,368 upregulated, 1,502 downregulated; false discovery rate [FDR] < 0.05) (Figure 1B; Table S1). Because of the large number of replicates, it was possible to achieve a high level of confidence in the statistical significance of the differences we observed. This contrasts with previous studies using a meta-analysis that allowed an examination of trends within the dataset but did not permit an accurate assessment of individual genes (Gabel et al., 2015). Importantly, no significantly misregulated genes were identified when the WT and KO genotypes were shuffled by randomly selecting 5 WT samples and 5 MeCP2-KO samples and comparing them with the remaining 10 samples (Figure 1B), indicating that the identified gene expression changes reflect the disruption of MeCP2 function as opposed to sample-to-sample variability.

The genes significantly upregulated in the MeCP2 KO include genes known to be important for normal brain development and function, such as *Auts2*, *Efna5*, and *Scn1b*, many of which have been observed to be upregulated in human individuals with RTT (Renthal et al., 2018). The genes significantly upregulated in the MeCP2 KO exhibit significantly increased gene length, gene body mCA levels, and gene body MeCP2 binding compared with genes whose expression is unaffected in the MeCP2 KO (Figures 1C and S1C). Moreover, the magnitude of the gene expression differences between the MeCP2-KO and WT samples correlates with both the level of gene body mCA and gene length such that the greater the amount of gene body mCA and the longer the genes, the greater the upregulation in the absence of MeCP2 (Figure 1D). The upregulation of these genes when MeCP2 function is disrupted likely reflects the loss of MeCP2 rather than sample variability, because there is no correlation with mCA levels or gene length for these same genes when the 20 WT and MeCP2-KO samples are shuffled (Figure 1D). Consistent with previous findings, the genes upregulated in MeCP2 KO compared with WT tend to have high levels of mCA, but not mCG, relative to unaffected genes (Figures 1C and S1C; Chen et al., 2015; Gabel et al., 2015). The increased gene length, mCA levels, and MeCP2 binding of the genes upregulated in MeCP2 mutants was also seen when compared with a set of expression-matched, unaffected genes (Figure S1D) as well as when fold-change cutoffs were applied (Figures S1E and S1F).

To determine if the observed length-dependent gene upregulation is due to amplification bias during library preparation (Raman et al., 2018), the same ten MeCP2 KO and ten WT whole-cell RNA samples that were used for RNA-seq and validated using qPCR (Figure S1G; Table S2) were analyzed using NanoString nCounter, an amplification-free gene expression analysis method (Geiss et al., 2008). NanoString was used to assay 200 genes, including 96 genes that are significantly upregulated and 96 genes that are significantly downregulated in the KO compared with WT by RNA-seq (Table S3). The fold change in gene expression between the KO and WT samples as measured using NanoString was highly correlated with that measured using RNA-seq, even for genes greater than 100 kb in length (genes < 100 kb, r = 0.937; genes > 100 kb, r = 0.961; Figures 1E and S1H).

Together, these results demonstrate that the small-magnitude but statistically significant upregulation of highly methylated long genes in MeCP2-KO compared with WT mice is caused by the loss of MeCP2 function and not sample-to-sample variability or amplification bias. Given the higher levels of gene body mCA and MeCP2 binding within the upregulated gene set (Figure 1C), MeCP2's known function as an NCoR-dependent transcriptional repressor, and that known RTT-causing missense mutations disrupt the MeCP2-NCoR interaction (Lyst et al., 2013; Nan et al., 1997), we focused our subsequent analysis on the genes upregulated upon disruption of MeCP2 function. We consider in the Discussion the contribution to RTT of genes that are downregulated upon MeCP2 disruption.

MeCP2 Represses Transcription of Highly Methylated Long Genes

The enrichment of MeCP2 binding and mCA in the genes that are significantly upregulated when MeCP2 function is disrupted

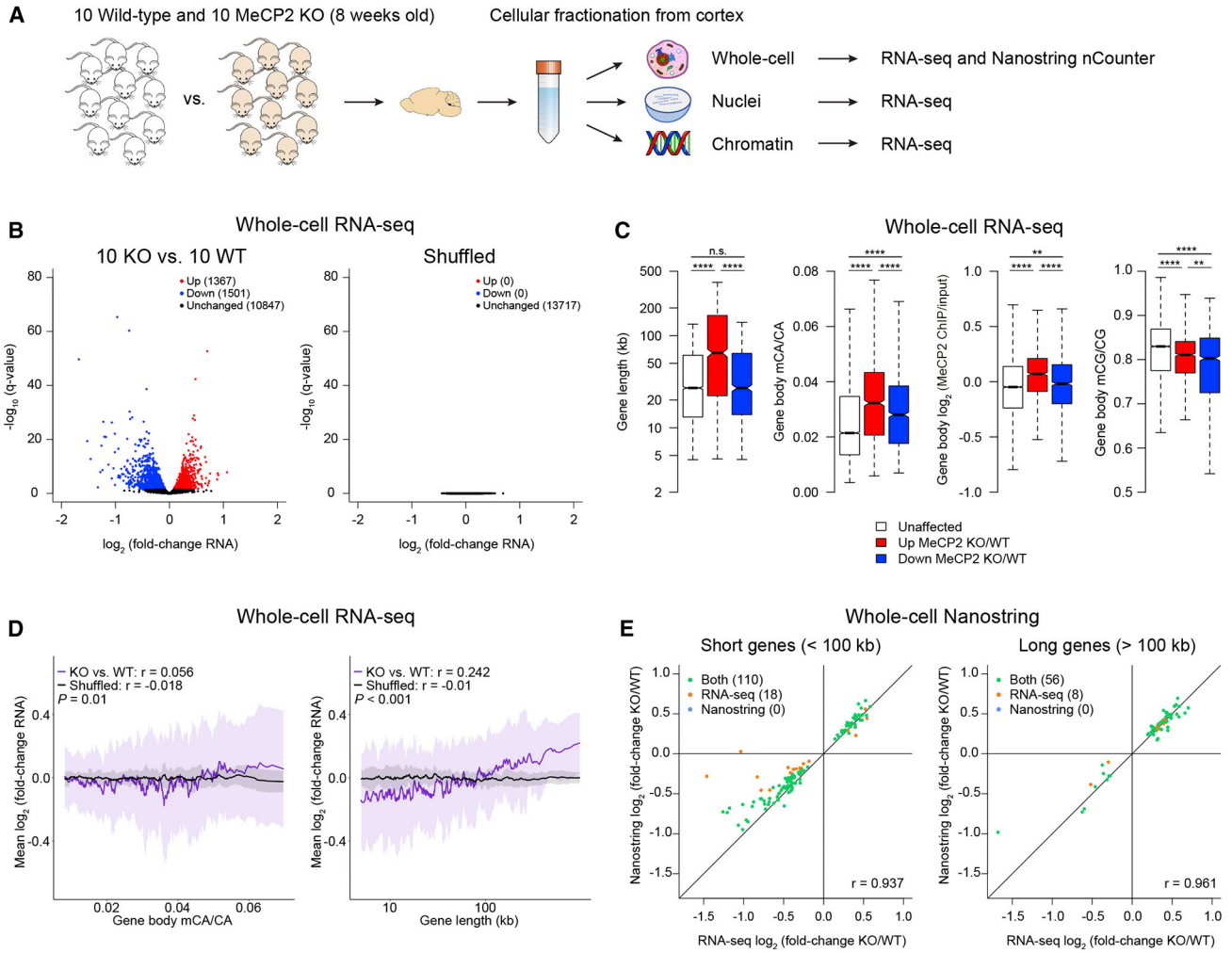


Figure 1. Upregulation of Highly Methylated Long Genes with Loss of MeCP2 Function Is Reproducible and Not Due to Amplification Bias

(A) Schematic of cellular fractionation for RNA-seq from ten WT and ten MeCP2-KO 8-week-old mice.

(B) Differential expression analysis from whole-cell RNA-seq of ten MeCP2 KO versus ten WT (left) or five randomly selected WT and five randomly selected KO versus the remaining ten samples (shuffled, right). Significantly misregulated genes (FDR < 0.05) shown in red (upregulated) or blue (downregulated).

(C) Boxplots of gene length, gene body mCA levels (mCA/CA), gene body MeCP2 ChIP/input, and gene body mCG levels (mCG/CG) of genes unaffected, upregulated, or downregulated in MeCP2 KO versus WT whole-cell RNA-seq. Gene body defined as 3 kb downstream of the TSS to the transcription termination site (TTS) for genes > 4.5 kb in length. n.s., $p > 0.05$; **$p < 0.01$ and ****$p < 0.0001$ (Kruskal-Wallis test with Dunn's post hoc test). Boxplots show median (line), interquartile range (IQR, box), 1.5 x IQR (whiskers), and 1.57 x IQR/$\sqrt{n}$ (notch).

(D) Mean fold changes of significantly misregulated genes from (B) displayed as a function of gene body mCA/CA (left) or gene length (right, x-axis is $\log_{10}$ scale). Pearson correlation between gene misregulation and mCA/CA was significantly greater in KO versus WT than shuffled (KO versus WT, r = 0.056; shuffled, r = −0.018; permutation test, p = 0.01). Pearson correlation between gene misregulation and gene length was significantly greater in KO versus WT than shuffled (KO versus WT, r = 0.242; shuffled, r = −0.01; permutation test, $p < 0.001$). Lines represent mean fold change in expression for 60-gene bins with 6-gene steps; ribbon is SD of each bin.

(E) Pearson correlation of fold change in expression in KO versus WT assayed by NanoString nCounter with fold change in expression in KO versus WT assayed by RNA-seq for short genes (<100 kb, left, r = 0.937) or long genes (>100 kb, right, r = 0.961). Genes significantly misregulated (FDR < 0.05) in both NanoString and RNA-seq are shown in green; genes significantly misregulated only in RNA-seq are shown in orange.

See also Figure S1 and Tables S1, S2, and S3.

suggests that MeCP2 directly binds to and represses the transcription of this set of genes. To investigate this possibility, we analyzed the nuclear and chromatin-associated RNA from ten MeCP2 KO and ten WT mice and identified 1,688 significantly misregulated genes in the nuclear RNA-seq and 1,626 significantly misregulated genes in the chromatin-associated RNA-seq (FDR < 0.05) (Figures 2A and 2B; Table S1). Notably, there were no significantly misregulated genes detected in either fraction upon sample shuffling.

Similar to the findings from our whole-cell RNA-seq, we found that genes significantly upregulated in the MeCP2-KO compared with WT nuclear and chromatin-associated fractions are

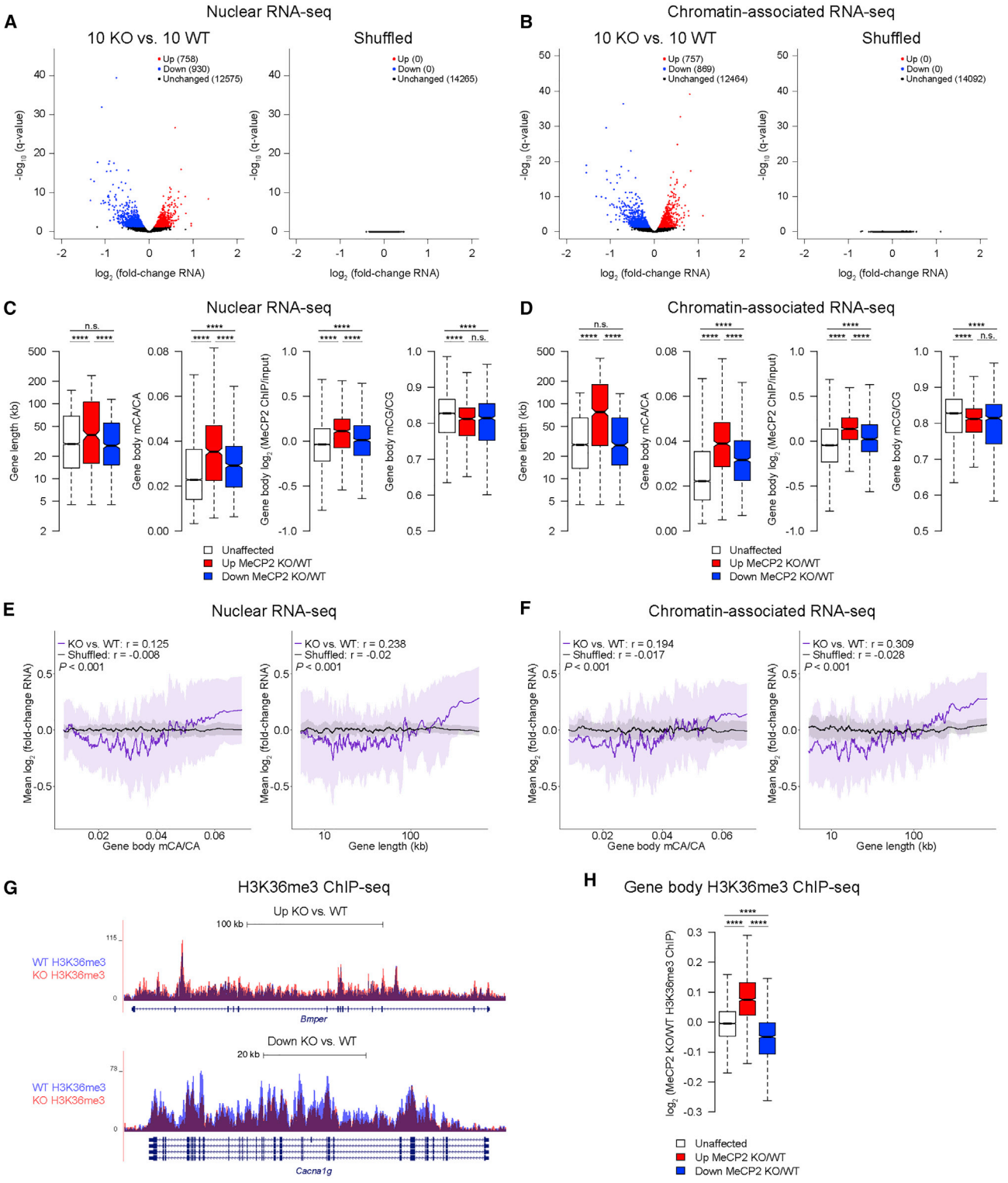


Figure 2. MeCP2 Represses Transcription of Highly Methylated Long Genes

(A and B) Differential expression analysis from nuclear (A) or chromatin-associated (B) RNA-seq of ten MeCP2 KO versus ten WT (left) or shuffled (right). Significantly misregulated genes (FDR < 0.05) shown in red (upregulated) or blue (downregulated).

(legend continued on next page)

significantly longer and have higher gene body mCA levels and MeCP2 binding than unaffected or downregulated genes (Figures 2C, 2D, S2A, and S2B). Moreover, in both the nuclear and chromatin-associated samples, the fold change in expression of the genes that are significantly misregulated in the MeCP2 KO compared with WT correlates with both gene length and gene body mCA levels, whereas there is no correlation with gene length or mCA levels for these same genes when the WT and MeCP2-KO samples are shuffled (Figures 2E and 2F). There is also significant overlap among the up- and downregulated genes across the three cellular fractions, and the fold changes of gene expression across these fractions are strongly correlated, indicating that genes that are transcriptionally upregulated upon MeCP2 disruption are also upregulated as the mRNA matures (Figure S2C).

To further investigate whether the upregulation of highly methylated long genes upon the disruption of MeCP2 function is due to a defect in transcriptional repression, we performed chromatin immunoprecipitation sequencing (ChIP-seq) for the histone modification H3K36me3, which coats transcribed gene bodies in a manner that is correlated with the expression of the gene (Pokholok et al., 2005). When gene body H3K36me3 ChIP-seq signal was compared between the forebrains of ten WT and ten MeCP2-KO mice, we found that relative to the genes that were downregulated or unaffected in the MeCP2-KO brain, there was a significant increase in H3K36me3 in the MeCP2 KO compared with WT across the transcribed regions of genes that are significantly upregulated in the MeCP2-KO brain (Figures 2G and 2H; Table S4). We detected 1,684 genes with significant changes in gene body H3K36me3 in the MeCP2 KO compared with WT, and the fold change in H3K36me3 correlated with both gene body mCA levels and gene length (Figures S2D and S2E). Thus, we find that disruption of MeCP2 function is associated with the upregulation of nascent RNA transcripts from highly methylated long genes, indicating that MeCP2 functions within the nucleus to restrict the transcription of these genes. We consider possible reasons for the discrepancy between our findings and those of Johnson et al. (2017) in the Supplemental Information.

