2017

Epigenomic And Nuclear Architectural Insights Into Rett Syndrome

Maria Danielle Fasolino
University of Pennsylvania, fasolino@mail.med.upenn.edu

Epigenomic And Nuclear Architectural Insights Into Rett Syndrome

Abstract
The importance of DNA methylation in neuronal function is highlighted by mutations in the neuronally enriched "reader" of DNA methylation, methyl-CpG-binding protein 2 (MECP2), causing Rett Syndrome (RTT), a severe neurodevelopmental disorder. Although MeCP2 displays broad genomic binding, gene expression changes in Mecp2 mutant mice are very subtle, and brain region-specific, making it difficult to determine how MeCP2 regulates gene expression. Therefore, we developed an approach to assess cell type-specific effects of Mecp2 mutations on the transcriptome, epigenome, and chromatin architecture to determine whether epigenomic features can explain gene misregulation in RTT. Differentially expressed genes (DEGs) in R106W Mecp2 mutants (R106W) are enriched for MeCP2 binding in the WT setting and are preferentially demethylated in R106W, suggesting that the loss of MeCP2 binding results in the exposure of unbound cytosines to demethylation, thus contributing to gene dysregulation. Given that DEGs are enriched for MeCP2 binding, we next determined unique features of DEGs to gain an understanding of why MeCP2 preferentially targets DEGs. We find that DEGs are cell type-specific, lowly expressed, and intragenically associated with heterochromatin, active enhancer, and CTCF chromatin states, suggesting that MeCP2 is essential for the regulation of lowly expressed genes. Upregulated and downregulated DEGs are differentially enriched for particular chromatin states, providing an insight into the directionality of gene dysregulation. Given the enrichment of DEGs for active enhancer and CTCF chromatin states, we next investigated transcription factor (TF) footprints and found thousands of altered TF footprints in R106W, with the CTCF motif being the most significantly associated. In WT, these sites are enriched for MeCP2 binding, and in R106W, these sites, which are associated with downregulated DEGs, become demethylated, enabling CTCF binding. This therefore suggests that MeCP2 can affect CTCF recruitment to chromatin. Given CTCF's known role in chromatin organization, we employed Oligopaint and found large-scale condensation of euchromatin and heterochromatin, as well as decondensation of long genes. Together, this work provides insight into why DEGs are differentially susceptible to dysregulation in RTT and posits MeCP2 as a key player in global maintenance of the methylome and chromatin architecture for the preservation of neuronal gene expression.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Neuroscience

First Advisor
Zhaolan Zhou

Keywords
DNA Methylation, Epigenome, methyl-CpG-binding protein 2 (MeCP2), Nuclear Architecture, Rett Syndrome, Transcriptome

EPIGENOMIC AND NUCLEAR ARCHITECTURAL INSIGHTS INTO RETT SYNDROME

Maria Fasolino

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

_____

Dr. Zhaolan Zhou

Associate Professor of Genetics

Graduate Group Chairperson

_____

Dr. Joshua I. Gold

Professor of Neuroscience

Dissertation Committee:

Dr. Klaus Kaestner, Thomas and Evelyn Suor Butterworth Professor in Genetics
Dr. Edward Lee, Assistant Professor of Pathology and Laboratory Medicine
Dr. Virginia Lee, John H. Ware 3rd Endowed Professor in Alzheimer's Research
Dr. Golnaz Vahedi, Assistant Professor of Genetics

ABSTRACT

EPIGENOMIC AND NUCLEAR ARCHITECTURAL INSIGHTS INTO RETT SYNDROME

Maria Fasolino

Dr. Zhaolan Zhou

The importance of DNA methylation in neuronal function is highlighted by mutations in the neuronally enriched "reader" of DNA methylation, methyl-CpG-binding protein 2 (*MECP2*), causing Rett Syndrome (RTT), a severe neurodevelopmental disorder. Although MeCP2 displays broad genomic binding, gene expression changes in *Mecp2* mutant mice are very subtle, and brain region-specific, making it difficult to determine how MeCP2 regulates gene expression. Therefore, we developed an approach to assess cell type-specific effects of *Mecp2* mutations on the transcriptome, epigenome, and chromatin architecture to determine whether epigenomic features can explain gene misregulation in RTT. Differentially expressed genes (DEGs) in R106W *Mecp2* mutants (R106W) are enriched for MeCP2 binding in the WT setting and are preferentially demethylated in R106W, suggesting that the loss of MeCP2 binding results in the exposure of unbound cytosines to demethylation, thus contributing to gene dysregulation. Given that DEGs are enriched for MeCP2 binding, we next determined unique features of DEGs to gain an understanding of why MeCP2 preferentially targets DEGs. We find that DEGs are cell type-specific, lowly expressed, and intragenically associated with heterochromatin, active enhancer, and CTCF chromatin states,

suggesting that MeCP2 is essential for the regulation of lowly expressed genes. Upregulated and downregulated DEGs are differentially enriched for particular chromatin states, providing an insight into the directionality of gene dysregulation. Given the enrichment of DEGs for active enhancer and CTCF chromatin states, we next investigated transcription factor (TF) footprints and found thousands of altered TF footprints in R106W, with the CTCF motif being the most significantly associated. In WT, these sites are enriched for MeCP2 binding, and in R106W, these sites, which are associated with downregulated DEGs, become demethylated, enabling CTCF binding. This therefore suggests that MeCP2 can affect CTCF recruitment to chromatin. Given CTCF's known role in chromatin organization, we employed Oligopaint and found large-scale condensation of euchromatin and heterochromatin, as well as decondensation of long genes. Together, this work provides insight into why DEGs are differentially susceptible to dysregulation in RTT and posits MeCP2 as a key player in global maintenance of the methylome and chromatin architecture for the preservation of neuronal gene expression.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 2

Chapter 3

Appendix 1

Appendix 2

Appendix 3

CHAPTER 1

Introduction

One Sentence Summary

This chapter provides an introduction to the concepts presented in subsequent chapters, with a focus on the outstanding questions in MeCP2 research, and has been accepted for publication as a review article in the journal *Genes*.

Overview

A neuron is unique in its ability to dynamically modify its transcriptional output in response to synaptic activity while maintaining a core gene expression program that preserves cellular identity throughout a lifetime that is longer than almost every other cell type in the body. A contributing factor to the immense adaptability of a neuron is attributed to its unique epigenetic landscape that elicits rapid, locus-specific alterations in chromatin architecture, which in turn influences gene expression. One such epigenetic modification that is sensitive to changes in synaptic activity, as well as essential for maintaining cellular identity, is DNA methylation. The focus of this chapter is on the importance of DNA methylation in neuronal function, summarizing the recent studies on critical players in the establishment of (the "writing"), the modifying or erasing of (the "editing"), and the mediation of (the "reading") DNA methylation in neuroplasticity. One "reader" of DNA methylation in particular, methyl-CpG-binding protein 2 (MeCP2), is highlighted given that understanding the molecular function of this protein, which has undisputed importance in neuronal function, has been the main focus of my graduate work.

The adaptability of gene expression programs in response to neuronal activity

Neuroplasticity is the ability of the brain to alter neural function in response to environmental input. This adaptability, also expressed as long-lasting plasticity, is the basis of learning and memory formation. Two neuronal mechanisms that are essential for long-lasting plasticity are long-term depression (LTD) and long-term potentiation (LTP), which decrease or increase synaptic strength, respectively, in response to activity (Malenka and Bear, 2004; McClung and Nestler, 2007). Here we will focus on LTP, as key molecules in this process are frequently referenced throughout this review.

LTP is generally broken down into two phases: a transient "early" phase (30-60 minutes) that is dependent upon the activation of pre-existing proteins and a stable "late" phase (hours, days, or even weeks) that is contingent upon new gene transcription and translation (Malenka and Bear, 2004). The induction of early phase LTP is dependent upon postsynaptic depolarization that leads to an influx of calcium (Ca^{2+}) through NMDA receptors (NMDARs). Ca^{2+} entry in turn triggers a biochemical cascade starting with the activation of calcium/calmodulin-dependent protein kinase II (CaMKII) via phosphorylation, which translocates to active synapses where it binds to NMDARs and phosphorylates AMPA receptors (AMPAR) and activates RAS-extracellular signal-regulated kinase (ERK) pathway that leads to an increase in the insertion of AMPAR into the synapse. This leads to an increase in the conductance of Ca^{2+} through AMPARs in stimulated spines, which subsequently activates a molecular cascade that strengthens potentiated spines (Lisman et al., 2012; Malenka and Bear, 2004). An important modulator of the CaMKII cascade is the protein phosphatase 1 (PP1), which at basal conditions dephosphorylates CaMKII; however, in response to neuronal activity and subsequent LTP induction, cyclic adenosine monophosphate (cAMP) activates protein kinase A, which in turn phosphorylates inhibitor 1 (I-1) that inhibits PP1, leading to an increase in CaMKII phosphorylation and activity (Blitzer, 1998; Brown et al., 2000; Lisman et al., 2012; Malenka and Bear, 2004).

Similarly to the case of CaMKII, many of the early phase LTP signaling cascades converge upon the activation of ERK, which in turn activates various biochemical pathways that are essential for late phase LTP. One group of proteins activated by ERK are transcription factors, such as cAMP response element binding protein (CREB), FBJ

osteosarcoma oncogene B (*Fosb*), and the transcription factor complex NF-κB, which subsequently induce gene expression changes essential for the potentiation of active synapses, such as immediate early response genes, ion channels, structural proteins, and neurotrophins (LYNCH, 2004; McClung and Nestler, 2007). One neurotrophin that is particularly important in the maintenance of LTP is brain-derived neurotrophic factor (*Bdnf*), which acts through mitogen-activated protein kinase (MAPK)/ERK kinase (MEK) to increase the expression of activity-regulated cytoskeleton-associated protein (*Arc*), a gene important in cytoskeleton organization at activated synapses (LYNCH, 2004; Ying et al., 2002). It is during this late phase of LTP where DNA methylation in thought to contribute to the regulation of the expression of key genes necessary for the maintenance of long-term neuronal plasticity, and in this review, we will summarize the substantial body of work supporting this notion. We will begin with an overview of the important forms of DNA methylation in the mammalian brain.

Types of DNA methylation in neurons

5mCG: a repressive epigenetic mark

DNA methylation is an epigenetic mechanism that allows for sustained adaptability of gene expression in response to developmental or environmental factors; it plays an essential role in various biological functions such as regulation of gene transcription and the establishment and maintenance of cellular identity (Golshani et al., 2007; Hutnick et al., 2009; Jaenisch and Bird, 2003). DNA methylation at the 5-carbon of cytosine (5mC) is widely distributed throughout the mammalian genome, with ~5% of cytosines being methylated in the adult mouse brain. In most mammalian cell types, methylation predominately occurs at cytosine-phosphate-guanine dinucleotides (CpG), with methylation occurring at 60-90% of CpGs depending on tissue type, and in the mammalian brain, ~62% of CpGs are methylated (methylated CpG will subsequently referred to as 5mCG) (Lister et al., 2013; Schultz et al., 2015; Varley et al., 2013; Xie et al., 2012). DNA methylation is deposited on cytosine by family of enzymes known as the DNA methyltransferases (DNMTs). This enzyme family is divided into two broad classes: the *de novo* DNMTs, DNMT3A and DNMT3B, that establish methylation on

unmethylated DNA and methylate the complementary strand of DNA strand on hemimethylated DNA, and the maintenance DNMT, DNMT1, which can only methylate hemimethylated DNA (Klose and Bird, 2006). Whereas *Dnmt3a* and *Dnmt1* are expressed in both the embryonic and adult stages, *Dnmt3b* is detectable only during early neurogenesis (Brooks et al., 1996; Feng et al., 2005), suggesting that particular DNMTs play a role in neuronal function at specific times over neuronal development and maturation.

DNA methylation at CGs in all somatic types, including the neurons, represses the expression of genes and repetitive DNA regions given whole genome bisulfite sequencing (WGBS) findings that 5mCG is enriched at intergenic regions, silenced genes, and repetitive DNA regions and is depleted at enhancers, promoters, and actively expressed gene bodies (Lister et al., 2013; 2009; Xie et al., 2012; Ziller et al., 2013)**.** Additionally, 5'-upstream, intragenic, and 3'-downstream 5mCG inversely correlates with gene expression in neurons (Mo et al., 2015). Furthermore, 5mCG patterns are neuronal type-specific, suggesting that this epigenetic mark regulates gene expression in a cell type-specific manner (Lister et al., 2013; Mo et al., 2015). This epigenetic mark is also important in nuclear organization and compartmentalization given that DNA methylation of satellite repeats is an important step in the formation of heterochromatic regions (Cedar and Bergman, 2009; Lehnertz et al., 2003).

Additional neuronally enriched forms of DNA methylation

Historically, DNA methylation was thought to be a stable, repressive covalent modification, existing predominately, if not exclusively, as 5mCG. However, this view has dramatically changed over the past few years with the discovery that neurons, unlike other adult somatic cells, accumulate two other forms of DNA methylation over development, 5-hydroxymethylcystosine (5hmC) and non-CG methylation (5mCH, where H = A, T, or C). 5hmC levels in neurons are up to 10 fold higher than other cell types, reaching ~17% of all DNA methylation in the mouse frontal cortex (Globisch et al., 2010; Kriaucionis and Heintz, 2009; Münzel et al., 2010; Song et al., 2011; Tahiliani et al., 2009). Even though 5mCH is nearly absent in other cell types, except embryonic stem cells (ESCs), 5mCH accounts for up to 53% of all of DNA methylation in adult

human neurons, making it the predominate form of DNA methylation, and ~25-40% of all DNA methylation in the adult mouse neurons. Although the overall percentage of all CH dinucleotides that are methylated is relatively low (1.4% of CH dinucleotides in adult mouse neurons) compared to those methylated in CG dinucleotides (62% of CG dinucleotides in adult mouse neurons), the relative depletion of CG dinucleotides in eukaryotic genomes enables the high abundance of methylation in the CH context in neurons (Guo et al., 2014; Kinde et al., 2015; Lister et al., 2013; Mo et al., 2015; Schultz et al., 2015; Varley et al., 2013; Xie et al., 2012; Ziller et al., 2011). Therefore, these 5hmC and 5mCH have added important new dimensions in understanding epigenetic regulation of neuronal function.

5hmCG: a neuronally enriched form of DNA methylation associated with gene activation

5mC is oxidized by the ten-eleven translocation (TET) family of enzymes to form 5hmC (Tahiliani et al., 2009). Three Tets are expressed in the brain (*Tet1-3*), with *Tet2* and *Tet3* mRNA levels being considerably higher than *Tet1* in the postnatally (Klose and Bird, 2006; Szwagierczak et al., 2010). Recently developed sequencing approaches have enabled the genomic detection of 5hmC, providing invaluable insight into the biological functions of this epigenetic mark. Approximately 0.9% of all cytosines are hydroxymethylated in the mouse brain, predominately (>98%) occurring in the CG context (Lister et al., 2013). 5hmC is enriched at transcriptionally active sites, such as gene bodies, transcriptional end sites, DNaseI-hypersensitive sites (DHSs), and both active and poised enhancers, and is depleted at promoters and major satellite regions (Lister et al., 2013; Ong and Corces, 2014; Wen et al., 2014). Additionally, gene expression levels strongly correlate with intragenic 5hmC levels (Lister et al., 2013; Mellén et al., 2012a; Song et al., 2011). These features suggest that 5hmC is involved in gene activation, which starkly contrasts with the biological function of 5mC. The importance of 5hmC in neuronal maturation and function is highlighted by the fact that 5hmC levels increase nearly tenfold over development (Lister et al., 2013), and genes enriched for 5hmC in the mammalian brain relative to other tissues are related to

synaptic function (Khare et al., 2012a). The substantial body of literature regarding the importance of the 5hmC and TETs in neuronal function will be reviewed below.

5mCH: a repressive DNA methylation mark in neurons

Whereas 5mCG is established by three DNMTs (DNMT1, DNMT3a, and DNMT3b) and levels remain unchanged during development (Klose and Bird, 2006), 5mCH is catalyzed by DNMT3a (Ramsahoye et al., 2000) and levels increase during synaptogenesis, which coincides with an increase in *Dnmt3a* mRNA levels during this same time period (Guo et al., 2014; Lister et al., 2013). Further supporting the unique role of *Dnmt3a* in 5mCH formation is a study in which exogenous *Dnmt3a* expression in *Drosophila* lead to the formation of *de novo* 5mCH (Ramsahoye et al., 2000). Additionally, conditional deletion of *Dnmt3a* from the brain during early development results in a significant reduction in 5mCH, specifically in the 5mCA and 5mCT contexts, but not in the 5mCG context (Gabel et al., 2015; Guo et al., 2014). Similarly to 5mCG, 5'-upstream, intragenic, and 3'-downstream 5mCH levels inversely correlate with gene expression, with gene body 5mCH outperforming 5mCG as an indicator transcriptional levels in the brain (Guo et al., 2014; Mo et al., 2015). Additionally, 5mCH patterns are cell type-specific, which is similar to 5mCG, but 5mCH is an even better indicator than 5mCG in this regard (Guo et al., 2014; Mo et al., 2015). 5mCH-mediated repression has been confirmed with the use of a methylated quantitative reporter assay in hippocampal neurons (Guo et al., 2014). Furthermore, 5mCH is abundant at regions with low CpG density and at linker regions of nucleosomes, whereas it is absent from sites bound by transcription factors and at nucleosomal cores (Burger et al., 2013; Guo et al., 2014; Mo et al., 2015). Taken together, this suggests that brain-specific establishment or mediation of this epigenetic mark may contribute to its unique regulatory effect in neurons (Kinde et al., 2015), and we will discuss these possibilities below.

