5-7-2024

Regulation of neuronal cell-type-specific transcription by non-CG methylation and MeCP2

James Russell Moore
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Division of Biology and Biomedical Sciences
Neurosciences

Dissertation Examination Committee:
Harrison Gabel, Chair
Joseph Corbo
Tim Holy
Jason Yi
Andrew Yoo

Regulation of Neuronal Cell-Type-Specific Transcription by Non-CG Methylation and MeCP2

by
J. Russell Moore

A dissertation presented to
Washington University in St. Louis
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2024
St. Louis, Missouri

Table of Contents

List of Figures

Acknowledgments

I would like to thank members of the Gabel lab. I am very grateful that I got to work with such an amazing group of people. Each person in their own unique way contributed to my growth professionally, scientifically, and personally. They also ensured that I had a great deal of fun along the way. Harrison, thank you for all the support, the scientific training, for the meetings that never end on time because there are always more crazy ideas to talk about, the movie quotes, and encouragement. I will always admire and appreciate your dedication, enthusiasm, and kindness. Thank you Mati for your diligence and sense of humor. Sabin for your toughness, energy, and laughter. Diana for your kindness and cheer. Nicole for your scientific enthusiasm and ambition, in addition to your silliness. Adam for the coffee breaks and scientific rigor. Dennis for your quirkiness and willingness to help. Jenna for your conscientiousness. To all others -Yasmin, Ryan, Chibueze, Aaron, Alyssa, Ladan, and more – thank you.

Thank you to the funding sources that have made the work in this dissertation possible. The Washington University MSTP and the National Institute of Child Health and Human Development for funding my F30 application (F30HD102147). 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 NIM R01MH117405 to HWG.

Thank you to my thesis committee: Andrew Yoo, Joe Corbo, Tim Holy, and Jason Yi. I greatly appreciate you investing your time to provide me with scientific feedback and professional advice.

Thank you to the members of the Department of Neuroscience, the Yi lab, the Bonni lab, the Li lab, and the Zhao lab who through seminars, lab meetings, coffee breaks, happy hours, parties, and hallway chats formed a stimulating scientific community.

Thank you to my friends for the support and good times. To all of my family, my many aunts, uncles, and cousins, who have supported me even though they never understood what I was up to. To my sister, Ellen, for being a champion of fun and adventure. Thank you to my wife Tina who proofread all my science writings and presentations. Also, who has cared for me during the ups and downs of graduate school. To Ziggy for being the best stress reliever. Thank you to my mom for endless advice, to my dad for always keeping me on my toes, and to both of you for great friendship, humor, and love.

With deep gratitude,
Russell

Washington University in St. Louis

May 2024

ABSTRACT OF THE DISSERTATION

Regulation of Neuronal Cell-Type-Specific Transcription by Non-CG Methylation and MeCP2

by

James Russell Moore

Doctor of Philosophy in Biology and Biomedical Sciences

Neurosciences

Washington University in St. Louis, 2024

Professor Harrison Gabel, Chair

Modulation of gene expression is fundamental to the exceptional complexity and diversity of cells in the nervous system. Within that last decade, evidence has emerged that neurons utilize unique forms of DNA methylation to regulate transcriptional programs, with multiple types of neurons displaying high levels of cytosine methylation in a non-CG context compared to non-neural cells. This non-CG methylation primarily occurs at CA dinucleotides (mCA) and is read out by methyl-CpG binding protein 2 (MeCP2). Inactivation of MeCP2 results in Rett Syndrome (RTT), a severe neurodevelopmental disorder, while overexpression of MeCP2 results in MeCP2 duplication syndrome, an autism spectrum disorder. MeCP2 has been identified as a critical reader of mCA and regulates transcription of long neuronal genes that are embedded in regions of high mCA. Both MeCP2 and mCA remarkably increase postnatally in the developing mammalian brain. Recent studies have identified that mCA patterns are highly cell-type specific, but how this methylation is read out by MeCP2 to dictate cell-type-specific transcriptional programs remains poorly understood. Here, I discuss the unique challenges to transcriptional regulation in the developing nervous system and how the mCA-MeCP2 pathway contributes to that process,

particularly to cell-type-specific transcription. I then systematically analyze the effects of loss of MeCP2 on distinct neuronal cell types. I show that the methylation level of a population of neurons determines the magnitude of dysfunction when MeCP2 is lost. I find evidence for both shared and distinct transcriptional regulation by MeCP2 across different cell types and identify enhancer elements as important sites of MeCP2 regulation to control cell-type-specific transcriptional programs. Finally, I provide evidence that MeCP2 and mCA are involved in the maturation and maintenance of stable transcriptomic subtype identities and discuss ongoing studies to follow up this exciting finding. These analyses provide a significant advance in our understanding of how MeCP2, along with mCA, regulate transcription within distinct types of neurons and allow for the specialization of neurons to drive the complex functions of the nervous system.

Chapter 1: The unique neuronal epigenome

1.1 - Challenges of neuronal transcriptional regulation

The mammalian nervous system comprises an immense web of cells and processes connected across trillions of synapses to produce cognition and behavior. In the face of this complexity, scientists have sought to organize the brain into its component parts. Santiago Ramon y Cajal, the so-called "father of modern neuroscience," began this process with the recognition that the brain contained individual units called neurons, rather than a single interspersed network. Using the new Golgi staining technique, he systematically characterized neurons by drawing sections of tissue by hand and was captivated by the breadth and diversity of cells in the nervous system (Ramón y Cajal, 1909). It was clear from his early drawings that there were stereotyped patterns of organization across the brain with different types of cells present in different regions. The cellular diversity and organization of neurons is integral for the multifaceted rhythms of the brain and is theorized to underlie the brain's impressive computational capacity (Kepecs & Fishell, 2014; Klausberger & Somogyi, 2008; MacNeil & Masland, 1998). How this complex biological system is created and maintained is the focus of developmental neuroscience.

As pioneering experiments have shown, the brain does not developmentally unfold strictly adhering to a pre-ordained orthodoxy. Famously, Hubel and Wiesel showed that sensory inputs are required to ensure proper development (Wiesel & Hubel, 1963). Maturing neurons and their resulting circuits are malleable and plastic by design. Postmitotic neurons dramatically migrate into the cortex in a cell autonomous process (Bystron et al., 2008; Lim et al., 2018). Upon settling into fixed positions, neurons expand and grow into their microenvironment, forming connections

with each other and refining those connections based on inputs. Some of these connections are transient. For instance, somatostatin (SST) neurons arrive in the developing cortex prior to parvalbumin (PV) interneurons and form temporary synapses with maturing excitatory cells that ultimately disappear after this early inhibitory role is fulfilled (Marques-Smith et al., 2016). Other early developmental stimuli cause more lasting changes in circuits. For example, the transcription factor Etv1 in PV interneurons reacts to activity inputs and acts as a molecular switch to divide PV interneurons into multiple subtypes with differing electrophysiological properties (Dehorter et al., 2015). Neurons are designed to receive local cues and adopt the characteristics necessary for the proper functioning of their surrounding circuit (Bandler et al., 2017; Mardinly et al., 2016; Wamsley & Fishell, 2017; Wonders & Anderson, 2006). If this learning on the job is by design, then how is it that neurons can encode these lessons and remember them throughout a lifetime?

At the molecular level, the morphologic and functional traits of individual neurons are products of expression of specific sets of genes. So how then is the phenotypic complexity of neurons reflected in their gene expression programs? To start, neurons express more genes than nonneuronal cells (Tasic et al., 2016) and specifically express the longest genes in the genome at a far greater level than other tissues (Gabel et al., 2015; Zylka et al., 2015). These long genes (e.g. TSS to TES > 100 kilobases) perform critical functions in the brain encoding ion channels and cell adhesion molecules (Gabel et al., 2015) and their expression serve as a marker of neuron maturity (McCoy et al., 2018). Targeted inhibition of these genes strongly impairs neuronal function (Mabb et al., 2014) and their dysfunction is implicated in neurological disorders (King et al., 2013).

The number of genes in the genome has not significantly increased with organismal complexity (King & Wilson, 1975), rather, there has been an increase in noncoding DNA during evolution

(Lander et al., 2001). This means that to generate more variety in cellular features there must be an increase in gene regulatory capability. Recently, it was discovered that neuronal genes utilize an expanded regulatory network involving greater number of regulatory elements, non-coding stretches of DNA such as enhancers that control transcription, from a larger genomic space (Closser et al., 2022). This increase appears to scale with the complexity of the organism and is the product of the evolution of novel vertebrate neuronal enhancers. Enhancers promote transcription of genes and show great specificity to cell-type and developmental stage (Heinz et al., 2015; Nord et al., 2013; Visel et al., 2013). The location and timing of their activation and inhibition must then be tightly regulated throughout development and the lifespan of the neuron by epigenetic regulators.

It is clear that the nervous system faces unique challenges to transcriptional regulation. An extended developmental regime must be dynamic and malleable, and a sophisticated gene expression network must be reliably coordinated. Given this, it should not be surprising that neurons have evolved unique epigenetic modifications to achieve dynamic stability in gene expression. Specialized epigenetic mechanisms allow neurons to engage a broader complement of regulatory elements in long genes that drive their exceptionally complex morphology and function.

1.2 - Unique neuronal methylation

The most common epigenetic modification to DNA in animal genomes comes in the form of methylation at the 5’ position of cytosine nucleotides (Schübeler, 2015). The addition of a methyl group can block binding of proteins such as transcription factors to DNA and function as a substrate for proteins to bind to it directly. Proteins that read out methylation are methyl-CpG

binding domain (MBD) proteins that are thought to contribute to methylation dependent repression (Schübeler, 2015). DNA methylation in mammals was initially characterized as occurring cytosines next to guanines (mCG) and shown to play an important role in differentiation, imprinting and development (Bird, 2002). This mCG accounts for nearly all the methylation in most tissues, but in recent years it was discovered that neurons contain noncanonical forms of methylation (Kriaucionis & Heintz, 2009; Lister et al., 2013). Methylation occurring at cytosines next to bases other than guanines is termed non-CG methylation (mCH, where H = A, C, or T), with the major form being methylation at cytosine-adenine dinucleotides (mCA). Whereas the percentage of methylated CG dinucleotides is high (60-90%), the percentage of mCH is much lower (2-6%), which caused it to be overlooked until the advent of well-controlled whole-genome base-resolution DNA methylation profiling (Lister et al., 2009; Xie et al., 2012). mCH in neurons reaches levels equivalent or greater to mCG levels, while being scarcely detected in other tissues (Lister et al., 2013). The exciting discovery of unique neuronal methylation initiated strong efforts to understand why mCH evolved as an essential feature of neurons and how it acts on the genome to regulate transcription.

mCH arose at the beginning of the vertebrate lineage and is deposited by the de novo methyltransferase DNMT3A (de Mendoza et al., 2021). The germline loss of *Dnmt3a* is lethal three to four weeks after birth (Okano et al., 1999). Deletion of *Dnmt3a* in the perinatal period using a floxed *Dnmt3a* and a Nestin-Cre driver blocks non-CG methylation buildup in neurons, resulting in severe neurological phenotypes and a shortened lifespan (Nguyen et al., 2007) In humans, heterozygous disruption of DNMT3A results in the Tatton-Brown-Rahman syndrome, an overgrowth and intellectual disability syndrome, and autism (Sanders et al., 2015; Satterstrom et

al., 2020; Tatton-Brown et al., 2018). Mouse models of this disorder show that while mCG is largely intact in the brains of these animals, non-CG methylation is globally reduced by half. The private use of non-CG methylation in neurons, and serious consequences upon disruption of the non-CG pathway indicate that this mark serves a novel and essential function in the mammalian brain, motivating continued studies to understand its mechanisms and function.

In euchromatin, mCH is found at its highest levels at regions surrounding genes within gene bodies of lowly expressed genes, while inactive, inaccessible heterochromatin lacks mCH (Kinde et al., 2016; Lister et al., 2013; Luo et al., 2017; Mo et al., 2015; Stroud et al., 2017). Like mCG, mCH shows signatures of acting as a repressive mark on the genome with an inverse correlation existing between mCH level and expression of a gene, and *in vitro* additions of mCH inhibit expression (Guo et al., 2014). In mice, mCH is deposited by DNMT3A in the first few weeks of life reaching its adult levels by 4-6 weeks. In humans, this postnatal increase is also observed with the majority of mCH accumulating in the first two years of life, followed a slow increased until approximately sixteen years of age (Lister et al., 2013). The increase in mCH occurs in postmitotic neurons that do not undergo DNA replication it is therefore maintained for the life of the neuron, with negligible change over the lifespan (Lister et al., 2013).

Two mechanisms have been identified for how DNMT3A is recruited to the genome and mCH patterns established. The first relates to the topology of the genome itself within the nucleus. Investigations into the three-dimensional structure of the genome using high-throughput chromatin conformation capture (Hi-C) have uncovered topologically associating domains (TADs), which are genomic regions with increased cis interactions (Dixon et al., 2012) TADs in neurons vary in

amounts of DNMT3A binding and mCH levels, such that genes and regulatory elements within a TAD are influenced by a baseline methylation "set-point." This points to a model where TADs organize DNMT3A activity to specific genomic stretches resulting in large regional differences in mCH levels (Clemens et al., 2020).

On a smaller scale, mCH levels at genes and regulatory elements are determined by the activity of these sites during the early postnatal period. Genes and regulatory elements that are active early in life are protected from mCH addition, whereas lowly expressed genomic elements accumulate mCH contributing to their maintained suppression (Stroud et al., 2017). This appears to be a fairly linear relationship with intermediately expressed genes accruing a median amount of mCH, indicating that large numbers of genes are influenced by mCH.

The discovery of topological and expression-associated influences on DNMT3A represent important insights into how mCH is added to the genome, but it is not yet clear how these two phenomena on different physical scales interact. For instance, is there a sequential regional addition of mCH followed by a more local, activity-mediated deposition? Investigations into the molecular machinery that influences DNMT3A to create these different methylation patterns are ongoing and will provide a better understanding of how different methylation patterns are established.

1.3 - Heterogeneity of mCH patterns across neuronal cell types

A defining aspect of mCH biology is that it displays remarkably cell-type-specific patterning. Cell-type differences in expression early in development result in cell-type-specific patterns of

methylation which are then maintained throughout the life of the neuron (Stroud et al., 2017). Genes that are differentially expressed between cell types are also differentially methylated such that a gene that is highly expressed in a cell type will be lowly methylated in that cell type and vice versa. This correlation between mCH and expression is so strong that mCH signal at a gene is a better predictor of expression than mCG or open chromatin (Mo et al., 2015). The technical advance to allow for methylation profiling by single cell bisulfite sequencing extended this point (Luo et al., 2017). Single cell methylome profiling from the brain has revealed numerous neuronal cell types can be distinguished by their methylomes alone (Luo et al. 2017). An artificial neural network trained on single cell mCH data can predict the regional and laminar position of a cell type across the entire mouse brain (H. Liu et al., 2021). The sensitivity to detect high resolution neuronal types based on methylation patterns is striking. For example, within excitatory neurons of the same class, such as Layer 5 extra telencephalic projecting (L5 ET), methylation differences can distinguish cells based on their brain region and, within that region, their projection targets (Z. Zhang et al., 2021). Integration of this single cell methylome data with single cell ATAC-seq allows for identification of regulatory elements and their associated transcription factors (Liu et al. 2021). The variability in local patterns of mCH at genes and regulatory elements points towards a highly specialized role for this unique methylation in neurons.

Different neuronal brain regions and cell populations also show variations in total amount of mCH. In mammals, high methylation brains regions can have double the amount of global methylation compared to low methylation regions (Christian et al., 2020; Gabel et al., 2015; Guo et al., 2014; Lister et al., 2013). Within brain regions, neuronal cell classes fluctuate in amounts of mCH with some cells, such as layer 4 excitatory neurons, containing less than half the amount of methylation

found in high mCH cell classes such as somatostatin-positive (SST) interneurons. Even within a class of neurons, such as SST interneurons, there are subtypes with different levels of mCH (Luo et al. 2017, Liu et al. 2021). As a general trend, neurons in lower layers contain more methylation than those in upper layers. In development, these lower layer neurons arrive first to the growing cortex. An intriguing hypothesis, yet to be tested, is that deep layer neurons begin to build up methylation earlier because they are born earlier, which results in the relative enrichment of mCH. While mCH shows dramatic global variability, mCG levels between cell types are stable (C. Luo et al., 2017; Mo et al., 2015). What leads to this unevenness is unknown, but it suggests that mCH and, by extension, MeCP2, the only known reader of mCH, play a larger regulatory role in some cell types than others.

1.4 - MeCP2 as a critical reader of mCH

The biological effects of DNA methylation are in part mediated by proteins that bind to it. An important reader of mCH is the methyl-CpG binding protein 2 (MeCP2). Mutations in the X-linked *MECP2* cause Rett syndrome, a developmental disorder almost exclusively affecting females (Amir et al., 1999). Rett syndrome patients display normal development until one to two years of age, when they begin to regress in social, language and motor skills (Hagberg et al., 1986; Leonard et al., 2017). Duplication of MeCP2 leads to an autism spectrum disorder, MeCP2 duplication syndrome (MDS) (del Gaudio et al., 2006; van Esch et al., 2005). Additional *MECP2* mutations have been linked to other psychiatric disorders, such as schizophrenia and Asperger's syndrome (D. Cohen et al., 2002; Curie et al., 2017).

MeCP2 contains a methyl-binding domain and was originally discovered based on its binding to mCG (Meehan et al. 1992). However, it has since been shown to possess high affinity for mCH sites in the genome, particularly the most prevalent site, mCA (Chen et al., 2015; Guo et al., 2014). ChIP-seq studies to measure MeCP2 binding show it is spread widely across the genome with an increase in methylated regions (L. Chen et al., 2015a; Kinde et al., 2016b; Lagger et al., 2017a; Mellén et al., 2017). Evidence indicates MeCP2 can bind unmethylated DNA a low affinity *in vitro* (Meehan et al., 1992), but that this does not occur *in vivo* (Connelly et al., 2020). MeCP2 levels also increase during the postnatal period along with mCH (Skene et al., 2010). Mice with an edited MeCP2 that can bind mCG but not mCA show severe phenotypes on par with complete MeCP2 KO, which provides strong evidence that mCA is the proximal substrate for MeCP2 not mCG (Tillotson et al., 2021). Ultimately, MeCP2 reaches near histone levels in neurons and is expressed ten times higher in neurons than other cell types (Skene et al., 2010). Multiple function for MeCP2 have been posited (Chahrour et al., 2008), but it is most highly validated as a transcriptional repressor (Lyst & Bird, 2015). This is in part due to its strongest interaction being with the Nuclear Co-Repressor complex (NCoR). Mutations in MeCP2 that interrupt its interaction with NcoR are known to cause Rett syndrome, and *in vitro* studies show that MeCP2 fails to repress gene expression in the absence of NcoR (Lyst et al., 2013a).

Given that MeCP2 is a methyl-binding protein and acts as a transcriptional repressor, functional studies of MeCP2 have focused on assessing alterations in gene expression when the protein is disrupted. Transcriptomic studies of mutant mice and human Rett syndrome patients show subtle expression changes in large numbers of genes (Boxer et al., 2020; Clemens et al., 2020; Gabel et al., 2015; Ip et al., 2018; Tudor et al., 2002). This, coupled with the broad binding of MeCP2 across

the genome, have made identifying specific large-effect targets of MeCP2 regulation difficult (Chen et al., 2015; Gabel et al., 2015; Kinde et al., 2016; Mellén et al., 2017; Stroud et al., 2017). Genes that are upregulated upon the loss of MeCP2 and downregulated when MeCP2 is overexpressed are termed MeCP2-repressed genes. Genes that show the opposite pattern are called MeCP2-activated genes. MeCP2-repressed genes consistently demonstrate a few notable characteristics. While they do not show enriched DNA methylation at promoter regions, the classical site of DNA methylation-mediated repression, they show a signal for enrichment of mCH within their gene body and in the region surrounding the gene (Clemens et al., 2020; Gabel et al., 2015; Kinde et al., 2016; Stroud et al., 2017). MeCP2-activated genes, on the other hand, do not share a robust enrichment of mCH signal. MeCP2-repressed genes, but not MeCP2-activated genes, show a modest enrichment for MeCP2 binding from ChIP-seq experiments (Boxer et al., 2020; L. Chen et al., 2015; Clemens et al., 2020; Lagger et al., 2017; Renthal et al., 2018; Stroud et al., 2017). MeCP2-repressed genes have been identified as unusually long and containing high numbers of enhancers within their gene body (Clemens et al., 2020; Gabel et al., 2015; Sugino et al., 2014). The degree of misregulation of a gene upon loss of MeCP2 is correlated with the amount of mCH at the gene and the length of the gene, with longer, higher methylated genes showing greater change in expression (Gabel et al., 2015). Further strengthening the connection between mCH and MeCP2 are data that demonstrate high similarity in gene expression changes between models lacking MeCP2 and those lacking mCH (Christian et al., 2020; Clemens et al., 2020; Lavery et al., 2020; Stroud et al., 2017, 2020). Additionally, without mCH present, MeCP2 fails to show enrichment for binding to the regions of DNA that contain high levels of mCH in wild-type tissue (Clemens et al., 2020). In total, these findings show that MeCP2 binds mCH to repress transcription of long neuronal genes.

1.5 - MeCP2 regulates transcriptional initiation through repression of intragenic enhancers

The mechanism by which mCH and MeCP2 regulate transcription has been an area of intense interest for the field. Given the presence of high levels of mCH and MeCP2 in the gene body of repressed genes but not in the promoter, it was hypothesized that mCH and MeCP2 may prevent transcriptional elongation of long genes by affecting polymerase's progression through the gene. This "speed-bump" model was supported by mathematical modeling of methylation on RNA polymerase II function (Cholewa-Waclaw et al., 2019). Studies from our lab and another group independently tested this model and surprisingly discovered that initiation of transcription, rather than elongation, is affected by binding of MeCP2 to mCA, the highest affinity mCH site, within the gene body (Boxer et al., 2020; Clemens et al., 2020). This motivated analysis of enhancers as regulatory elements which could affect transcriptional initiation at a distance. Using chromatin immunoprecipitation (ChIP) for histone H3 lysine 27 acetylation (H3K27ac), a marker for enhancers and promoters (Creyghton et al., 2010), we found consistently increased H3K27ac signal in the MeCP2 KO compared to wild-type littermates. These MeCP2-repressed enhancers overlapped with open chromatin and were significantly enriched for mCA. Additionally, MeCP2-repressed genes display significant upregulation of their enhancers in the MeCP2 KO. Interestingly, no changes in were found in the looping of enhancers to interact with their target promoters in the MeCP2 KO, suggesting that MeCP2 acts locally to inhibit enhancer activation of promoters but does not impact long range chromatin topology. From these experiments, we conclude that MeCP2 binds to mCA at intragenic enhancers to prevent their activity and block the subsequent induction of transcriptional initiation that these enhancers drive (Clemens et al., 2020).

1.6 - MeCP2 is critical to the maturation and maintenance of neurons

The severe consequences of disruption of MeCP2 in humans and mice confirm that it serves a critical function. In the most well-studied mouse model of Rett syndrome, male mice lacking MeCP2 progressively deteriorate after birth dying by six to eight weeks of age, while female heterozygous MeCP2-null mice, who contain wild-type MeCP2 in approximately half of their cells, survive into adulthood but progressively reproduce many Rett-like symptoms (Guy et al., 2001). In mice, deletion of Mecp2 in neurons only produces equivalent effects to broader loss in all cells (R. Z. Chen et al., 2001; Guy et al., 2001). It is clear from humans and mice that disruption of MeCP2 causes distinctly neuronal phenotypes (Lombardi et al., 2015).

Investigations into cellular and circuit phenotypes in MeCP2 mutant mouse models and patient brain tissue have revealed important insights into MeCP2 function. The brains of Rett syndrome patients are smaller, but this is not due to a loss of neurons but a shrinkage in neuronal size and dendritic arborization (D. Armstrong, 2005; Bauman et al., 1995; R. Z. Chen et al., 2001). Additional cellular phenotypes profiled in MeCP2 mutant mice and humans have found significant synaptic deficits (D. Armstrong et al., 1995; Bauman et al., 1995; Blackman et al., 2012; Chao et al., 2007; Jiang et al., 2013; Tropea et al., 2009). Systems level analysis into the consequences of loss of MeCP2 have produced a large body of literature indicating dysfunction across many domains including disinhibition, hypo- and hyper-connectivity, maternal rearing behavior, and critical period closure (Chao et al., 2010; Ito-Ishida et al., 2015; Krishnan et al., 2015; W. Li et al., 2016; Sceniak et al., 2016). While these studies have provided important evidence of the physiologic consequences of MeCP2 disruption, a broader model of MeCP2 that connects its molecular mechanism to effects at the cellular and circuit level is lacking.

MeCP2 is increasingly expressed in postnatal neurons as the cortex develops and is believed to be involved in the maturation of neurons rather than cell-fate decisions (Kishi & Macklis, 2004, 2010). In MeCP2 mutant mice, while neurons are smaller and have reduced dendritic arbors, they differentiate normally and achieve correct cortical positioning (Chao et al., 2007; R. Z. Chen et al., 2001; Kishi & Macklis, 2004, 2010). It has been proposed that the inability to progress through this postnatal terminal differentiation leads neurons to revert back to an immature state as some data have indicated (Fukuda et al., 2005; He et al., 2014a; X. Liu et al., 2020). Conditional deletion of *Mecp2* in the adult leads to many of the same phenotypes seen in the complete *Mecp2* null mouse implicating MeCP2 as critical for maintenance of nervous system function (McGraw et al., 2011). Transplant of MeCP2-null L2/3 pyramidal neurons into wild-type and MeCP2-null mouse cortices at P2-3 both show equivalent loss of dendritic complexity, indicating that these cellular phenotypes are largely cell-autonomous (Kishi & Macklis, 2010). The results from these studies point to a largely cell-autonomous function of MeCP2 in the maturation and maintenance of mature phenotypes in postmitotic neurons.

One of the most exciting experiments in Rett syndrome research to date demonstrated that re-expression of MeCP2 in the adult rescues Rett-like phenotypes (Guy et al., 2007; Luikenhuis et al., 2004). Reactivation of MeCP2 in male mice at three to four weeks of age, just before symptom onset, led to complete prevention of symptoms and a normal lifespan including the ability to breed. In female heterozygous MeCP2 mutant mice, reactivation in adulthood, after symptom onset, corrects deficits seen in the mutant animals. These results, along with those showing minimal cell death in Rett syndrome patients and mouse models (D. Armstrong et al., 1995; R. Z. Chen et al.,

2001), indicate that Rett syndrome is not a neurodegenerative condition but largely results from dysfunction of an intact complement of cells in the brain that may be reinstated to normal function. The ability to rescue phenotypes by reintroducing MeCP2 postnatally strengthens the claim that MeCP2 is involved in maturation and maintenance of neurons. Furthermore, it demonstrates that upon loss of MeCP2, the normal molecular state is intact as reintroduction of MeCP2 allows the system to function normally. These extraordinary findings provide intriguing insight into MeCP2 function and give hope that Rett syndrome pathology may be reversed.

1.7 - Cell-type-specific analyses of MeCP2 mechanism and function

To date, the investigations into the role of MeCP2 in neuronal transcription have largely been carried out in heterogenous tissue. This has proven fruitful for understanding the basics of MeCP2 biology but still faces limitations. The primary binding site for MeCP2, mCH, is highly cell-type specific as discussed above, which indicates that MeCP2 should have prominent features of its biology that are cell-type specific in nature. This theory has been demonstrated in functional and genomic analyses, but important questions remain.

Analyses of morphology and electrophysiology of distinct cell populations in the MeCP2 KO mouse models have found varying contributions to disease phenotypes. Several groups have used conditional deletion of MeCP2 within a cell population to attempt to understand those cells contributions to disease. Glutamatergic neuron MeCP2 deletion limits lifespan and produces mobility and sensory effects. Reinstatement in this excitatory population rescues many disease features (Meng et al., 2016). A broader deletion of *Mecp2* in all GABAergic neurons alone produces most Rett-like phenotypes such as ataxia, seizures, breathing issues, and early death

(Chao et al., 2010). Remarkably, re-expression in inhibitory neurons only was sufficient to rescue disease pathology (Ure et al., 2016). Conditional knockout of MeCP2 from two inhibitory neuron types, PV and SST cells, reproduced many key phenotypes observed in a global KO and led to premature death (Ito-Ishida et al., 2015), potentially indicating a strong contribution of this neuronal population to Rett pathology.

An additional study of the individual contributions of three interneuron types found contributions of each to different aspects of disease, with SST-specific deletion contributing to seizure symptoms and premature death in particular (Mossner et al., 2020). PV neuron-specific deletion of MeCP2 affects critical period plasticity in the mouse primary visual cortex and alters PV cell membrane properties to make them resemble immature PV cells (He et al., 2014a). Restoration of MeCP2 in these PV neurons at four weeks of age successfully normalizes their function (X. Liu et al., 2020). Removal of MeCP2 from hypothalamic neurons affects eating and stress response while its removal in cholinergic neurons affects anxiety and social behavior. Together, these findings inform cell-type-specific contributions to pathology but do not provide mechanistic insights. Further, while selective deletion in particular cells is a powerful approach, it is ultimately limited in scope given the huge number of cell types in the brain and challenging to integrate results from many different studies into a unified understanding of pathology.

Several studies have generated data on transcriptional consequences of MeCP2 dysfunction in distinct neuronal populations. An early study used microarray profiling to assess gene expression changes in PV neurons, Layer 5 neurons in the motor cortex, locus coeruleus noradrenergic neurons, and cerebellar Purkinje cells in male MeCP2-null mice (Sugino et al., 2014). They found

evidence that analysis at the level of refined cell types overcame some of the cellular composition dilution issues from using heterogenous tissue. For example, effect sizes for dysregulated genes increase to as high as several-fold, compared to less than two-fold for studies from heterogenous tissues. Interestingly, this study highlighted different sets of genes as affected in each cell type with very little overlap. This led them to conclude that MeCP2 regulates distinct genes in unrelated cell types. An additional study profiled excitatory and inhibitory cells in the context of MeCP2 disruption and found very little overlap in differential expressed genes between MeCP2 mutant and wild type in the two cell populations (Johnson et al., 2017). However, the novel experimental approach to disrupt MeCP2 used in the study failed to replicate key aspects of well-validated MeCP2 function, making it unclear how generalizable its conclusions are for the broader field. At the level of brain regions, one study found different sets of misregulated genes in the striatum compared to hypothalamus and cerebellum in MeCP2 KO mice (Zhao et al., 2013). Despite these findings of mostly cell-type distinct MeCP2-regulated genes, other studies report set of genes that are repeatedly found in datasets from various brain regions (Ben-Shachar et al., 2009; Clemens et al., 2020; Gabel et al., 2015). As such, it remains unaddressed whether there are distinct or shared effects of MeCP2 dysfunction in neuronal cell types. A single cell study of female heterozygous MeCP2 mutant mice and Rett syndrome patient brains found that features of genes regulated by MeCP2, such as increased length and high mCA, were consistently identified in excitatory cells and vasoactive intestinal peptide (VIP) neurons from both humans and mice (Renthal et al., 2018). Low depth of sequencing and cellular capture issues limited the number and resolution of cell types profiled, as well as the mechanistic conclusions that could be drawn about MeCP2 cell-type-specific regulation.

It is clear from this body of literature that MeCP2 is involved in cell-type-specific transcription and that different neuronal cell types contribute in individual ways to Rett syndrome pathology. However, there is still a gap in our understanding of cell-type-specific transcriptional regulation in the MeCP2 KO. What do the different patterns and levels of mCH in different cell types mean for MeCP2 regulation in those cells? The observation of largely distinct MeCP2-regulated gene sets in different cell types in some studies is interesting but conflicts with reports of consistently misregulated MeCP2 genes in data from brain regions. Evidence from mouse whole cortex points to MeCP2 regulating intragenic enhancers, but enhancers are notably cell-type specific in their activity patterns (Cusanovich et al., 2018; Heinz et al., 2015), so further investigation into this biology at cell-type resolution will be more informative. Functional studies of neurons in the MeCP2 KO indicate that there is a failure in maturation of neurons, but these observations have not been linked to transcriptional effects of MeCP2 disruption in cell types.