MeCP2-Dependent Repression of Highly Methylated Long Genes Requires Methylated DNA Binding and Interaction with the NCoR Co-repressor Complex

Missense mutations in either the MeCP2 MBD or the TRD are sufficient to cause RTT, so we asked whether these functions of MeCP2 are required for MeCP2 to repress the transcription of highly methylated long genes. Because mutations of the MeCP2 MBD result in an unstable protein (Brown et al., 2016), we turned to a mouse strain in which Dnmt3a, the enzyme responsible for depositing mCA, is disrupted in the brain (*Nestin-cre*; *Dnmt3a*$^{fl/fl}$ mice, referred to as Dnmt3a cKO) (Gabel et al., 2015; Kaneda et al., 2004). Dnmt3a-cKO mice have significantly reduced mCA levels, but only slightly reduced mCG levels, and have decreased MeCP2 binding at genes that have high mCA levels in WT mice (Gabel et al., 2015; Kinde et al., 2016; Stroud et al., 2017), thereby providing a method to explore the effects disrupted MeCP2 DNA binding on gene expression. To this end, we analyzed RNA-seq from the cortex of Dnmt3a-cKO mice and found that the fold change in gene expression in the Dnmt3a cKO compared with WT was significantly higher for genes that are upregulated in the MeCP2 KO versus WT mice (Figure S3A; Table S1). This suggests that MeCP2 binding to mCA is required for its transcriptional repression of highly methylated long genes.

The MeCP2 R306C point mutation gives rise to RTT and specifically disrupts the interaction between MeCP2 and the NCoR co-repressor complex while leaving MeCP2 protein expression and DNA binding intact (Figures 3A–3C; Kruusvee et al., 2017; Lyst et al., 2013). To investigate the role of NCoR recruitment in the regulation of highly methylated long genes, we performed nuclear, chromatin-associated, and whole-cell RNA-seq from the forebrains of ten 8- to 12-week-old male MeCP2 R306C mice and ten WT littermate control mice (Table S1). In all three cellular fractions, the genes that were significantly upregulated in R306C compared with WT were significantly longer and displayed increased gene body mCA levels and MeCP2 binding compared with the downregulated or unaffected genes (Figures 3D–3F and S3B–S3E), with significant overlap among the genes misregulated in MeCP2-KO, MeCP2 R306C, and Dnmt3a-cKO mice (Figure 3G). Moreover, MeCP2-dependent gene expression changes were not observed when comparing ten 1-week-old MeCP2 R306C and ten WT littermate control mice, an age at which mCA levels have not yet accumulated in neurons (Figures 3H–3J). Together, these findings support a model in which the presence of postnatal mCA dictates where MeCP2 and NCoR are recruited to repress transcription.

Loss of MeCP2 Function Leads to Increased Histone Acetylation in Highly Methylated Long Genes

The NCoR co-repressor complex includes the histone deacetylase (Hdac) Hdac3 (Li et al., 2000), suggesting that the upregulation of highly methylated long genes in the absence of MeCP2

(C and D) Boxplots of gene length, gene body mCA/CA, gene body MeCP2 ChIP/input, and gene body mCG/CG in genes upregulated, downregulated, or unaffected in MeCP2 KO versus WT nuclear (C) or chromatin-associated (D) RNA-seq. n.s., $p > 0.05$; ****$p < 0.0001$ (Kruskal-Wallis test with Dunn's post hoc test). Boxplots show median (line), IQR (box), 1.5 x IQR (whiskers), and 1.57 x IQR/$\sqrt{n}$ (notch).

(E and F) Mean fold changes of genes significantly misregulated in MeCP2 KO versus WT nuclear (E) or chromatin-associated (F) RNA-seq displayed as a function of gene body mCA/CA (left) or gene length (right, x-axis is $\log_{10}$ scale). Pearson correlation between gene misregulation and mCA/CA was significantly greater in KO versus WT than shuffled in nuclear (KO versus WT, $r = 0.125$; shuffled, $r = -0.008$; permutation test, $p < 0.001$) and chromatin-associated (KO versus WT, $r = 0.194$; shuffled, $r = -0.017$; permutation test, $p < 0.001$). Pearson correlation between gene misregulation and gene length was significantly greater in KO versus WT than shuffled in nuclear (KO versus WT, $r = 0.238$; shuffled, $r = -0.028$; permutation test, $p < 0.001$) and chromatin-associated (KO versus WT, $r = 0.309$; shuffled, $r = 0.079$; permutation test, $p < 0.001$). Lines represent mean fold change in expression for 40-gene bins with 4-gene steps; ribbon is SD of each bin.

(G) Genome browser tracks of H3K36me3 ChIP-seq from ten WT and ten MeCP2-KO forebrains for an example upregulated (top) or downregulated (bottom) gene in MeCP2 KO versus WT RNA-seq. Reads from ten replicates are combined and displayed as overlay tracks.

(H) Boxplot of KO versus WT H3K36me3 gene body (TSS to TTS) ChIP signal in genes unaffected, upregulated, or downregulated in KO versus WT whole-cell RNA-seq. ****$p < 0.0001$ (Kruskal-Wallis test with Dunn's post hoc test). Boxplot shows median (line), IQR (box), 1.5 x IQR (whiskers), and 1.57 x IQR/$\sqrt{n}$ (notch). See also Figure S2 and Tables S1 and S4.

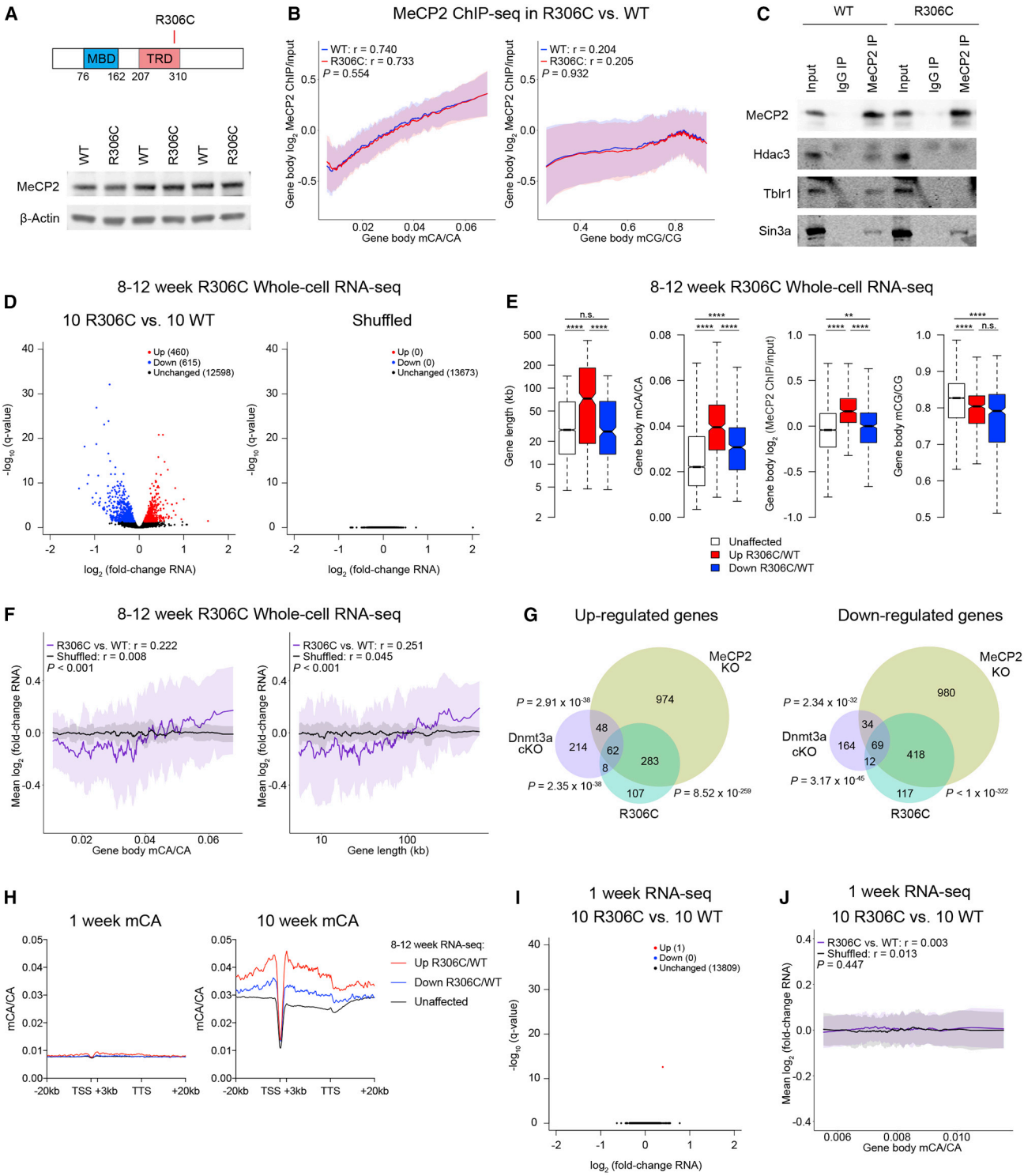


Figure 3. MeCP2-Dependent Repression of Highly Methylated Long Genes Requires Interaction with the NCoR Co-repressor Complex

(A) Top: diagram of MeCP2 protein and domains, showing location of R306C mutation. MBD, methyl-binding domain; TRD, transcriptional repressor domain. Bottom: western blot for MeCP2 compared with β-actin loading control from visual cortex of MeCP2 R306C and WT mice.

(B) Mean gene body MeCP2 ChIP/input from four R306C and four WT forebrains displayed as a function of gene body mCA/CA (left) or mCG/CG (right). Pearson correlation of MeCP2 ChIP with DNA methylation is unaffected by the R306C mutation (mCA/CA: WT, r = 0.740; R306C, r = 0.733; permutation test, p = 0.554; mCG/CG: WT, r = 0.204; R306C, r = 0.205; permutation test, p = 0.932). Lines represent mean $\log_2$ MeCP2 ChIP/input for 400-gene bins with 40-gene steps; ribbon is SD of each bin.

(legend continued on next page)

could be due to decreased Hdac activity (Koerner et al., 2018). Consistent with this idea, we observed significantly reduced Hdac activity associated with MeCP2 immunoprecipitates from R306C compared with WT forebrain (Figure 4A). The reduction in Hdac activity observed in R306C mice is specific to the MeCP2-associated NCoR complex, as no difference in Hdac activity was detected in lysates prior to MeCP2 immunoprecipitation (Figure 4A), and the overall histone acetylation levels were unchanged in R306C compared with WT mice (Figures S3F and S3G). To investigate whether loss of the MeCP2-NCoR interaction leads to changes in histone acetylation specifically at MeCP2-affected loci, we performed ChIP-seq for H3K27ac, H3K9ac, and H4K12ac, histone acetylation marks previously shown to be removed by Hdac3 (Bhaskara et al., 2010; Figure 4B; Table S4). Compared with WT tissue, R306C samples exhibited a small but significant increase in histone acetylation selectively at the TSS and gene bodies of genes upregulated in MeCP2 R306C mice (Figure 4C). Similar effects were also observed for H3K27ac at MeCP2-repressed loci in MeCP2-KO samples (Figures 4D and 4E; Table S4). These findings suggest that the loss of the MeCP2-NCoR interaction leads to increased histone acetylation across genes that are upregulated when MeCP2 function is disrupted, and that in the WT brain, MeCP2 likely represses transcription at least in part via the Hdac activity of its associated NCoR complex.

No Detectable Changes in Transcriptional Elongation Rates with Disruption of MeCP2 Function

We next investigated which steps of transcription are repressed by MeCP2. The enrichment of mCA levels and MeCP2 within gene bodies (Figure S1C) and the gene length-dependent transcriptional upregulation when MeCP2 function is disrupted (Figure 1D) are consistent with a cumulative repressive effect of MeCP2 bound across the gene body. Several possible mechanisms of MeCP2 action that might be important for transcriptional fidelity, but whose disruption could manifest as increased gene transcription when MeCP2 function is disrupted, can be envisioned; MeCP2 might (1) increase the frequency of premature Pol II termination, (2) slow the basal Pol II elongation rate, (3) slow the rate of co-transcriptional splicing, (4) suppress spurious intragenic transcriptional initiation, or (5) act at a distance to suppress transcriptional initiation (Figure 5A).

We sought evidence for one or more of these possible mechanisms, leveraging the data from our RNA-seq experiments. To investigate if MeCP2 increases the frequency of premature Pol II termination, we used the ratio of chromatin-associated RNA-seq reads mapping to the last intron compared with the first intron of each gene (Boswell et al., 2017). If MeCP2 mutations decrease the frequency of premature Pol II termination (increase Pol II processivity), we would expect an increase in the ratio of last to first intron reads in MeCP2-KO compared with WT mice. However, this analysis revealed no significant differences in these ratios between KO and WT (Figure S4A). To investigate whether MeCP2 regulates co-transcriptional splicing, we calculated the fraction of reads corresponding to the un-spliced transcript for each gene using the MeCP2-KO and WT chromatin-associated RNA-seq data (Herzel and Neugebauer, 2015; Khodor et al., 2011). This analysis revealed no significant changes in the fraction of un-spliced transcripts in KO compared with WT (Figure S4B). We next investigated whether MeCP2 suppresses spurious intragenic transcriptional initiation, as has been reported for Dnmt3b in embryonic stem cells (ESCs) (Neri et al., 2017), by asking if there is an increase in the ratio of sequencing reads between the last and first exons in the MeCP2-KO whole-cell RNA-seq. No shift in this ratio was apparent, even for genes that are significantly upregulated in the MeCP2 KO (Figure S4C). Thus, at the level of sensitivity of these assays, we find no evidence that MeCP2 represses transcription by increasing the

(C) Western blot of input, IgG immunoprecipitation (IP), and MeCP2 IP from WT or R306C forebrain probed with antibodies to MeCP2, Sin3a, and NCoR complex components Hdac3 and Tblr1. The interaction of MeCP2 with Hdac3 and Tblr1, but not Sin3a, is disrupted by the R306C mutation.

(D) Differential expression analysis from whole-cell RNA-seq of ten 8- to 12-week-old R306C versus ten WT (left) or shuffled (right). Significantly misregulated genes (FDR < 0.05) shown in red (upregulated) or blue (downregulated).

(E) Boxplots of gene length, gene body mCA/CA, gene body MeCP2 ChIP/input, and gene body mCG/CG of genes unaffected, upregulated, or downregulated in R306C compared with WT whole-cell RNA-seq. n.s., $p > 0.05$; **$p < 0.01$ and ****$p < 0.0001$ (Kruskal-Wallis test with Dunn's post hoc test). Boxplots show median (line), IQR (box), 1.5 x IQR (whiskers), and 1.57 x IQR/$\sqrt{n}$ (notch).

(F) Mean fold changes of the significantly misregulated genes described in (D) displayed as a function of gene body mCA/CA (left) or gene length (right, x-axis is $\log_{10}$ scale). Pearson correlation between gene misregulation and mCA/CA was significantly greater in R306C versus WT than shuffled (R306C versus WT, $r = 0.222$; shuffled, $r = 0.008$; permutation test, $p < 0.001$). Pearson correlation between gene misregulation and gene length was significantly greater in R306C versus WT than shuffled (R306C versus WT, $r = 0.251$; shuffled, $r = 0.045$; permutation test, $p < 0.001$). Lines represent mean fold change in expression for 40-gene bins with 4-gene steps; ribbon is SD of each bin.

(G) Venn diagram of the number of genes significantly upregulated (left) or downregulated (right) in MeCP2 KO, MeCP2 R306C, and Dnmt3a cKO whole-cell RNA-seq. p values describing the overlap between pairs of gene lists were calculated using Fisher's exact test.

(H) Mean mCA/CA at 1 week (left) or 10 weeks (right) across genes unaffected, upregulated, or downregulated in 8 to 12 week MeCP2 R306C versus WT whole-cell RNA-seq. Lines represent mean signal across genes in each list. The −20 to +3 kb and TTS to +20 kb regions are divided into 200 bp bins; TSS +3 kb to TTS is a metagene with 100 equally sized bins per gene.