The necessity of DNA methylation in neurodevelopment

The importance of DNA methylation in neurons has been demonstrated in studies in which the loss of key enzymes that regulate DNA methylation, DNMTs and

TETs, leads to deficits in neuronal function. Embryonic deletion of *Dnmt1* from forebrain neuronal progenitors in mice results in hypomethylation, deregulation of gene expression, deficits in cerebral cortical formation and maturation, increased dendritic branching, reductions in LTP, defects in learning and memory, and severe embryonic and early postnatal degeneration (Golshani et al., 2007; Hutnick et al., 2009). Similarly, embryonic deletion of *Dnmt3a* from the entire central nervous system (CNS) of mice also results in neuronal dysfunction, such as hypoactivity, motor abnormalities, decreased grip strength, a reduction in motor neuron number, and a shortened lifespan (Nguyen et al., 2007). Human genetics also supports the importance of DNMTs in neuronal function as *DNMT1* has been linked to hereditary sensory and neuropathy with dementia and hearing loss (Klein et al., 2011), *DNMT3A* with an overgrowth syndrome with intellectual disability (Tatton-Brown et al., 2014), *DNMT3B* with recessive mutations that cause immunodeficiency, centromere instability, and facial anomalies (ICF) syndrome in which a large percentage of patients have intellectual disability (Bestor et al., 1999; Bird, 2002; Okano et al., 1999).

Similarly to DNMTs, various studies have also demonstrated the importance of TETs in brain development. *Tet1* constitutive KO mice or mice lacking the catalytic dioxygenase domain of *Tet1* have memory deficits, a reduction in the expression of genes associated with neurogenesis and neuronal activity (along with hypermethylation at the promoters of this gene group), LTP deficits, increased LTD, and a reduction in the proliferating potential of neural progenitor cells (NPCs) (Kaas et al., 2013; Zhang et al., 2013). Proliferation potential is also influenced by the other two TET family members, TET2 and TET3, as the knockdown of *Tet2* and *Tet3* via electroporation of shRNAs into the cortex of mice leads to defects in the progression of differentiated neurons from the subventricular zone (Hahn et al., 2013).

Given that these are embryonic perturbations of DNMTs and TETS, these findings highlight the importance of these enzymes and DNA methylation dynamics during neuronal development. Studies specifically focusing on these enzymes in postmitotic neuron have also highlighted the importance of DNA methylation in neuroplasticity, as discussed below.

DNA methylation in neuronal maturation and neuroplasticity

Since DNMTs are usually expressed in frequently dividing cells for the establishment and maintenance of DNA methylation patterns, it was originally perplexing to find considerable levels of maintenance enzyme *Dnmt1* and *de novo* enzyme *Dnmt3a* in adult postmitotic neurons (Feng et al., 2005; Goto et al., 1994). However, several lines of evidence now support that these enzymes are essential for neuronal plasticity in postmitotic neurons.

Learning and memory is associated with alterations in the methylation status of various plasticity-related genes. Fear conditioning causes the memory-suppressor gene *Pp1* to become rapidly methylated, resulting in the repression of this gene. This response is dependent upon DNMT activity given that chemical inhibition of these enzymes abolishes this effect (Miller and Sweatt, 2007). Another plasticity gene whose methylation status is affected in response to contextual fear conditioning is *Bdnf,* which undergoes promoter specific methylation that correspond with isoform-specific changes in gene expression (Lubin et al., 2008; Martinowich, 2003). The regulation of isoform-specific expression of *Bdnf* is DNMT-dependent since chemical inhibition of DNMTs results in alterations in promoter methylation and isoform expression (Lubin et al., 2008). The importance of DNMTs in eliciting the changes in gene expression necessary for long-term memory formation is corroborated by the observed increase in *Dnmt3a* and *Dnmt3b* expression in the hippocampus in response to fear conditioning in both mice and rats (Miller and Sweatt, 2007; Morris et al., 2014), and inhibiting DNMT activity, via intra-hippocampal infusion of a DNMT inhibitor, abolishes memory formation (Miller and Sweatt, 2007).

Double conditional knockout mice lacking *Dnmt1* and *Dnmt3a* in postnatal forebrain excitatory neurons (deletion occurring at 2-3 weeks using the *Camk2a* promoter-drive Cre recombinase, *Camk2a-Cre*) results in a decrease in DNA methylation, misexpression of genes, reduction in soma size, alterations in synaptic function, and deficits in learning and memory (Feng et al., 2010). Although *Dnmt1* single conditional knockout in postnatal forebrain excitatory neurons (via the use of *Camk2a-Cre)* does not appear to affect learning and memory nor a variety of cellular and molecular characteristics in the hippocampus or cortex, including methylation levels, neuronal survival, gene expression, soma size, or synaptic function (LTP and LTD) (Fan

et al., 2001; Feng et al., 2010; Morris et al., 2014); however, this loss of *Dnmt1* does result in an anxiolytic and antidepressant phenotype, suggesting that cell types, brain regions, or neuronal circuits that underlie specify behaviors can be differentially vulnerable to the loss of certain DNMTs (Morris et al., 2016). A consensus on the consequence of *Dnmt3a* loss in postmitotic forebrain excitatory neurons remains to be determined, as one group reports no changes in any parameters assessed (DNA methylation, gene expression genes, soma size, synaptic function, and learning and memory), and another group finding deficits in LTP in conjunction with associative and episodic memory dysfunction (Feng et al., 2010; Morris et al., 2014).

Additionally, the importance of the TETs in postmitotic neurons has also been ascertained. *Tet1* is involved in neuronal activity-induced DNA demethylation and gene expression changes since short-hairpin mediated knockdown of endogenous *Tet1* in the dentate gyrus (DG) completely abolished electroconvulsive stimulation (ECS)-induced demethylation of *Bdnf IX* and *Fgf1B* promoters (Guo et al., 2011b). These findings are in agreement with another study that has shown that *Tet1* knockdown in hippocampal neurons leads to the hypermethylation of promoter IV of *Bdnf* and subsequent repression of transcript from this promoter (Yu et al., 2015). Transcript levels of *Tet1*, but not Tet2 or Tet3, is significantly downregulated in the dorsal CA1 of mice after fear learning (Kaas et al., 2013).

TET2 is also thought to play a role in the demethylation of developmentally dependent genomic loci. With the use of *Tet2* knockout mice, it was found that this TET family member is responsible for the oxidation of large fraction (19.7%) of CG genomic regions that gain hydroxymethylation status over development and aging (Lister et al., 2013). Studies on TET3 function in the brain have confirmed that this most highly expressed brain Tet family member is essential in mediating neuronal activity-dependent gene expression programs. When mice undergo extinction training, there is a significant increase in *Tet3* mRNA in the cortex. *Tet3* knockdown via lentiviral plasmids in the infralimbic prefrontal cortex (ILPFC) have normal fear memory acquisition but impaired fear memory extinction. Furthermore, inhibiting NMDAR activity blocked the increase in *Tet3* expression associated with fear memory extinction, suggesting that the rise in *Tet3* occurs via an NMDAR-mediated pathway. Fear acquisition and fear extinction result in genome-wide differences of 5hmC at locations that contain CA or CT

dinucleotide repeats instead of CGs (Li et al., 2014). *Tet3* expression levels correlate with neuronal activity as well; an increase in synaptic transmission correlates with an increase in *Tet3* mRNA and protein levels, but not *Tet1* or *Tet2*. When *Tet3* is knocked down from hippocampal neurons in culture, mEPSC amplitudes are significantly larger than controls, and the reciprocal effect occurs when Tet3 is overexpressed. Tet3 is also essential for the maintenance of homeostatic synaptic plasticity. In Tet3 knockdown neurons, promoter IV of *Bdnf* is hypermethylated, resulting in a decrease in expression from this promoter (Yu et al., 2015). Taken together, these findings implicate that there are essential locus-specific alterations in DNA methylation during neuronal plasticity that are mediated by DNMTs and TETs.

Deciphering DNA methylation: An introduction to methyl-CpG-binding proteins

Methylation is thought to influence transcription by altering DNA-protein interactions, such as the binding of transcription factors (Becker et al., 1987) or the recruitment of proteins that bind methylated DNA (Filion et al., 2005; Hendrich and Bird, 1998; Prokhortchouk et al., 2001; Unoki et al., 2004) The families of proteins that bind methylated DNA to mediate the molecular consequences of this epigenetic mark, canonically known as “readers” of DNA methylation, are the SET- and Ring finger-associated (SRA) domain family, the Kaiso family of proteins, and the methyl-CpG-binding domain (MBD) protein family (Hendrich and Bird, 1998; Nan et al., 1993). MBDs, consisting of six MBD proteins (MBDs 1-6) and methyl-CpG-binding protein 2 (MeCP2), were originally demarcated as a family due to their shared, highly conserved MDB domain (Hendrich and Bird, 1998; Roloff et al., 2003); however, it has since been discovered that two of the family members, MBD5 and MBD6, do not bind to methylated DNA (Laget et al., 2010) and that some of these family members also have the ability to bind to other forms of DNA methylation in addition to 5mCG through their MBD: a point mutation in the MBD of MBD3 renders it able to bind unmodified cytosine, 5mCG, and 5hmCG(Hashimoto et al., 2012; Spruijt et al., 2013) and MeCP2 has recently been demonstrated to bind with high affinity to 5mCA and 5hmCH (Chen et al., 2015; Gabel et al., 2015; Guo et al., 2014; Lagger et al., 2016). In this review, we will focus on the most-

well studied reader of DNA methylation, MeCP2, given its recognized importance in neuronal function and unique ability to bind a neuronal-enriched form of DNA methylation, 5mCH.

MeCP2's role in neuronal function

MeCP2 is a nuclear protein found in various tissues in mammals (Meehan et al., 1989; Shahbazian, 2002), with highest expression in the brain where it is half as abundant as nucleosomes (Skene et al., 2010). Within the brain, MeCP2 protein levels are seven times higher in neurons than glia, underscoring the importance of this protein in neuronal function (Shahbazian, 2002). However, various studies have shown that MeCP2 is also important for glial function, supporting glial dysfunction as a contributing factor in RTT pathology (Ballas et al., 2009; Delépine et al., 2016; Lioy et al., 2011; Maezawa et al., 2009; Nguyen et al., 2012). The levels of MeCP2 in the mammalian brain increase over postnatal development, suggesting that MeCP2 is important for synapse maturation (Balmer et al., 2002; Shahbazian, 2002). Clear evidence for the importance of this gene in CNS came with the discovery that mutations in *MECP2* cause Rett Syndrome (RTT), a severe neurological disorder that affects 1 in 10,000 live female births, making it one of the most common causes of intellectual disability in females (Amir et al., 1999). Given that *MECP2* is an X-linked gene, RTT is almost exclusive found in females, as hemizygous loss of MECP2 function in males leads to severe neonatal encephalopathy and death (Amir et al., 1999; Ellison et al., 1992; Schanen and Francke, 1998; Schanen et al., 1997; Zoghbi et al., 1990).

RTT is characterized by normal development for the first 6 months of life followed by regression and symptomatic presentation, which includes loss of acquired purposeful hand skills and spoken language, gait abnormalities, and hand stereotypies (Jeffrey L Neul et al., 2010; Percy et al., 2010). Postnatal deceleration of head growth is another clinical feature of RTT (Jeffrey L Neul et al., 2010; Percy et al., 2010), and radiological and pathological examination have determined that this is due to a reduction in brain size (Armstrong et al., 1995). Given the absence of degenerative features, this change is brain volume is attributed to a reduction the number and length of dendrites, resulting in more densely pack neurons (Armstrong et al., 1995; Belichenko et al., 1994). These changes are brain region- and cortical layer-specific, as different

groups have only observed changes in a subset areas examined (Armstrong et al., 1995; Wang et al., 2013). Additionally, the cortical layers (layer III and V) are indicative of deficits initiating during the second period of cortical development when lower layers are refining their connectivity with layer III postnatally. Furthermore, these dendritic alterations do not worsen over adulthood, indicating that RTT is not a progressive disease after development (Armstrong et al., 1995). Supporting a functional consequence of these anatomical deficits are electrophysiological studies in neurons derived from RTT induced pluripotent stem cells (iPSCs) that have found a decrease in the amplitude and frequency of spontaneous excitatory transmission (Marchetto et al., 2010). Taken together, these findings in RTT patients are clear indications of a disrupted neuronal network and suggest that RTT is a neuronal maturation and/or maintenance disorder.

To understand the molecular underpinnings of RTT, mice lacking *Mecp2* or carrying RTT patient-associated mutations in *Mecp2* have been developed, and the majority of neurological features are observed in these whole-body *Mecp2*-mutant mice, such as normal development during early life followed by symptomatic presentation of altered gait, motor incoordination, hindlimb clasping, and cognitive deficits (Chen et al., 2001; Goffin et al., 2012; Guy et al., 2001). Loss of *Mecp2* from the CNS either embryonically (embryonic day 12, E12) or postnatally (postnatal day 21, P21) resulted in phenotypes that were indistinguishable from whole-body loss of *Mecp2* (Chen et al., 2001), emphasizing the importance of this MBD protein in mature neuronal function and maintenance. In addition to behavioral phenotypes, *Mecp2*-mutant mice display additional pathological and histological features found in RTT patients, namely, a smaller brain size (Chen et al., 2001; Goffin et al., 2012; Guy et al., 2001) and a reduction in the number and length of dendrites (Belichenko et al., 2009a; 2009b). Furthermore, dendritic spines are reduced in number and altered in morphology, resembling immature spines, in *Mecp2*-null mice (Belichenko et al., 2009a; 2009b; Wood and Shepherd, 2010; Wood et al., 2009). The reduction in dendrite branches and dendritic spines is suggestive of neuronal connectivity deficits in *Mecp2*-null mice. And in fact, electrophysiological experiments corroborate these anatomical alterations. A reduction in excitatory synapses, as well as the quantal excitatory transmission at those synapses, contributes to a reduction in spontaneous firing rate of cortical neurons in *Mecp2*-null mice (Chao et al., 2007; Dani and Nelson, 2009; Dani et al., 2005; Goffin et al., 2012). In addition to synaptic dysfunction, there are also deficits in LTP and LTD, which is

thought to occur subsequently to the synaptic dysfunction (Dani and Nelson, 2009). This alteration in LTP and LTD harkens back to LTP and LTD deficits observed in DNMT and TET loss of function mice. Given these neuronal deficits, the molecular perturbations that occur with MeCP2 mutations or loss have been intensely studied in order to understand the biological function of this protein.

Proposed molecular functions of MeCP2

Multiple functions have been proposed for MeCP2

This year marks the 25th year anniversary of the discovery of MeCP2 (Lewis et al., 1992). Over this span of time, ~1000 papers have been published with *MeCP2* in the title. This large body of research has greatly advanced our understanding of MeCP2; however, a consensus on the specific molecular function of MeCP2 has yet to be reached, with evidence supporting MeCP2's role as a transcriptional repressor, transcriptional activator, chromatin organizer, regulator of alternative splicing, and miRNA processor (Lyst and Bird, 2015). In this review, we will focus on the three prevailing models, transcriptional repressor, transcriptional activator, and chromatin organizer.

MeCP2: a repressor?

A screen for proteins that bind methylated DNA led to the discovery of MeCP2, given its high affinity for DNA methylation, as it is capable of binding DNA with a symmetrically methylated 5mCG (Lewis et al., 1992) through its 85kda MBD (Nan et al., 1993). In this initial work, it was also found that MeCP2 nuclear localization mirrored that of 5mCG; it is concentrated at pericentromeric heterochromatin, which is known to contain ~40% of all genomic 5mCG at major satellite DNA (Lewis et al., 1992). Another functional domain of MeCP2 is the transcriptional repression domain (TRD), which has been show to exhibit long range repression, up to 2kb away from a transcriptional start site (Nan et al., 1997). A small portion of the TRD interacts with the histone deacetylase (HDAC)-containing co-repressor complexes nuclear receptor co-repressor (NCOR) and silencing mediator of retinoic acid and thyroid hormone receptor (SMRT), and bridge

this site of interaction is termed the NCOR-SMRT interaction domain (NID) (Kokura et al., 2001; Lyst et al., 2013; Stancheva et al., 2003). MeCP2 has also been shown to interact with the HDAC-containing repressor complex Sin3a (Bird et al., 1998; Jones et al., 1998). The importance of the MDB and NID is also supported from human genetics since RTT-causing missense mutations, but not neutral polymorphisms mutations, predominately cluster in these two protein domains (Lyst and Bird, 2015; Lyst et al., 2013). Notably, each of the four missense mutations within the NID abrogate MeCP2's ability to bind to the NCoR/SMRT co-repressor complex (Lyst et al., 2013), highlighting the importance of this interaction for MeCP2 function. Since deacetylation of histone tails leads to transcriptional repression (Grunstein, 1997), these findings on the functional domains of MeCP2 suggest a model in which MeCP2 acts as a molecular bridge to connect DNA methylation with chromatin changes that elicit repression (Bird et al., 1998).

