Understanding the Roles of DNA Methylation and MeCP2 in Neuronal Enhancer Control

Adam Clemens
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Developmental Biology Commons, and the Neuroscience and Neurobiology Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Understanding the Roles of DNA Methylation and MeCP2 in Neuronal Enhancer Control

by

Adam W. Clemens

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

January 2021
St. Louis, Missouri
# Table of Contents

List of Figures ............................................................................................................................ v
List of Tables .............................................................................................................................. vii
Acknowledgments ..................................................................................................................... viii
Abstract .................................................................................................................................. xi

Chapter 1: Emerging insights into the distinctive neuronal methylome ........................................ 1
  1.1 DNA methylation guides genomic regulation ........................................................................ 2
  1.2 The unique neuronal methylome .......................................................................................... 2
  1.3 mCH as a hallmark of cell types .......................................................................................... 6
  1.4 Readout of mCH by MeCP2. ............................................................................................... 9
  1.5 Gene regulation by MeCP2 and neuronal DNA methylation ................................................ 12
  1.6 Mechanistic insights into gene regulation by neuronal methylation and MeCP2. .............. 16
  1.7 Biological function of mCH-MeCP2-mediated gene regulation ............................................ 21
  1.8 Expanding roles for mCH and MeCP2 in disease ................................................................. 23
  1.9 Figures ................................................................................................................................ 25
    1.9.1 Figure 1. Non-CG methylation and MeCP2 in neuronal gene regulation ...................... 25
    1.9.2 Figure 2. The unique neuronal epigenome .................................................................... 28
    1.9.3 Figure 3. Molecular mechanisms of the mC-MeCP2-NCoR axis and its disruption in neurodevelopmental disorders .......................................................... 30
    1.9.4 Figure 4. Neuron-enriched DNA modifications and their enzymes ............................. 32
  1.10 Glossary ............................................................................................................................. 33

Endnote 1: Neuron-enriched DNA modifications and their enzymes ........................................ 36
Endnote 2: Global, graded gene regulation by mC and MeCP2. ................................................ 37

Chapter 2: MeCP2 represses enhancers through chromosome topology-associated DNA
methylation ................................................................................................................................. 40
  2.1 Introduction ......................................................................................................................... 41
  2.2 Results ................................................................................................................................ 45
    2.2.1 Large-scale non-CG DNA methylation profiles are linked to chromatin topology ........ 47
    2.2.2 Loss of MeCP2 leads to promoter-associated transcriptional activation ..................... 54
2.2.3 MeCP2 represses enhancers that are enriched for mCA and mCG binding sites. ... 58
2.2.4 MeCP2-mediated enhancer repression is associated with gene regulation. ........ 64
2.3 Discussion ............................................................................................................... 67
2.4 STAR Methods ...................................................................................................... 75
  2.4.1 Key Resources Table ...................................................................................... 75
2.5 Experimental Model and Subject Details .............................................................. 85
2.6 Figures .................................................................................................................. 101
  2.6.1 Figure 1. Regional non-CG methylation patterns are associated with domains of chromatin folding. ................................................................. 102
  2.6.2 Figure 2. Domain-associated recruitment of DNMT3A defines megabase- and kilobase-scale mCA levels ................................................................. 104
  2.6.3 Figure 3. Loss of MeCP2 leads to promoter-associated transcriptional activation of MeCP2-repressed genes ............................................................... 106
  2.6.4 Figure 4. MeCP2 represses enhancers that are enriched for mCA and mCG binding sites. ................................................................. 109
  2.6.5 Figure 5. mCA-associated de-repression of enhancers in DNMT3A Baf53b-cKO. ........................................................................................................ 112
  2.6.6 Figure 6. MeCP2-repressed enhancers are linked to MeCP2-repressed genes..... 115
  2.6.7 Figure S1. Identification of MeCP2-regulated genes in the cerebral cortex by combined RNA-seq analysis of MeCP2 KO and MeCP2 OE. Related to Figure 1. .... 119
  2.6.8 Figure S2. Chromatin topology is associated with non-CG DNA methylation in the cerebral cortex. Related to Figure 1 ......................................................... 123
  2.6.9 Figure S3. Binding of DNMT3A and recruitment of MeCP2 are shaped by topologically-associating domains. Related to Figure 2 ........................................ 127
  2.6.10 Figure S4. Changes in intronic RNA are consistent with promoter-associated transcriptional upregulation of long, highly methylated, MeCP2-repressed genes in the MeCP2 KO. Related to Figure 3 ......................................................... 130
  2.6.11 Figure S5. Analysis of enhancers dysregulated in MeCP2 mutants. Related to Figure 4 and Figure 6 ................................................................. 134
  2.6.12 Figure S6. Transcriptomic and epigenomic analysis of DNMT3A Baf53b-cKO. Related to Figure 5 ................................................................. 137
Chapter 3: Probing the mechanism of neuronal enhancer regulation by MeCP2 .... 139
  3.1 Introduction ........................................................................................................ 140
  3.2 Results .............................................................................................................. 143
3.2.1 NCoR binding to regulatory elements in the brain is not dependent on MeCP2... 143
3.2.2 Loss of MeCP2 leads to alterations in accessibility and enhancer priming upstream of histone acetylation at MeCP2-regulated enhancers.................................................. 145
3.2.3 Loss of MeCP2 differentially affects cofactors associated with enhancer-mediated promoter activation........................................................................................................ 147
3.3 Discussion .................................................................................................................. 148
3.4 Figures ......................................................................................................................... 152
  3.4.1 Figure 1. NCoR binds to enhancers independent of MeCP2.................................... 152
  3.4.2 Figure 2. mC-MeCP2-NCoR drives changes in multiple steps of enhancer activation.......................................................................................................................... 155
Chapter 4: Concluding remarks and future perspectives ................................................. 157
  4.1 Introduction ................................................................................................................. 158
  4.2 Mechanisms of unique non-CG methylation patterns in the brain......................... 158
  4.3 Understanding enhanced susceptibility of intragenic enhancers to repression by MeCP2 160
  4.4 Understanding cell-type specific roles for mCH and MeCP2................................... 161
  4.5 Disruption of mCH and MeCP2 in neurodevelopmental disease............................. 163
  4.6 Figures ......................................................................................................................... 166
    4.6.1 Figure 1. Early developmental H3K36me2 distribution correlates with adult patterns of mCH........................................................................................................... 166
References ......................................................................................................................... 167
Curriculum Vitae .............................................................................................................. 183
List of Figures

Chapter 1

1.9.1 Figure 1. Non-CG methylation and MeCP2 in neuronal gene regulation ........................................ 25
1.9.2 Figure 2. The unique neuronal epigenome .......................................................................................... 28
1.9.3 Figure 3. Molecular mechanisms of the mC-MeCP2-NCoR axis and its disruption in neurodevelopmental disorders .......................................................................................................................... 30
1.9.4 Figure 4. Neuron-enriched DNA modifications and their enzymes .................................................. 32

Chapter 2

2.6.1 Figure 1. Regional non-CG methylation patterns are associated with domains of chromatin folding .................................................................................................................................................. 102
2.6.2 Figure 2. Domain-associated recruitment of DNMT3A defines megabase- and kilobase-scale mCA levels ........................................................................................................................................ 104
2.6.3 Figure 3. Loss of MeCP2 leads to promoter-associated transcriptional activation of MeCP2-repressed genes ........................................................................................................................................... 106
2.6.4 Figure 4. MeCP2 represses enhancers that are enriched for mCA and mCG binding sites ................................................................................................................................................. 109
2.6.5 Figure 5. mCA-associated de-repression of enhancers in DNMT3A Baf53b-cKO ....................... 112
2.6.6 Figure 6. MeCP2-repressed enhancers are linked to MeCP2-repressed genes ........................................ 115
2.6.7 Figure S1. Identification of MeCP2-regulated genes in the cerebral cortex by combined RNA-seq analysis of MeCP2 KO and MeCP2 OE. Related to Figure 1 ................................................. 119
2.6.8 Figure S2. Chromatin topology is associated with non-CG DNA methylation in the cerebral cortex. Related to Figure 1 .......................................................................................................................................... 123
2.6.9 Figure S3. Binding of DNMT3A and recruitment of MeCP2 are shaped by topologically-associating domains. Related to Figure 2 .............................................................................................................. 127
2.6.10 Figure S4. Changes in intronic RNA are consistent with promoter-associated transcriptional upregulation of long, highly methylated, MeCP2-repressed genes in the MeCP2 KO. Related to Figure 3 ........................................................................................................................................... 130
2.6.11 Figure S5. Analysis of enhancers dysregulated in MeCP2 mutants. Related to Figure 4 and Figure 6 ................................................................................................................................................. 134
2.6.12 Figure S6. Transcriptomic and epigenomic analysis of DNMT3A Baf53b-cKO. Related to Figure 5 ................................................................................................................................................. 137
List of Tables

2.4.1 Key Resources Table ........................................................................................................ 76
Acknowledgments

I want to thank the members of the Gabel lab, all of whom have made it a joy to be in lab. You have provided valuable experimental advice, necessary coffee breaks, and a half-way decent sounding board for “jokes”. Thank you to Harrison for being an open-minded mentor who has always put his students first. Your drive for the advancement of science for science-sake is admirable and astounding, and thank you for the endless supply of coffee. Thank you to Dennis for being the cooler head and your constant willingness to offer help, Sabin for your the comedic-relief and selfless nature, Russell for never telling a true story, Diana for being the reasonable one, Mati for taking over responsibility for “jokes”, Nicole for your enthusiasm and willingness to take over my projects, Jenna for gracing us with your presence before your wildly-successful career in science, and Ryan for being instrumental in establishing the lab.

Thank you to the funding sources that have made the work discussed here possible. National Institutes of Health (NIH) through 5T32GM007067 and F31NS108574 to AWC and The Klingenstein-Simons Fellowship Fund, the G. Harold and Leila Y. Mathers Foundation, the Brain and Behavior Research Foundation, The Simons Foundation Autism Research Initiative 508034, and NIMH R01MH117405 to HWG.

I appreciate the time and scientific and professional advice provided by my thesis committee: Kristen Kroll, Ting Wang, Andrew Yoo, and Shiming Chen.

Thank you to those behind the scenes. Jim Skeath and Sally Voigt for offering assurance and guidance. Thank you to Debbie Pfeiffer for constantly smiling, offering candy, and keeping our floor together.
Thank you to friends. Without you, this would have been a bear. With you, my heart and stomach are full. You are all remarkable.

Thank you to Peaches for your companionship.

Thank you to my parents for your constant support and love.

Adam W. Clemens

*Washington University in St. Louis*

*January 2021*
Dedicated to curiosity.
Abstract

Understanding the roles of DNA methylation and MeCP2 in neuronal enhancer control

by

Adam Clemens

Doctor of Philosophy in Biology and Biomedical Sciences

Developmental, Regenerative, and Stem Cell Biology

Washington University in St. Louis, 2020

Professor Harrison Gabel, Chair

The development of organisms relies on complex spatial and temporal patterning of gene expression to define cell types and facilitate their functions. Cis-regulatory elements in our genome are responsible for the control of gene expression across tissues and cell types. Regulation of these elements themselves depends on a balance of activation and repression through epigenetic modifications and molecular regulatory components. Disruption of cis-regulatory element control is emerging as a cause of neurodevelopmental disease. An important contributor to gene regulation in development is methylation of cytosine in DNA, disruption of which has been associated with disease. Notably, while all cell types employ methylation at CG dinucleotides to control gene expression, mammalian neurons contain uniquely high levels of non-CG DNA methylation that are critical for proper nervous system function. MeCP2, the protein associated with Rett syndrome, in turn binds to non-CG methylation to regulate gene expression. Defining how non-CG methylation accumulates in neurons and is read out by MeCP2 will improve our understanding of the unique gene-regulatory environment in these cells and can begin to decipher the molecular
underpinnings of neurodevelopmental disorders. Here, I explore the role of gene expression and genome architecture in establishing patterns of non-CG methylation in neurons and highlight emerging mechanistic insights into how non-CG methylation and MeCP2 control transcription through neuronal enhancers. I define a nested pattern of methylation by which highly-methylated, mega-base scale topologically-associated domains contain genes of high methylation, which themselves contain enhancers of high methylation. MeCP2 represses enhancers found in these methylation-enriched domains, with the strongest repression occurring for enhancers located within genes repressed by MeCP2. I show that loss of methylation-dependent repression of enhancers can drive changes in gene expression in models of disease. Finally, I outline preliminary findings identifying upstream and downstream mechanisms of enhancer regulation through MeCP2 and DNA methylation. These insights provide clues as to how the distinctive epigenome in neurons facilitates the development and function of the complex mammalian brain.
Chapter 1: Emerging insights into the distinctive neuronal methylome

This chapter is adapted from a review published in *Trends in Genetics*.

Clemens, AW., Gabel, HW. Emerging insights into the distinctive neuronal methylome. *Trends in Genetics* in press.
1.1 DNA methylation guides genomic regulation.

Eukaryotic gene expression is guided by covalent chromatin modifications that facilitate temporal and spatial control of transcription in diverse cell types during development and across dynamic processes (Zhou et al., 2017). The addition of a methyl group to the 5’ position of cytosine nucleotides (mC) is a major epigenetic modification that contributes to gene regulation across phylogeny (Zhou et al., 2017). mC can block the binding of gene regulatory proteins such as transcription factors, or recruit “reader” proteins that affect chromatin structure and alter transcription. There is evidence that DNA methylation participates in both activation and repression of gene expression; however, in mammals it is predominantly associated with repeat and transposable element silencing, as well as gene repression (reviewed in (Greenberg and Bourc’his, 2019)). While mC at CG dinucleotides contributes to gene regulation in all cell types, in the last decade it has become clear that DNA methylation has unique and essential roles in the nervous system. Here, we review the discovery and characterization of prevalent mC at non-CG sequences in neurons and discuss the identification of Methyl CpG-binding Protein 2 (MeCP2), the protein disrupted in the neurodevelopmental disorder Rett syndrome (see Glossary), as an essential reader of this mark. We explore new results shedding light on mechanisms of neurodevelopmental disorders caused by disruption of DNA methylation and gene regulation mediated by MeCP2.

1.2 The unique neuronal methylome.

Classically, DNA methylation in mammals was described almost exclusively at cytosines followed by guanines (mCG), with symmetric methylation occurring at cytosines on both strands. mCG is
the predominant form of DNA methylation in most tissue types (reviewed in (Greenberg and Bourc’his, 2019)). However, in the brain, alternative forms of methylation are abundant. The neuronal methylome was first recognized as unique when high levels of oxidation of mCG sites to hydroxymethylation (hmCG) were discovered in Purkinje cells of the cerebellum and in brain tissue (Endnote 1) (Kriaucionis and Heintz, 2009). More recently, prevalent non-CG DNA methylation (mCH; where H=A, C, or T) has been identified, with this methylation primarily occurring at cytosines followed by adenine (mCA). mCH is highly enriched in neurons compared to other cell types in mouse and humans, and while the methylation rate of CH is lower than that of CG, in some classes of neurons the number of modified CH sites (~1.5-3% for 1.1 billion CH sites = 16-30 million modified CH sites) is equivalent to, or higher than, total modified CG sites (~80% for 21 million CG sites = ~17 million modified CG sites) (Figure 1A; 2A; discussed in Endnote 1) (Guo et al., 2014; Lister et al., 2013; Xie et al., 2012).

mCH is deposited by the de novo DNA methyltransferase 3A (DNMT3A) (Endnote 1), which is upregulated in neurons at birth and reaches peak expression at ~2 weeks in mice, before declining to lower expression levels in adulthood (Gabel et al., 2015; Guo et al., 2014; Lister et al., 2013). In frontal cortex, this expression leads to postnatal accumulation of mCH, which plateaus by 4-6 weeks in mice. In humans, mCH builds up primarily during the first two years, but requires sixteen years to fully accumulate (Figure 2B) (Lister et al., 2013). Like mCG, mCH occurs broadly across the genome. Postnatal mCH accumulation across the neuronal genome is influenced by pre-existing gene expression and chromatin structure (Clemens et al., 2020; Lister et al., 2013; Stroud et al., 2017). Little to no mCH is deposited at completely silent genes and inaccessible regions of constitutive heterochromatin (e.g. olfactory receptor genes clusters) (Lister et al., 2013). Within
euchromatic regions, DNMT3A is readily recruited and deposits mCH at repeated sequences, extragenic regions, lowly transcribed genes, and inactive regulatory elements. In contrast, DNMT3A binding and mCH accumulation are depleted from the transcribed region of highly expressed genes and active regulatory elements (Figure 1B; 2C,D) (Kozlenkov et al., 2018; Lister et al., 2013; Luo et al., 2017; Mo et al., 2015, 2016; Stroud et al., 2017). Experimental manipulation of gene expression in mouse cerebral cortex during the postnatal period indicates that high transcriptional activity blocks DNMT3A binding in genes and results in low mCH accumulation, which persists throughout adulthood (Stroud et al., 2017). Readout and repression of genes through mCH in adult neurons likely reinforces low expression of highly methylated genes, while lowly methylated genes escape repression and are highly transcribed (Stroud et al., 2017) (discussed below). Genes expressed at moderate levels postnatally build up intermediate levels of mCH, resulting in balanced activation by transcription-promoting machinery and repression by mCH in adult neurons that may tune gene expression levels (Stroud et al., 2017).

Non-CG methylation depletion at genes and regulatory elements largely parallels patterns of mCG at a local scale (kilobase), but mCH shows unique variations on a large scale (megabase) that have been linked to the folding of chromosomes within the nucleus (Clemens et al., 2020). Megabase-sized genomic regions show enrichment and depletion of mCH that correlate with Topologically Associating Domains (TADs) of chromatin folding (Clemens et al., 2020). TADs appear to be regions of consistent DNMT3A binding and accumulation of mCH, such that sequences found within individual TADs share similar mCH levels, or a “mCH set-point”, while sequences in neighboring TADs can have very different mCH levels (Clemens et al., 2020). This consistency of TAD mCH impacts genes and enhancer elements within TADs. For example, genes and
enhancers in a high-mCH TAD tend to have higher methylation than genes and enhancers in a low-mCH TAD (Figure 2D) (Clemens et al., 2020). Importantly, these differences in TAD methylation influence the regulation of genes in TADs by mCH. More details on the neuronal patterning and importance of mCH are further discussed below and within Chapter 2.

While transcription of genes and TAD structure are clearly associated with mCH deposition, molecular mechanisms controlling DNMT3A activity to create these patterns are not yet defined. Analysis of diverse histone modifications in mouse cortex suggests that chromatin structure during early postnatal development impacts mCH deposition (Stroud et al., 2017), and studies of mCG deposition by DNMT3A outside the nervous system may provide clues to mCH deposition mechanisms within neurons. DNMT3A can bind to unmethylated lysine 4 on histone H3 (H3K4me0) through its ATRX-DNMT3-DNMT3L (ADD) domain, which releases an auto-inhibitory conformation to allow cytosine methylation (Guo et al., 2015; Rondelet et al., 2016). This mechanism could restrict mCH deposition from active regulatory elements, which are marked by H3K4 methylation. The Pro-Trp-Trp-Pro (PWWP) domain of DNMT3A can bind methylated lysine 36 on histone H3 (H3K36me) (Dhayalan et al., 2010; Guo et al., 2015). Based on studies of the close paralog, DNMT3B, this domain has been thought to bind H3K36me (Greenberg and Bourc’his, 2019), but recent studies indicate similar or more robust binding to H3K36me2 compared to H3K36me3. Notably, H3K36me2 has been shown to accumulate in broad euchromatic regions and facilitate CG methylation of these domains in dividing mouse and human cells (Dukatz et al., 2019; Weinberg et al., 2019; Xu et al., 2020), while H3K36me3 is associated with the gene bodies of highly expressed genes. If similar patterns of H3K36me2 exist in neurons, DNMT3A may bind to broadly distributed H3K36me2 to guide TAD-scale methylation, while
conversion of H3K36me2 to H3K36me3 in the gene body could result in lower levels of DNMT3A recruitment and less mCH deposition in highly expressed genes. Future studies examining the effects of disruption of these histone marks on DNMT3A localization and activity in neurons will help to define the precise mechanisms that govern the deposition of mCH in the brain. For instance, it has yet to be determined how the ADD domain functions in neurons and if it shares the same auto-inhibitory role suggested from structural analysis. Moreover, it is unknown how gene expression, H3K36me2, and H3K36me3 function together to balance to the recruitment and repression of DNMT3A to gene bodies within neurons. Identifying these regulatory roles for gene expression and DNMT3A within neurons will provide valuable insight into how gene expression and chromatin structure dictate the distribution of neuronal DNA methylation.

1.3 mCH as a hallmark of cell types.

An intriguing feature of non-CG methylation is its high degree of cell-type specificity, both in global levels of mCH and in local patterns of demethylation at genes and regulatory elements. In both mouse and human, levels of mCH can vary by up to 2-fold between brain regions (Christian et al., 2020; Gabel et al., 2015; Guo et al., 2014; Lister et al., 2013; Mellén et al., 2017) and 1.5-fold among neuron subtypes in the same brain region (Luo et al., 2017; Mo et al., 2015, 2016). For example, somatostatin- (SST+) and parvalbumin-positive (PV+) inhibitory interneurons in the cerebral cortex are enriched approximately 30% for mCH compared to vasoactive intestinal polypeptide-positive (VIP+) neurons (Luo et al., 2017; Mo et al., 2015). mCH is also enriched 30-50% in deep layer cortical excitatory neurons compared to their upper layer counterparts (Luo et al., 2017). These large variations contrast with smaller global differences in mCG across brain
regions and cell classes (Figure 2E) (Luo et al., 2017; Mo et al., 2015). Studies of single neuron methylomes in hippocampus and cortex suggest that both the subtype of a neuron and its location influence mCH levels. For example, inhibitory neuron classes (e.g. PV+) from both cortex and hippocampus share similar mCH levels, but total amounts can be different for cells within the same class in different layers within a region (Liu et al., 2020; Luo et al., 2017). How these global variations in mCH originate has not been determined, but differential expression or activation of DNMT3A in cell types during postnatal development is a potential mechanism that future studies can explore. In all, the varying global mCH levels across cell types suggest that mCH may play a larger regulatory role in some brain regions and neuronal classes compared to others.

The role of MeCP2 in neuronal function has been extensively characterized across multiple brain regions through studies of mice carrying knockout, overexpression, and disease-associated missense mutations (Gulmez Karaca et al., 2019; L.M. et al., 2015; Lyst and Bird, 2015; Tillotson and Bird, 2019). When exogenous MeCP2 is expressed specifically in post-mitotic neurons of a mouse lacking MeCP2, the phenotypes associated with the MeCP2 knockout mice are rescued back to wild-type (Luikenhuis et al., 2004). However, there is evidence of some importance of MeCP2 in glia, in which loss of MeCP2 in glia show a poisoning effect on neurons, which suggests that while MeCP2 likely plays an important function in glia, the majority of the phenotypes may result from neuronal-dysfunction (Ballas et al., 2009; Jin et al., 2017). Moreover, re-expression of MeCP2 in glia or microglia of an MeCP2-null mouse only partially recovers phenotypes, unlike restoration of MeCP2 in neurons (Cronk et al., 2015; Nguyen et al., 2013). Researchers have remained focused on MeCP2 in neurons, not just because it is enriched in neurons over glia, but also because CA methylation does not accumulate in glia like it does in neurons (Lister et al.,
Additionally, the more prominent effects of MeCP2 disruption in neurons compared to glia underlie the importance of the presence of both MeCP2 and mCA to observe the functional effects associated with MeCP2.

Differential CG methylation at genes and regulatory sequences has historically been known to contribute to differentiation and maintenance of distinct cell types (reviewed in (Greenberg and Bourc’his, 2019)). In addition to global variations, local mCH profiles at genomic loci show even more robust cell-type specific patterning than canonical mCG (Luo et al., 2017; Mo et al., 2015). Gene body mCH patterns across cell types are tightly associated with gene expression (Luo et al., 2017; Mo et al., 2015; Stroud et al., 2017), such that mCH within genes varies to a greater degree and is more highly correlated with cell-type specific expression than mCG or open chromatin signatures, which show a less dynamic range in signal at genes across cell types (Luo et al., 2017; Mo et al., 2015). Emerging compendiums of single-cell methylomes across multiple mouse brain regions are further defining cell-type specific patterns of mCH (Liu et al., 2020; Yao et al., 2020). These data show that mCH profiles of individual neurons can be used to predict the precise location of a given neuron within one of five different brain regions and the laminar position within that region. The data also reveal patterns of gene expression and enhancer activity across increasingly refined neuronal subtypes. Cell-type specific mCH patterns appear to be dictated by existing gene expression patterns in the early postnatal period through the mechanisms described above (Figure 2C). Once established, these mCH patterns function with canonical mCG to maintain cell-type specific gene programs in the adult brain (Figure 2D) (Stroud et al., 2017) (discussed below).
1.4 Readout of mCH by MeCP2.

Insights into the functional importance of mCH in the brain have emerged through studies establishing MeCP2 as a major reader of this methyl mark. MeCP2 accumulates dramatically in neurons during postnatal development in parallel with the build-up of mCH (Figure 2B) (Kishi and Macklis, 2004; Shahbazian, 2002; Skene et al., 2010). In mature neurons, MeCP2 protein reaches expression levels nearly equivalent to that of histone H4 (Figure 2A) (Skene et al., 2010), and the expression of MeCP2 in neurons has been shown to be essential for nervous system function (Luikenhuis et al., 2004; Tillotson and Bird, 2019). While MeCP2 was originally identified as a reader of mCG sites, the discovery of high levels of mCH in the brain prompted close examination of its affinity for this methyl-mark. Indeed, several independent studies identified high-affinity binding of MeCP2 to mCH sites, specifically mCA (Chen et al., 2015; Gabel et al., 2015; Guo et al., 2014). MeCP2 binds to methylated DNA through its methyl-binding domain (MBD), a motif common amongst other methyl-binding proteins (Ginder and Williams, 2018; Hendrich and Bird, 1998), and shows the strongest binding to mCA compared to the other members of this protein family (Ginder and Williams, 2018; Sperlazza et al., 2017). Notably, mCAC is the most common site for non-CG methylation in the neuronal genome (Lister et al., 2013; Xie et al., 2012) and is the highest affinity trinucleotide non-CG site for MeCP2 binding (Lagger et al., 2017). The preference of MeCP2 for the most prevalent mCH site suggests a functional evolution of non-CG binding for MeCP2 (Lagger et al., 2017). Further support for readout of mCA as a critical player in MeCP2 function comes from an emerging study of mice carrying an engineered MeCP2 protein that can bind mCG, but not mCA. These mice recapitulate many neurologic phenotypes and gene expression changes seen in MeCP2 knockouts (Tillotson et al., 2020), indicating that mCG binding...
is not sufficient for normal nervous system function. Together, these studies emphasize readout of mCA by MeCP2 as essential in the brain.

Interestingly, the high levels of hydroxymethylation in neurons (Endnote 1) may increase the functional importance for mCH as a site of MeCP2 binding. Biochemical and structural studies indicate that MeCP2 has a lower affinity for hmCG than mCG, while conversion of mCH to hmCH appears to have little effect on MeCP2 binding (Buchmuller et al., 2020; Gabel et al., 2015; Hashimoto et al., 2012; Lagger et al., 2017; Mellén et al., 2017; Sperlazza et al., 2017; Valinluck et al., 2004). The large-scale conversion of mCG to hmCG (Endnote 1) (Kozlenkov et al., 2018) has been proposed to inactivate (“functionally demethylate”) high-affinity CG binding sites of MeCP2. Given that the number of MeCP2 molecules in neurons appears to be substantially lower than total numbers of mCG and mCH binding sites (Figure 2A, I) (Lister et al., 2013; Skene et al., 2010), hmCG accumulation could shift MeCP2 binding in favor of mCH or hmCH sites (Buchmuller et al., 2020; Gabel et al., 2015; Hashimoto et al., 2012; Lagger et al., 2017; Mellén et al., 2017; Sperlazza et al., 2017; Valinluck et al., 2004).

Chromatin-immunoprecipitation-sequencing (ChIP-seq) studies of MeCP2 in mouse and human brain and isolated neuronal cell populations have detected extremely broad occupancy of the protein across the genome, with relative enrichment of binding at methylated DNA (Baker et al., 2013; Chen et al., 2015; Cholewa-Waclaw et al., 2019; Cohen et al., 2011a; Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017; Mellén et al., 2017; Renthal et al., 2018; Rube et al., 2016; Skene et al., 2010; Stroud et al., 2017). At approximately 16 million molecules of MeCP2 per neuronal nucleus (Skene et al., 2010), there are sufficiently high numbers of MeCP2 molecules to
engage a substantial percentage of the ~24-44 million total mCG and mCH binding sites for the protein in typical neurons (Endnote 1). Indeed, ChIP signals from multiple studies reflect near-ubiquitous binding with high levels of enrichment (~10-100 fold) compared to Mecp2 knockout controls at all sites in the genome assessed (Boxer et al., 2020; Cohen et al., 2011b). Within the context of broad binding, MeCP2 ChIP signal is enriched at regions with high levels of mCG and mCH (e.g. extragenic regions) and is depleted at regions with low mCH and mCG and high levels of hmCG (e.g. promoters, enhancers, gene bodies for highly expressed genes). However, the magnitude of this depletion is minimal (~1-2-fold), even at regions that are essentially devoid of mC sites (Baker et al., 2013; Chen et al., 2015; Cholewa-Waclaw et al., 2019; Cohen et al., 2011a; Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017; Mellén et al., 2017; Renthal et al., 2018; Rube et al., 2016; Skene et al., 2010; Stroud et al., 2017). It is unclear if this limited dynamic range is a technical limitation of the ChIP method or if it indicates that MeCP2 binds substantially to unmethylated DNA in vivo. In vitro studies have shown that MeCP2 is capable of binding unmethylated DNA, with a preference for GTG residues, albeit with lower binding affinity than methylated DNA (Lei et al., 2019; Meehan et al., 1992). However, recent analyses of MeCP2 binding in cells indicate that ChIP-seq and footprint signals for MeCP2 are not enriched at unmethylated GT-rich DNA sequences (Connelly et al., 2020). A new study suggests that MeCP2 may undergo phase separation with DNA (Fan et al., 2020; Wang et al., 2020) (discussed below), raising the possibility that condensates of MeCP2 may drive multi-valent contacts with regions of the genome. These associations may contribute to the observed ubiquitous binding pattern of MeCP2, even at sites of low methylation. Together, the largely ubiquitous and low dynamic range patterns of MeCP2 ChIP-seq signal have not allowed researchers to definitively classify specific “target genes” of MeCP2 based on binding profiles alone. Rather, some genes and regulatory
elements display modest enrichment of MeCP2 binding compared to those with modest depletion (Baker et al., 2013; Boxer et al., 2020; Clemens et al., 2020; Gabel et al., 2015; Rube et al., 2016). This suggests MeCP2 may play a regulatory role at virtually every region of the genome (Endnote 2).