1.8 – Scope of this dissertation

My graduate work focused on understanding the mechanism and function of mCA and MeCP2 in regulating transcription in neuronal cell types. In Chapter 2, I share results from a series of experiments analyzing transcriptional dysregulation in diverse cell types in the context of loss of MeCP2. We identify how MeCP2 reads mCA to produce shared and distinct regulation in neuronal cell types and reveal a function of the mCA-MeCP2 pathway in maintaining neuronal subtype transcriptomic identity. In Chapter 3, I explore results from other models of disrupted mCA and MeCP2 with a focus on shared features with MeCP2 KO mouse models and other models of neurodevelopmental disease. In Chapter 4, I discuss our proposed model that mCA-MeCP2

regulation enables maintenance of stable transcriptomic states that define neuronal cell types and describe ongoing experiments to test this hypothesis.

Chapter 2: MeCP2 reads non-CG DNA methylation to stabilize neuronal transcriptomic identity

This chapter is in the form of a manuscript we are preparing for submission.

J. Russell Moore[1], Mati T. Nemera[1], Adam W. Clemens[1], Jared V. Goodman[1], Diana L. Christian[1], Guoyan Zhao[1], Harrison W. Gabel[1*]

[1]Department of Neuroscience, Washington University School of Medicine, St. Louis, MO 63110-1093, USA

[*]Correspondence: gabelh@wustl.edu

J.R.M. performed all experiments in collaboration with A.W.C. for ChIP-seq experiments and J.V.G. for Hi-C experiments, and performed RNAScope analysis. J.R.M and M.T.N performed RNA-seq analysis. M.T.N., A.W.C., and J.R.M. performed ChIP-seq analysis. J.V.G., G.Z., and M.T.N. performed Hi-C analysis. D.L.C. performed Bisulfite-seq experiments and analysis. H.W.G., J.R.M. and M.T.N. designed the experiments and analysis, and wrote the manuscript.

2.1 - Abstract

Mammalian neurons contain high levels of non-CG DNA methylation (mCA) that are read out by MeCP2 to control gene expression. mCA patterns are highly cell-type specific but the functional role of MeCP2 in individual cell types, and whether it impacts similar or distinct gene sets across cell types remains poorly understood. Here, we investigate the effects of loss of MeCP2 in isolated neuronal cell types to probe the function of the mCA-MeCP2 pathway. We find that the global methylation levels within each cell type determine the overall magnitude of transcriptional dysregulation upon MeCP2 perturbation. We define shared and distinct gene regulation by MeCP2 across cell types that are associated with large scale regional mCA patterns and gene specific depletion of mCA respectively, and identify intragenic enhancer elements as important sites by which MeCP2 stabilizes cell-type-specific transcriptional programs. Finally, we find that a major function of mCA and MeCP2 is to regulate genes that define neuronal subtypes at the highest level of cellular resolution. These findings implicate the mCA-MeCP2 pathway in maintenance of functional transcriptomic identities for diverse neuronal subtypes in the complex vertebrate brain.

2.2 - Main

The development and function of the mammalian central nervous system relies on precisely orchestrating expression of thousands of genes in a multitude of cell types. Transcriptional regulatory mechanisms are critical to this process, and neurons employ specialized epigenetic mechanisms to tune gene expression. While all mammalian cells utilize cytosine methylation at CG dinucleotides to control transcription, neurons contain high levels of non-CG methylation that are not present in other cell types, and this unique mark accounts for approximately half the total methylation sites in neurons (Lister et al., 2013). Non-CG methylation primarily occurs at CA

dinucleotides (mCA) and accumulates in the early postnatal period (Lagger et al., 2017; Lister et al., 2013). The methyl-CpG binding protein 2 (MeCP2) has been identified as a critical reader of mCA, working in concert with this mark to repress transcription of long neuronal genes that are embedded in regions of high mCA (Clemens et al., 2020; Gabel et al., 2015; Guo et al., 2014). Evidence suggests that the mCA-MeCP2 pathway is essential for brain function, as loss of mCA through impairment of the de novo methyltransferase DNMT3A has profound functional consequences in both mice and humans causing Tatton-Brown-Rahman syndrome (Tatton-Brown et al. 2014; Okano et al. 1999). Disruption of MeCP2 causes Rett syndrome (RTT) and MeCP2 duplication syndrome (MDS), respectively (Amir et al., 1999; del Gaudio et al., 2006; van Esch et al., 2005). Interestingly, MeCP2 undergoes a dramatic postmitotic increase in neurons concomitant with mCA accumulation and is understood to play a role in postnatal neuronal maturation (Kishi & Macklis, 2004, 2010; Skene et al., 2010). mCA varies in global levels in different brain regions and cell types and shows strikingly cell-type-specific patterning, revealing an important role in regulating cell-type-specific transcription (Mo et al. 2015; Luo et al. 2017; Liu et al. 2021). Recently, we found that genome topology organizes megabase-scale patterns of mCA across the genome, such that some genes are found in high mCA topologically associating domains (TADs), while others can be in lower mCA TADs (Clemens et al., 2020). MeCP2 preferentially represses genes within high mCA TADs by binding to mCA that is enriched within their intragenic enhancers, downregulating enhancer activation and preventing transcriptional initiation. These findings have defined mCA and MeCP2 as critical to normal nervous system function, but major questions about the mechanism and biological output of this pathway remain.

While important initial insights into MeCP2 and mCA have been reported, there is still a lack of understanding of the major functions of this unique neuron transcriptional regulation particularly as it pertains to regulating cell-type-specific gene expression programs. For example, some transcriptomic analyses of brain regions and cell populations in *Mecp2* mutant models have reported recurrent regulation of a set of MeCP2-repressed genes, those upregulated in *Mecp2*-null mutants, with key functions in neurons (Ben-Shachar et al., 2009; Gabel et al., 2015), while other studies have underscored distinct gene expression effects across samples, suggesting unique regulatory targets of mCA and MeCP2 in each cell type (Johnson et al., 2017; Sugino et al., 2014; Zhao et al., 2013). No quantitative analysis of the level of shared and distinct gene regulation across cell types has been performed to date, however, and these apparently discordant results leave it unclear as to whether the mCA-MeCP2 pathway regulates a core set of genes, or if it defines distinct programs in each cell type.

In addition to defining the functional importance of the mCA-MeCP2 pathway across cell types, the patterns of methylation that drive shared or distinct effects have not been evaluated. This is in part because, despite the fact that both MeCP2 and mCA both have been implicated in cell-type-specific transcription, mechanistic analyses to date have been performed largely in heterogenous tissue. Current evidence suggests that mCA patterns in neurons are established by two interacting mechanisms. First, genome topology leads to differences in methylation across large genomic regions defined by topological-associating domains (TADs) of chromatin folding (Clemens et al., 2020). This methylation "set-point" influences mCA levels at all genes and regulatory elements within each domain. Second, a transcription-mediated mechanism affects mCA deposition, whereby genes and regulatory elements that are active in the postnatal period accumulate low

amounts of mCA compared to genes that are not transcribed (Stroud et al., 2017). How these two mechanisms interact to impact gene regulation by MeCP2 at the level of individual cell types and whether they contribute to shared versus distinct gene regulation between cell types is unknown.

At a biological level, the functional significance of the postnatal establishment of the mCA-MeCP2 pathway remains enigmatic. High levels of mCA deposition and the presence of MeCP2 is a hallmark of vertebrates, suggesting that this pathway has a unique role in the development of complex nervous systems (de Mendoza et al., 2021). Furthermore, the accumulation of mCA and MeCP2 to high levels in the brain during a dynamic developmental period of circuit refinement suggests an important role for this pathway in defining mature circuits in the brain, but how the genes most strongly regulated by MeCP2 contribute to this process has not been explored. Thus, it is unclear what the overarching role of this vertebrate-specific regulatory pathway is in the mature brain.

In this study, we address these open questions to uncover a major role for mCA and MeCP2 in maintaining the distinct transcriptomic identity of neuronal subtypes in the adult brain. By interrogating gene regulation by mCA and MeCP2 across multiple isolated neuron populations in cerebral cortex, we demonstrate that the degree of transcriptomic dysregulation in each cell type upon loss of MeCP2 is proportional to the global levels of mCA. We assess the degree of shared gene regulation by MeCP2 in different cell types, finding highly significant overlap between MeCP2-regulated gene sets across cell types, as well as distinct targets within each cell type. Assessing the DNA methylation patterns associated with shared and distinct regulation, we find that large-scale regional mCA patterns, which are relatively invariant across cell types, predispose

genes to repression by MeCP2 and explain the consistent set of MeCP2 targets observed across cell types. In contrast, we observe that cell-type-specific depletion of mCA in gene bodies can result in differential intragenic enhancer methylation and distinct repression of genes between types. Investigating the underlying mechanisms of these effects, we find that repression of enhancers, particularly enhancers that are normally suppressed within a specific cell type, is a major driver of mCA-MeCP2 mediated gene regulation. Finally, we discover that a major function of MeCP2 is to regulate genes that define high-resolution neuronal subtypes. We find that there is considerable reuse of genes to distinguish subtypes and that MeCP2 regulation is critical to maintain this recurrent differential expression, providing an explanation for how MeCP2 can be critical for cell-type-specific transcription, while its target genes can often be shared across cell types. Together, our findings delineate a unifying framework to explain disparate observations across MeCP2 gene regulation studies and implicate the presence of mCA and MeCP2 in vertebrate neurons as a key mechanism to maintain the extreme cellular diversity required for the function of complex neural systems.

2.3 - Results

Global levels of mCA determine the functional impact of MeCP2 within each cell type

While major differences in levels of mCA exist between brain regions and cell types, the quantitative impact these differences have on MeCP2 regulation has not yet been explored. We hypothesized that because cell types differ in their total mCA levels, then, upon loss of MeCP2, brain regions and cell types with higher levels of mCA should display more transcriptional dysregulation than brain regions and cell types with lower levels. To examine if the degree of disruption upon loss of MeCP2 correlates with amount of mCA, we performed analysis of gene

expression changes in MeCP2 knockout (KO) and control mice across the cerebellum, striatum, and hypothalamus, three brain regions previously suggested to vary in amounts of global mCA (Figure 2.1A-C). We compared methylation levels by whole-genome bisulfite sequencing (WGBS) to alterations in gene expression measured across these brain regions that had been collected using the same microarray platform and batch normalization approach. This allowed us to compare effect sizes between regions. As a measure of effect size, we examined the fold-change in gene expression for long genes greater than 100kb that contain the highest levels of genic mCA in that tissue (top 10% highest methylated) which should represent direct targets of MeCP2 regulation. We find that the degree of dysregulation of these genes correlates with levels of mCA in the brain region (Figure 2.1C), suggesting that global mCA is a determinant of the level of dysregulation that occurs in each brain region.

The findings in brain regions suggest that mCA levels determine the scope and degree of gene repression by MeCP2, but analysis at the level of individual cell types is essential to understand the true quantitative impact of this pathway. We therefore carried out analysis of transcriptional regulation by mCA and MeCP2 in discrete cell types in the cerebral cortex using the INTACT nuclear isolation system with cell-type-specific Cre lines. We profiled four neuron cell types in the cerebral cortex with diverse physiologic functions and varying global amounts of mCA (C. Luo et al., 2017; Figure 2.1D). We selected two inhibitory interneuron populations, somatostatin-positive interneurons (SST-Cre) and fast-spiking parvalbumin interneurons (PV-Cre), which have been shown to contribute to Rett pathology in mice and are relatively enriched for mCA (Ito-Ishida et al., 2015; X. Liu et al., 2020). For comparison, we also chose two excitatory neuron populations, Layer 4 neurons (Nr5a1-Cre) which are relatively low in mCA, and Layer 5 neurons (Rbp4-Cre)

which show intermediate levels of mCA (Luo et al 2017; Figure S1A). We crossed these mice to MeCP2 knockout (KO) female mice and then collected MeCP2 KO and WT littermate pairs of each cell type for RNA-sequencing. For each cell type similar numbers of nuclei were isolated between MeCP2 KO and WT mice (Figure 2.1E, S1B), and high expression of distinct marker genes in each cell type validated the specificity of each isolation (Figure 2.1F, S1C). To analyze the correlation between mCA and gene expression effects, we compiled methylomes for each cell type by merging single-nucleus methylomes from the 8-week mouse cortex (C. Luo et al., 2017). Each of these methylomes reflected known patterns of cell-type-specific gene expression, with gene body methylation showing a robust anticorrelation with gene expression (Figure 2.S1D) and pairwise comparisons of cell-type gene-body mCA showing that differential expression in each cell type is inversely correlated with differential enrichment of mCA (Figure 2.S1E). Having obtained gene expression and methylation data for each cell type, we interrogated gene dysregulation and associated DNA methylation patterns in each cell type.

To further test the hypothesis that the magnitude of gene regulation in MeCP2 correlates with mCA level, we compared the effects in low mCA cell types to high mCA cell types. We performed differential expression analysis between MeCP2 KO and WT pairs for each cell type and used remove unwanted variation (RUVg) to normalize across cell types (Risso et al., 2014; Figure 2.1G). Identification of significantly dysregulated genes for each cell type (Figure 2.1H) revealed a correlation between the number of misregulated genes and the levels of mCA in that cell type, suggesting a relationship between methylation level and degree of dysfunction in each population of cells (Figure 2.1I).

Analysis of gene expression changes genome-wide in the MeCP2 KO have shown subtle subthreshold effects on gene expression that are not captured by analysis of significantly changed genes alone (Clemens et al., 2020; Tudor et al., 2002). To further analyze the overall effect size in each cell type in a manner that does not rely on statistical cut-offs, we assessed the fold change of the longest, highest methylated genes in each cell type. We found that the amount of mCA in the cell type correlates with the effect size on these genes (Figure 2.1J). Additionally, we assessed the effect on a set of genes identified as MeCP2-repressed in past gene expression analysis across multiple brain regions (Gabel et al., 2015). These "recurring MeCP2-repressed genes" are long and highly methylated in all cell types examined (Figure 2.S1F) and were found to be consistently upregulated in MeCP2 KO datasets. We find that this population of genes is also more dysregulated in the high mCA cell types compared to the low methylation cell types (Figure 2.1J). Finally, when we look genome-wide, we observe a greater spread in genic fold changes in higher mCA cell types (Figure 2.S1G). Together, our data from brain regions and neuronal cell types indicate that the level of mCA in a population of neurons correlates with the magnitude of dysregulation when MeCP2 is lost. The transcriptomes of cell types with the highest levels of mCA are therefore likely those to be most affected by loss of MeCP2 and may have outsized influence on pathology of Rett syndrome and other disorders involving the mCA-MeCP2 pathway.

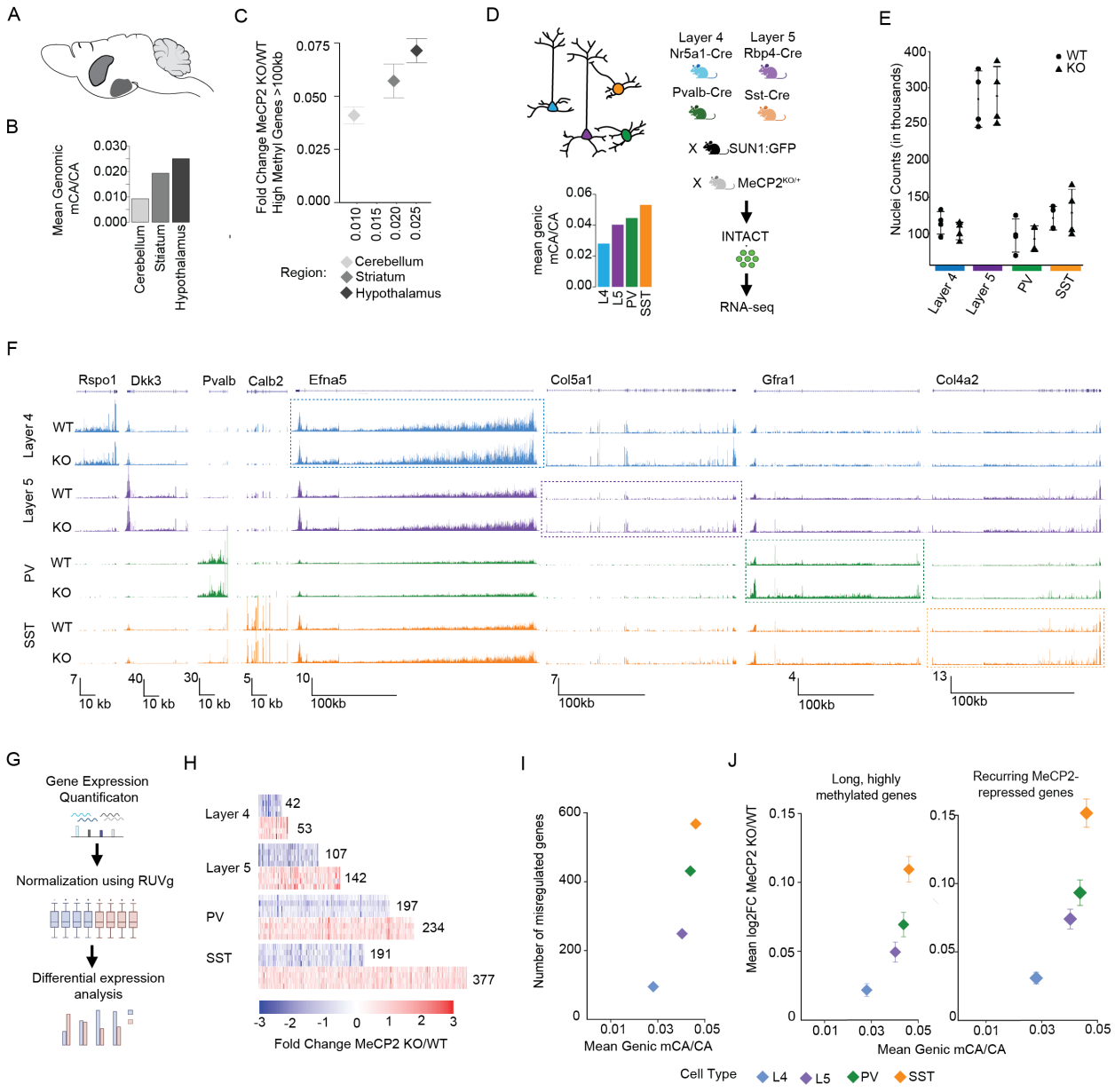


Figure 2.1: Global levels of mCA determine the functional impact of MeCP2 within each cell type

A. Depiction of the three brain regions, the cerebellum, striatum, and hypothalamus selected for analysis.
B. The three brain regions contain varying levels of mCA.
C. The amount of methylation in a brain region correlates with the effect size on long (greater than 100kb), high methylation (top decile of methylation) genes in the MeCP2 KO brain.

D. Four neuronal cell types with varying levels of methylation were selected for gene expression analysis in the context of MeCP2 KO or WT using the INTACT nuclear isolation system.
E. We detect no differences in the number of nuclei isolated by INTACT for KO and WT animals for each cell type profiled.
F. Genome browser view of RNA seq data for MeCP2 WT and KO in each cell type. Marker genes, *Rspo1, Dkk3, Pvalb, Calb2,* for each cell types profiled (Layer 4, Layer 5, PV, and SST, respectively) shown to the left. Example MeCP2-repressed gene for each cell type highlighted with box to the right (*Efna5* – Layer 4, *Col5a1* – Layer 5, *Gfna1* – PV, *Col4a2* – SST).
G. Overview of analysis workflow of gene expression data.
H. Heatmap of per rep fold-changes MeCP2 KO/WT of significantly down-regulated genes (blue) and up-regulated (red) genes identified in each cell type with total number of genes shown.
I. The number of mis-regulated genes in a cell type correlates with the overall methylation level in each cell type.
J. The level of methylation in each cell type correlates with the mean fold change on long (greater than 100kb), highly methylated (top decile of methylation) genes (left) and recurring MeCP2-repressed genes (right).

Regional and gene specific methylation patterning drive shared and distinct MeCP2 cell-type-specific gene regulation

Previous studies examining gene expression changes in discrete neuronal populations upon loss of MeCP2 have emphasized evidence of largely nonoverlapping misregulated gene sets between cell types and brain regions (Sugino et al., 2014; Zhao et al., 2013) suggesting that MeCP2 regulates genes in a cell-type-specific manner. However, other studies have detected shared gene targets of MeCP2 across brain regions, such as recurring MeCP2-repressed genes (Gabel et al., 2015; Ben-Shachar et al., 2009). To address this incongruence, we sought to assess the degree to which shared and distinct MeCP2 gene regulation occur across our profiled cell types and to explore the methylation patterns that could drive these effects. We quantified the amount of overlap between significantly misregulated genes in each cell type and discovered significantly more overlap than predicted by chance (Figure 2.2A-B, 2.S2A). Additionally, we observed high overlap between the MeCP2-repressed gene list in each cell type and the recurring MeCP2-repressed genes previously identified in non-cortical brain regions (Figure 2.2A-B). This finding shows that shared gene regulation is a major feature of MeCP2 repression, even when assessed at the level of individual cell types. In addition to shared effects, we noted uniquely misregulated genes in each cell type, showing that MeCP2 can mediate distinct as well as shared transcriptomic regulation (Figure 2.2A). These findings indicate that across cell types a major subset of genes is biased toward recurrent repression by MeCP2, but that in each cell type distinct gene regulation also occurs.

We next sought to understand how DNA methylation patterns could drive significant recurrent regulation across cell types while allowing distinct gene regulation in each cell type. Tissue-level analysis has shown that MeCP2-repressed genes are embedded in regions of high mCA, which in

turn leads to enhancers of these genes to be enriched for mCA and repressed by MeCP2 (Clemens et al. 2020). To allow analysis of methylation at enhancers associated with each gene, we obtained putative enhancers, or candidate cis-regulatory elements (cCREs) distal to the transcription start site (TSS), defined through single-cell ATAC-seq of the adult mouse brain (Y. E. Li et al., 2021). These putative enhancers were linked to target genes by chromatin co-accessibility with promoters and positively correlated accessibility and RNA expression in joint ATAC-RNA clusters. We then evaluated methylation in and around MeCP2-repressed genes and their linked cCREs. We find that MeCP2-repressed genes in each cell type are embedded in regions of high mCA and display high gene body and cCRE mCA levels (Figure 2.2C, 2.S2B). Analysis of mCG at MeCP2-repressed genes revealed weaker enrichments in mCG than mCA at gene bodies but detected significant enrichment of mCG sites in linked cCREs. Thus, site-specific patterning of CG sites, rather than regional effects, appear to contribute to regulation by MeCP2 (Figure 2.S2C-D). In contrast to MeCP2-repressed genes, MeCP2-activated genes tend to be depleted for methylation at their linked cCREs relative to unchanged genes in PV, SST, and L5 cells (Figure 2.2C, 2.S2B-D).

The impact of regional methylation on cell-type-specific gene regulation by MeCP2 has not been assessed. We focused on MeCP2-repressed genes because they have characteristics suggesting they are direct targets of a well-studied repressive mechanism mediated by the MeCP2-NCoR complex (Kokura et al., 2001; Lyst et al., 2013). We assessed the distribution of MeCP2-repressed genes across the genome, assessing how often they fall in high mCA regions. This analysis revealed a strong propensity for MeCP2-repressed genes to be embedded in high mCA regions, supporting a model in which genes found within regions of high mCA in each cell type are predisposed to repression by MeCP2 (Figure 2.2D). Similarly, analysis of the correlation between

regional and cCRE methylation supports a model in which high regional mCA set-points lead to high mCA at cCREs within those regions (Figure 2.S2E). Tissue-level analysis has suggested that intragenic enhancers are a major site of regulation by MeCP2 (Clemens et al., 2020), with long genes containing many highly methylated intragenic enhancers undergoing repression. We therefore examined the location and methylation of intragenic enhancers in MeCP2-repressed genes in each cell type. We find that MeCP2-repressed genes in each cell type contain significantly more intragenic enhancers and are longer than expression-matched controls (Figure 2.2E, 2.S2F). Furthermore, intragenic enhancers showed the highest degree of mCA enrichment of all elements associated with MeCP2 (Figure 2.S2B). These data support the notion that regional methylation influences the level of mCA at enhancers, and that mCA levels at intragenic enhancers are important for MeCP2 repression in individual cell types.

If high regional mCA predisposes intragenic enhancers and their target genes to MeCP2-mediated repression in each cell type, what might prevent some genes in high-mCA regions from being repressed by MeCP2 in specific cell types? To address this, we examined cell-type-specific methylation at the regional, gene and enhancer levels, and assessed how methylation profiles differ at these sites when differential MeCP2 regulation occurs between cell types. Comparison of genes that are significantly repressed in a given cell type (cell-type MeCP2-repressed genes) to genes that are not significantly repressed by MeCP2 in that cell type but are MeCP2-repressed in other cell types (other-cell-type MeCP2-repressed genes), revealed differential methylation profiles. We found that gene body and cCRE methylation was depleted in other-cell-type MeCP2-repressed genes relative to cell-type MeCP2-repressed genes (Figure 2.2F-G). Notably, the strongest signal for cell-type specificity of methylation occurs at intragenic cCREs linked to MeCP2-repressed

genes (Figure 2.2F), supporting differential mCA at these sites as a major driver of cell-type-specific regulation by MeCP2. In contrast, cell-type MeCP2-repressed genes and other-cell-type MeCP2-repressed genes show similar regional mCA levels (Figure 2.2F-G). This suggests that the regional mCA set-point predisposes genes to repression by MeCP2, but that demethylation of genes and intragenic enhancers relative to the regional methylation set-point in a given cell type can exclude genes found within these high mCA regions from repression.

To further investigate the methylation patterns driving cell-type-specific gene regulation by MeCP2, we examined the methylation of elements associated with recurring MeCP2-repressed genes. Although these genes are enriched for mCA in the region, gene bodies, and intragenic cCREs in each cell type (Figure 2.S2H), a subset of this population is not dysregulated in each cell type (Figure 2.2A). We therefore split these genes between those that are not MeCP2-repressed in a cell type (non-cell-type recurring MeCP2-repressed genes) and those that are (cell-type recurring MeCP2-repressed genes) and evaluated their methylation profiles. We find that recurring MeCP2-repressed genes that escape repression by MeCP2 within a cell type exhibited similar regional mCA to cell-type recurring MeCP2-repressed genes but were depleted for mCA at gene bodies and intragenic linked cCREs (Figure 2.S2I-J). This differential methylation was strongest at intragenic enhancers. These results further support the model of regional mCA predisposing genes to MeCP2-mediated repression, with cell-type-specific de-methylation at intragenic cCREs opting genes out of these repressive effects.

Since regional mCA for MeCP2-repressed genes is invariant across cell types, while gene body mCA can be selectively depleted, we sought to examine how regional and gene body mCA levels

vary genome-wide. Comparison of 1kb bins across genomic regions revealed that intragenic cCREs showed the greatest degree of variation in mCA fold difference between cell types compared to intragenic 1kb regions, while extragenic 1kb regions showed the most limited cell-type variability (Figure 2.2H and 2.S2K). One gene representative of these methylation patterns is *Cacna1i*, which is repressed by MeCP2 in PV cells but not in L5 cells. *Cacna1i* is embedded in a region with high mCA in both PV and L5 cells, but the transcribed region of *Cacna1i* itself is demethylated in L5 cells relative to PV cells (Figure 2.2I).

Together, the methylation profiles associated with MeCP2-repressed genes in each cell type support a model where genes that are repressed by MeCP2 tend to be in regions of high mCA that are invariant across cell types. This regional mCA predisposes these genes to regulation by MeCP2 and drives substantial shared regulation across cell types. Cell-type-specific demethylation at gene bodies and especially intragenic cCREs is driven by high gene expression in early development (Stroud et al., 2017) and can opt genes out of repression, creating unique sets of targets for MeCP2 in each cell type.

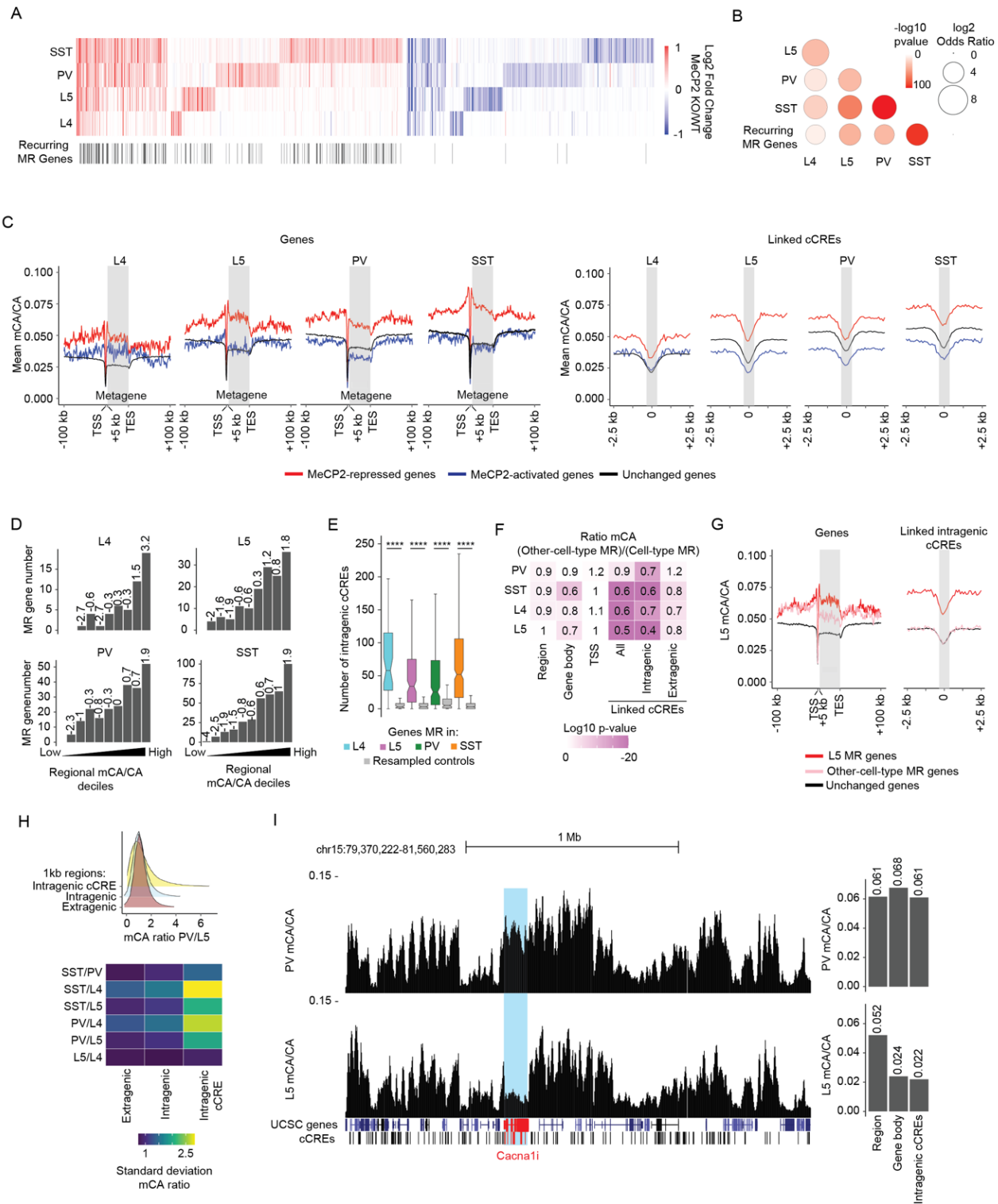

A
SST
PV
L5
L4
Recurring MR Genes
Log2 Fold Change MeCP2 KO/WT
B
-log10 pvalue
log2 Odds Ratio
Recurring MR Genes
C
Genes
Linked cCREs
Mean mCA/CA
Metagene
MeCP2-repressed genes
MeCP2-activated genes
Unchanged genes
D
MR gene number
MR genenumber
Low
High
Regional mCA/CA deciles
E
Number of intragenic cCREs
Genes MR in:
Resampled controls
F
Ratio mCA
(Other-cell-type MR)/(Cell-type MR)
Region
Gene body
TSS
Intragenic
Extragenic
Linked cCREs
Log10 p-value
G
Linked intragenic cCREs
L5 mCA/CA
L5 MR genes
Other-cell-type MR genes
Unchanged genes
H
1kb regions:
Intragenic cCRE
Intragenic
Extragenic
mCA ratio PV/L5
Standard deviation mCA ratio
I
chr15:79,370,222-81,560,283
1 Mb
PV mCA/CA
L5 mCA/CA
UCSC genes
cCREs
Cacna1i

Figure 2.2: Regional and gene specific methylation patterning drive shared and distinct MeCP2 cell-type-specific gene regulation

A. Heatmap of fold changes of shared and unique significantly mis-regulated genes in each cell type. MeCP2-recurring genes marked in black below.
B. Significance of overlap of MeCP2-repressed genes in each cell type and recurring MeCP2-repressed genes from multiple datasets.
C. Left: Aggregate mCA/CA levels for MeCP2-regulated genes in L4, L5, SST and PV neurons. Mean mCA/CA is reported for 1kb bins. "Metagene'' is 50 equally sized bins within gene bodies. Right: Aggregate mCA/CA levels centered at cCREs linked to MeCP2-regulated genes in L4, L5, SST and PV neurons. Mean mCA/CA is reported for 100 bp bins. Gray rectangle = 700 bp ~ median length of all cCREs.
D. Number of MR genes in deciles of regional mCA level.
E. Number of intragenic cCREs inside MR genes identified in L4, L5, SST and PV neurons.
F. Heatmap of mCA/CA enrichment in regions, gene bodies, and linked cCREs of other-cell-type MR genes over those of cell-type MR genes, colored by the log10 Wilcoxon rank-sum p-value. Numbers in the tiles represent the ratio of median mCA/CA of elements associated with cell-type MR genes to the median mCA/CA of elements associated with other-cell-type MR genes.
G. Aggregate mCA/CA levels at gene bodies (left) and linked cCREs (right) of L5 MR genes, other-cell-type MR genes, and unchanged genes.
H. Top: Density plot of mCA/CA ratio between PV and L5 neurons in 1kb extragenic regions, intragenic regions, and regions centered at intragenic cCREs. Bottom: heatmap of standard deviation mCA/CA ratios between pairs of cell types among L4, L5, PV and SST neurons.
I. Left: Genome browser view of PV (top) and L5 (bottom) mCA/CA at *Cacna1i*, a gene repressed by MeCP2 in PV neurons but not in L5 neurons. Right: Regional, gene body, and intragenic linked cCRE mCA levels of *Cacnai1* in PV (top) and L5 (bottom) neurons.