Various studies support that MeCP2 binds globally across the genome at methylated cytosine sites (Chen et al., 2015; Cohen et al., 2011; Gabel et al., 2015; Skene et al., 2010). Chromatin immunoprecipitation (ChIP) has demonstrated that MeCP2 binding profiles scale with CG methylation density (Baubec et al., 2013; Chen et al., 2015; Skene et al., 2010), with high levels of binding at CG methylation dense regions and lower levels of binding at unmethylated CG islands (CGIs) (Skene et al., 2010). It has recently been discovered that in addition to 5mCG, MeCP2 can also bind to 5mCH, specifically 5mCA, with high affinity (Chen et al., 2015; Gabel et al., 2015; Guo et al., 2014; Kinde et al., 2015; Lagger et al., 2016). When the genome-wide MeCP2 binding profiles in mouse cortical tissue were compared to base-pair resolution profiles of 5mC and 5hmC, it was found that MeCP2 binding globally correlates with 5mCG, 5mCA (or 5mCAC), and 5hmCAC density, but not 5hmCG. Additionally, misregulated genes in *Mecp2* KO mice are enriched for MeCP2 binding and 5mCA (Chen et al., 2015; Gabel et al., 2015; Guo et al., 2014; Kinde et al., 2015; Lagger et al., 2016). However, it remains to be determined which subsets of misregulated genes show this enrichment. One group finds that genes enriched with MeCP2 binding tend to be long and also enriched with 5mCA, and that in the absence of *Mecp2* in cortical tissue, upregulated genes tend to be long, implicating MeCP2 as a transcriptional repressor (Gabel et al., 2015). An independent group confirms this finding in the hypothalamus with the finding that MeCP2 binding is enriched at genes that are upregulated in *Mecp2* KO mice (Lagger et

al., 2016). However, another group reported that both upregulated and downregulated genes in *Mecp2*-null hypothalamus tissue are enriched for MeCP2 binding and 5mCH, suggesting a more complex model in which MeCP2 is both an activator and repressor (Chen et al., 2015). An important difference between the analyses carried out by the first two groups versus the third group is that the former analyzed MeCP2 and 5mCH enrichment within the gene bodies and surrounding regions of misregulated genes (Gabel et al., 2015; Lagger et al., 2016) whereas the later normalized intragenic MeCP2 and 5mCH enrichment to the surrounding regions of misregulated genes (Chen et al., 2015). Given that the raw data used by two opposing groups (Chen et al., 2015; Lagger et al., 2016) were identical, it is important to keep in mind that data analysis methods can critically influence the interpretation of MeCP2 function since gene expression changes are subtle in RTT.

MeCP2: an activator?

In direct opposition of the repressor model, MeCP2 has also been proposed to be a transcriptional activator. In mouse neurons differentiated from *Mecp2*-null ES cells, there is a significant reduction of overall RNA synthesis, which was detected with the use of a radioactive ribonucleotide incorporation assay (Yazdani et al., 2012). Similarly, in human neurons derived from *MECP2*-null ES cells, per-cell total RNA levels are significantly reduced, and this phenotype gets worse as time passes (Li et al., 2013). These findings are supported by an independent group that found a significant reduction of Serine5 phosphorylated RNA polymerase (a proxy for transcriptional activity) in *Mecp2*-null mouse neurons *in vivo* using array tomography (Linhoff et al., 2015). Additionally, gene expression studies from various brain regions of *Mecp2*-null mice have found that more genes are repressed than activated, supporting MeCP2's role as an activator (Ben-Shachar et al., 2009; Chahrour et al., 2008; Zhao et al., 2013). Furthermore, genome-wide promoter analysis revealed that the majority of MeCP2-bound promoters are actively expressed genes that are not highly methylated, which therefore indicates that MeCP2 binding isn't solely correlated with repression (Yasui et al., 2007). Molecularly, there is evidence supporting that MeCP2 mediates its activator role through its interaction with cyclic AMP-responsive element-binding protein 1

(CREB1) at genes that are activated, but not at those that are repressed (Chahrour et al., 2008). However, corroboration of this finding by independent groups is lacking. Related to this, MeCP2 has been found to bind to 5hmCH (more specifically, 5hmCA) *in vitro* with the use of the electrophoretic mobility shift assay (EMSA) (Gabel et al., 2015). However, whether 5hmCA is associated with gene activation, like 5hmCG, or whether the binding of MeCP2 to 5hmCA is biologically significant given the low percentage of 5hmC that is in the CH context, remains to be determined. Notably, multiple studies support that MeCP2 does not bind to 5hmCG, either *in vitro,* with the use of EMSAs, fluorescent polarization, or mass spectrometry (Gabel et al., 2015; Hashimoto et al., 2012; Khrapunov et al., 2014; Spruijt et al., 2013; Valinluck et al., 2004), or *in vivo* via correlating MeCP2 ChIP-seq profiles with base-pair resolution profiles of 5hmCG, which in fact shows an anticorrelation between the two (Gabel et al., 2015). Taken together with the previous section, these data indicate that more studies are necessary before classifying MeCP2 as a *bona fide* transcriptional activator or repressor.

MeCP2: a transcriptional mediator of neuronal activity?

In line with MeCP2 acting as a regulator of transcription, are findings demonstrating MeCP2's ability to regulate gene expression in an activity-dependent manner. In this model, at basal conditions, unmodified MeCP2 functions as a repressor when bound to gene regions. However, upon stimulation, post-translational modifications on MeCP2 lead to a reduction in MeCP2 binding and subsequent gene activation. This function of MeCP2 is particularly intriguing given the necessity of genes expression changes for the maintenance of long-term neuronal plasticity. One important neuronal plasticity-related gene affected by the loss of MeCP2 function is *Bdnf,* with transcript and protein levels being significantly reduced in the brains of *Mecp2*-knockout mice (Chahrour et al., 2008; Chang et al., 2006; Chen et al., 2015). The levels of *Bdnf* appear to be critically important for the RTT phenotype since increasing levels of *Bdnf* in *Mecp2*-null mice ameliorates electrophysiological dysfunction, improves locomotor deficits, and extends lifespan (Chang et al., 2006). In the wild type setting, 5mCH levels and MeCP2 binding are enriched across the *Bdnf* locus, suggesting that MeCP2 binding to developmentally regulated 5mCH sites is critical for the regulation of genes that are

important in the manifestation of RTT phenotypes (Chen et al., 2015). Under basal conditions in neuronal cultures, MeCP2 is bound to the methylated promoter III of *Bdnf* where it is thought to act as a repressor of transcription through the recruitment of the Sin3a repression complex which is found to be present via ChIP (Chen, 2003; Martinowich, 2003; Zhou et al., 2006). However, in response to membrane depolarization, MeCP2 and the Sin3a complex disassociates from the *Bdnf* promoter and a reduction of methylation at the promoter in conjunction with an increase in expression, dendritic complexity, and spine maturation (Chen, 2003; Martinowich, 2003; Zhou et al., 2006). The release of MeCP2 is mediated by calcium-dependent, CaMKII-mediated phosphorylation of MeCP2 at serine 421 (pS421) given that loss of this phosphorylation site on MeCP2 results in the loss of activity-dependent *Bdnf* expression (Chen, 2003; Zhou et al., 2006). This phosphorylated form of MeCP2 is brain-specific, as it is absent in a multitude of tissues, suggesting that this activity-dependent form of MeCP2 might underlie why RTT is predominately a neurological disorder.

In the mouse brain in response to neuronal activity, 10-30% of total MeCP2 becomes phosphorylated at S421 (Cohen et al., 2011). However, the genomic binding profile of MeCP2 remains virtually unchanged in response to neuronal activity in the mouse brain, in opposite to what was observed at *Bdnf*. To investigate the importance of MeCP2 phosphorylation at S421 *in vivo*, knock-in mice have been created in which S421 is converted to alanine (S421A), prevent the phosphorylation of this site. Cortical pyramidal neurons in these mutants have increased dendritic complexity and an increase in the amplitude of miniature inhibitory postsynaptic currents (mIPSCs), highlighting the importance of this phosphorylation site on MeCP2 in neuronal development and function. When MeCP2 binding is assessed in wild-type and MeCP2-S421A primary cultures in response to neuronal activity, it was found that the binding patterns are maintained in the mutant, implicating that this phosphorylation site is not necessary for the release of MeCP2 from bound sites in response to neuronal activation (Cohen et al., 2011). Similarly to what was observed in the mouse brain, when MeCP2 binding patterns are compared in basal and stimulated primary neuronal cultures, no changes are detectable, which does not support a model in which activity-dependent phosphorylation of MeCP2 releases repressive binding. Additionally, the employment of a more sensitive and targeted approach, ChIP-PCR of MeCP2 at activity-dependent genes in neuronal cultures in response to stimulation, MeCP2 binding remained similar. This suggests that

additional phosphorylation events might contribute to the release of MeCP2 from chromatin or that specific stimulation paradigms are require to elicit a release. Additionally, activity induced gene expression programs are not altered in the MeCP2-S241A mutant mice, suggesting that singular phosphorylation site is not responsible for the release of MeCP2 from genomic regions (Cohen et al., 2011). And in fact, three additional phosphorylation sites on MeCP2 have been described: S86, S247, and T308. Phosphorylation of these sites on MeCP2 are differentially induced in response to various forms of stimulation, such as robust neuronal activity, exogenous BDNF, and chemically induced elevations of cAMP, suggesting that these phosphorylation sites induce different expression patterns in response to external stimuli. Supporting the importance of one of these phosphorylation site is the finding that mice carrying a T308A mutation in MeCP2 display RTT phenotypes, constitutive NCoR/SMRT binding to MeCP2, and a reduction in the expression of activity-induced genes (Ebert et al., 2013). In addition to the S421, S86, S247, and T308 phosphorylation sites on MeCP2, additional phosphorylation sites and post-translational modifications on MeCP2 have been identified, suggesting that this protein is regulated in various ways (Gonzales et al., 2012; Tai et al., 2016). Taken together, these findings highlight the potential importance of MeCP2 as a mediator of gene expression changes in response to neuronal activity; however, much remains to be explored to determine the transcriptional effects of the various post-translation modifications on MeCP2.

MeCP2: an architectural protein?

In addition to the MBD, MeCP2 has three additional domains that enable it to bind to DNA. These are three basic clusters containing AT-hook-like domains, which bind to AT-rich sequences of DNA (Baker et al., 2013; Lewis et al., 1992). The recombinant form of one of these AT-Hook domains, AT-Hook 2 (amino acids 265-272), has been shown to independently bind to DNA, and an RTT-associated mutation in this domain (R270X) abolishes this ability. Additionally, the importance of this domain in forming higher order structures is highlighted by the finding that recombinant MeCP2-WT is able to compact nucleosomal arrays, but MeCP2-R106X is not (Baker et al., 2013). This also relates to *in vitro* work that has demonstrated the necessity of A/T rich DNA (four or

more A/T bases) adjacent to a methylated CpG site for efficient MeCP2 binding (Klose et al., 2005). Another region of MeCP2 implicated in chromatin organization is the C-terminal portion of the MBD that has been found to bind to ATRX, a chromatin remodeler that is mutated in α-thalassemia/mental retardation, X-linked syndrome (ATRX syndrome) (Nan et al., 2007). MeCP2 is thought to be important for the localization of ATRX to pericentric heterochromatin, as loss of MeCP2 function in *Mecp2*-null mice or in mice carrying several RTT-associated point mutations in *Mecp2* abolishes ATRX localization to pericentric heterochromatin specifically in the brain (Baker et al., 2013; Nan et al., 2007).

Recombinant MeCP2 binds, compacts, and oligomerizes nucleosomal arrays *in vitro* (Georgel et al., 2003; Ghosh et al., 2010; Nikitina et al., 2007b). Compaction by MeCP2 comes in two forms, the clustering of nucleosomes and the formation of DNA-MeCP2-DNA complexes (Nikitina et al., 2007b). One striking feature of nucleosomal array compaction by MeCP2 is the formation of chromatin loops (free DNA emanating from clusters of nucleosomes) (Ghosh et al., 2010; Horike et al., 2004), which are implicated in the silencing of loci *in vivo* (Horike et al., 2004)**.** Further supporting MeCP2's role in chromatin compaction is the finding that ectopic expression of the MBD of MeCP2 in mouse myoblasts is sufficient to cause the clustering of pericentric heterochromatin (Brero et al., 2005). In a comprehensive study evaluating the effect of 21 RTT-associated mutations to bind and cluster chromocenters in mouse myoblasts, it was found that half of the mutations lead to deficits in chromatin binding and two-thirds showed a decrease in the ability to cluster chromocenters (Agarwal et al., 2011). When the nuclear localization of heterochromatin protein 1 (HP1), a protein essential for heterochromatin packaging, and MeCP2 were compared over cellular differentiation using an *in vitro* system culture system for myogenesis, it was found that both HP1 (specifically the γ isoform (HP1γ) and MeCP2 redistributed to heterochromatin foci around the same time (Agarwal et al., 2007). Additionally, MeCP2 interacts through its N-terminal domain (the first 55 amino acids) to the chromo shadow domain of HP1 (Agarwal et al., 2007). Taken together, these *in vitro* and cell culture studies suggest that MeCP2 plays a role in the organization and compaction of nucleosomes, especially in heterochromatin, and suggest that loss of function of MeCP2 would lead to decondensation of heterochromatin.

However, although few in number, studies examining the *in vivo* effects of MeCP2 loss on heterochromatin compaction in mouse neurons have found an increase, rather than decrease, in heterochromatinization. Using array tomography (AT) in mosaic RTT mouse females, it was found that although the total amount of nuclear DAPI remained similar between WT and *MeCP2*-null neurons, there was an increase of DAPI density in heterochromatin, as well as an increase in heterochromatin size in CA1 pyramidal cells. Additionally, it was found that heterochromatin volume and nuclear volume are negatively correlated, suggesting that chromatin condensation reduces nuclear size. Interestingly, AT also allowed for quantitative analysis of the distribution of MeCP2 within the nucleus, finding that only 30% of MeCP2 is localized to heterochromatin, while the remain 70% localizes to euchromatin and nucleoplasm. AT imaging allowed for the detection of H4K20me3 overlapping with heterochromatin foci in *MeCP2*-null nuclei, which is not found in *MeCP2*-positive neurons, but rather H4K20me3 staining was found in a region adjacent to dense, major satellite rich heterochromatic foci in CA1 pyramidal neurons. The redistribution of H4K20me3 in *Mecp2*-null neurons is dramatic, with 65% changing it localization to dense heterochromatin, while there was a negligible increase (11%) in H4K20me3 nuclear levels. Additionally, there is a significant distribution of H3K9me3 to dense, major-satellite rich heterochromatin, as well as a slight increase (10%) in total H3K9me3 abundance (Linhoff et al., 2015). This change in nuclear organization is cell type-specific given that granule cells of the dentate gyrus show a similar pattern that was observed in CA1 pyramidal neurons, but this alteration is absent in granule cells of the cerebellum. Another study found that although heterochromatin size and numbers are affected embryonically and perinatally in *Mecp2*-null neurons, these affects are not found later during development. The importance of chromocenter size is highlighted by a study that found an increase in heterochromatin size after induced depolarization in primary neuronal cultures; however, this effect is absent in *Mecp2*-deficient primary neurons (Singleton et al., 2011). Given that the findings from the few studies on heterochromatic changes *in vivo* are disparate, more research is warranted to determine the organizational defects that occur when MeCP2

Further support for MeCP2's role in chromatin organization comes from studies that have compared the functions of MeCP2 to histone H1 (H1), an essential component of chromatin that binds to linker DNA and organizes nucleosomes into higher order

structures. H1 is present at the level of 1 molecule per nucleosome in most somatic cells, except for neurons, where it is present at the level of 1 molecule for every two nucleosomes (Allan, 1984; PEARSON et al., 1984; Skene et al., 2010). *In* vitro, MeCP2 is able to displace H1 from preassembled chromatin (Ghosh et al., 2010; Ishibashi et al., 2008; Nan et al., 1997), but not vice versa(Ghosh et al., 2010), suggesting that H1 and MeCP2 compete for binding to linker DNA regions. Additional *in vitro* support for this comes from studies with nucleosomal arrays in which human MeCP2 is found to bind to 11bp of the linker DNA entry-exit site, protecting this portion of linker DNA from micrococcal nuclease (MNase) digestion, and an RTT-associated mutation in the MDB, MeCP2-R106W, results in a loss of protection of the linker DNA from MNase digestion (Nikitina et al., 2007a). Furthermore, linker DNA is essential for MeCP2's ability to properly bind to nucleosomes, and this binding induces a variety of conformational changes to linker DNA (Nikitina et al., 2007a). Additionally, of all the canonically histone components, MeCP2 is in closest proximity to H3, whose N-terminal region is in close proximity to the linker region (Nikitina et al., 2007a). When mouse fibroblasts stably expressing fluorescent MeCP2 are challenged with recombinant H1, and vice versa (mouse fibroblasts stably expressing fluorescent H1 were challenged by recombinant MeCP2), it was found that MeCP2 and H1 compete for binding sites, but that similar to *in vitro* findings, MeCP2 is more effective at expelling H1 than the reverse (Ghosh et al., 2010). In mouse fibroblasts, H1 and MeCP2 colocalize at pericentromeric heterochromatin (Ghosh et al., 2010). MeCP2 induces nucleosomal array compaction in a zigzag folding pattern that is very similar to the manner in which H1 induces compaction, which is thought to be essential for 30nm fiber formation that is the basic building block of heterochromatin (Ghosh et al., 2010). Furthermore, in *Mecp2*-knockout neurons from the cortex, H1 protein levels double, reaching the abundance of 1 molecule per nucleosome in neurons (Skene et al., 2010). Histone H1 levels do not change in unsorted nuclei from the cortex, which contains both glia and neurons, suggesting that this change is specific to neurons (Skene et al., 2010). This suggests that lower levels of H1 in the neurons are due to the presence of MeCP2, and that in the absence of MeCP2, sites are now open for histone H1 occupancy (Skene et al., 2010). It is possible that the substitution of MeCP2 by H1 leads to less dynamic transcriptional changes in response to neuronal activity. However, this remains to be fully addressed.

MeCP2: a multifunctional protein

Twenty-five years of research on MeCP2 has led to vast insight into the complex, multifaceted function of this protein, generating numerous hypothesized molecular models. And yet another model is emerging—a synergistic model that treats many of the proposed functions as an integrative piece of MeCP2's function instead of treating them as mutually exclusive entities. It is possible that MeCP2 is predominately as an architectural protein that is important for regulating chromatin structural changes that aid in the activation or repression of genes necessary for long-lasting plasticity in response to neuronal activity. Therefore, in this model, the transcription activator or repressor effects that have been attributed to MeCP2 would be secondary to its organizational role, similar to the function of CCCTC-binding factor (CTCF) (Ong and Corces, 2014; Phillips and Corces, 2009). However, to prove such a model, additional studies are necessary to determine nuclear organizational defects, and the consequential effect on gene expression, elicited by loss of MeCP2 function. And therefore, the goal of this work was to carryout such analysis for the better understanding of the downstream molecular effects of MeCP2 loss of function to gain insight into RTT pathology.