Despite these challenges, in vitro and in vivo binding studies clearly indicate that MeCP2 binds with high affinity to mC and does exhibit enriched binding patterns by ChIP-seq (Gabel et al., 2015; Guo et al., 2014; Lagger et al., 2017; Rube et al., 2016). Building on this knowledge, researchers have employed genomic analysis of high affinity MeCP2 binding sites, mCA and mCG, together with transcriptomic studies to establish a functional role of MeCP2 in controlling transcription of methylation-rich genes (Baker et al., 2013; Chen et al., 2015; Cholewa-Waclaw et al., 2019; Cohen et al., 2011a; Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017; Mellén et al., 2017; Renthal et al., 2018; Rube et al., 2016; Skene et al., 2010; Stroud et al., 2017).

1.5 Gene regulation by MeCP2 and neuronal DNA methylation.

Since its initial identification as a reader of mCG, a myriad of putative protein binding partners for MeCP2 have been identified. These include proteins involved in transcriptional repression and activation, splicing regulation, and microRNA processing, suggesting diverse molecular functions of MeCP2 (reviewed in (Ip et al., 2018; Tillotson and Bird, 2019)). In addition, MeCP2 is heavily phosphorylated in response to neuronal stimulation, which modulates its activity (Yap and Greenberg, 2018). A large body of evidence supports gene repression as a major function of
MeCP2 (Tillotson and Bird, 2019), and this direct, repressive function of MeCP2 is the focus of discussion here. The best characterized interactor for MeCP2 is the **Nuclear Co-Repressor complex (NCoR)**. NCoR binding is critical for the repressive function of MeCP2 in *in vitro* assays, and mutations of *MECP2* that specifically disrupt this interaction have been shown to drive Rett syndrome (Kong et al., 2020; Tillotson and Bird, 2019). Identifying the mechanisms by which the MeCP2-NCoR complex affects gene expression is a major outstanding challenge for the field.

The function of MeCP2 in gene regulation has been intensely studied, but remains difficult to decode (Lavery and Zoghbi, 2019; Tillotson and Bird, 2019). Hundreds to thousands of genes can be detected as significantly dysregulated in transcriptomic studies of brain tissue and isolated cells from *Mecp2* knockout and missense mutants or *MECP2* over-expressing mice, as well as humans with Rett syndrome. However, the magnitude of dysregulation for these genes is subtle (less than 2-fold), and the near-ubiquitous binding patterns of MeCP2 do not provide sufficient evidence to suggest that these genes are the exclusive targets of regulation by the protein (Endnote 2).

Since its identification as a high-affinity reader of mCG, multiple lines of evidence have supported gene repression as a major function of MeCP2. It is indeed thought that MeCP2 largely functions as a repressor due to its observed interaction with the NCoR complex *in vitro* through MeCP2’s NCoR-interacting domain (NID), a critical sequence contained within the transcriptional repression domain (TRD) of MeCP2. Notably, the NID contains several sites of mutations associated with Rett syndrome, including R306C, the most common mutation observed in Rett syndrome that abolishes this NCoR interaction (the MBD also contains a large percentage of Rett-associated mutations) (Kruusvee et al., 2017; Lyst and Bird, 2015; Lyst et al., 2013). Moreover,
the repressive effects of MecP2 have been combined with the repressive KRAB domain in in vitro dCas9 experiments to generate more robust repression at targets sites (Yeo et al., 2018).

Recent integrated analysis of whole-genome bisulfite-sequencing DNA methylation maps and transcriptomic changes in MeCP2 knockout and missense mutants or MECP2 transgenic mice has detected methylation signatures on genes most highly affected by MeCP2 and has provided clues to the mechanism of MeCP2-mediated gene regulation (Boxer et al., 2020; Chen et al., 2015; Clemens et al., 2020; Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017; Lavery et al., 2020; Renthal et al., 2018; Rube et al., 2016; Stroud et al., 2017) (Figure 1B). These studies identify a reproducible enrichment of mCH compared to the genome average within the gene body and flanking sequences of MeCP2-repressed genes, those that show significant increases in expression when MeCP2 is disrupted and decreases when MeCP2 is overexpressed. The high levels of mCH in and around MeCP2-repressed genes arise from high mCH in the TADs of which these genes are found, indicating that genome topology plays a role in establishing high methylation at these genes (Clemens et al., 2020). MeCP2-activated genes, those that show significant decreases in expression when MeCP2 is lost and increases when it is overexpressed, are often found within lower mCH TADs, although they can show moderate gene body mCH enrichment in some studies (Figure 1B) (Boxer et al., 2020; Chen et al., 2015; Clemens et al., 2020). Though it is low in dynamic range, the MeCP2 ChIP-signal in and around MeCP2-repressed genes is enriched (Boxer et al., 2020; Chen et al., 2015; Lagger et al., 2017; Renthal et al., 2018), further supporting a direct role of MeCP2 in repression of these genes.
The function of mCH in MeCP2-mediated gene repression is supported by studies in which mCH accumulation was blocked through perinatal conditional deletion of \textit{Dnmt3a} in the brain or specifically in neurons of mice (Boxer et al., 2020; Clemens et al., 2020; Gabel et al., 2015; Kinde et al., 2016; Lavery et al., 2020; Stroud et al., 2017) (Figure 3A). These analyses detected alterations in gene expression in the absence of mCH that partially recapitulate the effects observed in \textit{Mecp2} knockout and missense mutants and demonstrated a loss in MeCP2 ChIP-signal at sequences that lose the most mCH upon deletion of \textit{Dnmt3a} (Clemens et al., 2020; Kinde et al., 2016). Importantly, because mCG is pre-established early in development by DNMT3A/B and is largely maintained in neurons by DNMT1 (Okano et al., 1999), the effects observed after \textit{Dnmt3a} conditional deletion can be attributed to the absence of mCH (Boxer et al., 2020; Clemens et al., 2020; Gabel et al., 2015; Kinde et al., 2016; Lavery et al., 2020; Stroud et al., 2017). This provides substantial in vivo evidence supporting mCH as a key site through which MeCP2 affects transcription.

Notably, while signatures of methylation can be detected at lists of genes significantly dysregulated in \textit{Mecp2} and \textit{Dnmt3a} mutants, neuronal DNA methylation and MeCP2 are present at every gene in the genome to varying degrees, and therefore may impact all genes. Indeed, gradients of gene dysregulation associated with genic methylation levels have been detected across all genes genome wide in \textit{Dnmt3a} and \textit{Mecp2} knockout, conditional knockout, and missense mutants (Endnote 2) (Boxer et al., 2020; Cholewa-Waclaw et al., 2019; Clemens et al., 2020; Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017; Lavery et al., 2020; Renthal et al., 2018; Stroud et al., 2017). This suggests that disruption of mCH or MeCP2 has subtle yet global effects on neuronal transcriptomes and therefore can have far-reaching impacts on circuit function.
1.6 Mechanistic insights into gene regulation by neuronal methylation and MeCP2.

The identification of DNA methylation signatures associated with MeCP2-mediated regulation has provided a starting point for studies dissecting the mechanism of this regulation. The observation that MeCP2-repressed genes are enriched for methylation within the gene body and flanking regions, rather than sequences at the transcription start site, has suggested that MeCP2 regulates transcription through binding to mC outside of the core promoter region (Figure 1B) (Chen et al., 2015; Gabel et al., 2015; Kinde et al., 2016). This finding, combined with the fact that MeCP2-repressed genes also tend to be expressed through extremely long pre-mRNAs, led to the initial hypothesis that MeCP2 binds mC and acts as a “speed-bump”, inhibiting processivity of RNA polymerase II (RNAPII) (Cholewa-Waclaw et al., 2019; Kinde et al., 2016; Lagger et al., 2017). However, recent studies in mice analyzing intronic RNA-seq and GRO-seq data, as well as ChIP-seq data for RNAPII and histone modifications associated with transcription (H3K4me3, H3K27ac, H3K36me3) (Boxer et al., 2020; Clemens et al., 2020) did not support this prediction. Rather than finding altered rates of RNAPII processivity, these studies detected altered promoter activity and transcription initiation in genes dysregulated upon mutation of Mecp2 (Boxer et al., 2020; Clemens et al., 2020).

How can binding of MeCP2 to methylation outside of the promoter control transcription initiation? One possibility is that MeCP2 bound to mCH and mCG within and outside of genes can loop to contact promoters, recruiting the NCoR complex and repressing transcription initiation (Figure
3B) (Boxer et al., 2020). In support of this possibility, recent Hi-C analysis in mice indicates that contacts between these regions and promoters do occur in the brain (Boxer et al., 2020). Substantial additional evidence suggests that MeCP2-NCoR binding to mC at distal sites can broadly block histone acetylation (Boxer et al., 2020; Clemens et al., 2020; Shahbazian et al., 2002; Skene et al., 2010), and genomic looping could bring this function to promoters.

In a parallel mechanism, MeCP2 bound to mC at enhancers locally represses the capacity of these elements to activate their cognate genes (Clemens et al., 2020). Loss of MeCP2 in mice leads to an increase in histone acetylation at sequences with highly methylated TADs, particularly within MeCP2-repressed genes (Boxer et al., 2020; Clemens et al., 2020). These effects are most robust at enhancer elements, where de-repression of acetylation in the absence of MeCP2 is correlated with the number of mCG and mCH sites. Notably, intragenic enhancers are more highly repressed by MeCP2 than extragenic enhancers, providing one explanation for the original identification of enriched intragenic methylation in MeCP2-repressed genes. A role for both mCG and mCH in mediating these effects at enhancers has been further supported through analysis of Dnmt3a conditional knockout mice lacking mCH (Clemens et al., 2020). In addition, a recent study in mouse cortex found that DNA methylation accumulates postnatally in embryonic-specific enhancers and MeCP2 represses these elements (Stroud et al., 2020). Thus, deacetylation at enhancers and potentially other genomic sequences is an important consequence of mC binding by MeCP2 that contributes to gene regulation in neurons.

By parsing out the effects of mCH and mCG, a clearer and more definitive understanding of the importance of both CH and CG methylation for the function of MeCP2 is revealed. Compared to
the effects seen in the MeCP2 KO model, only a partial effect on gene regulation and regulatory
element activity was observed due to the loss of roughly half of the binding sites of MeCP2
(Clemens et al., 2020). Since, mCG and MeCP2 both remain at similar levels between the
DNMT3A cKO and wild-type mice, there were no changes in enhancer activity as a function of
mCG at enhancers. Conversely, the loss of mCA sites drove changes in enhancer activity as a
function of mCA levels found at those regions in wild-type mice. Therefore, the partial effects
seen in the DNMT3A cKO compared to the MeCP2 KO are derived from the persisting mCA-
dependent changes and the lack of mCG-dependent changes. This illustrates that both mCA and
mCG are important for the full functional activity of MeCP2 and that neuronal DNA methylation
and MeCP2 have overlapping regulatory functions. Chapter 2 provides a deeper discussion of the
unique and congruent aspects of DNA methylation, MeCP2, and DNMT3A in neuronal gene
regulation.

The biological role of NCoR in the context of MeCP2 binding to mC is also coming into focus.
The NCoR complex is known to possess deacetylase activity mediated by its HDAC3 subunit
(Guenther et al., 2001). It is therefore intuitive to theorize that this activity could mediate
repression by MeCP2. Specific disruption of the MeCP2-NCoR interaction through an MeCP2
point mutation in mice leads to similar effects on histone acetylation and gene expression as the
Mecp2 knockout, underscoring the role of NCoR in these effects (Boxer et al., 2020; Gabel et al.,
2015). In addition, loss of MeCP2 and loss of HDAC3 result in shared social and motor
impairments as well as an overlap of dysregulated genes associated with neuronal function in mice,
supporting a role for HDAC3 in MeCP2-NCoR regulation (Nott et al., 2016). However, a recent
study tested the importance of NCoR-associated HDAC activity by assessing the severity of
phenotypes when MeCP2 is overexpressed in vivo. This study found that introducing an R306C missense mutation associated with Rett syndrome into overexpressed MeCP2 disrupts the NCoR-MeCP2 interaction and blocks the toxicity of MeCP2 overexpression. In contrast, introducing mutations into NCoR components that inhibit activation of HDAC3 did not rescue lethality (Koerner et al., 2018), suggesting the deacetylase activity of the NCoR complex is not required for the key functions of MeCP2-NCoR. This study relied primarily on gross organismal phenotypes for its interpretations. Studies directly testing changes in histone acetylation in similar mutants can confirm that deacetylation by NCoR is not the direct activity needed for MeCP2’s repressive function (Figure 3C). In Chapter 3 of this dissertation, I present initial results probing the mechanism by which MeCP2 and NCoR regulate enhancers. These findings suggest that while the interaction between MeCP2 and NCoR is critical for gene regulation by MeCP2, NCoR binding to the genome does not depend on MeCP2. In addition, these results suggest that MeCP2-NCoR act upstream of histone acetylation in the process of enhancer activation and drive downstream consequences of enhancer function. I address ongoing studies and future experiments that can further elucidate the precise mechanism of enhancer regulation by MeCP2.

A challenge for researchers going forward is to integrate findings on protein-protein interactions and chromatin modifying activity of MeCP2-NCoR with potential structural or biophysical roles of MeCP2 in chromatin that are re-emerging. Historic findings have shown that MeCP2 localizes to regions of heterochromatin and drives nucleosome aggregation in vitro (Agarwal et al., 2007; Brero et al., 2005; Nan et al., 1996). Recent studies have shed new light on this activity by showing that MeCP2 can undergo liquid-liquid phase separation, forming condensates with nucleosomal DNA in vitro (Fan et al., 2020; Wang et al., 2020). Rett syndrome causing mutations of MECP2
reduced condensate forming activity, suggesting that MeCP2 might affect compartmentalization of heterochromatic droplets in cells and that disruption of this activity may be indicative of MECP2 inactivation in disease. High resolution nuclear imaging in MeCP2 knockout mouse has also observed altered heterochromatin volume in neurons lacking MeCP2, suggesting altered chromatin condensation (Linhoff et al., 2015). However, these findings have not yet been linked to the epigenomic and transcriptomic consequences of mCH-MeCP2 disruption. Recent analyses of chromatin looping by 3C and Hi-C detected no dramatic changes in chromosome topology in Mecp2 knockout mouse brain tissue (Boxer et al., 2020; Clemens et al., 2020). The lack of changes in chromosome topology does not support changes in nuclear compartmentalization occurring upon loss of MeCP2, as might be predicted by the phase-separation experiments. However, a study of human ES-cell-derived interneurons carrying the MECP2 R133C Rett syndrome mutation detected altered global topology (Xiang et al., 2020). Notably, these condensation and human topology studies analyzed MECP2 truncation and missense mutants, while the mouse studies focused on complete loss of the protein. Different MECP2 mutations result in differential clinical severity (Lyst et al., 2013; Tillotson and Bird, 2019), and these differences may account for contrasting effects across studies. Additional analysis will be needed to determine how condensation characteristics of MeCP2 may impact genome topology and gene expression to manifest cellular dysfunction when disrupted.

Future studies can build on these recent findings to further decode the mechanism of gene regulation by the mC-MeCP2-NCoR pathway. For example, if epigenomic profiling experiments detect alterations in enhancer activation that occur upstream of histone acetylation at highly methylated enhancers upon loss of MeCP2 (e.g. transcription factor binding, nucleosome removal,
H3K4me1 deposition), it would suggest that MeCP2-NCoR indeed acts outside of histone acetylation in controlling chromatin structure. The role of NCoR in enhancer regulation through MeCP2 is further discussed in Chapter 3.

1.7 Biological function of mCH-MeCP2-mediated gene regulation.

In addition to these mechanistic insights, recent studies have shed light on the functional impact of mCH-MeCP2-mediated gene regulation. The highly cell-type-specific nature of mCH profiles suggests an important role for MeCP2 is to mediate neuron subtype-specific gene expression. In support of this, an initial study of hand-sorted cell populations from MeCP2 knockout mouse brains detected larger changes in mRNA levels in purified cell types compared to whole tissue and identified distinct sets of dysregulated genes in different cell types (Sugino et al., 2014). Recent integrated analyses of mCH profiles and RNA changes in isolated cell populations (Johnson et al., 2017; Lavery et al., 2020; Stroud et al., 2017) and single cells (Renthal et al., 2018; Stroud et al., 2017) from MeCP2 and Dnmt3a knockout, conditional knockout, and missense mouse mutants and humans with Rett syndrome have identified cell-type-specific de-repression of genes enriched for mCH. Notably, the genes most de-repressed in cell-specific and tissue-based studies of mCH and MeCP2 disruption tend to be long genes that encode protein with important roles in the establishment and maintenance of synaptic connectivity (e.g. cell-adhesion molecules, ion channels, and synaptic receptors) (Boxer et al., 2020; Clemens et al., 2020; Gabel et al., 2015; Lagger et al., 2017; Renthal et al., 2018; Sugino et al., 2014). Thus, mCH and MeCP2 appear to regulate genes with essential roles in establishing connectivity in a cell-type specific fashion.
These recent findings, coupled with the knowledge that MeCP2 reads postnatal mCH patterns to repress transcription in the adult, suggest a functional model: During development when MeCP2 expression is rising and mCH is being deposited in each neuronal subtype, genes and enhancers that are robustly expressed escape accumulation of mCH and subsequent repression by MeCP2. Conversely, genes that are lowly expressed accumulate mCH and MeCP2, which in turn maintains them in a repressed state later in life (Stroud et al., 2017). The absence of mCH-MeCP2 repression in the early postnatal period may allow for flexible expression of critical protein components of synaptic connectivity, as cells respond to extrinsic cues and integrate into circuits. Build-up of mCH and MeCP2 occurs primarily during the closure of postnatal hyperplastic periods, which could then maintain these gene expression patterns to allow for consolidation and refinement of cellular functions in the circuit. Thus, mCH-MeCP2 repression might effectively close an “epigenomic critical period” of plastic gene expression to stabilize functional circuits, in much the same way that build-up of extracellular matrix closes critical periods of plastic connectivity in the brain during this same period (Picard and Fagiolini, 2019). Once global patterns of mCH are established, stimulus-dependent inactivation of MeCP2-NCoR-mediated repression that results from MeCP2 phosphorylation (Yap and Greenberg, 2018) and activity-dependent alterations in DNA methylation (Bayraktar and Kreutz, 2018) could facilitate more limited, but important, dynamic gene expression during adult plasticity. In support of this role for mCH and MeCP2 in brain function, loss of MeCP2 disrupts critical period timing and synaptic plasticity (Ip et al., 2018; Krishnan et al., 2015; Patrizi et al., 2019; Picard and Fagiolini, 2019).
1.8 Expanding roles for mCH and MeCP2 in disease.

For over a decade, disruption of MeCP2 due to loss-of-function mutations or overexpression has been recognized as the cause of Rett syndrome and **MeCP2 duplication syndrome**, respectively (reviewed in (Lavery and Zoghbi, 2019)). The fact that either too much or too little MeCP2 manifests in severe neurologic dysfunction suggests that circuits require precise, dose-sensitive tuning of gene regulation by the mC-MeCP2-NCoR pathway (Lavery and Zoghbi, 2019; Tillotson and Bird, 2019) and raises the possibility that this pathway may be susceptible to additional insults in disease (Figure 1C). Indeed, exome sequencing studies have recently uncovered mutations in individuals with intellectual disability, autism, and related disorders that disrupt this pathway up- and down-stream of MeCP2. Heterozygous mutations in **DNMT3A** cause **Tatton-Brown-Rahman syndrome** and autism (Sanders et al., 2015; Satterstrom et al., 2020; Tatton-Brown et al., 2014). Strikingly, **Dnmt3a** heterozygous knockout mice that model this disorder exhibit ~50% global reductions in mCH across multiple brain regions (Christian et al., 2020; Sendžikaitė et al., 2019). This reduction in mCH drives alterations in enhancer histone acetylation and gene expression in the cerebral cortex that partially recapitulate MeCP2 loss of function (Christian et al., 2020). These findings indicate that mCH deposition and its role in neuronal regulation are highly sensitive to reduction in DNMT3A protein (Figure 3D).

Mutations of a component in the NCoR complex, **TBL1XR1**, have also emerged as causal for neurodevelopmental disease (Coe et al., 2019; Firth et al., 2009; Satterstrom et al., 2020). Notably, some missense mutations identified in **TBL1XR1** have been shown to specifically disrupt the NCoR-MeCP2 interaction, and a patient with a missense mutation in this interacting domain has been diagnosed with Rett syndrome, based on clinical criteria (Kruusvee et al., 2017; Zaghlula et
al., 2018). In further support of overlapping pathology arising from an absence of NCoR, loss-of-function mutations of the Deacetylase Activating Domain of NCOR1 within GABAergic neurons resulted in cognitive-, social-, and anxiety-related phenotypes in mice that have some similarities with MeCP2 knockout models (Zhou et al., 2019).

It is important to note that mutations in DNMT3A and NCOR1 also manifest multiple distinct clinical features from those of Rett syndrome (e.g. overgrowth, heart defects, joint hypermobility). Undoubtedly, these factors have gene regulatory roles early in development and outside of the mC-MeCP2-NCoR axis that contribute to the non-overlapping aspects of these disorders (Tatton-Brown et al., 2018; Zaghlula et al., 2018) (Figure 3E). The precise degree of overlapping molecular etiology between these disorders may have implications for the development of treatments. A striking feature of Mecp2 mutations is that reintroduction of exogenous MeCP2 in adults can dramatically reverse symptoms in mice (Guy et al., 2007), likely because restored MeCP2 reads out mCH and mCG patterns that were appropriately laid down during postnatal development. This finding has fueled development of gene therapies for Rett syndrome (Gadalla et al., 2013). If DNMT3A mutations cause deficits in DNA methylation during critical temporal windows of embryonic and postnatal development, these effects may not be as reversible as the absence of MeCP2. Likewise, changes in cellular compositions of the brain or structural changes that occur due to loss of early roles for NCoR or DNMT3A may be difficult to reverse later in life. Nonetheless, disruption of epigenomic regulation through the mC-MeCP2-NCoR axis in adults is likely a shared deficit that contributes to neurologic dysfunction in these disorders and may be a viable candidate to explore therapeutic approaches.
1.9 Figures

1.9.1 Figure 1. Non-CG methylation and MeCP2 in neuronal gene regulation.

A. Non-CG DNA methylation is enriched in neurons compared to glia and other cell types and can show substantial variations in global levels across neuronal subtypes (Lister et al., 2013; Luo et al., 2017; Mo et al., 2015).
B. Summary of mCH profiles detected at genes most impacted by mCH and MeCP2 mediated gene regulation. Two example “meta genes” and TADs depict local depletion of mCH within gene bodies and enhancer sequences. MeCP2-repressed genes (red) are enriched for mCH within their gene bodies, at associated enhancers, and throughout the TAD they are located within, while MeCP2-activated genes (blue) and enhancers associated with them are in regions of mCH depletion (Boxer et al., 2020; Clemens et al., 2020; Stroud et al., 2017).

C. Neuronal gene regulation by the mC-MeCP2-NCoR axis is impacted by disease-associated mutations at multiple levels: mCH deposition, MeCP2 expression, and NCoR complex components. Dysregulation of enhancers and transcriptional activity resulting from these mutations may contribute to disease pathology. The susceptibility of this pathway to disruption suggests that it may be affected in additional neurodevelopmental disorders caused by mutations of epigenetic regulatory genes. (ASD – Autism Spectrum Disorder; ID – Intellectual Disability; RTT- Rett Syndrome; TBRS – Tatton-Brown Rahman Syndrome).
1.9.2 Figure 2. The unique neuronal epigenome.

A. Summary of reported levels of DNA modification sites and MeCP2 molecules between neurons and non-neural cell types, illustrating that mCH, hmCG, and MeCP2 are uniquely enriched within neurons.

B. During postnatal neuronal maturation, non-CG methylation, hydroxymethylation, and MeCP2 build up, reaching high levels in mature neurons at the young adult stage.

C. Left, a model depicting the early postnatal deposition of non-CG methylation by DNMT3A, establishing megabase-scale domains of high and low methylation associated with TADs. Right, a model illustrating how high levels of gene expression block DNMT3A activity, as highly expressed genes exhibit less DNMT3A binding and mCH accumulation compared to lowly expressed genes.

D. Illustration of typical methylation profiles in adult brain tissue. Both mCG and mCH are depleted at promoters and enhancers, but only mCH exhibits robust gene body demethylation and megabase-scale variations. H3K27ac and H3K36me3 histone modifications illustrate the relationship between DNA methylation and regulatory elements and transcriptional activity.

E. Illustration of differential global levels of mCH detected in subtypes of neurons. A layer V excitatory neuron shows substantially more mCH than a layer IV excitatory neuron (Luo et al., 2017). These differences in methylation, combined with differential methylation of genes, set the stage for cell-type specific repressive effects by MeCP2 (e.g. Gene A is more strongly expressed in layer V neurons than in layer IV neurons and Gene B more so in layer IV than in layer V).
1.9.3 Figure 3. Molecular mechanisms of the mC-MeCP2-NCoR axis and its disruption in neurodevelopmental disorders.

A. Top, a model depicting that MeCP2 binds to mCH deposited by DNMT3A. Bottom, upon complete or conditional knockout of Dnmt3a, mCH binding sites for MeCP2 are lost, but binding of MeCP2 to mCG persists.

B. Top, model of MeCP2 binding to mC at enhancers and gene bodies to reduce the acetylation of enhancers, genes, and promoters, resulting in a reduction of transcription initiation. Bottom, upon MeCP2 knockout, restriction on acetylation and transcription are reduced.

C. Left, the MeCP2-NCoR complex bound to mC represses enhancer acetylation. This may occur directly through the HDAC component of the complex or indirectly through other undefined activities of NCoR. Right, loss of MeCP2 can lead to loss of NCoR recruitment to the genome, loss of co-repressor activities that result in reduced acetylation, or potentially a mechanism yet to be defined.

D. Schematic depicting a spectrum of disruption for MeCP2 repressive effects at enhancers across neurodevelopmental disorders; ranging from hyper repression in MeCP2 duplication, to intermediate disruption of repression occurring when heterozygous loss of DNMT3A leads to global reduction in mCH, to complete loss of repression when MeCP2 is knocked out.

E. Venn diagram illustrating mechanistic and phenotypic overlaps of molecular pathologies of neurological disorders. Disruption of enhancer repression by the mC-MeCP2-NCoR axis may be shared across these disorders and contribute to pathology. Notably, each genetic lesion results in unique phenotypes that are likely to drive loss of molecular functions outside of the overlapping mC-MeCP2-NCoR axis. (ASD – Autism Spectrum Disorder; ID – Intellectual
Disability; MDS – MeCP2 Duplication Syndrome; RTT- Rett Syndrome; TBRS – Tatton-Brown Rahman Syndrome)
Approximate levels of CG and CH dinucleotides and their levels of modifications across cell types. Numbers of mC sites are estimated based on measurements made in (Greenberg and Bourc’his, 2019; Hon et al., 2014; Kozlenkov et al., 2018; Kriaucionis and Heintz, 2009; Lister et al., 2013; Luo et al., 2017; Mellén et al., 2017; Mo et al., 2015; Sardina et al., 2018). mC vs hmC levels are inferred by combining results quantifying all modified mC (using bisulfite-sequencing) with studies using hmC-sensitive detection methods (TAB-seq, OxBS). Values are based on the mouse genome, but similar numbers occur in the human genome (Kent et al., 2002).
1.10 Glossary

**3C & Hi-C**: Chromatin Conformation Capture methods employing crosslinking and proximity ligation followed by PCR (3C) or high-throughput sequencing (Hi-C) to map 3D-interactions and the architecture of genome folding within the nucleus.

**Enhancer**: Regulatory element found outside of promoters that recruits transcription factors and co-activators and interacts with promoters to drive transcription.

**Gene body**: Region of the gene from transcription start site (TSS) to transcription end site (TES) that is transcribed during pre-mRNA production.

**Global run-on sequencing (GRO-seq)**: Method that utilizes co-transcriptional labeled nucleotide incorporation followed by sequencing to directly quantify gene transcription.

**Histone modifications**: Denoted by histone number, amino acid, and modification (e.g. monomethylation, me1; trimethylation, me3; acetylation, ac). Associated with steps in transcriptional regulation.

- **H3K4me1**: Present at active enhancers.
- **H3K4me3**: Present at active/bivalent promoters and enhancers.
- **H3K27ac**: Present at active promoters and enhancers.
- **H3K36me2**: Precursor to H3K36me3; marks open euchromatic regions.
- **H3K36me3**: Marks the 3’ portion of the gene body for actively transcribed genes.
**mCA set-point:** Hypothesis that levels of mCA are consistent across discrete mega-base scale domains of the genome, but can vary from domain to domain. mCA set-points appear to influence the levels of mCA at kilobase-scale sequence elements such as enhancers and gene bodies.

**MeCP2 duplication syndrome:** Neurological disorder caused by genetic duplication of *MECP2*. Causes developmental delay among other severe clinical features.

**MeCP2-repressed genes:** Genes significantly upregulated when MeCP2 is inactivated, and downregulated when MeCP2 is overexpressed.

**MeCP2-activated genes:** Genes significantly downregulated when MeCP2 is inactivated, and upregulated when MeCP2 is overexpressed.

**Methyl-binding domain (MBD):** Conserved protein domain that binds specifically to methylated cytosine. Found in MeCP2 as well as MBD1/2/3/4.

**Nuclear Co-Repessor complex (NCoR):** A complex comprised of NCoR1 and/or its paralog SMRT, with TBL1, HDAC3, and GPS2 as its core components. It is recruited to the genome by multiple DNA binding proteins, including MeCP2.
**Rett syndrome (RTT):** X-linked recessive neurological disorder predominantly occurring in females that is caused by loss of function *MECP2* mutations. Typified by phenotypically normal early development, followed by regression and decline in brain growth.

**Tatton-Brown Rahman syndrome (TBRS):** Autosomal dominant neurodevelopmental disorder characterized by overgrowth, intellectual disability, and autism. Caused by heterozygous mutations of *DNMT3A*.

**Topologically Associating Domains (TADs):** Linear regions of the genome that show enrichments for interactions in three-dimensional space. Can facilitate and/or result from contacts between enhancers and promoters that drive gene expression.

**Whole-genome bisulfite-sequencing (WGBS):** Base-resolution DNA methylation profiling method whereby methylated cytosines are protected from sodium bisulfite conversion to uracil (read as thymidine), thereby distinguishable from unmethylated cytosines during whole genome sequencing.
Endnote 1: Neuron-enriched DNA modifications and their enzymes.

DNA methylation in mammals is deposited at unmethylated CG and non-CG sites (CH) by the de novo methyltransferases DNMT3A and DNMT3B. Additionally, DNMT1 methylates DNA at existing hemi-methylated CG sites to maintain symmetric CG methylation after new strand synthesis, or to create a fully methylated CG site after de novo methylation of one strand (Gowher and Jeltsch, 2018; Hermann et al., 2004). The ten-eleven translocation enzymes (TET 1,2,3) oxidize DNA, creating 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fcC), or 5-carboxylcytosine (caC). These oxidized forms of DNA can drive active demethylation through the base excision repair pathway. However, hmC primarily leads to passive demethylation (dilution) in mitotic cells upon DNA replication by blocking DNMT1 from maintaining hemi-methylated CG dinucleotides (Wu and Zhang, 2017).