MeCP2 prevents aberrant activation of cell-type-specific enhancers

We next sought to interrogate the mechanisms of transcriptional regulation by mCA and MeCP2 within a defined cell type. For this analysis, we focused on PV interneurons, which have been shown to be functionally disrupted in MeCP2 mutant mice (He et al., 2014; Ito-Ishida et al., 2015) We utilized the INTACT system to assess chromosome topology and chromatin states in these cells, examining how genome structure impacts mCA and how enhancer activity is impacted by these patterns.

We have previously demonstrated that genome topology and regional mCA in the whole cortex are associated (Clemens et al., 2020) and that regional mCA influences cCRE mCA in PV cells (Figure 2.S2B). Therefore, it is possible that genome topology may be associated with regional mCA, and therefore cCRE mCA, in PV neurons specifically. To see if more refined cell types also exhibit this association between genome topology and regional mCA, we performed Hi-C analysis on PV neurons, identifying over 89 million contacts (Figure 2.S3A) and called TADs. We find that mCA levels in PV neurons are more highly correlated for regions within the same TAD than with regions in adjacent TADs. This effect is particularly robust for TADs containing MeCP2-repressed genes (Figure 2.S3B), in concordance with cortical tissue findings (Clemens et al., 2020). This demonstrates that in an isolated cell type, PV neurons, TADs organize mCA levels in genomic regions which has important implications for regulatory elements within the region by establishing their mCA set-point (Figure 2.S2B) and propensity to be regulated by MeCP2.

We next sought to directly interrogate the mechanism of enhancer regulation by MeCP2. The high mCA at enhancers associated with MeCP2-repressed genes supports a model in which MeCP2 binds this methylation to repress enhancer activity. To test this model, we profiled acetylation of histone H3 lysine 27 (H3K72ac), a histone modification associated with enhancer activation (Creyghton et al., 2010). To focus our analysis on the direct effects of MeCP2, we conducted combined analysis of mice in which MeCP2 is either deleted or overexpressed. We performed H3K27ac ChIP-seq on INTACT-isolated PV nuclei from MeCP2 KO and MeCP2 overexpression (OE) mice and their wild-type littermate controls (Figure 2.3A). Comparing H3K27ac ChIP-seq signal in PV nuclei to total cortical nuclei, we observe strong differential H3K27ac signal at genes that exhibit cell-type-specific expression, verifying our isolation of PV neurons (Figure 2.3B-C). To identify cCREs that are regulated by MeCP2, we performed differential analysis of H3K27ac ChIP-seq signal at cCREs in MeCP2 KO and MeCP2 OE PV nuclei using the edgeR (Robinson et al., 2010), identifying 5078 dysregulated cCREs (Figure 2.3D). Consistent with a role for MeCP2 in enhancer repression we found that MeCP2-repressed cCREs, defined as those with increased H3K27ac signal in the MeCP2 KO and decreased H3K27ac signal in the MeCP2 OE, are enriched for mCA, mCG, and MeCP2 binding compared to enhancers that are not significantly affected or affected in the opposite direction (Figure 2.3E and 2.S3C-E). If enhancer misregulation drive changes in gene expression, we would predict that MeCP2-repressed enhancers would be associated with MeCP2-repressed genes. We find that our MeCP2-repressed cCREs are indeed significantly enriched for being located inside and linked to PV MeCP2-repressed genes (Figure 2.3F). Furthermore, we find that analysis of all cCREs linked to MeCP2-repressed genes shows H32K7ac signal is upregulated in the MeCP2 KO and downregulated in the MeCP2 OE at these

sites (Figure 2.3G). These findings support a model in which enhancer regulation by MeCP2 drives cell-type-specific control of gene expression.

We next sought to interrogate the role of enhancer regulation in cell-type-specific gene repression by MeCP2. Our analysis above predicts that, for MeCP2-repressed genes selectively repressed in one cell type but not another, differential methylation between cCREs linked to those genes should drive differential enhancer repression (Figure 2.2F). We tested this prediction by profiling the H3K27ac fold changes between these groups of cCREs. Indeed, we find that cCREs linked to other-cell-type MeCP2-repressed genes have lower H3K27ac fold changes upon MeCP2 perturbation than cCREs linked to PV MeCP2-repressed genes (Figure 2.3H-I). cCREs linked to other-cell-type MeCP2-repressed genes are also depleted for MeCP2 binding relative to cCREs linked to cell-type MeCP2-repressed genes, suggesting that MeCP2 directly drives these differences in cCRE activation (Figure 2.S3F). Additionally, we find that intragenic cCREs linked to recurring MeCP2-repressed genes that are not repressed in PV cells have lower H3K27ac fold changes than cCREs intragenic and linked to recurring MeCP2-repressed genes that are repressed in PV (Figure 2.S3G). Together, these data suggest that MeCP2 regulates genes in a cell-type-specific manner by reading out cell-type-specific enhancer methylation patterns differentially regulating enhancer activation.

Enhancer activity is highly cell-type-specific, and the appropriate activation of these elements is key to cell-type-specific gene expression programs (Carullo & Day, 2019; Heinz et al., 2015). Previous analyses of H3K27ac changes in MeCP2 mutants found acetylation changes at regions outside of defined enhancers (Clemens et al., 2020). However, that study did not have the

resolution to assess enhancers known to be active in specific cell types. Examination of acetylation tracks in our PV-specific data revealed robust acetylation changes in regions with low baseline in wild-type samples that might represent activation of cCREs from other cell types that are not normally active in PV neurons (Figure 2.3J). We therefore assessed acetylation changes at cCREs defined as active in PV neurons and compared them to cCREs not detected in these cells. We separated cCREs based on these criteria and find that both PV and non-PV MeCP2 cCREs are enriched for mCA density and linked to PV MeCP2-repressed genes (Figure 2.S3H-I). However, we found that non-PV cCREs linked to PV MeCP2-repressed genes display greater methylation, acetylation fold changes, and MeCP2 binding than PV cCREs linked to MeCP2-repressed genes (Figure 2.3K and 2.S3I). These data suggest that MeCP2 plays a role in preventing enhancers from activating in a cell type where they are not supposed to be activated.

Why do enhancers normally inactive in a cell type change in activity upon MeCP2 perturbation? Since only a subset highly methylated non-PV cCREs are identified as MeCP2-repressed, we asked whether aberrant activation of enhancers in cell types where they are normally inactive depends on a specific driver of enhancer activity, like transcription factors. Therefore, we performed motif analysis on PV and non-PV MeCP2-repressed cCREs using HOMER (Heinz et al. 2010). We find that motifs from the ROR family are enriched in both PV and non-PV MeCP2-repressed cCREs (Figure 2.S3K). Interestingly, *rora* has been raised as a possible contributor to neuronal cell type diversity including distinguishing subtypes of excitatory projection neurons and PV neurons (Bakken et al., 2021; H. Liu et al., 2021). This raises the possibility that loss of MeCP2 in PV neurons creates a permissive environment that allows ROR transcription factors to bind to and activate enhancers normally off in PV neurons to drive expression patterns of other cell types.

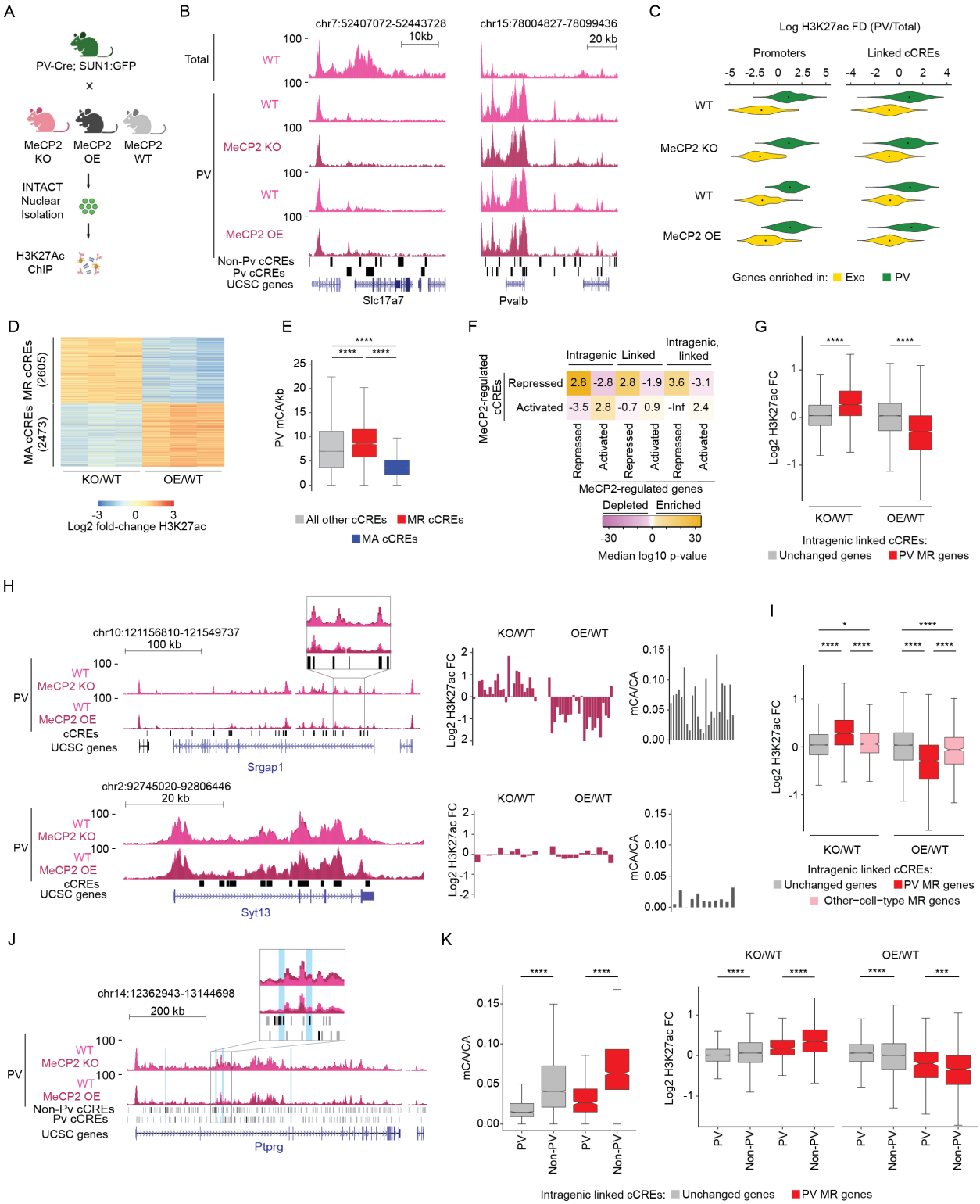

A
PV-Cre; SUN1:GFP
x
MeCP2 KO
MeCP2 OE
MeCP2 WT
INTACT Nuclear Isolation
H3K27Ac ChIP
B
chr7:52407072-52443728
10kb
chr15:78004827-78099436
20 kb
Total
PV
WT
MeCP2 KO
MeCP2 OE
Non-Pv cCREs
Pv cCREs
UCSC genes
Slc17a7
Pvalb
C
Log H3K27ac FD (PV/Total)
Promoters
Linked cCREs
Genes enriched in: Exc PV
D
MR cCREs (2605)
MA cCREs (2473)
KO/WT
OE/WT
Log2 fold-change H3K27ac
E
PV mCA/kb
All other cCREs
MR cCREs
MA cCREs
F
Intragenic
Linked
Intragenic, linked
MeCP2-regulated cCREs
Repressed
Activated
MeCP2-regulated genes
Depleted
Enriched
Median log10 p-value
G
Log2 H3K27ac FC
KO/WT
OE/WT
Intragenic linked cCREs:
Unchanged genes
PV MR genes
H
chr10:121156810-121549737
100 kb
Srgap1
chr2:92745020-92806446
20 kb
Syt13
Log2 H3K27ac FC
mCA/CA
I
Other−cell−type MR genes
J
chr14:12362943-13144698
200 kb
Ptprg
K
PV
Non-PV

Figure 2.3: MeCP2 prevents aberrant activation of cell-type-specific enhancers

A. Schematic of INTACT isolation and H3K27ac ChIP-seq profiling of PV MeCP2 wild-type (WT), MeCP2 knockout (KO), and MeCP2 overexpression (OE) neurons.
B. Genome browser views of H3K27ac ChIP-seq signal in cortical (total) and PV nuclei at Slc7a7, a gene expressed in excitatory neurons, and Pvalb, a marker gene for PV neurons.
C. Log2 H3K27ac fold difference between PV and total nuclei for the same genetic background (MeCP2 WT, MeCP2 KO, or MeCP2 OE) at promoters and linked cCREs of genes enriched PV neurons or in excitatory neurons.
D. H3K27ac ChIP fold changes in PV MeCP2 KO or MeCP2 OE for MeCP2-regulated cCREs identified in Fisher-combined analysis of H3K27ac ChIP-seq signal in MeCP2 KO and MeCP2 OE (FDR ≤ 0.1).
E. Boxplot PV mCA/kb of MeCP2-regulated cCREs.
F. Overlap between MeCP2-regulated cCREs and cCREs inside (intragenic) and/or linked to MeCP2-regulated genes. Median significance (color) and log2 enrichment (number) are shown for cCREs associated with MeCP2-regulated genes compared to cCREs associated with expression-resampled genes.
G. Log2 H3K27ac fold change in cCREs inside and linked to PV MR genes or unchanged genes in PV MeCP2 KO and MeCP2 OE. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001 Wilcoxon test.
H. Top left: Overlaid PV MeCP2 WT, MePC2 KO, and MeCP2 OE H3K27ac ChIP-seq tracks in the PV MR gene Srgap1. The inset zooms in on cCRE-containing regions with robust changes in H3K27ac upon MeCP2 perturbation. Top center: Log2 H3K27ac fold change of cCREs inside and linked to Srgap1 in PV MeCP2 KO and MeCP2 OE. Top right: PV mCA/CA of cCREs inside and linked to Srgap1. Bottom left: Overlaid PV MeCP2 WT, MePC2 KO, and MeCP2 OE H3K27ac ChIP-seq tracks in the other-cell-type MR gene Syt13. Bottom center: Log2 H3K27ac fold change of cCREs inside and linked to Syt13 in PV MeCP2 KO and MeCP2 OE. Bottom right: PV mCA/CA of cCREs inside and linked to Syt13.
I. Log2 H3K27ac fold change of cCREs inside and linked to PV MR genes, other-cell-type MR genes, or unchanged genes in PV MeCP2 KO and MeCP2 OE. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001 Wilcoxon test.
J. Genome browser view of H3K27ac ChIP-seq in Ptprg in PV wild-type, MeCP2 KO, and MeCP2 OE. Gray bars are all cCREs while black bars are cCREs inside and linked to genes. The inset zooms in on a region containing non-PV cCREs with robust changes in H3K27ac ChIP-seq signal in MeCP2 KO and MeCP2 OE relative to wild-type. Blue highlights are non-PV cCREs called as significantly dysregulated upon MeCP2 perturbation.
K. PV mCA/CA (left) and log2 H3K27ac fold change in PV MeCP2 KO and MeCP2 OE (right) of PV and non-PV cCREs inside and linked to unchanged genes or PV MR genes. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001 Wilcoxon test.

mCA and MeCP2 preferentially regulate genes that differentiate neuronal subtypes

Our analyses have shed light on the mechanism by which MeCP2 and mCA regulate cell-type-specific transcriptional programs, but the overarching function of this unique regulatory pathway remains elusive. We therefore examined the functional annotations of MeCP2-regulated genes in each cell type, searching for unifying characteristics. Gene ontology analysis of MeCP2-repressed genes in each cell type identified key functional classes of genes that were affected across cell types, including ion channels, cell-adhesion molecules, and extracellular matrix proteins (Figure 2.4A). Notably, functional categories of regulated genes in each cell type showed substantial overlap, suggesting that genes that precisely modulate the connectivity and physiology of cell types generally are fine-tuned by MeCP2. These genes are important for the unique physiology of each of the neuronal types profiled. For instance, in PV interneurons, we find that sets of ion channels (*Kcnb2*, *Kcng4*, *Kcnma1*, *Scn1b, Grik5*) known to contribute to their specialized firing properties of these neurons are affected (Olah et al., 2022; Otsu et al., 2020; Pelkey et al., 2017). We further noted that genes we detect as MeCP2-regulated in each cell type are functionally important for defining the identities of PV and SST neurons, including higher resolution neuronal subtypes within these types (Paul et al., 2017). In light of our finding above that, without MeCP2, regulatory elements normally dormant in a cell type become activated and drive aberrant gene expression., we considered whether mCA and MeCP2 are particularly important for regulating genes that define neuronal cell type identity.

To investigate if MeCP2 regulates genes that define neuronal cell types, we turned to hierarchically organized single cell RNA-sequencing data from the primary visual and anterior lateral motor cortices. We identified gene sets that distinguish cells at each level of the hierarchy – from genes

that distinguish neurons from non-neuronal cells to those that distinguish neuronal subtypes within a class of neurons – and assessed the overlap of these gene sets with our MeCP2-repressed genes in each cell type (Figure 2.4B). Strikingly, we find that MeCP2 repressed genes are enriched genes that define closely related cell types with the greatest overlap with genes that distinguish high resolution neuronal subtypes (Figure 2.4B). This finding is not dependent on performing gene expression analysis at the level of neuronal cell types as we find a similar pattern of overlap with recurring MeCP2-repressed genes as well as MR genes identified in whole cortex analysis (Figure 2.S4B). To test whether the control of subtype-defining genes is specific to disruption of MeCP2 in the Rett syndrome model, we overlapped the hierarchical neuronal gene sets with misregulated genes from other neurodevelopmental disorder mouse models (Figure 2.S4B). Gene sets from other neurodevelopmental models show overlap with neuronal genes but do not show the same strong association with genes that distinguish neuronal subtypes. Our results demonstrate that genes that distinguish neuronal subtypes are specifically regulated by MeCP2. This raises the intriguing hypothesis that loss of cell-type-specific repression via mCA-MeCP2 leads to a loss of distinction between gene expression patterns in normally discrete subtypes of neurons.

To directly test the hypothesis, we again turned to PV interneurons and conducted RNAScope *in situ* hybridization experiments to analyze expression of MeCP2-repressed genes in these cells. We used this approach to interrogate genes that distinguish three PV interneuron subtypes in layer 4 of the visual cortex of female heterozygous *Mecp2*$^{KO/+}$ mutant mice (Figure 2.4C-D, S4C). As an X-linked gene, female heterozygous *Mecp2*$^{KO/+}$ mutant mice contain a portion of their cells expressing the wildtype *Mecp2* allele and a portion expressing the mutant form, providing a powerful system to quantify cell autonomous effects on gene regulation while controlling for the

efficacy of *in situ* hybridization in wild-type and mutant cells. After calling cells as *Mecp2*-null or wildtype using *Mecp2* expression, we identified PV neurons using *Pvalb* signal (Figure 2.4D). We then in each PV neuron assessed the expression of marker genes (Figure 2.S4D). We find that genes predicted to be mutually exclusive in their expression across PV subtypes show strong segregation in MeCP2 wild-type cells. In contrast, we find that these genes are co-expressed improperly at a much higher rate in MeCP2 KO cells (Figure 2.4E). These results support a model in which MeCP2 functions to maintain transcriptomic identity of PV neuron subtypes.

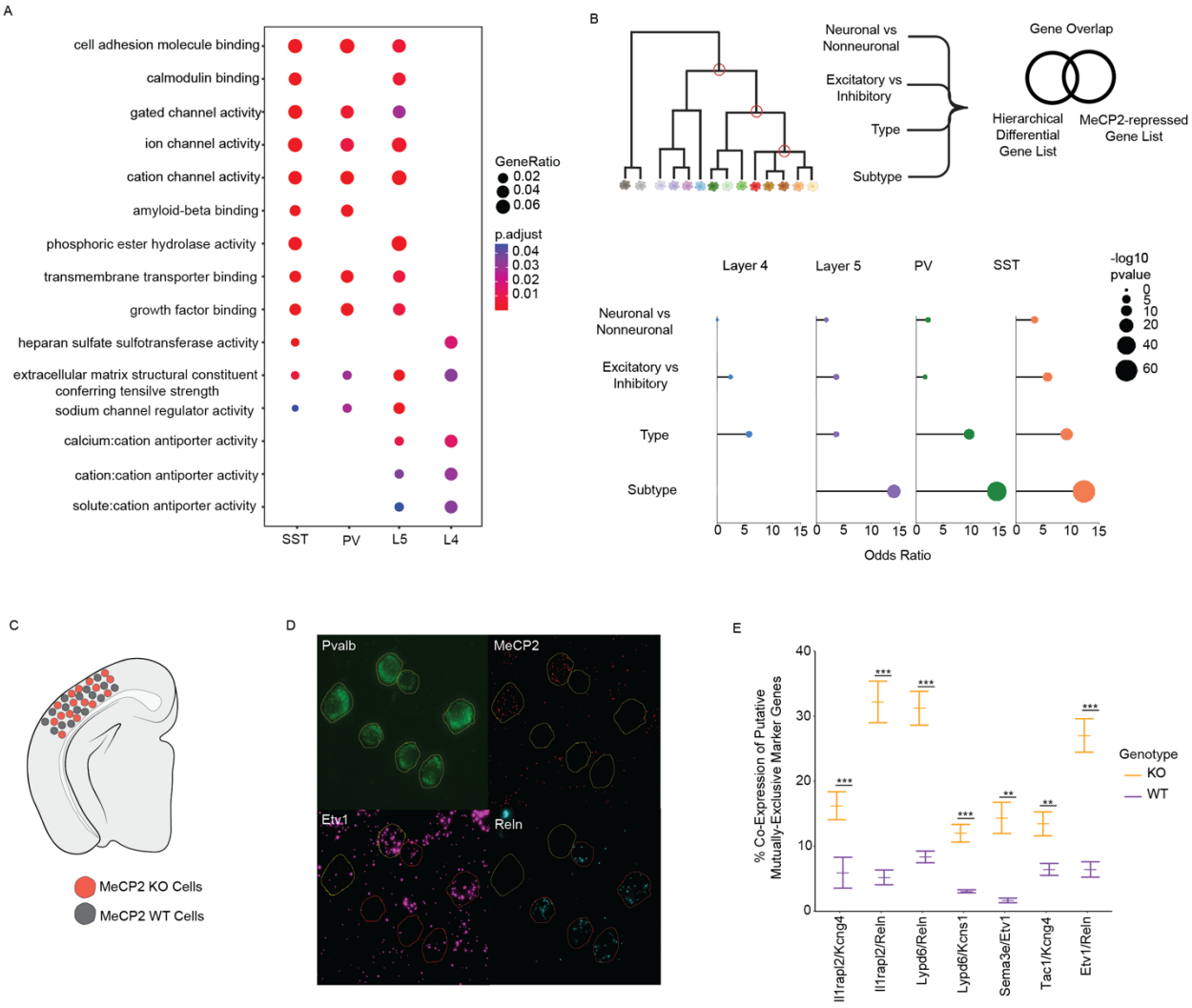


Figure 2.4: mCA and MeCP2 preferentially regulate genes that differentiate neuronal subtypes

A. Gene Ontology analysis of significantly misregulated genes in each cell type shows enrichment for synaptic proteins important for neuronal cell type function. Displayed are top Molecular Function terms that show overlap between cell types.

B. Gene overlap analysis of cell type MeCP2-repressed genes with gene sets distinguishing neuronal cell types at different levels of hierarchically organized single cell transcriptomic data. P-value determined by Fisher's exact test.

C. MeCP2 mutant heterozygous females contain wildtype and mutant cells.

D. RNAscope analysis of expression of mutually exclusive marker genes for visual cortex PV interneurons in MeCP2 KO and WT PV neurons. Identification of PV neurons using *Pvalb* and call of MeCP2 KO and WT cells using *Mecp2*.

E. Plots of rate of co-expression of putatively mutually exclusive PV marker genes in MeCP2 null and WT PV interneurons (n=3, 50-100 cells per experiment, two-side unpaired *t* test, **p<0.01, ***p<0.005). Center lines represent mean and erros bars represent standard error of the mean.

Genes that differentiate neuronal subtypes are repeatedly used and targets of MeCP2 regulation

Our data show that neuronal subtype genes are enriched for MeCP2-repressed genes; however, we have shown that MeCP2 can have graded effects on genes outside of those detected as significantly dysregulation. We therefore considered the possibility that, as a population, subtype genes may be preferentially targeted by MeCP2. We assessed the characteristics of subtype distinguishing genes to see if, as a population, they shared features that predispose them to MeCP2 regulation. Indeed, we find that neuronal subtype distinguishing genes are preferentially located in regions of high methylation, enriched for mCA in their gene bodies, and contain increased numbers of cCREs in their gene bodies (Figure 2.5A). Thus, these neuronal subtype defining genes contain the features of MeCP2-repressed genes making them likely targets of MeCP2 regulation.

While these data support a role for mCA and MeCP2 in defining subtype specific identity, they appear contradictory to the notion that there is shared regulation across cell types. We considered that one explanation for these paradoxical findings could be that genes that define high resolution neuronal subtypes (PV basket cell subtypes) may be repeatedly used in different major neuron types (PV, SST). Indeed, analysis of subtype specific genes revealed this to be the case, with subtype-defining genes being repeatedly used across classes several-fold more than would be expected by chance (Figure 2.5B). This observation suggests that neuronal identity is combinatorial in nature, with different sets of cell adhesion molecules and ion channels configured to endow neuronal cell types with their specific connectivity and function. Notably, we found that the more often a gene is differentially tuned across subtypes the longer the gene tends to be, and

the higher numbers of intragenic regulatory elements the gene contains (Figure 2.S5A). Furthermore, repeatedly-used subtype genes are enriched for having high levels of mCA at their gene bodies and flanking regions (Figure 2.S5B). These characteristics indicate that the more a gene is repeatedly differentially tuned between subtypes the more likely the gene is to fall under robust regulation by mCA and MeCP2. Indeed, we found that genes are repeatedly differentially regulated between subtypes overlap significantly with recurring MeCP2-repressed genes (Figure 2.5C). Together, these findings suggest that a major function of the mCA-MeCP2 pathway is to regulate neuronal subtype distinguishing genes, and that genes that are recurrently tuned across subtypes are strongly predisposed to come under regulation by this pathway (Figure 2.5D). In this way, the mCA-MeCP2 pathway appears to have arisen, in part, to stabilize the transcriptomic identity of diverse neuronal subtypes in the mammalian brain.

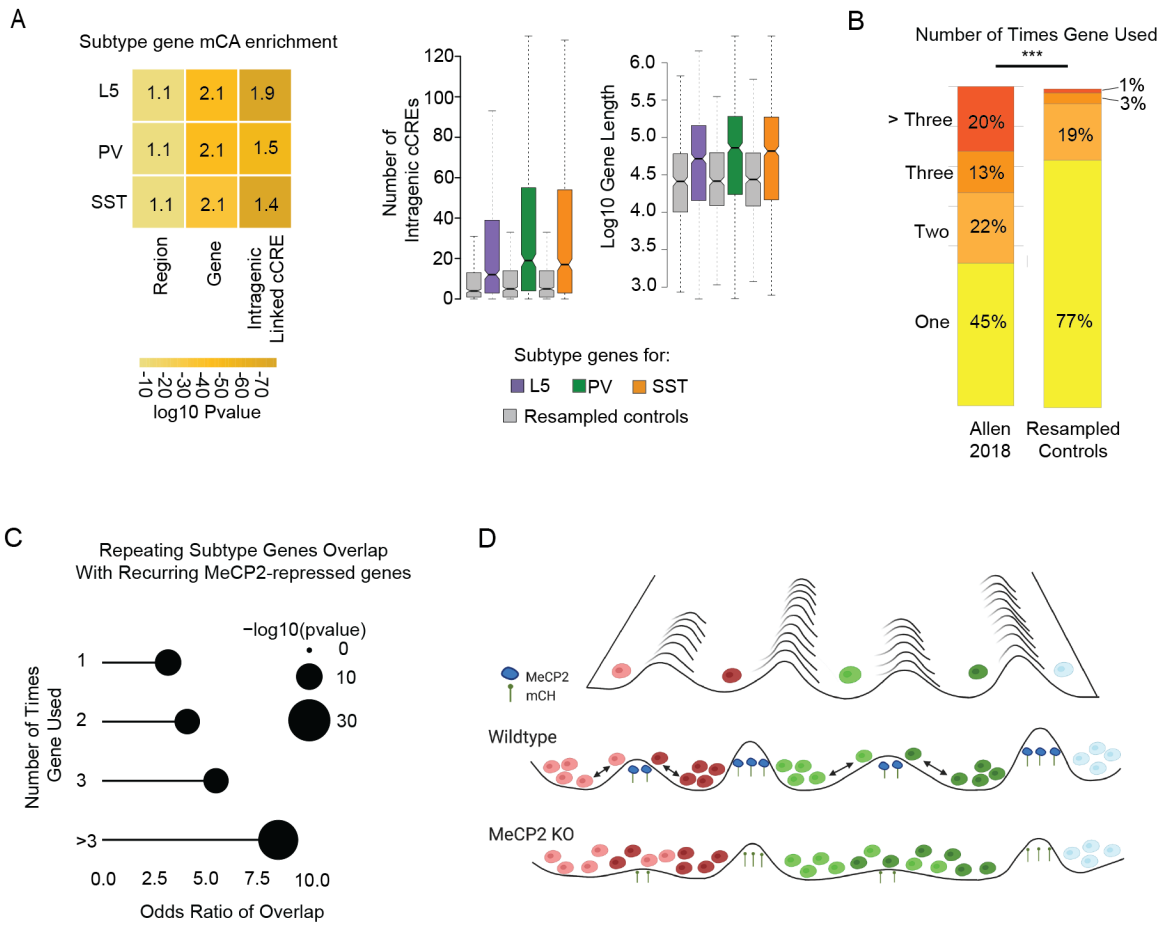


Figure 2.5: Genes that differentiate neuronal subtypes are repeatedly used and targets of MeCP2 regulation

A. Left, mCA enrichment analysis of regions, gene bodies, and intragenic-linked cCREs for genes that distinguish neuronal types within L5, PV, and SST subclasses. Enrichment relative to expression resampled control gene sets is shown numerically, significance is indicated by color. Right, quantification of number of intragenic enhancers and gene length for these neuron-type-specific genes. The center line of each boxplot is the median. Each box encloses the first and third quartile of the data. The whiskers extend to the most extreme values, excluding outliers which are outside 1.5 times the interquartile range. ****$p < 0.0001$ two-sided Wilcoxon rank-sum test

B. Percentage of subtype-distinguishing genes that are repeatedly used by multiple subtypes of PV, SST, and L5 neurons compared to expression-matched resampled controls.