(I) Differential expression analysis from whole-cell RNA-seq of ten 1-week-old R306C versus ten WT mice. Significantly misregulated genes (FDR < 0.05) shown in red (upregulated) or blue (downregulated). The one significantly misregulated gene, *Cd99l2*, contains six 129 strain-specific SNPs in its 3′ UTR that have remained in *cis* with the R306C allele despite extensive backcrossing with C57/Bl6 mice. Given that *Cd99l2* is not misregulated in RNA-seq from MeCP2-KO mice or chromatin-associated RNA-seq from R306C mice along with the known role of the 3′ UTR in miRNA-dependent mRNA stability, the effect on *Cd99l2* expression is likely independent of MeCP2 and reflects a difference in mRNA stability caused by the 3′ UTR SNPs.

(J) Mean fold changes of all expressed genes described in (I) displayed as a function of 1-week-old gene body mCA/CA. Pearson correlation between gene misregulation and mCA/CA is not significantly greater in R306C versus WT than Shuffled (R306C versus WT, $r = 0.003$; shuffled, $r = 0.013$; permutation test, $p = 0.447$). Lines represent mean fold change in expression for 500-gene bins with 50-gene steps; ribbon is SD of each bin.

See also Figure S3 and Tables S1 and S3.

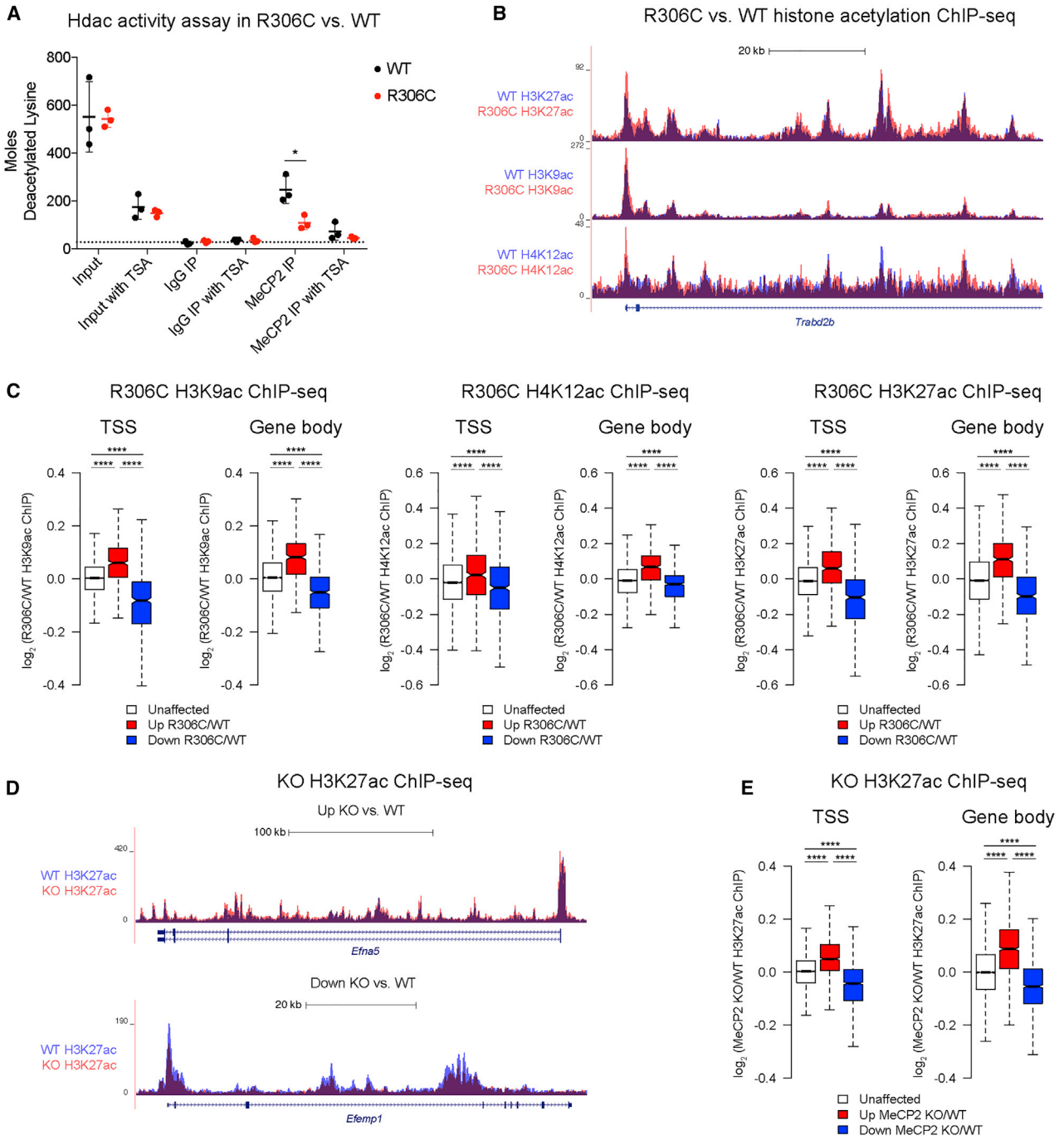


Figure 4. Loss of MeCP2 Function Leads to Increased Histone Acetylation in Highly Methylated Long Genes

(A) Histone deacetylase (Hdac) activity of nuclear input, IgG control IP, or MeCP2 IP from WT or R306C forebrain, either untreated or treated with the Hdac inhibitor trichostatin A (TSA, 1 μM). Dashed line indicates signal from wells without Hdac enzyme. Lines represent mean of three replicates; error bars are SD. Hdac activity is significantly reduced in R306C compared with WT MeCP2 IP ($p = 0.02$), but not in the input ($p = 0.93$). Two-tailed unpaired t test.

(B) Genome Browser tracks of H3K27ac (n = 2), H3K9ac (n = 5), and H4K12ac (n = 3) ChIP-seq from WT and R306C forebrain for an example gene upregulated in R306C versus WT RNA-seq. Reads from replicates are combined and displayed as overlay tracks.

(C) Boxplots of R306C versus WT H3K9ac (left), H4K12ac (middle), and H3K27ac (right) ChIP-seq signal in TSS (−1 to +1 kb) or gene body (TSS +3 kb to TTS) of genes unaffected, upregulated, or downregulated in R306C versus WT whole-cell RNA-seq. ****$p < 0.0001$ (Kruskal-Wallis test with Dunn's post hoc test). Boxplots show median (line), IQR (box), 1.5 x IQR (whiskers), and 1.57 x IQR/$\sqrt{n}$ (notch).

(legend continued on next page)

frequency of premature Pol II termination, slowing co-transcriptional splicing, or suppressing intragenic initiation.

To investigate a possible effect of MeCP2 on transcriptional pausing, we used precision run-on sequencing (PRO-seq) (Kwak et al., 2013). Chromatin from brain tissue of MeCP2-KO, R306C, and Dnmt3a-cKO mice and their respective WT controls was subjected to an *ex vivo* transcriptional run-on assay using chain-terminating biotin-NTPs, which, after purification and sequencing, identifies where Pol II is bound at single-base pair resolution at the time of tissue harvest (Figure S4D). If mCA-bound MeCP2 functions as a speed bump to slow transcriptional elongation, we would expect to observe increased Pol II occupancy directly upstream of mCA sites enriched for MeCP2 binding that is lost with disruption of MeCP2. After confirming that mCA patterns do not change when MeCP2 function is disrupted (Figure S4E) and that we could detect Pol II accumulation at known sites (Figure S4F; Mayer et al., 2015), we analyzed PRO-seq signal at intragenic mCA sites (Lister et al., 2013), but found no evidence of Pol II pausing or a change in occupancy when MeCP2 or Dnmt3a function was disrupted (Figure 5B).

Although the lack of change in pausing over mCA sites suggested that MeCP2 does not promote Pol II pausing at mCA sites within gene bodies, it remained possible that disruption of MeCP2 function and concomitant increases in histone acetylation accelerate transcriptional elongation by creating a more permissive chromatin environment throughout the entire gene body. To measure rates of transcriptional elongation directly, we performed PRO-seq after treating mice with flavopiridol (FP), an inhibitor of the transition of paused Pol II to active elongation. FP does not affect elongating Pol II, so at increasing time points after FP treatment, a depletion of Pol II (negative wave) is observed within the gene body at increasing distances from the TSS, with the rate of displacement of this wave front serving as a measure of elongation rate (Jonkers et al., 2014). To test whether MeCP2 regulates transcriptional elongation, FP was injected into MeCP2-KO or WT mice for 20, 40, or 60 min prior to nuclear isolation from forebrain tissue and PRO-seq analysis (Figures 5C and 5D). Elongation rates could be calculated for 341 genes greater than 200 kb in length, yielding values between 2 and 3.5 kb/min that correlated positively with gene body H3K27me2 and first intron length (Figures 5E–5G and S4G; Table S5), consistent with elongation rates and genome-wide correlations previously measured in cultured cells (Duffy et al., 2018; Jonkers et al., 2014; Veloso et al., 2014). Comparing elongation rates in the MeCP2 KO with those in WT identified no significant difference between genes upregulated in the MeCP2 KO compared with the downregulated or unaffected genes (Figure 5H). There was also no strong relationship between WT elongation rates and gene body mCA levels or MeCP2 ChIP signal (Figures 5G and S4G), further suggesting that MeCP2 does not function to measurably restrain transcriptional elongation.

MeCP2 Reduces the Rate of Pol II Initiation at Highly Methylated Long Genes

The lack of evidence for MeCP2-dependent regulation of transcriptional elongation led us to reconsider our hypothesis that MeCP2 functions as a speed bump within gene bodies. In this regard, in MeCP2-KO, R306C, and Dnmt3a-cKO mice compared with their WT controls, we observed significantly more PRO-seq signal at both the TSS and gene bodies of genes that are significantly upregulated in these mice compared with the downregulated or unaffected genes (Figures 6A–6C; Table S4). We also observed similar effects in MeCP2 KO and R306C mice using Pol II ChIP-seq (Figures S5A–S5C; Table S4). Together, these findings indicate that the loss of MeCP2 function leads to increased Pol II binding at the TSS of highly methylated long genes, raising the possibility that MeCP2 might repress transcriptional initiation.

To directly measure the rate of transcriptional initiation for MeCP2-regulated genes in the brains of WT and MeCP2-KO mice, we quantified TSS PRO-seq signal at various times after FP treatment. Because FP inhibits the transition of Pol II to active elongation, but not initiation, the rate of transcriptional initiation can be calculated as the increase in TSS PRO-seq signal over time (Figures 6D and 6E). The calculated initiation rates in WT mice positively correlate with gene expression, TSS H3K27ac, and Pol II binding, and negatively correlate with TSS MeCP2 binding and methylation (Figure 6F). Comparing initiation rates in MeCP2-KO and WT samples, we found increased initiation rates in the MeCP2 KO specifically for the set of genes upregulated in the MeCP2 KO (Figures 6G and S5D; Table S5). In addition, the fold change in gene misregulation in MeCP2 KO compared with WT directly correlated with the fold change in initiation rate in MeCP2 KO compared with WT (Figure 6H). These findings, together with the observation by ChIP-seq that there is increased Pol II and H3K27ac at the TSS of genes that are upregulated in the MeCP2 KO, suggest that MeCP2 primarily functions to repress the rate of transcriptional initiation of highly methylated long genes.

MeCP2 May Act from a Distance to Repress Transcriptional Initiation

The observation that MeCP2 represses transcriptional initiation led us to revisit the distribution of MeCP2 binding with respect to the TSS of each gene. In this regard, relative to the MeCP2 KO, we observe an enrichment of MeCP2 ChIP signal in WT forebrain at the TSS and regions flanking the TSS of genes upregulated in MeCP2 KO compared with WT (Figures S5E–S5H). Although one interpretation of these findings is that MeCP2 at the TSS directly represses transcriptional initiation, mCA and MeCP2 are also enriched within gene bodies and in Mb-scale domains up- and downstream of gene bodies of the genes that are upregulated in the MeCP2 KO (Figures S6A–S6C). Moreover, the chromatin in these large domains of mCA and MeCP2 becomes hyperacetylated in MeCP2-KO compared with WT mice (Figure S6D), suggesting

(D) Genome Browser tracks of H3K27ac ChIP-seq from ten WT and ten MeCP2-KO forebrain for an example upregulated (top) or downregulated (bottom) gene in MeCP2 KO versus WT RNA-seq. Reads from ten replicates are combined and displayed as overlay tracks.

(E) Boxplots of MeCP2 KO versus WT H3K27ac ChIP-seq signal in TSS (−1 to +1 kb) or gene body (TSS +3 kb to TTS) for genes unaffected, upregulated, or downregulated in KO versus WT whole-cell RNA-seq. ****$p < 0.0001$ (Kruskal-Wallis test with Dunn's post hoc test). Boxplots show median (line), IQR (box), 1.5 x IQR (whiskers), and 1.57 x IQR/$\sqrt{n}$ (notch).

See also Figure S3 and Table S4.

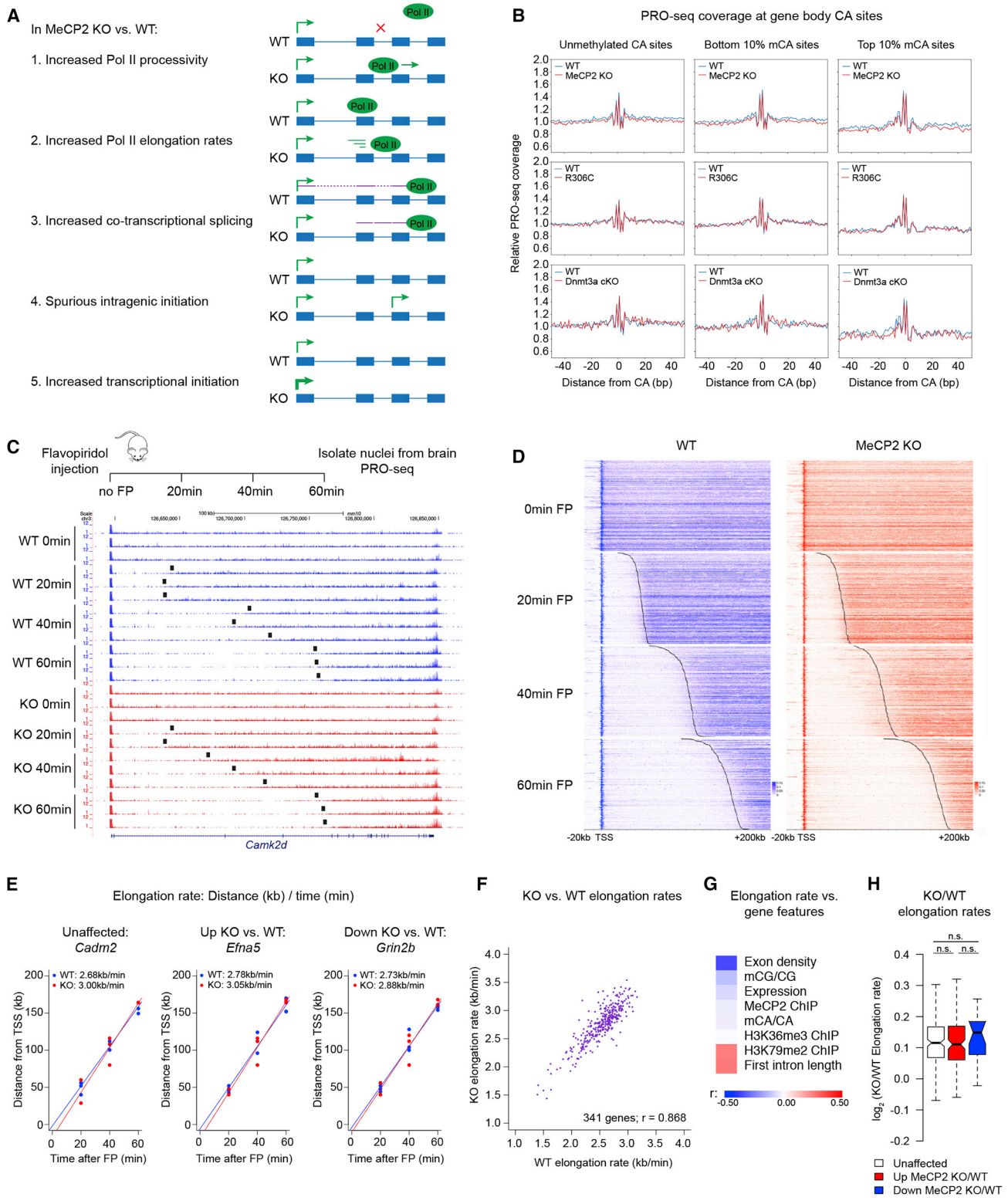


Figure 5. No Detectable Changes in Transcriptional Elongation with Disruption of MeCP2 Function

(A) Possible mechanisms for MeCP2-dependent transcriptional repression of highly methylated long genes.