In all cells, CG sites are highly methylated due to early activity of de novo methyltransferases and subsequent active maintenance of mCG by DNMT1 (Figure 4) (Hermann et al., 2004). However, both mCH and hmC are very low in most dividing cells. This is likely due to both low expression of the DNMT3A/B (Okano et al., 1998) and TET (Wu and Zhang, 2017) enzymes and the lack of an efficient mechanism to maintain mCH and hmC after DNA replication. In contrast, it appears that increased expression of TET (Szulwach et al., 2011) and DNMT3A enzymes (Guo et al., 2014; Lister et al., 2013), in conjunction with a lack of DNA replication, leads to accumulation of hmC and mCH in neurons. Substantial evidence now suggests that hmC is a stable epigenetic mark in
neurons, reaching high levels compared to non-neural cell types (Figure 4) (Globisch et al., 2010; Kriaucionis and Heintz, 2009; Lister et al., 2013; Wu and Zhang, 2017). Therefore, while levels of modified CG sites are maintained in neurons, TET enzymes drive large scale conversion of mCG into hmCG (e.g. total modified CG = mCG + hmCG, Figure 4). Additionally, there appears to be limited turnover of mCH in neurons due to low hmC conversion or active demethylation at these sites.

More efficient methylation at CG sites compared to CH sites leads to a substantially higher percent of mCG compared to mCH (Figure 4). However, the depletion of CG dinucleotides from the genome resulting from mutation of methylated cytosine to thymine over evolution (Bird, 1980) results in nearly equivalent numbers of mC events occurring at CG and non-CG sites in neuronal genomes (Figure 4) (Lister et al., 2013). Together, the high, stable levels of hmC and mCH create a unique environment for epigenetic regulation in neurons that can affect the binding of regulatory factors to DNA and impact gene expression.

**Endnote 2: Global, graded gene regulation by mC and MeCP2.**

Unlike transcription factors, where sparse genomic binding sites can be linked to “target” genes, MeCP2 binds to millions of mCH and mCG sites, which are present in varying amounts at every gene and regulatory element. This genome-wide binding suggests that MeCP2 influences transcription of all genes to some extent. Indeed, genome-wide analysis of RNA changes in MeCP2 knockout and missense mutants, MECP2 over expressing mice, and brain-specific Dnmt3a
conditional knockout mice have detected trends in which the degree of gene dysregulation is proportional to the number of mCH and mCG sites found in the gene body (Boxer et al., 2020; Clemens et al., 2020; Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017; Lavery et al., 2020; Renthal et al., 2018; Stroud et al., 2017). In addition, genome-wide upregulation of expression associated with gene body length has been detected upon loss of MeCP2 (Gabel et al., 2015; Johnson et al., 2017; Kinde et al., 2016; Lagger et al., 2017; Sugino et al., 2014), providing the first clues that intragenic binding by MeCP2 is an important aspect of its regulatory mechanism (Gabel et al., 2015; Stroud et al., 2017). Care must be taken to ensure that apparent genome-wide effects do not result from technical noise (Raman et al., 2018), but studies have now verified these trends in large-replicate datasets, using multiple RNA quantification methods (Boxer et al., 2020; Cholewa-Waclaw et al., 2019; Clemens et al., 2020; Renthal et al., 2018). Notably, the investigation of these trends has led to the generation of new molecular models for gene regulation by mCH, mCG, and MeCP2 (Boxer et al., 2020; Cholewa-Waclaw et al., 2019; Clemens et al., 2020; Renthal et al., 2018; Stroud et al., 2017).

While these global trends provide mechanistic clues, limitations of gene expression analysis also present challenges for interpretation of transcriptomic data. For instance, normalization procedures used to quantify relative RNA levels between samples assume that no global changes in the distribution of gene expression values occur across conditions (Love et al., 2014). As a result, genome-wide fold-changes will effectively be re-centered around zero during data processing, and this can possibly switch the sign of perceived changes in gene expression. Similar effects can occur in RT-qPCR experiments, where total RNA and house-keeping gene normalization are employed. Thus, in one plausible model, loss of repression in the Mecp2 knockout leads to upregulation of
nearly all genes in the genome, with the most highly methylated regions being most de-repressed. However, upon normalization, genes that are lowly methylated and the least de-repressed are re-centered below zero and quantified as being down-regulated. In such a paradigm, a substantial portion of MeCP2-activated genes that appear to decrease in expression when MeCP2 is lost may in fact represent genes that normally escape repression by MeCP2 rather than being genes that are directly activated by the protein.

In addition to these issues, standard transcriptomic approaches do not detect changes in global RNA levels per cell, and several reports indicate that *MeCP2-null* neurons, which are reduced in size, contain less total RNA than normal cells (Lagger et al., 2017; Li et al., 2013). In all, relative quantification methods can combine with secondary effects on RNA levels and gene expression, as well as disruption of potential direct gene-activating functions for MeCP2 (Ip et al., 2018), to result in the overall changes in gene expression observed in MeCP2 mutants. These complexities highlight the need to integrate transcriptomic findings with biochemical insights into the direct function of MeCP2 to build accurate models of transcriptional regulation by DNA methylation and MeCP2.
Chapter 2: MeCP2 represses enhancers through chromosome topology-associated DNA methylation

This chapter is adapted from a manuscript published in Molecular Cell.


AWC carried out all of the experiments in collaboration with JRM and DLC and performed ChIP-seq analysis; DYW performed RNA-seq, Bisulfite-seq, and TAD analysis; GZ and D.Y.W. performed the Hi-C analysis; AWC, DYW, and HWG designed the experiments and analysis, and wrote the manuscript.
2.1 Introduction

The development and function of the mammalian brain requires precise control of gene expression to specify neuronal sub-types, form and refine circuits, and respond to dynamic changes in neuronal activity (Cholewa-Waclaw et al., 2016; Ebert and Greenberg, 2013; Molyneaux et al., 2007; Ziats et al., 2015). Emerging evidence indicates that neurons utilize a unique form of DNA methylation to regulate these transcriptional programs, with multiple types of CNS neurons displaying high levels of cytosine methylation in a non-CG context compared to non-neural cells (Guo et al., 2014; Lister et al., 2013; Luo et al., 2017; Mo et al., 2015; Xie et al., 2012). This non-CG DNA methylation is deposited by the DNA methyltransferase, DNMT3A, and accumulates specifically in neurons during the early postnatal period, until the number of methylated cytosines in the non-CG context is nearly equivalent to that of canonical methyl-CG sites (mCG) (Guo et al., 2014; Kinde et al., 2015; Lister et al., 2013). Disruption of DNMT3A specifically in mouse brain completely blocks the accumulation of non-CG methylation, leading to de-repression of genes that normally contain high levels of non-CG methylation and resulting in severe neurological phenotypes in these mice (Gabel et al., 2015; Nguyen et al., 2007; Stroud et al., 2017). These findings underscore the importance of non-CG methylation in neural development, motivating efforts to understand how this methylation works together with mCG to control gene expression in neurons.

Recent studies have demonstrated that Methyl-CpG binding Protein 2 (MeCP2) binds tightly to non-CG methylation, showing a high affinity for methylated cytosines within CA dinucleotides (mCA) that is similar to its affinity for its classically defined substrate, mCG (Chen et al., 2015; Gabel et al., 2015; Guo et al., 2014; Kinde et al., 2015; Lagger et al., 2017). Inactivation of MeCP2
causes the severe neurological disorder Rett syndrome, and duplication of MeCP2 leads to autism spectrum disorder (Amir et al., 1999; Chahrour and Zoghbi, 2007; Van Esch et al., 2005; Hagberg et al., 1983). While MeCP2 has been proposed to be both a repressor and an activator of transcription, biochemical analyses have identified a critical interaction between MeCP2 and the NCoR-HDAC3 transcriptional corepressor complex, and in vitro experiments have demonstrated that MeCP2 can repress transcription of reporter plasmids. Together these findings indicate that MeCP2 plays a critical role in nervous system function, and suggest that one important role for MeCP2 is to bind to DNA methylation in neurons and repress gene expression (Ip et al., 2018; Lyst and Bird, 2015; Lyst et al., 2013).

Because of its important role in brain function, identifying the genes that MeCP2 directly regulates in neurons has been an area of intense investigation. Genomic analyses have revealed that MeCP2 is expressed at near histone levels and binds extremely broadly across the neuronal genome. In the context of this broad binding, enrichment of MeCP2 can be detected in regions with high numbers of mCG and mCA sites (Chen et al., 2015; Cohen et al., 2011a; Gabel et al., 2015; Kinde et al., 2015; Lagger et al., 2017; Skene et al., 2010). Studies of gene expression changes in brain tissue from human Rett syndrome patients, MeCP2 knockout mice (MeCP2 KO), and MeCP2 overexpression mice (MeCP2 OE) detect only subtle changes across many genes upon MeCP2 disruption, making it difficult to differentiate direct targets of MeCP2 regulation from secondary effects on gene expression (Baker et al., 2013; Ben-Shachar et al., 2009; Chahrour et al., 2008; Gabel et al., 2015; Kinde et al., 2016; Samaco et al., 2012; Tudor et al., 2002; Zhao et al., 2013). Defining the mechanism by which MeCP2 mediates its broad but subtle effects on gene expression
has therefore been recognized as an important challenge (Ip et al., 2018; Lyst and Bird, 2015), and despite continued studies, has remained difficult to decipher.

Recent integrated analysis of gene expression and DNA methylation has revealed enrichment of methylation at CA dinucleotide sites (mCA/CA) in and around “MeCP2-repressed” genes--those genes most significantly upregulated upon loss of MeCP2 and downregulated upon overexpression of MeCP2 (Chen et al., 2015; Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017). Notably, these genes are not substantially enriched for mCA at promoters, a canonical site of action for DNA methylation, but instead show high mCA within the transcribed region and flanking sequences of the gene. Furthermore, genome-wide analysis has revealed that loss of MeCP2 leads to a graded upregulation of genes that is proportional to both the level of DNA methylation within the transcribed region of the gene (the “gene body”) and to the length of the pre-mRNA transcript. Thus, genes that are very long (e.g. >100kb) and contain high numbers of methylation sites are upregulated relative to shorter genes that contain fewer methylation sites (Gabel et al., 2015; Kinde et al., 2015, 2016; Lagger et al., 2017; Sugino et al., 2014). This association with gene length, and the observation that upregulation of gene expression upon loss of MeCP2 is best correlated with the DNA methylation within the gene, has led to the proposal that MeCP2 regulates expression by binding to methylated DNA in gene bodies and repressing transcription (Kinde et al., 2016).

While these findings have provided insight into the role of mCA, mCG, and MeCP2 in the brain, they raise new questions. For example, high levels of mCA are detected at regions extending megabases up- and downstream of MeCP2-repressed genes (Kinde et al., 2016; Lagger et al., 2017), but it has not been determined what delineates the large regions of mCA enrichment in which these genes are found. In addition, while multiple studies now support a role for MeCP2 in
repressing long, highly-methylated genes, it has not been demonstrated that MeCP2 directly affects transcription when bound to mCG and mCA within these genes, and alternative post-transcriptional mechanisms have been proposed (Johnson et al., 2017). Finally, the precise mechanism by which binding of MeCP2 to DNA methylation outside of promoter regions results in subtle, but functionally critical gene repression has not been determined.

In this study, we have investigated how high levels of CA methylation are established at genes that are repressed by MeCP2 and examined how MeCP2 functions with the methylation at these loci to control gene transcription. Through integrated genomic analysis we identified an association between chromatin folding in the nucleus and the patterning of mCA across the neuronal genome, and we show that genes repressed by MeCP2 land within topologically-associating domains (TADs) enriched for mCA. We further investigated how the high levels of TAD-associated mCA are read out by MeCP2 to control gene transcription, uncovering new evidence that binding of MeCP2 to mCA and mCG at enhancer elements controls the activity of these regulatory sequences. We find that MeCP2-repressed enhancers tend to be located within the same highly methylated TADs as MeCP2-repressed genes, and often reside within these genes. Thus, dysregulation of enhancers can contribute to altered transcription that occurs upon mutation of MeCP2. Together, our study defines a mechanism that shapes the deposition of neuron-enriched mCA in the genome and uncovers a new role for mCA and MeCP2 in the regulation of enhancers. These findings provide insight into how disruption of MeCP2 causes altered gene expression to drive neurological dysfunction in Rett syndrome and related disorders.
2.2 Results

To probe how high levels of mCA are established at MeCP2-repressed genes and to determine how DNA methylation is read out by MeCP2 to control gene expression, we carried out integrated genomic analysis in the cerebral cortex, a brain region in which loss of MeCP2 has been shown to affect gene expression, alter neuronal morphology, and disrupt physiology (Gabel et al., 2015; Kinde et al., 2016; Kishi and Macklis, 2004, 2010; Pacheco et al., 2017; Sceniak et al., 2016; Shepherd and Katz, 2011). By studying the cortex, we were able to integrate our genomic datasets with a compendium of data on gene expression, DNA methylation, chromatin structure, and genome topology for this brain region compiled across multiple studies (Dixon et al., 2012; ENCODE Consortium, 2012; Mo et al., 2015; Shen et al., 2012; Sloan et al., 2016; Stroud et al., 2017). This allowed us to explore features of the genome and chromatin structure that are associated with high mCA levels at MeCP2-repressed genes. While we focused on patterns of mCA, the most prevalent and highest affinity non-CG methylation site for MeCP2 binding (Gabel et al., 2015; Lagger et al., 2017), we obtained similar results when we included minor forms of non-CG methylation, mCT and mCC, in analysis throughout our studies.

To identify the genes most robustly dysregulated upon disruption of MeCP2 in the cortex, we used combined RNA-seq analysis of MeCP2 KO and MeCP2 OE mice, an approach that has been powerful for detecting MeCP2-regulated genes in other brain regions (Ben-Shachar et al., 2009; Chahrour et al., 2008; Chen et al., 2015). Differential mRNA expression analysis of cortex from six MeCP2 KO-wild-type and five MeCP2 OE-wild-type littermate pairs identified 884 “MeCP2-repressed genes” that are significantly upregulated upon loss of MeCP2 and downregulated upon overexpression of MeCP2, and 843 “MeCP2-activated genes” that are significantly downregulated...
upon loss of MeCP2 and upregulated upon overexpression of MeCP2 (Figure 1A; Table S1). These genes significantly overlap with genes identified through meta-analysis of MeCP2 mutant gene expression studies across multiple non-cortical brain regions (Gabel et al., 2015), indicating that a substantial number of MeCP2-regulated genes are shared across functionally distinct brain regions made up of diverse neuronal subtypes (Figure S1A,B).

We then assessed the characteristics of MeCP2-regulated genes, including gene length and DNA methylation levels as measured by whole-genome bisulfite sequencing (Stroud et al., 2017, see methods). In agreement with previous studies (Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017; Renthal et al., 2018; Stroud et al., 2017; Sugino et al., 2014), MeCP2-repressed genes in the cortex are significantly longer than the genome average (Figure S1D). These genes are also enriched for CA methylation (mCA/CA, referred to as “mCA level”) in the gene body and surrounding the gene, but show little to no mCA enrichment at the transcriptional start site (TSS) (Figure 1B,C; S1C). As previously observed, we find a global gene-length- and mCA-associated relative upregulation of genes in MeCP2 KO cortex and a reciprocal relative downregulation in the MeCP2 OE cortex (Figure S1E,F). Analysis of expression changes in all genes showed that the degree of dysregulation in MeCP2 KO and MeCP2 OE cortex genome-wide was more highly correlated with gene body DNA methylation than it was with gene-flanking regions (Figure S1G,H). We did not detect evidence of enriched mCG levels in and around MeCP2-repressed genes, and MeCP2-activated genes displayed limited differences in gene length and mCA or mCG profiles compared to the genome average or control sets of unchanged genes (Figure 1B; S1C). These results validate conclusions from previous studies that MeCP2 preferentially represses long genes that contain high levels of methylation within their gene body (Gabel et al., 2015; Kinde et
al., 2016; Lagger et al., 2017; Renthal et al., 2018; Stroud et al., 2017; Sugino et al., 2014) and provide a high confidence set of MeCP2-regulated genes in the cerebral cortex to facilitate our further analyses.

2.2.1 Large-scale non-CG DNA methylation profiles are linked to chromatin topology.

Having established that MeCP2-repressed genes in the cerebral cortex are enriched for mCA in and around the gene, we explored how patterns of high mCA may be established at these loci. Broad-scale analysis of mCA around MeCP2-repressed genes revealed that, like gene-sets identified in other brain regions (Kinde et al., 2016; Lagger et al., 2017), enrichment of this methylation can extend megabases away from the gene (Figure 1B,C). This suggested that high regional methylation might be associated with chromosomal features that vary at the hundred-kilobase to megabase scale. We therefore searched the compendium of datasets for transcription factor binding, chromatin structure, and chromosome topology in the cerebral cortex (Dixon et al., 2012; ENCODE Consortium, 2012; Lister et al., 2013; Shen et al., 2012; Stroud et al., 2017) looking for sequence features or epigenetic marks that correlate with the large-scale patterns of mCA observed at MeCP2-repressed genes and across the genome. Examination of high-throughput chromatin conformation capture (Hi-C) interaction maps of nuclear topology generated for the mouse cerebral cortex (Dixon et al., 2012, 2016) revealed notable instances in which a region of high mCA surrounding MeCP2-repressed genes dropped off at transition points in Hi-C interactions, similar to those that delineate topologically associating domains (TADs) (Figure 1D) (Dixon et al., 2012, 2016).
TADs are defined in Hi-C analysis as stretches of the genome that are enriched for cis interactions compared to background (Dixon et al., 2012). TAD structures delineate interactions between cis-regulatory enhancer elements and genes, such that enhancer elements within a TAD are more likely to interact with promoters of genes within the same TAD than they are to interact with promoters of genes outside of the TAD (Spielmann et al., 2018). A correlation between mCA patterns and TAD structure would implicate genome folding in the deposition of this unique methylation, and could have important implications for regulation of gene expression through methylation at promoters and enhancer elements within each TAD. We therefore computationally defined TAD structures in the genome of adult mouse cerebral cortex (Dixon et al., 2012) using the TADtree algorithm (Weinreb and Raphael, 2016) and systematically examined DNA methylation in regions containing MeCP2-repressed genes (Table S2). TADs defined in the cerebral cortex by this method are ~400kb on average and range in size from ~160kb to ~2mb, a scale that is similar to regions that we observed with enriched mCA levels.

Analysis of mCA levels across TADs revealed that MeCP2-repressed genes are found within TADs that are significantly enriched for mCA compared to TADs genome-wide (Figure 1E). We observed a significant drop-off of mCA levels at the boundaries of TADs containing MeCP2-repressed genes that was not observed for control TADs that were generated by shuffling TADs across the genome (Figure 1F; S2B,C). While the shift in methylation across all of these boundaries was a return toward the genome average (Figure 1F), more dramatic step-changes in methylation occurring specifically at the boundary could be detected in genomic regions where a high-mCA TAD containing an MeCP2-repressed gene was juxtaposed with a genomic region with substantially lower mCA levels (Figure 1G). Such strong enrichments in mCA and discrete step-
changes in signal at boundaries were not observed when TAD locations were shuffled and these resampled boundaries were examined (Figure 1F,G; S2B, see methods). In addition, similar signals could not be detected in coverage rates for Bisulfite-sequencing or GC composition of the sequences on either side of TAD boundaries (Figure S2B), indicating that the patterns we observed were specific to TAD boundaries and are not likely to be driven by sequencing biases.

To further test if TAD structures impact the patterns of mCA at MeCP2-repressed genes and throughout the genome, we performed genome-wide, cross-correlation analysis of mCA levels within and across TAD boundaries using a procedure similar to analysis carried out for histone modifications (Rao et al., 2014, see methods). We found that the mCA levels for genomic regions located within the same TAD are more highly correlated with each other than they are with regions in neighboring TADs (Figure 1H). This relationship was prominent for high-mCA TADs that contain MeCP2-repressed genes, but also robustly detectable when examining all TADs in the genome. Similar signals of correlation within the TAD were not detected when analyzing shuffled, control TADs randomly placed throughout the genome (Figure 1H, see methods). These findings suggest that TADs are units of organization for mCA levels across the genome and that MeCP2-repressed genes often occupy TADs with high levels of mCA.

In contrast to MeCP2-repressed genes, MeCP2-activated genes reside in TADs that have lower mCA levels compared to the genome average (Figure 1E; S2A). This finding is consistent with the limited enrichment of mCA in gene-flanking sequences that we and others observe for these genes (Kinde et al., 2016; Lagger et al., 2017). mCG levels across the genome do not exhibit prominent megabase-scale variation. No clear enrichment of mCG levels was detectable for TADs that
contain MeCP2-repressed genes, and a subtle depletion of mCG was present in TADs that contain MeCP2-activated genes (Figure 1D; S2A,B). Some evidence of more consistent mCG levels within TADs compared to across TAD boundaries was observed however, suggesting that while TADs do not vary dramatically in average mCG levels, TAD structures are associated with the modest fluctuations in the levels of regional mCG that do exist (Figure S2B).

As an additional test of the association between chromosome topology and regional mCA levels, we analyzed high-resolution Hi-C data generated from an independent study of the cerebral cortex in early development (Bonev et al., 2017), prior to the deposition of mCA. We defined TAD-related “contact domains” of interaction across the genome using the Arrowhead algorithm as previously described (Rao et al., 2014) (Table S3). These analyses of mCA revealed similar enrichments of mCA levels in contact domains that contain MeCP2-repressed genes, a drop-off of mCA levels at the boundaries of domains containing MeCP2-repressed genes, and consistent levels of mCA across contact domains (Figure S2E-H). This shows that regions of close chromatin folding early in development are predictive of mCA patterns in the adult cortex. We further assessed if these findings can be generalized by repeating our analysis using an independently derived cortical DNA methylation dataset (Lister et al., 2013), as well as Hi-C and DNA methylation from the cerebellum (Mellén et al., 2017; Yamada et al., 2019). This analysis detected similar topology-associated mCA patterns (Figure S2I-K). These findings demonstrate a reproducible association between genome topology and large-scale mCA levels in the brain that is not dependent on analysis algorithms that are employed or the brain regions analyzed, suggesting that the topological structure that is established in neurons early in development influences the patterns of mCA deposition that occur during the postnatal period.
We next probed the mechanism by which mCA levels are established within TADs during development. Previous studies have demonstrated that the de novo methyltransferase DNMT3A increases in expression and deposits mCA across the genome during the postnatal period in mice from birth to six weeks of age (Gabel et al., 2015; Guo et al., 2014; Lister et al., 2013). Consistent with its role in patterning methylation during this time, binding of DNMT3A across kilobase-scale regions of the genome at two weeks is predictive of mCA levels in the cortex at eight weeks of age (Figure S3A; Stroud et al., 2017). We therefore investigated whether hundred-kilobase- to megabase-scale patterns of DNMT3A binding during the early postnatal period support our finding that mCA levels are associated with TAD structures. Analysis of DNMT3A ChIP-seq signal in the cortex at two weeks showed that DNMT3A binding is more highly correlated within TADs than across TAD boundaries (Figure 2A). Levels of binding for DNMT3A within TADs at two weeks are well correlated with levels of mCA observed at eight weeks (Figure 2B). Consistent with the idea that TAD boundaries delineate regions of higher and lower DNMT3A binding, aggregate DNMT3A ChIP signal showed distinct reductions at the boundaries of TADs containing MeCP2-repressed genes (Figure 2C). This signal was specific to DNMT3A ChIP, as no such profile was observed for sequencing of the input controls (Figure S3B).

The association between DNMT3A binding and mCA levels for TADs called in the postnatal cortex is also present in contact domains identified in embryonic neurons (Figure S3C, S3D). Notably, the low level of DNMT3A that is expressed in the cortex and bound to the genome at eight weeks is also consistent across TAD structures (Figure S3E), but the degree of binding is not as closely associated with mCA levels in TADs compared to DNMT3A at two weeks (Figure S3F).
These findings indicate that the level of DNMT3A that is recruited to the genome during mCA accumulation is shaped by established TAD structures, and they suggest that the activity of this enzyme in a TAD during postnatal development defines a long-term “set-point” for mCA across this region into adulthood.

To examine the consequences of TAD-associated mCA patterns on MeCP2 binding and function we quantified MeCP2 ChIP-seq signal (Kinde et al., 2016) across TADs in the adult cortex. MeCP2 ChIP signal is specific for MeCP2 but extremely broad and shows only very modest fluctuations across the genome (Chen et al., 2015; Gabel et al., 2015; Lagger et al., 2017). In the context of these small fluctuations however, we found that the level of MeCP2 for TADs is associated with mCA levels within each TAD (Figure 2D). In addition, MeCP2 signal shows higher cross-correlation within TADs compared to across TAD boundaries (Figure S3G). To test the importance of mCA for MeCP2 binding we examined how patterns of MeCP2 binding change when mCA is completely blocked from building up in neurons by conditional deletion of DNMT3A from the brain (DNMT3A Nestin-cKO) during the perinatal period (Figure S3H) (Kinde et al., 2016). This analysis revealed that TADs that normally contain high mCA levels in wild-type conditions displayed subtle reductions in MeCP2 binding in the DNMT3A mutant compared to TADs that contain lower mCA levels in wild-type conditions (Figure S3I). This indicates that high mCA levels in TADs are associated with some enrichment of MeCP2 binding, and upon loss of mCA from the neuronal genome, large-scale MeCP2 profiles are measurably altered.

We next considered how TAD-associated mCA levels across the genome could influence gene regulation within each TAD. Given that TADs coordinate the function of regulatory elements and
genes within a genomic region, we investigated if the set-point level of mCA within a TAD could influence the methylation of these sequences. Levels of methylation for regulatory elements and genes are primarily thought to be established locally, with the methylation status determined by sequence features and the activity level at these sites (Lister et al., 2013; Schübeler, 2015). However, the differences in set-point mCA between TADs could act on top of these local determinants of mCA, driving consistently higher or lower mCA across all of the smaller scale elements within each TAD. In this way, the TAD mCA set-point could affect the regulation of genes by MeCP2.

We therefore investigated if set-point mCA level in each TAD is associated with mCA at these smaller scale genomic elements within the TAD. For this analysis, we defined TSS regions and gene bodies based on Ensembl gene models (see methods). Enhancers are distal regulatory sequences that are bound by transcription factors and can loop to interact with promoters, activating gene transcription (Bulger and Groudine, 2010; Heintzman et al., 2007; Rada-Iglesias et al., 2011). We defined enhancers as sequences that show peaks of ChIP-seq signal for the enhancer-associated Histone H3 lysine 27 acetylation (H3K27ac) and Histone H3 lysine 4 monomethylation (H3K4me1) and do not overlap with known TSS regions or ChIP-seq peaks for promoter-associated Histone H3 lysine 4 trimethylation (H3K4me3) (see methods). We then quantified mCA and mCG levels at these genomic elements and examined whether these levels correlate with the average mCA and mCG levels across the TADs in which they reside. Strikingly, this analysis detected a robust correlation between the set-point level of mCA across each TAD and the mCA level for gene bodies and enhancers found inside the TAD (Figure 2E). This correlation occurred only for elements found within the TAD, breaking down for elements found immediately outside
of the TAD (Figure S3J). Correlations in mCA levels between elements found within the same TAD were also stronger than for elements found on different sides of a TAD boundary (Figure S3K), underscoring that boundaries delineate regions of similar mCA levels for these sequences.

In contrast to gene bodies and enhancers, mCA levels in TSS regions showed much more limited correlation with overall TAD mCA levels, suggesting that these elements largely escape the TAD-associated mCA set-point. Control analysis in which we resampled the location of TADs genome wide, confirmed that TAD boundaries delineate transitions between TAD mCA set-points for enhancers and genes (Figure S3J,K). Randomizing the location of TSS-, gene-, and enhancer-sized sequences within TADs showed that that these small-scale random regions are correlated with TAD mCA levels, and only true TSS regions escape TAD mCA levels (Figure S3M). In contrast to mCA, mCG levels did not show strong TAD-associated signals (Figure S3L,M). Together these findings indicate that mCA levels at enhancers and genes within the TAD are linked to the mCA set-point for each TAD, while TSS regions are protected from these effects. By influencing mCA at genes and enhancers, high TAD-associated mCA levels may directly impact mCA- and MeCP2-mediated regulation of genes within a TAD.

2.2.2 Loss of MeCP2 leads to promoter-associated transcriptional activation.

We next investigated the mechanism by which methylation in high-mCA TADs is read out by MeCP2 to affect mRNA expression. Consistent with our TAD methylation analysis above (Figure 2E), we find that the high TAD mCA set-point for MeCP2-repressed genes is associated with high
mCA in and around the genes, but little to no signature of mCA enrichment is observed at their TSS (Figure 1B; S1C) (Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017). This finding suggests that regulation of these genes occurs through binding of MeCP2 to methylation at regions outside of the proximal promoter. Based on the long length of MeCP2-repressed genes and the observation that the levels of mCA within gene bodies are better correlated with dysregulation of RNA expression upon disruption of MeCP2 compared to extragenic regions (Figure S1D,G,H) (Kinde et al., 2016), we considered several different mechanisms by which MeCP2 might repress transcription. For example, one possible explanation is that binding of MeCP2 to methylation within the gene body inhibits transcriptional elongation by interfering with transcribing RNA polymerase II, resulting in premature termination and reduced production of mature mRNA. Alternatively, binding of MeCP2 to methylation within the gene body could act at a distance to repress transcription initiation or polymerase release occurring near promoter elements. To test these and other possible mechanisms of action, we carried out integrative genomic analyses of the cerebral cortex from MeCP2 KO and wild-type mice, assessing changes in intronic RNA and histone marks associated with gene transcription.

Because intronic RNA is quickly destroyed after pre-mRNA transcription and splicing, analysis of reads corresponding to introns in RNA-seq data can provide read-outs of changes in different stages of transcription (Boswell et al., 2017; Gray et al., 2014). For example, if MeCP2 reduces the processivity of RNA polymerase II leading to a significant level of premature termination under wild-type conditions, knockout of MeCP2 could lead to more efficient completion of transcription elongation and result in an increase in intronic reads at the 3’ end of the pre-mRNA transcript in the MeCP2 knockout, but cause little to no change in intronic reads at the 5’ end. If MeCP2 represses or
activates promoters to modulate transcription, knockout of MeCP2 would result in a consistent change in levels of intronic reads starting at the beginning of MeCP2-repressed genes and continuing along their entire length. Alternatively, if transcription is not upregulated for these genes upon loss of MeCP2, as suggested by one study (Johnson et al., 2017), we would expect no increase or even a reduction in intronic RNA corresponding to these genes.