C. Plot of overlap of repeating subtype genes of PV, SST, and L5 with recurring MeCP2-repressed genes. P-values calculated by two-sided Fisher's exact test.

D. Proposed model of MeCP2 and mCA regulation of high resolution neuronal transcriptional states.

2.4 - Discussion

The exciting discovery that neurons utilize a unique form of transcriptional regulation through mCA and MeCP2 has opened important questions of how the pathway mediates gene expression and what it functionally contributes to the brain. In this study, we analyze methylation patterns and MeCP2 regulation in distinct neuronal cell types to provide important insights. Our results show that the cell types with the highest levels of mCA are likely those most affected by loss of MeCP2. We observe distinct gene misregulation upon loss of MeCP2 in different cell types, but also find highly significant overlap between misregulated gene sets across cell types including a set of recurring-MeCP2-regulated genes. Large-scale regional mCA patterns, which are invariant across cell types, predispose genes to repression by MeCP2, and explain the high degree of recurrence for MeCP2 repression observed across cell types. In contrast, distinct repression of genes between cell types is driven by cell-type-specific depletion of mCA in gene bodies which results in differential intragenic enhancer methylation. Investigating the underlying mechanisms of these effects, we find that repression of enhancers is a major driver of mCA-MeCP2 mediated gene regulation and there is strong signal for de-repression of enhancers in incorrect cell types without MeCP2. This altered repression of cell-type-specific enhancers is part of a broader regulation by mCA and MeCP2 of functionally important genes that define the identity of neuronal subtypes. Strikingly, we find that high mCA genes with complex regulatory domains are repeatedly used to distinguish subtypes, and consistently regulated by mCA and MeCP2. These findings present a model in which cell-type invariant regional methylation predisposes neuronal subtype distinguishing genes to be regulated by MeCP2. Then, in each subtype MeCP2 reads mCA at enhancers to maintain neuronal subtype-specific transcriptional programs.

Identifying if there are particular neuronal cell types that could be principal drivers of nervous system dysfunction in MeCP2 disorders has been a major goal of the field. We show that the amount of mCA within cell populations dictates the magnitude of effects when MeCP2 is lost. The PV and SST interneuron types analyzed in our study are among the highest methylated cell types in the brain, and our data demonstrates their strong transcriptomic disruption upon loss of MeCP2. Functional experiments have shown these two populations to be significant contributors to pathologic phenotypes in Rett syndrome mouse models (Chao et al., 2010; Ito-Ishida et al., 2015). As tools for cell-type-specific targeting and interventions grow (Goertsen et al., 2022; Graybuck et al., 2021; Vormstein-Schneider et al., 2020), these analyses can help direct therapeutic approaches for Rett syndrome aimed at particular neuronal classes.

An open question regarding mCA is why do some neuronal types have higher mCA levels than others? Our finding that mCA is read out by MeCP2 to stabilize neuronal subtype transcriptomes may provide a clue as to the significance of these differences. Notably, multiple single cell transcriptomic studies have shown that high PV and SST interneurons contain some of the highest number of distinct subtypes compared to other cell types, while excitatory neurons, particularly L4 neurons, have few to no identifiable subtypes (Tasic et al., 2016, 2018; Yao et al., 2021). It is intriguing to consider that the presence of high levels of mCA within a cell type may facilitate the diversification and stabilization of high-resolution subtypes, and that interneurons in fact can maintain their diverse transcriptomic states because of the robust effects of mCA-MeCP2 pathway in these cells. Future systematic analysis of the number of subtypes below each neuronal type can test this hypothesis.

The question of whether MeCP2 regulates mutually exclusive genes in different cell types or if there is shared regulation between cell types has produced apparently conflicting conclusions for the field. Some studies found very little overlap of MeCP2 regulated genes in disparate cell populations (Johnson et al., 2017; Sugino et al., 2014; Zhao et al., 2013). Meanwhile, others have identified sets of genes consistently misregulated in the context of MeCP2 perturbation across brain regions (Ben-Shachar et al., 2009; Gabel et al., 2015). Here, we discover that while cell-type-specific MeCP2 misregulated genes are a substantial portion of genes affected in any cell type, there is still significant shared effects across cell types and brain tissues. We find that genes recurrently repressed by MeCP2 are located in cell-type invariant high mCA regions predisposing them to MeCP2 repression unless they opt-out in individual cell types through local depletion of their intragenic enhancers.

Our data show that MeCP2 controls enhancers in PV neurons to regulate cell-type-specific gene expression. This fits with recent discoveries in whole tissue that MeCP2 blocks transcriptional initiation, rather than elongation, by inhibiting enhancers. Interestingly, only half of our PV MeCP2-repressed cCREs are active in PV neurons in single cell data. Neuronal genes contain a multitude of enhancers which provide different transcriptional instructions in different cell types and contexts (Cusanovich et al., 2018; Y. E. Li et al., 2021; Nord & West, 2020; Visel et al., 2013). Without MeCP2 regulation, the activation of cCREs in inappropriate cell types could lead to errors in gene expression that have deleterious effects that cell's physiology. How multiple cell-type-specific enhancers function within genes to govern diverse neuronal transcription dynamics, and how this system is regulated by mCA and MeCP2 will be an interesting area to explore going forward.

A striking finding of our analysis is that MeCP2 regulates genes that distinguish subtypes of neurons. This discovery sheds new light on the function of mCA and MeCP2 regulation. Previous studies found that differential postnatal expression of genes between broad cell types leads to differential methylation and readout by MeCP2 (Stroud et al., 2017), but lacked the resolution to account for neuronal subtype diversity which our data indicate are the most susceptible to MeCP2 regulation. Future studies will need to expand on this model through additional modalities and perturbations. Tests in related models, such as those in which mCA is altered by DNMT3A disruption, will be informative as to what methylation alone contributes to this process.

Our proposed model fits with data from functional studies which demonstrate that MeCP2 is important for neuronal maturation and not progenitor cell fate decisions (Kishi & Macklis, 2004, 2010). A growing body of literature indicates that neurons undergo significant transcriptional maturation during the postnatal period, with the full collection of neuronal cell types not developing until neurons settle into their mature positions in the cortex and refine their synaptic inputs (Allaway et al., 2021; Dehorter et al., 2015; di Bella et al., 2021; Mayer et al., 2018). Developing neurons modulate their characteristics based on local cues in the early postnatal brain to configure themselves to function in their resident circuit (Dehorter et al., 2015; Mardinly et al., 2016). After this period of plasticity, they must stabilize their properties in order to maintain functional circuits into adulthood. They do this by fine-tuning their transcriptional programs to control proper expression of genes that encode synaptic proteins critical to their specialized functions. The distinct characteristics and gene expression programs molded by this process defines highly resolved neuronal subtypes. We propose that mCA and MeCP2 accumulate in neurons during the postnatal period to stabilize subtype transcriptomes, providing epigenetic

robustness to this terminal differentiation process. When MeCP2 is lost in Rett syndrome neurons, fluctuations between states ultimately lead to circuit dysfunction. Taken together, our results provide evidence for a role whereby mCA-MeCP2 regulate the formation and maintenance of stable transcriptomic states in high resolution neuronal subtypes. The evolution of mCA and MeCP2 in the nervous system is thus a critical component in the extraordinary diversity and complexity seen in neurons.

2.5 – Materials and Methods

Mice

Pvalb-Cre mice (B6.129P2-Pvalb$^{tm1(cre)Arbr}$/J) and Sst-IRES-Cre mice (Sst$^{tm2.1(cre)Zjh}$/J) were obtained from The Jackson Laboratory. Nr5a1-Cre mice (FVB-Tg(Nr5a1-cre_2Lowl/J) were generously shared by the Allen Brain Institute and Rbp4-Cre mice (Tg(Rbp4-cre)KL100Gsat) were generously provided by Bernardo Sabatini (Harvard University) and Yevgenia Kozorovitskiy (Northwestern University). Each of these cre lines were cross to Sun1:GFP mice (B6;129-Gt(ROSA)26Sor$^{tm5(CAG-Sun1/sfGFP)Nat}$/J) obtained from The Jackson Laboratory. MeCP2 knockout mice (B6.129P2(C)-MeCP2$^{tm1.Bird}$/J) were obtained from The Jackson Laboratory. For PV-Cre, SST-Cre, and Nr5a1-Cre, female heterozygous mice (MeCP2$^{-/+}$) were crossed to Cre:Sun1:GFP mice to generate hemizygous male knockout mice and wild-type male littermates. For Rbp4-Cre we noticed recombination in Rbp4-Cre:Sun1:GFP mice so we crossed Rbp4-Cre to MeCP2:Sun1:GFP heterozygous females.*MeCP2* overexpression mice (FVB-Tg(MECP2)3Hzo/J) were cryo-recovered from The Jackson Laboratory. Female heterozygous mice (*MeCP2*$^{Tg3/+}$) were crossed to Pvalb-Cre:Sun1:GFP mice to generate hemizygous male transgenic mice (*MeCP2*$^{Tg3/y}$) and wild-type male litter mates (*MeCP2*$^{+/y}$).

INTACT

The cortex from 8-week old mice was quickly dissected in ice-cold homogenization buffer (0.25M sucrose, 25mM KCl, 5mM MgCl2, 20mM Tricine-KOH) and flash frozen in liquid nitrogen and stored at -80C. Tissue was thawed on ice in homogenization buffer containing 1mM DTT, 0.15 spermine, 0.5 spermidine, EDTA-free protease inhibitor, and RNasin Plus RNase Inhibitor (Promega N2611) at 60U/mL for RNA experiments. Tissue was minced using razor blades then dounce homogenized in homogenization buffer using 5 strokes with the loose pestle and tight pestle. A 5% IGELPAL-630 solution was added and the homogenate further dounced 10 times with the tight pestle. The homogenized sample was filtered through a 40um strainer and underlaid with a density gradient. The sample was then slowly spun at 8,000g on a swinging bucket rotor and the nuclei collected from the density interface. Nuclei were then isolated using GFP antibody (Fisher G13062) and Protein G Dynabeads (Invitrogen 10003D) with all immunoprecipation steps being performed in a 4C cold room.

Nuclear RNA-seq

RNA from SUN1-purified nuclei was extract using RNeasy micro kit (QIAGEN) following manufacturer instruction and sequencing libraries prepared using the Nugen/Tecan Ovation SoLo RNA-Seq Library Preparation Kit. Libraries for PV samples were sequenced using Illumina HiSeq 3000 (GTAC). All other experimental libraries were NextSeq 500 (Center for Genome Sciences at Washington University).

Chromatin immunoprecipitation protocol

INTACT isolated nuclei were input into chromatin immunoprecipitation experiment following previously described protocol (Clemens et al., 2020; S. Cohen et al., 2011). ChIP was performed for H3K27ac (0.03ug; Abcam ab4729). ChIP libraries 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 to acquire 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-seq signal at cCRE locations (Y. E. Li et al., 2021). The TSS was defined as a 1kb region surrounding the TSS (+/-500bp). Differential ChIP-seq signal across genotypes was determined using edgeR.

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. The nominal p values from edgeR 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.

Hi-C

Hi-C was performed as previously described (Goodman et al., 2020). INTACT nuclei were placed in a 1.01% formaldehyde solution (4.5 mM HEPES-KOH pH 7.9, 9.1 mM NaCl, 0.09 mM EDTA, 0.05 mM EGTA, 0.9X PBS) while rotating for 15 min at room temperature (RT). The formaldehyde was quenched with the addition of Tris and glycine (final 113 mM glycine, 0.91 mM Tris-HCl) while rotating for 5 min at RT. Nuclei were then resuspended in 2.5 mL 0.5% SDS and incubated at 62 °C for 10 min to permeabilize nuclei. 100-250k nuclei from this suspension (25 µL) of this nuclear suspension was then quenched with a Triton-X100 solution (final 1% Triton-X100, 1.2% Cutsmart buffer [NEB]) and incubated at 37 °C for 15 min. Nuclei were then treated with MboI (50U; NEB) and spun at 300 rpm at 37 °C for 4 h, followed by incubation at 65 °C for 20 min to inactivate the enzyme.

DNA blunting was performed by incubating nuclei with Biotin-14-dATP and other dNTPs (final 30 µM) with Klenow (20U; NEB) at 300 rpm at 37 °C for 4 h. Proximity ligation was performed by incubating nuclei in a ligation buffer (final 1X T4 DNA ligase buffer [NEB], 0.1 mg/mL Bovine Serum Albumin [BSA], 1% Triton-X100) with T4 DNA Ligase (4000U) at 300 rpm at 16 °C overnight. Nuclei were then pelleted and resuspended in 1X Cutsmart buffer (NEB). SDS (final 0.8%), NaCl (final 217 mM), and proteinase K (3.2U; NEB) were then added and spun at 1200 rpm at 55 °C for 1 h, then at 1200 rpm at 65 °C for >12 h. RnaseA (0.02 mg; ThermoFisher Scientific) was then added and incubated at 37 °C for 1 h. DNA was purified by phenol-chloroform purification followed by ethanol precipitation in the presence of glycogen. Biotin

was removed from free ends in a dATP solution (100 μM dATP, 1X Buffer 2.1 [NEB]) with 1 U/μg DNA T4 DNA Polymerase (NEB) at 20 °C for 4 h. DNA was then purified using a Monarch PCR & DNA Cleanup Kit (NEB). Purified DNA was sonicated to 300 bp using a Covaris E220 instrument. Sonication tubes were washed with an additional volume of TE to capture DNA stuck to side of tubes. Right-sided size selection was performed using SPRIselect beads. Biotin-labeled DNA was captured using Dynabeads MyONE Streptavidin T1 (ThermoFisher). Beads were then resuspended in 40 μL Low-EDTA TE (Swift Biosciences) and used in the Swift NGS 2 S Plus Library Prep Kit (Swift Biosciences) with minor modifications. For all washes, beads were resuspended in 2X TBW (10 mM Tris-HCl pH 8, 1 mM EDTA, 2 M NaCl, 0.05% Tween-20), incubated for 5 min at RT, then washed twice with 1X TBW. Beads were then resuspended in the appropriate volume of enzyme master mix (Swift Biosciences) for each step. Prior to amplification, DNA was eluted from beads by incubation in Low-EDTA TE at 98 °C for 10 min. DNA was then amplified using 14 cycles of PCR according to kit instructions. Following amplification, cDNA was sequenced on the NextSeq 500 (Center for Genome Sciences at Washington University).

RNA sequencing analysis

RNA sequencing analysis was performed as previously described (Clemens et al., 2020). Briefly, raw FASTQ files were trimmed with Trim Galore and rRNA sequences were filtered out with Bowtie. Remaining reads were aligned to mm9 using STAR (Dobin et al., 2013) with the default parameters. Reads mapping to multiple regions in the genome were then filtered out, and uniquely mapping reads were converted to BED files and separated into intronic and exonic reads. Finally, reads were assigned to genes using bedtools coverage -counts (Quinlan and Hall 2010).

For gene annotation we defined a "flattened" list of longest transcript forms for each gene, generated on Ensgene annotations and 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. Based on these Ensembl gene models, we defined TSS regions and gene bodies. Exonic reads were filtered for non- and lowly-expressed coding genes (minimum of 5 counts across samples) and then DESeq2 performed using adaptive shrinkage. To enable comparisons across cell types we used RUVg to normalize data from each cell type on "in silico" defined negative control genes. These were determined using RUVg recommendations as unaffected genes (bottom 5% in significant change) in KO to WT comparisons shared across all cell types. Significantly MeCP2-repressed genes were those that had a DESeq padj. < 0.1 and log2 fold-change > 0, while MeCP2-activated genes DESeq padj. < 0.1 and log2 fold-change < 0.

Methylation analysis

Pseudo-bulk methylomes for L4, L5, PV and SST cells were obtained by pooling single-cell methylation data from publicly available data (C. Luo et al., 2017). The pseudo-bulk methylomes were then lifted over from mm10 to mm9. The methylation level for an element was assessed by dividing the total number of reads mapping to Cs that supported mC by the total coverage in that region, using bedtools map -o sum. Gene body methylation was calculated using 3kb downstream of the TSS to the TES. Regional methylation for a gene was calculated using the region 10 kb to 210 kb upstream of the TSS and the region 200 kb downstream of its TES, removing the signal from genes overlapping these regions.

Candidate cis-regulatory elements (cCREs)

Candidate cis-regulatory elements (cCREs) and their linkages to genes were identified through chromatin co-accessibility analysis and RNA expression correlation analysis by the BICCN (Li et al., 2021). The coordinates of these cCREs were lifted over from mm10 to mm9 using the UCSC LiftOver tool.

cCREs most robustly regulated by MeCP2 were identified by combining analysis of H3K27ac ChIP-seq signal in cCREs in MeCP2 KO and MeCP2 OE PV nuclei. Nominal p values and fold changes were calculated for the cCREs using edgeR. The p-values were combined using the Fisher method (log-sum) and were Benjamini-Hochberg corrected. cCREs with a combined adjusted p-value ≤ 0.1, log2 fold change > 0 in the MeCP2 OE, and log2 fold change < 0 in the MeCP2 KO were considered MeCP2-repressed cCREs. cCREs with a combined adjusted p-value ≤ 0.1, log2 fold change < 0 in the MeCP2 OE, and log2 fold change > 0 in the MeCP2 KO were considered MeCP2-activated cCREs.

Motif analysis

Transcription factor motif enrichment analysis was performed using HOMER (Heinz et al., 2010) on cCREs using the following parameters: findMotifsGenome.pl input.bed mm10 output -size 200 -len 8.

TAD analysis

The juicer-tools (v.1.19.02) Arrowhead algorithm was used to call topologically associated domains from Knight-Ruiz normalized contact matrices as previously described (Rao et al., 2014).

Cross-correlation matrices were generated as previously described (Clemens et al., 2020). Briefly, each domain is divided into 10 equally sized bins, and 10 equally sized bins are prepended and appended to the domain to make a number of domains x 30 correlation matrix. Each column of this matrix was then correlated against each other to make a 30x30 correlation matrix. TAD-sized regions were shuffled randomly across the genome to generate a negative control. To retain TAD structure, these "shuffled TADs" were separated by similar distances as actual TADs.

TAD/cCRE methylation correlations were calculated by first intersecting TADs with cCREs. For a given TAD-cCRE pair, the TAD methylation signal was calculated by subtracting the methylation signal of the cCRE from the methylation across the entire TAD region.

Controlled resampling

A similar resampling approach was used as previously described (Clemens et al., 2020). Briefly, for every entry in a sample set (e.g., MeCP2-repressed genes), an entry in the control set (e.g., all other genes) with a similar desired characteristic (e.g., expression) was selected, generating a control set of the same size and variable distribution as the sample set.

Tissue preparation for RNAScope

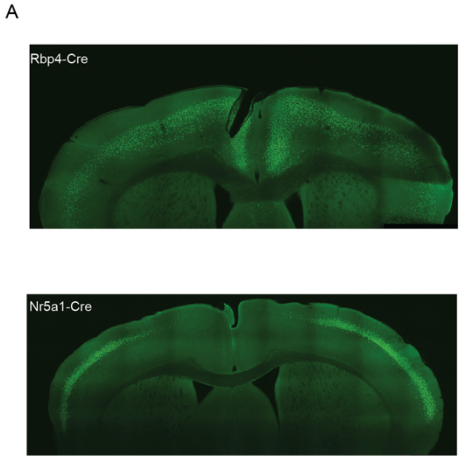
MeCP2$^{KO/+}$ female mice at 8 weeks of age were perfused with ice-cold saline. The brain was removed and placed in cryomold filled with OCT which was flash frozen in a isopentane bath that was pre-cooled in liquid nitrogen. Tissue was stored at −80C for up to 3 months. The night before slicing, the brain was placed in a −20C freezer and allowed to equilibrate to −20C. The day of slicing, it was placed in a pre-cooled cryostat. All chambers and tools were cleaned with RNase-away and 70% ethanol. Slices were cut coronally at 12-14 μm with the mouse visual cortex

collected using Allen Brain Atlas Mouse P56 Coronal as reference. Slides were stored at -80C in a slide box wrapped in plastic and used within two weeks.

RNAscope *in situ* hybridization

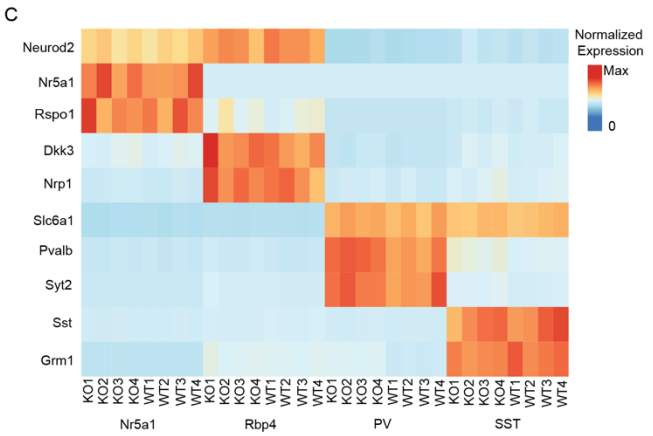
Prior to starting the experiment all tools and surfaces were cleaned with RNase-away. Standard protocol from ACDBio for fresh-frozen tissue was performed. Brain slices were fixed in 4% PFA, dehydrated using increasing ethanol concentrations, and then treated with Protease IV. Probes were warmed to 40C and then mixed and added to the tissue for 2 hours at 40C in an RNAScope hybridization over. Amplification reagents were then applied followed by Round 1 fluorescent probes (T1-T4) and DAPI. Imaging was performed for first round and then slide inserted into 4X SSC for at least an hour until coverslip easily slide off of slide with great care taken to not damage tissue. 10% cleaving solution was then applied to cleave Round 1 fluorophores. After washing, Round 2 fluorophores (T5-T8) were added and the sample imaged. This process was repeated to image the final four probes T9-T12.

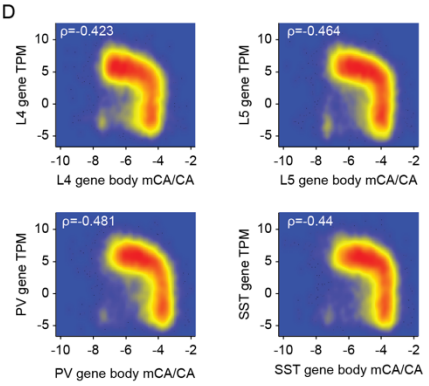
2.6 - Supplemental Information

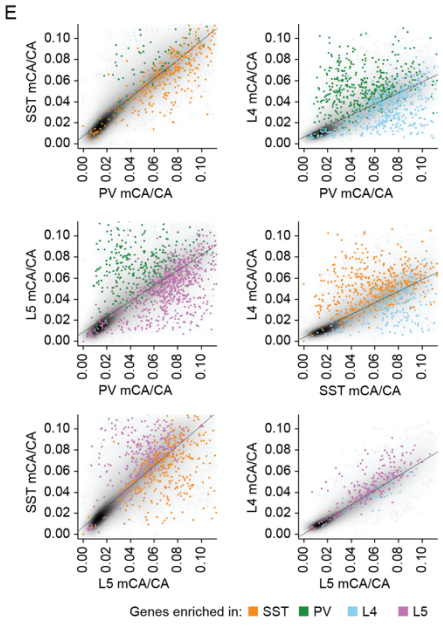


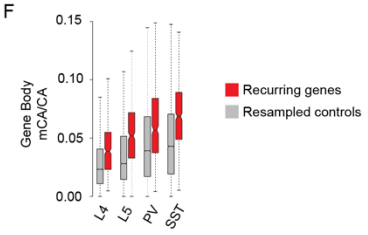
B

Cre Line	Genotype	% GFP Positive Input Nuclei	% GFP Positive IP Nuclei	Average Number of Isolated Nuclei
Nr5a1-Cre (Layer 4)	WT	3.27 (3.11 - 3.48)	99.3 (98,33-100)	103,600 (88,800 - 115,200)
	KO	3.85 (3.51-4.19)	99.6 (98.84-100)	114,950 (95,400-132,900)
Rbp4-Cre (Layer 5)	WT	8.42 (6.84-9.63)	100	289,900 (251,200-336,560)
	KO	9.40 (7.31-11.94)	100	284,495 (246,600-326,100)
PV-Cre	WT	3.44 (3.26-3.62)	99.4 (98.73-100)	93,475 (79,400-108,900)
	KO	3.63 (3.41-3.85)	99.6 (98.46-100)	97,550 (70,200-125,200)
SST-Cre	WT	3.39 (3.03-3.89)	100	128,875 (99,200 - 166,800)
	KO	3.54 (2.99-4.17)	99.8 (99.08-100)	121,325 (107,200-137,700)









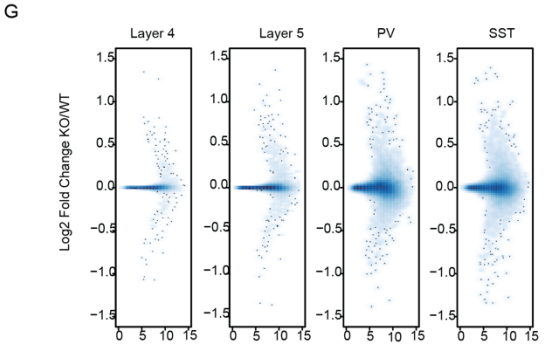


Figure 2.S1: Related to Figure 2.1, Global levels of mCA determine the functional impact of MeCP2 within each cell type

A. Representative images of Rbp4-Cre;SUN1:GFP labeling of Layer 5 excitatory neurons and Nr5a1-Cre;SUN1:GFP labeling of Layer 4 excitatory neurons.
B. Summary statistics of INTACT experiments organized by Cre-line and genotype (no significant differences found between genotypes).
C. Heatmap of normalized counts for marker genes in RNA sequencing data from each cell type profiled.
D. Log2 gene snmc-seq mCA/CA vs log2 gene TPMs for L4, L5, SST, and PV cells.
E. Pairwise comparisons of gene body mCA/CA across cell types. Genes enriched >5 fold in one cell type over another are colored.
F. Gene body mCA level of recurring MeCP2-repressed genes in each cell type compared to expression-matched resampled controls.
G. Smooth scatter plot of all DESeq shrunken log2 fold-changes (MeCP2 KO/WT) of gene expression in L4, L5, SST, and PV cells.

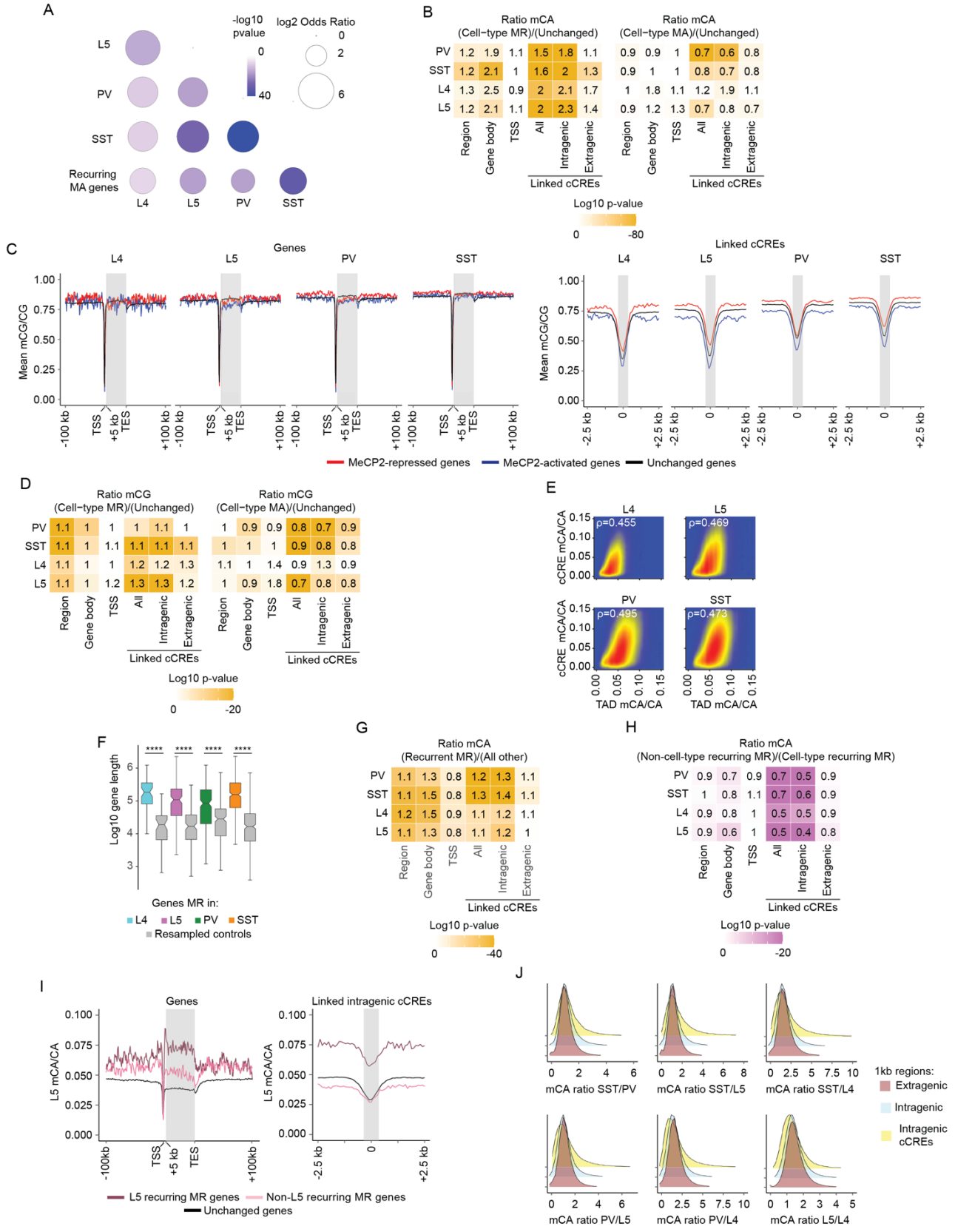
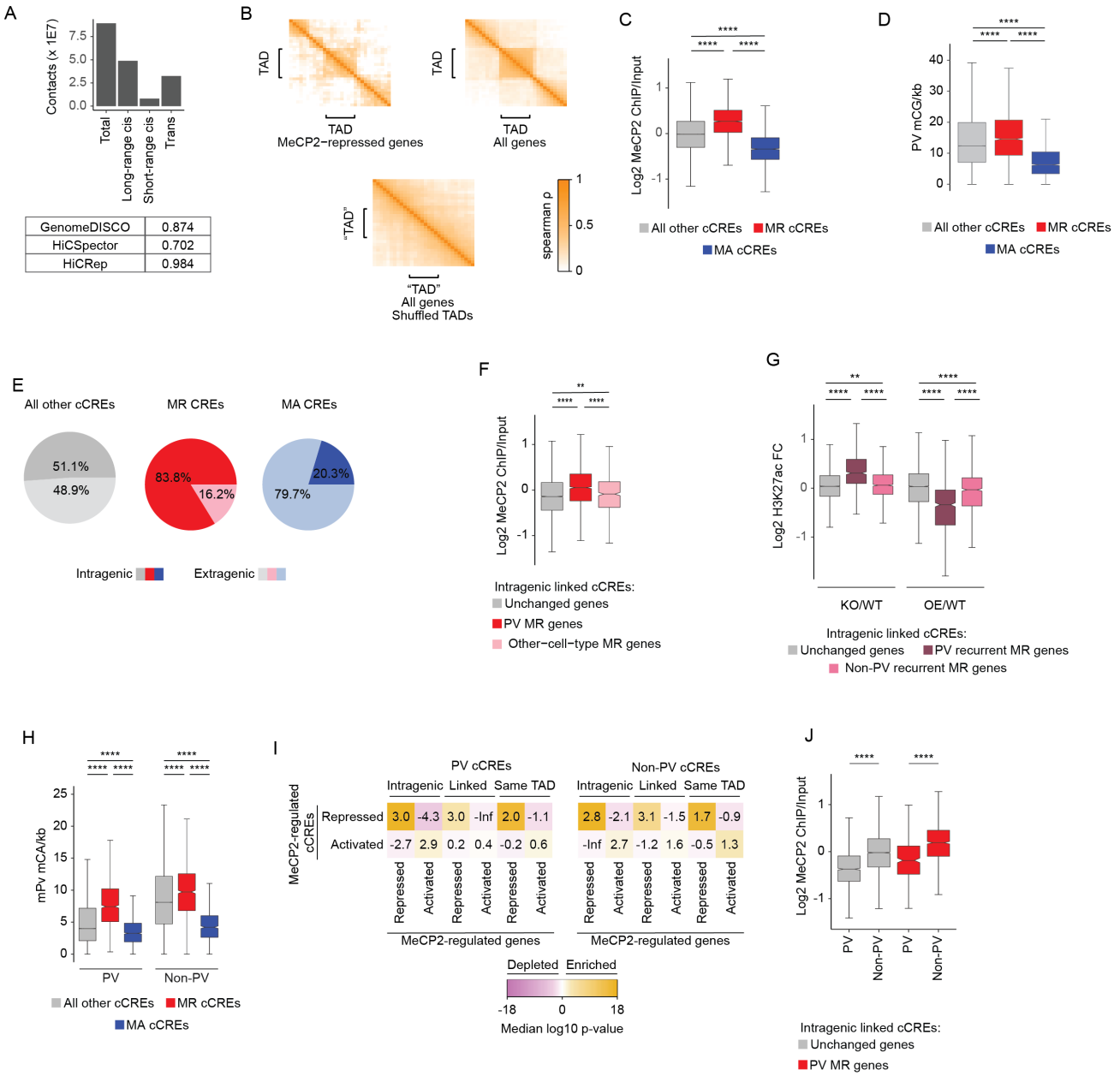