(B) Analysis of Pol II pausing at gene body mCA sites. Aggregate plots of PRO-seq read coverage relative to gene average PRO-seq coverage for WT and MeCP2 KO (top), WT and R306C (middle), and WT and Dnmt3a cKO (bottom), in introns of expressed genes, centered on CA sites that are unmethylated (left), the 10% lowest mCA/CA levels (center), and the 10% highest mCA/CA levels (right).

(legend continued on next page)

that MeCP2 may repress transcription across Mb-scale domains of chromatin proportionally to the level of mCA and amount of MeCP2 bound, possibly by acting on the TSS from a distance through three-dimensional contacts.

To further test this hypothesis, we performed Hi-C on the forebrains of MeCP2-WT and KO mice. MeCP2 does not appear to regulate the formation of higher order chromatin structure, as the genome-wide distribution of chromatin contacts, A-B compartments, and topologically associating domains (TADs) was highly similar between WT and KO mice (Figures 7A–7D, S7A, and S7B). We next investigated if there are contacts between the TSS and up- and downstream regions that might explain how MeCP2 bound there might be positioned to repress initiation. We observed an enrichment of TSS contacts, particularly in the direction of the gene body, that appeared strongest for the genes upregulated in MeCP2-KO compared with WT mice (Figure 7E). These TSS-gene body contacts are correlated with the expression level of genes, as suggested previously (Lee et al., 2016; Larkin et al., 2012), and are observed in the adult brain and embryonic cortical neurons but not in ESCs (Bonev et al., 2017; Figure S7C). In both MeCP2 WT and KO forebrain, we observed significantly greater observed/expected TSS-gene body contacts in the genes upregulated in MeCP2-KO mice compared with expression-matched, unaffected genes (Figures 7F and 7G). In addition, because the genes upregulated in MeCP2-KO mice tend to be long, they have a greater number of total observed TSS-gene body contacts than downregulated or unaffected genes (Figure S7D). Interestingly, there is no difference in TSS-gene body contacts between MeCP2-KO and WT mice (Figures 7F and 7G), suggesting that MeCP2 is not required for the formation of TSS-gene body contacts but that these contacts may position MeCP2 bound within the gene body so that it can repress initiation at the TSS. Thus, genes with more TSS-gene body contacts may be particularly susceptible to the repressive effects of gene body MeCP2 on initiation.

DISCUSSION

Given the broad distribution of MeCP2 binding to chromatin and the relatively subtle changes in mRNA levels when MeCP2 function is disrupted, it has been difficult to understand the mechanisms by which MeCP2 regulates gene expression. The previous finding that genes that are long and have high levels of gene body mCA are preferentially upregulated with loss of MeCP2 (Chen et al., 2015; Gabel et al., 2015; Kinde et al., 2016; Sugino et al., 2014) provided a potential inroad to address this issue and led to the hypothesis that MeCP2 binding to gene body mCA sites might inhibit transcriptional elongation (Cholewa-Waclaw et al., 2019; Kinde et al., 2016). However, these observations largely relied on a meta-analysis of previously published studies or small sample sizes. Here, we generated a resource of high-confidence MeCP2-regulated genes using dozens of biological replicates and multiple independent methods of analysis. Together, these approaches across two animal models of RTT all lead to the same conclusion: MeCP2 is a weak repressor of the transcription of highly methylated long genes (see also Clemens et al., 2019 [this issue of *Molecular Cell*]).

Although MeCP2 can bind mCG with high affinity to regulate gene expression in some contexts, we find that MeCP2-dependent gene expression changes in mouse cortex are primarily associated with higher levels of mCA and occur only after the accumulation of mCA in neurons. Moreover, there is significant overlap between the genes misregulated in MeCP2 and Dnmt3a mutant mice, suggesting that mCA deposition is required for MeCP2-dependent gene repression in the mature brain. Recent mathematical modeling of MeCP2 function suggested that MeCP2 might function by slowing transcriptional elongation (Cholewa-Waclaw et al., 2019), but our genome-wide experimental measurements of transcriptional elongation and initiation rates in the brain led to the surprising observation that MeCP2 inhibits transcriptional initiation of highly methylated long genes through its interaction with the NCoR co-repressor, without having any detectable effect on transcriptional elongation. Consistent with this idea, decreased binding of the NCoR complex component Hdac3 at the TSS has been reported in MeCP2-deficient hippocampus (Nott et al., 2016).

There are several mutually compatible mechanisms by which MeCP2 might regulate transcriptional initiation. Given that in WT mice, MeCP2 is present at the TSS of genes upregulated in MeCP2-mutant mice, MeCP2 could act locally at the TSS via NCoR to repress initiation, possibly by inhibiting the recruitment of Pol II to the TSS. Such a model, however, does not account for MeCP2 enrichment in gene bodies and over Mb-scale domains up- and downstream of genes that are upregulated in

(C) MeCP2-KO and WT mice were injected with flavopiridol (FP) for 0, 20, 40, or 60 min, and nuclei were isolated for PRO-seq. Genome Browser tracks show PRO-seq signal from MeCP2 KO and WT forebrain after FP treatment. Black bars indicate location of identified transcription wave front.
(D) Heatmap of PRO-seq signal in MeCP2 KO and WT after 0, 20, 40, or 60 min of FP. Each row represents an individual gene. Black lines indicate location of identified transcription wave front. Scale bar indicates PRO-seq read density per base pair.
(E) Pol II elongation rates were calculated for each gene as the change in distance of the transcription wave front (kb) over time of FP treatment (min). Plots show WT and KO transcription wave fronts at each time point for representative unaffected, upregulated, or downregulated genes. The slope of the linear fit represents the elongation rate.
(F) Scatterplot of calculated elongation rates in MeCP2 KO versus WT. Genes > 200 kb with an R^2 of the linear fit > 0.9 are included. r value corresponds to Pearson correlation.
(G) Heatmap of Pearson correlations between WT elongation rates and gene features. Exon density is number of exons per kilobase; expression is RPKM (reads per kilobase of transcript per million mapped reads) from whole-cell RNA-seq. mCG/CG, mCA/CA, and MeCP2 ChIP/input levels calculated from TSS +3 kb to TTS; H3K36me3 and H3K79me2 ChIP constitute the TSS to TTS regions.
(H) Boxplot of MeCP2 KO versus WT elongation rates for genes unaffected (194 genes), upregulated (126 genes), or downregulated (21 genes) in MeCP2 KO versus WT whole-cell RNA-seq. n.s., $p > 0.05$ (Kruskal-Wallis test with Dunn's post hoc test). Boxplot shows median (line), IQR (box), 1.5 x IQR (whiskers), and 1.57 x IQR/$\sqrt{n}$ (notch).
See also Figure S4 and Table S5.

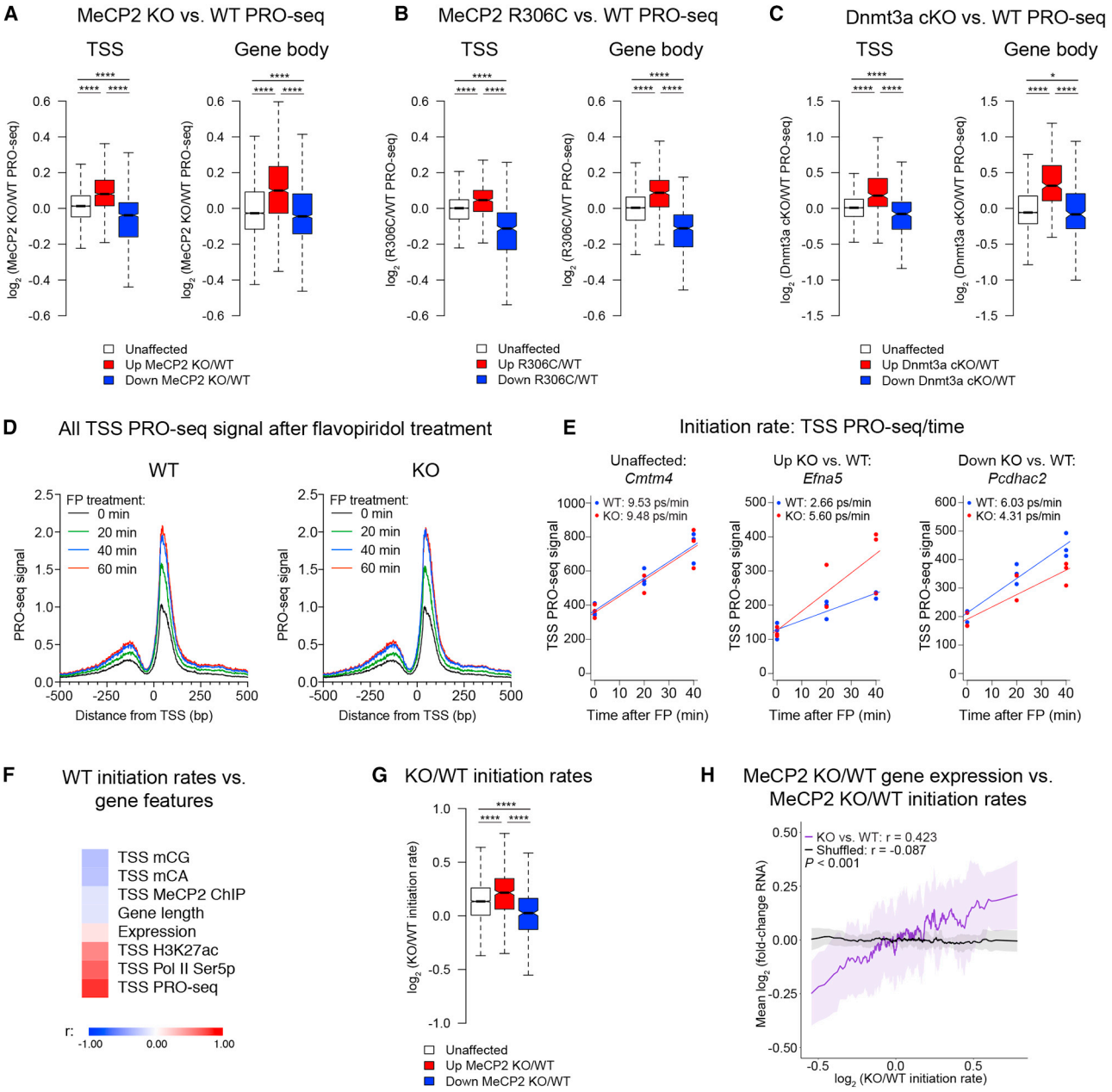


Figure 6. MeCP2 Reduces the Rate of Pol II Initiation at Highly Methylated Long Genes

(A–C) Boxplots of MeCP2 KO versus WT (A), R306C versus WT (B), and Dnmt3a cKO versus WT (C) PRO-seq signal in TSS (−1 to +1 kb, left) or gene body (+1 kb to TTS, right) of genes unaffected, upregulated, or downregulated in MeCP2 KO versus WT (A), R306C versus WT (B), and Dnmt3a cKO versus WT (C) whole-cell RNA-seq. *$p < 0.05$ and ****$p < 0.0001$ (Kruskal-Wallis test with Dunn's post hoc test). Boxplots show median (line), IQR (box), 1.5 x IQR (whiskers), and 1.57 x IQR/$\sqrt{n}$ (notch).

(D) Aggregate plots of TSS PRO-seq signal in MeCP2 KO or WT at 0, 20, 40, or 60 min after FP injection. Lines represent mean signal of two or three replicates.

(E) Pol II initiation rates were calculated as change in TSS PRO-seq signal (ps) over time of FP treatment (min). Plots show WT and KO TSS PRO-seq signal at each time point for representative unaffected, upregulated, or downregulated genes. The slope of the linear fit represents the rate of initiation.

(F) Heatmap of Pearson correlations between WT initiation rates and gene features. mCG, mCA, MeCP2 ChIP, H3K27ac ChIP, Pol II Ser5p ChIP, and 0 min FP TSS PRO-seq signal are shown for −1 to +1 kb relative to TSS. Expression is RPKM from whole-cell RNA-seq.

(G) Boxplot of MeCP2 KO versus WT initiation rate for genes unaffected (3667 genes), upregulated (412 genes), or downregulated (339 genes) in MeCP2 KO versus WT whole-cell RNA-seq. ****$p < 0.0001$ (Kruskal-Wallis test with Dunn's post hoc test). Boxplot shows median (line), IQR (box), 1.5 x IQR (whiskers), and 1.57 x IQR/$\sqrt{n}$ (notch).

(H) Mean fold changes of genes significantly misregulated in MeCP2 KO versus WT whole-cell RNA-seq displayed as a function of the fold change in MeCP2 KO versus WT initiation rates. Pearson correlation between gene misregulation and initiation rate misregulation was significantly greater in KO versus WT than shuffled (KO versus WT, $r = 0.423$; shuffled, $r = -0.087$; permutation test, $p < 0.001$). Lines represent mean fold change in expression for 30-gene bins with 3-gene steps; ribbon is SD of each bin.

See also Figures S5 and S6 and Tables S4 and S5.

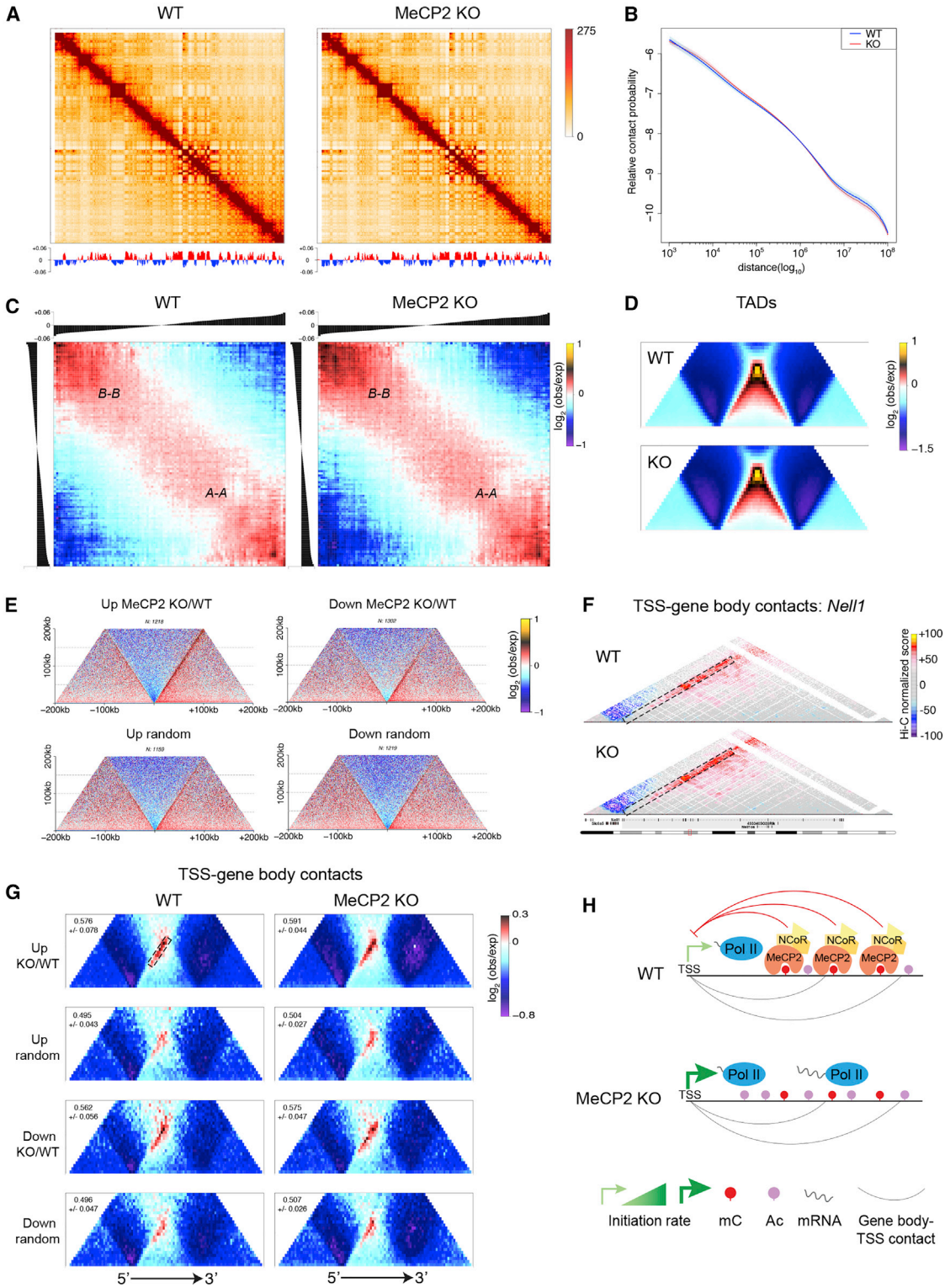


Figure 7. MeCP2 May Act from a Distance to Repress Transcriptional Initiation

(A) Hi-C from WT and MeCP2-KO forebrain tissue (n = 3). Knight-Ruiz normalized observed contact matrices for chr3 at 250 kb resolution and the first eigenvector at 100 kb resolution for WT (left) and MeCP2 KO (right). Scale bar adjusted to account for total coverage on chr3 in each condition.