To determine if these or other effects on transcription can be observed upon mutation of MeCP2 we carried out RNA-seq analysis of intronic RNA in MeCP2 KO and wild-type controls, assessing the direction and distribution of changes in pre-mRNA transcripts. Nuclear RNA was isolated from the cerebral cortex to enrich for short-lived, intron-containing transcripts and total RNA sequencing was performed (Figure 3A). We found that changes in nascent intronic RNA are highly concordant with exonic RNA, both for genes called as significantly dysregulated at the exon level and all genes genome-wide (Figure 3B; S4A,B). Robust coverage of intronic RNA in these data also allowed us to perform differential expression analysis, defining genes that are significantly dysregulated at the intron RNA level upon loss of MeCP2 (Table S1). We find that these dysregulated genes significantly overlap with genes identified by quantification of exonic reads in whole tissue RNA-seq (Figure S4B). In addition, genome-wide changes in nuclear intronic RNA as a function of gene length and gene body mCA levels mirror those observed for total cell exonic reads, with long and highly methylated genes displaying intronic RNA upregulation relative to shorter and lower methylated genes (Figure S4C-F). To examine if MeCP2 regulates transcription initiation or successful completion of transcription elongation, we assessed changes in intronic RNA across the length of genes that are significantly upregulated at the intronic level in the MeCP2 KO. We observed a consistent increase of intronic reads across these long, highly-methylated genes in the MeCP2 KO (Figure 3C; S4G,H). These results suggest that subtle
changes in mRNA observed in the MeCP2 KO result from a concomitant subtle change in pre-mRNA transcription and that promoter-associated transcriptional activation, rather than RNA polymerase II processivity, is increased in MeCP2-repressed genes upon loss of MeCP2.

We further examined the effects of MeCP2 disruption on transcription by performing ChIP-seq analysis of the cerebral cortex in wild-type and MeCP2 KO mice, assessing histone modifications that report on promoter activity and transcription through the gene. To determine if promoters of MeCP2-regulated genes are affected upon loss of MeCP2 we examined profiles of H3K27ac and H3K4me3, two marks closely associated with promoter activation (Heintzman et al., 2007; Santos-Rosa et al., 2002; Schübeler et al., 2004). To test if transcription through genes is altered, we profiled Histone H3 lysine 36 trimethylation (H3K36me3), a modification that is closely associated with levels of transcription (Bannister et al., 2005; Guenther et al., 2007; Krogan et al., 2003). Quantification of changes in gene body-associated H3K36me3 revealed that MeCP2-repressed genes display a subtle but significant increase in this signal upon loss of MeCP2 (Figure 3A,D) that is consistent with the magnitude of the increased mRNA expression that occurs for these genes. H3K27ac and H3K4me3 at the TSS of these genes showed comensurately subtle, but significant upregulation in the MeCP2 KO (Figure 3A,D). Together with our analysis of intronic RNA changes, these findings support a model in which loss of MeCP2 leads to promoter activation of long, highly methylated, MeCP2-repressed genes.

Given the small magnitude of the effects of MeCP2 disruption on gene expression, we sought to further test the consequences of MeCP2 disruption on transcription in an independent set of analyses. In light of the opposite effects on mRNA levels that we and others observe in the MeCP2 OE compared to the MeCP2 KO, we predicted that intronic RNA and transcription-associated chromatin marks should be
reciprocally affected in the MeCP2 OE. We therefore repeated our integrated RNA-seq and ChIP-seq analyses in the MeCP2 OE mice, examining how genes identified as dysregulated in the MeCP2 KO are affected in this strain. Consistent with our predictions, analysis of intronic RNA, H3K27ac, H3K4me3, and H3K36me3 profiles in the MeCP2 OE revealed reciprocal effects on promoter activity and transcription compared to the MeCP2 KO (Figure 3E-G; S4I, Table S1). These findings serve as independent validation that MeCP2 represses long, highly-methylated genes through down-regulation of promoter activity.

2.2.3 MeCP2 represses enhancers that are enriched for mCA and mCG binding sites.

We next considered how binding of MeCP2 to high levels of methylation found outside of the TSS could impact promoter-associated transcriptional activation. One mechanism by which sequences outside of the TSS can impact transcription is through enhancers. Activation of enhancers is associated with high levels of H3K27ac at these sequences, and changes in acetylation are linked to increases and decreases in enhancer activity (Creyghton et al., 2010; Malik et al., 2014; Rada-Iglesias et al., 2011; Spiegel et al., 2014). Our TAD methylation analysis indicates that the high mCA set-point of TADs that contain MeCP2-repressed genes is likely associated with high mCA at the enhancers within the TAD (Figure 1E; 2E). Binding of MeCP2 at these highly methylated enhancers and repression of their activity could result in reduced activation of the MeCP2-repressed genes the enhancers regulate. To date however, no study has examined the effects of loss of MeCP2 on enhancer activity. Notably, visualization of putative enhancers in the vicinity of promoters of MeCP2-repressed genes showed that these regions displayed subtly higher levels of H3K27ac signal in the MeCP2 KO cortex compared to
wild type (Figure 4A), further suggesting that the activity of these elements may be affected by MeCP2. We therefore systematically assessed the effects of loss of MeCP2 on enhancer activity genome-wide.

Through combined analysis of differential H3K27ac ChIP-seq signal for MeCP2 KO and MeCP2 OE cortex samples at enhancer-associated H3K27ac peaks (see methods), we identified enhancers that are significantly altered in the MeCP2 KO and MeCP2 OE cerebral cortex relative to controls (Figure 4B; S5A; Table S4). To assess if the presence of mCA or mCG at enhancers might drive their regulation by MeCP2, we quantified DNA methylation at these sequences. We detected an enrichment of mCA levels at MeCP2-repressed enhancers compared to the genome average and a relative depletion of mCA levels at MeCP2-activated enhancers. mCG levels showed limited enrichment and depletion in MeCP2-repressed and MeCP2-activated enhancers respectively (Figure 4C; S5B).

Given the well-established high affinity of MeCP2 for mCG (Meehan et al., 1989), but the lack of robust enrichment of mCG/CG levels at MeCP2-regulated enhancers, we further investigated if the presence of mCG sites at enhancers is associated with MeCP2-mediated repression. The number of mC binding sites for MeCP2 within an enhancer region is determined by both per-base cytosine methylation at CA and CG dinucleotides (e.g. mCA/CA) and the frequency of each dinucleotide sequence that occurs in the enhancer region. The contribution of dinucleotide frequency to mC site abundance is particularly relevant for mCG, because CG dinucleotides are depleted from the genome and non-uniform in their distribution (Bird, 1980; McClelland and Ivarie, 1982). We therefore examined the number of mC binding sites per kilobase (mC/kb), or “density” of mC sites at enhancers (Figure 4C). In comparison to the limited signal for mCG levels (mCG/CG) at dysregulated enhancers, the density of mCG sites at these enhancers showed more robust association with repression and
activation (Figure 4C; S5B). The larger enrichment and depletion of mCG density compared to mCG levels was driven primarily by enrichment of CG dinucleotides at repressed enhancers (Figure S5B). Our findings therefore suggest a role for both mCA and mCG sites in enhancer regulation by MeCP2, but while enrichment for mCA sites is largely driven by high per-base methylation (mCA/CA), enrichment of mCG sites is primarily driven by higher frequencies of CG dinucleotides (CG/kb) at these sequences.

While mCA is the highest affinity non-CG dinucleotide binding site for MeCP2 (Gabel et al., 2015; Lagger et al., 2017), a recent study has indicated that the third nucleotide in the mCNN context can also affect MeCP2 binding, with mCAC sites showing the highest affinity for MeCP2 binding amongst non-CG sites, and other trinucleotides showing lower affinity (Lagger et al., 2017). We reasoned that if MeCP2 directly binds enhancers to repress their activity, then repressed enhancers should be most robustly enriched for mCAC sites compared to other mC trinucleotide contexts. We therefore quantified the relative enrichment of all mCNN trinucleotide sites at significantly dysregulated enhancers, comparing them to control sets of enhancers. We find that mCAC, of all mC trinucleotides, is most significantly enriched and depleted in MeCP2-repressed and MeCP2-activated enhancers respectively (Figure 4D; S5C). This enrichment of the highest affinity non-CG site for MeCP2 further supports a role for MeCP2 binding in regulation of these enhancers. Consistent with these patterns of high affinity binding sites for MeCP2 at dysregulated enhancers, MeCP2 ChIP-seq signal is enriched at MeCP2-repressed enhancers (Figure 4E). Analysis of a subset of high-stringency MeCP2-repressed enhancers that are selected based on enrichment for MeCP2 binding yields similar changes in histone acetylation and enrichment of mC/kb (Figure S5A,B), further supporting a direct role for MeCP2 in enhancer repression.
While the enhancers that are significantly dysregulated upon loss of MeCP2 are enriched for mCA, mCG, and MeCP2 binding, the levels of methylation at enhancers genome-wide occur in a continuous distribution (Figure S5D), and MeCP2 binds broadly across the genome at enhancers and other regions (Chen et al., 2015; Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017). This raises the possibility that every enhancer may be regulated by MeCP2 to some degree, with the number of mC sites determining the level of repression. To examine this possibility, we performed analysis of H3K27ac changes at all enhancers genome-wide, assessing if the levels of mC at these regulatory regions are correlated with the degree of H3K27ac dysregulation observed in the MeCP2 KO and MeCP2 OE. In this “unthresholded” analysis, we observed a genome-wide association between mCA and mCG sites at enhancers and the dysregulation of H3K27ac at these regulatory elements. This effect is positively correlated in the MeCP2 KO and negatively correlated in the MeCP2 OE (Figure 4F,G). We noted that mCA and mCG density is enriched in and around MeCP2-repressed enhancers (Figure 4C), but if binding of MeCP2 to methylation specifically at the enhancer is required for repression of these elements, we reasoned that the methylation of the enhancer sequence itself should be most predictive of enhancer de-repression upon loss of MeCP2. We therefore examined the correlation between the density of mCA and mCG sites and change in H3K27ac at enhancers at 500bp spatial resolution. We observe that the degree of H3K27ac upregulation of enhancers in the MeCP2 KO is most robustly correlated with the density of mCA and mCG sites occurring at the centers of the enhancers (Figure 4H), and the reciprocal relationship is observed in the MeCP2 OE (Figure 4I). This finding points to methylation specifically within the regulatory sequence itself as the primary determinant of regulation by MeCP2. Together, these findings support a model in which binding of MeCP2 to mCA and mCG
sites within enhancers genome-wide leads to downregulation of acetylation that is proportional to the density of methyl sites at each sequence.

Examination of profiles of increased acetylation in the MeCP2 KO suggested some increases in signal occurred at sites outside of our stringently defined enhancer regions (Figure S5E). We therefore further investigated the potential for sub-threshold regulation of H3K27ac at putative regulatory sites that were not detected by our initial peak calling. Examination of a comprehensive set of possible regulatory elements defined using a compendium of ATAC-seq peaks across cells types from 13 different mouse tissues (Cusanovich et al., 2018) showed that these sites often overlap with high-confidence enhancer peaks (97.1% of enhancers contain an ATAC peak) and are measurably enriched for acetyl signal above control, non-peak regions (Figure S5E,F). Comparison of acetylation in MeCP2 KO and MeCP2 OE to controls at these ATAC-peaks genome-wide revealed noisier, but detectable changes in acetylation that were associated with mCA and mCG density (Figure S5F,G,H). These findings indicate that MeCP2 also regulates histone acetylation at these sub-threshold putative regulatory sites.

Given the higher level and larger absolute change of H3K27ac that we observe in the MeCP2 KO and MeCP2 OE at enhancers compared to sub-threshold regions (Figure S5F), the dysregulation of called enhancers is likely to have the largest impact on gene expression. However, alterations in acetylation at sub-threshold sites may also contribute to the transcriptional effects observed in MeCP2 mutants.

To directly examine the importance of methylation in enhancer repression and dissect the contribution of mCA and mCG in these effects, we next investigated the effects of selective disruption of mCA on enhancer activity. Disruption of DNA methylation by conditional knockout of DNMT3A using brain-specific Nestin-Cre leads to overlapping, but more limited effects on MeCP2-repressed genes than...
those seen in MeCP2 mutants (Gabel et al., 2015). Given that conditional deletion of DNMT3A leads to loss of mCA, but leaves mCG largely intact (Gabel et al., 2015), we reasoned that these partial effects on gene expression might arise due to loss of mCA-mediated enhancer repression, but preservation of mCG-mediated effects. To test this hypothesis, we assessed the effects of conditional deletion of DNMT3A specifically from post-mitotic neurons (DNMT3A Baf53-cKO, see methods). Bisulfite-seq of cerebral cortex from these mice revealed ablation of mCA, but retention of the majority of mCG at genomic elements, including TADs and enhancers (Figure 5A, S6A-C). Total RNA-seq analysis detected significant, but partial dysregulation of MeCP2-regulated genes (Figure 5B) that is consistent with previous DNMT3A brain-specific Nestin-cKO studies (Gabel et al., 2015). Quantification of H3K27ac signal at enhancers genome wide, similarly detected robust, but partial, dysregulation of MeCP2-repressed and –activated enhancers in DNMT3A mutants (Figure 5B). Analysis of normal methylation levels at enhancers defined as significantly dysregulated in the DNMT3A Baf53b-cKO mice revealed enrichment for mCA at upregulated enhancers, while mCG was equivalent between up and downregulated enhancers (Figure 5C, S6D,E). In genome-wide, unthresholded analysis, we detected an association between the change in H3K27ac signal and wild-type mCA levels at enhancers that is similar to the trends observed in MeCP2 mutant mice (Figure 5D). In contrast, no robust association was detected for mCG at enhancers (Figure 5D). Together these findings support the role for mCA in the repression of MeCP2-regulated neuronal enhancers. Furthermore, the lack of mCG-associated changes in the DNMT3A Baf53b-cKO mice, together with the quantitatively lower dysregulation of MeCP2-repressed genes and enhancers compared to MeCP2 mutants, supports an important role for mCG in MeCP2-mediated enhancer regulation.
2.2.4 MeCP2-mediated enhancer repression is associated with gene regulation.

We next examined if the effects we observe on enhancers can explain changes in gene transcription in the MeCP2 KO. Interactions of promoters and enhancers are enriched within TADs, such that enhancers tend to interact with, and activate, promoters within the same TAD (Bell et al., 1999; Ong and Corces, 2014). We reasoned therefore that if MeCP2-repressed enhancers regulate MeCP2-repressed genes they should be found within the same TAD as these genes. In addition, based on our analysis of TAD methylation above (Figure 2E), we predicted that the high mCA set-point in TADs that contain MeCP2-repressed genes would result in high levels of mCA at enhancers within these TADs. Indeed, we found that enhancers within TADs that contain MeCP2-repressed genes are more highly methylated as a population than the genome average (Figure 6A), supporting the notion that these enhancers are likely to be repressed by MeCP2. To directly examine the link between enhancer dysregulation in TADs and changes in gene expression, we analyzed the distribution of MeCP2-regulated enhancers in TADs across the genome. We found that MeCP2-repressed enhancers are significantly more likely than by chance to be found within TADs that contain MeCP2-repressed genes (Figure 6B). MeCP2-repressed enhancers are also significantly associated with MeCP2-repressed genes by GREAT analysis (McLean et al., 2010) and are more likely to show enrichment of interactions with promoters of MeCP2-repressed genes by Hi-C analysis (Figure 6B). Because our unthresholded analysis of H3K27ac changes revealed a genome-wide association between DNA methylation levels and repression of H3K27ac at enhancers (Figure 4F,G), we considered that highly-methylated enhancers linked to MeCP2-repressed genes may display widespread upregulation of H3K27ac in the MeCP2 KO that is below the statistical threshold for significance. Indeed, inspection of changes in acetylation at enhancers in TADs that contain prominent MeCP2-repressed genes showed that while only some were called as significantly changed by our differential acetylation...
analysis, many of these enhancers display a trend toward upregulation. Conversely, this pattern did not hold within TADs devoid of MeCP2-repressed genes (Figure 6C). Quantification of acetylation changes for the population of enhancers found in TADs containing MeCP2-repressed genes confirmed this observation, revealing that the H3K27ac signal in this population of enhancers is significantly upregulated in the MeCP2 KO (Figure 6D). Analysis of MeCP2 OE mice further showed reciprocal downregulation of acetylation for the population of enhancers found within TADs that contain MeCP2-repressed genes (Figure 6D).

This analysis links dysregulation of enhancers to changes in gene expression upon loss of MeCP2. We next wanted to understand if regulation of enhancers by MeCP2 can explain our previous finding that the levels of methylation within gene bodies are better correlated with changes in mRNA expression upon disruption of MeCP2 compared to regions flanking genes (Figure S1F-H; S4E,F) (Kinde et al., 2016). We reasoned that the gene-body-associated signal we have detected may in part reflect the regulation of enhancers located within genes, and that MeCP2 may preferentially repress intragenic enhancers. Indeed, examination of enhancers within TADs containing prominent MeCP2-repressed genes (e.g. Zmat4, Auts2) revealed that enhancers within repressed genes show a stronger trend toward dysregulation upon loss or overexpression of MeCP2 compared to enhancers found outside these genes (Figure 6C). Analysis of all MeCP2-repressed enhancers showed that they are preferentially found within genes (Figure 6E), and significantly enriched in MeCP2-repressed genes (Figure 6B). In contrast, MeCP2-repressed enhancers are depleted from MeCP2-activated genes (Figure 6B). Examination of all enhancers located within MeCP2-repressed genes revealed that they are more robustly upregulated as a population in the MeCP2 KO and downregulated in the MeCP2 OE compared to enhancers within the same TAD as these genes (Figure 6D). Furthermore, analysis of
enhancer changes genome-wide revealed that while changes in acetylation of both intragenic and extragenic enhancers are correlated with mCA and mCG sites at these enhancers, for a given level of methylation, intragenic enhancers show more upregulation than extragenic enhancers (Figure 6F). Thus, intragenic enhancers appear to be more susceptible to repression by MeCP2 than enhancers outside of genes.

Given the susceptibility of intragenic enhancers to repression by MeCP2, we investigated the degree to which these enhancers interact with the promoters of their cognate genes and whether disruption of MeCP2 affects the strength of these interactions to affect gene expression. Interrogation of Hi-C interaction maps revealed that intragenic sequences preferentially interact with promoters compared to extragenic sequences (Figure 6G), and intragenic enhancers make more contacts with promoters of the gene they reside in than enhancers that are equidistant from the promoter but extragenic (Figure 6H). This suggests that intragenic enhancers are privileged compared to extragenic enhancers for regulation of their cognate genes. We next examined if disruption of MeCP2 alters enhancer-promoter interactions. Quantitative 3C analysis of a representative set of enhancers that are altered in MeCP2 mutants and associated with MeCP2-regulated genes showed no evidence of changes in interaction frequencies between these enhancers and their target promoters upon disruption of MeCP2 (Figure S5I). This suggests that MeCP2 does not regulate the contact frequencies between enhancers and promoters, but rather the degree to which enhancers drive gene activation once looped to their target promoter.

If the repression of genes by MeCP2 is mediated through repression of intragenic enhancers, then we reasoned that gene body methylation and upregulation of gene expression in the MeCP2 KO cortex
should be more tightly linked for genes that contain enhancers than genes that do not contain enhancers. Because long genes are more likely to contain enhancers than shorter genes (Figure S5J), robust regulation of intragenic enhancers by MeCP2 binding could also help to explain the global trend toward upregulation of longer genes with high gene body mCA levels upon loss of MeCP2 that we and others have observed (Figure S1F-H) (Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017; Rube et al., 2016; Sugino et al., 2014). We therefore examined the effects of MeCP2 knockout and MeCP2 overexpression on gene expression for genes that do or do not contain intragenic enhancers. We find little to no correlation between dysregulation of gene expression and the density of gene body mCA sites for genes that do not contain enhancers, but we observe a robust correlation for genes containing one or more intragenic enhancer (Figure 6I, S5L). Importantly, these effects were seen even when we controlled for gene length in this analysis (Figure S5K). Consistent with the idea that enhancers substantially contribute to the gene-body mCA signature that we have previously observed for MeCP2-repressed genes, we found that the H3K27ac changes in intragenic enhancers was more predictive of the change in gene expression in MeCP2 mutants than H3K27ac changes in intragenic non-peak sequences (Figure S5M). These results point to dysregulation of intragenic enhancers as an underlying cause for the dysregulation of long, highly methylated genes that we and others have observed in MeCP2 mutant mice.

2.3 Discussion

The recent discovery of high levels of non-CG methylation in neurons and the realization that MeCP2 can bind to this neuron-enriched methyl-mark have uncovered a new pathway of gene regulation that is critical for neuronal function. However, our understanding of how patterns of non-CG methylation
are established across the neuronal genome is still in its infancy. In this study, we have uncovered a link between genome topology in the nucleus of neurons and the levels of mCA occurring at the hundred-kilobase to megabase scale. Our findings suggest a model in which TAD structures in neurons dictate the distribution of DNMT3A across the genome during early postnatal development to establish higher or lower mCA set-points for genes and enhancers within each TAD (Figure 7A). In mature neurons, MeCP2 then reads-out this methylation to repress histone acetylation at enhancer elements and control gene expression. Because MeCP2 most potently represses highly-methylated, intragenic enhancers, its repressive effects have the largest impact on genes that contain multiple enhancers and lie within high-mCA TADs. Upon disruption of MeCP2 or loss of mCA, dysregulation of intragenic enhancers leads to altered expression of these genes, driving nervous system dysfunction (Figure 7B).

The varying levels of CA methylation in TADs appear to be facilitated by the dynamic and unimodal distribution of mCA across the genome that allows for large relative fluctuations in levels between megabase scale TADs and contrast with megabase scale mCG levels, which tend to be very high and consistent across domains. The consistently high levels of DNMT3A recruitment and resulting high set-point of CA methylation that we observe in TADs that contain MeCP2-repressed genes provide insight into the origin of high mCA that we and others have previously observed at these genes (Kinde et al., 2016; Lagger et al., 2017). Future studies will be needed to determine how DNMT3A recruitment and mCA becomes associated with TAD structures during postnatal development.

Our finding that the mCA set-point for TADs correlates well with the mCA levels for enhancers and genes in each domain suggests that, while sequence features and the regulatory state of each enhancer are known to affect their methylation status (Schübeler, 2015), regional mCA levels also influence total levels of mCA at these important regulatory elements. In contrast, mCA levels at TSSs are
generally extremely low and are not well correlated with TAD-associated mCA levels. Thus, while TSSs are not likely to be directly affected by TAD-associated mCA, enhancers within high-mCA TADs are enriched for mCA and are more likely to be repressed by MeCP2 compared to enhancers within low-mCA TADs. In this way, TADs can dictate mCA patterns across the genome and set the stage for MeCP2-mediated gene regulation in the brain.

Previously, based on the long length of MeCP2-repressed genes, as well as the observed correlation between gene-body methylation and changes in gene expression upon loss of MeCP2 (e.g. Figure 1B; S1D,G,H), we had proposed a mechanism whereby binding of MeCP2 within genes interrupts RNA polymerase II processivity to downregulate expression (Kinde et al., 2016). Surprisingly, our analysis here did not uncover evidence of altered polymerase processivity or termination for long, highly-methylated genes in the MeCP2 mutants, but instead pointed to changes in promoter activation for these MeCP2-repressed genes. In parallel with our study, Boxer, Renthal, Greenberg, and colleagues (Boxer et al., co-submitted) have employed several high-resolution methodologies to examine alterations in gene expression in MeCP2 mutants on multiple levels (e.g. initiation, elongation, splicing). This analysis identified alterations in transcription initiation as the primary change in MeCP2-regulated genes upon loss of MeCP2, which is consistent with our findings here. Our search for how mCA and mCG sites found outside the TSS of the gene can affect promoter activation led to the identification of intragenic enhancers that are dysregulated in MeCP2 mutants, and showed that the methylation of these enhancers is most tightly linked to changes in gene expression (Figure 6I, S5J-M). Hence, our new findings support the model in which MeCP2 preferentially represses long, highly-methylated genes through binding within the gene body, but provide an unexpected mechanism of MeCP2 enhancer repression that acts upstream of transcriptional initiation to mediate this regulation.
While MeCP2 has been shown to act as a transcriptional repressor \textit{in vitro}, the protein binds nearly ubiquitously across the genome \textit{in vivo} and its most relevant sites of action have not been clear. Through genome-wide analysis of changes in histone acetylation, we have uncovered a role for mCA, mCG, and MeCP2 in enhancer regulation. Multiple biochemical studies have shown that MeCP2 binds with high affinity to both mCG and mCA, suggesting that each of these marks are important for gene regulation by MeCP2 (Chen et al., 2015; Gabel et al., 2015; Guo et al., 2014; Lagger et al., 2017). Previous genomic analyses have primarily detected enrichments of per-cytosine mCA levels (mCA/CA) at sequences in and around MeCP2-repressed genes however (Chen et al., 2015; Gabel et al., 2015; Kinde et al., 2016; Lagger et al., 2017), with limited evidence of higher mCG levels (mCG/CG) at these genes. Our findings show that TAD-scale fluctuations in mCA/CA are a major determinant of the number of mCA sites at enhancers that are repressed by MeCP2. In contrast, fluctuations in the number of CG dinucleotides present in enhancer sequences, rather than mCG/CG, primarily contribute to the enrichment of mCG sites found at MeCP2-repressed enhancers. Thus, our findings uncover a novel function for MeCP2 in regulating enhancers and support a role for the high affinity of MeCP2 to both mCG and mCA in its repressive function at these sites.

Consistent with some role for mC sites in recruiting MeCP2 to the genome, MeCP2 ChIP-seq profiles show a measurable association with mCA and mCG binding sites (Figure 2D, S3I). However, the binding of the protein is very broad and signal is present at highly demethylated sites. Therefore, it is possible that the presence or absence of mC binding sites for MeCP2 may modulate its repressive activity, rather than serving as a strict determinant of its binding. Future analyses will be necessary to dissect the importance of mC sites for MeCP2 recruitment versus modulation of its repressive activity.
Our identification of genes and enhancers defined as changed in MeCP2 mutants using statistical thresholds revealed critical determinants for enhancer regulation by MeCP2 (e.g. number of mCA and mCG sites, presence within a gene), but our findings suggest that these lists of significantly altered enhancers represent only the extreme examples in a continuum of genome-wide effects. DNA methylation and MeCP2 binding occur nearly ubiquitously across the genome and show only limited fluctuations in levels from region to region. Our unthresholded, genome-wide analysis of enhancer methylation supports a graded repressive effect of MeCP2 that is proportional to the number of mC sites at each enhancer. Thus, nearly every gene-regulatory element may be regulated to some extent by MeCP2. Alterations in this pervasive tuning of enhancers may explain why disruption of non-CG methylation or MeCP2 leads to profound nervous system dysfunction, despite the small magnitude of gene expression effects that are observed for individual genes.

Analysis of more than a dozen human and mouse MeCP2 mutant datasets carried out across multiple laboratories support a role for MeCP2 in repressing long, highly methylated genes (Gabel et al., 2015; Kinde et al., 2015, 2016; Lagger et al., 2017; Rube et al., 2016; Sugino et al., 2014). The mechanism by which this repression occurs has been an area of open investigation however, and two recent studies have suggested that a model of direct transcriptional repression of these genes requires further testing. One study proposed that upregulation of long genes upon loss of MeCP2 is a consequence of post-transcriptional effects rather than disrupted transcriptional repression (Johnson et al., 2017), while another suggested that the subtle dysregulation of long genes observed in individual MeCP2 mutant datasets does not reach a threshold for statistical significance (Raman et al., 2018). Here, we have addressed these issues by confirming these effects in a new set of MeCP2 KO and MeCP2 OE RNA-
seq analyses and by using ChIP-seq analysis to independently assess effects on transcription-associated histone modifications. Our new findings further support the conclusion that MeCP2 mediates methylation-associated transcriptional repression that preferentially influences long genes. Notably, parallel analyses from an independent laboratory using multiple methodologies to assess transcriptional changes in large-replicate-number MeCP2 mutant datasets (Boxer et al., co-submitted), have further verified that mutation of MeCP2 leads to bona fide gene-length- and methylation-associated dysregulation of gene transcription. Our additional insights implicating enhancers in this gene regulation not only validate these effects, but also provide one mechanism by which MeCP2 can carry out this regulation.

While our study focuses on understanding the origin and function of mCG and mCA at MeCP2-repressed genes, our findings suggest that enhancer repression by MeCP2 may also influence MeCP2-activated genes. A distinguishing feature of MeCP2-activated enhancers identified in this study is that they are depleted of mCA and mCG sites compared to the genome average (Figure 4C; S5B-D). This suggests that acetylation changes that we observed for these activated enhancers may actually reflect a relative lack of de-repression upon loss of MeCP2 and an escape from repression upon increased MeCP2 levels, respectively. Notably, MeCP2-activated enhancers are significantly associated with MeCP2-activated genes (Figure 6B). This raises the possibility that a lack of de-repression at some enhancers upon loss of MeCP2 drives the apparent relative downregulation of MeCP2-activated genes that is detected. Future studies will be necessary to further delineate whether the downregulation effects that we observe upon loss of MeCP2 are caused by a lack of de-repression at MeCP2-activated enhancers or whether MeCP2 can directly activate genes through independent molecular mechanisms.
Previous studies have implicated the interaction between MeCP2 and the NCoR histone deacetylase co-repressor complex as critical for MeCP2 to mediate gene repression (Ebert et al., 2013; Kokura et al., 2001; Lyst et al., 2013; Nott et al., 2016). Indeed, analysis from Boxer, Renthal, and colleagues (Boxer et al., co-submitted) shows that specific disruption of the MeCP2-NCoR interaction in the MeCP2 R306C mutant mouse leads to similar effects on transcription and chromatin as deletion of MeCP2. Our findings detecting regulation of enhancers by MeCP2 suggest an important site of action for the MeCP2-NCoR complex. In one model, recruitment or regulation of enhancer-associated NCoR by MeCP2 induces histone deacetylation to block the activating effect of the enhancer on its target promoter. Alternatively, MeCP2 may modulate NCoR activity on non-histone proteins. Stimulus-dependent phosphorylation of MeCP2 may selectively disrupt the MeCP2-NCoR interaction and modulate NCoR function at enhancers. Such a mechanism could facilitate de-repression of enhancer activity upon neuronal activation, and help drive activity-dependent gene expression in the brain (Ebert et al., 2013; Yap and Greenberg, 2018).