CHAPTER 2:

Biotin tagging of MeCP2 reveals contextual insights into the Rett syndrome transcriptome

One Sentence Summary

This chapter summarizes our approach that enables cell type-specific biotinylation of MeCP2 across various contexts, revealing MeCP2 as a unique global modulator of gene transcription and pointing to new directions for translational research in Rett Syndrome.

Abstract

Mutations in *MECP2* cause Rett syndrome (RTT), an X-linked neurological disorder characterized by regressive loss of neurodevelopmental milestones and acquired psychomotor deficits. However, the cellular heterogeneity of the brain impedes understanding of how *MECP2* mutations contribute to RTT. We therefore developed cell type-specific biotin tagging of MeCP2 in mice bearing RTT-associated mutations and profiled their nuclear transcriptomes. Although most gene expression changes are largely specific to each mutation and cell type, lowly expressed cell type-enriched genes are preferentially disrupted by MeCP2 mutations, with upregulated and downregulated genes reflecting distinct functional categories. Subcellular RNA analysis in MeCP2 mutant neurons further reveals reductions in the nascent transcription of long genes and uncovers widespread post-transcriptional compensation at the cellular level. Finally, we overcame cellular mosaicism in female RTT models and identified distinct gene expression changes between neighboring wild-type and mutant neurons, altogether providing contextual insights into RTT etiology that support personalized therapeutic interventions.

Introduction

RTT is a progressive X-linked neurological disorder that represents one of the most common causes of intellectual disability among young girls. Patients experience a

characteristic loss of acquired social and psychomotor skills and develop stereotyped hand movements, hypotonia, breathing irregularities, and seizures after 6-18 months of normal development[1]. Approximately 95% of RTT cases are mapped to the X-linked gene encoding methyl-CpG binding protein 2 (MeCP2), a ubiquitously expressed protein that is highly enriched in postmitotic neurons[2,3]. The majority of RTT-associated mutations cluster within two functionally distinct domains of MeCP2. The Methyl-CpG Binding Domain (MBD) allows MeCP2 to recognize and bind to methylated cytosines[4]. The Transcriptional Repression Domain (TRD) mediates protein-protein interactions with histone deacetylase-containing co-repressors, such as the NCoR-SMRT and mSin3A complexes[5–7]. These domains support MeCP2 as a chromatin factor that mediates transcriptional repression[7,8], although transcriptional activation by MeCP2 has also been reported[9–11].

Different types of mutations in *MECP2*, together with random X-chromosome inactivation (XCI), are thought to underlie a spectrum of mild to severe clinical symptoms among RTT patients[12]. Mouse models carrying RTT-associated mutations recapitulate this phenotypic variability, although most studies are limited to hemizygous male mice[13–16]. Given that MeCP2 is a chromatin-bound nuclear protein, the identification of MeCP2 transcriptional targets in the brain remains key towards illuminating RTT etiology[17]. However, target identification is confounded by the cellular heterogeneity of the brain, which contains multiple intermixed cell types that differ in morphology, function, electrophysiological properties, and transcriptional programs[18–22]. Analyses using heterogeneous brain tissues obscures cell type-specific gene expression changes, impeding the assessment of MeCP2 function at the transcriptional level. The identification of transcriptional targets is further complicated by the widespread binding patterns of MeCP2 to methylated cytosines (mCpG and mCpA)[8,23,24], hydroxymethylated cytosines (hmCpG)[25], or unmethylated GC-rich regions[26] throughout the genome. Furthermore, although RTT predominantly affects heterozygous females, an experimental strategy to selectively identify gene expression changes from *Mecp2* mutant-expressing cells in a mosaic female brain has not yet been developed.

In this study, we addressed the confounding effects of cellular heterogeneity by engineering genetically modified mice whereby nuclear MeCP2 is labeled with biotin using Cre-Lox recombination. To understand the molecular impact of RTT-associated

mutations on cell type-specific gene expression *in vivo*, we also developed an allelic series of knockin mice bearing one of two frequent RTT missense mutations, T158M and R106W. When combined with Fluorescence-Activated Cell Sorting (FACS), this strategy effectively circumvents cellular heterogeneity in the brain and allows for the isolation of neuronal nuclei from targeted cell types. By examining MeCP2-mediated gene expression changes in different cell types, we identified underlying transcriptional features that correlate with the severity of the MeCP2 mutation. We also found that genome-wide transcriptional changes in the nucleus are opposed by post-transcriptional compensation of RNAs in a gene length-dependent manner. Furthermore, our approach allows us to circumvent cellular mosaicism and profile the transcriptome of neighboring wild-type (WT) and mutant neurons in females, thereby discerning cell and non-cell autonomous transcriptional effects. This comprehensive study across different neuronal settings allows us to propose a contextualized model by which cell and non-cell autonomous transcriptional changes in different cell types contribute to the molecular severity of neuronal deficits observed in RTT, uncovering new directions for therapeutic development.

Results

Engineering a System to Genetically Biotinylate MeCP2 In Vivo

Biotin-mediated affinity tagging has been widely utilized in cell and animal models for multiple experimental approaches because of the strong (K_d = 4x10^{-14}M) and specific interaction between biotin and avidin protein[27]. We exploited this approach to investigate MeCP2 function by using homologous recombination to insert a short 23-amino acid affinity tag immediately upstream of the *Mecp2* stop codon (Fig. 1a and Supplementary Fig. 1a). This tag comprises a TEV protease cleavage site and a 15-amino acid biotinylation consensus motif (termed Tavi, *T*EV and *Avi*din-binding) that can be post-translationally labeled with biotin by the *E. coli* biotin ligase, BirA. To biotinylate the tag in cell types of interest, we also generated Cre-dependent BirA transgenic mice (herein *R26*$^{STOP\text{-}BirA}$; Supplementary Fig. 1b). Therefore, upon crossing mice to a cell type-specific Cre line, BirA is expressed and subsequently biotinylates MeCP2-Tavi (Fig.

1b). We used *EIIa-Cre*[28] to ubiquitously express BirA (herein *R26*^{BirA}) and confirmed that MeCP2 is specifically biotinylated *in vivo* under conditions where BirA is expressed and the Tavi tag is present (*R26*$^{BirA/+}$;*Mecp2*$^{Tavi/y}$; Fig. 1c and Supplementary Fig. 1c).

To exclude the possibility that tagging MeCP2 adversely affects its molecular function, we assessed MeCP2 RNA and protein expression levels from age-matched *Mecp2*$^{Tavi/y}$ (herein TAVI) and *Mecp2*$^{+/y}$ mice (herein WT; Supplementary Figs. 1d-f). We also examined MeCP2 binding to methylated DNA and its known protein interactions. We found that tagged and untagged MeCP2 both exhibit similar levels of binding at methylated major satellite repeats and IAP elements (Supplementary Fig. 1g), and MeCP2-Tavi remains associated with NCoR-SMRT (Supplementary Fig. 1h). We also found that MeCP2-Tavi protein, but not RNA, is reduced by ~40% when compared to untagged MeCP2 in mice at 20 weeks of age (Supplementary Fig. 1e-f). Given that we observed similar levels of MeCP2 chromatin binding between WT and TAVI mice, the diminution likely affects soluble, rather than chromatin-bound, MeCP2. Notably, a 50% reduction in MeCP2 expression leads to hypoactivity and altered behavioral phenotypes in mice by one year of age[29]. Despite reduced MeCP2 protein expression, we found that TAVI mice are phenotypically indistinguishable from WT mice (Fig. 1g-j and Supplementary Fig. 1i), and do not display RTT-like features over an observational period of 20 weeks (Fig. 1i; data not shown). These data support that TAVI mice appear functionally equivalent to WT mice, at least up to 20 weeks of age.

MeCP2 Missense Mutations Recapitulate RTT-like Phenotypes in Mice

Four of the eight most frequent RTT mutations are missense mutations, three of which are located in the MBD of MeCP2 and include R106W (2.76% of RTT patients), R133C (4.24%), and T158M (8.79%)[30]. Typical RTT patients bearing the R133C mutation display milder clinical symptoms, whereas patients carrying the T158M or R106W mutation exhibit moderate or severe symptoms, respectively[12]. Although the clinical severity of these mutations scales with their effects on MeCP2 binding affinity to methylated DNA[13,31–33], this relationship is not fully understood on a molecular level.

We thus generated independent *Mecp2*$^{T158M\text{-}Tavi}$ (herein T158M) and *Mecp2*$^{R106W\text{-}Tavi}$ (herein R106W) knock-in mice in parallel with TAVI mice (Fig. 1a). Relative to TAVI

mice, we found that both *Mecp2* mutant mice display significant reductions in MeCP2 protein expression despite comparable levels of mRNA (Fig. 1e-f and Supplementary Fig. 1j), similar to other RTT mutations in the MBD[13,14]. We further performed MeCP2 immunofluorescent staining (IF) using hippocampal sections of BirA-expressing mice to examine the subcellular localization of MeCP2 mutant protein. In contrast to MeCP2 WT and TAVI protein, MeCP2 T158M and R106W are diffuse throughout the nucleus and not properly localized to heterochromatic foci, indicative of a reduced capacity to bind mCpGs *in vivo* (Fig. 1d). Streptavidin IF is also noticeably reduced in *Mecp2* mutant mice, likely due to the reduced protein stability associated with these mutations (Fig. 1d and Supplementary Fig. 1k). Although streptavidin IF confirmed a loss of MeCP2 localization to heterochromatic foci, this channel also revealed a portion of mutant MeCP2 redistributed to the nucleolus (Supplementary Fig. 1k-l), similar to GFP-tagged MeCP2 upon deletion of the MBD[34]. Upon site-by-site comparison, we found that T158M and R106W mice also exhibit RTT-like phenotypes similar to that of *Mecp2*-null mice, including decreased brain and body weight, and an age-dependent increase in phenotypic score (Fig. 1g-i). Although lifespan is significantly reduced in all three *Mecp2* mutant mice, the median survival of R106W mice more closely resembles that of *Mecp2*-null than T158M mice (Fig. 1j). Statistical analysis revealed a significant difference in the survival curves of T158M (median survival = 14 weeks) and R106W mice (median survival = 10 weeks; Mantel Cox $P = 0.012$). Thus different mutations within the MBD differentially affect the severity of RTT-like phenotypes in mice.

Given that MeCP2 T158M is known to retain partial affinity for mCpGs, and that our T158M mice exhibit higher MeCP2 protein levels and longer survivability when compared to R106W mice, we infer that this mutation represents a partial loss-of-function. In contrast, the R106W mutation, which abolishes affinity for mCpGs, exhibits lower MeCP2 protein levels and lower survivability on par with *Mecp2*-null mice, represents a complete loss-of-function.

Genetic Biotinylation Permits Cell Type-specific Transcriptional Profiling

Given MeCP2's abundant expression throughout the CNS, we devised a biotinylation-based strategy for cell type-specific nuclei isolation and transcriptional profiling (Fig. 2a-

b). As an example, we used the *NeuroD6/NEX-Cre* line[35] to drive BirA expression and MeCP2-Tavi biotinylation in forebrain excitatory neurons (Fig. 2a and Supplementary Fig. 2a-h). Quantification of pan-neuronal (NeuN), pan-inhibitory (GAD67), and inhibitory-specific (parvalbumin, somatostatin and calretinin) neuronal markers in the somatosensory cortex of $Mecp2^{Tavi/y}$;$R26^{cBirA/+}$;$NEX^{Cre/+}$ (NEX-Cre) mice demonstrated that biotinylation occurs in ~80% of NeuN+ cortical neurons devoid of inhibitory markers, consistent with NEX-Cre-mediated recombination in excitatory neurons (Supplementary Fig. 2h). FACS using stained cortical nuclei from NEX-Cre mice identified three distinct nuclear populations (Fig. 2c). RT-PCR for cell type-specific markers confirmed that NeuN+Biotin+ nuclei reflect excitatory neurons, whereas NeuN+Biotin- nuclei represent a mixture of inhibitory interneuron subtypes (Fig. 2c-d). Excitatory and inhibitory populations are both depleted of astrocytic, microglial and oligodendrocytic markers, which are restricted to the third, non-neuronal population of NeuN-Biotin- nuclei (Fig. 2d). We also used the Dlx5/6-Cre line[36] to drive BirA expression in forebrain GABAergic neurons and obtained results inverse to that of NEX-Cre mice (Fig. 2a and Supplementary Fig. 2a-j), confirming that MeCP2-Tavi is reliably biotinylated in Cre-defined cell types.

Having established a cell type-specific nuclei isolation approach, we next performed transcriptional profiling in mice near the onset of RTT-like phenotypes. We employed the NEX-Cre driver and isolated 120,000 or 250,000 cortical excitatory and inhibitory nuclei from 6-week male T158M, R106W and TAVI mice via FACS, followed by total RNA-seq. We found biological replicates to be well correlated (Fig. 2e and Supplementary Fig. 3a), and ~74% of total reads mapped to introns. Intron-mapped reads represent chromatin-associated primary transcripts that are commonly used as a proxy for transcriptional activity[37,38]. Because MeCP2 is thought to function as a transcriptional modulator[17], nuclear RNA-seq thus affords an unique opportunity to study the primary effects of RTT mutations on gene expression.

We first analyzed gene expression profiles from cortical excitatory and inhibitory neurons in TAVI mice to confirm that a nuclear transcriptome analysis of genic-mapped reads can be used to study cell types. Unsupervised hierarchical clustering shows that replicate transcriptomes are highly correlated by cell type, and genic-mapped reads illustrate genes that are selectively expressed for each cell type (Fig. 2g and

Supplementary Fig. 3a). We identified 9,379 differentially expressed genes (DEGs, FDR < 0.05) between excitatory and inhibitory neurons, the majority (86.9%) of which comprise protein-coding genes (Fig. 2f and Supplementary Fig. 3b). Approximately half of these cell type-enriched genes display Gene Ontology (GO) functions consistent with glutamatergic pyramidal cell types (herein EXC-enriched; Supplementary Fig. 3c), including cell signaling and post-synaptic functions, whereas the remaining half exhibit GO functions consistent with metabolically active GABAergic interneurons (herein INH-enriched; Supplementary Fig. 3d), including cellular respiration and mitochondrial function. These functional associations demonstrate that total RNA-seq from cell type-specific nuclei discerns known functional differences between excitatory and inhibitory neurons.

Protein-Coding Genes are More Severely Affected in R106W Mice

We next compared nuclear gene expression profiles in excitatory and inhibitory neurons between 6-week old mutant (T158M, R106W) and control (TAVI) mice to identify and characterize DEGs associated with the appearance of RTT-like phenotypes (Fig. 3a). We identified more DEGs in R106W excitatory and inhibitory neurons than T158M neurons, indicating that the number of misregulated genes positively scales with the severity of the *Mecp2* mutation (Fig. 3b). More than 90% of MeCP2 DEGs are protein-coding genes (Supplementary Fig. 4a), significantly higher than the percentage of protein-coding genes from genomic (60.4%), actively expressed (77.7-78.3%) and cell type-enriched (86.2-87.7%) gene distributions (Supplementary Fig. 3b). We therefore excluded non-coding genes from further analyses. We note that the number and percentage of protein-coding DEGs overlapping between T158M and R106W genotypes is greater in inhibitory (74.8% of T158M DEGs) than excitatory neurons (40% of T158M DEGs; Fig. 3c). Moreover, overlapping DEGs tend to be misregulated in the same direction (Fig. 3c).

The median fold change of T158M and R106W DEGs is consistently small in mutant neurons, particularly when compared to overall differences in gene expression between excitatory and inhibitory neurons. Because the R106W mutation leads to a more severe phenotype, we examined whether this mutation impacts gene expression to a greater extent than the T158M mutation. We compared fold changes between T158M

and R106W DEGs, limiting our analysis to protein-coding genes that overlap between genotypes to account for disproportionate numbers of DEGs. Within this subset, the median fold change among upregulated and downregulated DEGs is consistently higher in both cell types of R106W mice than those of T158M mice (Fig. 3d). The differences between mutations are more apparent in inhibitory neurons as the cumulative fold change in R106W neurons is significantly shifted to the right of that in T158M neurons (Fig. 3e). A similar trend is also observed in excitatory neurons (Fig. 3e).

These data suggest that the number of genes and the degree to which they are misregulated positively correlate with RTT phenotypic severity, reflecting the molecular consequence of differentially impaired MeCP2 binding to methylated DNA. Moreover, inhibitory neurons appear to be more sensitive to the transcriptional effects of such mutations than excitatory neurons.

Transcriptional Features of T158M and R106W DEGs

The genetic removal of MeCP2 from different brain regions or cell types in mice is associated with distinct, non-overlapping phenotypes that reflect the specific neural circuit being impaired[16,17]. In a similar fashion, we investigated whether MeCP2-dependent transcriptional changes are specific to each cell type. Indeed, we found only 20 T158M DEGs (6.2% of total) and 114 R106W DEGs (10.7% of total) that overlap between excitatory and inhibitory neurons (Fig. 3f), similar to a study using *Mecp2*-null mice[39]. This degree of overlap is unexpectedly low, given that 74.2% of expressed genes are shared between excitatory and inhibitory neurons (data not shown). The finding that MeCP2-dependent gene expression changes are specific to each cell type prompted us to examine the association between MeCP2 DEGs and EXC/INH-enriched genes. We found that cell type-enriched genes are significantly overrepresented and comprise 70-80% of DEGs in each cell type (Fig. 3g). Moreover, EXC- and INH-enriched genes are preferentially downregulated and upregulated, respectively, in each cell type (Figure 3h).