(B) Decay of relative contact probability as function of genomic distance. Lines represent mean values from three replicates; ribbons are SEM.

(C) Average contact enrichment between pairs of 100 kb loci arranged by eigenvalue (shown on top) for WT (left) and MeCP2 KO (right).

(legend continued on next page)

MeCP2-mutant mice. Our observation that there are frequent contacts between the TSS and gene body provides a possible explanation for how MeCP2 beyond the TSS could repress initiation. One possibility is that MeCP2 bound in the gene body represses transcriptional initiation through the interaction of intragenic enhancers with promoters, as suggested by Clemens et al. (2019). However, although our Hi-C data suggest that the strongest TSS-gene body interactions occur at sites with enriched H3K27ac (i.e., regions that are likely intragenic enhancers) (Figure S7F), such contacts occur throughout the gene body, not just with enhancers. It is therefore possible that longer genes tend to be upregulated in MeCP2-mutant mice because longer genes have more TSS-gene body contacts (Figure S7D). These interactions are independent of MeCP2 but likely function to position gene body MeCP2 and NCoR close to the promoter where they would then attenuate transcriptional initiation. Finally, it should also be acknowledged that it is possible that large domains of MeCP2 binding, which Clemens et al. (2019) observed are delineated by TAD boundaries, may recruit NCoR/Hdac3 to deacetylate and compact the surrounding chromatin, and as a result indirectly reduce initiation of genes within these domains (Figure S6D). However, we favor a model in which the full array of gene body-TSS contacts, including intragenic enhancers, deliver MeCP2 to the TSS to repress transcriptional initiation (Figure 7H). Additional studies are needed to resolve whether MeCP2 represses transcriptional initiation by binding directly to the TSS, gene body, enhancers, Mb domains, or a combination of these genomic elements.

It is important to note that, as in most other studies to date, MeCP2-regulated genes were identified in the brains of mature MeCP2-mutant compared with WT mice. It is likely that some of the observed changes in gene expression are caused by indirect transcriptional or developmental phenotypes in MeCP2 mutants rather than direct effects of MeCP2 loss on transcription. We focused our study on the genes that are upregulated in MeCP2-mutant mice because these genes have the highest levels of mCA and MeCP2 binding and because MeCP2 represses reporter gene transcription through its interaction with the NCoR co-repressor (Lyst et al., 2013; Nan et al., 1997). However, the genes that are downregulated by the disruption of MeCP2 function still have higher mCA levels, MeCP2 binding, and gene body-TSS contacts than genes unaffected by the disruption of MeCP2. Indeed, MeCP2 has been suggested in one report to function as a transcriptional activator (Chahrour et al., 2008). Future studies that investigate the kinetics of gene misregulation following acute loss of MeCP2 will be critical to distinguish the direct targets of MeCP2 from those that are misregulated as a secondary consequence of the loss of MeCP2 function.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
- METHOD DETAILS
  - Cellular fractionation and RNA sequencing
  - Nanostring nCounter gene expression assay
  - Quantitative PCR validation of RNA-seq
  - Chromatin immunoprecipitation sequencing
  - MeCP2 Immunoprecipitation
  - Hdac assay
  - Western blotting
  - Precision run-on sequencing (PRO-seq)
  - Flavopiridol PRO-seq
  - Oxidative bisulfite sequencing (OxBS-seq)
  - Hi-C
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis
  - RNA sequencing analysis
  - Nanostring nCounter gene expression analysis
  - Chromatin immunoprecipitation sequencing analysis
  - PRO-seq analysis
  - OxBS-seq analysis
  - Hi-C analysis
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2019.10.032.

ACKNOWLEDGMENTS

We would like to thank the Rett Syndrome Research Trust for support of this work, along with NIH F32NS101739 and K99NS112415 (L.D.B.), K08NS101064 (W.R.), and R01NS048276 (M.E.G.). We would like to thank

(D) Average TAD representation in WT (top) and MeCP2 KO (bottom).
(E) Aggregate Hi-C maps in WT forebrains centered on TSS of genes upregulated in MeCP2 KO versus WT (top left), downregulated in MeCP2 KO versus WT (top right), or unaffected and expression-matched to those up in MeCP2 KO versus WT (up random, bottom left) or down in MeCP2 KO versus WT (down random, bottom right).
(F) Example TSS-gene body contact enrichment for a gene upregulated in MeCP2 KO versus WT.
(G) Scaled aggregate Hi-C maps of gene bodies for genes 50 kb to 2 Mb in length, expanded up- and downstream by gene length, oriented 5′ to 3′, and divided into 100 bins in WT (left) and MeCP2 KO (right). Values in upper left corners indicate the average and SD of TSS-gene body log contact enrichment (observed/expected) respectively, quantified between a 10 kb window at the TSS and the second half of the gene body (black dashed rectangle). TSS-gene body contacts are enriched in genes upregulated in MeCP2 KO versus WT compared with unaffected expression-matched genes in both WT and MeCP2 KO (WT, $p = 0.019$; KO, $p = 0.006$; permutation test), and in genes downregulated in MeCP2 KO versus WT compared with unaffected expression-matched genes in both WT and MeCP2 KO (WT, $p = 0.013$; KO, $p = 0.028$; permutation test).
(H) Model for MeCP2-dependent repression of highly methylated long genes.
See also Figure S7 and Table S6.

Charles Danko and Ed Rice for PRO-seq protocols; Hojoong Kwak for code for analyzing elongation rates; Matt Lyst for MeCP2 immunoprecipitation protocols; and Adrian Bird and Gail Mandel as well as members of their laboratories, Drs. Karen Adelman, Telmo Henriques, Erin Duffy, David Harmin, Liming Liang, and Harrison Gabel, for helpful discussions.

AUTHOR CONTRIBUTIONS

L.D.B. and W.R. designed, performed, and analyzed most experiments in the study. A.W.G., T.W., A.S., H.S., E.L., M.G.Y., and B.K. performed additional experiments. B.B. analyzed Hi-C data. L.D.B, W.R., E.C.G., and M.E.G. wrote the manuscript. M.E.G. advised on all aspects of the study.

DECLARATION OF INTERESTS

M.E.G is on the Board of Directors and holds equity in Allergan, plc.

Received: February 26, 2019
Revised: September 9, 2019
Accepted: October 23, 2019
Published: November 26, 2019

REFERENCES

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat. Genet. *23*, 185–188.

Bhaskara, S., Knutson, S.K., Jiang, G., Chandrasekharan, M.B., Wilson, A.J., Zheng, S., Yenamandra, A., Locke, K., Yuan, J.L., Bonine-Summers, A.R., et al. (2010). Hdac3 is essential for the maintenance of chromatin structure and genome stability. Cancer Cell *18*, 436–447.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A., and Cavalli, G. (2017). Multiscale 3D genome rewiring during mouse neural development. Cell *171*, 557–572.e24.

Boswell, S.A., Snavely, A., Landry, H.M., Churchman, L.S., Gray, J.M., and Springer, M. (2017). Total RNA-seq to identify pharmacological effects on specific stages of mRNA synthesis. Nat. Chem. Biol. *13*, 501–507.

Brown, K., Selfridge, J., Lagger, S., Connelly, J., De Sousa, D., Kerr, A., Webb, S., Guy, J., Merusi, C., Koerner, M.V., and Bird, A. (2016). The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. Hum. Mol. Genet. *25*, 558–570.

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T.C., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. Science *320*, 1224–1229.

Chen, W.G., Chang, Q., Lin, Y., Meissner, A., West, A.E., Griffith, E.C., Jaenisch, R., and Greenberg, M.E. (2003). Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science *302*, 885–889.

Chen, L., Chen, K., Lavery, L.A., Baker, S.A., Shaw, C.A., Li, W., and Zoghbi, H.Y. (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. U S A *112*, 5509–5514.

Cholewa-Waclaw, J., Shah, R., Webb, S., Chhatbar, K., Ramsahoye, B., Pusch, O., Yu, M., Greulich, P., Waclaw, B., and Bird, A.P. (2019). Quantitative modelling predicts the impact of DNA methylation on RNA polymerase II traffic. Proc. Natl. Acad. Sci. U S A *116*, 14995–15000.

Chu, T., Rice, E.J., Booth, G.T., Salamanca, H.H., Wang, Z., Core, L.J., Longo, S.L., Corona, R.J., Chin, L.S., Lis, J.T., et al. (2018). Chromatin run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. Nat. Genet. *50*, 1553–1564.

Clemens, A.W., Wu, D.Y., Zhao, G., and Gabel, H.W. (2019). MeCP2 represses enhancers through chromosome topology-associated DNA methylation. Mol. Cell. *77*. Published: November 26, 2019. https://doi.org/10.1016/j.molcel.2019.10.033.

Cohen, S., Gabel, H.W., Hemberg, M., Hutchinson, A.N., Sadacca, L.A., Ebert, D.H., Harmin, D.A., Greenberg, R.S., Verdine, V.K., Zhou, Z., et al. (2011). Genome-wide activity-dependent MeCP2 phosphorylation regulates nervous system development and function. Neuron *72*, 72–85.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Duffy, E.E., Canzio, D., Maniatis, T., and Simon, M.D. (2018). Solid phase chemistry to covalently and reversibly capture thiolated RNA. Nucleic Acids Res. *46*, 6996–7005.

Gabel, H.W., Kinde, B., Stroud, H., Gilbert, C.S., Harmin, D.A., Kastan, N.R., Hemberg, M., Ebert, D.H., and Greenberg, M.E. (2015). Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature *522*, 89–93.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., et al. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat. Biotechnol. *26*, 317–325.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Gu, Z., Eils, R., Schlesner, M., and Ishaque, N. (2018). EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. BMC Genomics *19*, 234–237.

Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat. Genet. *27*, 322–326.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Herzel, L., and Neugebauer, K.M. (2015). Quantification of co-transcriptional splicing from RNA-Seq data. Methods *85*, 36–43.

Johnson, B.S., Zhao, Y.T., Fasolino, M., Lamonica, J.M., Kim, Y.J., Georgakilas, G., Wood, K.H., Bu, D., Cui, Y., Goffin, D., et al. (2017). Biotin tagging of MeCP2 in mice reveals contextual insights into the Rett syndrome transcriptome. Nat. Med. *23*, 1203–1214.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife *3*, e02407.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900–903.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H.A., Marr, M.T., 2nd, and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. *25*, 2502–2512.

Kinde, B., Wu, D.Y., Greenberg, M.E., and Gabel, H.W. (2016). DNA methylation in the gene body influences MeCP2-mediated gene repression. Proc. Natl. Acad. Sci. U S A *113*, 15114–15119.

Koerner, M.V., FitzPatrick, L., Selfridge, J., Guy, J., De Sousa, D., Tillotson, R., Kerr, A., Sun, Z., Lazar, M.A., Lyst, M.J., and Bird, A. (2018). Toxicity of overexpressed MeCP2 is independent of HDAC3 activity. Genes Dev. *32*, 1514–1524.

Kruusvee, V., Lyst, M.J., Taylor, C., Tarnauskaitė, Ž., Bird, A.P., and Cook, A.G. (2017). Structure of the MeCP2-TBLR1 complex reveals a molecular basis for Rett syndrome and related disorders. Proc. Natl. Acad. Sci. U S A *114*, E3243–E3250.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.

Lagger, S., Connelly, J.C., Schweikert, G., Webb, S., Selfridge, J., Ramsahoye, B.H., Yu, M., He, C., Sanguinetti, G., Sowers, L.C., et al. (2017). MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide

sequences to tune transcription in the mammalian brain. PLoS Genet. *13*, e1006793.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Larkin, J.D., Cook, P.R., and Papantonis, A. (2012). Dynamic Reconfiguration of Long Human Genes during One Transcription Cycle. Mol Cell Biol. *32*, 2738–2747.

Lee, K., and Bobel, G.A. (2016). Chromatin Architecture Underpinning Transcription Elongation. Nucleus *7*, 1–8.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*, 905–914.

Li, J., Wang, J., Wang, J., Nawaz, Z., Liu, J.M., Qin, J., and Wong, J. (2000). Both corepressor proteins SMRT and N-CoR exist in large protein complexes containing HDAC3. EMBO J. *19*, 4342–4350.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. Science *341*, 1237905.

Lyst, M.J., Ekiert, R., Ebert, D.H., Merusi, C., Nowak, J., Selfridge, J., Guy, J., Kastan, N.R., Robinson, N.D., de Lima Alves, F., et al. (2013). Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. Nat. Neurosci. *16*, 898–902.

Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., and Lis, J.T. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat. Protoc. *11*, 1455–1476.

Malik, A.N., Vierbuchen, T., Hemberg, M., Rubin, A.A., Ling, E., Couch, C.H., Stroud, H., Spiegel, I., Farh, K.K.H., Harmin, D.A., and Greenberg, M.E. (2014). Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. Nat. Neurosci. *17*, 1330–1339.

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell *161*, 541–554.

Nan, X., Campoy, F.J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471–481.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386–389.

Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., et al. (2012). BEDOPS: high-performance genomic feature operations. Bioinformatics *28*, 1919–1920.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72–77.

Nott, A., Cheng, J., Gao, F., Lin, Y.T., Gjoneska, E., Ko, T., Minhas, P., Zamudio, A.V., Meng, J., Zhang, F., et al. (2016). Histone deacetylase 3 associates with MeCP2 to regulate FOXO and social behavior. Nat. Neurosci. *19*, 1497–1505.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517–527.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Raman, A.T., Pohodich, A.E., Wan, Y.W., Yalamanchili, H.K., Lowry, W.E., Zoghbi, H.Y., and Liu, Z. (2018). Apparent bias toward long gene misregulation in MeCP2 syndromes disappears after controlling for baseline variations. Nat. Commun. *9*, 3225.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44* (W1), W160–W165.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Renthal, W., Boxer, L.D., Hrvatin, S., Li, E., Silberfeld, A., Nagy, M.A., Griffith, E.C., Vierbuchen, T., and Greenberg, M.E. (2018). Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing. Nat. Neurosci. *21*, 1670–1679.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Andrews, R., and Bird, A.P. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol. Cell *37*, 457–468.

Stroud, H., Su, S.C., Hrvatin, S., Greben, A.W., Renthal, W., Boxer, L.D., Nagy, M.A., Hochbaum, D.R., Kinde, B., Gabel, H.W., and Greenberg, M.E. (2017). Early-life gene expression in neurons modulates lasting epigenetic states. Cell *171*, 1151–1164.e16.