Our quantification of H3K27ac shows that its highest signal is located at enhancers and promoters and indicates that robust mCA- and mCG-associated changes occur at enhancers in MeCP2 mutants. MeCP2 binds broadly across the genome however, including at non-regulatory sequences, and disruption of MeCP2 has effects on acetylation in broad domains that are not strictly confined to our stringently defined enhancer sequences (Figure S5E) (Boxer et al., co-submitted). Our analysis of ATAC-seq peaks suggest that additional regulatory sequences that were not detected as H3K27ac ChIP-seq peaks may be regulated by MeCP2, and that acetylation changes in enhancers and these regions contribute to increases in gene body acetylation detected. However, MeCP2 is likely to have repressive effects on chromatin structure outside of enhancers that can contribute to gene regulation.
A striking feature of the enhancer repression that we observe is that MeCP2 appears to have larger repressive effects on intragenic enhancers compared to extragenic enhancers, and de-repression of intragenic enhancers upon loss of MeCP2 is closely linked to upregulation of the gene in which the enhancers reside. Our analysis of Hi-C data further suggests that, in general, intragenic enhancers more readily contact and regulate their cognate promoters than extragenic enhancers, making the dysregulation of intragenic enhancers in MeCP2 mutants particularly impactful for gene expression. Notably, the larger MeCP2-mediated effects on intragenic enhancer acetylation and preferential Hi-C interactions that we observe provide an explanation for previous observations in which the mC sites specifically within longer genes are most tightly linked to the level of repression imposed on the gene by MeCP2 (Kinde et al., 2016; Lagger et al., 2017). Future studies to examine the mechanism by which MeCP2 preferentially represses intragenic enhancers will allow us to further understand the origins of these effects.

Our findings have important implications for the molecular etiology of Rett syndrome and MeCP2 duplication disorders. Our results indicate that when MeCP2 is lost or overexpressed in disease, this disrupts the precise control of enhancers normally maintained by MeCP2 through its interaction with genome-topology-associated DNA methylation. The resulting dysregulation of enhancers that occurs can contribute to altered gene expression, driving disease pathology. In this way, our study suggests that Rett syndrome and MeCP2 duplication syndrome are disorders that stem in part from disruption of enhancer control.
# 2.4 STAR Methods

## 2.4.1 Key Resources Table

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit polyclonal anti-Histone H3 (acetyl K27)</td>
<td>Abcam</td>
<td>Cat# ab4729; RRID:AB_2118291</td>
</tr>
<tr>
<td>Mouse monoclonal anti-Histone H3 (tri methyl K4)</td>
<td>Abcam</td>
<td>Cat# ab1012; RRID:AB_442796</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-Histone (tri methyl K36)</td>
<td>Active Motif</td>
<td>Cat# 61101; RRID:AB_2615073</td>
</tr>
<tr>
<td><strong>Bacterial and Virus Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Biological Samples</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OptiPrep Density Gradient Medium</td>
<td>Sigma-Aldrich</td>
<td>Cat#: D1556</td>
</tr>
<tr>
<td>IGEPAL CA-630</td>
<td>Sigma-Aldrich</td>
<td>Cat#: I8896</td>
</tr>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNeasy Micro Kit</td>
<td>Qiagen</td>
<td>Cat#: 74004</td>
</tr>
<tr>
<td>NEBNext Ultra Directional RNA Library Prep Kit for Illumina</td>
<td>NEB</td>
<td>Cat#: E7420S</td>
</tr>
<tr>
<td>NEBNext Multiplex Oligos for Illumina (Index Primers Set 1)</td>
<td>NEB</td>
<td>Cat#: E7335S</td>
</tr>
<tr>
<td>Product Description</td>
<td>Manufacturer</td>
<td>Catalog Number</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>NEBNext rRNA Depletion Kit (Human/Mouse/Rat)</td>
<td>NEB</td>
<td>Cat#: E6310L</td>
</tr>
<tr>
<td>Ovation Ultralow Library System V2</td>
<td>NuGEN</td>
<td>Cat#: 0344-32</td>
</tr>
<tr>
<td>Ovation Ultralow Methyl-Seq Library System</td>
<td>NuGEN</td>
<td>Cat# 0535-32</td>
</tr>
<tr>
<td>2S Set A Indexing Kit</td>
<td>Swift Biosciences</td>
<td>Cat#: 26148</td>
</tr>
</tbody>
</table>

### Deposited Data

<table>
<thead>
<tr>
<th>Type</th>
<th>Source</th>
<th>GEO Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA-sequencing data</td>
<td>This paper</td>
<td>GSE123373</td>
</tr>
<tr>
<td>ChIP-sequencing data (H3K27Ac, H3K4Me3, H3K36me3)</td>
<td>This paper</td>
<td>GSE123373</td>
</tr>
<tr>
<td>ENCODE H3K4me1 ChIP-seq</td>
<td>ENCODE</td>
<td>GSM769022</td>
</tr>
<tr>
<td>Bisulfite-sequencing data</td>
<td>This paper</td>
<td>GSE123373</td>
</tr>
<tr>
<td>MeCP2 ChIP-seq</td>
<td>(Kinde et al., 2016)</td>
<td>GSE90704</td>
</tr>
<tr>
<td>DNMT3A ChIP-seq, Bisulfite-seq</td>
<td>(Stroud et al., 2017)</td>
<td>GSE104298</td>
</tr>
<tr>
<td>Dixon Hi-C contact matrices</td>
<td>(Dixon et al., 2012)</td>
<td><a href="http://chromosome.sdsc.edu/mouse/hi-c/cortex.norm.tar.gz">http://chromosome.sdsc.edu/mouse/hi-c/cortex.norm.tar.gz</a></td>
</tr>
<tr>
<td>Bonev Hi-C data</td>
<td>(Bonev et al., 2017)</td>
<td>GSE96107</td>
</tr>
<tr>
<td>Bisulfite-seq</td>
<td>(Lister et al., 2013)</td>
<td>GSE47966</td>
</tr>
<tr>
<td>Bisulfite-seq</td>
<td>(Mellén et al., 2017)</td>
<td>GSE95628</td>
</tr>
<tr>
<td>Mus musculus mm9 genome assembly</td>
<td>UCSC</td>
<td><a href="http://hgdownload.soe.ucsc.edu/goldenPath/mm9/">http://hgdownload.soe.ucsc.edu/goldenPath/mm9/</a></td>
</tr>
<tr>
<td>Experimental Models: Cell Lines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse: MeCP2 KO: B6.129P2(C)-MeCP2&lt;sup&gt;tm1.Bird&lt;/sup&gt;/J</td>
<td>The Jackson Laboratory</td>
<td>JAX: 003890; RRID:IMSR_JAX: 003890</td>
</tr>
<tr>
<td>Mouse: MeCP2 OE: FVB-Tg(MECP2)3Hzo/J</td>
<td>The Jackson Laboratory</td>
<td>JAX: 008680; RRID:IMSR_JAX: 008680</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experimental Models: Organisms/Strains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mus musculus mm9 genome assembly</td>
</tr>
<tr>
<td>Ensembl gene models</td>
</tr>
</tbody>
</table>

Mouse: MeCP2 KO: B6.129P2(C)-MeCP2<sup>tm1.Bird</sup>/J
The Jackson Laboratory
JAX: 003890; RRID:IMSR_JAX: 003890

Mouse: MeCP2 OE: FVB-Tg(MECP2)3Hzo/J
The Jackson Laboratory
JAX: 008680; RRID:IMSR_JAX: 008680
| Mouse: DNMT3A Baf53b-cKO: Dnmt3a^-/- | Derived from DNMT3A fl/fl strain (described in Kaneda et al., 2004), mice provided by M. Goodell) and Baf53b-Cre (Zhan et al., 2015), mice obtained from The Jackson Laboratory (IMSR Cat# JAX:027826, RRID:IMSR_JAX:027826). | This study |
Mouse: DNMT3A Nestin-cKO: Dnmt3a<sup>fl/fl</sup> Tg(Nes-cre)<sup>1Kln/J</sup>

Derived from DNMT3A fl/fl strain (described in (Kaneda et al., 2004), mice provided by M. Goodell) and Nes-Cre (Tronche et al., 1999), mice obtained from The Jackson Laboratory (IMSR Cat# JAX:003771, RRID:IMSR_JAX:003771).

(Gabel et al., 2015)

<table>
<thead>
<tr>
<th>Oligonucleotides</th>
<th>IDT</th>
<th>Custom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zmat4 anchor primer:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGCTGTTTGGAAATGTAGCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zmat4 – e1 primer:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGGTGAGAAGGGAAGAGACAGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zmat4 – e2 primer:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCTTGCCAGCTCCTGTCTTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zmat4 – e3 primer:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGTGTGCCACTACCCTGCTCAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primer Name</td>
<td>Sequence</td>
<td>Supplier</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Zmat4 – e1 negative region primer:</td>
<td>CGTGAATGCCAAATGTGTTTC</td>
<td>IDT</td>
</tr>
<tr>
<td>Zmat4 – e2 negative region primer:</td>
<td>CAAGGCTGGGCACTCTCTGT</td>
<td>IDT</td>
</tr>
<tr>
<td>Zmat4 – e3 negative region primer:</td>
<td>ACAGTATTGTGTTGGCATTACAG</td>
<td>IDT</td>
</tr>
<tr>
<td>Efna5 anchor primer:</td>
<td>TGCTACCGTTGTTTGGTTTG</td>
<td>IDT</td>
</tr>
<tr>
<td>Efna5 – e1 primer:</td>
<td>CCCTAAATGTCTCAGAACATGTGG</td>
<td>IDT</td>
</tr>
<tr>
<td>Efna5 – e2 primer:</td>
<td>GCTGAGCCTGCTGAAATATGTA</td>
<td>IDT</td>
</tr>
<tr>
<td>Efna5 – e1 negative region primer:</td>
<td>GCCCAAGGTGATTTCTAAACTGT</td>
<td>IDT</td>
</tr>
<tr>
<td>Efna5 – e2 negative region primer:</td>
<td>TTGCTTTCAAGAACTCCTATTTCA</td>
<td>IDT</td>
</tr>
<tr>
<td>Snx24 anchor primer:</td>
<td>TAGAAACCACAGATGCTGTA</td>
<td>IDT</td>
</tr>
<tr>
<td>Snx24 – e1 primer:</td>
<td>GGCTAAGGTAGATCCAGAC</td>
<td>IDT</td>
</tr>
<tr>
<td>Snx24 – e1 negative region primer:</td>
<td>CATTTGAGACTTTCCAGTTCC</td>
<td>IDT</td>
</tr>
<tr>
<td>Primer Type</td>
<td>Sequence</td>
<td>IDT</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Tshz1 anchor primer:</td>
<td>ATGTCCCCAAATGTCTGTGTTT</td>
<td></td>
</tr>
<tr>
<td>Tshz1 – e1 primer:</td>
<td>CCAGGCTTGAAGTAGTTTCTGG</td>
<td></td>
</tr>
<tr>
<td>Tshz1 – e2 primer:</td>
<td>ACAGTGCCAGCAAGGTATG</td>
<td></td>
</tr>
<tr>
<td>Tshz1 – e1/e2 negative region primer:</td>
<td>TTGAAACAAAAATGTCATCCAG</td>
<td></td>
</tr>
<tr>
<td>Actb anchor primer:</td>
<td>TGTTACCAAACTGGGACGACAT</td>
<td></td>
</tr>
<tr>
<td>Actb – e1 primer:</td>
<td>AGACATGCCAGAAGCGGTG</td>
<td></td>
</tr>
<tr>
<td>Actb – e1 negative region primer:</td>
<td>TAGACTTAGCCAGCCTGGAACT</td>
<td></td>
</tr>
<tr>
<td>Ank2 anchor primer:</td>
<td>GGCACCTCCTCTGTCTATGAGG</td>
<td></td>
</tr>
<tr>
<td>Ank2 – e1 primer:</td>
<td>ATATGTGTAGGGGCCAGTGT</td>
<td></td>
</tr>
<tr>
<td>Ank2 – e1 negative region primer:</td>
<td>ACGGTTTCTGTGCAGTTAGT</td>
<td></td>
</tr>
<tr>
<td>Zfp60 anchor primer:</td>
<td>GATGGTGTACACGTCAGATG</td>
<td></td>
</tr>
</tbody>
</table>
Zfp60 – e1 primer:  
GGTGGTCATCTATAACTGTCTTGAG  
| IDT | Custom |

Zfp60 – e1 negative region primer:  
TAAGTGGGACGTTCATTCC  
| IDT | Custom |

**Recombinant DNA**

<table>
<thead>
<tr>
<th>BAC:</th>
<th>Source</th>
<th>Cat#</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP23-21B20</td>
<td>Life Technologies</td>
<td>21B20 RPCI-23 MM BAC CLONE</td>
<td></td>
</tr>
<tr>
<td>RP23-320D14</td>
<td>Life Technologies</td>
<td>320D14 RPCI-23 MM BAC CLONE</td>
<td></td>
</tr>
<tr>
<td>RP23-51F6</td>
<td>Life Technologies</td>
<td>51F6 RPCI-23 MM BAC CLONE</td>
<td></td>
</tr>
<tr>
<td>RP23-309F7</td>
<td>BACPAC Resources</td>
<td>Cat#: RP23-309F7</td>
<td></td>
</tr>
<tr>
<td>RP23-207I5</td>
<td>BACPAC Resources</td>
<td>Cat#: RP23-207I5</td>
<td></td>
</tr>
<tr>
<td>RP23-153P8</td>
<td>BACPAC Resources</td>
<td>Cat#: RP23-153P8</td>
<td></td>
</tr>
<tr>
<td>RP23-35M10</td>
<td>BACPAC Resources</td>
<td>Cat#: RP23-35M10</td>
<td></td>
</tr>
<tr>
<td>RP23-399L5</td>
<td>BACPAC Resources</td>
<td>Cat#: RP23-399L5</td>
<td></td>
</tr>
<tr>
<td>RP23-181E8</td>
<td>BACPAC Resources</td>
<td>Cat#: RP23-181E8</td>
<td></td>
</tr>
</tbody>
</table>

**Software and Algorithms**
<table>
<thead>
<tr>
<th>Tool</th>
<th>Version</th>
<th>Authors</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAMtools</td>
<td>v1.3</td>
<td>Li and Durbin, 2009</td>
<td><a href="https://sourceforge.net/projects/samtools/files/">https://sourceforge.net/projects/samtools/files/</a></td>
</tr>
<tr>
<td>BEDtools2</td>
<td>v2.25.0</td>
<td>Quinlan and Hall, 2010</td>
<td><a href="https://github.com/arq5x/bedtools2">https://github.com/arq5x/bedtools2</a></td>
</tr>
<tr>
<td>STAR</td>
<td></td>
<td>Dobin et al., 2013</td>
<td><a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a></td>
</tr>
<tr>
<td>fastQC</td>
<td></td>
<td></td>
<td><a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a></td>
</tr>
<tr>
<td>Tool</td>
<td>Reference</td>
<td>Website</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------</td>
<td>-------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>MACS2 (v2.1.0)</td>
<td>(Zhang et al., 2008)</td>
<td><a href="https://github.com/taoiliu/MACS">https://github.com/taoiliu/MACS</a></td>
<td></td>
</tr>
<tr>
<td>Trim galore</td>
<td></td>
<td><a href="https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/">https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/</a></td>
<td></td>
</tr>
<tr>
<td>Tadtree</td>
<td>(Weinreb and Raphael, 2016)</td>
<td><a href="http://compbio.cs.brown.edu/projects/tadtree/">http://compbio.cs.brown.edu/projects/tadtree/</a></td>
<td></td>
</tr>
<tr>
<td>Juicer</td>
<td>(Durand et al., 2016)</td>
<td><a href="https://github.com/theaidenlab/juicer">https://github.com/theaidenlab/juicer</a></td>
<td></td>
</tr>
<tr>
<td>BS-seeker2</td>
<td>(Guo et al., 2013)</td>
<td><a href="https://github.com/BSSeeker/BSseeker2">https://github.com/BSSeeker/BSseeker2</a></td>
<td></td>
</tr>
<tr>
<td>GREAT</td>
<td>(McLean et al., 2010)</td>
<td><a href="http://great.stanford.edu">http://great.stanford.edu</a></td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Contact for Reagent and Resource Sharing**

Requests for reagents and resources should be directed towards the Lead Contact, Harrison Gabel (gabelh@wustl.edu).
2.5 Experimental Model and Subject Details

Mice

MeCP2 knockout mice (B6.129P2(C)-MeCP2tm1.Bird/J) were obtained from The Jackson Laboratory. Female heterozygous mice (MeCP2-/+) were crossed with C57BL/6J male mice to generate hemizygous male knockout mice (MeCP2-/+y) and wild-type male litter mates (MeCP2+/y). MeCP2 overexpression mice (FVB-Tg(MECP2)3Hzo/J) were cryo-recovered from The Jackson Laboratory. Female heterozygous mice (MeCP2Tg3/+) were crossed with FVB/NJ male mice to generate hemizygous male transgenic mice (MeCP2Tg3+y) and wild-type male litter mates (MeCP2+/y). Female Dnmt3aflox were provided by M. Goodell and crossed to male B6.Cg-Tg(Nestin-cre)1Kln/J (Nestin-Cre+) to generate Dnmt3afl/+; Nestin-Cre+. Male Dnmt3afl/+; Nestin-Cre+ were then crossed to female Dnmt3aflox to generate Dnmt3afl/+ Tg(Nestin-cre)1Kln/J conditional knockout mice (DNMT3A Nestin-cKO) (Gabel et al., 2015). To generate Dnmt3afl/+ Tg(Actl6b-cre)4092Jiwu/J conditional knockout mice (DNMT3A Baf53b-cKO), Dnmt3afl/+ were crossed to Tg(Actl6b-cre)4092Jiwu/J (Baf53b-Cre+/+) to generate Dnmt3afl/+; Baf53b-Cre+. Dnmt3afl/+; Baf53b-Cre+ were then crossed to Dnmt3afl/+ producing experimental and control animals for analysis.

Method Details

Topologically associated domain analysis

Topologically associating domains were called using TADtree on interaction matrices from (Dixon et al., 2012), with the following parameters: gamma (Sensitivity vs specificity tradeoff) = 200, M (Number of hierarchical layers) = 1, p and q (minimum scale of interaction shift) = 3 and 12 respectively, and N (maximum number of TADs detected) = 500. To allow for manual model
selection, TADtree also computes TADs for the range of 1:500, in this case. We leveraged this to apply an additional filter for consistency, wherein we selected for TADs that are called in at least 30% of all runs. When we applied this filter, we found that the effects of changing gamma were moderated, and the program performed consistently at many ranges of sensitivity vs specificity. 8-week cortex Hi-C data was obtained as pre-processed observed/expected contact matrices, from http://chromosome.sdsc.edu/mouse/hi-c/download.html (Dixon et al., 2012). For analysis of neocortex Hi-C data from Bonev and colleagues (Bonev et al., 2017), raw FASTQ files were downloaded from GEO. HiC-Pro (Servant et al., 2015) was used to generate contact matrices using the mm10 mouse genome as reference. Juicer (Durand et al., 2016) was then run on the contact matrices to generate Hi-C contact matrices at 1.5kb, 5kb, and 40kb resolutions using KR normalization. Arrowhead domains were then called on the data within Juicer.

For cerebellum analysis, contact domains called from 10kb-resolution Hi-C data were used (Yamada et al., 2019).

**Total and nuclear RNA isolation**

Cerebral cortex was dissected on ice in phosphate buffered saline from 1) MeCP2 knockout and wild-type male litter mates at 7-8 weeks old, 2) MeCP2 overexpression and wild-type male litter mates at 7-10 weeks old, and 3) DNMT3A Baf53b-cKO and control mice at 7-8 weeks old. Total RNA was extracted from 1/16th of a whole cortex using RLT buffer following RNeasy Micro Kit (Qiagen). Nuclear RNA was isolated following a modified version of the protocol described by (Mo et al., 2015). Briefly, half of a cortex was homogenized in 0.25M sucrose, 25mM KCl, 5mM MgCl$_2$, 20mM Tricine-KOH using a glass dounce homogenizer. Nuclei were isolated via centrifugation at 10,000g for 18 minutes at 4°C (Sorvall HB-4) by pelleting through a 30%
iodixanol density gradient (Sigma D1556). RNA was isolated from nuclei by resuspending pellet in RLT buffer following the RNeasy Micro Kit (Qiagen).

**RNA sequencing**

RNA libraries were generated from 250ng total and nuclear RNA with NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB) using a modified amplification protocol (37°C, 15 minutes; 98°C, 30 seconds; (98°C, 10 seconds; 65°C, 30 seconds; 72°C, 30 seconds)x13; 72°C, 5 minutes; 4°C hold. RNA libraries were pooled at a final concentration of 8-10nM and sequenced using Illumina HiSeq 2500 or 3000 with the Genome Technology Access Center (GTAC) at Washington University in St. Louis, typically yielding 20-30 million single-end reads per sample.

**RNA sequencing analysis**

Raw FASTQ files were trimmed with Trim Galore, using a quality filter of 20, then rRNA sequences were filtered out using Bowtie, using rRNA sequences from Mus Musculus obtained from the NCBI sequence database (see Key Resources Table). The unaligned reads from this step were then aligned to mm9 using STAR (Dobin et al., 2013), using the default parameters. Reads mapping to multiple regions in the genome were then filtered out, and uniquely mapping reads were converted to BED files. Intronic and exonic reads were then separated. To do this, splice-site reads were first filtered out of the BED read files, then reads that mapped entirely within exons were added to the splice-site reads to make the exonic read file. All remaining reads that overlapped introns were considered intronic reads. Finally, reads were assigned to genes using bedtools coverage -counts.
For gene annotation we defined a "flattened" list of longest transcript forms for each gene, generated on Ensgene annotations, obtained from the UCSC table browser. For each gene, ensembl IDs were matched up to MGI gene names. Then, for each unique MGI gene name, the most upstream Ensgene TSS and the most downstream TES were taken as that gene's start and stop.

**Chromatin immunoprecipitation protocol**

Cerebral cortex was dissected on ice in phosphate buffered saline from 1) *MeCP2* knockout and wild-type male litter mates at 7-8 weeks old, 2) *MeCP2* overexpression and wild-type male litter mates at 7-10 weeks old, 3) DNMT3A Baf53b-cKO and control mice at 7-8 weeks old. The tissue was flash-frozen in liquid nitrogen and stored at -80°C. ChIP experiments were performed on half a cortex as previously described (Cohen et al., 2011a), using an alternative chromatin fragmentation method. Chromatin were fragmented with Covaris E220 sonicator (5% Duty Factory, 140 Peak Incidence Power, 200 cycles per burst, milliTUBE 1mL AFA Fiber). ChIP was performed with H3K27ac (0.025-0.1µg; Abcam ab4729), H3K4me3 (2µg; Abcam ab1012), H3K36me3 (0.2µg; Active Motif 61101). ChIP libraries for H3K27ac, H3K4me3, and H3K36me3 were generated using Ovation Ultralow Library System V2 (NuGEN). Libraries were pooled to a final concentration of 8-10nM and sequenced using Illumina HiSeq 3000 with GTAC, yielding 15-30 million single-end reads per sample.

**Chromatin immunoprecipitation analysis**

Sequenced reads were mapped to the mm9 genome using bowtie2 alignment, and reads were extended based on library sizes and deduplicated to consolidate PCR duplicate reads. Deduplicated reads were used to quantify read density normalized by the number of reads per sample and by
read length in basepairs. Bedtools coverage –counts was used to quantify ChIP signal at the transcriptional start site (TSS), gene body (GB), and transcriptional end site (TES). For consistency with methylation analysis, the TSS was defined as a 1kb region surrounding the TSS (+/-500bp), the GB was defined as 3kb downstream of the TSS to the end of the transcript, and the TES was defined as 2kb upstream through 3kb downstream of the end of the transcript. edgeR was then used to determine differential ChIP-signal across genotypes.

Whole-genome bisulfite sequencing

Cerebral cortex was dissected from DNMT3A Nestin-cKO or DNMT3A Baf53-cKO and controls at 7-8 weeks, flash-frozen in liquid nitrogen, and stored at -80°C. DNA extraction and bisulfite conversion were performed as in (Gabel et al., 2015). Briefly, genomic DNA was extracted and bisulfite libraries were generated using the Ovation Ultralow Methyl-Seq Library System (NuGEN). Libraries were pooled and sequenced using Illumina MiSeq 2x150 with the Spike-In Cooperative at Washington University in St. Louis.

Chromatin conformation capture

3C assays were adapted from previously described procedures (Kim and Dekker, 2018; Lieberman-Aiden et al., 2009; Yamada et al., 2019). Half of a cerebral cortex was dounced 10x with a loose pestle in cross-linking buffer (50mM Hepes-KOH, pH 7.9; 100mM NaCl; 1mM EDTA; 0.5mM EGTA) with 1% formaldehyde at room temperature for 15 minutes. Formaldehyde was quenched with 125mM glycine and incubated for 5 minutes at room temperature. The suspension was pelleted at 1150xg for 5 minutes at 4°C and the pellet was subsequently washed with PBS and frozen at -80°C. The frozen pellet was thawed and lysed (10mM Tris-HCl, pH 8;
10mM NaCl, 0.2% IGEPAL-630, Protease Inhibitor (Complete Protease Inhibitor Cocktail, Roche)) for 15 minutes followed by douncing 25x with a loose pestle on ice. The lysed cells were strained through a 70um cell strainer (Falcon) and pelleted at 200xg for 5 minutes at 4°C. Nuclei were resuspended in 1mL of 0.5% SDS and incubated at 62°C for 10 minutes to permeabilize the nuclei. 25ul (approx. 90-100k nuclei) were used for the 3C library construction, and SDS was quenched with Triton X-100 (1% Triton, 1x NEB 3.1 Buffer). Nuclei were digested overnight at 37°C with 200U of BglII. The digested nuclei were incubated at 65°C for 15 minutes to inactivate the enzyme. Samples were then ligated for 6 hours at 16°C with 4000U of T4 DNA ligase (NEB Ligase Buffer, 0.1mg/ml BSA, Triton X-100, NEB T4 ligase). Digested and ligated 3C samples were pelleted at 3500rpm for 5 minutes at 4°C and resuspended in 200ul NEB 3.1 Buffer and brought to 1% SDS, 250mM NaCl. 3C libraries were de-crosslinked overnight at 65°C. They were then incubated with 40ug RNase A at 37°C for 1 hour, followed by 80ug of Proteinase K at 55°C for 1 hour. The samples were purified with Phenol/Chloroform/Isoamyl Alcohol (25:24:1) and subsequently ethanol precipitated (100mM Sodium Acetate) overnight at -20°C. Control libraries were generated from Bacterial Artificial Chromosomes (BACs) as previously published (Kim and Dekker, 2018). BACs (see STAR methods) were isolated using PureLink HiPure Plasmid Maxiprep Kit (Invitrogen), following the modified protocol for BAC isolation.

Quantification and Statistical Analysis

Whole-genome bisulfite analysis

Bisulfite-sequencing cannot distinguish between hydroxymethylation and methylation at cytosines, detecting both only as modified sites during sequencing. Thus, measures of methylation included in this study represent the aggregate of both forms of methylation at sites across the
genome. Bisulfite data for the DNMT3A Nestin-cKO and DNMT3A Baf53b-cKO, and data obtained from GEO as FASTQ files, were adapter-trimmed, mapped to mm9, then deduplicated and called for methylation using BS-seeker2 (with bowtie2). Nonconversion rate was set to be (--XS=.3,2), and default settings were used otherwise. The methylation levels for genes and regions was assessed by summing up the number of reads mapping to Cs that supported mC and dividing that by the number of reads mapping to Cs that supported non-mC, using bedtools map -o sum. This allows sites with more read information to contribute more in determining the methylation level of the surrounding region. In order to avoid confounding effects of promoter-associated depletion of methylation, genic methylation was assessed 3kb downstream of promoters, to the TES. In order to assess any potential for C site coverage bias influencing TAD-associated Bisulfite-seq results, a bed file of every C/G in the genome was generated, and bedtools coverage -counts was run on it. Bisulfite coverage bias was assessed on informative sites for methylation with bedtools map -o mean on the number of reads mapping to Cs in the region.

**RNA-sequencing quantification**

MeCP2-repressed and activated genes were identified by quantitative analysis of exonic reads from total RNA from the MeCP2 KO and OE. For this analysis we applied an approach similar to previous studies of other brain regions that combined results of gene expression in these two strains in order to identify the most-robustly MeCP2-regulated genes (Ben-Shachar et al., 2009; Chahrour et al., 2008; Chen et al., 2015). DESeq2 was run using default parameters on exonic reads from MeCP2 KO and their littermate control animals (n=6 per genotype). Separately, we ran DESeq2 analysis on exonic reads from MeCP2 OE and their littermate control animals (n=5). The nominal p-values output by DESeq2 for each gene in each mutant-control comparison were then combined
using the Fisher method (log-sum). The resulting combined p-values were then Benjamini-Hochberg corrected, and genes with a q-value < .1 and a log2 fold-change > 0 in the KO and a log2 fold-change < 0 in the OE were labeled as MeCP2-repressed, while genes with a q-value < .1 and a log2 fold-change < 0 in the KO and a log2 fold-change > 0 in the OE were labeled as MeCP2-activated. Notably, similar results for enrichment of mCA, and gene length were observed when examining lists of genes called as significantly dysregulated in the MeCP2 KO and OE on their own. However, a smaller gene list was identified in each independent analysis, likely due to the reduced statistical power. For comparison between changes in intronic RNA and changes in exonic RNA (Figure 3; S1) a list of significantly changed genes in the MeCP2 KO compared to littermate controls based on intronic fold-changes was generated by applying the same DESeq2 analysis to intronic reads derived from total RNA-seq of nuclear RNA.

RNA-seq aggregate plots examining changes in expression over lengths of genes (Figure S4G-I) were performed by binning genes into 1kb windows, then calculating nuclear intronic coverage over each bin using bedtools coverage -hist. Each gene was then normalized by the median amount of combined coverage from the MeCP2 KO and wild-type. Finally, genes were aligned by their TSS's, and median expression levels were plotted for each bin. For any given graph, the genes are filtered such that the lengths of genes plotted is equal to or greater than the aggregate length being plotted (Boswell et al., 2017; Gray et al., 2014). Termination ratios (adapted from (Boswell et al., 2017)) were calculated in a manner similar to the aggregate plots of expression. Windows were made for every gene longer than 100kb, from the TSS to 25kb downstream of the TSS (window A), and from the TES to 25kb upstream of the TES (window B). Coverage over these windows was calculated with bedtools coverage -hist, then coverage of window A was divided by the
coverage of window B, within each sample. Then, wild type/MeCP2 KO samples were divided against each other, and a normalized metric for each wild type/MeCP2 KO replicate was generated. Finally, median metrics were plotted for each wild type/MeCP2 KO replicate. Simulated metrics were generated by progressively applying exonic fold-changes to the wild-type expression throughout the gene. For example, a gene 50kb long, with an observed exonic log2 fold-change (MeCP2 KO/wild type) of -1 would be divided into 50 1kb bins. A simulated KO expression pattern would then be generated by progressing down each of the 50 bins, applying a fold-change of 1/50 * -1, 2/50 * -1, etc, to the expression of the corresponding bin in the WT.