A
-log10 pvalue
log2 Odds Ratio
L5
PV
SST
Recurring MA genes
L4
L5
PV
SST
B
Ratio mCA (Cell−type MR)/(Unchanged)
Ratio mCA (Cell−type MA)/(Unchanged)
Region
Gene body
TSS
All
Intragenic
Extragenic
Linked cCREs
Log10 p-value
C
Genes
Linked cCREs
Mean mCG/CG
MeCP2-repressed genes
MeCP2-activated genes
Unchanged genes
D
Ratio mCG (Cell−type MR)/(Unchanged)
Ratio mCG (Cell−type MA)/(Unchanged)
E
cCRE mCA/CA
TAD mCA/CA
ρ=0.455
ρ=0.469
ρ=0.495
ρ=0.473
F
Log10 gene length
Genes MR in:
Resampled controls
G
Ratio mCA (Recurrent MR)/(All other)
H
Ratio mCA (Non-cell-type recurring MR)/(Cell-type recurring MR)
I
L5 mCA/CA
Linked intragenic cCREs
L5 recurring MR genes
Non-L5 recurring MR genes
Unchanged genes
J
mCA ratio SST/PV
mCA ratio SST/L5
mCA ratio SST/L4
mCA ratio PV/L5
mCA ratio PV/L4
mCA ratio L5/L4
1kb regions:
Extragenic
Intragenic
Intragenic cCREs

Figure 2.S2: Related to Figure 2.2, Regional and gene specific methylation patterning drive shared and distinct MeCP2 cell-type-specific gene regulation

A. Significance of overlap of MeCP2-activated genes from each cell type and recurring MeCP2-activated genes from multiple datasets.
B. Heatmap of mCA/CA enrichment in regions, gene bodies, and linked cCREs of cell-type MR genes or cell-type MA genes over those of unchanged genes, colored by the log10 Wilcoxon rank-sum p-value. Numbers in the tiles represent the ratio of median methylation level of elements associated with cell-type MR or MA genes to the median methylation level of elements associated with unchanged genes.
C. Left: Aggregate mCG/CG levels for MeCP2-regulated genes in L4, L5, SST and PV neurons. Mean mCG/CG is reported for 1kb bins. "Metagene'' is 50 equally sized bins within gene bodies. Right: Aggregate mCG levels centered at cCREs linked to MeCP2-regulated genes in L4, L5, SST and PV neurons. Mean mCG/CG is reported for 100 bp bins. Gray rectangle = 700 bp ~ median length of all cCREs.
D. Comparison of TAD mCA/CA levels and cCRE mCA/CA levels in each TAD in each cell type.
E. Heatmap of mCG/CG enrichment in regions, gene bodies, and linked cCREs of cell-type MR genes or cell-type MA genes over those of unchanged genes, colored by the log10 Wilcoxon rank-sum p-value. Numbers in the tiles represent the ratio of median methylation level of elements associated with cell-type MR or MA genes to the median methylation level of elements associated with unchanged genes.
F. Log 10 gene length of genes MeCP2-repressed in L4, L5, PV and SST genes. The gray box next to each cell type represents the expression-matched genes resampled from that cell type's list of unchanged genes.
G. Heatmap of mCA/CA enrichment in regions, gene bodies, and linked cCREs recurrent MR genes over those of all other genes, colored by the log10 Wilcoxon rank-sum p-value. Numbers in the tiles represent the ratio of median mCA/CA of elements associated with recurrent MR genes to the median mCA/CA of elements associated with all other genes.
H. Heatmap of mCA/CA enrichment in regions, gene bodies, and linked cCREs of non-cell-type recurring MR genes over those of cell-type recurring MR genes, colored by the log10 Wilcoxon rank-sum p-value. Numbers in the tiles represent the ratio of median mCA/CA of elements associated with cell-type MR genes to the median mCA/CA of elements associated with other-cell-type MR genes.
I. Aggregate mCA/CA levels at gene bodies (left) and linked cCREs (right) of L5 recurring MR genes, non-L5 recurring MR genes, and unchanged genes.
J. Density plots of pairwise mCA/CA ratios between cell types in 1kb extragenic regions, intragenic regions, and regions centered at intragenic cCREs.



GenomeDISCO	0.874
HiCSpector	0.702
HiCRep	0.984

K

PV MR cCREs

Motif Name	Consensus	P-value	% of Target Sequences with Motif	% of Background Sequences with Motif
RORa(NR)/Liver-Rora-ChIP-Seq(GSE101115)/Homer	AAWCTAGGTCARDNN	1.00E-30	5.78%	1.91%
RORgt(NR)/EL4-RORgt.Flag-ChIP-Seq(GSE56019)/Homer	AAYTAGGTCA	1.00E-25	5.16%	1.79%
RORg(NR)/Liver-Rorc-ChIP-Seq(GSE101115)/Homer	WAABTAGGTCAV	1.00E-21	4.22%	1.41%
AR-halfsite(NR)/LNCaP-AR-ChIP-Seq(GSE27824)/Homer	CCAGGAACAG	1.00E-04	50.00%	46.23%
ARE(NR)/LNCAP-AR-ChIP-Seq(GSE27824)/Homer	RGRACASNSTGTYCYB	1.00E-04	4.61%	3.19%

Non-PV MR cCREs

Motif Name	Consensus	P-value	% of Target Sequences with Motif	% of Background Sequences with Motif
RORa(NR)/Liver-Rora-ChIP-Seq(GSE101115)/Homer	AAWCTAGGTCARDNN	1.00E-05	2.57%	1.40%
E2F7(E2F)/Hela-E2F7-ChIP-Seq(GSE32673)/Homer	VDTTTCCCGCCA	1.00E-04	0.83%	0.29%
RORg(NR)/Liver-Rorc-ChIP-Seq(GSE101115)/Homer	WAABTAGGTCAV	1.00E-04	1.89%	0.99%
RORgt(NR)/EL4-RORgt.Flag-ChIP-Seq(GSE56019)/Homer	AAYTAGGTCA	1.00E-04	2.19%	1.25%
Tgif1(Homeobox)/mES-Tgif1-ChIP-Seq(GSE55404)/Homer	YTGWCADY	1.00E-03	41.06%	37.69%

Figure 2.S3: MeCP2 prevents aberrant activation of cell-type-specific enhancers

A. Top: Distribution of pooled Hi-C contacts in PV cells. Bottom: reproducibility scores of PV HiC replicates given by GenomeDISCO, HiCSpector, and HiCRep.
B. PV mCA/CA cross-correlation for regions in and around TADs containing MeCP2-repressed genes, all genes, and shuffled control TADs.
C. Log2 input-normalized MeCP2 ChIP signal at MeCP2-regulated cCREs in PV cells.
D. Boxplot of PV mCG/kb in MeCP2-regulated cCREs.
E. Genic distributions of MeCP2-regulated cCREs.
F. Log2 input-normalized MeCP2 ChIP-seq signal in cCREs inside and linked to PV MR genes, other-cell-type MR genes, or unchanged genes.
G. Log2 H3K27ac ChIP fold change (MeCP2 mutant/wild-type) in cCREs inside and linked to PV recurrent MR, non-PV recurrent MR, or unchanged genes.
H. PV mCA/kb of MeCP2-regulated PV cCREs and non-PV cCREs.
I. Association of PV cCREs (left) and non-PV cCREs (right) with MeCP2-regulated genes in PV. The associations are cCREs inside genes (Intragenic), cCREs linked to genes (Cicero), and cCREs in the same TAD as genes (Same TAD).
J. Log2 input-normalized MeCP2 ChIP-seq signal in PV cCREs and non-PV cCREs inside and linked to PV MR genes or unchanged genes.
K. Top enriched motifs from Homer in PV and non-PV MeCP2-repressed cCREs.

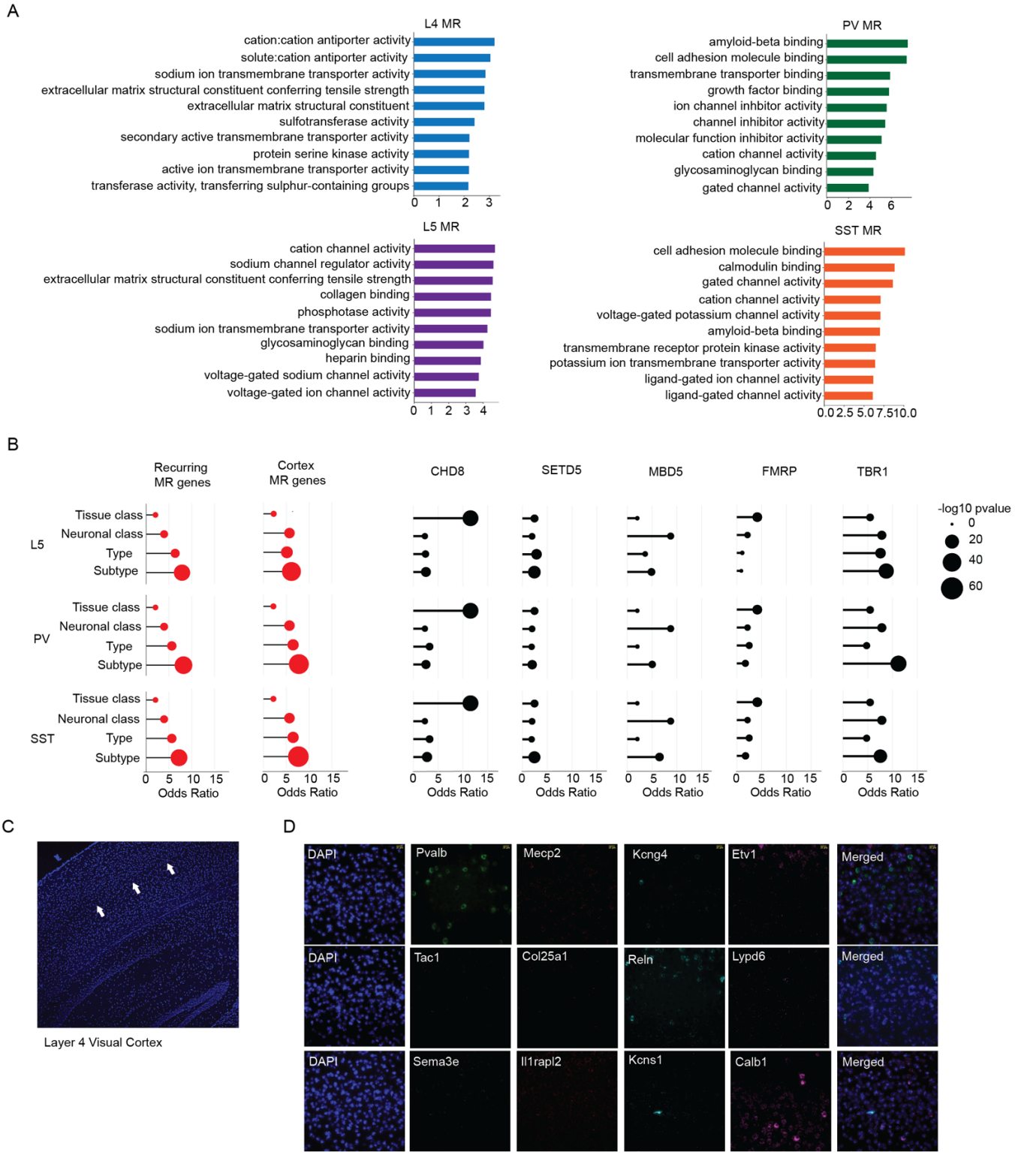

A
L4 MR
cation:cation antiporter activity
solute:cation antiporter activity
sodium ion transmembrane transporter activity
extracellular matrix structural constituent conferring tensile strength
extracellular matrix structural constituent
sulfotransferase activity
secondary active transmembrane transporter activity
protein serine kinase activity
active ion transmembrane transporter activity
transferase activity, transferring sulphur-containing groups
0 1 2 3
PV MR
amyloid-beta binding
cell adhesion molecule binding
transmembrane transporter binding
growth factor binding
ion channel inhbitor activity
channel inhibitor activity
molecular function inhibitor activity
cation channel activity
glycosaminoglycan binding
gated channel activity
0 2 4 6
L5 MR
cation channel activity
sodium channel regulator activity
extracellular matrix structural constituent conferring tensile strength
collagen binding
phosphotase activity
sodium ion transmembrane transporter activity
glycosaminoglycan binding
heparin binding
voltage-gated sodium channel activity
voltage-gated ion channel activity
0 1 2 3 4
SST MR
cell adhesion molecule binding
calmodulin binding
gated channel activity
cation channel activity
voltage-gated potassium channel activity
amyloid-beta binding
transmembrane receptor protein kinase activity
potassium ion transmembrane transporter activity
ligand-gated ion channel activity
ligand-gated channel activity
0.0 2.5 5.0 7.5 10.0
B
Recurring MR genes
Cortex MR genes
CHD8
SETD5
MBD5
FMRP
TBR1
-log10 pvalue
0
20
40
60
L5
PV
SST
Tissue class
Neuronal class
Type
Subtype
0 5 10 15
Odds Ratio
C
Layer 4 Visual Cortex
D
DAPI
Pvalb
Mecp2
Kcng4
Etv1
Merged
DAPI
Tac1
Col25a1
Reln
Lypd6
Merged
DAPI
Sema3e
Il1rapl2
Kcns1
Calb1
Merged

Figure 2.S4: mCA and MeCP2 preferentially regulate genes that differentiate neuronal subtypes

A. Gene ontology of MeCP2-repressed (MR) genes in L4, L5, PV, and SST neurons. Top 10 terms for Molecular Function shown.
B. Overlap of L5, PV, and SST hierarchy datasets with recurring MeCP2-repressed genes, MeCP2-repressed genes from the cortex, and misregulated genes from other NDD mouse models. Fisher's Exact Test of overlap significance for p-value.
C. Example of Visual Cortex area of analysis. White arrows point to layer 4 region where PV neurons were analyzed.
D. Representative images from all 12 target probes and DAPI stain for each of the three imaging rounds. Merged images of each round shown.

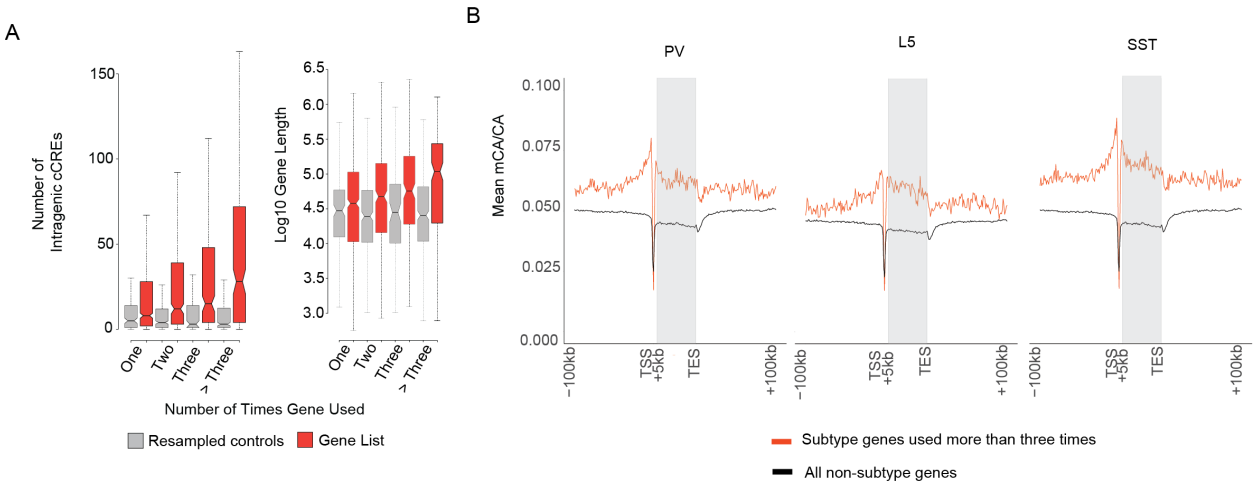


Figure 2.S5: Genes that differentiate neuronal subtypes are repeatedly used and targets of MeCP2 regulation

A. Left: Number of intragenic cCREs in recurring genes. Right: Gene length of recurring genes.
B. Aggregate plots of mean mCA/CA for PV, L5, and SST of subtype genes used more than three times and non-subtype genes.

Chapter 3: Alternative models of mCA-MeCP2 dysfunction reveal overlapping features

This chapter comprises part of a published study that I contributed to (Christian et al., 2020) as well as independent work I completed.

Methylation, RNA, and protein analysis used in Figure 3.1A and 3.1B was performed by Diana Christian and Dennis Wu. RNA-sequencing experiments for DNMT3A$^{KO/+}$ whole cortex conducted by Diana Christian. I collaborated with Diana Christian, Dennis Wu, and Harrison Gabel for the writing of the paper sections.

3.1 - Introduction

MeCP2 has been well recognized for its role in disease with its identification as the principal cause of Rett syndrome resulting in extensive attention from the research community. However, other disruptions of the mCA-MeCP2 pathway lead to pathological conditions. With this in mind, integrated findings across multiple models that test different manipulations of the mCA-MeCP2 strengthen our ability to draw conclusions about shared mechanisms. For instance, data from complete postmitotic deletion of DNMT3A using a Baf53b-Cre mouse revealed overlap with MeCP2 that was dependent on mCA, not mCG (Clemens et al., 2020). Our lab works to elucidate the primary mechanisms of this pathway and then to apply to those insights into understanding of disease. This next chapter will discuss work that I conducted in mouse models with disrupted DNTM3A and overexpression of MeCP2.

3.2 - DNMT3A haploinsufficiency results in epigenomic dysregulation shared across neurodevelopmental disorders

The de novo methyltransferase, DNMT3A, deposits mCA to the neuronal genome during the postnatal period which is then read out by MeCP2. While mutations in MeCP2 have been known to contribute to neurodevelopmental disorders such as Rett syndrome, alterations in its substrate, mCA, through mutations in its depositor DNMT3A were not identified as causative of disease until more recently. Notably, human exome sequencing studies have recently identified de novo mutations in DNMT3A in individuals with autism spectrum disorder (ASD) (Feliciano et al., 2019; Sanders et al., 2015; Satterstrom et al., 2019). Separate studies have also defined heterozygous disruption of DNMT3A as the underlying cause of Tatton-Brown-Rahman syndrome (TBRS), a

heterogeneous NDD characterized by intellectual disability, overgrowth, craniofacial abnormalities, anxiety, and high penetrance of ASD (Tatton-Brown et al., 2014, 2018).

Our lab sought to understand the effects of heterozygous inactivation of DNMT3A in a mouse model carrying a constitutive heterozygous deletion of exon 19 of *Dnmt3a* (Kaneda et al., 2004). We find that this mutation leads to 50% reduction of RNA and protein expression, allowing us to study the *in vivo* effects of heterozygous null mutation of DNMT3A (referred to as DNMT3A$^{KO/+}$ (Figure 3.1A). We performed whole-genome bisulfite sequencing on cortical tissue from DNMT3A$^{KO/+}$ mice at different timepoints and found mCA levels were reduced across postnatal development (Figure 3.1B). Behavioral and morphological analyses of DNMT3A$^{KO/+}$ demonstrate that they display key features of DNMT3A disorders such as increased long-bone length, increased weight, deficits in pro-social communication, and alterations in repetitive behaviors (Christian et al., 2020).

To understand the transcriptional effects of a 50% reduction in mCA we assessed changes in gene expression in DNMT3A$^{KO/+}$ mice. RNA sequencing (RNA-seq) of the DNMT3A$^{KO/+}$ cerebral cortex identified subtle mRNA that are consistent in magnitude with small effects observed in other heterozygous NDD models (Fazel Darbandi et al., 2018; Gompers et al., 2017; Katayama et al., 2016; Figure 3.1C).

To determine whether transcriptional dysregulation upon disruption of DNMT3A and MeCP2 results in shared effects, we examined the overlap of significantly dysregulated genes across mutant mouse lines. We found that the genes dysregulated in the DNMT3A$^{KO/+}$ overlap extensively

with the gene sets dysregulated in a neuron-specific DNMT3A conditional deletion (DNMT3A *Baf53b*-cKO) and MeCP2 mutants (Clemens et al., 2020; Figure 3.1D). For example, all three models show upregulation of Shroom3, whose mutation is associated with developmental defects in humans (Deshwar et al., 2020) and latrophilin-2, a post-synaptic adhesion molecule with roles in synapse targeting and assembly in multiple brain regions (Anderson et al., 2017; R. S. Zhang et al., 2020). Shared dysregulation of genes like these in DNMT3A and MeCP2 disorders may contribute to synaptic dysfunction and disease pathology.

Although a limited gene set is detected as significantly dysregulated in DNMT3A$^{KO/+}$, we considered whether genome-wide enhancer changes could lead to widespread, subtle dysregulation of gene expression below the significance threshold for detection. Such transcriptional pathology would mirror subthreshold genome-wide effects observed upon loss of neuronal mCA (DNMT3A *Baf53b*-cKO) and in models of Rett syndrome (MeCP2 KO) and ASD (MeCP2 OE) (Clemens et al., 2020; Gabel et al., 2015). Importantly, evidence from MeCP2 mutants suggests that pathology arises from thousands of small changes in gene expression rather than large changes in a few significant genes (Kinde et al., 2016; Lavery & Zoghbi, 2019). Disruption of the neuronal methylome in DNMT3A$^{KO/+}$ could similarly lead to numerous gene expression changes that escape statistical significance but still contribute to pathology. To assess whether sub-significance level changes in DNMT3A$^{KO/+}$ phenocopy the MeCP2 mutant and DNMT3A *Baf53b*-cKO models, we performed generally applicable gene set enrichment (GAGE) analysis (W. Luo et al., 2009). GAGE analysis revealed highly significant, concordant changes in gene expression in DNMT3A$^{KO/+}$ for gene sets dysregulated in DNMT3A *Baf53b*-cKO and MeCP2 mutants (Figure 3.1E-F).

Gene Ontology analysis of gene expression changes detected terms associated with neurodevelopment processes, such as axon guidance and recognition, that occur before and during the period when DNMT3A deposits mCA in the brain (Figure 3.2A). This is consistent with a role of DNMT3A in regulating transcriptional programs as neurons mature. Additionally, we see significantly reduced expression of relevant ASD genes, Shank2 and Shank3, which suggests dysfunction at the synaptic level (Guilmatre et al., 2014). These results suggest that genes critical for development of neural circuits are affected in DNMT3A$^{KO/+}$ mice.

We next explored whether shared gene expression signatures in DNMT3A$^{KO/+}$ mice extend to models of neurodevelopmental disease (NDD) without established mechanistic links to DNMT3A. GAGE analysis detected overlap of DNMT3A$^{KO/+}$ gene dysregulation with neurodevelopmental gene co-expression modules identified in the human brain (Parikshak et al., 2013) including M13, M16, and M17, which increases during cortical development and are enriched for ASD risk genes (Figure 3.2B). Modules involved in gene regulation that are expressed early in development and decrease over time (M2 and M3) are also increased in DNMT3A$^{KO/+}$ (Figure 3.2B). We detected significant alterations in genes identified as dysregulated in CHD8 and PTEN mouse models of overgrowth and ASD (Gompers et al., 2017; Katayama et al., 2016; Tilot et al., 2016) as well as the SETD5 NDD model (Sessa et al., 2019; Figure 3.2 C). These findings suggest that overlapping gene dysregulation could underlie common symptomology in individuals with mutations in distinct epigenomic regulatory genes. Extending GAGE analysis to human ASD data, we observed significant changes of gene sets dysregulated in ASD postmortem brains (Gandal et al., 2018; Voineagu et al., 2011) in the DNMT3A$^{KO/+}$ cortex (Figure 3.2D) and also detected upregulation of

genes linked to ASD from human genetics studies (Banerjee-Basu and Packer 2010; Abrahams et al. 2013; Figure 3.2D). Resampling analysis indicated that significant dysregulation of these mouse and human gene sets in DNMT3A$^{KO/+}$ is not driven by enriched expression in the cortex (Figure 3.2E-F). These findings suggest that the DNMT3A$^{KO/+}$ mouse shares overlapping transcriptional pathology with gene expression changes underlying NDD.

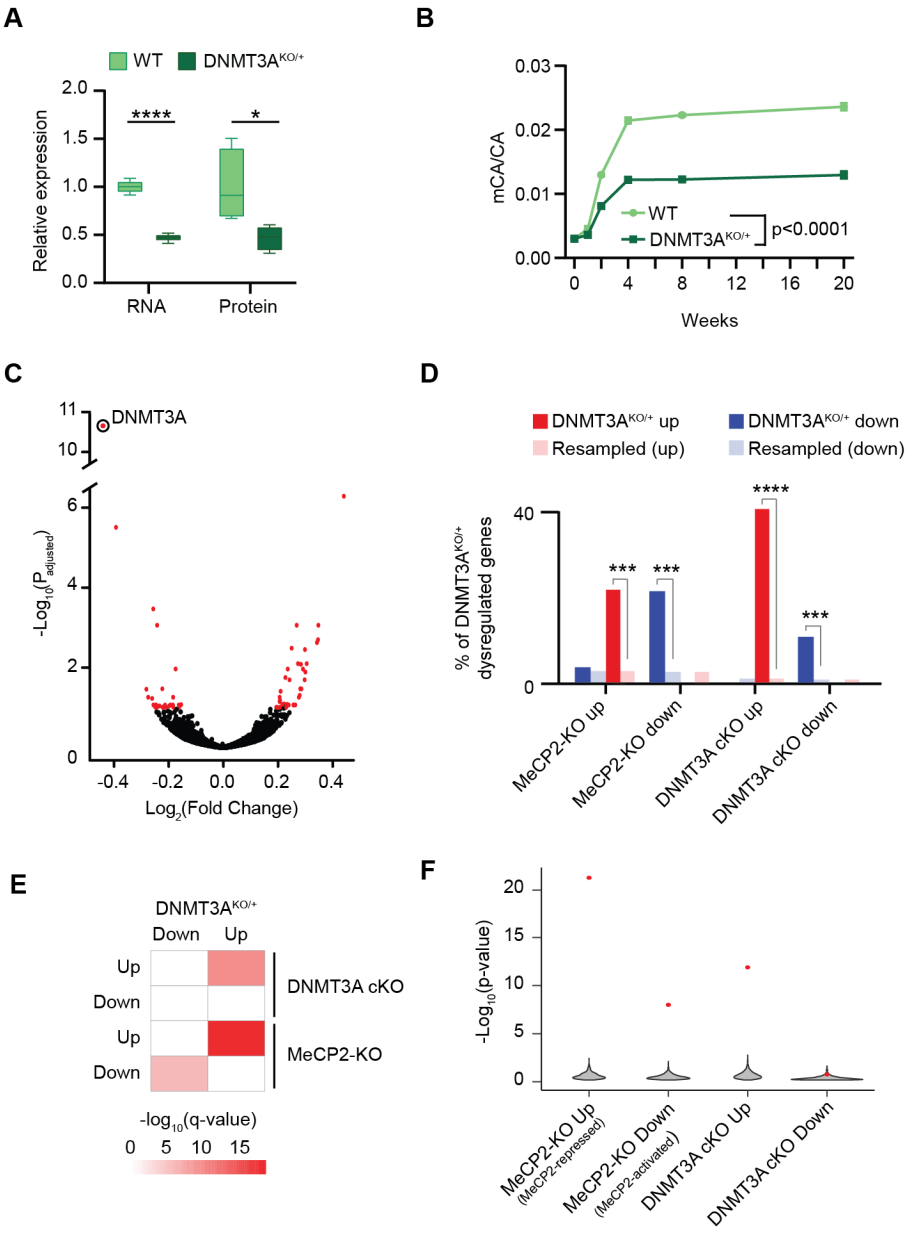


Figure 3.1: Transcriptomic dysregulation in the DNMT3A$^{KO/+}$ cortex overlaps with MeCP2 mutants

(A) Level of DNMT3A RNA and protein expression in the DNMT3A$^{KO/+}$ versus WT.

(B) Developmental time course of global mCA in the cerebral cortex (right), as measured by sparse whole genome bisulfite sequencing (WGBS).

(C) Volcano plot of DESeq $\log_2$ fold changes of the DNMT3A$^{KO/+}$ versus WT. Genes reaching a significance of $p_{adj.}$<0.1 are colored in red.

(D) Overlap of significantly dysregulated genes (padj. < 0.1) in the DNMT3A$^{KO/+}$ and genes dysregulated in DNMT3A Baf53b-cKO or MeCP2 mutants (***, P<e^{-5}; ****, P<e^{-10}; hypergeometric test, observed vs background estimated by resampling, see methods).

(E) Significance of gene set expression changes in the DNMT3A$^{KO/+}$ for GAGE analysis of gene sets dysregulated in DNMT3A Baf53b-cKO or MeCP2 mutants (Clemens et al., 2020). Box plots indicate median and quartiles. Bar graphs are mean and SEM.
(F) Significance of GAGE analysis of gene dysregulation compared to expression matched resample controls. Expression matched resampling of each gene set was performed 1,000 times and analyzed using GAGE for enrichment in DNMT3A$^{KO/+}$ fold-change data (gray violin). This was compared with the true gene set p-value (red point) to test for significance (right).