Furthermore, upon examination of relative expression levels of MeCP2 DEGs using Fragments Per Kilobase of transcript per Million mapped reads (FPKM), we found that T158M, R106W, and overlapping DEGs all displayed significantly lower median

FPKM values than overall expressed genes in each cell type (Fig. 3i). To confirm whether low-expressing genes are specifically enriched for MeCP2 DEGs, we divided actively expressed genes from each cell type into four equally sized bins according to the FPKM expression level. Among these bins, EXC- and INH-enriched genes display expected distributions for each cell type (Supplementary Fig. 4b). In contrast, T158M and R106W DEGs are preferentially enriched in Q1, the bottom 25th percentile of actively expressed genes, in both excitatory and inhibitory neurons (Fisher Exact one-tailed *P*, T158M EXC = 1.11e-07, T158M INH = 2.03e-04, R106W EXC = 4.04e-08, R106W INH = 1.50e-02; Supplementary Fig. 4b). Between both mutations, we found that T158M DEGs are more likely to be enriched in Q1 (Fisher Exact Odds Ratio (OR) for Q1, T158M EXC = 3.1, T158M INH = 3.2) than R106W DEGs (Fisher OR for Q1, R106W EXC = 2.0, R106W INH = 1.3). Accordingly, R106W DEGs consistently display significantly higher FPKM values than T158M DEGs in both cell types, and these higher FPKM genes are predominantly downregulated in R106W neurons (Fig. 3i and Supplementary Fig. 4c).

Thus, in excitatory and inhibitory neurons of mice at 6 weeks of age, low-expressing cell type-enriched genes are particularly sensitive to MeCP2 dysfunction regardless of mutation and cell type. However, when compared to the T158M mutation, the increased severity of the R106W mutation is consequently associated with an increased number of high-expressing genes that are preferentially downregulated in both cell types.

Upregulated and Downregulated Genes Demarcate Distinct Cellular Functions

We noticed that DEGs are preferentially downregulated in excitatory neurons and upregulated in inhibitory neurons, and this trend is preserved among the DEGs shared between both mutations (Fig. 3c). The increased severity of the R106W mutation also correlates with increased numbers of high-expressing DEGs that are preferentially downregulated, whereas upregulated DEGs similarly comprise low-expressing genes in both mutants (Supplementary Fig. 4c). Given these characteristics, we next performed a pre-ranked Gene Set Enrichment Analysis (GSEA, FDR < 0.1) to determine whether upregulated and downregulated DEGs represent functionally distinct categories. We found that upregulated genes in T158M and R106W mice are both primarily associated

with transcriptional regulation (Fig. 3j). These include DNA-binding transcriptional activators, co-activators, repressors, and chromatin remodelers, most of which tend to be INH-enriched genes (Fig. 3j and Supplementary Fig. 4d). Significant functional categories associated with downregulated genes, however, are specifically detected in R106W excitatory neurons and enriched for post-synaptic membrane proteins, including Na^{2+}, K^{+}, Ca^{2+} and Cl^{-} channels, synaptic scaffolding proteins, and ionotropic glutamate receptors (Fig. 3j). The specific loss of these synaptic functions appears to be consistent with the preferential downregulation of high-expressing genes in R106W excitatory neurons and the development of more severe phenotypes in R106W than T158M mice. Although significant gene functions were not identified among downregulated DEGs in inhibitory neurons using our established GSEA FDR cutoff, gene functions associated with upregulated DEGs in R106W inhibitory neurons are related to cellular metabolism and signal transducer activity (Supplementary Fig. 4e).

Subcellular RNA Fractions Reveal Global Transcriptional and Post-transcriptional Changes

Several recent reports implicate MeCP2 in the transcriptional regulation of long genes, which are preferentially upregulated in the neurons of multiple RTT animal models[23,39]. We therefore examined the possibility that genome-wide transcriptional changes may correlate with T158M and R106W phenotypic and molecular severity. Similar to these studies, we sorted and binned expressed protein-coding genes according to gene length and measured the mean fold change in *Mecp2* mutant neurons at 6 weeks of age. Nuclear transcriptomes revealed a striking inversion of previously reported gene expression changes whereby short (≤ 100kb in gene length) and long (> 100kb in gene length) genes are upregulated and downregulated, respectively, in a length-dependent manner (Supplementary Fig. 5a).

Although most nuclear RNAs comprise intron-containing pre-mRNA transcripts on chromatin, the presence of processed mRNA transcripts awaiting nuclear export may confound the assessment of transcriptional events using nuclear RNA alone[40]. We therefore performed global nuclear run-on with high-throughput sequencing (GRO-seq[41]) to directly assess *de novo* transcriptional activity by RNA polymerase in cortical

nuclei of TAVI and R106W mice. Similar to sorted nuclear RNA, the nascent transcription of short and long genes in R106W neurons is predominantly increased and decreased, respectively (Fig. 4a). LOESS local regression of DEGs that were identified in R106W excitatory and inhibitory neurons, which are individually upregulated or downregulated, also revealed a similar overall trend towards the preferential downregulation of long genes (Fig. 4a). The genome-wide trend we observe in sorted nuclear RNA thus represents a primary effect at the transcriptional level, prompting us to further investigate if the length-dependent upregulation of long genes that was previously reported may represent an indirect effect of MeCP2-dependent transcriptional deregulation. To test this, we resected cortical tissue from TAVI and R106W mice at 6 weeks of age. Each cortical half was used to isolate whole cell RNA or nuclear RNA in parallel, followed by sequencing. Whole cell RNA from mutant cortices display a length-dependent increase in the mean expression of long genes (Fig. 4b), similar to what was previously described[23,39]. In contrast, cortical nuclear RNA isolated from the same TAVI and R106W mice exhibited a length-dependent upregulation of short genes and downregulation of long genes genome-wide (Fig. 4c), corroborating the transcriptional changes we observed from nascent RNA (Fig. 4a) and sorted nuclear RNA preparations (Supplementary Fig. 5a). Using the 10,390 expressed genes associated with *de novo* transcription by GRO-seq (Fig. 4d), we observed that genes upregulated in nascent and nuclear RNA fractions were cumulatively shorter in length relative to those upregulated in whole cell RNA, and the inverse was observed among downregulated genes (Fig. 4e). Together, these data demonstrate that gene expression changes in *Mecp2*-mutant neurons are substantially different between subcellular RNA fractions.

To directly compare individual genes across subcellular fractions, we next classified all 10,390 expressed genes into eight groups depending on the direction in which genes are misregulated across fractions. Groups B and D comprise 38.4% of expressed genes and represent expression changes that are misregulated in the same direction across nascent, nuclear, and whole cell RNA fractions (Fig. 4f). Among these groups of genes, $\log_2$ fold changes measured from the whole cell are significantly smaller than fold changes in the nuclear compartment, suggesting that gene expression changes in the nucleus are post-transcriptionally minimized in the cell (Fig. 4f and Supplementary Fig. 5b). The majority of genes (48%), however, exhibit expression

changes in nuclei that are significantly inverted in whole cell RNA (Groups A,C,G,H; Fig. 4f). Groups A and C consist of relatively long, EXC-enriched genes that are transcriptionally downregulated in nascent RNA but post-transcriptionally upregulated using whole cell RNA (Fig. 4f and Supplementary Fig. 5c-d). DAVID gene ontology revealed that Group A genes are associated with synaptic functions and protein phosphorylation signaling at the plasma membrane (Fig. 4g). Groups G and H consist of considerably shorter, INH-enriched genes that are transcriptionally upregulated in nascent RNA but post-transcriptionally downregulated in whole cell RNA (Fig. 4f and Supplementary Fig. 5c-d). Notably, Group G genes are functionally associated with cellular energy and metabolism in mitochondria (Fig. 4g). Finally, we used the GRO-seq dataset to also filter for genes in age-matched T158M and R106W sorted neurons that are associated with *de novo* transcriptional activity in cortical nuclei (Supplementary Fig. 5e). This revealed a trend towards long genes being more severely downregulated in both excitatory and inhibitory neurons bearing the R106W mutation when compared to the T158M mutation (Fig. 4h and Supplementary Fig. 5f-g).

Taken together, these data support that global gene expression changes in *Mecp2*-mutant mice differ between subcellular RNA fractions, and that many long genes are transcriptionally downregulated in the nucleus of *Mecp2* mutant neurons. Furthermore, the molecular severity of missense mutations correlates with the extent to which these long genes are downregulated in both excitatory and inhibitory neurons at 6 weeks of age.

Female RTT Mouse Models Reveal Cell and Non-Cell Autonomous DEGs

RTT is an X-linked disorder that primarily affects heterozygous females. However, the extent to which intermixed *Mecp2* WT and mutant (MUT) neurons in cellular mosaic RTT females affect each other at the level of gene expression remains unknown. The reduced stability and expression of T158M and R106W mutant protein allowed us to use our genetic tagging and sorting strategy to isolate and profile WT (denoted by subscript: $T158M_{WT}$, $R106W_{WT}$) and MUT (denoted by subscript: $T158M_{MUT}$, $R106W_{MUT}$) excitatory neurons from mosaic female mice. By comparing the gene expression profiles of WT or MUT neurons from heterozygous mutant mice to those from control mice ($TAVI_{WT}$), we

could discern both cell and non-cell autonomous gene expression changes as a result of MeCP2 mutations. We thus generated TAVI (*Mecp2*$^{Tavi/+}$;*R26*$^{cBirA/+}$;*NEX*$^{Cre/+}$), T158M (*Mecp2*$^{Tavi/T158M\text{-}Tavi}$;*R26*$^{cBirA/+}$;*NEX*$^{Cre/+}$), and R106W (*Mecp2*$^{Tavi/R106W\text{-}Tavi}$;*R26*$^{cBirA/+}$;*NEX*$^{Cre/+}$) mosaic females that each carry one copy of the Tavi-tagged WT allele and one copy of the tagged T158M, tagged R106W, or untagged WT allele.

We first aged females to ~18 weeks, when T158M and R106W females both display RTT-like phenotypes relative to TAVI females (Fig. 5a). We next subjected female cortical nuclei for FACS isolation (Fig. 5b-c and Supplementary Fig. 6a). From the number of females sampled, we did not detect skewed XCI (> 75%) among excitatory neurons in TAVI, T158M, or R106W mice (Fig. 5d). When compared to $TAVI_{WT}$ neurons, we identified a total of 526 and 678 unique protein-coding DEGs in $T158M_{WT\ and\ MUT}$ and $R106W_{WT\ and\ MUT}$ neurons, respectively (Fig. 5e and Supplementary Fig. 6b). Most DEGs represent cell autonomous gene expression changes that occur in mutant neurons alone (Fig. 5e). However, R106W DEGs contain a larger proportion of indirect DEGs that are found in both WT and MUT neurons (43.4%; Fig. 5e), revealing a mutation-specific susceptibility of WT neurons to non-cell autonomous gene expression changes in heterozygous females. We further visualized differences in transcriptomes using principal component analysis (PCA) to plot the first two major axes of variation (PC1 vs. PC2; Fig. 5f). PC2 separates neuronal populations by *Mecp2* allele status (WT vs. MUT neurons) irrespective of genotype, indicating that *Mecp2* mutations induce cell autonomous changes that are transcriptionally distinct from neighboring wild-type neurons. However, PC1 accounts for twice the variation as PC2 and clusters R106W-derived populations away from other genotypes, revealing the additive extent to which indirect DEGs associate with this mutation. Against PC1 and PC2, $T158M_{WT}$ neurons closely resemble $TAVI_{WT}$, suggesting that the mere presence of mutant neurons is not sufficient to generate the indirect gene expression changes apparent in R106W mice (Fig. 5f). Thus, indirect DEGs are likely a result of the increased molecular severity associated with the R106W mutation.

We found that 194 DEGs overlap between T158M and R106W female mice, most of which are misregulated in the same direction (Fig. 5g). Among these genes, cell autonomous transcriptional changes (149 genes, 76.8%) are more likely to be shared across independent *Mecp2* mutations than non-cell autonomous changes (9 genes, 4.6%;

Fig. 5h). These overlapping DEGs also show higher fold changes in R106W than T158M female mice, similar to observations from male mice (Fig. 5i). However, this difference is mainly driven by indirect DEGs in R106W neurons (Fig. 5h).

Using the R106W female mice to further characterize features that distinguish *Mecp2*-dependent DEGs in WT and MUT neurons, we noticed that non-cell autonomous DEGs are predominantly upregulated (~60%) in contrast to cell autonomous DEGs (~48%; Supplementary Fig. 6b). The absolute fold change among cell autonomous gene expression changes is also significantly smaller than non-cell autonomous changes, particularly among upregulated genes (Supplementary Fig. 6c). Furthermore, cell autonomous DEGs are considerably longer in gene length, specifically among upregulated genes (Supplementary Fig. 6d). To determine if cell and non-cell autonomous DEGs represent distinct biological processes, we next performed pre-ranked GSEA (FDR < 0.1) and found that non-cell autonomous gene expression changes primarily affect cell-to-cell signaling and negative regulation of protein-kinases (Supplementary Fig. 6e). Indeed, these DEGs include several immediate early and late response genes that are induced by neuronal activity and modulate signaling pathways associated with synaptic plasticity[42]. In contrast, cell autonomous DEGs are significantly associated with transcriptional regulation (Supplementary Fig. 6f). These functional categories demonstrate a marked resemblance to those observed in excitatory neurons of male T158M and R106W mice (Fig. 5j). The striking consistency with which these functional annotations characterize *Mecp2*-mutant neurons, despite apparent differences in age and sex, supports the cell autonomous disruption of these functions as a key, contributing factor to RTT pathogenesis.

Discussion

The complexity of MeCP2 molecular function, coupled with the cellular heterogeneity of the brain, complicates the identification and interpretation of transcriptional changes in RTT. To overcome these challenges, we developed a genetic strategy to biotinylate MeCP2 and its mutant variants in different cell types of adult mice. This strategy couples *in vivo* biotinylation with Cre-Lox technology and extends the use of the Tavi tag for cell type-specific biochemical purification and molecular profiling studies[43]. Notably, the

small size of the Tavi tag makes it ideal for targeted gene insertion using CRISPR-Cas9 technology. Our genetic approach can thus be expanded to any protein-coding gene of interest for cell type-specific *in vivo* applications beyond those conveyed in this study.

By using an allelic series of mutations in mice to perform a transcriptome analysis of cortical neurons that vary by cell type, subcellular compartment, and sex, we discovered underlying transcriptional features that correlate with impairments in MeCP2 binding to chromatin and have therapeutic implications for RTT. Our characterization of T158M and R106W mice demonstrates that both mutations are associated with a similar age-dependent onset and progression of RTT-like phenotypes. We also found that both mutations exhibit similar molecular features among upregulated genes, including those in male and female *Mecp2*-mutant neurons that encode INH-enriched transcription factors and chromatin remodelers. Because T158M and R106W mutations both impair MeCP2 binding to chromatin and lead to RTT-like phenotypes, the misexpression of transcriptional regulators may contribute to shared etiology in T158M and R106W mice. In support, both GRO-seq and nuclear RNA-seq show a trend towards the increased transcription of short, INH-enriched genes across the genome in T158M and R106W mice. Many of these genes are functionally associated with cellular respiration and energy metabolism, which could provide a transcriptional basis for several clinical features shared among RTT patients that notably resemble mitochondrial and metabolic disorders[44].

We also show that lowly-expressed, cell type-enriched genes are sensitive to the effects of MeCP2 dysfunction, which likely contributes to the specificity of MeCP2-mediated gene expression changes in different neuronal cell types. However, regardless of cell type, the R106W mutation affects a larger number of genes that tend to be both highly expressed and preferentially downregulated relative to the T158M mutation, consistent with the greater impairment of MeCP2 R106W binding to methylated DNA[32]. GRO-seq and nuclear RNA-seq further demonstrates that transcriptional differences between these two mutations extend to most long genes throughout the genome, which tend to be highly expressed in neurons[45]. Our data is thus in partial agreement with global reductions of Ser5-phosphorylated RNA polymerase in *Mecp2*-null neuronal nuclei[46] and supports MeCP2 as a global modulator of gene transcription. Given that transcription is impeded by chromatin-mediated physical constraints[47], loss of MeCP2

occupancy may induce alterations to chromatin structure or organization that decreases the efficiency of transcriptional elongation at long genes. Alternatively, the loss of MeCP2 binding may reduce HDAC3-mediated transcription factor de-acetylation that could be required for long gene transcriptional activation[48]. Of note, downregulated genes are associated with synaptic morphology and function, which is stereotypically reduced in *Mecp2*-deficient neurons[16]. Because R106W mice also display reduced lifespan compared to T158M mice, it is possible that reductions in long gene transcription may act as a modifier in specific neuronal cell types, worsening subsets of RTT-like phenotypes. RTT patients with mutations that preserve MeCP2 binding do exhibit milder clinical features than patients for whom binding is disrupted[12]. Additional transcriptional assessments of milder mutations that preserve MeCP2 binding to chromatin, coupled with our biotin tagging and nuclei sorting approach, are necessary to further standardize and refine genotype-phenotype correlations at the transcriptional level.

RTT transcriptional changes in neuronal nuclei complement the reported upregulation of long genes across multiple RTT mouse models[23,39]. By analyzing subcellular distributions of RNA, we found that the upregulation of most long genes in *Mecp2* mutant neurons is absent in nuclear and nascent RNA but present in whole cell RNA. Notably, whole cell RNA is enriched for cytoplasmic mRNAs whose steady-state abundance and turnover is modulated by post-transcriptional regulatory mechanisms, including RNA nuclear retention, miRNA-mediated decay, or sequestration into cytoplasmic RNA granules[40,49]. Gene expression changes using whole cell RNA may thus be compensatory and not fully reflective of transcriptional activity. This aptly questions the therapeutic benefit surrounding the use of small molecules that decrease long gene transcription for treating RTT patients. Rather, fold changes across subcellular RNA fractions appear consistent with post-transcriptional mechanisms that could abate cellular consequences arising from global alterations in synaptic, mitochondrial, and metabolic gene transcription. Identifying molecular players that underlie a cellular compensation of RTT-associated transcriptional changes may yield a novel class of interventional therapies that can be administered prior to or during the initial regression phase of RTT, minimizing its pathological impact during neurodevelopment.