**TAD boundary analysis**

For Figures 1F, 2C, S2B, and S2F, all TADs that intersected MeCP2-repressed, MeCP2-activated, and unchanged genes were selected, boundaries phased, and values for mC/C, coverage, GC percent, or ChIP/Input plotted in aggregate using R, python, and bedtools. Shuffling analysis was performed in these cases by moving each TAD randomly around their target genes (TADs containing MeCP2-repressed genes were shuffled around MeCP2-repressed genes, etc.). This was done once to generate a single example (e.g. Figure 1F), or up to 20 times to generate a resampling ribbon (e.g. Figure S2B). Figures 1G and S2K were generated by selecting boundaries through the following method: TADs were assigned a score, based on the difference between its mCA level and the mCA level of the next downstream TAD. Then, the top 33% top scoring TADs were taken and plotted in aggregate analysis. For shuffled TAD control plots, TAD locations were randomized and the selection process for these shuffled TADs was repeated. We note that because both true TADs and shuffled TADs in these plots were selected based on high differences in average methylation between them, a reduction of signal going from left to right is predicted. However,
true TADs display two aspects of mCA/CA signal that are not present in shuffled TADs, which demonstrate the organizing effects associated with TAD boundaries: 1) True TADs are more highly enriched for mCA, showing that resampling TADs eliminates an organized enrichment of mCA/CA in specific regions of the genome. 2) A steep step-down in mCA/CA occurs at true TAD boundaries, while a more gradual fall off in signal occurs at reshuffled boundaries. These two differences illustrate that ~30% shifts in mCA/CA levels are common between adjacent TADs, and they demonstrate that TAD boundaries delineate a sharp transition between mCA/CA levels at these regions. Plotted in all figures is the average of the upstream and downstream TAD boundaries, taking into account TAD orientation. Upstream boundaries (where the TAD would be on the right, instead of the left), have their orientation flipped horizontally along the boundary. The value plotted is the mean of the flipped upstream boundaries and the unflipped downstream boundaries.

**TAD methylation cross-correlation**

Cross-correlation matrices (e.g. Figure 1H) were generated by dividing each domain into 10 equally-sized bins, then prepending and appending 10 identically sized bins up and down-stream of the domain, making a number of domains x 30 matrix. Each column of this matrix was then correlated against each other column, making a 30 x 30 correlation matrix, which was plotted in heatmap form. Shuffled TADs were generated as a negative control by randomly placing TAD-sized regions around the genome, separated from each other by similar distances as actual TADs to retain TAD structure.
In order to calculate heatmaps of correlations between mCA/CA levels of genomic elements in and outside of TADs and mean TAD mCA/CA levels (Figure S3J), TADs were again divided into 10 equally sized bins, with equal-sized regions placed upstream and downstream the domain. Enhancers, genes, and TSSs in the genome were then intersected with these regions, and spearman correlations between TAD methylation (subtracting out the methylation of the element if necessary) and element methylation within each region were calculated.

To assess the similarity of mCA levels between individual elements inside and outside of the same TAD (i.e. TSS regions, enhancers, gene bodies; Figure S3K), each element was paired to each other element on the same chromosome. Each pair was then assessed if they paired within, or between TADs. Because mCA varies with genomic distance, each intra-TAD pair was matched to the extra-TAD pair with the most similar distance between elements, and Spearman correlations were calculated on the two distance-matched sets.

**Identification of enhancers**

Bed files of H3K27ac and H3K4me3 ChIP-seq were pooled by replicate. Peaks of H3K27ac and H3K4me3 ChIP-seq were identified using the MACS2 peak calling algorithm on the pooled bed files using the pooled ChIP Input as background signal (macs2 callpeak --nomodel -q 0.05). MeCP2 KO and wild type peak files were then combined using bedtools unionbedg, and overlapping peaks were merged into single peaks using bedtools merge. Bedtools intersect was used to identify H3K27ac peaks that did not overlap with gene promoter regions (1kb around annotated TSS) or with H3K4me3 peaks from MACS2. These non-overlapping H3K27ac peaks were then further filtered for landing within an H3K4me1 peak, as called in the ENCODE-
generated broadpeak file for H3K4me1 ChIP from 8-week old cortex. All H3K27ac peaks that remained after these rounds of filtering were defined as enhancers.

To identify the enhancers most robustly regulated by MeCP2, we used combined analysis of the MeCP2 KO and OE, similar to the approach used for determining gene-expression changes. We ran differential ChIP-seq analysis on H3K27ac from the MeCP2 KO and their littermate control animals (n=5), and from the MeCP2 OE and their littermate controls (n=3). Reads were quantified in all merged acetyl peaks, and edgeR was used to calculate nominal p-values and fold-changes for these peaks. These p-values were then combined using the Fisher method (log-sum), and were Benjamini-Hochberg corrected. Acetyl peaks with a combined q-value < 0.1, and a log2 fold-change > 0 in the KO and a log2 fold-change < 0 in the OE were called as MeCP2-repressed peaks, while peaks with a combined q-value < 0.1, and a log2 fold-change < 0 in the KO and a log2 fold-change > 0 in the OE were called as MeCP2-activated peaks.

Enhancers were also called as misregulated in the DNMT3a Baf53b-cKO, using edgeR. H3K27ac ChIP-seq reads from the DNMT3A Baf53-cKO and control (n=6) were quantified in the merged acetyl peaks called from the MeCP2 KO/control. edgeR was then run on these regions, and peaks with a q-value < 0.1 were called as misregulated in the DNMT3a Baf53-cKO.

**ATAC peak analysis**

To sensitively detect sub-peak-threshold histone acetylation signal that could correspond to putative regulatory elements, a compendium of all ATAC peaks detected in the genome was obtained from http://atlas.gs.washington.edu/mouse-atac/data/
(atac_matrix.binary.qc_filtered.peaks.txt) and acetylation was quantified in these regions. ATAC peaks which landed within 1kb of a TSS were filtered out, and methylation and acetylation analysis was performed on them as described for enhancers. Equal-sized control regions for enhancers and ATAC peaks were generated through a structured resampling approach. Enhancers and ATAC peaks, if located within a gene, are shuffled within that gene. If the peak is extragenic, it is shuffled between the nearest upstream and downstream enhancer/gene/ATAC peak. In each case, enhancers and ATAC peaks are restricted from landing within existing enhancers and ATAC peaks. Normalized pseudocounts of acetylation for enhancers, ATAC peaks, and resampled ATAC peaks were generated from edgeR common dispersions, running on each dataset and mouse strain separately.

**Associations between enhancers and genes using Hi-C and GREAT**

GREAT 3.0 (McLean et al., 2010) was used to determine potential enhancer-promoter interactions. NCBI build 37 of Mus Musculus was used as the species assembly, and bed files of enhancer regions were uploaded into the web tool. Enhancer-promoter interactions were identified for MeCP2-repressed, MeCP2-activated, and all other enhancers. All identified genes linked to these enhancers were used for further analysis. To link enhancers to promoters by Hi-C, intrachromosomal Hi-C matrices (KR-normalized) were extracted at 1.5 kb resolution using Straw (Durand et al., 2016), and mean average interactions for all distances were calculated per-chromosome. Each matrix was then filtered down to only interactions between promoters and enhancers within 3mb of each other, and enhancers-promoter pairs with greater than 3 observed interactions and an observed/expected ratio over 1.5 were linked, making Hi-C-linked enhancers.
Enhancers that were linked to promoters of MeCP2-repressed or MeCP2-activated genes were then used in analysis in Figure 6B.

For analysis of intragenic versus extragenic contact frequencies (Figure 6G), the same normalized contact matrices were analyzed. Genes greater than 50kb were extended on either side by their respective gene lengths, and contacts were mapped to them using bedtools intersect -wao -F 1. Each region was then split into 60 equally-sized bins (20 upstream, 20 intragenic, 20 downstream), making 3600 possible regions of interactions. Average interaction frequencies within these regions were calculated from the interaction matrix, and each gene’s intragenic and extragenic interactions were aggregated by calculating the mean of each bin of interaction.

For Figure 6H, intragenic and extragenic enhancer-promoter interactions were distance matched to control for the greater number and variability of extragenic interactions. To do this, a similar resampling approach to Figure S3K was used: for each intragenic interaction, an extragenic interaction with a similar distance was selected for comparison.

To assess the relationship between enhancer acetylation fold-change and gene fold-change (Figure S5M), partial correlations between enhancer/control region acetyl fold-change and gene fold-change were calculated by averaging the acetyl fold-changes of elements within the gene and correlating this aggregate value to that gene’s fold-change.

**Quantitative 3C analysis**
Quantitative PCR was employed, as previously described (Joo et al., 2016; Schaukowitch et al., 2014; Tolhuis et al., 2002), to determine enhancer-promoter interactions of enhancers using the primers listed (see STAR Methods). Negative regions were selected from genomic restriction fragments that did not contain detectable enhancer sequences and were located nearby to the enhancer being tested but in closer proximity to the anchor TSS. Relative concentrations of enhancers and corresponding negative regions were calculated from a standard curve of BAC 3C libraries containing the targeted loci. Enrichments of interactions were then calculated as the relative concentration of targeted enhancers divided by the relative concentration of the nearby negative region.

**mC context enrichment analysis**

Trinucleotide contexts and methylation status (a rational number determined through bisulfite coverage - # of Cs mapping to site / # of Ts mapping to site) were determined for each cytosine and guanine in the genome, and then sites were assigned to enhancers based on proximity (a site was assigned to an enhancer if it was within 1 kb of the enhancer’s center). T-tests were then run for each context, summing up methyl-weighted or unweighted (for total mC per kb / C sites per kb respectively), or averaging methylation statuses (for percent mC), comparing test sets (up/downregulated enhancers, or enhancers within certain genes) to a set of resampled control enhancers with similar acetylation levels 5 times the size of the test set.

**mC vs acetylation change local correlation analysis at enhancers**

For both enhancers and gene local correlation analysis, 1 kb-sized bins were assessed for average methylation around and within enhancers/genes, making an N x M sized matrix of methylation,
where \( N = \# \) of enhancers/genes, and \( M = \# \) of bins. Each column of this matrix was then correlated against the matching enhancer/gene’s fold-change in the MeCP2 KO/MeCP2 OE. For enhancers, correlation analysis was centered on summits of wild-type acetylation and MeCP2 KO acetylation and analyzed separately through MACS2.

**Running-average plots**

Running-average plots of mCA/CA and genic fold-change, as well as total mC sites and genic fold-change were generated from means of 201 gene bins, with a 1 gene step, using the `rollMean` command in the zoo package of R. Length-controlled resampling was performed by selecting a gene within \( .75 – 1.25 \) times the length of each gene in the test set.

**Data and Software Availability**

All genomic data generated in this study have been uploaded to the NCBI GEO archive GSE123373.
2.6 Figures
2.6.1 Figure 1. Regional non-CG methylation patterns are associated with domains of chromatin folding.

A. Left, heatmap of changes in exonic RNA for genes detected as significantly dysregulated (FDR <0.1) in combined analysis of total RNA-seq from MeCP2 KO and MeCP2 OE cerebral cortex (grey indicates no data for replicate). Right, fold-change of exonic RNA in MeCP2 KO or OE vs wild-type control for an example MeCP2-repressed gene, *Sdk1*. ***, Benjamini-Hochberg adjusted pvalue < 10⁻⁸ DESeq2 Wald test.

B. Aggregate plot of cerebral cortex mCA/CA levels for MeCP2-repressed, MeCP2-activated, and all other genes. Mean mCA/CA for 1kb bins in the TSS and regions surrounding genes is shown at kilobase (left) and megabase (right) scale. For “Metagene” region, mean mCA/CA was calculated for 50 equally-sized bins within the gene bodies.

C. DNA methylation levels around the *Sdk1* gene as assessed by bisulfite-sequencing of the cerebral cortex.

D. Hi-C interaction matrix data from cerebral cortex for a genomic region including two MeCP2-repressed genes, *Sdk1* and *Auts2* (red), above mCA/CA and mCG/CG levels. Blue highlights indicate regions within TAD-like structures that are visible in the interaction matrix.

E. A boxplot of mCA/CA levels of TADs that contain MeCP2-repressed genes or MeCP2-activated genes, or TADs that do not contain dysregulated genes. Notches estimate the 95% confidence interval for the median (see methods). ***, p < 10⁻⁸ Wilcoxon rank sum test.

F. Aggregate plot of mean mCA/CA for 4kb bins, phased on boundaries of TADs that contain MeCP2-repressed genes. Red, analysis of true TADs; pink, analysis repeated on shuffled TAD boundaries (see methods). Drop off in mCA/CA at boundaries is significantly different from TAD boundaries shuffled around MeCP2-repressed genes, p < 10⁻³ (see Figure S2C, methods).
G. Aggregate plot of mean mCA/CA phased on boundaries that contain the top 33% most differential levels of mCA/CA between TADs that contain MeCP2-repressed genes and neighboring TADs. Red, true TADs; pink, shuffled TAD boundaries.

H. Cross correlation analysis of mCA/CA signal within and outside of TADs. mCA/CA values were calculated for 10 intra-domain regions and 10 equally-sized regions up and downstream of each TAD. Correlation between mCA/CA in these regions across TADs containing MeCP2-repressed genes, all genes, and shuffled control TADs is shown.

Data from cerebral cortex of 7-10 week old animals. n=6 per genotype (MeCP2 KO, wild type) and n=5 per genotype (MeCP2 OE, wild type) for RNA-seq, n=2 wild type for DNA methylation (Stroud et al., 2017). Hi-C data was obtained as contact matrices, from http://chromosome.sdsc.edu/mouse/hi-c/download.html.
2.6.2 Figure 2. Domain-associated recruitment of DNMT3A defines megabase- and kilobase-scale mCA levels.

A. Cross correlation within and outside of TADs for DNMT3A ChIP signal from the cerebral cortex at 2 weeks of age. DNMT3A ChIP/Input values were calculated for 10 intra-domain regions and 10 equally-sized regions up and downstream of each TAD. Correlation between these regions for all TADs is shown (see methods).

B. Density scatter plot of DNMT3A ChIP/Input signal at 2 weeks of age and mCA/CA levels at 8 weeks of age for each TAD. Spearman rho shown for correlation between DNMT3A signal and mCA/CA levels within TADs. ***, p < 10^{-8}.

C. Aggregate plots of ChIP/Input for DNMT3A at 2 weeks phased on boundaries of TADs that contain MeCP2-repressed genes. Black line and ribbon for each plot indicates the mean and standard deviation of 20 sets of resampled boundaries generated by shuffling TADs in the genome (see methods).

D. Boxplot of MeCP2 ChIP/Input signal within TADs, split up by quartile of TAD mCA/CA. Spearman correlation shows significant association between these signals. ***, p < 10^{-8}.
E. Density scatter plots of TAD mCA/CA levels and mCA/CA levels at various kilobase-scale genomic elements in the cerebral cortex at 8 weeks of age. Values for every TSS, gene body, or enhancer in the genome are plotted against the value for the TAD in which the element resides. Data is subtracted from TADs such that element methylation does not contribute to plotted TAD methylation. Spearman rho shown for the mCA/CA levels of each genomic element vs the mCA/CA level of the TAD that the element is in. ***, p < 10^{-8}.

Data from cerebral cortex of 2 or 8 week old animals. n=2-3 per time point for DNMT3A ChIP- seq, n=2 wild type per time point for DNA methylation (Stroud et al., 2017).
2.6.3 Figure 3. Loss of MeCP2 leads to promoter-associated transcriptional activation of MeCP2-repressed genes.

A. Left, genome browser view of nuclear total RNA-seq coverage, as well as H3K36me3, H3K4me3, and H3K27ac ChIP-seq signals from the MeCP2 KO and wild type at an example
MeCP2-repressed gene, Zmat4. Right, close-up view of overlaid MeCP2 KO and wild type aggregate signal illustrating subtle increases in RNA-seq and ChIP-seq signal in regions indicated in blue at left.

B. Boxplot of fold-changes in gene expression for MeCP2 KO versus wild type for MeCP2-repressed, MeCP2-activated, and all other genes defined in combined analysis of total RNA from MeCP2 KO and MeCP2 OE versus wild type cerebral cortex (see Figure 1A). Shown are fold changes as measured by RNA-seq quantification of exonic regions from whole cortex RNA (Exonic) or intronic regions of genes for RNA from isolated nuclei (Intronic). ***, p < 10^-8 Wilcoxon rank-sum test.

C. Profile of intronic expression and median fold-change for the first 100kb of genes identified as upregulated in intronic RNA analysis of MeCP2 KO versus wild type cortex (Figure S4B). Normalized median expression of one hundred 1kb bins in the wild type and MeCP2 KO is plotted for genes >100kb (see methods). The fold-change of the median expression of the bins between the MeCP2 KO versus wild type is shown in orange. The consistent differences between MeCP2 KO and wild type across the length of these genes suggests there is promoter-associated upregulation of transcription for these genes in the MeCP2 KO (see also Figure S4G,H).

D. Boxplots of mean fold-changes in ChIP signal for MeCP2 KO versus wild type at indicated regions of MeCP2-repressed, MeCP2-activated, and all other genes (defined by analysis of intronic RNA sequencing from MeCP2 KO and wild type cortex). Value for Zmat4 gene shown in panel A is indicated as a point on each plot. ** p < 10^-3; ***, p < 10^-8 Wilcoxon rank-sum test.
E. Boxplots of fold-changes in gene expression for MeCP2 OE versus wild type for gene sets defined in combined analysis of total RNA from MeCP2 KO and MeCP2 OE versus wild type cerebral cortex (see Figure 1A). Fold-changes of reads corresponding to exonic or intronic RNA are shown. ***, p < 10^{-8} Wilcoxon rank-sum test.

F. Profile of intronic expression and fold-change in the MeCP2 OE versus wild type as in 3C. The consistent differences between MeCP2 OE and wild type across the length of these genes suggests there is promoter-associated downregulation of transcription for these genes in the MeCP2 OE (see also Figure S4I).

G. Boxplots of mean fold-changes in ChIP-signal for MeCP2 OE versus wild type as in panel D.
   *, p < 0.05 ; **, p<10^{-3} ; ***, p < 10^{-8} Wilcoxon rank-sum test.

Data generated from cerebral cortex of 7-10 week old animals. For MeCP2 KO analysis: n=6 per genotype for RNA-seq, n=3 per genotype for H3K36me3, n=4 per genotype for H3K4me3, n=5 per genotype for H3K27ac. For MeCP2 OE analysis: n= 5 per genotype for RNA-seq, n=2 per genotype for H3K36me3, n=2 per genotype for H3K4me3, n=3 per genotype for H3K27ac.
2.6.4 Figure 4. MeCP2 represses enhancers that are enriched for mCA and mCG binding sites.

A. Left, genome browser view of overlaid wild type and MeCP2 KO aggregate H3K27ac ChIP-seq signal, H3K4me1 ChIP-seq peaks, and H3K27ac peaks called as enhancers at two example MeCP2-repressed genes, Zmat4 and Efna5. Right, relative scale and close-up view of enhancers indicated in blue at left.

B. Heatmap of per replicate fold-changes of H3K27ac ChIP signal between MeCP2 KO and MeCP2 OE, for enhancers identified as significantly changed (FDR < 0.1) in combined analysis of H3K27ac ChIP-seq from MeCP2 KO and MeCP2 OE.
C. Aggregate plots of mCA/CA or mCG/CG levels (left) or number of mCA and mCG sites per kilobase (right) of MeCP2-repressed, MeCP2-activated, and all other enhancers. Plots are centered at the midpoint for each enhancer. Average values for 100 bp bins are shown.

D. Heatmap of enrichment significance for methylated trinucleotide sites in MeCP2-repressed and MeCP2-activated enhancers. Significance of enrichment or depletion was calculated by comparing mC sites per kb in the enhancers to resampled sets of enhancers that are not significantly changed (see methods).

E. Boxplots of MeCP2 ChIP/Input signal in MeCP2-repressed, MeCP2-activated, and all other enhancers. ***, p < 10^{-8} Wilcoxon rank-sum test.

F. Boxplots of fold-changes of H3K27ac ChIP signal in MeCP2 KO versus wild type across deciles of mCA/kb (left) and mCG/kb (right) for all enhancers genome-wide. Spearman rho shown for correlations of each methylation mark with the change in H3K27ac at enhancers. ***, p < 10^{-8}.

G. Boxplots of fold-changes of H3K27ac ChIP signal in MeCP2 OE versus wild type for all enhancers across deciles of mCA/kb (left) and mCG/kb (right). Spearman rho shown for the correlation of each methylation mark with the change in H3K27ac at enhancers. ***, p < 10^{-8}.

H. Plot of spearman correlations between H3K27ac fold-change at enhancers in the MeCP2 KO and mC/kb for 500bp bins in and around the enhancers. Plots are centered at the summit of enhancer H3K27ac ChIP peaks.

I. Plot of spearman correlations for H3K27ac fold-change at enhancers in the MeCP2 OE and mC/kb for 500bp bins in and around the enhancers. Analysis as described in panel H.
Data from cerebral cortex of 7-10 week old animals. For MeCP2 KO: n=5 per genotype for H3K27ac. For MeCP2 OE: n=3 per genotype for H3K27ac. n=2 wild type for DNA methylation (Stroud et al., 2017) and MeCP2 ChIP-seq data (Kinde et al., 2016).
2.6.5 Figure 5. mCA-associated de-repression of enhancers in DNMT3A

Baf53b-cKO.

A. Quantification of mCG/CG and mCA/CA levels in enhancers genome-wide for wild type and DNMT3a Baf53b-cKO cerebral cortex. Error bars = SEM, **, p < 10^{-3}, ***, p < 10^{-8} two-tailed t-test.

B. Fold-change for mRNA of MeCP2-regulated genes (left) and H3K27ac for MeCP2-regulated enhancers (right) in MeCP2 KO and DNMT3A Baf53b-cKO mice. ***, p < 10^{-8} Wilcoxon rank-sum test.
C. Profiles of mC/C and mC/kb for enhancers detected as significantly dysregulated in the DNMT3A Baf53b-cKO (see Figure S6D).

D. Boxplots of fold-changes of H3K27ac ChIP signal in DNMT3A Baf53b-cKO versus wild type across deciles of mCA/kb (left) and mCG/kb (right) for all enhancers genome-wide. Spearman rho shown for correlations of each methylation mark with the change in H3K27ac at enhancers. * p < 0.05, ***, p < 10^-8.

Data from cerebral cortex of 7-10 week old animals. n=6 per genotype for DNA methylation, n=4 per genotype for RNA-seq, n= 6 per genotype for H3K27ac ChIP-seq.
2.6.6 Figure 6. MeCP2-repressed enhancers are linked to MeCP2-repressed genes.

A. Boxplot of enhancer mCA/kb for enhancers within TADs containing MeCP2-repressed genes, MeCP2-activated genes, or no dysregulated genes. **, p < 10^{-3}, ***, p < 10^{-8} Wilcoxon rank-sum test.

B. Heatmap of enrichment for the associations between MeCP2-repressed and MeCP2-activated enhancers with MeCP2-repressed and MeCP2-activated genes. Comparisons are for enhancers within the same TAD as MeCP2-repressed/activated genes (“Same TAD”), assigned to MeCP2-repressed/activated genes by GREAT analysis (McLean et al., 2010) (“GREAT”), enriched for enhancer promoter contacts by Hi-C (Bonev et al., 2017) (“Hi-C-linked”), or found within MeCP2-repressed/activated genes (“Intragenic”). Significance of enrichment or depletion was calculated from the median p-value of multiple Fisher’s exact tests on 1000 sets of resampled, acetylation-matched control enhancers (see methods). Value shown for each comparison is the median log2 enrichment (odds ratio) for the overlap of true MeCP2-regulated enhancers compared to the overlap of control enhancers.

C. Barplots of log2 fold-changes of H3K27ac ChIP-signal in MeCP2 KO versus wild type for enhancers near MeCP2-repressed genes, Zmat4 and Auts2. TAD regions overlapping the MeCP2-repressed genes are indicated in blue above. Red bars, enhancers within MeCP2-repressed gene; pink bars, other enhancers within the same TAD as MeCP2-repressed gene; gray and black bars, extragenic and intragenic enhancers in other TADs. *, indicates enhancers identified as significantly changed in genome-wide analysis (FDR < 0.1).

D. Boxplots of fold-changes in H3K27ac ChIP signal in the MeCP2 KO (top) and the MeCP2 OE (bottom) for enhancers within TADs that contain MeCP2-repressed, MeCP2-activated genes,
or no dysregulated genes ("Same TAD", left), and for enhancers found within these genes ("Intragenic", right). ***, p < 10^{-8} Wilcoxon rank-sum test.

E. Pie charts displaying intragenic and extragenic distributions of MeCP2-repressed, MeCP2-activated, and all other enhancers. **, p < 10^{-3}, ***, p < 10^{-8} chi-squared test comparing distribution of dysregulated enhancers to distribution of enhancers that are not significantly changed.

F. Boxplots of fold-changes of H3K27ac ChIP signal in MeCP2 KO (top) or MeCP2 OE (bottom) versus wild type for intragenic and extragenic enhancers across deciles of enhancer mCA/kb (left) and mCG/kb (right). Spearman rho shown for correlations of each methylation mark in enhancers and their change in H3K27ac. ***, p < 10^{-8}.

G. Heatmap of aggregate observed/expected Hi-C interaction frequencies for regions inside and outside of genes, genome-wide (see methods).

H. Boxplot of observed/expected Hi-C interactions detected for enhancers and promoters genome-wide, comparing enhancers found within (Intragenic) and outside (Extragenic) of genes (see methods). To facilitate comparison, distances of enhancer-promoter interactions were controlled for (see methods). ***, p<10^{-8} Wilcoxon rank-sum test.

I. Running-average plot of fold-change in intronic RNA expression in the MeCP2 KO (top) and MeCP2 OE (bottom) versus mCA/kb for genes containing various numbers of enhancers. Mean fold-changes are plotted for genes sorted by level of gene body mCA/kb. Mean shown for bins of 201 genes with a 1-gene step.

Data from cerebral cortex of 7-10 week old animals. For MeCP2 KO: n=6 per genotype for RNA-seq, n=5 per genotype for H3K27ac. For MeCP2 OE: n=3 per genotype for H3K27ac, n=5 per
genotype for RNA-seq. n=2 wild type for DNA methylation (Stroud et al., 2017). Hi-C interaction data from neurons isolated from E14.5 cortex (Bonev et al., 2017).
2.6.7 Figure 7. TAD-associated mCA dictates enhancer repression by MeCP2.

A. During early postnatal development, TAD structures in neurons dictate DNMT3A activity leading to the establishment of higher or lower mCA set-points within each TAD.

B. In mature neurons, MeCP2 reads-out this methylation, repressing histone acetylation at enhancer elements. MeCP2 most strongly represses highly-methylated intragenic enhancers, resulting in repression of genes that contain multiple enhancers and lie within high mCA TADs. Upon disruption of MeCP2 or loss of mCA, dysregulation of intragenic enhancer activity leads to altered expression of these genes.
2.6.7 Figure S1. Identification of MeCP2-regulated genes in the cerebral cortex by combined RNA-seq analysis of MeCP2 KO and MeCP2 OE. Related to Figure 1.

A. Venn diagram of the overlap between genes identified as MeCP2-repressed and MeCP2-activated in combined analysis of MeCP2 KO and OE versus wild type cerebral cortex and genes previously identified as consistently dysregulated across multiple datasets and brain regions (Gabel et al., 2015). *** p < 10^{-8} hypergeometric test. Analysis was performed using coding genes found in annotation sets of both studies (see methods).
B. Boxplot of fold-changes in exonic RNA in the cortex of MeCP2 KO and wild type mice, for genes previously identified as consistently dysregulated across multiple datasets and brain regions (Gabel et al., 2015). ***, p < 10^{-8} Wilcoxon rank-sum test.

C. Boxplots of mCA/CA and mCG/CG levels at the TSS (left) and gene bodies (right) of MeCP2-repressed, MeCP2-activated, and all other genes. Light colored boxplots are plots generated for a control set of genes matched for the distribution of gene expression for each dysregulated gene set (see methods). * p < 0.05 ; **, p <10^{-3}; ***, p < 10^{-8} Medians of Wilcoxon rank-sum tests on 100 gene-set resamplings.

D. Boxplot of gene lengths for MeCP2-repressed, MeCP2-activated, and all other genes identified in combined analysis of total RNA-seq from MeCP2 KO and OE cerebral cortex. ***, p<10^{-8} Wilcoxon rank-sum test.

E. Running average plots of exonic RNA fold-changes versus gene length, for MeCP2 KO (left) and OE (right) versus wild type. Mean fold-changes are plotted for bins of 201 genes sorted by gene length with a 1-gene step (see methods).

F. Running average plots of exonic RNA fold-change vs gene body mCA/CA for the MeCP2 KO (left) and OE (right) versus wild type. mCA/CA levels are calculated for the gene body defined as +3kb from the TSS to the TES. Mean fold-changes are plotted for bins of 201 genes sorted by mean mCA/CA per gene with a 1-gene step (see methods).

G. Plot of genome-wide correlations between methylation levels for 1kb regions in and around genes and fold-changes in exonic RNA expression in the MeCP2 KO versus wild type for each gene.
H. Plot of genome-wide correlations between methylation levels for 1kb regions in and around genes and fold-changes in exonic RNA expression in the MeCP2 OE versus wild type for each gene.

Data from cerebral cortex of 7-10 week old animals. n=6 per genotype for RNA-seq (MeCP2 KO, wild type), n=5 per genotype for RNA-seq (MeCP2 OE, wild type), n=2 wild type for DNA methylation (Stroud et al., 2017). In G and H analysis was carried out for genes over 75kb to visualize correlation signal within gene bodies.
2.6.8 Figure S2. Chromatin topology is associated with non-CG DNA methylation in the cerebral cortex. Related to Figure 1.

A. Boxplots of mCA/CA and mCG/CG within TADs that contain MeCP2-repressed genes, MeCP2-activated genes, or no dysregulated genes. ***, p < 10^-8 Wilcoxon rank-sum test.

B. Aggregate plots of mCA/CA (left), mCG/CG (center), and BS-sequencing coverage rates and GC composition (right) phased on boundaries of TADs defined in Hi-C data from the cerebral cortex at eight weeks of age. TADs containing MeCP2-repressed (red), MeCP2-activated (blue), or all other genes (purple) are shown. “Inside” indicates TADs containing genes of interest. Black line and ribbon for each plot indicates the mean and standard deviation of 20 sets of resampled boundaries generated by shuffling TAD locations in the genome (see methods).

C. A histogram of -log10 p-values for paired t-tests, comparing the difference in DNA methylation on either side of the TAD boundaries or 1000 resampled TAD boundaries as in B. Histogram shows the distribution of p-values for resampled TADs, red line indicates p-value for true TADs.

D. Cross correlation analysis of mCG/CG signal within and across TAD boundaries for all genes. mCG/CG values were calculated for 10 intra-domain regions and 10 equally-sized regions up and downstream of each TAD. Correlation between these regions across all TADs is shown (see methods).