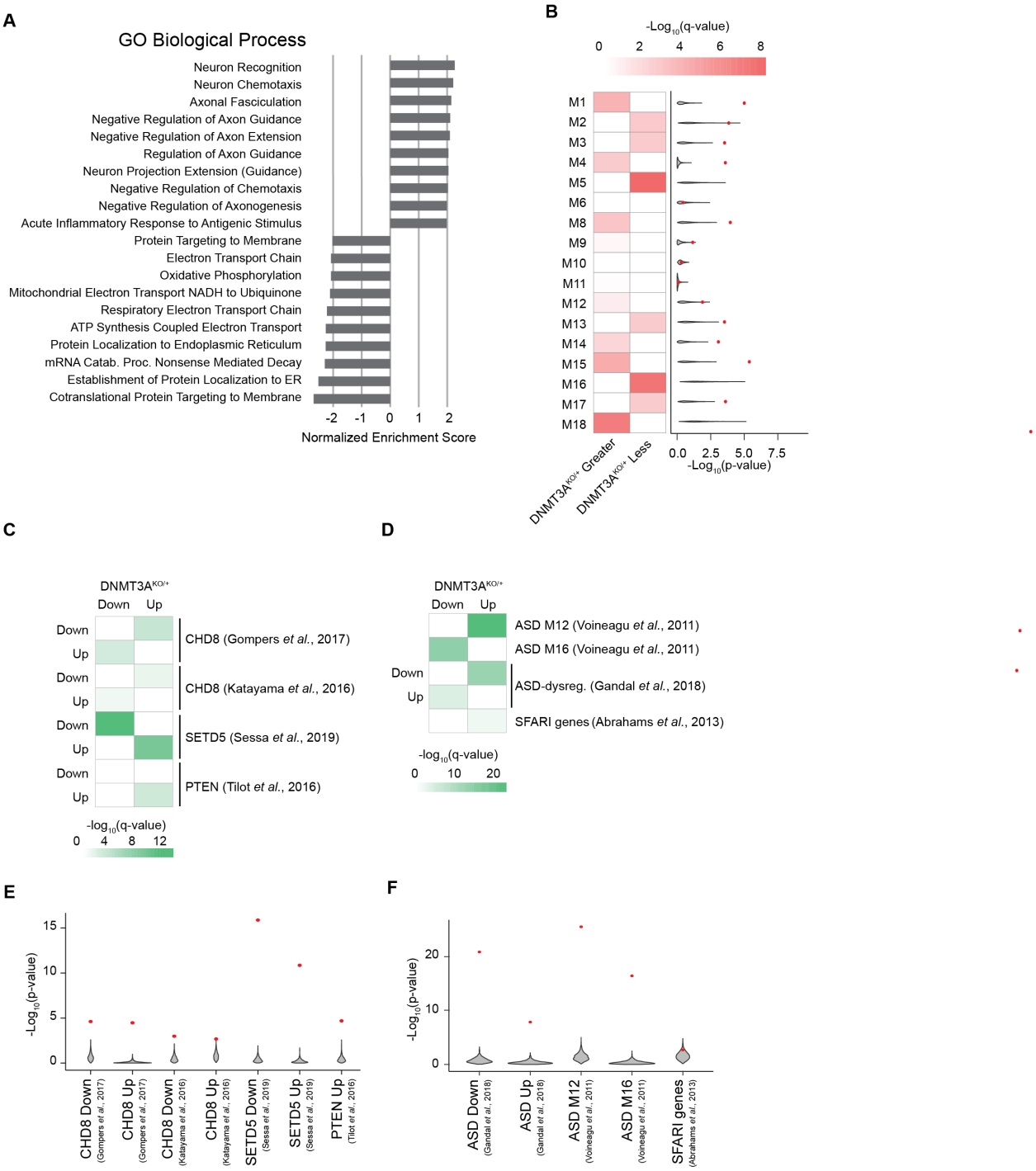


Figure 3.2: Gene dysregulation in the DNMT3A$^{KO/+}$ overlaps with effects in other NDDs.
(A) Top ten up- and down-regulated Gene Ontology terms from Broad GSEA Molecular Signatures Database version 7.0 (Subramanian et al., 2005) All terms are significant at an FDR<0.1.

(B) GAGE analysis of developmental expression modules (Parikshak et al., 2013) Significant modules (q-value<0.1) are colored in red (left). Expression matched resampling of each gene set was performed 1,000 times and analyzed using GAGE for enrichment in DNMT3A$^{KO/+}$ fold-change data (gray violin). This was compared with the true gene set p-value (red point) to test for significance (right). Only the direction of dysregulation in which the gene sets showed the highest significance (i.e., DNMT3A$^{KO/+}$ greater or less) is shown.
(C) GAGE analysis of expression changes in DNMT3A$^{KO/+}$ for dysregulated gene sets in studies of NDD mouse models (Gompers et al., 2017; Katayama et al., 2016; Sessa et al., 2019; Tilot et al., 2016).
(D) GAGE analysis of expression changes in DNMT3A$^{KO/+}$ for gene sets identified in studies of human ASD. ASD module 12 (synaptic) and 16 (immune) were previously identified in weighted-gene coexpression analysis of human ASD brain (Voineagu et al., 2011), and ASD-dysregulated genes were previously identified (Abrahams et al., 2013; Gandal et al., 2018).
(E, F) Expression matched resampling of GAGE analysis for gene sets displayed in C and D.. Only the direction of dysregulation in which the gene set showed significance (i.e. DNMT3A$^{KO/+}$ greater or less) is shown. *, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001.

3.3 - Transcriptional profiling of PV neurons in the DNMT3A$^{KO/+}$ model

The cell-type-specific nature of mCA-MeCP2 led us to analyze dysfunction in the DNMT3A$^{KO/+}$ model within a refined population. We elected to study parvalbumin (PV) interneurons as they are relatively enriched for mCA (Mo et al., 2015). I performed INTACT to isolate nuclei from PV-Cre:Sun1 mice crossed to DNMT3A$^{KO/+}$ mutant mice followed by bisulfite and RNA sequencing. Consistent with the whole cortex PV neurons show a 50% reduction in mCA, suggesting that heterozygous loss of DNMT3A does not disproportionally lead to a larger loss in high mCA cell types (Figure 3.3A). Using differential expression analysis between mutant and wild-type pairs, we find very few dysregulated genes, indicating that the small transcriptional effects we see at the whole cortex level are reproduced in a more homogenous population. Despite the small transcriptional effects in DNMT3A$^{KO/+}$ PV neurons, we hypothesized that genome-wide analysis would reveal similar transcriptomic responses to PV neurons lacking MeCP2. We performed GAGE analysis on global gene expression changes and find strong concordance with MeCP2 KO PV neuron transcriptional alterations (Figure 3.3B). This indicates similar effects at the cell-type-specific level when DNA methylation and its reader, MeCP2, are disrupted.

Our MeCP2 analysis indicates a function for mCA and MeCP2 in regulating transcriptional programs of high-resolution neuronal subtypes (Chapter 2, Figures 2.4-5). We hypothesize that disruption of MeCP2 binding substrate should result in similar effects. To study this in our DNMT3A$^{KO/+}$ model, we performed gene set enrichment analysis (GSEA) of gene sets that distinguish neurons at different hierarchical levels in our PV DNMT3A$^{KO/+}$ differential expression data. While a subtle effect, we detect significant disruption of neuronal subtype distinguishing

gene sets in our data set, indicating that the cellular confusion phenotype we detect in PV neurons of MeCP2 mutant mice is similarly observed in DNMT3A mutant mice (Figure 3.3C).

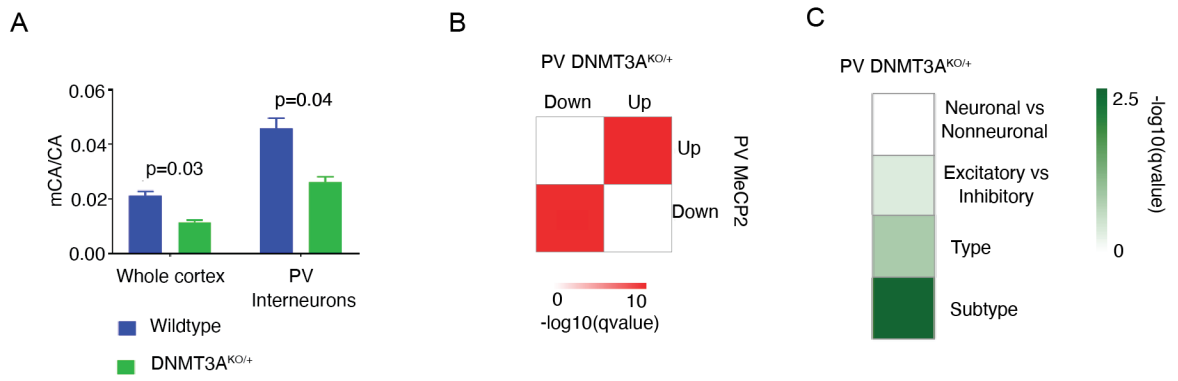


Figure 3.3: PV DNMT3A$^{KO/+}$ neurons share transcriptional effects with PV MeCP2 KO neurons
(A) Global mCA levels in the DNMT3A$^{KO/+}$ is reduced by half compared to WT for whole cortex and PV interneurons.
(B) GAGE analysis of PV MeCP2-repressed and MeCP2-activated gene sets in PV DNMT3A$^{KO/+}$ differential expression fold changes between mutant and WT.
(C) GAGE analysis of single cell PV hierarchy distinguishing genes in PV DNMT3A$^{KO/+}$ RNA sequencing data.

3.4 - MeCP2 overexpression in PV and Layer 4 excitatory neurons produces inverted transcriptomic effects to MeCP2 knockout

While loss of *MECP2* was first characterized as causative of Rett syndrome in females, duplications of *MECP2* were later identified as causing a severe form of intellectual disability, *MECP2* duplication syndrome (MDS), predominantly affecting males (van Esch et al., 2005). Mouse models of MDS were generated through insertion of the human *MECP2* gene leading to 2-8 fold higher expression (Collins et al., 2004). MeCP2 overexpression (OE) mice develop progressive neurobehavioral abnormalities and premature death. Overexpression of MeCP2 has been shown to alter gene expression programs in an opposite manner compared to the MeCP2 KO (Clemens et al., 2020). Measuring the effects of excess MeCP2 allows researchers to corroborate results from MeCP2 KO studies, with the rationale that direct effects of MeCP2 should have opposing valences in MeCP2 KO versus MeCP2 OE mice. The effect of MeCP2 OE on transcription within neuronal cell types has not been explored. With this in mind, we crossed PV-Cre:Sun1 and Nr5a1-Cre:Sun1 mice to MeCP2 OE mice and performed INTACT to isolate nuclei from PV and Layer 4 neurons from mutant and wild-type littermate pairs followed by RNA sequencing. Comparing the gene expression effects genome wide using Rank-Rank Hypergeometric Overlap analysis (RRHO) to our MeCP2 KO data from each cell type shows that genes upregulated in the MeCP2 KO are downregulation in MeCP2 OE in PV and Layer 4 neurons (Figure 3.4A). This shows that within two neuronal populations with different levels of mCA, PV and Layer 4 neurons, MeCP2 OE and MeCP2 KO consistently produce inverse effects.

Next, we wanted to determine if MeCP2 overexpression causes misregulation of neuronal subtype distinguishing genes as we observe in the MeCP2 KO. We performed the same overlap analysis

as described previously (Chapter 2 Figure 2.4) using MeCP2-repressed genes, those that significantly downregulated in the MeCP2 OE compared to wild-type littermates, from PV and Layer 4 neurons. We detect significant overlap with MeCP2 OE downregulated genes and cell-type gene sets, with PV subtype genes showing the most significant overlap. This suggests that in the context of excess MeCP2, these subtype genes are significantly dysregulated, and strengthens the argument that they are direct targets of MeCP2 regulation.

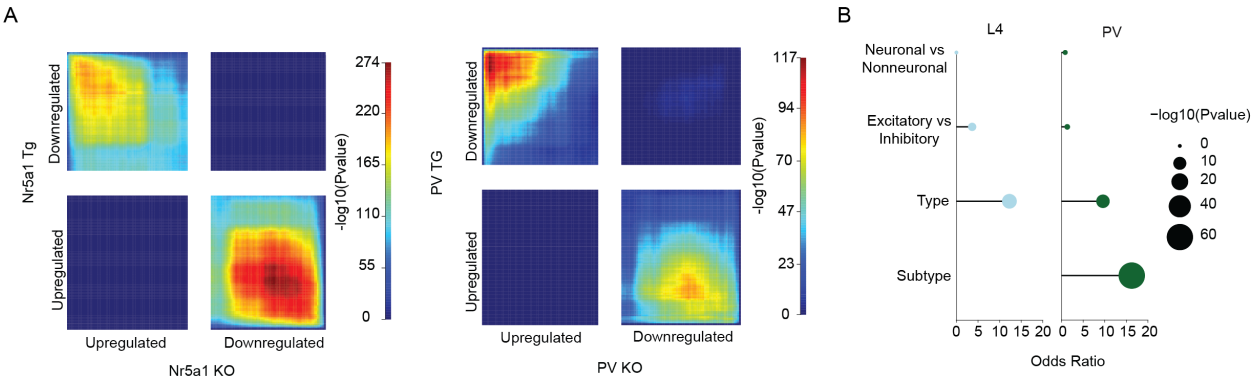


Figure 3.4: MeCP2 overexpression in PV and Layer 4 excitatory neurons produces inverted transcriptomic effects to MeCP2 knockout
(A) RRHO analysis of gene expression changes between for MeCP2 OE/WT and KO/WT in Layer 4 and PV neurons. Genes ranked by sign corrected significance values from DESeq with color representing -log10(p-value) of RRHO overlap.
(B) Overlap of MeCP2-repressed genes (downregulated) in Layer 4 and PV MeCP2 OE neurons with single cell hierarchy distinguishing genes from Allen 2018 data.

3.5 - Discussion

Heterozygous disruption of DNMT3A leads to subtle effects in gene expression and regulatory elements that demonstrate shared mechanisms with conditional deletion of DNMT3A and loss of MeCP2 (Christian et al., 2020). Our transcriptomics analysis of changes in ASD/NDD gene sets in DNMT3A$^{KO/+}$ mice has detected overlap beyond MeCP2 disorders, including mouse NDD/ASD models (e.g., CHD8) and gene sets identified in human ASD. As additional NDD transcriptomics studies of mouse and human brain emerge, systematic analyses can identify shared aspects of transcriptional pathology across diverse causes of NDD. Notably, the large number of chromatin-modifying enzymes mutated in NDD suggests that shared transcriptomic effects may emerge from common chromatin pathology. Future studies may identify additional gene disruptions in which alterations in mCA and enhancer dysregulation contribute to molecular pathology, expanding the role of "methylopathies" in NDD.

Whether the dysfunction in maintaining neuronal subtype transcriptomic identity we detect in the MeCP2 KO is present upon disruption of mCA is unclear. We hypothesize that loss of mCA which is known to be fundamental to MeCP2 function would produce a similar effect on subtype gene sets as MeCP2 KO. Here the data from PV neurons in the DNMT3A$^{KO/+}$ model give preliminary indications that is the case. However, we show shared regulation of ASD gene sets in the DNMT3A$^{KO/+}$ in the whole cortex data, yet we previously failed to detect a strong signal for subtype-defining genes being misregulated in other NDD models as we do for MeCP2 (Chapter 2, Figure S4B). DNMT3A has functions early in prenatal development that are unrelated to MeCP2 and likely the cause of nonoverlapping aspects of DNMT3A and MeCP2 disorders (Lavery et al., 2020). Perhaps this explains why we see significant overlap of DNMT3A$^{KO/+}$ disruption with other

ASD models and an effect on subtype-defining genes that appears to have specificity to the mCA-MeCP2 pathway. Follow-up analyses will investigate whether conditional loss of DNMT3A in postmitotic neurons, and therefore near total loss of mCA, produces dysfunction in subtype identity maintenance.

Our results at the level of cell types in MeCP2 OE mice are consistent with our data from MeCP2 KO studies and indicate that MeCP2 regulates neuronal subtype defining gene sets. While this is strong data, it is difficult to conceptualize in our proposed model of mCA-MeCP2 transcriptomic identity regulation. It is relatively straight forward to imagine a loss of epigenetic regulation allowing overly permissive state changes in neurons. However, the converse would be neurons that are overly constrained and fail to respond to environmental stimuli. Such an effect would potentially be perceptible in single-cell RNA sequencing data with new computational approaches to detect cellular dynamics (la Manno et al., 2018; Riba et al., 2022). Future studies in the MeCP2 overexpression mouse model can test the hypothesis that excess MeCP2 reduces plasticity in transcriptional programs.

Caution must be used when interpreting data from multiple mouse models that come from different genetic backgrounds and vary in ability to replicate pathology. However, attempts to integrate multiple lines of evidence into a larger picture of disease remains a worthwhile endeavor. This becomes even more important when considering a new function for the mCA-MeCP2 pathway as we have proposed. Future experiments elaborating how alternative disruptions of mCA and MeCP2 affect cell-type-specific transcriptional programs will be necessary to improve our understanding of biological mechanisms.

3.6 - Materials and Methods

Mice

Pvalb-Cre mice (B6.129P2-Pvalb$^{tm1(cre)Arbr}$/J) were obtained from The Jackson Laboratory. Nr5a1-Cre mice (FVB-Tg(Nr5a1-cre_2Lowl/J) were generously shared by the Allen Brain Institute. Each of these cre lines were cross to Sun1:GFP mice (B6;129-Gt(ROSA)26Sor$^{tm5(CAG-Sun1/sfGFP)Nat}$/J) obtained from The Jackson Laboratory. DNMT3A$^{KO/+}$ were generated as described in Christian et al. 2020. PV-Cre Sun1:GFP mice were crossed to DNMT3A$^{KO/+}$ knockout mice and wild-type male littermates. MeCP2 overexpression mice (FVB-Tg(MECP2)3Hzo/J) were cryo-recovered from The Jackson Laboratory. Female heterozygous mice (MeCP2$^{Tg3/+}$) were crossed to Pvalb-Cre:Sun1:GFP and Nr5a1-Cre:Sun1:GFP mice to generate hemizygous male transgenic mice (MeCP2$^{Tg3/y}$) and wild-type male litter mates (MeCP2$^{+/y}$).

INTACT

The mouse cortex was quickly dissected in ice-cold homogenization buffer (0.25M sucrose, 25mM KCl, 5mM MgCl2, 20mM Tricine-KOH) and flash frozen in liquid nitrogen and stored at -80C. Tissue was thawed on ice in homogenization buffer containing 1mM DTT, 0.15 spermine, 0.5 spermidine, EDTA-free protease inhibitor, and RNasin Plus RNase Inhibitor (Promega N2611) at 60U/mL for RNA experiments. Tissue was minced using razor blades then dounce homogenized in homogenization buffer using 5 strokes with the loose pestle and tight pestle. A 5% IGELPAL-630 solution was added and the homogenate further dounced 10 times with the tight pestle. The homogenized sample was filtered through a 40um strainer and underlaid with a density gradient. The sample was then slowly spun at 8,000g on a swinging bucket rotor and the nuclei collected from the density interface. Nuclei were then isolated using GFP antibody (Fisher G13062) and

Protein G Dynabeads (Invitrogen 10003D) with all immunoprecipitation steps being performed in a 4C cold room.

Nuclear RNA-seq

RNA from SUN1-purified nuclei was extract using RNeasy micro kit (QIAGEN) following manufacturer instruction and sequencing libraries prepared using the NuGEN/Tecan Ovation SoLo RNA-Seq Library Preparation Kit. Libraries for PV samples were sequenced using NextSeq 500 (Center for Genome Sciences at Washington University).

Rank-Rank Hypergeometric Overlap

RRHO analysis was performed using RRHO2 (Cahill et al., 2018) using hypergeometric distribution of ranked lists of pvalues with sign adjusted by direction of fold change.

Controlled resampling

A similar resampling approach was used as previously described (Clemens et al., 2020) Briefly, for every entry in a sample set (e.g., DNMT3A-dysregulated genes), an entry in the control set (e.g., all other genes) with a similar desired characteristic (e.g., expression) was selected, generating a control set of the same size and variable distribution as the sample set.

GAGE

Gene set enrichment analysis for the gene sets described was performed using the Generally Applicable Gene-set Enrichment (GAGE) program (W. Luo et al., 2009) The NDD models for comparison were chosen by searching for gene expression datasets meeting the following criteria:

1) NDD/ASD models that have at least some similar features to DNMT3A disorders (e.g., ID, ASD) 2) generated with the RNA-seq approach 3) analyzed brain tissue with enough similarity to our cortical analysis to justify a reasonable comparison. Analysis was performed directionally on the shrunken, log-normalized exonic fold changes from DESeq2 analysis of DNMT3A$^{KO/+}$ versus WT RNA-seq data. For each gene set, fold changes of genes in that set were compared to a background of all expressed genes. Gene sets with an FDR q-value below 0.1 and an adjusted p value below 0.5 following expression matched resampling repeated 1,000 times were considered statistically significant. Gene sets were selected for analysis from both human and mouse studies of autism associated genes. SFARI genes (Abrahams et al., 2013) with scores of equal to or less than 3 were considered. Date accessed: 6/20/2019.

GSEA

Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) version 7.0, the Broad Institute of MIT and Harvard; https://www.gsea-msigdb.org/gsea/downloads.jsp) was performed on shrunken, log-normalized exonic fold changes from DESeq2 between DNMT3AKO+ and WT RNA-seq data. GSEA calculated a gene set Enrichment Score (ES) that analyzed genes were enriched in the biological signal conduction on the MsigDB (Molecular Signatures Database; https://www.gsea-msigdb.org/gsea/msigdb). Background was set to all expressed genes in this study and 1,000 permutations were set to generate a null distribution for enrichment score in the hallmark gene sets and functional annotation gene sets. The gene sets database used for enrichment analysis were ‘c5.all.v7.0.symbols.gmt’, ‘c5.bp.v7.0.symbols.gmt’, ‘c5.cc.v7.0.symbols.gmt’and ‘c5.mf.v7.0.symbols.gmt’ and FDR < 0.1 was defined as the cut-off criteria for significance.

Chapter 4: Conclusions and future directions, a testable model of the MeCP2-mCA pathway as a regulator of transcriptional states

4.1 - MeCP2 and mCA as regulators of terminal transcriptomic states

The classification of specific neuronal cell types is hierarchical and single cell RNA sequencing is now defining subclasses of cell types based on high resolution transcriptomes (Ecker et al., 2017). These transcriptomic types are shedding new light on the cellular diversity of the nervous system. High-resolution transcriptomic types represent functionally important cell states in cortical circuits that require different gene expression programs to subserve distinct functions (Economo et al., 2018). There is considerable evidence that neurons complete their final maturation postnatally as they integrate into circuits and receive environmental inputs (Cocas et al., 2016; de Marco García et al., 2015; Tuncdemir et al., 2016; Wamsley & Fishell, 2017). The timing of mCA and MeCP2 buildup coincides with the closure of this early postnatal period of cellular and circuit refinement and has been shown to respond to the transcriptional activity of the cell in the first few weeks of life (Stroud et al., 2017, 2020). Cellular phenotypes following MeCP2 disruption point towards function in the maturation of postmitotic developing neurons (Fukuda et al., 2005; Kishi & Macklis, 2004, 2010; Shahbazian, 2002). Our findings show that the genes that distinguish high-resolution neuronal subtypes are regulated by mCA and MeCP2. We see that neurons lacking MeCP2 demonstrate activation of normally suppressed enhancers from other cell types and that PV subtypes inappropriately express distinguishing marker genes. We term this loss of transcriptional regulation of identity defining genes as a “cellular confusion” phenotype. Our data

lead to the intriguing model that mCA and MeCP2 are critical for the stabilization and maintenance of neuronal diversity at the subtype level. In this way, unique neuronal epigenetic features that arise during postnatal development would contribute to specialization of neurons within microcircuits.

4.2 - Maturation of terminal neuronal identity occurs during the postnatal period

In our proposed model of MeCP2-mCA as critical mediators of terminal transcriptomic identity maturation and maintenance, we hypothesize that the diversification of neuronal subtypes, or terminal transcriptomic differentiation, takes place during the postnatal period when mCA and MeCP2 are increasing. This question of when the full diversity of neuronal cell types is determined has been a focus of developmental neuroscience research and holds high importance for considering the involvement of MeCP2 and mCA in the differentiation process.

Much of this work has concentrated on inhibitory interneurons which can be roughly grouped into cardinal cell types (also referred to as subclasses), such as PV, SST, VIP, and Lamp5 (Kepecs & Fishell, 2014; Paul et al., 2017; Tasic et al., 2018). Cardinal inhibitory cell types originate in different ganglionic eminences, with PV and SST cells coming from progenitors in the medial ganglionic eminence (MGE), and other types, such as VIP and Lamp5, originating from the caudal and lateral ganglionic eminences (Miyoshi, 2019; Nery et al., 2002). However, whether cardinal types, and their respective subtypes, come from distinct progenitor populations within embryonic proliferative zones was unknown until recently. A series of recent studies have sought to answer this question using single cell transcriptomics and epigenomics across the interneuron

developmental period. Two groups performed single cell RNA sequencing at successive embryonic and postnatal timepoints to track the emergence of interneuron diversity (Mayer et al., 2018; Mi et al., 2018). They identified different progenitor populations within MGE and CGE at embryonic stages but were unable to classify them as early precursors to the adult mature subtypes. Once postmitotic, cardinal types could be distinguished but subtypes diversity could not. At postnatal day 10, PV and SST neurons separated transcriptomically but the further diversity within these types was unable to be identified, indicating that subtypes within these broader classes were not yet established. A follow-up study used joint single cell RNA-sequencing and ATAC-seq at embryonic and postnatal timepoints in the developing MGE-derived population (Allaway et al., 2021). They found that it is during the postmitotic stage after birth that PV and SST progressively differentiate themselves from each other and develop discrete subtype populations. This matches functional literature (discussed in Chapter 1) that these neuronal types undergo significant maturation in their morphologic and electrophysiologic properties during this critical developmental period (Bandler et al., 2017; de Marco García et al., 2011; Dehorter et al., 2015; Wamsley & Fishell, 2017).

The story is similar for excitatory neurons. A recent study seeking to address the same question but for excitatory neurons performed joint single cell RNA-seq and ATAC-seq at embryonic and early postnatal timepoints (di Bella et al., 2021). Their data argue against strictly pre-committed progenitors and conclude that while cardinal types can be identified at postnatal day four, the diversity of subtypes that exist within them is established in the subsequent developmental period. Together, these studies support the timing of the emergence of neuronal subtypes, or maturation of terminal transcriptomic identities, coincident with mCA and MeCP2 activity. These results

largely confirm what has been known from functional studies (Molyneaux et al., 2007; Wamsley & Fishell, 2017) but still provide important evidence that the overlapping timing of mCA and MeCP2 with the terminal transcriptomic differentiation process.

4.3 - Theory of cell types as stable attractor states and considerations for our proposed model

When attempting to understand a complex system like the brain, an important step is identifying its component parts. This fundamental truth has motivated efforts to characterize all the cell types in the human brain and relevant model organisms such that a complete "periodic table" of cell types is created. In recent years, studies to accomplish this ambitious task have made astounding progress. However, this explosion in data has raised new questions about what constitutes a discrete "cell type." The BRAIN Initiative Cell Census Network (BICCN) has considered cell types by various dimensions, including morphology, physiology, transcriptomic, positioning and function (Ecker et al., 2017). This inclusive definition has practical benefits but faces issues with integrating multiple modalities (Fishell & Heintz, 2013; Vlasits et al., 2019; Zeng & Sanes, 2017). Approaching the cell type question by focusing on genetic and developmental processes offers an interesting perspective and is most useful for considering the contributions of mCA and MeCP2.

Traditionally, cell-type-specific gene expression programs are largely thought to result from the activity of a unique set of transcription factors belonging to each cell type. This holds that transcription factors are combinatorial in nature, and it is through multiplexing that cell-type diversity is achieved. The search for evidence of these transcription factor networks works well for cardinal neuron types, such as PV or SST, but struggles to explain the subtypes within these

broader subclasses that emerge developmentally late in postmitotic neurons (Fishell & Kepecs, 2020). Studies have found that late acting transcription factors, defined by their expression in postmitotic maturing neurons, when deleted postnatally do not lead to broad loss of neuronal types (Batista-Brito et al., 2009, Close et al., 2012, Miyoshi et al., 2015). In *C. elegans* genetic studies have indicated a terminal selector model better describes neuronal identity (Hobert, 2008, 2016). In such a model, expression of "effector genes" endows a neuron with its defining characteristics such as a set of ion channels to produce an electrophysiological profile or cell adhesion molecules to create its connectivity. Collections of transcription factors or "terminal selectors" act independently to activate these effector genes, but they are configured rather than combinatorial, meaning that not every neuronal type has its own unique barcoded combination of terminal selectors. Rather transcription factors are redundantly and iteratively used across cell types. This theory highlights the need for repressive influences to limit the targets of broad terminal selectors to generate a unique expression program for highly related subtypes. A cell type at the highest resolution is a self-stabilizing system that maintains a cell-type-specific gene regulatory program. Newer theoretical arguments incorporating knowledge of dynamical systems have proposed that cell types reflect stable attractor states (Fishell & Kepecs, 2020; Mukamel & Ngai, 2019). In the face of perturbations cells are recurrently influenced by attractor networks to return to local energy minima or basins.

Considering this theory in the context of mCA and MeCP2 provides an appealing explanation for our results and why this unique regulatory mechanism evolved in neurons. In one possible framework, strong transcription factor networks produce cardinal neuron types that then undergo further diversification postnatally as they adapt to their surroundings and circuit influences. These

subtypes, or terminal transcription states, are only weakly configured by late acting transcription factors and as they develop require mCA and MeCP2 to tune their transcriptional program and deepen the energy basin separating states. Thus, a self-stabilizing system is created that maintains a cell-type-specific gene expression program against extrinsic perturbations with mCA and MeCP2 providing an epigenetic robustness that returns neurons to an equilibrium. When the regulation by MeCP2 and mCA is lost cells can be pushed out of the equilibrium state by stochastic stimuli, resulting in a chaotic loss of transcriptomic stability. Our data show that neuronal subtype effector genes tend to be misregulated when MeCP2 is lost and that these effector genes as a population show characteristics of MeCP2-regulated genes. However more work is required to test this theory experimentally. Next, I will describe ongoing and future experiments to do so.

4.4 - Single cell RNA sequencing in MeCP2$^{KO/+}$ heterozygous mutant brains

Our analysis of PV cells lacking MeCP2 shows that genes that distinguish PV subtypes are dysregulated suggesting that there a failure to maintain stable terminal transcriptomic states. To broaden this analysis beyond the subtypes and marker genes tested by RNAScope, we have turned to single-cell RNA sequencing. Analysis of changes in high-resolution transcriptotypes requires high-depth sequencing with low technical noise and batch variability between wild-type and mutant expression datasets. We will exploit the genetics of MeCP2 to overcome these technical challenges and provide unprecedented sensitivity to cellular transcriptome changes at a subtype level. As an X-linked gene, female *Mecp2*$^{KO/+}$ mice contain a portion of their cells expressing the wild-type *Mecp2* allele and a portion expressing the mutant form. In *Mecp2*$^{KO/+}$ mice, the mutant and wild-type alleles can be distinguished by single nucleotide polymorphisms (SNPs) maintained in cis with the *Mecp2* mutant allele (Renthal et al., 2018). This presents a powerful system for

probing cell autonomous versus non-cell autonomous effects while controlling for batch effects in cell isolation and sequencing procedures. We hypothesize that the transcriptomic diversity of PV neuron subtypes will be decreased upon loss of MeCP2.

I have performed single cell dissociations from visual cortices from in female *Mecp2*$^{KO/+}$; PV-Cre:Sun1-GFP mice. I sorted GFP+ positive cells into 96-well plates for preparation of RNA-sequencing libraries using the Smart-Seq v2 protocol in collaboration with Tristan Li's lab. These libraries are then deep sequenced on the NovaSeq to a target depth of 500 thousand to one million reads per cell to capture the full transcriptome of cells and match the current largest single-cell RNA sequencing datasets from the visual cortex (Tasic et al. 2016; 2018). These experiments are ongoing, but early data analysis of a trial dataset reveals that we are generating quality single cell libraries. The majority of cells collected express PV markers indicating that we are capturing our cells of interest and very few cells show signs that they are doublets (high expression of excitatory or nonneuronal markers, total reads above two standard deviations from the mean). Our ability to call the genotype of cells as either MeCP2 KO or WT is still being assessed but may require additional sequencing to the identifying SNPs using specific primers.

Once the full data is acquired, we plan to analyze the MeCP2 KO and WT neurons to look for evidence of cellular confusion phenotypes. We hypothesize that the transcriptomic diversity of PV neuron subtypes will be decreased upon loss of MeCP2. We will perform iterative hierarchical clustering analysis to determine transcriptomic clusters for MeCP2 KO and WT populations (Tasic et al., 2018). We expect to detect fewer distinct transcriptional clusters in the cells lacking MeCP2 compared to wild-type cells indicating that there is decreased ability to parse these cells based on

their transcriptomes. Our deep sequencing approach will allow us to assess continuity in addition to discreteness. Using a validated approach (Tasic et al., 2018), we will perform analysis using a nearest-centroid classifier to classify cells as core (classifies to the same cluster > 90% of the time) or intermediate (classifies to the same cluster < 90% of the time). We hypothesize that the MeCP2 KO cells will contain greater numbers of intermediate cells classifying between clusters compared to WT cells. To further demonstrate reduced cellular diversity in the MeCP2 KO cells, we will apply the MetaNeighbor algorithm to assess the cross-dataset replicability of clusters of KO and WT cells to the consensus single-cell RNA sequencing dataset from the Allen Brain Institute (Crow et al., 2018; Tasic et al., 2018). This analysis will test whether the KO and WT cells can be predicted using the closest matching (nearest neighbor) cells in the consensus dataset, providing a quantitative measurement (AUROC) indicating the replicability score. We hypothesize that WT cells will correlate with the consensus dataset much more than KO cells. This will show that cells lacking MeCP2 fail to maintain mature, wild-type transcriptional identities.