Importantly, our approach allows for the novel isolation and molecular examination of *Mecp2* WT and MUT neurons from cellular mosaic female mice, which represent more accurate pre-clinical models of RTT. We demonstrate that WT neurons are also susceptible to the effects of *Mecp2* mutations in neighboring cells, and that these non-cell autonomous gene expression changes are dependent on the molecular severity of the MeCP2 mutation. Additionally, non-cell autonomous gene expression changes also occur in *Mecp2* MUT neurons, indicating that DEGs in RTT neurons arise from MeCP2-dependent and independent molecular processes. Further investigation, encompassing a wide range of X-inactivation ratios across multiple ages, cell types, and *Mecp2* mutations, are thus required to elucidate direct and indirect contributions to RTT.

The non-cell autonomous gene expression changes we found in R106W females markedly include genes whose expression is stereotypically induced by neuronal activity. These genes are known to negatively modulate synaptic plasticity by reducing responsiveness to excessive neuronal stimuli[42]. The selective upregulation of these indirect genes in WT and MUT neurons of R106W, but not T158M, mice may be a response to increased neuronal activity among severely affected mosaic neurons. This would be strikingly consistent with morphological reductions in dendritic branching and synaptic spine density exhibited by both WT and MUT neurons of heterozygous *Mecp2*-null female mice[50]. Accordingly, electrophysiological abnormalities observed in phenotypically severe *Mecp2*-null mice might not fully reflect the milder spectrum of RTT-associated mutations. Furthermore, two of these late-response genes, *Bdnf* and *Igf1*, are indirectly upregulated in both WT and MUT neurons of 18-week old females and encode neuroprotective peptides that improve RTT synaptic, cellular, and behavioral deficits[16,51]. Because these molecules are in active clinical trials[52], pathways associated with non-cell autonomous DEGs may act to ameliorate RTT-associated neuronal phenotypes and could represent a novel source of therapeutic targets for treating RTT patients.

Author Contributions

Conceptualization, B.S.J. and Z.Z.; Methodology, B.S.J., Y.Z., M.F., J.M.L., D.G., and Z.Z.; Investigation, B.S.J., Y.Z., M.F., J.M.L., K.H.W., Y.J.K., and D.B.; Formal Analyses, B.S.J., Y.Z., and T.H.K.; Validation, G.V.; Resources, B.S.J., Y.Z. and Y.C.; Data Curation,

Y.Z.; Writing – Original Draft, B.S.J.; Writing – Review & Editing, B.S.J., Y.Z., M.F., G.V., T.H.K. and Z.Z.; Visualization, B.S.

Materials and Methods

Generation of Mouse Lines

The targeting construct used for homologous recombination at the *Mecp2* locus in murine ES cells was cloned in two arms by PCR amplification of sv129 genomic DNA. The 5′ arm was PCR amplified with 5′-AGGAGGTAGGTGGCATCCTT-3′ and 5′-CGTTTGATCACCATGACCTG-3′ primers, whereas the 3′ arm was PCR amplified with 5′-GAAATGGCTTCCCAAAAAGG-3′ and 5′-AAAACGGCACCCAAAGTG-3′ primers. Restriction sites at the ends of each arm were created using nested primers for cloning into a vector containing a *loxP*-flanked neomycin cassette (Neo) and a diphtheria toxin A negative-selection cassette. QuikChange (Stratagene) insertional mutagenesis was used to generate the *Mecp2-Tavi* targeting construct by inserting the Tavi tag immediately upstream of the *Mecp2* stop codon within the 5' arm.

The portion of the Tavi tag containing the biotinylation consensus sequenced flanked by 5' NaeI and 3' BspHI restriction sites was inserted through two rounds of mutagenesis:

Round 1 Forward: 5'-GACCGAGAGAGTTAGCGCCGGCCTGAACGACATCTTCGAGTCATGACTTTACATAGAGCG-3'

Round 1 Reverse: 5'-CGCTCTATGTAAAGTCATGACTCGAAGATGTCGTTCAGGCCGGCGCTAACTCTCTCGGTC-3'

Round 2 Forward: 5'-CTGAACGACATCTTCGAGGCTCAGAAAATCGAATGGCACGAATCATGACTTTACATAGAG-3'

Round 2 Reverse: 5'-CTCTATGTAAAGTCATGATTCGTGCCATTCGATTTTCTGAGCCTCGAAGATGTCGTTCAG-3'

The portion of the tag containing the TEV protease cleavage site was inserted upstream of the NaeI restriction site with a third round of mutagenesis:

Round 3 Forward: 5'-
GACCGAGAGAGTTAGCGAAAACCTGTATTTTCAGGGCGCCGGCCTGAACGACATC-3'

Round 3 Reverse: 5'-

GATGTCGTTCAGGCCGGCGCCCTGAAAATACAGGTTTTCGCTAACTCTCTCGGTC-3'

To generate *Mecp2-Tavi* targeting constructs bearing independent RTT-associated point mutations, QuikChange site-directed mutagenesis was used to mutate MeCP2 arginine 106 to tryptophan and MeCP2 threonine 158 to methionine within the 3'arm and 5'arm, respectively. A single nucleotide at codon T160 also underwent site-directed mutagenesis for a silent mutation to introduce a BstEII restriction site to correctly identify targeted ES cells.

To generate conditional *BirA* transgenic mice, PCR primers containing AscI restriction sites and a Kozak consensus sequence were used to subclone the *BirA* coding sequence and insert it downstream of both a CAG promoter and a floxed transcriptional attenuator, *Neo-STOP*, within pROSA26-1, a transgenic targeting vector that has previously been characterized[53].

After confirmation by Sanger sequencing and linearization with NotI (*Mecp2-Tavi* targeting construct and its mutant variants) or SgfI (*BirA* targeting construct), the constructs were electroporated into sv129-derived murine ES cells. Correctly targeted ES cells were independently injected into C57BL/6 blastocysts and subsequently implanted into pseudopregnant females. Agouti offspring were screened by southern blot and PCR genotyping to confirm germline transmission of the *Mecp2-Tavi*, *Mecp2^{T158M}-Tavi*, *Mecp2^{R106W}-Tavi*, and *R26^{cBirA}* alleles. In the case of the *Mecp2-Tavi* allele and its mutant variants, the resulting offspring were mated with C57BL/6 *EIIa-cre* mice to ensure germline deletion of the floxed Neo cassette between *Mecp2* exons 3 and 4.

Additional Mouse lines

Dlx5/6-Cre (Stock 008199) and *EIIa-Cre* (Stock 003724) mice were obtained from Jackson Laboratories[28,36]. *NeuroD6/NEX-Cre* mice were obtained with permission from the Nave Laboratory[35].

Animal Husbandry

Experiments were conducted in accordance with the ethical guidelines of the US National Institutes of Health and with the approval of the Institutional Animal Care and Use Committee of the University of Pennsylvania. All of the experiments described were performed using mice on a congenic sv129:C57BL/6J background with the knock-in/transgenic alleles backcrossed to C57BL/6J mice (Charles River) for at least five generations, unless otherwise stated. Mice were housed in a standard 12h light/12h dark cycle with access to ample amounts of food and water. Mice bearing the Tavi tag were genotyped using a bipartite primer PCR-based strategy to detect the Tavi tag at the 3'-end of the endogenous *Mecp2* gene (Forward: 5'-CACCCCGAAGCCACGAAACTC-3', Reverse: 5'-TAAGACTCAGCCTATGGTCGCC-3') and give rise to a 318-bp product from the wild-type allele and a 388-bp product from the tagged allele. Mice bearing the *BirA* transgene were genotyped using a tripartite primer PCR-based strategy to detect the presence or absence of the CAG promoter at the *Rosa26* locus (Forward:5'-TGCTGCCTCCTGGCTTCTGAG-3', Reverse #1: 5'-GGCGTACTTGGCATATGATACAC-3', Reverse #2: 5'-CACCTGTTCAATTCCCCTGCAG-3') and give rise to a 173-bp product from the wild-type allele and a 477-bp product from the transgene-bearing allele. Mice bearing Cre-recombinase (either *NeuroD6/NEX-Cre* or *Dlx5/6-Cre*) were genotyped using PCR-based strategies as previously described[35,36].

Phenotypic Assessment

For tagged *Mecp2* knock-in mice, phenotypic scoring was performed on a weekly basis for the presence or absence of RTT-like symptoms as previously described[54]. Investigator was blinded to genotypes during phenotypic assessment of mice. For BirA transgenic

mice, no formal scoring was performed. However, *R26*^{BirA} heterozygous and homozygous mice are viable, fertile, and devoid of any gross abnormalities, consistent with previously engineered transgenic mice that express BirA either ubiquitously or within restricted tissues using cell type-specific promoters[55,56].

Immunofluorescence and Microscopy

Mice were anesthetized with 1.25% Avertin (wt/vol), transcardially perfused with 4% paraformaldehyde (wt/vol) in 0.1M sodium-potassium phosphate buffered saline and postfixed overnight at 4°C. Brains were coronally or sagittally sectioned at 20µm using a Leica CM3050 S cryostat. Immunofluorescence on free-floating sections was performed as previously described, except sections were permeabilized with 0.5% Triton without methanol for 20 minutes, and sections were blocked overnight with 10% Normal Goat Serum and 1:100 unconjugated goat anti-mouse IgG (Sigma M5899). The following primary antibodies were incubated at 4°C overnight: rabbit anti-MeCP2 C-terminus (1:1000, in house), rabbit anti-nucleolin (1:1000, Abcam ab22758), mouse anti-parvalbumin (1:500, Millipore MAB1572), rabbit anti-calretinin (1:1000, Swant 7699/3H), mouse anti-GAD67 (1:500, Millipore MAB5406), mouse anti-NeuN (1:500, Millipore MAB377). For rat anti-somatostatin (1:250, Millipore MAB354MI), primary incubation was performed for 48 hours at 4°C. Fluorescence detection of primary antibodies was performed using Alexa 488-conjugated goat anti-rabbit (1:1000, Invitrogen A11008), Alexa 488-conjugated goat anti-mouse (1:1000, Invitrogen A11029), and Alexa 488 goat anti-rat (1:1000, Invitrogen A11006). Fluorescence detection of biotin was performed simultaneously with secondary antibody incubations, using Streptavidin Dylight 650 (1:1000, Fisher 84547) for fluorescence microscopy and Streptavidin Dylight 550 (1:1000, Fisher 84542) for confocal microscopy. Sections were counterstained with DAPI (1:1000, Affymetrix 14564) to visualize DNA before mounting with Fluoromount G (SouthernBiotech). Images were acquired using a Leica DM5500B fluorescent microscope with a Leica DFC360 FX digital camera (region-specific biotinylation, quantification of neuronal cell type-specific markers) or a Leica TCS SP8 Multiphoton confocal microscope (representative images of neuronal cell type specific

markers, subcellular localization of MeCP2). Images were acquired using identical settings for laser power, detector gain amplifier offset and pinhole diameter in each channel. Image processing was performed using ImageJ and Adobe Photoshop, including identical adjustments of brightness, contrast, and levels in individual color channels and merged images across genotypes.

Quantitative western analysis

Quantitative western blot was performed using Odyssey Infrared Imaging System (Licor). Primary antibodies include rabbit anti-MeCP2 C-terminus (1:4000, in house), mouse anti-MeCP2 N-terminus (1:4000, Sigma M7433), mouse anti-NeuN (1:500, Millipore MAB377), and rabbit anti-Histone H3 (1:1000, Abcam ab1791). Secondary antibodies include anti-rabbit IRDye 680LT (1:10,000, Licor), anti-mouse IRDye 800CW (Licor), Streptavidin Dylight 650 (1:10,000, Fisher 84547) and Streptavidin Dylight 800 (1:10,000, Fisher 21851). Quantification of protein expression levels was carried out following Odyssey Infrared Imaging System protocols.

Co-immunoprecipitation using nuclear extracts

Tissues were mined on ice and homogenized in ice cold lysis buffer (10 mM HEPES pH 7.9, 1.5mM $MgCl_2$, 10mM KCl, 0.5% NP-40, 0.2mM EDTA, protease inhibitors). Nuclei were pelleted, washed and resuspended in nuclear extract (NE) buffer (20mM HEPES pH 7.9, 1.5mM $MgCl_2$, 500mM KCl, 0.2mM EDTA, 10% glycerol, protease inhibitors). Nuclei were incubated in NE buffer at 4°C for two hours with rotation. Samples were cleared by ultracentrifugation with a TLA 100.3 rotor (Beckman Optima TL) at 4°C for 30 minutes and the supernatant taken for nuclear extract. Protein concentration was quantified using a modified Bradford assay (Bio-Rad). 1mg of nuclear extract was adjusted to 300µl total volume with NE buffer to perform IP in duplicate. Protein G Dynabeads or Streptavidin M-280 Dynabeads (Life Technologies) were washed three times in PBS with 0.1% Tween-20 and 0.1% BSA. Nuclear extracts were cleared for 30

minutes at 4°C with 25µl Protein G Dynabeads. For streptavidin pulldown, 50µl of Streptavidin M-280 Dynabeads were added to the nuclear extract and incubated at 4°C for two hours with rotation. To test if the Tavi tag was required for streptavidin pulldown, nuclear extracts were split and incubated with or without 200U TEV protease (Invitrogen) in the absence of a reducing agent and without agitation at 4°C for ≥ 4 hours prior to IP. For antibody immunoprecipitation, 5µg of (rabbit anti-MeCP2, in house) was added to the nuclear extract and incubated overnight at 4°C with rotation. Protein G beads were blocked in wash buffer overnight at 4°C with rotation. Blocked beads were then incubated with antibody-bound nuclear extract for two hours at 4°C with rotation. Beads were washed four times in PBS with 0.1% Tween-20 and split into two equal volumes. Each sample was resuspended in 25µl loading buffer with 50mM DTT and boiled for 10 minutes at 95°C prior to loading on a 4-12% Bis-Tris NuPage gel (Life Technologies).

Chromatin immunoprecipitation

Forebrain tissues from male mice at 8 weeks were homogenized in cross-linking buffer (1% formaldehyde (wt/vol), 10mM HEPES (pH 7.5), 100mM NaCl, 1mM EDTA, 1mM EGTA) and cross-linked for 5 minutes at RT. After quenching with 125mM glycine, cross-linked tissue was washed with ice-cold PBS and dounced with 16 strokes in lysis buffer (50mM HEPES (pH 7.5), 140mM NaCl, 1mM EDTA, 1mM EGTA, 10% glycerol (vol/vol), 0.5% NP-40 (vol/vol), and 0.25% Triton X-100 (vol/vol) with protease inhibitors). Nuclei were pelleted, washed and resuspended in chromatin buffer (10mM Tris-HCl (pH 8.0), 1mM EDTA, and 0.5mM EGTA with protease inhibitors). Chromatin was sonicated using a Diagenode Bioruptor, and salt and detergent were added to adjust the chromatin buffer to 0.5% Triton X-100, 150mM NaCl, 10mM EDTA, and 0.1% sodium deoxycholate (DOC, vol/vol), and precleared at 4°C with Protein A Dynabeads (Invitrogen). For immunoprecipitation, 3µg of purified rabbit anti-MeCP2 IgG (in house) or non-specific rabbit IgG control (Millipore NI01) was incubated with 45µg of chromatin for 4 hours, followed by an overnight incubation with pre-blocked Protein A Dynabeads, at 4°C with rotation. Bead-bound chromatin was washed with low salt buffer (50mM HEPES pH 7.5,

150mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% DOC), high salt buffer (50mM HEPES pH 7.5, 500mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% DOC), LiCl buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 1mM EDTA, 0.5% NP-40, 0.5% DOC) and TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA). Chromatin was eluted with elution buffer (50mM Tris-HCl pH 8.0, 10mM EDTA, and 1% SDS (wt/vol)), digested with proteinase K (0.5mg ml^{-1}), and reversed crosslinked at 65°C overnight. After RNase A treatment, DNA fragments were extracted with phenol/chloroform and ethanol-precipitated.

Quantitative real-time PCR (qPCR) analysis was carried out using SYBR green detection (Life Technologies) on an ABI Prism 7900HT Real-Time PCR System (Applied Biosystems). The percent input for each amplicon was determined by comparing the average threshold cycle of the immunoprecipitated DNA to a standard curve generated using serial dilutions of the input DNA and interpolating the "fraction of input" value for this sample.

FACS Isolation of Neuronal Nuclei for RT-PCR and RNA-seq

Nuclei were isolated from fresh cortical tissue for FACS as previously described under ice-cold and nuclease-free conditions[57]. Mouse cortices were rapidly resected on ice and subjected to dounce homogenization in homogenization buffer (0.32M sucrose, 5mM CaCl2, 3mM MgAc2, 10mM Tris-HCl pH 8.0, 0.1% Triton, 0.1mM EDTA, Roche Complete Protease Inhibitor without EDTA). Homogenates were layered onto a sucrose cushion (1.8M sucrose, 10mM Tris-HCl pH 8.0, 3mM MgAc2 Roche Complete Protease Inhibitor without EDTA) and centrifuged in a Beckman Coulter L7 Ultracentrifuge at 25,000 rpm at 4°C for 2.5 hours using a Beckman Coulter SW28 swinging bucket rotor. Nuclei were resuspended & washed once in blocking buffer (1x PBS, 0.5% BSA (Sigma A4503), RNasin Plus RNase Inhibitor (Promega)) and pelleted using a tabletop centrifuge at 5000 RCF at 4°C for 10 minutes. Nuclei were resuspended in blocking buffer to a concentration of ~$6x10^6$ nuclei/ml, blocked for 20 minutes at 4°C with rotation, then incubated with Streptavidin Dylight 650 (1:1000, Fisher 84547) and Alexa 488-conjugated anti-NeuN antibody (1:1000, Millipore MAB377X) for 30 minutes at 4°C with rotation. After a 5-minute incubation with 1:1000 DAPI to enable singlet detection

during FACS, labeled nuclei were washed for an additional 30 minutes at 4°C with blocking buffer, pelleted and resuspended in blocking buffer with 1% BSA. A BD Biosciences Influx cell sorter at the University of Pennsylvania Flow Cytometry and Cell Sorting Facility was used to identify cell type-specific populations of nuclei, and 1.2 – 2.5 x10^5 singlet nuclei from specified populations were directly sorted into Qiagen Buffer RLT Plus for immediate lysis and stabilization of RNA transcripts. Total nuclear RNA was processed using the Qiagen AllPrep DNA/RNA mini kit according to manufacturer instructions, with exception to the on-column DNaseI treatment. RNA was eluted from RNeasy mini spin columns and treated with DNaseI (Qiagen 79254) for 25 minutes at room temperature, then precipitated with glycogen/NaOAc and stored in ethanol at -80°C. Ethanol precipitation of nuclear RNA was carried out to completion prior to initiating RT-PCR or RNA-seq library construction.