Sugino, K., Hempel, C.M., Okaty, B.W., Arnson, H.A., Kato, S., Dani, V.S., and Nelson, S.B. (2014). Cell-type-specific repression by methyl-CpG-binding protein 2 is biased toward long genes. J. Neurosci. *34*, 12877–12883.

Tillotson, R., Selfridge, J., Koerner, M.V., Gadalla, K.K.E., Guy, J., De Sousa, D., Hector, R.D., Cobb, S.R., and Bird, A. (2017). Radically truncated MeCP2 rescues Rett syndrome-like neurological defects. Nature *550*, 398–401.

Tronche, F., Kellendonk, C., Kretz, O., Gass, P., Anlag, K., Orban, P.C., Bock, R., Klein, R., and Schütz, G. (1999). Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety. Nat. Genet. *23*, 99–103.

Tudor, M., Akbarian, S., Chen, R.Z., and Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. Proc. Natl. Acad. Sci. U S A *99*, 15536–15541.

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., and Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res. *24*, 896–905.

Vierbuchen, T., Ling, E., Cowley, C.J., Couch, C.H., Wang, X., Harmin, D.A., Roberts, C.W.M., and Greenberg, M.E. (2017). AP-1 transcription factors and the BAF complex mediate signal-dependent enhancer selection. Mol. Cell *68*, 1067–1082.e12.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185.

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics *10*, 232.

Young, J.I., Hong, E.P., Castle, J.C., Crespo-Barreto, J., Bowman, A.B., Rose, M.F., Kang, D., Richman, R., Johnson, J.M., Berget, S., and Zoghbi, H.Y. (2005). Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. Proc. Natl. Acad. Sci. U S A *102*, 17551–17558.

Yusufzai, T.M., and Wolffe, A.P. (2000). Functional consequences of Rett syndrome mutations on human MeCP2. Nucleic Acids Res. *28*, 4172–4179.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-Histone H3K27ac	Abcam	Cat# ab4729; RRID:AB_2118291
Rabbit anti-Histone H3K36me3	Abcam	Cat# ab9050; RRID:AB_306966
Rabbit anti-Histone H3K79me2	Abcam	Cat# ab3594; RRID:AB_303937
Rabbit anti-Histone H3K9ac	Millipore	Cat# 06-942; RRID:AB_310308
Rabbit anti-Histone H3	Abcam	Cat# ab1791; RRID:AB_302613
Rabbit anti-Histone H4K12ac	Millipore	Cat# 07-595; RRID:AB_310740
Rabbit anti-Histone H4	Abcam	Cat# ab10158; RRID:AB_296888
Mouse anti-RNA Pol II	Abcam	Cat# ab817; RRID:AB_306327
Rat anti-RNA Pol II Phospho-Ser5	Millipore	Cat# 04-1572; RRID:AB_11213421
Rabbit anti-MeCP2	(Chen et al., 2003)	N/A
Mouse anti-MeCP2	Sigma	Cat# M6818; RRID:AB_262075
Mouse IgG control	Sigma	Cat# M5284; RRID:AB_1163685
Rabbit anti-Gapdh	Sigma	Cat# G9545; RRID:AB_796208
Rabbit anti-Snrp70	Abcam	Cat# ab83306; RRID:AB_10673827
Rabbit anti-Hdac3	Sigma	Cat# H3034; RRID:AB_260037
Rabbit anti-Tblr1	Bethyl	Cat# A300-408A; RRID:AB_420967
Rabbit anti-mSin3a	Santa Cruz	Cat# sc-994; RRID:AB_2187760
Rabbit anti-YY1	Abcam	Cat# ab109237; RRID:AB_10890662
Chemicals, Peptides, and Recombinant Proteins		
Flavopiridol hydrochloride	Tocris Bioscience	Cat# 309410; CAS: 131740-09-5
Critical Commercial Assays		
RNeasy Micro Kit	QIAGEN	Cat# 74004
NEBNext Ultra Directional Library Prep Kit for Illumina	NEB	Cat# E7420L
NEBNext rRNA Depletion Kit	NEB	Cat# E6310X
NEBNext Multiplex Oligos for Illumina	NEB	Cat# E7335L
NuGEN Ovation Ultralow System V2	NuGEN	Cat# 0344-32
Fluorescent HDAC Activity Assay	Active Motif	Cat# 56200
Deposited Data		
Raw and analyzed RNA-seq, NanoString, ChIP-seq, PRO-seq, OxBS-seq, and Hi-C data	This study	GEO: GSE128186
Bisulfite-sequencing data	(Lister et al., 2013)	GEO: GSE47966
Hi-C data from ES cells and cortical neurons	(Bonev et al., 2017)	GEO: GSE96107
Experimental Models: Organisms/Strains		
Mouse: *Mecp2* KO	The Jackson Laboratory	JAX: 003890
Mouse: *Mecp2* R306C	(Lyst et al., 2013)	N/A
Mouse: Nestin-cre	The Jackson Laboratory	JAX: 003771
Mouse: *Dnmt3a*$^{fl/fl}$	(Kaneda et al., 2004)	N/A
Oligonucleotides		
qPCR primers for gene expression (Table S2)		N/A
qPCR primers for ChIP (Table S2)		N/A
Oligonucleotides for PRO-seq	(Mahat et al., 2016)	N/A
Software and Algorithms		
Trimmomatic (v0.36)	(Bolger et al., 2014)	http://www.usadellab.org/cms/?page=trimmomatic
STAR (v2.5.2b)	(Dobin et al., 2013)	https://github.com/alexdobin/STAR

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RSeqQC (v2.6.4)	(Wang et al., 2012)	http://rseqc.sourceforge.net/
Subread featureCounts (v1.6.0)	(Liao et al., 2014)	http://subread.sourceforge.net/
edgeR (v3.18.1)	(Robinson et al., 2010)	https://bioconductor.org/packages/release/bioc/html/edgeR.html
deepTools (v3.0.2)	(Ramírez et al., 2016)	https://github.com/deeptools/deepTools
BEDTools (v2.26)	(Quinlan and Hall, 2010)	https://github.com/arq5x/bedtools2
SAMtools (v0.1.19)	(Li et al., 2009)	http://samtools.sourceforge.net/
nSolver (v4.0)	NanoString	https://www.nanostring.com/products/analysis-software/nsolver
Bowtie2 (v2.2.9)	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
UCSC tools (v363)	UCSC-tools	http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/
BBTools (v37.90)	BBTools	https://sourceforge.net/projects/bbmap/
BEDOPS (v2.4.30)	(Neph et al., 2012)	https://github.com/bedops/bedops
Homer (v4.9)	(Heinz et al., 2010)	http://homer.ucsd.edu/homer/
MACS2 (v2.1.1.20160309)	(Zhang et al., 2008)	https://github.com/taoliu/MACS
FIMO (meme v5.0.3)	(Grant et al., 2011)	http://meme-suite.org/doc/fimo.html
EnrichedHeatmap (v1.14)	(Gu et al., 2018)	https://bioconductor.org/packages/release/bioc/html/EnrichedHeatmap.html
BSmap (v2.74)	(Xi and Li, 2009)	https://code.google.com/archive/p/bsmap/
Shaman (v2.0)	Shaman	https://bitbucket.org/tanaylab/shaman

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by Michael Greenberg (michael_greenberg@hms.harvard.edu). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

All animal experiments were approved by the National Institutes of Health and the Harvard Medical School Institutional Animal Care and Use Committee and were conducted in compliance with the relevant ethical regulations. 8-week-old male *Mecp2* knockout (KO) mice ($Mecp2^{-/y}$) and their wild-type controls ($Mecp2^{+/y}$) were obtained from Jackson Labs (Stock No. 003890). This line was originally generated by Adrian Bird (Guy et al., 2001). *Mecp2* R306C knock-in mice were generated previously (Lyst et al., 2013). Female heterozygous $Mecp2^{R306C/+}$ mice were crossed with male wild-type C57BL/6J mice to generate hemizygous male $Mecp2^{R306C/y}$ and $Mecp2^{+/y}$ littermate controls that were analyzed at either 1 week or 8-12 weeks of age. Dnmt3a conditional knock-out (cKO) mice were generated by crossing $Dnmt3a^{fl/fl}$ mice (Kaneda et al., 2004) with *Nestin-cre* mice (Tronche et al., 1999). Littermate control comparisons were between wild-type ($Dnmt3a^{fl/fl}$) and Dnmt3a cKO (*Nestin-cre*; $Dnmt3a^{fl/fl}$) mice, with an equal number of male and female mice in each group. Mice were housed under a standard 12-hour light cycle. *Mecp2* KO and R306C mice all demonstrated decreased locomotor activity at time of analysis (8-12 weeks old).

METHOD DETAILS

Cellular fractionation and RNA sequencing

Forebrain (cortex and hippocampus) was dissected from 10 8-week-old male MeCP2 knock-out and 10 WT littermate controls, 10 8-to-12-week-old male R306C and 10 WT littermate controls, and 10 1-week-old male R306C and 10 WT littermate controls and flash frozen. Tissue was removed from dry ice and placed directly into a dounce with 5 mL of Homogenization buffer (HB: 250 mM sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM Tricine KOH pH 7.8, 0.04% BSA, 5 μg/mL Actinomycin D, and RNAsin (Promega)). After 10 strokes with the tight pestle, a 5% Igepal (Sigma) solution in HB was added to a final concentration of 0.32%, followed by five additional strokes with the tight pestle. At this point, 200 μL of lysate was pipetted into a tube with 600 μL Trizol LS (Invitrogen) as the "whole cell" sample. The remaining lysate was filtered first through a 40 μm cell strainer and then through a 20 μm Celltrics filter. Lysate was diluted 1:2 in HB to dilute Igepal to 0.16%, then centrifuged at 500 g for 5 minutes at 4°C. The

nuclear pellet was washed twice with 1 mL HB by resuspending and centrifuging at 500 g for 5 minutes at 4°C. Nuclei were resuspended in 1mL HB and 200 μL was pipetted into a separate tube, spun at 500 g for 5 minutes at 4°C, and the pellet was resuspended in Trizol as the "nuclear fraction." The remaining 800 μL was spun at 500 g for 5 minutes at 4°C, then resuspended in 200 μL Glycerol buffer (20 mM Tris pH 8, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol, 0.85 mM DTT). 200 μL of Nuclear lysis buffer (1% Igepal, 20 mM HEPES pH 7.5, 300 mM NaCl, 0.2 mM EDTA) was added and nuclei were resuspended, vortexed, and incubated at 4°C for 5 minutes. Chromatin was spun at 16,000 g for 10 minutes at 4°C, and the pellet was resuspended in 800 μL Trizol and homogenized to solubilize. For Dnmt3a cKO mice, visual cortex was dissected and homogenized in Trizol. For all samples, total RNA was chloroform extracted and purified with the QIAGEN RNeasy Micro Kit with on-column DNase treatment. RNA-seq libraries were generated with the NEBNext Ultra Directional Library Prep Kit with rRNA depletion. Libraries were sequenced on an Illumina Nextseq 500 with 85 bp single-end reads. All libraries to be compared as a set for differential expression analysis (e.g., 10 KO and 10 WT whole-cell) were prepared and sequenced at the same time to reduce batch effects.

Nanostring nCounter gene expression assay

Nanostring nCounter probes were designed to 200 genes based on fold-changes and significance from the MeCP2 KO versus WT whole-cell RNA-seq (96 significantly upregulated genes, 96 significantly downregulated genes, 6 housekeeping genes, and positive controls *Mecp2* and *Irak1*). Nanostring gene expression analysis was performed at the Boston Children's Hospital Molecular Genetics Core Facility on the same 10 MeCP2 KO and 10 WT and 10 R306C and 10 WT whole-cell RNA used for RNA-seq. RNA (100 ng) was hybridized to capture probes and reporter probes for 16 hours at 65°C. Hybridized RNA was scanned with the nCounter Sprint Profiler following the manufacturer's instructions.

Quantitative PCR validation of RNA-seq

cDNA was generated from 500 ng of whole-cell MeCP2 KO and WT RNA with the Superscript III First-strand synthesis system (Life Technologies) with random hexamer priming. Quantitative PCR was performed with gene-specific primers (Table S2) with SYBR green detection on a Roche Lightcycler 480. Relative transcript levels and fold-changes were calculated by normalizing to two housekeeping genes with unaltered expression by RNA-seq or Nanostring (*Aars* and *Tada2b*).

Chromatin immunoprecipitation sequencing

Forebrain (cortex and hippocampus) was dissected from 8-to-12-week-old male mice and flash frozen. Tissue was homogenized in 1% formaldehyde and cross-linked for 10 minutes at room temperature. Cross-linking was quenched with 125 mM glycine for 5 minutes at room temperature. Tissue was pelleted by spinning at 500 g for 5 minutes at 4°C, washed once with PBS, then spun again at 500 g for 5 minutes at 4°C. Cells were lysed by resuspending in L1 buffer (50 mM HEPES pH 7.5, 140mM NaCl, 1 mM EDTA pH 8, 1 mM EGTA pH 8, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, protease inhibitors) and rotating at 4°C for 10 minutes, then spun at 500 g for 5 minutes at 4°C. Nuclei were washed for 10 minutes at 4°C in L2 buffer (10 mM Tris pH 8, 200 mM NaCl, protease inhibitors), then spun at 500 g for 5 minutes at 4°C. Nuclei were resuspended in LB3 buffer (10 mM Tris pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-Lauroylsarcosine, protease inhibitors), and sonicated in a Diagenode Bioruptor (High Power, 60 cycles, 30 s on/45 s off). Insoluble material was removed by spinning at 16,000 g for 10 minutes at 4°C, and Triton X-100 was added to soluble chromatin at a final concentration of 1%. Chromatin was pre-cleared for two hours with Protein A (for rabbit antibodies) or Protein G (for mouse and rat antibodies) Dynabeads, then incubated with Protein A or Protein G Dynabeads conjugated to antibodies overnight at 4°C. Antibodies used were: H3K27ac (Abcam ab4729), H3K36me3 (Abcam ab9050), H4K12ac (Millipore 07-595), H3K9ac (Millipore 06-942), H3K79me2 (Abcam ab3594), Pol II (Abcam ab817), Pol II Ser5p (Millipore 04-1572), MeCP2 antibody 1 (Chen et al., 2003), and MeCP2 antibody 2 (Sigma M6818). Beads were washed twice with Low Salt Buffer (20 mM Tris pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), twice with High Salt Buffer (20 mM Tris pH 8, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), twice with LiCl Wash Buffer (10 mM Tris pH 8, 1 mM EDTA, 1% NP-40, 250 mM LiCl, 1% sodium deoxycholate) and once with TE Buffer (50 mM Tris pH 8, 10 mM EDTA) at 4°C. Chromatin was eluted off beads by incubating in TE Buffer with 1% SDS at 65°C for one hour, and crosslinks were reversed by incubating overnight at 65°C. Chromatin was treated with RNase A for 30 minutes at 37°C and Proteinase K for 2 hours at 55°C. DNA was phenol-chloroform extracted and purified with the QIAGEN PCR purification kit. Quantitative PCR was performed with gene-specific primers (Table S2) with SYBR green detection on a QuantStudio 3 (Applied Biosystems). ChIP-seq libraries were generated using the NuGEN Ovation Ultralow System V2 following manufacturer instructions. Libraries were sequenced on an Illumina Nextseq 500 with 85 bp single-end reads.