4.5 - Cellular confusion analysis using spatial transcriptomics

When considering questions of neuronal cell type identity, the spatial environment of a neuron must be taken into context. As many have noted, cellular positioning can mean neurons participate in different information streams (Bugeon et al., 2022; D'Souza et al., 2016). Spatial transcriptomic technologies that assess single cell gene expression profiles with position information included represent an exciting new frontier in neuroscience (Lein et al., 2017). Scientists are already using this technology combined with imaging of neuronal activity to powerfully link molecular identification with functional roles in neural circuits (Bugeon et al., 2022; Condylis et al., 2022). A challenge to these approaches is reliably detecting large numbers of genes while retaining single-

cell resolution. This technical feat was accomplished through the development of multiplexed-error-robust-fluorescent-in-situ hybridization or MERFISH (Chen et al., 2015). MERFISH is a single molecule FISH method that uses combinatorial labeling and sequential imaging with error-robust encoding schemes to greatly expand the number of genes profiled at single cell resolution. A company, Vizgen, now offers this technology through their MERSCOPE platform which allows for imaging the transcripts of up to 500 genes in cells with *in situ* positioning information. The BICCN recently used MERSCOPE to probe cell types in the motor cortex and found it successfully identified the complement of neuronal subtypes characterized by single cell RNA-sequencing (Zhang et al., 2021). Our initial findings using RNAScope of improper marker gene expression in MeCP2 KO neurons led us to turn to MERSCOPE to greatly expand our findings to more genes across all cell types with spatial context. We found that of the marker genes analyzed by the BICCN MERSCOPE motor cortex paper, nearly half were misregulated in our MeCP2 datasets from cell type and brain region analysis. Considering this, we can use their validated probe set to test our hypothesis that MeCP2 regulates subtype-defining genes. To compare knockout and wild-type cells next to each other in the same sample, we will perform this experiment in the *Mecp2*$^{KO/+}$ heterozygous female. We designed probes for *Mecp2* that are specific for the deleted portion of the transcript in the MeCP2 KO mouse so we can identify MeCP2 KO and WT cells. We hypothesize that we will see co-expression of putatively mutually exclusive markers genes in the MeCP2 KO cells at a far greater frequency than we do in WT cells. Using MERSCOPE we can extend our findings to more genes and cell types and consider spatial context with regards to MeCP2 regulation of cell-type identity.

4.6 - Developmental timing of transcriptional dysfunction in MeCP2 KO

While the vast majority of evidence indicates a role for MeCP2 in regulating neuronal maturation in the postnatal brain, there are reported effects at earlier timepoints (Lozovaya et al., 2019; Orefice et al., 2016). An important test to strengthen our model that mCA and MeCP2 regulation in the first few weeks of life are critical to the stabilization of mature transcriptional states is to analyze an early developmental time point prior to mCA and MeCP2 buildup. At postnatal day 10 (P10), when mCA and MeCP2 levels are low in neurons, we hypothesize that we will find no differences in the transcriptomes of MeCP2 KO and WT cells. Using single-cell RNA sequencing of PV neurons isolated from $MeCP2^{KO/+}$ cortices at P10, we predict we will identify equivalent clusters in KO and WT cells and few misregulated genes. We also expect KO and WT cells to map equally well onto a published P10 dataset of PV interneurons using the MetaNeighbor algorithm (Crow et al., 2018; Mayer et al., 2018). Additionally, consistent with previous literature discussed above, we expect to detect a lack of PV subtypes in both *Mecp2*-null and wild-type populations at the P10 timepoint compared to the P56 timepoint. A lack of distinction between MeCP2 KO and WT cells at an early developmental timepoint would be an important demonstration of the specificity of the mCA-MeCP2 pathway to neuronal subtype diversification.

Given the observations that MeCP2 plays a role in critical period closure and that PV-neurons with deleted *Mecp2* resemble immature neurons (He et al., 2014; Krishnan et al., 2015; Picard & Fagiolini, 2019), a possible third timepoint at the height of plasticity in the visual cortex must be considered. We hypothesize that MeCP2 acts to reduce transcriptomic plasticity, and this seems to be reflected at the level of circuits with its involvement in critical period closure. Thus, perhaps in the absence of MeCP2 neurons fail to constrain their plastic states. We would hypothesize then

that our adult MeCP2 KO neurons analyzed at P56 would most resemble an earlier hyper-plastic state in development such as postnatal day 28, P28, in the visual cortex. We could test this through single cell RNA-seq of WT neurons at P28 and then using MetaNeighbor to map our P56 MeCP2 KO neurons, with the prediction that our P56 MeCP2 KO neurons will show stronger mapping to the WT P28 timepoint that to WT P56 neurons. If found, this would produce convincing evidence for failure to stabilize hyperplastic states without MeCP2.

4.7 - Rescue of MeCP2 cellular confusion phenotype

While our data provide evidence that cells lacking MeCP2 fail to maintain mature transcriptional identities, questions remain about whether this dysfunction is a specific consequence of the loss of MeCP2 transcriptional regulation or rather representative of a more general disease state. Amazingly, re-expression of MeCP2 in the adult rescues Rett-like phenotypes (Guy et al., 2007; Luikenhuis et al., 2004). This, along with the observation that substantial cell loss does not occur in patient samples and mouse models of Rett syndrome (Armstrong et al., 1995; Chen et al., 2001; Kishi & Macklis, 2004; Renthal et al., 2018), indicates that lack of MeCP2 does not lead to neurodegeneration. However, it is not clear what transcriptional regulation by MeCP2 can be rescued, and the importance of neuronal subtype "cellular confusion" to pathology has not been tested. We predict that reinstatement of MeCP2 in the adult brain will lead to appropriate read out of mCA and lead to reinstatement of high-resolution subtypes. This observation would support direct, cell-autonomous involvement of MeCP2 in the stabilization of subtype-identity.

For this experiment, we will use MeCP2 lox-STOP mice that possess a *loxP*-flanked STOP cassette in intron 2 of the *Mecp2* gene that upon Cre-mediated excision leads to normal expression of MeCP2 (Guy et al., 2007). This mice will be crossed to Sun1:GFP mice to allow for isolating these neurons. Female heterozygous MeCP2$^{KO/+}$mice will be injected with AAV-Syn-Cre and control AAV-Syn-Cre virus into the visual cortex at P30-35. Four weeks following injection, brain tissue will be collected to test for rescue of the cellular confusion phenotype either by RNAScope or single cell RNA sequencing. We hypothesize that this rescue in the will lead to infected PV neurons appearing identical to wild-type cells. By RNAScope we predict that the inappropriate subtype gene expression we observe will be ameliorated for infection neurons (see Chapter 2 Figure 2.4E). For single cell RNA-sequencing, we plan to perform MetaNeighbor analysis on our rescued neurons compared to our control injected neurons, correlating both groups with wild-type neurons from the same experiment and with the consensus single-cell transcriptomic dataset. We expect that our rescued neurons will demonstrate significantly higher correlation with the wild-type and consensus datasets compared to non-rescued neurons, indicating that re-expression of MeCP2 in the adult neurons stabilize their transcriptomes into mature subtypes that mirror their wild-type counterparts.

4.8 – Role of intragenic enhancers in MeCP2 cell-type identity regulation

MeCP2 inhibits transcription by repressing enhancers, primarily those within genes. We find that when MeCP2 is lost there is activation of enhancers in cell types where they are normally suppressed, and this may be a central cause of the disruption of cell-type-specific transcription we observe in the MeCP2 KO. New approaches for joint profiling of open chromatin and RNA simultaneously in single cells (Cao et al., 2018), may be of use in examining this phenomenon in

the MeCP2 KO at the single cell level. Joint ATAC-RNA profiling in the MeCP2 KO would identify putative enhancers linked to genes in MeCP2 KO and WT cells (Y. E. Li et al., 2021). We would expect in *Mecp2*-null cells to see aberrant activation of putative enhancers in cell types not found in the wild-type cells. While this experiment would be technically challenging, if effective it would allow for the investigation of the inappropriate cell type enhancer activation we discovered at scale for many cell-types across the brain.

A curious observation from our data finds that intragenic rather than extragenic enhancers as central sites of MeCP2 regulation. MeCP2-repressed genes also show an enrichment for having large numbers of enhancers located within their gene bodies. This raises the question as to why enhancers located inside genes are especially targeted by mCA-MeCP2. We have shown that enhancers within a gene interact more with their promoter than enhancers located outside of genes at an equivalent distance from the promoter (Clemens et al., 2020). Enhancers can function in diverse ways, from promoting transcriptional bursting to operating as theorized "super-enhancers" that drive more stable transcription (Nord & West, 2020; Panigrahi & O'Malley, 2021; Shlyueva et al., 2014). Intragenic enhancers may serve an important role for long, complex neuronal genes.

What does this bias towards mCA and MeCP2 regulating intragenic enhancers mean for our proposed model of neuronal subtype stabilization? Are the regulatory elements targeted by terminal selectors preferentially located inside of genes? This final question likely could be addressed through careful analysis of available single cell ATAC, methylation, and RNA sequencing data to identify terminal selectors and their target enhancers. An intriguing bit of data found that neurons, but not other tissue types, accumulate enhancers inside of genes over postnatal

development (Nord et al., 2013). This timing of this increase raises the possibility that in the nervous system intragenic enhancers serve a privileged transcriptional role related to the processes regulated by mCA and MeCP2. The increase in number of enhancers regulating a gene could allow for greater variability in transcription regulation, cell-type specificity, or responsiveness to stimuli (Nord & West, 2020; Tyssowski & Gray, 2019). Perhaps the postnatal transition to a self-stabilizing transcriptional state in neurons relies on an increase in intragenic enhancers, which in turn explains why they are preferential targets of MeCP2. Fundamental insights into neuronal enhancer biology in the coming years will need to be carefully attended to in order to understand of how the regulation of intragenic enhancers by mCA and MeCP2 contributes to their proposed role in subtype identity stabilization.

4.9 - Conclusion

What the insights described here mean for the ultimate goal of our research, understanding and remedying disease, is unclear. It has been proposed that mutations in chromatin regulators that cause neurodevelopmental disorders, broadly represent state changes in neuronal gene networks (Sullivan et al., 2019). We offer up a similar idea here for mCA and MeCP2 but link the state change induced upon dysfunction of mCA-MeCP2 to a specific developmental period and biological process. The postnatal buildup of mCA and MeCP2, as well as the reversibility of phenotypes through reintroduction of MeCP2 in adulthood, point to a special function in maturing neurons. Further, the planned experiments can directly test our hypotheses. Future experiments will need to explore whether the reduced transcriptomic maturation we detect in the MeCP2 KO is reproduced in other models affecting the MeCP2-mCA regulatory axis, such as mice overexpressing MeCP2 and deletion of DNMT3A. Detecting shared mechanisms between disease

states is an important path forward in the efforts to develop therapeutics that can broadly target multiple pathologies.

In considering our model I am reminded of the George Box quote: “all models are wrong, but some are useful.” There are many factors at play in the formation of high resolution neuronal subtypes beyond our model, such as the contributions of isoform specificity (Booeshaghi et al., 2021) and alternative splicing (X. Zhang et al., 2016). Questions about what constitutes a discrete cell type versus a transient cell state are difficult to decipher. However, the data described in this dissertation uncover exciting new insights and open up many significant new research areas. The experiments outlined in the discussion should produce highly informative data on the function of mCA and MeCP2 as critical epigenomic factors in the formation and maintenance of stable transcriptomic states. This will shed light on the mystery of why these epigenomic features developed only in the nervous system and perhaps how they contributed to the explosion of diverse neuronal cell types.

I would argue the most impressive accomplishment of the brain is the ability to balance both plasticity and stability and it is fascinating that one way this may be achieved is through transcriptional regulation by mCA and MeCP2.

References

Abrahams, B. S., Arking, D. E., Campbell, D. B., Mefford, H. C., Morrow, E. M., Weiss, L. A., Menashe, I., Wadkins, T., Banerjee-Basu, S., & Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Molecular Autism*, *4*(1), 36. https://doi.org/10.1186/2040-2392-4-36

Allaway, K. C., Gabitto, M. I., Wapinski, O., Saldi, G., Wang, C. Y., Bandler, R. C., Wu, S. J., Bonneau, R., & Fishell, G. (2021). Genetic and epigenetic coordination of cortical interneuron development. *Nature*. https://doi.org/10.1038/s41586-021-03933-1

Amir, R. E., van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, *23*(2), 185–188. https://doi.org/10.1038/13810

Anderson, G. R., Maxeiner, S., Sando, R., Tsetsenis, T., Malenka, R. C., & Südhof, T. C. (2017). Postsynaptic adhesion GPCR latrophilin-2 mediates target recognition in entorhinal-hippocampal synapse assembly. *Journal of Cell Biology*, *216*(11), 3831–3846. https://doi.org/10.1083/jcb.201703042

Armstrong, D. D. (2005). Neuropathology of Rett syndrome. *Journal of Child Neurology*, *20*(9), 747–753.

Armstrong, D., Dunn, J. K., Antalffy, B., & Trivedi, R. (1995). Selective Dendritic Alterations in the Cortex of Rett Syndrome. *Journal of Neuropathology & Experimental Neurology*, *54*(2), 195–201. https://doi.org/10.1097/00005072-199503000-00006

Bakken, T. E., Jorstad, N. L., Hu, Q., Lake, B. B., Tian, W., Kalmbach, B. E., Crow, M., Hodge, R. D., Krienen, F. M., Sorensen, S. A., Eggermont, J., Yao, Z., Aevermann, B. D., Aldridge, A. I., Bartlett, A., Bertagnolli, D., Casper, T., Castanon, R. G., Crichton, K., … Lein, E. S. (2021). Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature*, *598*(7879), 111–119. https://doi.org/10.1038/s41586-021-03465-8

Bandler, R. C., Mayer, C., & Fishell, G. (2017). Cortical interneuron specification: the juncture of genes, time and geometry. *Current Opinion in Neurobiology*, *42*, 17–24. https://doi.org/10.1016/J.CONB.2016.10.003

Banerjee-Basu, S., & Packer, A. (2010). SFARI Gene: an evolving database for the autism research community. *Disease Models & Mechanisms*, *3*(3–4), 133–135. https://doi.org/10.1242/dmm.005439

Bauman, M. L., Kemper, T. L., & Arin, D. M. (1995). Microscopic observations of the brain in Rett syndrome. *Neuropediatrics*, *26*(02), 105–108.

Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C. A., & Zoghbi, H. Y. (2009). Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus [JOUR]. *Human Molecular Genetics*, *18*(13), 2431–2442.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. In *Genes and Development*. https://doi.org/10.1101/gad.947102

Blackman, M. P., Djukic, B., Nelson, S. B., & Turrigiano, G. G. (2012). A Critical and Cell-Autonomous Role for MeCP2 in Synaptic Scaling Up. *The Journal of Neuroscience*, *32*(39), 13529 LP – 13536. https://doi.org/10.1523/JNEUROSCI.3077-12.2012

Booeshaghi, A. S., Yao, Z., van Velthoven, C., Smith, K., Tasic, B., Zeng, H., & Pachter, L. (2021). Isoform cell-type specificity in the mouse primary motor cortex. *Nature*, *598*(7879), 195–199. https://doi.org/10.1038/s41586-021-03969-3

Boxer, L. D., Renthal, W., Greben, A. W., Whitwam, T., Silberfeld, A., Stroud, H., Li, E., Yang, M. G., Kinde, B., Griffith, E. C., Bonev, B., & Greenberg, M. E. (2020). MeCP2 Represses the Rate of Transcriptional Initiation of Highly Methylated Long Genes. *Molecular Cell*, *77*(2), 294-309.e9. https://doi.org/10.1016/J.MOLCEL.2019.10.032

Bugeon, S., Duffield, J., Dipoppa, M., Ritoux, A., Prankerd, I., Nicoloutsopoulos, D., Orme, D., Shinn, M., Peng, H., Forrest, H., Viduolyte, A., Reddy, C. B., Isogai, Y., Carandini, M., & Harris, K. D. (2022). A transcriptomic axis predicts state modulation of cortical interneurons. *Nature*, *607*(7918), 330–338. https://doi.org/10.1038/s41586-022-04915-7

Bystron, I., Blakemore, C., & Rakic, P. (2008). Development of the human cerebral cortex: Boulder Committee revisited. *Nature Reviews Neuroscience*, *9*(2), 110–122. https://doi.org/10.1038/nrn2252

Cahill, K. M., Huo, Z., Tseng, G. C., Logan, R. W., & Seney, M. L. (2018). Improved identification of concordant and discordant gene expression signatures using an updated rank-rank hypergeometric overlap approach. *Scientific Reports*. https://doi.org/10.1038/s41598-018-27903-2

Cao, J., Cusanovich, D. A., Ramani, V., Aghamirzaie, D., Pliner, H. A., Hill, A. J., Daza, R. M., McFaline-Figueroa, J. L., Packer, J. S., Christiansen, L., Steemers, F. J., Adey, A. C., Trapnell, C., & Shendure, J. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science*, *361*(6409), 1380–1385. https://doi.org/10.1126/SCIENCE.AAU0730

Carullo, N. V. N., & Day, J. J. (2019). Genomic Enhancers in Brain Health and Disease. In *Genes* (Vol. 10, Issue 1). https://doi.org/10.3390/genes10010043

Chahrour, M., Sung, Y. J., Shaw, C., Zhou, X., Wong, S. T. C., Qin, J., & Zoghbi, H. Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science*, *320*(5880), 1224–1229. https://doi.org/10.1126/SCIENCE.1153252

Chao, H.-T., Chen, H., Samaco, R. C., Xue, M., Chahrour, M., Yoo, J., Neul, J. L., Gong, S., Lu, H.-C., Heintz, N., Ekker, M., Rubenstein, J. L. R., Noebels, J. L., Rosenmund, C., & Zoghbi, H. Y. (2010). Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature*, *468*(7321), 263–269. https://doi.org/10.1038/nature09582

Chao, H.-T., Zoghbi, H. Y., & Rosenmund, C. (2007). MeCP2 Controls Excitatory Synaptic Strength by Regulating Glutamatergic Synapse Number. *Neuron*, *56*(1), 58–65. https://doi.org/10.1016/j.neuron.2007.08.018

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., & Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, *348*(6233), aaa6090.

Chen, L., Chen, K., Lavery, L. A., Baker, S. A., Shaw, C. A., Li, W., & Zoghbi, H. Y. (2015a). MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), 5509–5514. https://doi.org/10.1073/PNAS.1505909112

Chen, L., Chen, K., Lavery, L. A., Baker, S. A., Shaw, C. A., Li, W., & Zoghbi, H. Y. (2015b). MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proceedings of the National Academy of Sciences*, *112*(17), 5509–5514. https://doi.org/10.1073/pnas.1505909112

Chen, R. Z., Akbarian, S., Tudor, M., & Jaenisch, R. (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nature Genetics*, *27*(3), 327–331. https://doi.org/10.1038/85906

Cholewa-Waclaw, J., Shah, R., Webb, S., Chhatbar, K., Ramsahoye, B., Pusch, O., Yu, M., Greulich, P., Waclaw, B., & Bird, A. P. (2019). Quantitative modelling predicts the impact of DNA methylation on RNA polymerase II traffic. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(30), 14995–15000. https://doi.org/10.1073/PNAS.1903549116

Christian, D. L., Wu, D. Y., Martin, J. R., Moore, J. R., Liu, Y. R., Clemens, A. W., Nettles, S. A., Kirkland, N. M., Papouin, T., Hill, C. A., Wozniak, D. F., Dougherty, J. D., & Gabel, H. W. (2020). DNMT3A Haploinsufficiency Results in Behavioral Deficits and Global Epigenomic Dysregulation Shared across Neurodevelopmental Disorders. *Cell Reports*, *33*(8), 108416. https://doi.org/10.1016/j.celrep.2020.108416

Clemens, A. W., Wu, D. Y., Moore, J. R., Christian, D. L., Zhao, G., & Gabel, H. W. (2020). MeCP2 Represses Enhancers through Chromosome Topology-Associated DNA Methylation. *Molecular Cell*, *77*(2), 279-293.e8. https://doi.org/10.1016/j.molcel.2019.10.033

Closser, M., Guo, Y., Wang, P., Patel, T., Jang, S., Hammelman, J., de Nooij, J. C., Kopunova, R., Mazzoni, E. O., Ruan, Y., Gifford, D. K., & Wichterle, H. (2022). An expansion of the non-coding genome and its regulatory potential underlies vertebrate neuronal diversity. *Neuron*. https://doi.org/10.1016/j.neuron.2021.10.014

Cocas, L. A., Fernandez, G., Barch, M., Doll, J., Zamora Diaz, I., & Pleasure, S. J. (2016). Cell Type-Specific Circuit Mapping Reveals the Presynaptic Connectivity of Developing Cortical Circuits. *Journal of Neuroscience*, *36*(11), 3378–3390. https://doi.org/10.1523/JNEUROSCI.0375-15.2016

Cohen, D., Lazar, G., Couvert, P., Desportes, V., Lippe, D., Mazet, P., & Héron, D. (2002). MECP2 mutation in a boy with language disorder and schizophrenia. *American Journal of Psychiatry*, *159*(1), 148-a.

Cohen, S., Gabel, H. W., Hemberg, M., Hutchinson, A. N., Sadacca, L. A., Ebert, D. H., Harmin, D. A., Greenberg, R. S., Verdine, V. K., Zhou, Z., Wetsel, W. C., West, A. E., & Greenberg, M. E. (2011). Genome-Wide Activity-Dependent MeCP2 Phosphorylation Regulates Nervous System Development and Function. *Neuron*, *72*(1), 72–85. https://doi.org/10.1016/j.neuron.2011.08.022

Collins, A. L., Levenson, J. M., Vilaythong, A. P., Richman, R., Armstrong, D. L., Noebels, J. L., David Sweatt, J., & Zoghbi, H. Y. (2004). Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Human Molecular Genetics*, *13*(21), 2679–2689. https://doi.org/10.1093/hmg/ddh282

Condylis, C., Ghanbari, A., Manjrekar, N., Bistrong, K., Yao, S., Yao, Z., Nguyen, T. N., Zeng, H., Tasic, B., & Chen, J. L. (2022). Dense functional and molecular readout of a circuit hub in sensory cortex. *Science*, *375*(6576), eabl5981. https://doi.org/10.1126/science.abl5981

Connelly, J. C., Cholewa-Waclaw, J., Webb, S., Steccanella, V., Waclaw, B., & Bird, A. (2020). Absence of MeCP2 binding to non-methylated GT-rich sequences in vivo. *Nucleic Acids Research*, *48*(7), 3542–3552.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., & Sharp, P. A. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, *107*(50), 21931–21936.

Crow, M., Paul, A., Ballouz, S., Huang, Z. J., & Gillis, J. (2018). Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nature Communications*, *9*(1), 884. https://doi.org/10.1038/s41467-018-03282-0

Curie, A., Lesca, G., Bussy, G., Manificat, S., Arnaud, V., Gonzalez, S., Revol, O., Calender, A., Gérard, D., & des Portes, V. (2017). Asperger syndrome and early-onset schizophrenia associated with a novel MECP2 deleterious missense variant. *Psychiatric Genetics*, *27*(3), 105–109.

Cusanovich, D. A., Hill, A. J., Aghamirzaie, D., Daza, R. M., Pliner, H. A., Berletch, J. B., Filippova, G. N., Huang, X., Christiansen, L., DeWitt, W. S., Lee, C., Regalado, S. G., Read, D. F., Steemers, F. J., Disteche, C. M., Trapnell, C., & Shendure, J. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell*, *174*(5), 1309-1324.e18. https://doi.org/10.1016/j.cell.2018.06.052

de Marco García, N. v, Karayannis, T., & Fishell, G. (2011). Neuronal activity is required for the development of specific cortical interneuron subtypes. *Nature*, *472*(7343), 351–355. https://doi.org/10.1038/nature09865

de Marco García, N. v, Priya, R., Tuncdemir, S. N., Fishell, G., & Karayannis, T. (2015). Sensory inputs control the integration of neurogliaform interneurons into cortical circuits. *Nature Neuroscience*, *18*(3), 393–401. https://doi.org/10.1038/nn.3946

de Mendoza, A., Poppe, D., Buckberry, S., Pflueger, J., Albertin, C. B., Daish, T., Bertrand, S., de la Calle-Mustienes, E., Gómez-Skarmeta, J. L., Nery, J. R., Ecker, J. R., Baer, B., Ragsdale, C. W., Grützner, F., Escriva, H., Venkatesh, B., Bogdanovic, O., & Lister, R. (2021). The emergence of the brain non-CpG methylation system in vertebrates. *Nature Ecology & Evolution*, *5*(3), 369–378. https://doi.org/10.1038/s41559-020-01371-2

Dehorter, N., Ciceri, G., Bartolini, G., Lim, L., Pino, I. del, & Marín, O. (2015). Tuning of fast-spiking interneuron properties by an activity-dependent transcriptional switch*. *Science (New York, N.Y.)*, *349*(6253), 1216. https://doi.org/10.1126/SCIENCE.AAB3415

del Gaudio, D., Fang, P., Scaglia, F., Ward, P. A., Craigen, W. J., Glaze, D. G., Neul, J. L., Patel, A., Lee, J. A., Irons, M., Berry, S. A., Pursley, A. A., Grebe, T. A., Freedenberg, D., Martin, R. A., Hsich, G. E., Khera, J. R., Friedman, N. R., Zoghbi, H. Y., … Roa, B. B. (2006). Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *8*(12), 784–792. https://doi.org/10.109701.gim.0000250502.28516.3c

Deshwar, A. R., Martin, N., Shannon, P., & Chitayat, D. (2020). A homozygous pathogenic variant in SHROOM3 associated with anencephaly and cleft lip and palate. *Clinical Genetics*, *98*(3), 299–302.

di Bella, D. J., Habibi, E., Stickels, R. R., Scalia, G., Brown, J., Yadollahpour, P., Yang, S. M., Abbate, C., Biancalani, T., Macosko, E. Z., Chen, F., Regev, A., & Arlotta, P. (2021). Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature*, *595*(7868), 554–559. https://doi.org/10.1038/s41586-021-03670-5

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

D'Souza, R. D., Meier, A. M., Bista, P., Wang, Q., & Burkhalter, A. (2016). Recruitment of inhibition and excitation across mouse visual cortex depends on the hierarchy of interconnecting areas. *ELife*, *5*. https://doi.org/10.7554/eLife.19332

Ecker, J. R., Geschwind, D. H., Kriegstein, A. R., Ngai, J., Osten, P., Polioudakis, D., Regev, A., Sestan, N., Wickersham, I. R., & Zeng, H. (2017). The BRAIN Initiative Cell Census Consortium: Lessons Learned toward Generating a Comprehensive Brain Cell Atlas. *Neuron*, *96*(3), 542–557. https://doi.org/10.1016/j.neuron.2017.10.007

Economo, M. N., Viswanathan, S., Tasic, B., Bas, E., Winnubst, J., Menon, V., Graybuck, L. T., Nguyen, T. N., Smith, K. A., Yao, Z., Wang, L., Gerfen, C. R., Chandrashekar, J., Zeng, H., Looger, L. L., & Svoboda, K. (2018). Distinct descending motor cortex pathways and their roles in movement. *Nature*, *563*(7729), 79–84. https://doi.org/10.1038/s41586-018-0642-9

Fazel Darbandi, S., Robinson Schwartz, S. E., Qi, Q., Catta-Preta, R., Pai, E. L.-L., Mandell, J. D., Everitt, A., Rubin, A., Krasnoff, R. A., Katzman, S., Tastad, D., Nord, A. S., Willsey, A. J., Chen, B., State, M. W., Sohal, V. S., & Rubenstein, J. L. R. (2018). Neonatal Tbr1 Dosage Controls Cortical Layer 6 Connectivity. *Neuron*, *100*(4), 831-845.e7. https://doi.org/https://doi.org/10.1016/j.neuron.2018.09.027

Fishell, G., & Heintz, N. (2013). The neuron identity problem: Form meets function. *Neuron*, *80*(3), 602–612. https://doi.org/10.1016/j.neuron.2013.10.035

Fishell, G., & Kepecs, A. (2020). Interneuron types as attractors and controllers. *Annual Review of Neuroscience*, *43*, 1.

Fukuda, T., Itoh, M., Ichikawa, T., Washiyama, K., & Goto, Y. (2005). Delayed maturation of neuronal architecture and synaptogenesis in cerebral cortex of Mecp2-deficient mice. *Journal of Neuropathology & Experimental Neurology*, *64*(6), 537–544.

Gabel, H. W., Kinde, B., Stroud, H., Gilbert, C. S., Harmin, D. A., Kastan, N. R., Hemberg, M., Ebert, D. H., & Greenberg, M. E. (2015). Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature*, *522*(7554), 89–93. https://doi.org/10.1038/nature14319

Gandal, M. J., Zhang, P., Hadjimichael, E., Walker, R. L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., Shieh, A. W., Haney, J., Parhami, S., Belmont, J., Kim, M., Moran Losada, P., Khan, Z., Mleczko, J., Xia, Y., … Abyzov, A. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*, *362*(6420), eaat8127. https://doi.org/10.1126/science.aat8127

Goertsen, D., Flytzanis, N. C., Goeden, N., Chuapoco, M. R., Cummins, A., Chen, Y., Fan, Y., Zhang, Q., Sharma, J., Duan, Y., Wang, L., Feng, G., Chen, Y., Ip, N. Y., Pickel, J., & Gradinaru, V. (2022). AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. *Nature Neuroscience*, *25*(1), 106–115. https://doi.org/10.1038/s41593-021-00969-4

Gompers, A. L., Su-Feher, L., Ellegood, J., Copping, N. A., Riyadh, M. A., Stradleigh, T. W., Pride, M. C., Schaffler, M. D., Wade, A. A., Catta-Preta, R., Zdilar, I., Louis, S., Kaushik, G., Mannion, B. J., Plajzer-Frick, I., Afzal, V., Visel, A., Pennacchio, L. A., Dickel, D. E., … Nord, A. S. (2017). Germline Chd8 haploinsufficiency alters brain development in mouse. *Nature Neuroscience*. https://doi.org/10.1038/nn.4592

Goodman, J. v, Yamada, T., Yang, Y., Kong, L., Wu, D. Y., Zhao, G., Gabel, H. W., & Bonni, A. (2020). The chromatin remodeling enzyme Chd4 regulates genome architecture in the mouse brain. *Nature Communications*, *11*(1), 3419. https://doi.org/10.1038/s41467-020-17065-z

Graybuck, L. T., Daigle, T. L., Sedeño-Cortés, A. E., Walker, M., Kalmbach, B., Lenz, G. H., Morin, E., Nguyen, T. N., Garren, E., Bendrick, J. L., Kim, T. K., Zhou, T., Mortrud, M., Yao, S., Siverts, L. A., Larsen, R., Gore, B. B., Szelenyi, E. R., Trader, C., … Tasic, B. (2021). Enhancer viruses for combinatorial cell-subclass-specific labeling. *Neuron*, *109*(9), 1449-1464.e13. https://doi.org/https://doi.org/10.1016/j.neuron.2021.03.011

Guilmatre, A., Huguet, G., Delorme, R., & Bourgeron, T. (2014). The emerging role of SHANK genes in neuropsychiatric disorders. *Developmental Neurobiology*, *74*(2), 113–122. https://doi.org/https://doi.org/10.1002/dneu.22128

Guo, J. U., Su, Y., Shin, J. H. J., Shin, J. H. J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G. L., & Song, H. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience*, *17*(2), 215–222. http://www.ncbi.nlm.nih.gov/pubmed/24362762

Guy, J., Gan, J., Selfridge, J., Cobb, S., & Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. *Science*, *315*(5815), 1143–1147. https://doi.org/10.1126/science.1138389

Guy, J., Hendrich, B., Holmes, M., Martin, J. E., & Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nature Genetics*, *27*(3), 322–326. https://doi.org/10.1038/85899

Hagberg, B., Witt-Engerström, I., Opitz, J. M., & Reynolds, J. F. (1986). Rett syndrome: a suggested staging system for describing impairment profile with increasing age towards adolescence. *American Journal of Medical Genetics*, *25*(S1), 47–59.