For RT-PCR, total RNA was prepared from FACS-isolated cortical nuclei of TAVI male mice at 6 weeks (2-3 mice pooled per biological replicate, 3 replicates each). To validate cell type-specific cortical nuclei populations, total RNA from 120,000 sorted nuclei was converted to cDNA with random hexamers using the SuperScript III First-Strand Synthesis System (Invitrogen). RT-PCR was performed on a ABI Prism 7900HT Real-Time PCR System (Applied Biosystems) using exon-spanning Taqman gene expression assays to detect mRNA transcripts for the following genes: *CRE* (Mr00635245_cn), *Mecp2* (Mm01193537_g1), *Rbfox3* (Mm01248771_m1), *Gfap* (Mm01253033_m1), *Aif1* (Mm00479862_g1), *Mog* (Mm00447824_m1), *Slc17a7* (Mm00812886_m1), *Tbr1* (Mm00493433_m1), *Gad1* (Mm04207432_g1), *Slc35a1* (Mm00494138_m1), *Ht3ar* (Mm00442874_m1), *Pvalb* (Mm00443100_m1), *Sst* (Mm00436671_m1), *Pgk1* (Mm00435617_m1), *Actb* (Mm00607939_s1), *ß2m* (Mm00437762_m1). A geometric mean was calculated to normalize mRNA expression levels to multiple housekeeping genes (*Actb*, *ß2m*, and *Pgk1*), and cell type-enrichment for each sorted population was determined relative to the total mixed population of DAPI+ nuclei.

For RNA-seq, total RNA was prepared from FACS-isolated cortical nuclei of male mice at 6 weeks (TAVI, T158M, R106W, 2-3 mice pooled per biological replicate, 4 replicates each) and female mice at 18 weeks (TAVI, T158M, R106W, single mouse per biological replicate, 2 replicates each). No method of randomization was used to

determine how animals were allocated to experimental groups, which was determined by genotype. The numbers of biological replicates used for differential gene expression analysis are in compliance with ENCODE consortium long RNA-seq recommendations (≥2 replicates). Furthermore, the total amount of RNA isolated from 120,000-250,000 sorted nuclei was used as input for library construction; hence differential gene expression comparisons between FACS-isolated *Mecp2* control and mutant neurons are performed using RNA from equivalent numbers of neuronal nuclei. Total RNA was depleted of ribosomal RNAs, subjected to 5 minutes of heat fragmentation, and converted to strand-specific cDNA libraries using the TruSeq Total RNA library prep kit with RiboZero depletion (Illumina). Multiplexed libraries were submitted for 100 paired-end sequencing on the Illumina HiSeq 2000/2500 platform at the University of Pennsylvania Next Generation Sequencing Core facility, yielding approximately 30-40M total reads per library. 90-95% of total reads were uniquely mapped to the mouse Ensembl GRCm38/mm10 mouse genomic assembly.

GRO-seq

Nuclei were isolated from fresh cortical tissue under ice-cold and nuclease-free conditions as described in the preceding section. After ultracentrifugation, nuclei were resuspended & washed once in PBS (1x PBS, RNasin Plus RNase Inhibitor (Promega)) and pelleted using a tabletop centrifuge at 5000 RCF at 4°C for 10 minutes. Nuclei were resuspended in PBS, pipetted through a 0.22µm filter and counted using a hemocytometer. Nuclei were then pelleted, resuspended to a concentration of 5×10^6 - 10×10^6 nuclei/100µl in glycerol storage buffer (50mM Tris pH 8.3, 40% glycerol, 5mM $MgCl_2$, 0.1 mM), and flash frozen in liquid N_2 for storage until needed.

For each nuclear run-on (NRO), 100µl of nuclei was mixed with 46.5µl NRO Reaction Buffer (10mM Tris pH 8.0, 5mM $MgCl_2$, 1 mM DTT, 300mM KCl), 3.5µl Nucleoside Mix (50µM ATP, 50µM GTP, 2µM CTP, 50µM Br-UTP, 0.4U/µl RNasin), and 50µl 2% Sarkosyl

Nuclear Run On Stop Solution (20mM Tris pH 7.4, 10mM EDTA, 2% SDS). The NRO reaction was performed at 30°C for 5 minutes, then terminated by a 20 minute incubation with DNAse I at 37°C, followed by a hour-long incubation with 225µl NRO Stop Buffer (20mM Tris, pH 7.4, 10mM EDTA, 2% SDS) and Proteinase K at 55°C. Phenol-extracted RNA was fragmented with 0.2N NaOH, and BrdU-RNA was isolated three consecutive times with BrdU-antibody beads, with enzymatic TAP and PNK treatments to remove the cap and 3'-phosphate and to add a 5'-phosphate, as well as Illumina TruSeq small RNA sample prep kit adapter ligations between BrU-RNA isolation steps as described[41,58].

RNA-seq Mapping, Read Counting, and Differential Expression Analysis

The mouse mm10 genomic sequence (Mus_musculus.GRCm38.75.dna.primary_assembly.fa.gz) and gene information (Mus_musculus.GRCm38.75.gtf.gz) were downloaded from Ensembl release 75. The genome files used for mapping were built by STAR using the parameters "STAR --runMode genomeGenerate --runThreadN 12

--genomeDir ./ --genomeFastaFiles Mus_musculus.GRCm38.75.dna.primary_assembly.fa.gz --sjdbGTFfile Mus_musculus.GRCm38.75.gtf --sjdbOverhang 100". The FASTQ files were mapped to the mouse Ensembl GRCm38/mm10 genome assembly by STAR (version 2.3.0) [59] using the parameters "--genomeDir ENSEMBL_75_mm10 --runThreadN 10 --outFilterMultimapNmax 1 --outFilterMismatchNmax 3". Perl scripts generated in-house were used to count the number of read pairs that mapped to across the entire gene body (exon + intron) for each gene. If one end of a read pair overlapped with the annotated genomic region of a given gene and the other did not, the read pair was included in the final count for that gene. Therefore, the total number of read pairs that overlapped within a given gene represented the final read count for that gene. All intron and exon-mapped reads were used for differentially expressed gene comparisons, which were all performed using the edgeR (v3.10.0) and DEseq2 (v1.8.0) R packages[60,61]. Genes with a total edgeR CPM ≥2 were included in the edgeR comparison. A false discovery rate <

0.05 was set to identify differentially expressed genes. No fold change cutoff was applied. For each comparison, the results of edgeR and DESeq2 analyses were merged to avoid method-based biases. The mean fold change and the mean FDR generated from both methods were used for generating plots and heatmaps.

Functional Enrichment of Differentially Expressed Genes

For DAVID gene ontology, a list of differentially expressed protein-coding genes was compared to a background list of actively expressed protein-coding genes from their respective cell type. Statistically significant terms (Benjamini $P < 0.01$, FDR < 0.05) were plotted for Figures S3C-D. For Gene Set Enrichment Analysis (GSEA), we performed a seeded, pre-ranked GSEA from lists of differentially expressed protein-coding genes (ranked by fold change) using the September 2015 Mouse GO Gene Set Release (http://download.baderlab.org/EM_Genesets/September_24_2015/Mouse/). GSEA network associations (*P-value* < 0.1, *Q-value* < 0.1) were visualized using the Enrichment Map application (v2.0.1) in Cytoscape (v3.2.1)[62,63], and clustered using gene set overlap coefficients.

Determination of Actively Expressed Genes

Actively expressed genes for excitatory and inhibitory neurons were determined by calculating the normalized FPKM (zFPKM) and using ZFPKM ≥ 3 for the active gene cutoff as previously described[64].

Statistical Analyses

Statistical analyses were performed using Graphpad Prism version 6.0 for Mac (GraphPad Software, La Jolla California USA, www.graphpad.com) and R[65]. No

statistical method was used to estimate sample size, as pre-specified effect sizes were not assumed. No animals or samples were excluded from analyses. Individual statistical tests are fully stated in the main text or figure legends. Comparisons of normally distributed data consisting of two groups with equal variances (F-test equality of variance $P > 0.05$) were analyzed using Student's T-test, and unequal variances (F-test equality of variance $P < 0.05$) using Students T-test with Welch's correction for unequal variance. Comparisons of normally distributed data consisting of three or more groups were analyzed using One-way ANOVA with the appropriate *post-hoc* test. Comparison of two or more factors across multiple groups was analyzed using a Two-way ANOVA with Sidak's correction for multiple comparisons. Comparisons of non-normally distributed data were analyzed using the Mann-Whitney/Wilcoxon test (two groups) or the Kruskal-Wallis test (three or more groups) with the appropriate *post-hoc* test. For multiple comparisons, all p-values are adjusted using the Holm-Bonferroni correction unless otherwise indicated.

Main Figure Statistical Analyses

Figure 1 Utilization and characterization of *Mecp2*^{Tavi} mice and associated RTT variants

(f) $n_{replicates} = 4$, One-way ANOVA [$F = 25.55$, $P = 0.0012$]; Tukey's multiple comparisons correction applied. **(g)** $n_{WT} = 20$, $n_{TAVI} = 11$, $n_{KO} = 6$, $n_{T158M} = 6$, $n_{R106W} = 12$; One-way ANOVA [$F = 20.05$, $P < 0.0001$]; Tukey's multiple comparison correction applied. **(h)** $n_{WT} = 31$, $n_{TAVI} = 23$, $n_{KO} = 17$, $n_{T158M} = 39$, $n_{R106W} = 26$, Mantel-Cox [$\chi^2 = 109.3$, df = 4, $P < 0.0001$].

Figure 2 Cell type-specific transcriptional profiling of neuronal nuclei

(d) $n_{replicates} = 3$, Two-way ANOVA, Control [Cell Type-Gene Interaction, $F = 42.68$, $P < 0.0001$; Cell Type, $F = 222.0$, $P < 0.0001$; Gene, $F = 80.03$, $P < 0.0001$], Non-Neuronal [Cell Type-Gene Interaction, $F = 12.47$, $P < 0.0001$; Cell Type, $F = 109.8$, $P < 0.0001$; Gene, $F = 7.655$, $P = 0.0027$], EXC-specific [Cell Type-Gene Interaction, $F = 4.376$, $P = 0.0198$; Cell Type, $F = 1227$, $P < 0.0001$; Gene, $F = 0.3267$, $P = 0.5756$], INH-spectific

[Cell Type-Gene Interaction, F = 3.047, P = 0.0040; Cell Type, F = 646.5, P < 0.001; Gene, F = 2.916, P = 0.033]; Dunnett's multiple comparisons correction applied.

Figure 3 T158M and R106W differentially expressed genes at 6 weeks of age

(e) One-tailed Wilcoxon Signed Rank, Excitatory $P_{\text{Upregulated}}$ = 4.357e-3, Excitatory $P_{\text{Downregulated}}$ = 7.345e-3, Inhibitory $P_{\text{Upregulated}}$ = 4.575e-09, Inhibitory $P_{\text{Downregulated}}$ = 1.684e-05. **(e)** One-tailed Two-sample Kolmogorov-Smirnov, Excitatory P = 0.1241 [D^+ = 0.17391], Inhibitory P = 2.032e-05 [D^+ = 0.31776]. **(g)** Chi-square Goodness-of-Fit, Excitatory P_{T158M} < 2.2e-16 [χ^2 = 182.2, df = 2], Excitatory P_{R106W} < 2.2e-16 [χ^2 = 401.11, df = 2], Inhibitory P_{T158M} < 2.2e-16 [χ^2 = 119.94, df = 2], Inhibitory P_{R106W} < 2.2e-16 [χ^2 = 346.86, df = 2]. **(i)** Two-tailed Kruskal-Wallis Rank Sum, Excitatory P < 2.2e-16 [χ^2 = 418.2, df = 3], Inhibitory P < 2.2e-16 [χ^2 = 1026.9, df = 3]; Pairwise Wilcoxon Rank Sum P displayed.

Figure 4 Genome-wide length-dependent transcriptional changes in RTT mutant mice

(e) *Top*, n = 10,390 genes, Kolmogorov-Smirnov P < 2.2e-16 for each nascent or nuclear RNA versus whole cell RNA comparison, no correction for multiple comparisons. **(f)** n = 10, 390 genes, Kruskal-Wallis $P_{\text{Group A}}$ < 2.2e-16 [χ^2 = 2664.8, df = 2], $P_{\text{Group B}}$ < 2.2e-16 [χ^2 = 290.18, df = 2], $P_{\text{Group C}}$ < 2.2e-16 [χ^2 = 2403.3, df = 2], $P_{\text{Group D}}$ < 2.2e-16 [χ^2 = 319.36, df = 2], $P_{\text{Group E}}$ < 2.2e-16 [χ^2 = 1483.8, df = 2], $P_{\text{Group F}}$ < 2.2e-16 [χ^2 = 1385.8, df = 2], $P_{\text{Group G}}$ < 2.2e-16 [χ^2 = 2522.9, df = 2], $P_{\text{Group H}}$ < 2.2e-16 [χ^2 = 2442.7, df = 2]; Pairwise Wilcoxon Rank Sum P displayed.

Figure 6 T158M and R106W differentially expressed genes in mosaic female mice

(a) Two-way ANOVA [Genotype-Time Interaction, F = 2.987, P = 0.0712; Genotype, F = 41.14, P < 0.0001; Time, F = 7.332, P = 0.0129; Subjects (matching), F = 1.873, P = 0.0744]. **(b)** FACS isolation of cortical mosaic excitatory neuronal nuclei from heterozygous TAVI, T158M, or R106W female mice. **(c)** n_{T158M} = 4, n_{R106W} = 9, Two-way

ANOVA [Population-Genotype Interaction, F = 0.3320, P = 0.5703; Population, F = 111.1, P < 0.0001; Genotype, F = 0.332, P = 0.5703]. **(d)** n_{TAVI} = 12, n_{T158M} = 4, n_{R106W} = 9, One-way ANOVA [F = 0.9376, P = 0.4067]. **(h)** One-tailed Fisher's Exact Test [Odds Ratio = 19.3, P = 2.43e-05]. **(i)** One-tailed Wilcoxon Signed Rank, $P_{\text{Total Overlap}}$ = 0.0331, $P_{\text{Cell. Auto.}}$ = 0.5778, $P_{\text{Non-Cell Auto.}}$ = 8.825e-06.

Data availability

All sequencing data reported in this study has been deposited in the NCBI Gene Expression Omnibus (GSE83474).

References (Specific to Chapter 2)

1. Chahrour, M. & Zoghbi, H. Y. The story of Rett syndrome: from clinic to neurobiology. *Neuron* **56,** 422–37 (2007).
2. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23,** 185–188 (1999).
3. Shahbazian, M. D., Antalffy, B., Armstrong, D. L. & Zoghbi, H. Y. Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. *Hum. Mol. Genet.* **11,** 115–124 (2002).
4. Lewis, J. D. *et al.* Purification, sequence, and cellular localization of a novel chromosomal protein that binds to Methylated DNA. *Cell* **69,** 905–914 (1992).
5. Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19,** 187–191 (1998).
6. Lyst, M. J. *et al.* Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16,** 898–902 (2013).
7. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393,** 386–9 (1998).
8. Skene, P. J. *et al.* Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. *Mol. Cell* **37,** 457–468 (2010).
9. Chahrour, M. *et al.* MeCP2, a Key Contributor to Neurological Disease, Activates and Represses Transcription. *Science* **320,** 1224–1229 (2008).
10. Chen, L. *et al.* MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc. Natl. Acad. Sci.* **112,** 5509–5514 (2015).
11. Li, Y. *et al.* Global Transcriptional and Translational Repression in Human-Embryonic-Stem-Cell-Derived Rett Syndrome Neurons. *Cell Stem Cell* **13,** 446–458 (2013).
12. Cuddapah, V. A. *et al.* Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in Rett syndrome. *J. Med. Genet.* **51,** 152–158 (2014).
13. Brown, K. *et al.* The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Hum. Mol. Genet.* ddv496 (2015). doi:10.1093/hmg/ddv496
14. Goffin, D. *et al.* Rett syndrome mutation MeCP2 T158A disrupts DNA binding, protein stability and ERP responses. *Nat Neurosci* **15,** 274–283 (2012).
15. Baker, S. A. *et al.* An AT-Hook Domain in MeCP2 Determines the Clinical Course of Rett Syndrome and Related Disorders. *Cell* **152,** 984–996 (2013).
16. Katz, D. M. *et al.* Preclinical research in Rett syndrome: setting the foundation for translational success. *Dis. Model. Mech.* **5,** 733–745 (2012).
17. Lyst, M. J. & Bird, A. Rett syndrome: a complex disorder with simple roots. *Nat. Rev. Genet.* **16,** 261–275 (2015).
18. Fishell, G. & Heintz, N. The Neuron Identity Problem: Form Meets Function. *Neuron* **80,** 602–612 (2013).
19. Lodato, S. & Arlotta, P. Generating Neuronal Diversity in the Mammalian Cerebral Cortex. *Annu. Rev. Cell Dev. Biol.* **31,** 699–720 (2015).
20. Doyle, J. P. *et al.* Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135,** 749–62 (2008).