E. Boxplots of mCA/CA and mCG/CG within contact domains that contain MeCP2-repressed genes, MeCP2-activated genes, or no dysregulated genes. Contact domains were defined by analysis of Hi-C data generated from cerebral cortex neurons isolated from fetal brain (Bonev et al., 2017). **, p < 10^-3; ***, p < 10^-8 Wilcoxon rank-sum test.
F. Aggregate plots of mCA/CA (left), mCG/CG (center), and BS-sequencing coverage rates and GC composition (right) phased on boundaries of contact domains defined in Hi-C data from fetal cortical neurons. Presented as in panel B.

G. A histogram of -log10 p-values for paired t-tests, comparing the difference in DNA methylation on either side of contact domain boundaries or 100 resampled contact domain boundaries as in panel F. Histogram shows the distribution of p-values for resampled contact domains, red line indicates p-value for true contact domains.

H. Cross correlation analysis of mCA/CA and mCG/CG signal within and outside of contact domains as in panel D. Resampling was performed by shuffling contact domain-sized regions around the genome and repeating the analysis of all genes (see methods).

I. Boxplots of mCA/CA within TADs defined in cerebral cortex (top) and cerebellum (bottom) that contain MeCP2-repressed genes, MeCP2-activated genes, or no dysregulated genes. ***, p < 10^-8 Wilcoxon rank-sum test. Cerebral cortex TADs were compared to genes from this study. Cerebellum TADs were compared to genes previously identified as misregulated in the cerebellum and multiple other brain regions (Gabel et al., 2015).

J. Cross correlation analysis of mCA/CA signal within and outside of TADs defined in cerebral cortex (top) and cerebellum (bottom), as performed in panel H.

K. Aggregate plots of mCA/CA from frontal cortex (top) and granule neurons (bottom) phased on boundaries of MeCP2 repressed TADs defined in cerebral cortex (top) and cerebellum (bottom). Analysis performed on TADs with highly differing mCA levels as in Figure 1G.

A-D analysis of Hi-C interaction data (Dixon et al., 2012), and DNA methylation, n=2 (Stroud et al., 2017) from the cerebral cortex at 8 weeks of age. E-H analysis of Hi-C interaction data from
neurons isolated from E14.5 cortex (Bonev et al., 2017), and DNA methylation, n=2 (Stroud et al., 2017) from the cerebral cortex at 8 weeks of age. I-K analysis of bisulfite data from 6-week frontal cortex and granule neurons isolated from 7-12 week old cerebellum (Lister et al., 2013; Mellén et al., 2017)(Mellén et al., 2017). Hi-C data from 6-8 week old Cerebellum (Yamada et al., 2019). Note that some panels from Figure 1 are repeated here to allow for comparisons.
2.6.9 Figure S3. Binding of DNMT3A and recruitment of MeCP2 are shaped by topologically-associating domains. Related to Figure 2.

A. Density scatter plot of DNMT3A ChIP/Input signal at 2 weeks of age and mCA/CA levels at 8 weeks of age for random 1kb regions of the genome. Spearman rho shown for the correlation between DNMT3a signal and mCA/CA level. ***, p < 10^{-8}.

B. Aggregate plot of Input CPM at 2 weeks phased on boundaries of TADs that contain MeCP2-repressed genes (TADS from cerebral cortex at eight weeks of age). Black line and ribbon for each plot indicates the mean and standard deviation of 20 resampled boundaries generated by shuffling contact domain locations in the genome (see methods).

C. Cross correlation analysis of DNMT3A ChIP-seq signal at 2 weeks of age for contact domains defined in fetal cortical neurons. DNMT3A ChIP/Input values were calculated for 10 intra-domain regions and 10 equally-sized regions up and downstream of each domain. Correlation between these regions across all domains is shown (see methods).

D. Density scatter plot of DNMT3A ChIP/Input signal at 2 weeks of age and mCA/CA levels at 8 weeks of age for contact domains defined in fetal cortical neurons. Spearman rho shown for the correlation between DNMT3a signal and mCA/CA level. ***, p < 10^{-8}.

E. Cross correlation analysis (performed as in panel C) for DNMT3A ChIP-seq signal at 8 weeks of age in TADs defined in the cortex at 8 weeks of age.

F. Density scatter plot of DNMT3A ChIP/Input signal at 8 weeks of age and mCA/CA levels at 8 weeks of age for TADs defined in the cortex at 8 weeks of age. Spearman rho shown for the correlation between DNMT3a signal and mCA/CA level. ***, p < 10^{-8}.

G. Cross correlation analysis (performed as in panel C) for MeCP2 ChIP/Input signal at 8 weeks of age in TADs defined in the cortex at 8 weeks of age.
H. Barplots of genome-wide mC/C levels in a DNMT3A Nestin-cKO and control cerebral cortex at 8 weeks of age. *, p < 0.05 two-tailed t-test.

I. Boxplot of fold-changes in MeCP2 ChIP-seq signal within TADs upon ablation of mCA in the DNMT3A Nestin-cKO. Change in signal is shown for TADs separated by quartiles of mCA/CA (left) and mCG/CG (right) under wild-type conditions. Spearman rho shown for the correlation between DNMT3a signal and mCA/CA level. *** p < 10⁻⁸.

J. Heatmap of correlation between mCA/CA levels of gene body (top), enhancer (middle), and TSS (bottom) regions found inside and outside of TADs and the average mCA/CA level for each TAD. Distinct drop off in correlation coefficient for these elements when they are found outside the TAD boundary illustrates influence of intra-TAD mCA/CA consistency. Similar correlation level and drop-off in signal is not seen for resampled TADs, placed in randomized positions in the genome (see methods).

K. Barplots of correlations of mCA/CA levels for gene bodies (left), enhancers (middle) and TSSs (right), located in either the same TAD or separate TADs. To facilitate comparison, distances between pairs of elements in different TADs was matched to distances between pairs of elements in the same TAD (see methods).

L. Density scatter plots of TAD methylation levels and methylation at kilobase-scale genomic elements in the cerebral cortex at 8 weeks of age. Values for every TSS, gene body, or enhancer in the genome are plotted against the TAD in which the element resides. As in Figure 2E, data is subtracted from TADs such that element methylation does not contribute to plotted TAD methylation. Spearman rho shown for the methylation levels of each genomic element versus the methylation levels of the TAD that the element is in. ***, p < 10⁻⁸.
M. Density scatter plots of TAD methylation levels and methylation at resampled random locations for regions size-matched to kilobase-scale genomic elements (as in panel L). Values for every resampled TSS-, gene body-, or enhancer-sized region in the genome are plotted against the TAD in which the element resides. Spearman rho shown for the methylation levels of each randomized genomic element versus the methylation levels of the TAD that the element is in. ***, p < 10^-8.

A,C,D Analysis of Hi-C interaction data from neurons isolated from E14.5 cortex (Bonev et al., 2017), DNA methylation, n=2, from the cerebral cortex at 8 weeks of age, and DNMT3A ChIP-seq data, n=3, from the cerebral cortex at 2 weeks of age (Stroud et al., 2017). B,E-G,I-M Analysis of Hi-C interaction data (Dixon et al., 2012), DNA methylation (n=2 wild type), DNMT3A ChIP-seq data, (n=2 wild type), (Stroud et al., 2017), and MeCP2 ChIP-seq data (n=2 per genotype) (Kinde et al., 2016) from the cerebral cortex at 8 weeks of age. H n=2 per genotype for DNA methylation DNMT3A Nestin-cKO versus wild-type, 8 weeks of age. Note that some panels from Figure 2 are repeated to allow for comparisons.
2.6.10 Figure S4. Changes in intronic RNA are consistent with promoter-associated transcriptional upregulation of long, highly methylated, MeCP2-repressed genes in the MeCP2 KO. Related to Figure 3.

A. Scatterplot of the log2 fold-changes in the MeCP2 KO versus wild type for exonic RNA measured by total RNA sequencing of cerebral cortex tissue compared to intronic RNA measured by total RNA sequencing of isolated nuclei from this tissue. Genes identified as MeCP2-repressed and MeCP2-activated in combined analysis of exonic RNA in MeCP2 KO
and MeCP2 OE mice (Figure 1; S1) are indicated by red and blue dots respectively, all other genes indicated as gray points.

B. Left, heatmap of changes in intronic RNA for genes detected as significantly dysregulated (FDR <0.1) in nuclear RNA-seq from MeCP2 KO cerebral cortex (grey indicates, no data for replicate). Right, venn diagram of the overlap between genes called as MeCP2-repressed or MeCP2-activated in analysis of nuclear intronic RNA expression in the MeCP2 KO (pale colors) with genes called as dysregulated in combined analysis of exonic RNA in the MeCP2 KO and MeCP2 OE mice (dark colors) (Figure 1; S1). ***. p < 10^{-8} hypergeometric test.

C. Aggregate plot of cerebral cortex mCA/CA levels for MeCP2-repressed, MeCP2-activated, and all other genes defined by analysis of intronic RNA in the MeCP2 KO (see panel B). Mean mCA/CA for 1kb bins in the TSS and regions surrounding genes is shown. For “Metagene” region, mean mCA/CA was calculated for 50 equal-sized bins within the body of each gene.

D. Boxplot of gene lengths for MeCP2-repressed or MeCP2-activated, and all other genes called by analysis of nuclear intronic RNA in the MeCP2 KO and wild type mice. *, p < 0.05; ***, p < 10^{-8} Wilcoxon rank-sum test.

E. Running-average plot of fold-changes in intronic RNA expression in the MeCP2 KO and wild type versus mean gene length (top) or mean gene body mCA/CA (bottom). Mean fold-changes are plotted for bins of 201 genes sorted by length of mCA/CA per gene with a 1-gene step (see methods).

F. Plot of spearman correlations between fold-changes in intronic RNA expression in the MeCP2 KO and wild type versus mCA/CA or mCG/CG for regions in and around genes. Analysis performed for 1kb bins across 200kb (top) and 2 Mb (bottom) regions.
G. Scheme to assess changes in pre-mature transcription termination in the MeCP2 KO versus wild type. Similar to analysis carried out by Boswell et al., 2017, a “termination ratio” for each gene above 50kb in each sample is defined as the ratio of read counts in the first 25kb of each gene to the read counts in the last 25 kb of that gene. The mean fold-change in the ratios between MeCP2 KO and wild type is calculated across all genes in paired replicates.

H. Dotplot showing the change in termination ratio between the nuclear intronic RNA-seq data in the MeCP2 KO and wild type for genes that are not significantly changed and MeCP2-repressed genes. A prediction for the change in this ratio that would be expected if the effects on mRNA in the MeCP2 KO were due entirely to changes in transcription termination rate was generated for comparison, “Predicted” (see methods).

I. Dotplot as in panel H, showing the change in termination ratio between the intronic RNA-seq in the MeCP2 OE and wild type for genes that are not significantly changed and MeCP2-repressed genes.

Data from cerebral cortex of 7-10 week old animals. For MeCP2 KO: n= 6 per genotype for RNA-seq; for MeCP2 OE: n=5 per genotype for RNA-seq; n=2 wild-type for DNA methylation (Stroud et al., 2017).
2.6.11 Figure S5. Analysis of enhancers dysregulated in MeCP2 mutants.

Related to Figure 4 and Figure 6.

A. Boxplot of fold-changes in H3K27ac ChIP signal in MeCP2 KO and OE (left) and MeCP2 ChIP/Input signal (right) for enhancers identified with combined ChIP-seq analysis in KO and OE as MeCP2-repressed, MeCP2-repressed with high MeCP2/Input (see methods), MeCP2-activated, or all other enhancers. Fold-changes were calculated by edgeR analysis of H3K27ac ChIP-seq signal at enhancer regions (see methods). Significantly dysregulated enhancers were defined as FDR < 0.1. *, p < 0.05, ***, p < 10^{-8} Wilcoxon rank-sum test.

B. Left, boxplots of mC/C, mC/kb, and C sites/kb at MeCP2-repressed, MeCP2-repressed with high MeCP2, MeCP2-activated, and all other enhancers, as displayed in panel A. Right, heatmap of enrichments for median levels of mC/C, mC/kb, and C sites/kb for MeCP2-repressed and MeCP2-activated enhancers compared to all other enhancers. *, p < 0.05, **, p < 10^{-3}; ***, p < 10^{-8} Wilcoxon rank-sum test.

C. Heatmap of enrichment significance for number of mC sites (mC/kb, left), methylation level (mC/C, middle) and sequence occurrence (right) for trinucleotide sites in MeCP2-repressed and MeCP2-activated enhancers. Significance of enrichment or depletion was calculated by comparing occurrences of mC/kb, mC/C, or sequence alone for each trinucleotide in the 2kb region surrounding enhancers to resampled sets of enhancers that are not significantly changed and matched to changed enhancers for H3K27ac signal (see methods).

D. Histograms of mCA/kb and mCG/kb in enhancers (black). Blue and red lines indicate the distributions of methylation for MeCP2-activated and MeCP2-repressed enhancers respectively.
E. Genome browser view of an example MeCP2-repressed gene, *Zmat4* (as in Figure 4A), showing overlaid wild-type and MeCP2 KO aggregate H3K27ac ChIP-seq signal, H3K4me1 ChIP-seq peaks, H3K27ac peaks called as enhancers, and peaks from a compendium of ATAC-seq peaks identified across 13 mouse tissues. Blue highlights ATAC-seq peaks that correspond to sub-thresholded enrichment of H3K27ac, not identified in peak calling analysis of H3K27ac ChIP-seq.

F. Boxplots of H3K27ac ChIP signal in MeCP2 KO versus wild type (left) and MeCP2 OE vs wild type (right) for enhancers defined in this study, a compendium of detectable ATAC-seq peaks, and non-peak regions that are sized matched to ATAC-peaks but selected to not overlap enhancers of ATAC-seq peaks. Values for the highest mCA/CA decile are shown for each class, illustrating total levels of H3K27ac and mCA-dependent dysregulation ***, p < 10^-8 Wilcoxon rank-sum test.

G. Boxplots of fold-changes of H3K27ac ChIP signal in MeCP2 KO versus wild type across deciles of mCA/kb (left) and mCG/kb (right) for ATAC-peaks described in panel E. Spearman rho shown for correlations of each methylation mark with change of H3K27ac at enhancers. ***, p < 10^-8 Wilcoxon rank-sum test.

H. Boxplots of fold-changes of H3K27ac ChIP signal in MeCP2 OE versus wild type across deciles of mCA/kb (left) and mCG/kb (right) for ATAC-peaks described in E. Spearman rho shown for correlations of each methylation mark with change of H3K27ac at enhancers. ***, p < 10^-8.

I. Quantitative 3C analysis of enhancers found in MeCP2 repressed genes, or control loci. No significant factors were detected in a 2-way ANOVA, using enhancer locations and genotype.
J. Density scatter plot of log10 number of intragenic enhancers versus log10 gene length for all genes in the genome. Spearman rho shown for the number of intragenic enhancers and gene length for all genes with at least 1 enhancer. ***, p < 10^{-8}.

K. Running-average plot of fold-change in intronic RNA expression in the MeCP2 KO or MeCP2 OE versus mCA/kb for genes containing enhancers (red line) and length-matched sets of genes that do not contain enhancers (gray lines). Mean fold-changes are plotted for bins of 201 genes sorted by mCA/kb with a 1-gene step (see methods).

L. Local correlation analysis correlating mCA/kb (left) and mCG/kb (right) for 1 kb windows in and around genes to intronic RNA fold-changes for the gene in the MeCP2 KO (top) and OE (bottom). Genes at least 75kb in length are analyzed to allow visualization of correlations in gene bodies. Genic windows of methylation go from the TSS to 75kb downstream.

M. Heatmap of correlation between fold-change in H3K27ac at intragenic regions and gene expression fold-change for MeCP2 KO and MeCP2 OE mice. Values are calculated as partial correlation for enhancers or control non-enhancer regions of the same size, removing the signal from the other class of element (see methods). Larger correlation for enhancers over control regions illustrates the link between enhancer regulation and transcriptional control.

Data from cerebral cortex of 7-10 week old animals. For MeCP2 KO: n=5 per genotype for H3K27ac, n=6 per genotype for RNA-seq; for MeCP2 OE: n=3 per genotype for H3K27ac, n=5 per genotype for RNA-seq; n=2 wild-type for DNA methylation (Stroud et al., 2017). ATAC-seq peaks (Cusanovich et al., 2018). Note that some panels in C are repeated from Figure 4 to allow for comparisons.
2.6.12 Figure S6. Transcriptomic and epigenomic analysis of DNMT3A Baf53b-cKO. Related to Figure 5.

A. Barplots of genome-wide methylation levels in the DNMT3A Baf53b-cKO at different CN dinucleotides. *, p < 0.05, **, p < 10^{-3}; ***, p < 10^{-8} two-tailed t-test.

B. Barplots of mCG (left) and mCA (right) levels in the DNMT3A Baf53b-cKO at genes, enhancers, and TSSs. **, p < 10^{-3}; ***, p < 10^{-8} two-tailed t-test.

C. Boxplots of mCG (left) and mCA (right) levels in the DNMT3A Baf53b-cKO within TADs that contain MeCP2-repressed genes, MeCP2-activated genes, or no dysregulated genes. ***, p < 10^{-8} Wilcoxon rank-sum test.
D. Heatmap of H3K27ac fold-changes for enhancers detected as significantly dysregulated in DNMT3A Baf53b-cKO mice by edgeR analysis (see methods).

E. Heatmap of enrichment significance for the number of mC sites (mC/kb) within enhancers detected as significantly dysregulated in the DNMT3A Baf53b-cKO. Significance of enrichment or depletion was calculated by comparing occurrences of mC/kb for each trinucleotide in the 2kb region surrounding enhancers to resampled sets of enhancers that are not significantly changed but matched to changed enhancers for H3K27ac signal (see methods).

F. Plot of spearman correlations between H3K27ac fold-change at enhancers in the DNMT3A Baf53b- cKO and mC/kb levels in the wild-type brain for 500bp bins in and around the enhancers. Plots are centered at the summit of enhancer H3K27ac ChIP peaks.

Data from cerebral cortex of 7-10 week old animals. n=6 per genotype for Bisulfite-seq analysis, n=6 per genotype for H3K27ac.
Chapter 3: Probing the mechanism of neuronal enhancer regulation by MeCP2

Experiments and analysis discussed in this chapter were done in collaboration between Nicole Hamagami (NH) and myself (AWC). AWC carried out all of the ChIP-seq and ATAC-seq experiments and analysis. NH performed the CUT&Tag experiments and AWC performed the analysis.
3.1 Introduction

MeCP2 and methylation-dependent regulation of enhancer activity has been identified in mouse cerebral cortex (Clemens et al., 2020). However, the direct mechanism by which MeCP2 mediates changes in H3K27ac to regulate enhancers remains to be determined. As discussed in Chapter 1, mutations of MeCP2 within the NID, particularly R306C, demonstrate the importance of NCoR for the regulatory function of MeCP2. This implicates the NCoR complex in the dysregulation of enhancer activity and gene expression that have been observed in studies of MeCP2 mutants. However, whether NCoR is directly responsible for controlling H3K27ac levels or a secondary aspect of enhancer activity remains to be determined. In the primary model that has been proposed (Tillotson and Bird, 2019), MeCP2 is required to recruit NCoR to sites of the genome in order to mediate histone deacetylation and gene repression. However, these studies are based on heterologous assays in non-neuronal cells (Lyst et al., 2013), and no studies have examined how NCoR recruitment to the genome is affected by MeCP2 mutations in vivo. Furthermore, a recent study testing the role of HDAC3 activity in MeCP2-NCoR mediated effects provides evidence of a mechanism for repression that occurs outside of the NCoR-associated deacetylase activity (Koerner et al., 2018). This study dissected the impacts of mutations of the catalytic site of HDAC3 within the NCoR complex and found that loss of HDAC3 deacetylase activity did not rescue phenotypes in mice caused by overexpression of MeCP2. The lack of rescue by reduced deacetylation suggests that the regulatory effects of MeCP2-NCoR are not directly associated with the deacetylase activity (Koerner et al., 2018). Thus, the recruitment and deacetylation model for MeCP2-NCoR may not fully explain the results to date, and additional studies will be needed to identify the importance of NCoR to the function of MeCP2.
Progress towards understanding where in enhancer regulation MeCP2-NCoR acts can be made by assessing which steps in enhancer activation are affected by loss of MeCP2. Enhancer activation is thought to be initiated by binding of pioneering transcription factors, enabling the recruitment of chromatin remodelers and histone modifying enzymes. Nucleosome eviction through modifications and chromatin remodeling allows for additional transcription factors and cofactors to bind, leading to the deposition of histone H3K4me1 at enhancers “poised” for activation (Local et al., 2018; Pundhir et al., 2015; Vernimmen and Bickmore, 2015). Histone acetyl transferase enzymes are then recruited to deposit H3K27ac, which mark active enhancers (Creyghton et al., 2010; Lai et al., 2017). Once marked for activation, BRD4 co-activator interacts with acetylated histones, which then drives further co-activator recruitment (Calo and Wysocka, 2013; Lee et al., 2017). Mediator is a critical cofactor recruited to active enhancers by BRD4 (Lee et al., 2017; Quevedo et al., 2019), which then functions as an anchor protein at enhancers and super enhancers to recruit transcription factors and other cofactors (Allen and Taatjes, 2015). Cohesin is one cofactor that interacts with Mediator at enhancers to in turn drive enhancer-promoter looping to connect active enhancers to their target gene and activate transcriptional initiation (Kagey et al., 2010; Quevedo et al., 2019; Vernimmen and Bickmore, 2015). This multi-step process of enhancer activation creates many potential regulatory targets for MeCP2.

If MeCP2-NCoR mediates deacetylation as a repressor of enhancer activity, it would be predicted that steps upstream of this process might not be affected by loss of MeCP2. Alternatively, changes upstream in enhancer activation may drive the changes of H3K27ac previously observed at enhancers. If these alterations are observed, it would support a model in which MeCP2-NCoR acts at a step before histone deacetylation, instead of directly deacetylating histones itself. To
understand how loss of MeCP2 repression affects gene expression, it is important to assess how direct or indirect alterations in H3K27ac cause downstream changes in coactivator binding and chromosome architectural proteins, linking enhancer activation to gene activation. Therefore, interrogation of each of the steps both upstream or downstream of H3K27ac, upon MeCP2 mutation may provide further information on the mechanism of gene regulation by the mC-MeCP2-NCoR pathway. For example, measuring upstream changes of chromatin accessibility or H3K4me1 of primed enhancers and downstream changes of binding profiles of BRD4, Mediator, Cohesin, and other cofactors in MeCP2 mutants, will provide direct readouts of epigenetic changes that occur outside of H3K27ac dysregulation.

To date, limited analyses have been performed to assess which steps in enhancer regulation are affected by MeCP2 disruption, but some analyses suggest selective effects. For example, changes in promoter-enhancer interactions were not observed genome-wide from Hi-C or in targeted analysis of de-repressed enhancers with 3C upon loss of MeCP2 (Boxer et al., 2020; Clemens et al., 2020), so measuring changes in enhancer activation may provide insight into the regulatory role of MeCP2 outside of alterations in genomic architecture. By dissecting multiple steps of enhancer activation, we can develop a clear indication of how each step is affected by MeCP2-NCoR, thereby predicting a more accurate model of gene regulation.

Here, we perform systematic chromatin analyses to examine the mechanism of enhancer regulation by MeCP2. We interrogate how genomic binding of the NCoR complex is affected by loss and overexpression of MeCP2 and which steps in enhancer activation are altered by MeCP2 disruption. We show that NCoR binds to the genome independent of MeCP2 and that multiple steps in
enhancer activation are regulated by MeCP2, including chromatin accessibility, H3K4me1 presence, and Mediator binding. However, Cohesin binding remains largely unchanged by MeCP2 disruption. These findings suggest a model in which MeCP2 regulates the pre-bound NCoR complex to modulate enhancer activation directly at the regulatory element, but that this regulation occurs independent of alterations in enhancer-promoter looping.

3.2 Results

3.2.1 NCoR binding to regulatory elements in the brain is not dependent on MeCP2.

To directly test the model that MeCP2 is required for NCoR recruitment, we examined if changes in levels of MeCP2 drive differences in binding profiles of the NCoR complex through ChIP-seq in adult cerebral cortex of MeCP2 mutants. If MeCP2 is required for NCoR recruitment, we would expect that ChIP-seq signal for NCoR1, a core component of the NCoR complex, would follow patterns of MeCP2, where sites normally enriched for methylation and MeCP2 would show loss of signal upon loss of MeCP2 and gain in signal with overexpressed MeCP2. Alternatively, if recruitment is not affected by MeCP2 presence, we would not expect significant changes in NCoR1 ChIP-seq signal with MeCP2 mutants compared to wild-type cortex, potentially invoking an alternative regulatory mechanism outside of MeCP2-mediated recruitment. We performed ChIP-seq for NCoR1 in 8-week cortex of Mecp2 knockout, MECP2 overexpression, and wild-type mice to determine if MeCP2 is required for its binding to the genome (Figure 1A).
MeCP2 binds broadly across the genome in a pattern that mirrors DNA methylation, and is depleted from, although still present at, lowly methylated, highly active regulatory elements. Thus, MeCP2-dependent binding of NCoR would be expected to show similar patterns as MeCP2 and DNA methylation across the genome. Surprisingly however, genome-wide binding profiles of NCoR1 heavily overlap with regulatory elements, with nearly half of NCoR1 peaks corresponding with defined enhancers in the cortex (Figure 1B). In fact, patterns of total NCoR1 binding across genotypes significantly correlate with the total amount of H3K27ac (Figure 1C). This finding is supported by previous studies that show ChIP-seq signal for NCoR and other co-repressors are in fact found at active promoters and enhancers, invoking an unknown and paradoxical mechanism of recruitment for co-repressors at sites of high activity (Siersbæk et al., 2017; Whyte et al., 2012). The profiles of NCoR1 suggest that MeCP2 is not a major determinant of NCoR complex binding across the genome, but rather the complex occupies active regulatory regions.

To directly interrogate the impact of MeCP2 disruption on NCoR complex binding, we examined how NCoR profiles change in MeCP2 knockout and MECP2 overexpression mice. If MeCP2 was necessary for recruitment of the NCoR complex, we would observe decreased ChIP signal of NCoR1 at enhancers most misregulated by MeCP2 mutations. Instead we observe the opposite finding, where at MeCP2-repressed enhancers, which are a group of enhancers previously identified to be de-repressed in the absence of MeCP2 (see Chapter 2), there is a subtle, but significant increase in ChIP signal of NCoR1 (Figure 1D). Additionally, upon overexpression of MeCP2, we observe decreased ChIP signal (Figure 1D), suggesting that MeCP2 is not required for NCoR binding and may even restrict the binding of the NCoR complex to regulatory elements.
3.2.2 Loss of MeCP2 leads to alterations in accessibility and enhancer priming upstream of histone acetylation at MeCP2-regulated enhancers.

We next sought to examine how disruption of MeCP2 affects each step in enhancer activation, in order to determine where in this process MeCP2 may collaborate with the independently recruited NCoR complex to repress activity. As previously mentioned, recent evidence suggests a role for NCoR-HDAC3 independent of deacetylase activity in neuronal gene regulation (Koerner et al., 2018). In this way, the complex would not directly remove histone acetylation at enhancers to block gene activation, but instead may affect chromatin accessibility or early, priming histone modifications and/or the enzymes that deposit them. Regulation of these components could subsequently lead to changes in histone acetylation we and others observe. We therefore sought to investigate which steps in enhancer regulation MeCP2-NCoR may influence to control gene expression.

To detect changes that occur early in the enhancer activation pathway, we performed ATAC-seq on MeCP2 knockout and MECP2 overexpression mice, assessing if changes in chromatin accessibility are driven by MeCP2 at enhancers (Figure 2A). Quantification of ATAC-seq signal within our enhancers defined by H3K27ac, illustrates a significant increase in MeCP2-repressed enhancers compared to non-misregulated enhancers in MeCP2 knockout mice, as well as a decrease in MECP2 overexpression compared to wild-type mice (Figure 2B). This trend was observed genome-wide as well, with changes in ATAC-seq correlating with total amounts of methylation (i.e. more MeCP2 binding sites) across all enhancers between MeCP2 knockout and wild-type mice (Figure 2C, MECP2 overexpression not shown). These changes in ATAC-seq suggest that MeCP2
decreases chromatin accessibility at enhancers, showing MeCP2 has a repressive effect on enhancer activity upstream of H3K27ac.

In order to get a refined measurement of changes in accessibility that occur directly at the core sequences of enhancers, we analyzed subnucleosomal regions as determined by ATAC-seq fragments less than the length of a mononucleosome (<100bp) (Figure 2A). Using this approach, we still found a significant change of ATAC signal at MeCP2-repressed enhancers, and highly-methylated enhancers genome-wide, but to a lesser degree than all ATAC data (Figure 2D, E). Smaller magnitude changes of nucleosome-free ATAC signal suggest that the nucleosome-free region of enhancers may be less susceptible to reductions in accessibility driven by MeCP2 compared to surrounding nucleosomal regions of enhancers.

To assess the “priming” of enhancers for activation we next assessed how the H3K4me1 histone modification at these sites changes in response to loss of MeCP2. H3K4me1 has been described to be present at active regulatory elements as a precursor to H3K27ac (Creighton et al., 2010; Lai et al., 2017). One potential mechanism of the effect of MeCP2 on enhancers is that it promotes demethylation of H3K4me1, which then causes a subsequent effect on enhancer-associated H3K27ac. Recent advances in genomic profiling have developed more sensitive techniques such as CUT&RUN and CUT&Tag that reduce background signal to measure protein binding with lower sequencing depth (Kaya-Okur et al., 2019; Skene and Henikoff, 2017). We have therefore implemented CUT&Tag to generate high signal-to-noise profiles for this modification genome-wide to detect changes in H3K4me1 in Mecp2 knockout mouse cerebral cortex. Within MeCP2-repressed enhancers, that are defined by their change in H3K27ac caused by mutation of MeCP2
(H3K27ac increases in MeCP2 knockout and decreases in MECP2 overexpression), we also observe increased H3K4me1 in the MeCP2 knockout and a decrease in MECP2 overexpression (Figure 2F, MECP2 overexpression not shown). Additionally, as previously observed for H3K27ac dysregulation by MeCP2, the enhancers with the most methylation exhibit the greatest upregulation of H3K4me1 signal in the MeCP2 knockout (Figure 2G). Notably, the effect size of H3K4me1 was more pronounced than H3K27ac, wherein for enhancers with the highest level of methylation, the change in H3K4me1 was greater in magnitude than the change in H3K27ac (compare Chapter 2 Figure 4F to Figure 2G).