He, L., Liu, N., Cheng, T., Chen, X., Li, Y., Shu, Y., Qiu, Z., & Zhang, X. (2014). Conditional deletion of Mecp2 in parvalbumin-expressing GABAergic cells results in the absence of critical period plasticity. *Nature Communications*, *5*(1), 5036. https://doi.org/10.1038/ncomms6036

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/https://doi.org/10.1016/j.molcel.2010.05.004

Heinz, S., Romanoski, C. E., Benner, C., & Glass, C. K. (2015). The selection and function of cell type-specific enhancers. *Nature Reviews. Molecular Cell Biology*, *16*(3), 144–154. https://doi.org/10.1038/nrm3949

Hobert, O. (2008). Regulatory logic of neuronal diversity: Terminal selector genes and selector motifs. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(51), 20067–20071. https://doi.org/10.1073/PNAS.0806070105

Hobert, O. (2016). Terminal Selectors of Neuronal Identity. *Current Topics in Developmental Biology*, *116*, 455–475. https://doi.org/10.1016/BS.CTDB.2015.12.007

Ip, J. P. K., Mellios, N., & Sur, M. (2018). Rett syndrome: Insights into genetic, molecular and circuit mechanisms. *Nature Reviews Neuroscience*, *19*(6), 368–382. https://doi.org/10.1038/S41583-018-0006-3

Ito-Ishida, A., Ure, K., Chen, H., Swann, J. W., & Zoghbi, H. Y. (2015). Loss of MeCP2 in Parvalbumin-and Somatostatin-Expressing Neurons in Mice Leads to Distinct Rett Syndrome-like Phenotypes. *Neuron*, *88*(4), 651–658. https://doi.org/10.1016/j.neuron.2015.10.029

Jiang, M., Ash, R. T., Baker, S. A., Suter, B., Ferguson, A., Park, J., Rudy, J., Torsky, S. P., Chao, H.-T., Zoghbi, H. Y., & Smirnakis, S. M. (2013). Dendritic Arborization and Spine Dynamics Are Abnormal in the Mouse Model of MECP2 Duplication Syndrome. *The Journal of Neuroscience*, *33*(50), 19518 LP – 19533. https://doi.org/10.1523/JNEUROSCI.1745-13.2013

Johnson, B. S., Zhao, Y.-T. T., Fasolino, M., Lamonica, J. M., Kim, Y. J., Georgakilas, G., Wood, K. H., Bu, D., Cui, Y., Goffin, D., Vahedi, G., Kim, T. H., & Zhou, Z. (2017). Biotin tagging of MeCP2 in mice reveals contextual insights into the Rett syndrome transcriptome. *Nature Medicine*, *23*(10), 1203–1214. http://www.nature.com/articles/nm.4406

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., & Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, *429*(6994), 900–903. https://doi.org/10.1038/nature02633

Katayama, Y., Nishiyama, M., Shoji, H., Ohkawa, Y., Kawamura, A., Sato, T., Suyama, M., Takumi, T., Miyakawa, T., & Nakayama, K. I. (2016). CHD8 haploinsufficiency results in autistic-like phenotypes in mice. *Nature*. https://doi.org/10.1038/nature19357

Kepecs, A., & Fishell, G. (2014). Interneuron cell types are fit to function. *Nature*, *505*(7483), 318–326. https://doi.org/10.1038/nature12983

Kinde, B., Wu, D. Y., Greenberg, M. E., & Gabel, H. W. (2016). DNA methylation in the gene body influences MeCP2-mediated gene repression. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(52), 15114–15119. https://doi.org/10.1073/pnas.1618737114

King, I. F., Yandava, C. N., Mabb, A. M., Hsiao, J. S., Huang, H.-S., Pearson, B. L., Calabrese, J. M., Starmer, J., Parker, J. S., Magnuson, T., Chamberlain, S. J., Philpot, B. D., & Zylka, M. J. (2013). Topoisomerases facilitate transcription of long genes linked to autism. *Nature*, *501*(7465), 58–62. https://doi.org/10.1038/nature12504

King, M.-C., & Wilson, A. C. (1975). Evolution at Two Levels in Humans and Chimpanzees. *Science*, *188*(4184), 107–116. https://doi.org/10.1126/science.1090005

Kishi, N., & Macklis, J. D. (2004). MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Molecular and Cellular Neuroscience*, *27*(3), 306–321. https://doi.org/10.1016/j.mcn.2004.07.006

Kishi, N., & Macklis, J. D. (2010). MeCP2 functions largely cell-autonomously, but also non-cell-autonomously, in neuronal maturation and dendritic arborization of cortical pyramidal neurons. *Experimental Neurology*, *222*(1), 51–58. https://doi.org/10.1016/j.expneurol.2009.12.007

Klausberger, T., & Somogyi, P. (2008). Neuronal diversity and temporal dynamics: The unity of hippocampal circuit operations. In *Science* (Vol. 321, Issue 5885, pp. 53–57). https://doi.org/10.1126/science.1149381

Kokura, K., Kaul, S. C., Wadhwa, R., Nomura, T., Khan, M. M., Shinagawa, T., Yasukawa, T., Colmenares, C., & Ishii, S. (2001). The Ski Protein Family Is Required for MeCP2-mediated Transcriptional Repression*. *Journal of Biological Chemistry*, *276*(36), 34115–34121. https://doi.org/https://doi.org/10.1074/jbc.M105747200

Kriaucionis, S., & Heintz, N. (2009). The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science*, *324*(5929), 929–930. https://doi.org/10.1126/science.1169786

Krishnan, K., Wang, B.-S., Lu, J., Wang, L., Maffei, A., Cang, J., & Huang, Z. J. (2015). MeCP2 regulates the timing of critical period plasticity that shapes functional connectivity in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(34), E4782. https://doi.org/10.1073/PNAS.1506499112

la Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnerberg, P., Furlan, A., Fan, J., Borm, L. E., Liu, Z., van Bruggen, D., Guo, J., He, X., Barker, R., Sundström, E., Castelo-Branco, G., … Kharchenko, P. v. (2018). RNA velocity of single cells. *Nature*, *560*(7719), 494–498. https://doi.org/10.1038/s41586-018-0414-6

Lagger, S., Connelly, J. C., Schweikert, G., Webb, S., Selfridge, J., Ramsahoye, B. H., Yu, M., He, C., Sanguinetti, G., Sowers, L. C., Walkinshaw, M. D., & Bird, A. (2017). *MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain*. *13*(5), e1006793. https://dx.plos.org/10.1371/journal.pgen.1006793

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., … Trust:, T. W. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Lavery, L. A., Ure, K., Wan, Y.-W., Luo, C., Trostle, A. J., Wang, W., Jin, H., Lopez, J., Lucero, J., Durham, M. A., Castanon, R., Nery, J. R., Liu, Z., Goodell, M., Ecker, J. R., Behrens, M. M., & Zoghbi, H. Y. (2020). Losing Dnmt3a dependent methylation in inhibitory neurons impairs neural function by a mechanism impacting Rett syndrome. *ELife*, *9*, e52981. https://doi.org/10.7554/eLife.52981

Lavery, L. A., & Zoghbi, H. Y. (2019). The distinct methylation landscape of maturing neurons and its role in Rett syndrome pathogenesis. *Current Opinion in Neurobiology*, *59*, 180–188. https://doi.org/10.1016/J.CONB.2019.08.001

Lein, E., Borm, L. E., & Linnarsson, S. (2017). The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science*, *358*(6359), 64–69. https://doi.org/10.1126/SCIENCE.AAN6827

Leonard, H., Cobb, S., & Downs, J. (2017). Clinical and biological progress over 50 years in Rett syndrome. *Nature Reviews Neurology*, *13*(1), 37–51.

Li, W., Xu, X., & Pozzo-Miller, L. (2016). Excitatory synapses are stronger in the hippocampus of Rett syndrome mice due to altered synaptic trafficking of AMPA-type glutamate receptors. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(11), E1575-84. https://doi.org/10.1073/pnas.1517244113

Li, Y. E., Preissl, S., Hou, X., Zhang, Z., Zhang, K., Qiu, Y., Poirion, O. B., Li, B., Chiou, J., Liu, H., Pinto-Duarte, A., Kubo, N., Yang, X., Fang, R., Wang, X., Han, J. Y., Lucero, J., Yan, Y., Miller, M., … Ren, B. (2021). An atlas of gene regulatory elements in adult mouse cerebrum. *Nature*, *598*(7879), 129–136. https://doi.org/10.1038/s41586-021-03604-1

Lim, L., Mi, D., Llorca, A., & Marín, O. (2018). Development and Functional Diversification of Cortical Interneurons. *Neuron*, *100*(2), 294–313. https://doi.org/10.1016/J.NEURON.2018.10.009

Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., Lucero, J., Huang, Y., Dwork, A. J., Schultz, M. D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J. C., Rao, A., Esteller, M., He, C., Haghighi, F. G., … Ecker, J. R. (2013). Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science*, *341*(6146), 1237905–1237905. https://doi.org/10.1126/science.1237905

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences . *Nature*, *462*(7271), 315–322. https://doi.org/10.1038/nature08514

Liu, H., Zhou, J., Tian, W., Luo, C., Bartlett, A., Aldridge, A., Lucero, J., Osteen, J. K., Nery, J. R., Chen, H., Rivkin, A., Castanon, R. G., Clock, B., Li, Y. E., Hou, X., Poirion, O. B., Preissl, S., Pinto-Duarte, A., O'Connor, C., … Ecker, J. R. (2021). DNA methylation atlas of the mouse brain at single-cell resolution. *Nature*, *598*(7879), 120–128. https://doi.org/10.1038/s41586-020-03182-8

Liu, X., Ma, L., Liu, H., Gan, J., Xu, Y., Zhang, T., Mu, P., Wu, J., Shi, Y., Zhang, Y., Gong, L., & He, M. (2020). Cell-type-specific gene inactivation and in situ restoration via recombinase-based flipping of targeted genomic region. *Journal of Neuroscience*, *40*(37), 7169–7186. https://doi.org/10.1523/JNEUROSCI.1044-20.2020

Lombardi, L. M., Baker, S. A., & Zoghbi, H. Y. (2015). MECP2 disorders: from the clinic to mice and back. *The Journal of Clinical Investigation*, *125*(8), 2914–2923.

Lozovaya, N., Nardou, R., Tyzio, R., Chiesa, M., Pons-Bennaceur, A., Eftekhari, S., Bui, T.-T., Billon-Grand, M., Rasero, J., Bonifazi, P., Guimond, D., Gaiarsa, J.-L., Ferrari, D. C., & Ben-Ari, Y. (2019). Early alterations in a mouse model of Rett syndrome: the GABA developmental shift is abolished at birth. *Scientific Reports*, *9*(1), 9276. https://doi.org/10.1038/s41598-019-45635-9

Luikenhuis, S., Giacometti, E., Beard, C. F., & Jaenisch, R. (2004). Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. *Proceedings of the National Academy of Sciences*, *101*(16), 6033–6038. https://doi.org/10.1073/pnas.0401626101

Luo, C., Keown, C. L., Kurihara, L., Zhou, J., He, Y., Li, J., Castanon, R., Lucero, J., Nery, J. R., Sandoval, J. P., Bui, B., Sejnowski, T. J., Harkins, T. T., Mukamel, E. A., Behrens, M. M., & Ecker, J. R. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science*, *357*(6351), 600 LP – 604.

Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D., & Woolf, P. J. (2009). GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*, *10*(1), 1–17.

Lyst, M. J., & Bird, A. (2015). Rett syndrome: a complex disorder with simple roots. *Nature Reviews Genetics*, *16*(5), 261–275. https://doi.org/10.1038/nrg3897

Lyst, M. J., Ekiert, R., Ebert, D. H., Merusi, C., Nowak, J., Selfridge, J., Guy, J., Kastan, N. R., Robinson, N. D., de Lima Alves, F., Rappsilber, J., Greenberg, M. E., & Bird, A. (2013). Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nature Neuroscience*, *16*(7), 898–902. https://doi.org/10.1038/nn.3434

Mabb, A. M., Kullmann, P. H. M., Twomey, M. A., Miriyala, J., Philpot, B. D., & Zylka, M. J. (2014). Topoisomerase 1 inhibition reversibly impairs synaptic function. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1413204111

MacNeil, M. A., & Masland, R. H. (1998). Extreme Diversity among Amacrine Cells: Implications for Function. *Neuron*, *20*(5), 971–982. https://doi.org/10.1016/S0896-6273(00)80478-X

Mardinly, A. R., Spiegel, I., Patrizi, A., Centofante, E., Bazinet, J. E., Tzeng, C. P., Mandel-Brehm, C., Harmin, D. A., Adesnik, H., Fagiolini, M., & Greenberg, M. E. (2016). Sensory experience regulates cortical inhibition by inducing IGF1 in VIP neurons. *Nature*, *531*(7594), 371–375. https://doi.org/10.1038/nature17187

Marques-Smith, A., Lyngholm, D., Kaufmann, A.-K., Stacey, J. A., Hoerder-Suabedissen, A., Becker, E. B. E., Wilson, M. C., Molnár, Z., & Butt, S. J. B. (2016). A Transient Translaminar GABAergic Interneuron Circuit Connects Thalamocortical Recipient Layers in Neonatal Somatosensory Cortex. *Neuron*, *89*(3), 536–549. https://doi.org/10.1016/j.neuron.2016.01.015

Mayer, C., Hafemeister, C., Bandler, R. C., Machold, R., Batista Brito, R., Jaglin, X., Allaway, K., Butler, A., Fishell, G., & Satija, R. (2018). Developmental diversification of cortical inhibitory interneurons. *Nature*, *555*(7697), 457–462. https://doi.org/10.1038/nature25999

McCoy, M. J., Paul, A. J., Victor, M. B., Richner, M., Gabel, H. W., Gong, H., Yoo, A. S., & Ahn, T.-H. (2018). LONGO: an R package for interactive gene length dependent analysis for neuronal identity. *Bioinformatics*, *34*(13), i422–i428. https://doi.org/10.1093/bioinformatics/bty243

McGraw, C. M., Samaco, R. C., & Zoghbi, H. Y. (2011). Adult neural function requires MeCP2. *Science*, *333*(6039), 186.

Meehan, R., Lewis, J. D., & Bird, A. P. (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Research*, *20*(19), 5085–5092. https://doi.org/10.1093/nar/20.19.5085

Mellén, M., Ayata, P., & Heintz, N. (2017). 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. *Proceedings of the National Academy of Sciences*, *114*(37), E7812–E7821. https://doi.org/10.1073/pnas.1708044114

Meng, X., Wang, W., Lu, H., He, L., Chen, W., Chao, E. S., Fiorotto, M. L., Tang, B., Herrera, J. A., Seymour, M. L., Neul, J. L., Pereira, F. A., Tang, J., Xue, M., & Zoghbi, H. Y. (2016). Manipulations of MeCP2 in glutamatergic neurons highlight their contributions to Rett and other neurological disorders. *ELife*, *5*, e14199. https://doi.org/10.7554/eLife.14199

Mi, D., Li, Z., Lim, L., Li, M., Moissidis, M., Yang, Y., Gao, T., Hu, T. X., Pratt, T., Price, D. J., Sestan, N., & Marín, O. (2018). Early emergence of cortical interneuron diversity in the mouse embryo. *Science*, *360*(6384), 81–85. https://doi.org/10.1126/science.aar6821

Miyoshi, G. (2019). Elucidating the developmental trajectories of GABAergic cortical interneuron subtypes. *Neuroscience Research*, *138*, 26–32.

Mo, A., Mukamel, E. A., Davis, F. P., Luo, C., Henry, G. L., Picard, S., Urich, M. A., Nery, J. R., Sejnowski, T. J., Lister, R., Eddy, S. R., Ecker, J. R., & Nathans, J. (2015). Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron*, *86*(6), 1369–1384. https://doi.org/10.1016/j.neuron.2015.05.018

Molyneaux, B. J., Arlotta, P., Menezes, J. R. L., & Macklis, J. D. (2007). Neuronal subtype specification in the cerebral cortex. *Nature Reviews Neuroscience*, *8*(6), 427–437. https://doi.org/10.1038/nrn2151

Mossner, J. M., Batista-Brito, R., Pant, R., & Cardin, J. A. (2020). Developmental loss of MeCP2 from VIP interneurons impairs cortical function and behavior. *ELife*, *9*, e55639. https://doi.org/10.7554/eLife.55639

Mukamel, E. A., & Ngai, J. (2019). Perspectives on defining cell types in the brain. *Current Opinion in Neurobiology*, *56*, 61–68. https://doi.org/10.1016/J.CONB.2018.11.007

Nery, S., Fishell, G., & Corbin, J. G. (2002). The caudal ganglionic eminence is a source of distinct cortical and subcortical cell populations. *Nature Neuroscience*, *5*(12), 1279–1287. https://doi.org/10.1038/nn971

Nguyen, S., Meletis, K., Fu, D., Jhaveri, S., & Jaenisch, R. (2007). Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, *236*(6), 1663–1676. https://doi.org/10.1002/dvdy.21176

Nord, A. S., Blow, M. J., Attanasio, C., Akiyama, J. A., Holt, A., Hosseini, R., Phouanenavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V., Rubenstein, J. L. R., Rubin, E. M., Pennacchio, L. A., & Visel, A. (2013). Rapid and Pervasive Changes in Genome-wide Enhancer Usage during Mammalian Development. *Cell*, *155*(7), 1521–1531. https://doi.org/10.1016/J.CELL.2013.11.033

Nord, A. S., & West, A. E. (2020). Neurobiological functions of transcriptional enhancers. *Nature Neuroscience*, *23*(1), 5–14. https://doi.org/10.1038/s41593-019-0538-5

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3), 247–257.

Olah, V. J., Goettemoeller, A. M., Rayaprolu, S., Dammer, E. B., Seyfried, N. T., Rangaraju, S., Dimidschstein, J., & Rowan, M. J. M. (2022). Biophysical K(v)3 channel alterations dampen excitability of cortical PV interneurons and contribute to network hyperexcitability in early Alzheimer's. *ELife*, *11*, e75316. https://doi.org/10.7554/eLife.75316

Orefice, L. L., Zimmerman, A. L., Chirila, A. M., Sleboda, S. J., Head, J. P., & Ginty, D. D. (2016). Peripheral Mechanosensory Neuron Dysfunction Underlies Tactile and Behavioral Deficits in Mouse Models of ASDs. *Cell*, *166*(2), 299–313. https://doi.org/https://doi.org/10.1016/j.cell.2016.05.033

Otsu, Y., Donneger, F., Schwartz, E. J., & Poncer, J. C. (2020). *Cation-chloride cotransporters and the polarity of GABA signaling 2 in mouse hippocampal parvalbumin interneurons Running title: GABA signaling in hippocampal parvalbumin interneurons*.

Panigrahi, A., & O'Malley, B. W. (2021). Mechanisms of enhancer action: the known and the unknown. *Genome Biology*, *22*(1), 108. https://doi.org/10.1186/s13059-021-02322-1

Parikshak, N. N., Luo, R., Zhang, A., Won, H., Lowe, J. K., Chandran, V., Horvath, S., & Geschwind, D. H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*, *155*(5), 1008–1021.

Paul, A., Crow, M., Raudales, R., He, M., Gillis, J., & Huang, Z. J. (2017). Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity. *Cell*. https://doi.org/10.1016/j.cell.2017.08.032

Pelkey, K. A., Chittajallu, R., Craig, M. T., Tricoire, L., Wester, J. C., & McBain, C. J. (2017). Hippocampal GABAergic Inhibitory Interneurons. *Physiological Reviews*, *97*(4), 1619–1747. https://doi.org/10.1152/physrev.00007.2017

Picard, N., & Fagiolini, M. (2019). MeCP2: an epigenetic regulator of critical periods. *Current Opinion in Neurobiology*, *59*, 95–101. https://doi.org/10.1016/j.conb.2019.04.004

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842.

Ramón y Cajal, S. (1909). Histologie du Systeme Nerveux de l'Homme et des Vertebres. Maloine, Paris: 1911. chap. II. In *Demography*.

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, *159*(7), 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Renthal, W., Boxer, L. D., Hrvatin, S., Li, E., Silberfeld, A., Nagy, M. A., Griffith, E. C., Vierbuchen, T., & Greenberg, M. E. (2018). Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing. *Nature Neuroscience*, *21*(12), 1670–1679. http://www.ncbi.nlm.nih.gov/pubmed/30455458

Riba, A., Oravecz, A., Durik, M., Jiménez, S., Alunni, V., Cerciat, M., Jung, M., Keime, C., Keyes, W. M., & Molina, N. (2022). Cell cycle gene regulation dynamics revealed by RNA velocity and deep-learning. *Nature Communications*, *13*(1), 2865. https://doi.org/10.1038/s41467-022-30545-8

Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, *32*(9), 896–902. https://doi.org/10.1038/nbt.2931

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140.

Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., Murtha, M. T., Bal, V. H., Bishop, S. L., Dong, S., Goldberg, A. P., Jinlu, C., Keaney, J. F., Klei, L., Mandell, J. D., Moreno-De-Luca, D., Poultney, C. S., Robinson, E. B., Smith, L., … State,

M. W. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*, *87*(6), 1215–1233. https://doi.org/10.1016/j.neuron.2015.09.016

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., de Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., & Klei, L. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, *180*(3), 568–584.

Sceniak, M. P., Lang, M., Enomoto, A. C., Howell, C. J., Hermes, D. J., & Katz, D. M. (2016). Mechanisms of Functional Hypoconnectivity in the Medial Prefrontal Cortex of Mecp2 Null Mice. *Cerebral Cortex (New York, NY)*, *26*(5), 1938. https://doi.org/10.1093/CERCOR/BHV002

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature*, *517*(7534), 321–326. https://doi.org/10.1038/nature14192

Sessa, A., Fagnocchi, L., Mastrototaro, G., Massimino, L., Zaghi, M., Indrigo, M., Cattaneo, S., Martini, D., Gabellini, C., & Pucci, C. (2019). SETD5 regulates chromatin methylation state and preserves global transcriptional fidelity during brain development and neuronal wiring. *Neuron*, *104*(2), 271–289.

Shahbazian, M. D. (2002). Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. *Human Molecular Genetics*. https://doi.org/10.1093/hmg/11.2.115

Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews Genetics*, *15*(4), 272–286. https://doi.org/10.1038/nrg3682

Skene, P. J., Illingworth, R. S., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Andrews, R., & Bird, A. P. (2010). Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. *Molecular Cell*, *37*(4), 457–468. https://doi.org/10.1016/j.molcel.2010.01.030

Stroud, H., Su, S. C., Hrvatin, S., Greben, A. W., Renthal, W., Boxer, L. D., Nagy, M. A., Hochbaum, D. R., Kinde, B., Gabel, H. W., & Greenberg, M. E. (2017). Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic States. *Cell*, *171*(5), 1151-1164.e16. https://doi.org/10.1016/J.CELL.2017.09.047

Stroud, H., Yang, M. G., Tsitohay, Y. N., Davis, C. P., Sherman, M. A., Hrvatin, S., Ling, E., & Greenberg, M. E. (2020). An Activity-Mediated Transition in Transcription in Early Postnatal Neurons. *Neuron*, *107*(5), 874-890.e8. https://doi.org/10.1016/j.neuron.2020.06.008

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., & Lander, E. S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550.

Sugino, K., Hempel, C. M., Okaty, B. W., Arnson, H. A., Kato, S., Dani, V. S., & Nelson, S. B. (2014). Cell-Type-Specific Repression by Methyl-CpG-Binding Protein 2 Is Biased toward Long Genes. *Journal of Neuroscience*, *34*(38), 12877–12883. https://doi.org/10.1523/JNEUROSCI.2674-14.2014

Sullivan, J. M., de Rubeis, S., & Schaefer, A. (2019). Convergence of spectrums: neuronal gene network states in autism spectrum disorder. *Current Opinion in Neurobiology*, *59*, 102–111. https://doi.org/https://doi.org/10.1016/j.conb.2019.04.011

Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., Levi, B., Gray, L. T., Sorensen, S. A., Dolbeare, T., Bertagnolli, D., Goldy, J., Shapovalova, N., Parry, S., Lee, C., Smith, K., Bernard, A., Madisen, L., Sunkin, S. M., … Zeng, H. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience*, *19*(2), 335–346. https://doi.org/10.1038/nn.4216

Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M. N., Viswanathan, S., Penn, O., Bakken, T., Menon, V., Miller, J., Fong, O., Hirokawa, K. E., Lathia, K., Rimorin, C., Tieu, M., … Zeng, H. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, *563*(7729), 72–78. https://doi.org/10.1038/s41586-018-0654-5

Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Aksglaede, L., Baralle, D., Dabir, T., Gener, B., Goudie, D., Homfray, T., Kumar, A., Pilz, D. T., Selicorni, A., Temple, I. K., … Rahman, N. (2014). Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nature Genetics*, *46*(4), 385–388. https://doi.org/10.1038/NG.2917

Tatton-Brown, K., Zachariou, A., Loveday, C., Renwick, A., Mahamdallie, S., Aksglaede, L., Baralle, D., Barge-Schaapveld, D., Blyth, M., Bouma, M., Breckpot, J., Crabb, B., Dabir, T., Cormier-Daire, V., Fauth, C., Fisher, R., Gener, B., Goudie, D., Homfray, T., … Rahman, N. (2018a). The Tatton-Brown-Rahman Syndrome: A clinical study of 55 individuals with de novo constitutive DNMT3A variants. *Wellcome Open Research*, *3*, 46. http://www.ncbi.nlm.nih.gov/pubmed/29900417

Tillotson, R., Cholewa-Waclaw, J., Chhatbar, K., Connelly, J. C., Kirschner, S. A., Webb, S., Koerner, M. v., Selfridge, J., Kelly, D. A., de Sousa, D., Brown, K., Lyst, M. J., Kriaucionis, S., & Bird, A. (2021). Neuronal non-CG methylation is an essential target for MeCP2 function. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2021.01.011

Tilot, A. K., Bebek, G., Niazi, F., Altemus, J. B., Romigh, T., Frazier, T. W., & Eng, C. (2016). Neural transcriptome of constitutional Pten dysfunction in mice and its relevance to human idiopathic autism spectrum disorder. *Molecular Psychiatry*, *21*(1), 118–125.

Tropea, D., Giacometti, E., Wilson, N. R., Beard, C., McCurry, C., Fu, D. D., Flannery, R., Jaenisch, R., & Sur, M. (2009). Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice. *Proceedings of the National Academy of Sciences*, *106*(6), 2029–2034. https://doi.org/10.1073/pnas.0812394106

Tudor, M., Akbarian, S., Chen, R. Z., & Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proceedings of the National Academy of Sciences*, *99*(24), 15536–15541.

Tuncdemir, S. N., Wamsley, B., Stam, F. J., Osakada, F., Goulding, M., Callaway, E. M., Rudy, B., & Fishell, G. (2016). Early Somatostatin Interneuron Connectivity Mediates the Maturation of Deep Layer Cortical Circuits. *Neuron*, *89*(3), 521–535. https://doi.org/10.1016/J.NEURON.2015.11.020

Tyssowski, K. M., & Gray, J. M. (2019). The neuronal stimulation–transcription coupling map. *Current Opinion in Neurobiology*, *59*, 87–94. https://doi.org/https://doi.org/10.1016/j.conb.2019.05.001

Ure, K., Lu, H., Wang, W., Ito-Ishida, A., Wu, Z., He, L., Sztainberg, Y., Chen, W., Tang, J., & Zoghbi, H. Y. (2016). Restoration of Mecp2 expression in GABAergic neurons is sufficient to rescue multiple disease features in a mouse model of Rett syndrome. *ELife*, *5*, e14198. https://doi.org/10.7554/eLife.14198

van Esch, H., Bauters, M., Ignatius, J., Jansen, M., Raynaud, M., Hollanders, K., Lugtenberg, D., Bienvenu, T., Jensen, L. R., Gécz, J., Moraine, C., Marynen, P., Fryns, J. P., & Froyen, G. (2005). Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *American Journal of Human Genetics*. https://doi.org/10.1086/444549

Visel, A., Taher, L., Girgis, H., May, D., Golonzhka, O., Hoch, R. v., McKinsey, G. L., Pattabiraman, K., Silberberg, S. N., Blow, M. J., Hansen, D. v., Nord, A. S., Akiyama, J. A., Holt, A., Hosseini, R., Phouanenavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V., … Rubenstein, J. L. R. (2013). A High-Resolution Enhancer Atlas of the Developing Telencephalon. *Cell*, *152*(4), 895–908. https://doi.org/10.1016/J.CELL.2012.12.041

Vlasits, A. L., Euler, T., & Franke, K. (2019). Function first: classifying cell types and circuits of the retina. *Current Opinion in Neurobiology*, *56*, 8–15. https://doi.org/10.1016/j.conb.2018.10.011

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., Mill, J., Cantor, R. M., Blencowe, B. J., & Geschwind, D. H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, *474*(7351), 380–384.

Vormstein-Schneider, D., Lin, J. D., Pelkey, K. A., Chittajallu, R., Guo, B., Arias-Garcia, M. A., Allaway, K., Sakopoulos, S., Schneider, G., Stevenson, O., Vergara, J., Sharma, J., Zhang, Q., Franken, T. P., Smith, J., Ibrahim, L. A., M astro, K. J., Sabri, E., Huang, S., … Dimidschstein, J. (2020). Viral manipulation of functionally distinct interneurons in mice, non-human primates and humans. *Nature Neuroscience*. https://doi.org/10.1038/s41593-020-0692-9

Wamsley, B., & Fishell, G. (2017). Genetic and activity-dependent mechanisms underlying interneuron diversity. *Nature Reviews Neuroscience*, *18*(5), 299–309. https://doi.org/10.1038/nrn.2017.30

Wiesel, T. N., & Hubel, D. H. (1963). EFFECTS OF VISUAL DEPRIVATION ON MORPHOLOGY AND PHYSIOLOGY OF CELLS IN THE CAT'S LATERAL GENICULATE BODY. *Journal of Neurophysiology*, *26*(6), 978–993. https://doi.org/10.1152/jn.1963.26.6.978

Wonders, C. P., & Anderson, S. A. (2006). The origin and specification of cortical interneurons. *Nature Reviews Neuroscience*, *7*(9), 687–696.

Xie, W., Barr, C. L., Kim, A., Yue, F., Lee, A. Y., Eubanks, J., Dempster, E. L., & Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell*, *148*(4), 816–831.

Yao, Z., Liu, H., Xie, F., Fischer, S., Adkins, R. S., Aldridge, A. I., Ament, S. A., Bartlett, A., Behrens, M. M., van den Berge, K., Bertagnolli, D., de Bézieux, H. R., Biancalani, T., Booeshaghi, A. S., Bravo, H. C., Casper, T., Colantuoni, C., Crabtree, J., Creasy, H., … Mukamel, E. A. (2021). A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature*, *598*(7879), 103–110. https://doi.org/10.1038/s41586-021-03500-8

Zeng, H., & Sanes, J. R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Publishing Group*, *18*. https://doi.org/10.1038/nrn.2017.85

Zhang, M., Eichhorn, S. W., Zingg, B., Yao, Z., Cotter, K., Zeng, H., Dong, H., & Zhuang, X. (2021). Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature*, *598*(7879), 137–143. https://doi.org/10.1038/s41586-021-03705-x

Zhang, R. S., Liakath-Ali, K., & Südhof, T. C. (2020). Latrophilin-2 and latrophilin-3 are redundantly essential for parallel-fiber synapse function in cerebellum. *Elife*, *9*, e54443.

Zhang, X., Chen, M. H., Wu, X., Kodani, A., Fan, J., Doan, R., Ozawa, M., Ma, J., Yoshida, N., Reiter, J. F., Black, D. L., Kharchenko, P. v., Sharp, P. A., & Walsh, C. A. (2016). Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell*, *166*(5), 1147-1162.e15. https://doi.org/10.1016/j.cell.2016.07.025

Zhang, Z., Zhou, J., Tan, P., Pang, Y., Rivkin, A. C., Kirchgessner, M. A., Williams, E., Lee, C.-T., Liu, H., Franklin, A. D., Miyazaki, P. A., Bartlett, A., Aldridge, A. I., Vu, M., Boggeman, L., Fitzpatrick, C., Nery, J. R., Castanon, R. G., Rashid, M., … Callaway, E. M. (2021). Epigenomic diversity of cortical projection neurons in the mouse brain. *Nature*, *598*(7879), 167–173. https://doi.org/10.1038/s41586-021-03223-w

Zhao, Y.-T., Goffin, D., Johnson, B. S., & Zhou, Z. (2013). Loss of MeCP2 function is associated with distinct gene expression changes in the striatum. *Neurobiology of Disease*, *59*, 257–266. https://doi.org/10.1016/j.nbd.2013.08.001

Zylka, M. J., Simon, J. M., & Philpot, B. D. (2015). Gene Length Matters in Neurons. *Neuron*, *86*(2), 353–355. https://doi.org/10.1016/j.neuron.2015.03.059