MeCP2 Immunoprecipitation

Forebrain tissue from 6 male R306C mice and 6 wild-type littermates (ages 8-to-15-weeks-old) was dissected and flash frozen. Tissue was removed from dry ice and added to a dounce with 5 mL buffer A (10 mM HEPES pH 7.5, 25 mM KCl, 0.15 M Spermine, 0.5 M Spermidine, 1 mM EDTA, 2 M Sucrose, 10% glycerol, and 1X protease inhibitor cocktail (Sigma Aldrich 5056489001)). Each brain was stroked 20 times using a tight pestle, and the resulting homogenate was combined with another homogenate from a second brain of the same genotype, yielding a 10 mL homogenate of two brains. Homogenates were rotated at 4°C for 10 minutes and layered on top of 2 mL Buffer A and spun at 24,000 RPM for 40 minutes at 4°C in an SW-41 rotor. The nuclear pellet was resuspended in 1 mL NE1 Buffer (20 mM HEPES pH 7.5, 10 mM NaCl, 1 mM $MgCl_2$, 0.1% Triton X-100, 1 mM DTT, and 1X protease

inhibitor cocktail), and spun again at 800 g for 5 minutes at 4°C. The nuclear pellet was resuspended in 100 μL NE1, split into two tubes and 1 μL Benzonase was added to each tube for 5 minutes at room temperature. The tubes were returned to ice and the volume was raised to 200 μL with NE1. NaCl was added to a final concentration of 150 mM. Samples were rotated at 4°C for 20 minutes, then spun at 16,000 g for 20 minutes at 4°C to remove insoluble material. 150 μL of NE1-150 (20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$, 0.1% Triton X-100, 1 mM DTT, and 1X protease inhibitor cocktail) was added to each supernatant. Two 17.5 μL-aliquots of each supernatant were set aside for downstream use in the HDAC kit and for western blots ("input"). NE1-150 (12.5 μL) was added to each input aliquot for total volumes of 30 μL; these were stored at 4°C until the HDAC assay, or until LDS-BME was added, respectively. For each tube of nuclear supernatant, either 5 μg of MeCP2 antibody (Sigma M6818) or 5 μg of IgG control (Sigma M5284) was added, and samples were rotated for 1 hour at 4°C. Protein G Sepharose Beads (Sigma P3296) were washed 4 times with 1 mL NE1-150. 10 μL of beads were added to each IgG and MeCP2 IP and rotated for another hour at 4°C. The beads were washed four times with NE1-150 and resuspended in NE1-150 to a total volume of 62 μL. A 42 μL aliquot was taken from each IP tube and set aside for use during the HDAC assay. The remaining 20 μL of each IP fraction and the input aliquots previously set aside for western blots were eluted in 20 μL of LDS-BME mix (LDS-BME mix: 1X LDS, 5% 2-mercaptoethanol, using PBS as the solvent). The eluates were boiled at 95°C for 15 minutes, and any sample with beads was spun for 2 minutes at 8000 g at room temperature. The 20 μL supernatant was placed into a new tube for running western blots.

Hdac assay

Fluorescent HDAC Activity assay (Active Motif 56200) was performed immediately following IP elution as described in the assay manual. Input samples consisted of the 30 μL input aliquots previously set aside for the HDAC assay mixed with 95 μL HDAC Assay Buffer. 30 μL of this mix was used per well of input condition. IP samples were plated as 20 μL of each IP aliquot added directly to every well of IP condition followed by 10 μL of HDAC Assay Buffer. All input and IP samples were plated in wells with or without the presence of Trichostatin A, 1 μM final concentration. The assay plate was incubated for 60 minutes at 37°C and then incubated at room temperature for 15 minutes. Fluorescence was measured using a Biotek Synergy 4 microplate reader and Biotek Gen5 software (version 1.1) at 360 nm excitation and 460 nm emission (Tungsten bulb, sensitivity = 100, top probe vertical offset = 4.0 mm, column offset = 0 mm). pmol of deacetylated lysine product formed was interpolated from a linear standard curve of known amounts of deacetylated lysine product (HDAC Assay Standard) plotted against fluorescence intensity.

Western blotting

Tissue was homogenized in RIPA buffer (50 mM Tris pH 8, 150 mM NaCl, 1% NP-40, 0.5% Sodium deoxycholate, 0.1% SDS, protease inhibitors) and incubated on ice for 15 minutes. Lysates were sonicated in a Diagenode Bioruptor (High Power, 10 cycles, 30 s on/30 s off), then spun at 16,000 g for 10 minutes at 4°C to remove insoluble material. LDS and β-Me (5% final concentration) were added to lysates and samples were boiled at 95°C for 15 minutes. Samples from tissue lysate, immunoprecipitation, or cellular fractionation were run on 4%–12% Bis-Tris gels (Life Technologies) and transferred to nitrocellulose membranes. Membranes were blocked for 1 hour at room temperature, then incubated with primary antibodies overnight at 4°C, washed 3 times with TBS-T, incubated with secondary antibodies for 45 minutes at room temperature, and then washed 3 times with TBS-T. Blocking and antibody incubations were performed in 4% BSA in TBS-T. Western blots were imaged on a Li-Cor Odyssey. Primary antibodies used were: Gapdh (Sigma G9545, 1:5000), Snrp70 (Abcam ab83306, 1:1000 from 0.9 μg/μL variable stock), Histone H3 (Abcam ab1791, 1:5000), MeCP2 (Chen et al., 2003), β-actin (Abcam ab8226, 1:2000), Hdac3 (Sigma H3034, 1:2500), Tblr1 (Bethyl A300-408A, 1:2000), mSin3a (Santa Cruz sc-994, 1:500), H3K9ac (Millipore 06-942 1:2000), H3K27ac (Abcam ab4729, 1:10,000), H4K12ac (Millipore 07-595 1:2000), and Histone H4 (Abcam ab10158 1:2000). Western blots were quantified with ImageJ.

Precision run-on sequencing (PRO-seq)

Chromatin Isolation for PRO-seq

Forebrain (cortex and hippocampus) was dissected from 8-to-12-week-old MeCP2 KO, MeCP2 R306C, Dnmt3a cKO mice and their respective WT littermate controls. Tissue was homogenized in ice-cold PBS using a Dounce homogenizer, and nuclei were pelleted by spinning at 17,000 g for 5 minutes at 4°C. Chromatin isolation was performed as previously described (Chu et al., 2018). Nuclei were resuspended in lysis buffer (NUN: 0.3 M NaCl, 1 M Urea, 20 mM HEPES pH 7.5, 7.5 mM $MgCl_2$, 0.2 mM EDTA, 1% NP-40, 1 mM DTT, cOmplete Protease Inhibitor (Roche), 40 units/mL SUPERase In RNase Inhibitor (Invitrogen)) and incubated on ice for 30 minutes. Lysed nuclei were centrifuged at 12,500 g for 30 minutes at 4°C. The resulting insoluble chromatin pellet was washed in 50 mM Tris-HCl pH 7.5 and placed in 50 μL chromatin storage buffer (25% glycerol, 50 mM Tris-HCl pH 8.0, 5 mM $MgAc_2$, 0.1 mM EDTA, 5 mM DTT, 40 units/mL SUPERase In RNase Inhibitor). Chromatin was solubilized by sonicating on a Diagenode Bioruptor (High Power, 40 cycles, 30 s on/30 s off). Solubilized chromatin was flash-frozen in liquid nitrogen and stored at −80°C.

RNA purification for PRO-seq

PRO-seq RNA purification and library preparation was performed as previously described (Kwak et al., 2013; Mahat et al., 2016). Frozen chromatin was thawed on ice and combined with 50 μL 2X Run-on Buffer (10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 300 mM KCl, 40 μM Biotin-11-ATP, 40 μM Biotin-11-GTP, 40 μM Biotin-11-CTP, 40 μM Biotin-11-UTP, 1% Sarksosyl, 1 mM DTT, 0.8 units/μL SUPERase In RNase Inhibitor). The reaction was incubated at 37°C for 5 minutes and then quenched with 500 μL Trizol LS, combined with 130 μL chloroform, and centrifuged at 17,000 g for 5 minutes at 4°C. The aqueous layer containing RNA was

combined with 1 mL ice-cold ethanol and 2.5 μL GlycoBlue (Invitrogen) and centrifuged at 17,000 g for 15 minutes at 4°C. The RNA pellet was washed in 75% ethanol, air-dried, and resuspended in 20 μL nuclease-free water. RNA was hydrolyzed by adding 5 μL 1 N NaOH and incubating for 4 minutes at 4°C, then neutralized with 30 μL 1 M Tris-HCl pH 7.0. Hydrolyzed RNA was run on a Micro Bio-Spin P-30 Column (Bio-Rad) to remove excess salts. Biotinylated RNA was purified with magnetic streptavidin beads (NEB) at room temperature for 20 minutes and washed 2 times each with 500 μL High-Salt (2 M NaCl, 50 mM Tris-HCl pH 7.4, 0.5% Triton X-100, 4 units/mL SUPERase In RNase Inhibitor), Binding (300 mM NaCl, 10 mM Tris-HCl pH 7.4, 0.1% Triton X-100, 4 units/mL SUPERase In RNase Inhibitor), and Low-Salt (5 mM Tris-HCl pH 7.4, 0.1% Triton X-100, 4 units/mL SUPERase In RNase Inhibitor) wash buffers. Beads were resuspended in 300 μL Trizol, and combined with 60 μL chloroform, and RNA was purified as above.

Library Preparation for PRO-seq

The dried RNA pellet was resuspended in 4 μL of 12.5 μM Reverse 3′ RNA adaptor and used to make a T4 RNA Ligase I reaction mix (5 μM Reverse 3′ RNA Adaptor, 1X T4 RNA Ligase I Buffer, 1 mM ATP, 10 units T4 RNA Ligase I, 10% PEG 8000, 2 units/μL SUPERase In RNase Inhibitor). The reaction was incubated overnight at 20°C. The ligated RNA was purified with magnetic streptavidin beads, washed, and extracted with Trizol as above. The RNA was treated with RNA 5′ Pyrophosphohydrolase (RppH) to remove the 5′ cap (1X NEBuffer 2, 10 units RppH, 0.2 units/μL SUPERase In RNase Inhibitor) and incubated for 1 hour at 37°C. A hydroxyl repair reaction mix (1X PNK Buffer, 1 mM ATP, 25 units PNK, 0.2 units/μL SUPERase In RNase Inhibitor) was added directly to the Rpph reaction and incubated for an additional hour at 37°C. RNA was extracted with Trizol as above. The dried RNA pellet was resuspended in 4 μL of 12.5 μM Reverse 5′ RNA adaptor and used to make a T4 RNA Ligase I reaction mix (5 μM Reverse 5′ RNA Adaptor, 1X T4 RNA Ligase I Buffer, 1 mM ATP, 10 units T4 RNA Ligase I, 10% PEG 8000, 2 units/μL SUPERase In RNase Inhibitor). The reaction was incubated overnight at 20°C. The ligated RNA was purified with magnetic streptavidin beads, washed, and extracted with Trizol as above. The resulting RNA library was reverse transcribed using SuperScript III Reverse Transcriptase (Invitrogen). The cDNA libraries were amplified using Q5 DNA Polymerase (NEB) for 14 cycles or more, as determined by a test amplification with diluted cDNA. Amplified libraries were purified on Agencourt RNAClean XP beads. Libraries were sequenced on an Illumina Nextseq 500 with 75 bp single-end reads.

Flavopiridol PRO-seq

Flavopiridol hydrochloride (Tocris Bioscience) was prepared at 4 mg/mL in a 10% DMSO solution in PBS and injected intraperitoneally at a dose of 50 mg/kg into 8-week-old male WT or MeCP2 KO mice. 20, 40, or 60 minutes after flavopiridol injection, forebrain tissue was dissected and flash frozen. For the "0 minute" time point, mice were injected with a control solution of 10% DMSO in PBS and dissected after 40 minutes. One side of each forebrain was used for PRO-seq. Cells were permeabilized as described previously (Mahat et al., 2016), with minor modifications. Tissue was dounce homogenized in 5 mL cold PBS, then spun at 800 g for 4 minutes at 4°C. The pellet was washed with 10 mL PBS, then spun at 800 g for 4 minutes at 4°C. The pellet was resuspended in 1 mL Buffer W (10 mM Tris pH 8, 10 mM KCl, 250 mM sucrose, 5 mM $MgCl_2$, 0.5 mM DTT, 10% glycerol) and then 9 mL Buffer P was added (Buffer W with 0.1% Igepal). Lysate was filtered through a 40 μm cell strainer and spun at 800 g for 4 minutes at 4°C. The pellet was resuspended in 1 mL Buffer W and then 9 mL Buffer P was added. Lysate was incubated on ice for 1 minute, then spun at 800 g for 4 minutes at 4°C. The pellet was resuspended in 500 μL Buffer F (50 mM Tris pH 8, 40% glycerol, 5 mM $MgCl_2$, 0.5 mM DTT) and spun at 1000 g for 4 minutes at 4°C. The pellet was resuspended in 50 μL Buffer F with 0.5 μL RNase inhibitor (SUPERase In, Ambion), flash frozen, and stored at −80°C. PRO-seq run-on and library preparation was performed as described above. Libraries were sequenced on an Illumina Nextseq 500 with 75 bp single-end reads.

Oxidative bisulfite sequencing (OxBS-seq)

OxBS-seq was performed and analyzed as described previously (Stroud et al., 2017), with minor modifications. Genomic DNA was extracted from 8-week-old MeCP2 KO and WT forebrain with the QIAGEN DNeasy kit. DNA was sonicated with a Covaris M220 to 200 bp and ligated to NEXTflex methylated adapters (Bioo Scientific). Bisulfite conversions and oxidation reactions were performed using the TrueMethyl oxBS-seq kit following the manufacturer instructions (Cambridge Epigenetix). Half of the sample was used for BS-seq without oxidation treatment, and half the sample was used for oxBS-seq as recommended by the manufacturer. Libraries were sequenced on an Illumina Nextseq 500 with 75 bp paired-end reads.

Hi-C

Forebrain (cortex and hippocampus) was dissected from 8-week-old MeCP2 KO and WT mice and flash frozen. *In situ* Hi-C was performed as described in Rao et al. (2014) with minor modifications. Tissue was homogenized in 1% formaldehyde in PBS and rotated for 10 minutes at room temperature. Cross-linking was quenched with 125 mM glycine for 5 minutes at room temperature. Tissue was pelleted by spinning at 500 g for 5 minutes at 4°C, washed once with PBS, then spun again at 500 g for 5 minutes at 4°C. Cells were lysed by resuspending in 500 μL Hi-C lysis buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.2% Igepal CA630) with protease inhibitors (Sigma P8340), incubated on ice for 15 minutes, then spun at 2500 g for 5 minutes at 4°C. Pelleted nuclei were washed with Hi-C lysis buffer, then spun at 2500 g for 5 minutes at 4°C. Nuclei were resuspended in 50 μL 0.5% sodium dodecyl sulfate (SDS) and incubated at 62°C for 7.5 minutes. After heating, 115 μL of water and 25 μL of 10% Triton X-100 were added to quench the SDS, and samples were incubated at 37°C for 15 minutes. Restriction digest was performed by adding 25 μL of 10X NEB Buffer 2 and 100 U of MboI (NEB R0147) and rotating 16 hours at 37°C. Samples were incubated at 62°C for 20 minutes to inactivate MboI.