3.2.3 Loss of MeCP2 differentially affects cofactors associated with enhancer-mediated promoter activation.

The changes in H3K27ac observed at enhancers are likely to drive changes in downstream mechanisms in order to result in changes of gene activation. To investigate changes driven by MeCP2 mutations that may occur downstream of histone modifications, we measured how the presence of co-activators that mediate gene activation between promoters and enhancers are affected across MeCP2 mutants using ChIP-seq and CUT&Tag. We determined binding profiles of two complexes associated with enhancer activity, Mediator and Cohesin, across MeCP2 knockout adult mouse cortex and found significant increases of Mediator in MeCP2 knockout compared to wild-type cortex at MeCP2-repressed enhancers compared to unaffected enhancers (Figure 2F). Conversely, ChIP-seq measurements of Cohesin identified only slight but significant increases in signal in MeCP2 knockout cortex compared to wild-type. In comparison to changes in Mediator, the changes in Cohesin at MeCP2-repressed enhancers are subtle. These findings suggest that loss of MeCP2 repression at enhancers leads to increased recruitment of Mediator, but less substantial effects on Cohesin as the chromatin looping complex.
3.3 Discussion

Here we present initial analysis of the molecular mechanisms influenced by the mC-MeCP2-NCoR pathway to control enhancer activity. Contrary to a popularly proposed model for mC-MeCP2-NCoR repression (Tillotson and Bird, 2019), we find that NCoR recruitment to the genome is independent of MeCP2 and that disruption of MeCP2 leads to dysregulation of multiple steps of the enhancer activation pathway, upstream and downstream of H3K27ac regulation.

Given that we find that the NCoR complex does not require MeCP2 for recruitment, and this result is at odds with in vitro interaction assays, it will be important in future analyses to address antibody specificity for the NCoR complex in our genome profiling assays. Preliminary analysis of ChIP-seq and CUT&Tag of a second component of the NCoR complex, TBLR1, indicates a high-degree of overlap between NCoR1 peaks and TBRL1 peaks, further supporting our NCoR findings (data not shown). Future experiments will expand on this TBLR1 data, allowing all conclusions to be based on two independent antibodies (NCoR1, TBLR1) and two methods (ChIP-seq and CUT&Tag) to assess NCoR complex binding. To further validate these results, we can use shRNAs to knockdown NCoR expression in vitro. ChIP-seq in control and NCoR-knockdown cells should show a measured depletion of NCoR signal in knockdown cells compared to control cells with normal levels. These control experiments will then provide further support for in vivo results of NCoR from mouse cortical tissue.
Because we observe changes in enhancer activation steps upstream of histone acetylation, our data support a mechanism through MeCP2 and NCoR that does not act through the HDAC3 component of this regulatory complex, but rather a yet to be defined role that drives these changes. We have identified steps upstream of histone acetylation in the enhancer activity pathway that are dysregulated through MeCP2 mutations that consist of increased chromatin accessibility surrounding enhancers and increased signal for the “primed” enhancer mark, H3K4me1. As these changes are canonically thought to occur upstream of H3K27ac modification, it may be feasible that these changes are driving the observed increases in H3K27ac upon MeCP2 loss. The greater effect size for H3K4me1 in MeCP2 mutants compared to H3K27ac changes suggests that H3K4me1 may be more directly affected by MeCP2. However, it is difficult to distinguish between technical differences (CUT&Tag vs ChIP-seq) or whether H3K4me1 is actually more affected by MeCP2 than H3K27ac. Distinguishing biological effects from technical differences will allow us to pinpoint a more direct target of MeCP2, where if H3K4me1 has a greater effect size, this suggests the changes in enhancer activity by H3K27ac are a downstream effect of MeCP2-mediated regulation of H3K4me1. The same technical considerations are true for ATAC-seq as well. Therefore, while the cross-platform approach used for this study adds support for the findings on increased enhancer activity, it increases the difficulty of interpreting the magnitude of effects to identify the directly affected mechanism by the MeCP2-NCoR pathway. Finally, while increases in accessibility and H3K4me1 are thought to occur upstream of enhancer acetylation, it is possible that MeCP2-NCoR affect acetylation and that these changes feedback to affect accessibility and H3K4me1 in mechanisms that are not yet known. Future experiments measuring H3K27ac with CUT&Tag will add sensitivity to our previous ChIP-seq findings as well as be more directly comparable to H3K4me1 in determining the regulatory target of MeCP2.
While recruitment and binding of NCoR does not appear to be affected by the presence or absence of MeCP2, studies have begun examining if MeCP2 can regulate the activity of HDAC3 as a component of the NCoR complex. Mutations of the deacetylase activating domain (NS-DAD) cause a reduction in deacetylase activity of HDAC3, but fail to rescue phenotypes associated with MeCP2 overexpression (Koerner et al., 2018). To directly test the role of MeCP2 in HDAC activity, future experiments can measure NCoR-HDAC3 associated-deacetylase activity in vitro in the presence of wild-type MeCP2, NCoR-interacting deficient mutant MeCP2, or no MeCP2. If MeCP2 functions to control the activity of HDAC3, we would expect higher levels of deacetylase activity with wild-type MeCP2 compared to controls. Moreover, the presence of R306C mutated MeCP2 (which disrupts its ability to interact with the NCoR-TBL1-HDAC3 complex) should produce comparable deacetylase activity to the absence of MeCP2. Such an approach allows for a more direct measurement of the involvement in the interaction of MeCP2 with NCoR-TBL1-HDAC3, and the results from these studies could corroborate the findings from Koerner et al.

As much as upstream molecular mechanisms may be regulated through MeCP2, the changes in H3K27ac caused by MeCP2 mutations appear to lead to downstream effects on regulatory elements that recognize acetylated histone modifications. Our finding that Mediator shows a robust upregulation that is similar in magnitude to changes in H3K4me1 and H3K27ac at misregulated enhancers, suggests it is recruited to enhancers upon loss of MeCP2. MeCP2 may likely regulate the binding of other co-factors as well. In fact, a recent study of cultured human induced pluripotent stem cell neurons suggests that MeCP2 represses binding of BRD4 (Xiang et al., 2020), a co-regulatory protein that forms condensates with acetylated enhancers to mediate gene activation.
(Sabari et al., 2018). Future studies can examine if alterations in BRD4 occur downstream of the mC-MeCP2-NCoR pathway at enhancers in adult neurons.

In contrast to our Mediator findings, we observed minimal changes in Cohesin at misregulated enhancers in MeCP2 mutant cortical tissue. In line with these limited changes in Cohesin are previous findings of 3C and Hi-C experiments in the same mouse model that did not identify significant changes in enhancer-promoter contacts (Boxer et al., 2020; Clemens et al., 2020). Alterations in Mediator binding, but not Cohesin, upon mutation of MeCP2 further suggests that the primary role of MeCP2 may be to control enhancer-associated modifications and cofactors (i.e. H3K4me1, H3K27ac, Mediator) but not enhancer-promoter looping.

The identification of upstream and downstream effects on enhancer-associated histone modifications driven by MeCP2 has furthered the understanding of the role of MeCP2 in neurons. Future studies addressing potential additional downstream effects of H2K27ac dysregulation, like BRD4 recruitment at enhancers, may identify new functional consequences of MeCP2 mutations. For example, examination of MeCP2 in phase separating and condensate forming structures will show how the presence or absence of MeCP2 in condensates can influence mechanisms of molecular crowding and condensation of BRD4 (Bancaud et al., 2009) to affect other components from gaining access to these regions of high-regulatory activity. Continued studies could link biophysical properties of MeCP2 to these epigenomic effects to create a neuronal-specific gene regulatory network.
3.4 Figures

3.4.1 Figure 1. NCoR binds to enhancers independent of MeCP2.

A. Genome browser view of MeCP2-repressed genes illustrating H3K27ac and NCoR ChIP-seq between *MeCP2* knockout and wild-type cortex as overlaid tracks. Below tracks are regions identified as enhancers from H3K27ac ChIP-seq and peaks of NCoR ChIP-seq signal.

B. Pie chart of NCoR ChIP-seq peak distribution across regulatory and non-regulatory elements.

C. Boxplot of quartiles of H3K27ac ChIP-seq quantified within NCoR ChIP-seq peaks. Spearman rho is shown as correlation between these two ChIP-seq marks. *** p<10^-8.

D. Boxplots of log2 fold changes in NCoR ChIP-seq in *MeCP2* knockout (left) and *MECP2* overexpression mice (right) compared to wild-type at enhancers identified as misregulated by MeCP2 by differential H3K27ac ChIP signal. *** p<10^-8, ** p<10^-3.
Data are from cerebral cortex 8-week-old wild-type, MeCP2 knockout, and MECP2 overexpression animals. n=2 for all NCoR ChIP-seq data.
3.4.2 Figure 2. mC-MeCP2-NCoR drives changes in multiple steps of enhancer activation.
A. Genome browser view of an MeCP2-repressed gene illustrating sequencing comparisons of H3K27ac ChIP-seq, ATAC-seq, and ATAC-seq filtered for nucleosome-free (NF) regions (sub 100bp). Below the tracks are highlighted regions denoting peaks called as NF ATAC peaks.
B. Boxplots of log2 fold-changes in ATAC-seq in Mecp2 knockout (left) and MECP2 overexpression (right) mice compared to wild-type at misregulated enhancers identified by differential H3K27ac ChIP-seq. ***p<10^{-8}, **p<10^{-3}.
C. Boxplots of log2 fold-changes in ATAC-seq across all enhancers genome-wide sorted into deciles based on total amounts of mCA per kilobase (left) and mCG per kilobase (right). Spearman correlations listed show the correlation between ATAC-seq fold-changes and levels of methylation at enhancers. ***p<10^{-8}.
D. Boxplots of log2 fold-changes in nucleosome-free (NF) ATAC-seq in Mecp2 knockout (left) and MECP2 overexpression (right) mice compared to wild-type that overlap with misregulated enhancers identified by differential H3K27ac ChIP-seq. ***p<10^{-8}, **p<10^{-3}, *p<0.05.
E. Boxplots of log2 fold-changes in nucleosome-free (NF) ATAC-seq across all enhancers genome-wide that a NF ATAC peak overlaps. The enhancers are sorted into deciles based on total amounts of mCA per kilobase (left) and mCG per kilobase (right). Spearman correlations listed show the correlation between ATAC-seq fold-changes and levels of methylation at enhancers that the ATAC peaks overlap with. ***p<10^{-8}.
F. Boxplots of log2 fold-changes in CUT&Tag signal for H3K4me1 (left) and Mediator (MedI) (center) and ChIP-seq for Cohesin (Smc1) (right) in Mecp2 knockout compared to wild-type mice at misregulated enhancers identified by differential H3K27ac ChIP-seq. ***p<10^{-8}, **p<10^{-3}.

155
C. Boxplots of log2 fold-changes in H3K4me1 CUT&Tag signal across all enhancers genome-wide sorted into deciles based on total amounts of mCA per kilobase (left) and mCG per kilobase (right). Spearman correlations listed show the correlation between CUT&Tag fold-changes and levels of methylation at enhancers. ***p<10^{-8}.

Data are from cerebral cortex of 8-week-old wild-type, Mecp2 knockout, and MECP2 overexpression animals. n=2 for ATAC-seq and H3K4me1 CUT&Tag data; n=1 for Mediator CUT&Tag and Cohesin ChIP-seq data.
Chapter 4: Concluding remarks and future perspectives

Ongoing experiments and analyses discussed in this chapter were done as a collaboration between Dennis Wu (DYW), James Moore (JRM), and myself (AWC). AWC performed the ChIP-seq experiments and analysis, and DYW has been fundamental in furthering the understanding of neuronal patterns of mCH. JRM has pioneered the cell type-specific studies using the INTACT method and performed the RNA-seq experiments and analysis following isolation. AWC has performed the ChIP-seq experiments and analysis following INTACT isolation.
4.1 Introduction
Recent studies have discovered large quantities of mCH in neurons, defined cell-type specific profiles for this methyl mark, and linked it to the essential functions of MeCP2 (see Chapters 1-3). These findings have demonstrated that mCH is a critical epigenetic component of the mammalian brain. However, important questions remain to be answered. Progress has been made in identifying how chromatin architecture and gene expression states define mCH profiles in neurons, but insights into the molecular mechanisms that recruit and activate DNMT3A are still needed. Integrated methylomic, transcriptomic, and epigenomic studies have revealed strong candidate models for the long-enigmatic mechanism of MeCP2-mediated transcriptional control. However, rather than working through a previously proposed NCoR recruitment model, it appears that MeCP2 and NCoR may act to regulate early steps of enhancer activation and affect downstream recruitment of key coactivators. However, the precise site of direct action by MeCP2-NCoR remains to be determined. Identification of this mechanism may also shed light on why intragenic enhancers are more affected than extragenic enhancers. Finally, accumulating evidence has implicated neuronal mC in diseases beyond MeCP2 disorders, suggesting other candidate genes for additional involvement exist that have not yet been tested. Here I discuss these outstanding questions, preliminary results addressing them, and future studies that can shed light on mechanisms of enhancer and gene expression control through the mC-MeCP2-NCoR pathway.

4.2 Mechanisms of unique non-CG methylation patterns in the brain
Our finding that the enhancers and genes most repressed by MeCP2 are not only enriched for mCH themselves, but are found within large megabase-sized regions of high CH methylation (Clemens
and Gabel, in press; Clemens et al., 2020) raises the question of how these large regions of methylation are put in place. DNMT3A is highest expressed at 2 weeks in mouse cortex, however, it is unknown what factors during this early timepoint determine how DNTM3A will deposit methylation across the neuronal genome. One candidate that is emerging from published studies and my preliminary work is the recruitment of DNMT3A to histone H3 methylated at K36. It has previously been shown that DNMT3A binds strongly to H3K36me3, and more recently has been identified to bind with greater affinity to H3K36me2 in embryonic stem cells (Rondelet et al., 2016; Weinberg et al., 2019). Examination of the relationship between DNTM3A and H3K36me2 within neurons will provide valuable information on the origin of neuronal mCH patterning.

To understand if the affinity of DNMT3A to these histone modifications are relevant to the establishment of the neuronal epigenome, it will be interesting to assess the genomic profiles of H3K36me2 and DNMT3A in 2-week wild-type mouse cortex, when DNMT3A is at its peak expression. If H3K36me2 is a strong binding target of DNMT3A, we would expect that H3K36me2 would mark regions of the genome for DNA methylation deposition by DNMT3A and that regions with enriched ChIP-seq for H3K36me2 and DNMT3A at 2 weeks would correlate with enriched regions of mCH in adult, 8-week mouse cortex. To assess if this is the case, I performed ChIP-seq for DNMT3A and H3K36me2 in 2-week mouse cortex. This analysis shows a strong correlation of 2-week DNMT3A ChIP signal and 8-week mCH levels across TADs identified in mouse cortex that is consistent with our previous analysis (Clemens et al., 2020) (Figure 1A). Assessment of H3K36me2 indicates robust correlations between 2-week H3K36me2 and DNMT3A in TADs, as well as 2-week H3K36me2 and 8-week mCH in TADs (Figure 1B, C). Together, these results may suggest that the mechanism of DNMT3A recruitment by H3K36me2
is conserved in neurons and that this facilitates the unique, broad domains of mCH deposition that occur in these cells.

4.3 Understanding enhanced susceptibility of intragenic enhancers to repression by MeCP2
A notable aspect of enhancer regulation by mC-MeCP2-NCoR that we have observed is that intragenic enhancers are more affected by MeCP2 repression and DNA methylation than enhancers found outside of genes. Classically, distance and position of enhancers has not been known to play a critical role in the capability of an enhancer to activate its target promoter (Nord and West, 2020; Schoenfelder and Fraser, 2019). However, we unexpectedly identified several location-dependent effects of MeCP2 on acetylation at enhancers and non-regulatory sequences:
1) Enhancers that are located within a gene were more susceptible to dysregulation compared to enhancers outside of genes upon loss or overexpression of MeCP2 and loss of mCA from DNMT3A cKO. 2) Enhancers within a gene contact their promoter more than equal-distanced enhancers outside the gene. 3) Genes that contain more enhancers are more susceptible to regulation by MeCP2 than genes that contain fewer enhancers (Clemens et al., 2020). These findings indicate that intragenic enhancers may play a leading role in gene regulation through MeCP2.

One possibility for the observed bias towards intragenic enhancers is the environment created by promoter-enhancer loops created during transcription. A recently-proposed model of transcriptional extrusion suggests that instead of polymerase progressing along the transcribed chromatin, polymerase machinery “reels-in” DNA, bringing the chromatin to the site of
transcriptional initiation (Zheng et al., 2019). In this scenario, intragenic enhancer-promoter contacts would be selectively increased compared to extragenic contacts as genes are transcribed and DNA is pulled in through the polymerase, creating more contacts. One newly proposed mechanism of regulation by MeCP2 depends on the formation of MeCP2 condensates within regions of high protein density (Fan et al., 2020; Li et al., 2020; Wang et al., 2020). For MeCP2 bound to enhancers, if enhancer-promoter loops are being pulled towards the promoter during transcription, this could increase the number of contacts of loops with the promoter in a region of compacted chromatin. Thus, the collection of enhancers created at the promoter will also raise the density of MeCP2 and restrict the access of other coregulators within this region through molecular crowding (Bancaud et al., 2009). Enhancers outside of genes would not be exposed to this increased density as transcriptional extrusion would only occur within the gene itself, thereby allowing extragenic enhancers to escape high concentrations of MeCP2 and stronger repression by MeCP2. Comprehensive assessment of epigenomic changes in MeCP2 mutants and studies exploring how the role of MeCP2 in liquid-liquid phase separation or chromatin condensation contributes to its functions in gene regulation promise to further solidify our understanding of MeCP2.

4.4 Understanding cell-type specific roles for mCH and MeCP2
On a finer resolution than brain regions composed of heterogenous populations of cells, there exists a clear role for mCH and MeCP2 in cell-type specific gene regulation. Levels of mCH are highly cell type-specific, particularly methylation within gene bodies, which strongly correlates to cell type-specific gene expression (Lister et al., 2013; Luo et al., 2017; Mo et al., 2015; Stroud et al.,
However, systematic studies across cell types are needed to understand the relative importance of mCH in each cell type and to determine the complements of genes most impacted by its regulation. For example, there are large differences in total mCH levels across the genomes of neuronal subtypes, and it is possible that mCH and MeCP2 have larger impacts on gene regulation in high-mCH subtypes compared to lower-mCH subtypes. Therefore, loss of mCH or MeCP2 in disease may disproportionately impact these high-mCH subtypes, manifesting in specific circuit defects. Moreover, it has not yet been determined how the mechanism of MeCP2 and DNA methylation identified in whole cortex to control enhancer function is utilized at a cell type-specific level. Studies using cell type-specific chromatin profiling, or single cell methods can identify how mCH and MeCP2 affect cell type-specific enhancers.

In preliminary studies, we have begun to explore effects of cell type methylation within MeCP2 knockout and MECP2 overexpression mice. Using the INTACT (Isolation of Nuclei Tagged in Specific Cell-Types) method (Deal and Henikoff, 2011; Mo et al., 2015) to isolate individual classes of neurons, we have initially profiled changes of gene expression and enhancer activity within Parvalbumin-positive (PV) interneurons isolated from MeCP2 mutant and wild-type cerebral cortex. PV+ neurons contain high levels of mCH compared to other neuronal classes, therefore providing a valuable substrate to detect mC-MeCP2 regulatory effects (Luo et al., 2017; Mo et al., 2015). In PV+ neurons, we found a significant overlap of misregulated genes and enhancers with whole cortex studies, as well as PV-specific changes mediated by MeCP2. Continued studies will profile Somatostatin-positive (SST) interneurons, another class with high levels of mCH (Luo et al., 2017; Mo et al., 2015), as well as two classes of excitatory neurons. These include layer V Rpb4+ neurons with relatively high levels of mCH and layer IV Nr5a+
neurons with relatively low levels of mCH (Luo et al., 2017; Mo et al., 2015; Tasic et al., 2018). Results from these diverse classes of neurons with varying levels of CH methylation will provide valuable information for the biological function of mCH and MeCP2 within individual cell types.

4.5 Disruption of mCH and MeCP2 in neurodevelopmental disease
We now have evidence across multiple studies that MeCP2 does affect the transcription of many, if not all genes, within neurons. However, it has been difficult to identify the importance of any one gene due to the small degree to which most genes are affected. Notably, this pattern of gene dysregulation follows the developmental pattern of MeCP2 expression in neurons; gene expression studies in MeCP2 knockout and missense mutation mice have found that the classically observed changes of gene expression in adult mouse brain are not present early in development between 1-4 weeks (Baker et al., 2013; Boxer et al., 2020). Because MeCP2 is very lowly expressed during this time period and there is no genome-wide change in gene expression, this highlights the importance MeCP2 plays in regulating the expression of these genes as it accumulates during neuronal maturation. The lack of gene dysregulation earlier in development corresponds with the delayed deterioration of neuronal function in individuals with Rett syndrome that occurs after largely normal development in the period prior to the buildup of MeCP2. As has been observed, Rett syndrome has a late onset after a seemingly normal development, which corresponds with the late onset appearance of mCH and MeCP2 in neurons. These events suggest that early neuronal maturation is affected when MeCP2 is disrupted. Critically, it has been shown in mice that re-expression of MeCP2 between birth and 6-weeks can restore phenotypes back to wild-type state. This includes re-expression of MeCP2 after Rett-associated phenotypes manifest, suggesting a
reversible nature of this disease (Katz et al., 2016; Sinnett and Gray, 2017; Vashi and Justice, 2019). These promising results in mice provide hope that an effective therapeutic can be used to treat children with Rett syndrome, even after symptoms arise.

As discussed in Chapter 1, the balance of the mC-MeCP2-NCoR pathway has been shown to be essential for proper neuronal regulation, as mutations within this pathway drive neurodevelopmental disease, including Rett syndrome, MeCP2 duplication syndrome, autism spectrum disorder, and intellectual disability. We have also illustrated the sensitive nature of neuronal DNA methylation through studies of heterozygous disease-causing mutations of DNMT3A causing behavioral defects and recapitulating effects of neuronal enhancer dysregulation (Christian et al., 2020).

The role of histone methyltransferases has also emerged as a critical factor in driving proper neuronal regulation. ASH1L and NSD1 are proteins responsible for H3K36 di-methyl modifications (Huang and Zhu, 2018; Weinberg et al., 2019), which our preliminary evidence has linked to patterning of mCH within neurons. If the DNA methylation regulatory system relies on H3K36me2 at an early stage in development, disruption of this histone methyl-mark should impact the ability of DNMT3A to deposit mCH. Disruption of ASH1L and NSD1 in neurons should prevent methylation of H3K36, which in turn may disrupt normal patterns of transcription and normal patterning and buildup of mCH by DNMT3A. These results would further support a connection between early postnatal patterns of H3K36me2 and adult patterns of DNA methylation and would raise the possibility that mCH is disrupted in neurodevelopmental disorders associated with ASH1L and NSD1.
It will also be important to determine if mechanisms of DNA methylation deposition in neurons are specific to particular cell types or a common pan-neuronal mechanism. Studies identifying mechanisms that drive patterns of neuronal methylation and determining if disruption of molecular components involved in this process occur in disease may provide valuable information for potential therapeutics for neurodevelopmental disorders.
4.6 Figures

4.6.1 Figure 1. Early developmental H3K36me2 distribution correlates with adult patterns of mCH.
A. Comparison of TAD mCA/CA levels at 8 weeks and DNMT3A ChIP/Input in TADs at 2 weeks. ***p < 10^-8.

B. Comparison of DNMT3A ChIP/Input at 2 weeks and H3K36me2 ChIP/Input in TADs at 2 weeks. ***p < 10^-8.

C. Comparison of TAD mCA/CA levels at 8 weeks and H3K36me2 ChIP/Input in TADs at 2 weeks. ***p < 10^-8.

Data are from cerebral cortex of 2- or 8-week-old wild-type animals. n=2 for DNMT3A and H3K36me2 ChIP-seq and for DNA methylation (Stroud et al., 2017).
References


Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson,


Rahman Syndrome: A clinical study of 55 individuals with de novo constitutive DNMT3A variants [version 1; referees: 3 approved]. Wellcome Open Res. 3.


Xiang, Y., Tanaka, Y., Patterson, B., Hwang, S.-M., Hysolli, E., Cakir, B., Kim, K.-Y., Wang,


Curriculum Vitae

ADAM W. CLEMENS
814 McDonnell Science ● St. Louis, MO 63110
https://www.linkedin.com/in/adamclemens/

EDUCATION
Washington University in St. Louis, St. Louis, MO, 2015 – 2020
PhD Candidate, Division of Biology and Biomedical Sciences

Pennsylvania State University, Eberly College of Science, University Park, PA
B.S. with Honors, Biotechnology, 2013 (Schreyer Honors College)
Minor, Microbiology, 2013

PROFESSIONAL EXPERIENCE
Ascidian Therapeutics, Inc., Brighton, MA
Scientist I, starting September 2020

• Developing gene therapy programs.

JBS Science, Inc., Doylestown, PA
Research Assistant, Advisor – Dr. Wei Song, 2013 - 2015
Research Focus: Identifying DNA biomarkers of colorectal cancer in urine to develop a noninvasive urine DNA test for early detection of cancer.

• Developing PCR-based and amplicon NGS-based assays with a panel of genetic and epigenetic biomarkers for early detection of colorectal cancer.
• Responsible for laboratory inventory and ordering.

RESEARCH EXPERIENCE
Washington University in St. Louis, St. Louis, MO
Doctoral Research, Advisor – Dr. Harrison Gabel, 2016 – present, anticipated graduation August 2020
Research Focus: Identifying novel neuron-specific mechanisms of transcription control using Rett Syndrome as a disease model.

• Designed and performed next-generation sequencing experiments of RNA-seq, ChIP-seq, ATAC-seq, and Cut-and-Tag to prepare libraries for gene expression and chromatin analysis in disease models.
• Developed R and Python cluster-computing analysis pipelines to analyze and interpret self-generated sequencing data.
• Designed sgRNA and generated dCas9 constructs to manipulate DNA modifications through in vitro viral infections in cell culture to determine role in gene regulation.
• Identified novel mechanism of gene regulation through MeCP2 at enhancers \textit{in vitro} and \textit{in vivo} with mouse models.
• Worked in highly collaborative environment to produce a co-first-author paper, a middle-author paper under review, foundations of a second co-first-author paper, and a review article under review.

\textbf{Pennsylvania State University}, University Park, PA  
\textbf{Honors Thesis Dissertation}, Advisor – Dr. Mark Guiltinan, 2010 - 2013  
Research Focus: \textit{Honors Thesis Research - Molecular Cloning and Functional Analysis of LEAFY COTYLEDON2 Gene of Theobroma cacao.}

• Collaborated with graduate students on the experimental design of the project.
• Independently carried out experimental pipeline from gene isolation and cloning to analyzing functions of transgene in tissue culture.
• Increased somatic embryogenic potential of cacao tissue through over-expression of transgene.

\textbf{Pennsylvania State University}, University Park, PA  
\textbf{Research Assistant}, Advisor – Dr. Charlie Anderson, 2013  
Research Focus: Developing assays to culture Tobacco BY-2 cells in micro-volumes for high throughput genetic analysis.

• Maintain cultures of BY-2 tobacco cells under sterile conditions.
• Analyze growth rates of BY-2 cells under varying culture conditions using confocal microscopy.

\textbf{The Ohio State University}, College of Pharmacy, Columbus, OH  
\textbf{Summer Undergraduate Research Fellow}, Advisor – Dr. Robert Brueggemeier, 2011  
Research Focus: Identifying protein-drug interaction of potential Estrogen-receptor positive breast cancer drug.

• Maintained cultures of breast cancer cell lines under sterile conditions.
• Performed protein purifications via ATP-affinity column chromatography to analyze specific drug-protein interactions.

\textbf{Publications}
\textbf{Clemens, AW.}, Gabel, HW., Emerging insights into the distinct neuronal methylome. \textit{Trends in Genetics} in press.

DOI: https://doi.org/10.1101/2020.07.10.195859


**PROVISIONAL PATENT**


**PRESENTATIONS**


**Clemens, A.,** Wu, D., Zhao, G., Gabel, HW. MeCP2 Represses Enhancers through Chromosome Topology-Associated DNA Methylation. Heidelberg, Germany, EMBO Wokshop: Chromatin and Epigenetics, May 2019 (Poster).

**Clemens, A.,** Wu, D., Gabel, HW. Dissecting MeCP2-mediated repression of neuronal transcription. Washington University in St. Louis, Developmental Biology Poster Session, June 2017 (Poster).

**Clemens, A.,** Lin, S., Jain, S., Su, YS., Song, W. Detection of Colorectal Cancer-Associated Genetic Alterations in Urine of Patients with CRC. AACR Annual Meeting, April 2014; Sidney Kimmel Cancer Center Consortium Symposium, October 2014; Drexel University Discovery Day, October 2014 (Honorable Mention – Technician). (Poster).


**TEACHING AND MENTOR EXPERIENCE**

Teaching Assistant – Washington University – Genomics graduate-level course, 2017

- Designed and presented teaching lessons for 1st-year graduate students in genomic applications.
• Guided and assisted students with computational genomics techniques and concepts.

Research Mentor – Washington University – Gabel Lab, 2016-2020
• Mentored four graduate students in next-generation sequencing techniques for them to independently perform experimental and computation analysis.

Honors and Professional Activities
NIH F31 Ruth L. Kirschstein Predoctoral Individual, 2018 - 2020
National Research Service Award – NINDS
• Awarded for 2 years of funding for proposed research
• Provides funds for career development and training

Science Communication Credential – Washington University
In progress, to be completed by Fall of 2020
• Gain experience in scientific communication through coursework, practical experience, and building a writing-centric portfolio.

InPrint – a scientific editing network – Washington University
Advertising Coordinator, 2020 – present
• Responsible for ensuring awareness of our services and availability to our Medical School campus and our Main campus.
Member, 2019 – present
• Edit and review manuscripts, grants, and abstracts from internal faculty or students

ProSPER – promotes science policy, education, and research – Washington University
Member, 2019 – present
• Organized Science Communication Career Panel for graduate students

NSF Graduate Research Fellowship Program – Honorable Mention, 2017

NIH T32 GM007067 Cell and Molecular Biology Training Program, 2016 – 2018

American Association for Cancer Research (AACR) Associate Member, 2013 – 2015

Post-Baccalaureate Apprenticeship Program at Baruch S. Blumberg Inst., 2013 – 2015
• Provides an opportunity to experience the research environment similar to graduate school.
• Present a recent peer-reviewed publication every 6 months at weekly journal clubs.
• Present progress of research project every 12 months in a 1-hour long seminar format.
• Provides opportunities to meet expert guest speakers.

• Provide writing and editing assistance for the preparation of NIH grants and peer-reviewed manuscripts.

Biotechnology Student Marshal – Penn State University, 2013
• First in Biotechnology graduating class.

Co-founder and Secretary of Biotech Club – Penn State University, 2012-13
• Organized guest lecture series to introduce interested students to faculty research and other opportunities related to biotechnology.

Foster Memorial Scholarship – Penn State University, 2012-13
• Recognition for outstanding academic achievement in Biochemistry and Molecular Biology.

Summer Discovery Research Grant – Penn State University, 2012
• Additional support from Eberly College of Science.
• Funding to support full-time undergraduate summer research.

College of Agricultural Sciences Grant – Penn State University, 2011-12
• Funding opportunity for independent undergraduate research.