To fill in restriction fragment overhangs and mark DNA ends with biotin, 50 μL of fill-in master mix was added (37.5 μL of 0.4 mM biotin-14-dATP, 1.5 μL of 10mM dCTP, 15 μL of 10mM dGTP, 1.5 μL of 10mM dTTP, and 8 μL 5U/μL DNA Polymerase I, Large (Klenow) Fragment (NEB, M0210)), and samples were rotated for 75 minutes at 37°C. 900 μL of Ligation master mix was added (669 μL of water, 120 μL of 10X NEB T4 DNA ligase buffer (NEB, B0202), 100 μL 10% Triton X-100, 6 μL 20 mg/mL Bovine Serum Albumin, and 5 μL of 400 U/μL T4 DNA Ligase (NEB, M0202)), and samples were rotated at room temperature for 4 hours. Protein was degraded by adding 50 μL 20 mg/mL proteinase K (NEB, P8102) and 120 μL of 10% SDS and incubating at 55°C for 30 minutes on a Thermomixer at 350 rpm. 130 μL of 5M NaCl was added and samples were incubated at 68°C for 16 hours to reverse cross-links. DNA was purified by adding 1.6X volumes of ethanol and 0.1X volumes of 3M sodium acetate, pH 5.2 and incubated at −80°C for 15 minutes. Samples were spun at 17,720 g for 15 minutes at 4°C, washed with 70% ethanol, and spun at 17,720 g for 5 minutes at 4°C. Pellet was re-suspended in 130 uL 1X Tris buffer (10 mM Tris-HCl, pH 8) and incubated at 37°C to fully dissolve DNA. To make biotinylated DNA suitable for high-throughput sequencing, DNA was sheared with a Covaris M220 (Peak incident power: 50, Duty factor: 10%, Cycles per burst: 200, Treatment time: 70 s). DNA fragments in the range of 300-500 bp were selected using AMPure XP beads. 150 μL of 10 mg/mL MyOne Streptavidin T1 beads were washed with 400 μL 1X Tween Washing Buffer (TWB: 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1M NaCl, 0.05% Tween 20). Beads were resuspended in 300 μL 2X Binding Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2M NaCl), added to samples, and rotated for 15 minutes at room temperature. Beads were separated on a magnet and washed twice with 600 μL TWB, incubating at 55°C with mixing at 350 rpm for 2 minutes on a Thermomixer, and transferring to a new tube with each wash. Beads were washed once with 100ul 1X NEB T4 DNA ligase buffer, then resuspended in 100 μL of master mix (88ul 1X NEB T4 DNA ligase buffer, 2 ul 25 mM dNTP mix, 5 ul 10U/ul NEB T4 PNK (NEB, M0201), 4 ul of 3U/ul NEB T4 DNA polymerase I (NEB, M0203), and 1 ul of 5U/ul NEB DNA polymerase I, Large (Klenow) Fragment (NEB, M0210)), and incubated at room temperature for 30 minutes. Beads were washed twice with TWB as described above, washed once with 100 ul 1X NEBuffer 2, and resuspended in 100ul dATP attachment master mix (90ul 1X NEBuffer 2, 5ul 10mM dATP, and 5ul of 5U/ul NEB Klenow exo minus (NEB, M0212)), and incubated at 37°C for 30 minutes. Beads were washed twice with TWB as described above and washed once with 100ul 1X Quick ligation reaction buffer (NEB, B6058). Beads were resuspended in 50ul of 1X NEB Quick ligation reaction buffer, and 2ul NEB DNA Quick ligase (NEB, M2200) and 3 ul of NEBNext adaptor (NEB E7335) diluted 1:10 in 10mM Tris pH 7.5 were added, and incubated at room temperature for 15 minutes. Beads were washed twice with TWB as described above, and washed once with 100ul 1X Tris buffer. Beads were resuspended in 20ul 1X Tris buffer and boiled at 98°C for 10 minutes to elute DNA off beads. Hi-C libraries were amplified by adding to 17 ul DNA: 25 ul NEBNext Q5 Hot Start HiFi PCR Master Mix (M0543), 3 ul NEBNext USER Enzyme (M5505), 2.5 ul 10uM Index (i7) primer, and 2.5ul Universal (i5) primer (E7335), and amplified for 8 cycles. Libraries were purified with AMPure beads (2 purifications with 0.7X beads) and sequenced on an Illumina Nextseq 500 with 75 bp paired-end reads to a read depth of at least 500 million reads per library.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

Pearson correlations between gene expression and gene length or DNA methylation were calculated in R (v3.4.2). Pearson correlations of MeCP2 mutant versus WT and Shuffled were compared by permutation. P values for these comparisons were estimated by calculating: (number of events where $|corr1^{permutation} - corr2^{permutation}| > |corr1^{observed} - corr2^{observed}|$)/1,000 iterations. Wilcoxon Rank Sum test and Kruskal-Wallis test with Dunn's post hoc test were performed in R (v3.4.2). Two-tailed unpaired t tests were performed in GraphPad Prism (v7.0b). Fisher's exact test was used to calculate the significance of overlap of gene lists: https://www.langsrud.com/fisher.htm.

RNA sequencing analysis

Read mapping, quantification, and differential expression

Reads were trimmed with Trimmomatic (v0.36) (Bolger et al., 2014) to remove Illumina adaptors and low quality sequence (settings: LEADING:5 TRAILING:5 SLIDINGWINDOW:4:20 MINLEN:50). Trimmed reads were mapped to the mm10 Refseq transcriptome and genome using STAR (v2.5.2b) (Dobin et al., 2013) with–outFilterMismatchNoverReadLmax 0.04. The percentage of intronic reads was quantified with RSeQC read_distribution.py (v2.6.4) (Wang et al., 2012). For whole-cell RNA-seq, reads were counted in all exons for each gene using Subread featureCounts (v1.6.0) (Liao et al., 2014) with default parameters (reversely stranded, multi-mapping reads not counted, multi-overlapping reads not counted). For nuclear and chromatin-associated RNA-seq, reads were counted in the full gene body (TSS to TTS of longest isoform, in order to include intronic reads) using Subread featureCounts. Similar results were also obtained when nuclear and chromatin-associated RNA-seq reads were counted in introns only. Differential expression was performed with the R package edgeR (v3.18.1) (Robinson et al., 2010). Genes with low counts were filtered by keeping only genes with rowSums(cpm(y) > 1) > = (number of samples per group). Differentially expressed genes were defined by an FDR < 0.05 with no fold-change cut-off, except where fold-change cut-offs are indicated. For the "shuffled" comparisons, 5 WT and 5 MeCP2 mutant samples were randomly selected using the sample function in R and assigned to one group, and the remaining 5 WT and 5 MeCP2 mutant samples were assigned to the other group, and differential expression was performed with edgeR. Random shuffling was performed 10 times for each comparison and one representative example is shown in plots. For downstream analysis of the

genes significantly misregulated in MeCP2 KO versus WT, *Mecp2* and the adjacent gene *Irak1* were excluded, as the expression of these genes is altered as a direct consequence of the mutations in the MeCP2 KO mouse line. For generating sets of expression-matched, unaffected genes, the logCPM from edgeR was converted to RPKM using the formula RPKM = 2^(logCPM-$\log_2$(gene length in kb)), and custom R scripts were used to obtain for each significantly up- or downregulated gene, the gene from the unaffected list with the closest RPKM, not allowing duplicate genes.

Comparison of MeCP2-dependent gene expression to gene length and DNA methylation

Gene length was defined as the difference between TSS and TTS for the longest isoform of each mm10 Refseq gene. For DNA methylation, bisulfite-sequencing from 10-week-old or 1-week-old mouse cortex from Lister et al. (2013) was used to calculate DNA methylation density as the number of cytosines/(number of cytosines + thymines) in either the CA or CG dinucleotide context in each genomic region. To examine the effects of gene body DNA methylation independently of promoter regions, which are relatively depleted for DNA methylation, the promoter region was removed by excluding the first 3 kb of genes and removing genes < 4.5 kb in length. Gene bodies with fewer than 5 total CA or CG reads, or fewer than 10 MeCP2 ChIP or input reads were excluded from analysis. To generate smooth-line correlation plots, genes that were significantly up- or downregulated between MeCP2 mutant and WT were sorted by gene length or DNA methylation, and a sliding window was defined by the indicated bin and step sizes for each analysis. The bin and step sizes were adjusted to the length of each gene list. The average $\log_2$ fold-change and standard deviation for each bin were plotted for either the MeCP2 mutant versus WT comparison or the Shuffled comparison. Since there were no significantly misregulated genes in the Shuffled comparisons, the Shuffled fold-changes of the genes significantly misregulated in the MeCP2 mutant compared to WT comparison are shown. For aggregate plots of DNA methylation over gene bodies, mCA/CA and mCG/CG were calculated in 1 kb genomic bins using BEDTools (Quinlan and Hall, 2010) (v2.26) map, then metagene plots were generated with deepTools (v3.0.2) (Ramírez et al., 2016).

Pol II premature termination analysis

Chromatin-associated RNA-seq reads from 10 WT and 10 MeCP2 KO mice were counted in individual mm10 Refseq introns using Subread featureCounts (feature level). Individual introns were defined as the difference between the full gene body of the longest isoform for a gene and exons from all isoforms with BEDTools subtract (v2.26). Custom R scripts were used to calculate the RPKM of each intron as the read counts in the intron normalized to the intron length in kb and the total mapped reads in millions. Genes with at least 4 introns and RPKM expression of at least 1 were included in the analysis. The last/first intron ratio was calculated as the last intron RPKM / first intron RPKM for each gene.

Co-transcriptional splicing analysis

3′ splice site ratio analysis was performed as described in (Herzel and Neugebauer, 2015; Khodor et al., 2011) with minor modifications. Read coverage in chromatin-associated RNA-seq from 10 WT and 10 MeCP2 KO mice was calculated using SAMtools (v0.1.19) (Li et al., 2009) depth and read coverage in the 3′ 25 bp of each intron and 5′ 25 bp of the following exon was calculated using BEDTools intersect (v2.26). For genes with multiple isoforms, the longest isoform was used for analysis. Custom R scripts were used to sum the coverage in each individual exon and intron and calculate the fraction un-spliced for each 3′ splice site as the 25 bp intron coverage / 25 bp exon coverage. 3′ splice sites with exon coverage less than 200 were excluded.

Spurious intragenic initiation analysis

Whole-cell RNA-seq reads from 10 WT and MeCP2 KO mice were counted in individual mm10 Refseq exons using Subread featureCounts (feature level, multi-overlapping reads counted). Analysis was performed with custom R scripts as described previously (Neri et al., 2017). Genes with at least 5 exons and RPKM expression of at least 1 were included in the analysis. For genes with multiple isoforms, the longest isoform was used for analysis. The RPKM of each exon was calculated by the read counts in the exon normalized to the exon length in kb and the total mapped reads in millions. The last/first exon ratio was calculated as the last exon RPKM / first exon RPKM for each gene.

Nanostring nCounter gene expression analysis

Nanostring data were analyzed with the nSolver software (v4.0) using the default mRNA parameters, including quality control, no background subtraction or thresholding, and normalization to positive controls and housekeeping genes. Differentially expressed genes were defined by an FDR < 0.05 with no fold-change cut-off.

Chromatin immunoprecipitation sequencing analysis

Sequencing reads were trimmed with Trimmomatic (v0.36) to remove adapters and low-quality sequence (settings: LEADING:5 TRAILING:5 SLIDINGWINDOW:4:20 MINLEN:50). Trimmed reads were mapped to the mm10 genome with Bowtie2 (v2.2.9) (Langmead and Salzberg, 2012) with default parameters. PCR duplicate reads were removed with SAMtools (v0.1.19) rmdup. Reads were extended to 250 bp (to approximate fragment length) with UCSC tools (v363) bedextendranges. For each ChIP antibody, reads for each sample were randomly down-sampled to equal numbers using the GNU shuf utility. For ChIP-seq track visualization figures, the down-sampled reads from each replicate were merged, and bigwig files were generated with deepTools bamCoverage with a bin size of 25 bp, and were visualized with the UCSC genome browser. ChIP-seq reads were quantified in the genomic regions specified in figure legends using BEDTools map (v2.26). For histone modifications and Pol II ChIP-seq, differential ChIP analysis was performed on ChIP read counts in specified regions with the R package edgeR (v3.18.1) (Robinson et al., 2010). Genes with low ChIP read

counts were filtered by keeping only genes with rowSums(cpm(y) > 1) > = (number of samples per group). Boxplots show log_2 fold-change in MeCP2 mutant compared to WT ChIP from edgeR output for each gene list. H3K27ac peaks were defined using MACS2 with the ChIP input as background (–nomodel -q 0.05). Intragenic H3K27ac peaks were defined using BEDTools intersect as peaks that were within gene bodies but excluded from the first 3 kb of genes. For MeCP2 ChIP-seq, MeCP2 ChIP read counts were normalized to input read counts in the specified genomic regions.

PRO-seq analysis

Read mapping and quantification

Random hexanucleotide barcodes included in sequencing adaptors permitted the identification and removal of PCR duplicates using Clumpify (BBTools v37.90). Reads were trimmed to 25 bp and mapped to the mm10 reference genome using Bowtie2 (v2.2.9). Reads were converted to BED format using BEDTools (v2.26.0) and trimmed to the width of a single nucleotide at the second-to-3′ position, corresponding to the position of RNA Polymerase II before the incorporation of the terminal 11-biotin-NTP. PRO-seq reads were quantified in Refseq transcription start sites (TSS: −1 kb to +1 kb) and gene bodies (TSS +1 kb to TTS) using BEDOPS (v2.4.30) (Neph et al., 2012). The fold-change in mutant compared to WT PRO-seq signal was calculated with edgeR (v3.18.1). Genes with low PRO-seq read counts were filtered by keeping only genes with rowSums(cpm(y) > 1) > = (number of samples per group). Boxplots show log_2 fold-change in mutant compared to WT PRO-seq from edgeR output for each gene list.

Pausing analysis

To specifically analyze the pausing of productively elongation RNA Pol II molecules, CA sites were identified from a set of neuronally-expressed, non-overlapping genes (> 0.2 RPKM in the PRO-seq data, 7616 genes). To avoid the impact of gene features known to regulate rates of pausing and elongation, CA sites were only considered that were within an intron of at least 100 bp and not within 10 bp of a known splice site. CA sites were separated based on their levels of CA methylation (high, low, none), and PRO-seq reads were mapped to the surrounding regions using Homer (v4.9) (Heinz et al., 2010) and plotted as an averaged histogram. To reduce the impact of cryptic promoters and other sites with abnormally high PRO-seq signal, the top 0.1% of sites by PRO-seq density were removed from analyses. To correct for differences in PRO-Seq signal resulting from gene expression levels, the read density at a given position was divided by the average PRO-seq density across the gene in which it resides. For the YY1 motif analysis, ChIP-seq was performed for YY1 with Abcam ab109237 antibody as detailed (Malik et al., 2014) from E16.5 primary cortical neuron cultures cultured for 7 days *in vitro*. Peaks were defined with MACS2 (Zhang et al., 2008) as described (Vierbuchen et al., 2017). YY1 motifs (MA0095.2) within ± 250 bp of peak summits were annotated with FIMO (MEME Suite) (Grant et al., 2011). YY1 motifs were then oriented by the direction of the motif sequence and PRO-seq reads were mapped to the surrounding regions using Homer and plotted as an averaged histogram. This histogram was plotted using a moving window of 10 bp for smoothing.

Elongation rate calculation

The wavefront of elongating RNA Polymerase II was identified using a Hidden Markov Model as implemented previously (Jonkers et al., 2014). Elongation rates were calculated for genes greater than 200 kb as the linear fit of the Pol II wavefronts at 20, 40, and 60 minutes after flavopiridol treatment over time. Genes with a linear fit R-squared greater than 0.9 were included in downstream analysis. Heatmaps of PRO-seq signal were created in R using the EnrichedHeatmap package (Gu et al., 2018).

Initiation rate analysis

Aggregate plots of PRO-seq signal in −500 to +500 bp of the TSS were generated with Homer (v4.9). To quantify initiated Pol II signal, PRO-seq reads were counted in −500 to +500 bp of the TSS using BEDTools map. Initiation rates were calculated as the linear fit of 0 minute, 20 minute, and 40 minute initiated Pol II PRO-seq signal over time using the lm function in R. The 60 minute time point was excluded because the inhibition of Pol II pause-release by flavopiridol had begun to wear off, as evident from the lack of increase of initiated Pol II signal from 40 to 60 minutes. Genes with initiation rates greater than zero, TSS PRO-seq signal greater than 10, and a linear fit R-squared greater than 0.5 were included in analysis.

OxBS-seq analysis

BS-seq and OxBS-seq reads were mapped to the mm10 genome using BSmap (Xi and Li, 2009). OX and BS C and T reads were counted in 1 kb genomic bins using BEDTools map. mC levels were calculated as OX: #C/(#C+#T). hmC levels were calculated as BS: #C/(#C+#T) – OX: #C/(#C+#T). Metagene plots were generated with deepTools (v3.0.2).

Hi-C analysis

Hi-C analysis was performed as described in (Bonev et al., 2017) with minor modifications below.

To calculate the average gene contact enrichment map, we extend each gene (with a length x) both upstream and downstream by x. We then generated a 100x100 grid from the beginning until the end of the extended 2D interval. As each gene has a different length the constructed grid had varying bin sizes. We then calculated log ratio of the observed and the expected contacts in each bin of the 100x100 grid per gene. To account for the direction of transcription we further flipped the resulting matrices for genes transcribed in the reverse orientation, so that the TSS relative coordinates were matching. Figure 7G represents the mean values in each grid bin across different gene types.

To quantify the TSS – gene body enrichment, we calculated the log ratio of observed and expected contacts within a 2D intervals formed by TSS:TSS+10kb and the second half of the gene body (TSS+x/2: TSS+X).

DATA AND CODE AVAILABILITY

All genomic data reported in this study have been deposited in the NCBI Gene Expression Omnibus under accession GSE128186. Custom R scripts for individual analyses are available upon request.