21. Molyneaux, B. J. *et al.* DeCoN: Genome-wide Analysis of In Vivo Transcriptional Dynamics during Pyramidal Neuron Fate Selection in Neocortex. *Neuron* **85,** 275–288 (2015).
22. Mo, A. *et al.* Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86,** 1369–1384 (2015).
23. Gabel, H. W. *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522,** 89–93 (2015).
24. Guo, J. U. *et al.* Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* **17,** 215–222 (2014).
25. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Cell* **151,** 1417–1430 (2012).
26. Rube, H. T. *et al.* Sequence features accurately predict genome-wide MeCP2 binding in vivo. *Nat. Commun.* **7,** 11025 (2016).
27. Malik, H. S. & Henikoff, S. A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. *Cell* **18,** 1030–1040 (2010).
28. Lakso, M. *et al.* Efficient in vivo manipulation of mouse genomic sequences at the zygote stage. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 5860–5865 (1996).
29. Samaco, R. C. *et al.* A partial loss of function allele of Methyl-CpG-binding protein 2 predicts a human neurodevelopmental syndrome. *Hum. Mol. Genet.* **17,** 1718–1727 (2008).
30. RettBASE: Rett Syndrome Variation Database. Available at: http://mecp2.chw.edu.au/. (Accessed: 8th April 2016)
31. Ghosh, R. P., Horowitz-Scherer, R. A., Nikitina, T., Gierasch, L. M. & Woodcock, C. L. Rett syndrome-causing mutations in human MeCP2 result in diverse structural changes that impact folding and DNA interactions. *J Biol Chem* **283,** 20523–34 (2008).
32. Ho, K. L. *et al.* MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. *Mol. Cell* **29,** 525–531 (2008).
33. Ballestar, E., Yusufzai, T. M. & Wolffe, A. P. Effects of Rett syndrome mutations of the methyl-CpG binding domain of the transcriptional repressor MeCP2 on selectivity for association with methylated DNA. *Biochemistry (Mosc.)* **39,** 7100–7106 (2000).
34. Kumar, A. *et al.* Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. *J. Cell Sci.* **121,** 1128–1137 (2008).
35. Goebbels, S. *et al.* Genetic targeting of principal neurons in neocortex and hippocampus of NEX-Cre mice. *Genes. N. Y. N 2000* **44,** 611–621 (2006).
36. Monory, K. *et al.* The Endocannabinoid System Controls Key Epileptogenic Circuits in the Hippocampus. *Neuron* **51,** 455–466 (2006).
37. Bhatt, D. M. *et al.* Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions. *Cell* **150,** 279–290 (2012).
38. Werner, M. S. & Ruthenburg, A. J. Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. *Cell Rep.* **12,** 1089–1098 (2015).
39. Sugino, K. *et al.* Cell-Type-Specific Repression by Methyl-CpG-Binding Protein 2 Is Biased toward Long Genes. *J. Neurosci.* **34,** 12877–12883 (2014).
40. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416,** 499–506 (2002).

41. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* **322,** 1845–1848 (2008).
42. Flavell, S. W. & Greenberg, M. E. Signaling Mechanisms Linking Neuronal Activity to Gene Expression and Plasticity of the Nervous System. *Annu. Rev. Neurosci.* **31,** 563–590 (2008).
43. Gong, S. *et al.* Targeting Cre Recombinase to Specific Neuron Populations with Bacterial Artificial Chromosome Constructs. *J. Neurosci.* **27,** 9817–9823 (2007).
44. Müller, M. & Can, K. Aberrant redox homoeostasis and mitochondrial dysfunction in Rett syndrome. *Biochem. Soc. Trans.* **42,** 959–964 (2014).
45. Zylka, M. J., Simon, J. M. & Philpot, B. D. Gene Length Matters in Neurons. *Neuron* **86,** 353–355 (2015).
46. Linhoff, M. W., Garg, S. K. & Mandel, G. A High-Resolution Imaging Approach to Investigate Chromatin Architecture in Complex Tissues. *Cell* **163,** 246–255 (2015).
47. King, I. F. *et al.* Topoisomerases facilitate transcription of long genes linked to autism. *Nature* **501,** 58–62 (2013).
48. Nott, A. *et al.* Histone deacetylase 3 associates with MeCP2 to regulate FOXO and social behavior. *Nat. Neurosci.* **advance online publication,** (2016).
49. Buxbaum, A. R., Yoon, Y. J., Singer, R. H. & Park, H. Y. Single-molecule insights into mRNA dynamics in neurons. *Trends Cell Biol.* **25,** 468–475 (2015).
50. Belichenko, N. P., Belichenko, P. V. & Mobley, W. C. Evidence for both neuronal cell autonomous and nonautonomous effects of methyl-CpG-binding protein 2 in the cerebral cortex of female mice with Mecp2 mutation. *Neurobiol. Dis.* **34,** 71–77 (2009).
51. Castro, J. *et al.* Functional recovery with recombinant human IGF1 treatment in a mouse model of Rett Syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 9941–9946 (2014).
52. Khwaja, O. S. *et al.* Safety, pharmacokinetics, and preliminary assessment of efficacy of mecasermin (recombinant human IGF-1) for the treatment of Rett syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 4596–4601 (2014).
53. Xiao, C. *et al.* MiR-150 Controls B Cell Differentiation by Targeting the Transcription Factor c-Myb. *Cell* **131,** 146–159 (2007).
54. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of neurological defects in a mouse model of Rett syndrome. *Science* **315,** 1143–7 (2007).
55. Boer, E. de *et al.* Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci USA* **100,** 7480–5 (2003).
56. Driegen, S. *et al.* A generic tool for biotinylation of tagged proteins in transgenic mice. *Transgenic Res* **14,** 477–82 (2005).
57. Lister, R. *et al.* Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* **341,** 1237905 (2013).
58. Greer, C. B. *et al.* Histone Deacetylases Positively Regulate Transcription through the Elongation Machinery. *Cell Rep.* **13,** 1444–1455 (2015).
59. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* bts635 (2012). doi:10.1093/bioinformatics/bts635
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** (2014).
61. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* **26,** 139–140 (2010).

62. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLOS ONE* **5,** e13984 (2010).
63. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinforma. Oxf. Engl.* **27,** 431–432 (2011).
64. Hart, T., Komori, H. K., LaMere, S., Podshivalova, K. & Salomon, D. R. Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics* **14,** 778 (2013).
65. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2014)

Figures

Figure 1

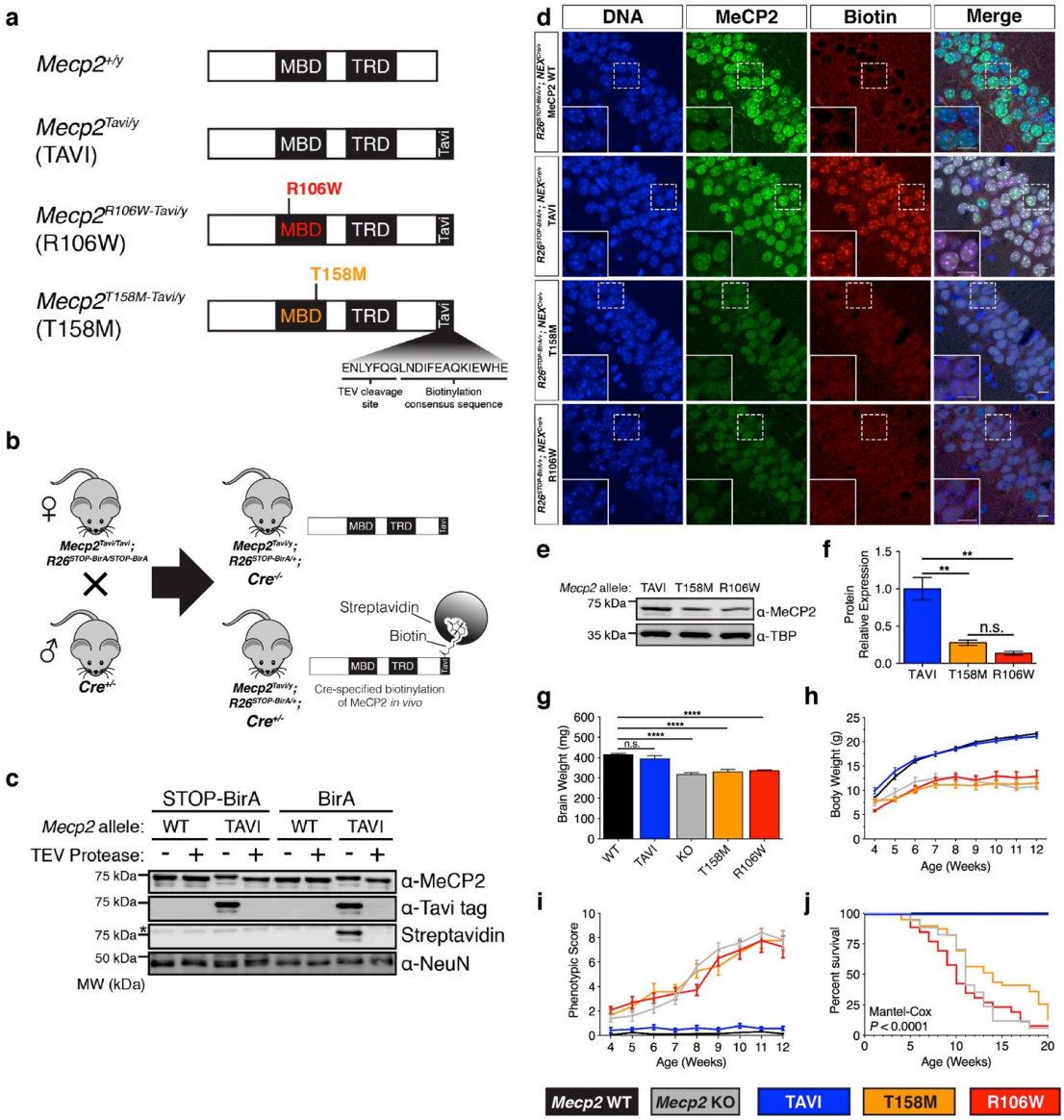


Figure 1 Utilization and characterization of *Mecp2*^{Tavi} mice and associated RTT variants **(a)** Diagram of wild-type and tagged MeCP2 showing R106W or T158M missense mutations. MBD, Methyl-CpG Binding Domain; TRD, Transcriptional Repression Domain. **(b)** Breeding strategy to biotinylate the Tavi tag in a Cre-dependent manner. **(c)** Representative western blot showing the conditions in which the Tavi tag is

biotinylated using whole brain nuclear extracts. Blot is probed with streptavidin for biotin detection and antibodies against MeCP2 N-terminus, Tavi tag, and NeuN. **(d)** Representative images showing immunofluorescent detection of biotinylated MeCP2 and mutant variants in hippocampal tissues of untagged (WT) and tagged (TAVI, T158M, R106W) male mice at 6 weeks of age. Tissue is probed with streptavidin for biotin detection and antibody against the MeCP2 C-terminus. Scale bars represent 10 μm. **(e)** Representative western blot comparing MeCP2 protein expression levels between Tavi-tagged control (TAVI) and mutant (T158M, R106W) male mice at 6 weeks of age. Blot is probed with antibodies against the MeCP2 C-terminus and TBP. **(f)** Quantification of western blot in (e) ($n_{replicates}$ = 4, One-way ANOVA). **(g)** Brain weights from untagged (WT, KO) and tagged (TAVI, T158M, R106W) male mice at 6 weeks of age (n_{WT} = 20, n_{TAVI} = 11, n_{KO} = 6, n_{T158M} = 6, n_{R106W} = 12; One-way ANOVA). **(h)** Body weight over postnatal age in untagged (WT, KO) and tagged (TAVI, T158M, R106W) male mice. Data points consist of at least 6 observations each. Total number of mice assessed: n_{WT} = 31, n_{TAVI} = 23, n_{KO} = 15, n_{T158M} = 14, n_{R106W} = 28. **(i)** RTT-like phenotypic score across postnatal development in untagged (WT, KO) and tagged (TAVI, T158M, R106W) male mice. Data points over time consist of at least 6 observations each. Total number of mice assessed: n_{WT} = 31, n_{TAVI} = 23, n_{KO} = 15, n_{T158M} = 14, n_{R106W} = 28. **(j)** Kaplan-Meier survival curve for untagged (WT, KO) and tagged (TAVI, T158M, R106W) male mice (n_{WT} = 31, n_{TAVI} = 23, n_{KO} = 17, n_{T158M} = 39, n_{R106W} = 26). $^{*}P < 0.5$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, n.s. = not significant; all pooled data depicts mean ± SEM. See also **Supplementary Figure 1.**

Figure 2

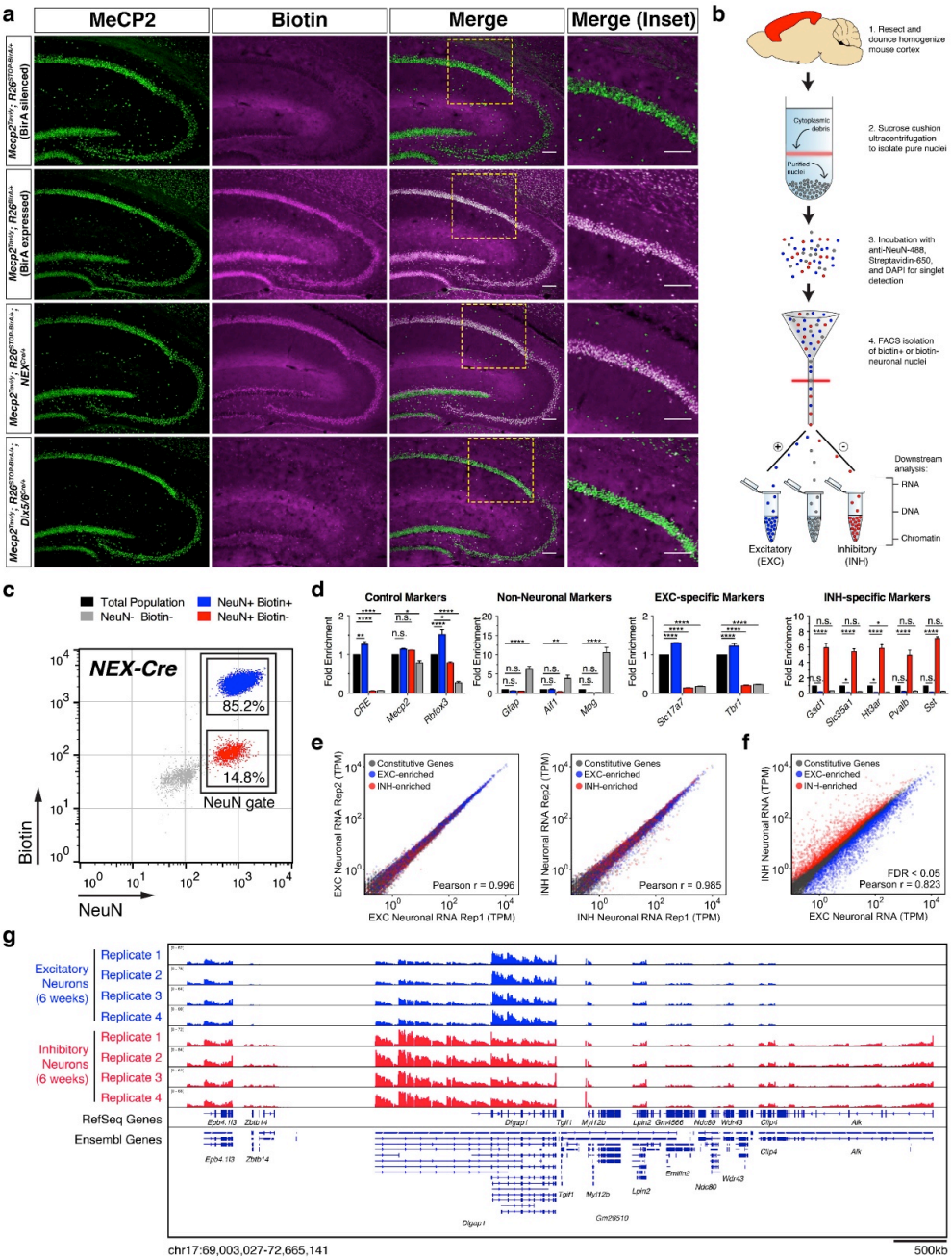


Figure 2 Cell type-specific transcriptional profiling of neuronal nuclei **(a)** Representative images showing immunofluorescent detection of biotinylated MeCP2-Tavi protein in Cre-specified neuronal populations of the mouse hippocampus. Probed

using streptavidin for biotin detection and antibody against the MeCP2 C-terminus. Scale bars represent 100μm. **(b)** Schematic of cortical nuclei preparation and FACS isolation. **(c)** FACS analysis of labeled cortical nuclei populations. Data shown is representative of nine independent experiments using NEX-Cre mice. Percentages indicate the mean distribution of neurons that are biotin+ (excitatory; 85.2% ± 0.35) or biotin- (inhibitory; 14.8% ± 0.35). **(d)** RT-PCR validation of FACS-isolated populations depicted in (c) ($n_{replicates}$ = 3, Two-way ANOVA). **(e)** Pearson correlation of biological replicate nuclear RNA-seq libraries from FACS-isolated populations depicted in (c). Colors correspond to EXC-enriched (blue) and INH-enriched (red) genes identified through differential expression analysis of excitatory and inhibitory neurons. **(f)** Pearson correlation of excitatory and inhibitory RNA-seq libraries. Data shown is representative of 4 biological replicates. Note lower Pearson correlation and clear dispersal of cell type-enriched genes. **(g)** IGV browser snapshot of *Dlgap1* genomic locus in excitatory and inhibitory neurons of TAVI male mice at 6 weeks of age. RefSeq and Ensembl gene annotations are both shown. $*P < 0.5$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$, n.s. = not significant; all pooled data depicts mean ± SEM. See also **Supplementary Figures 2 and 3**.

Figure 3

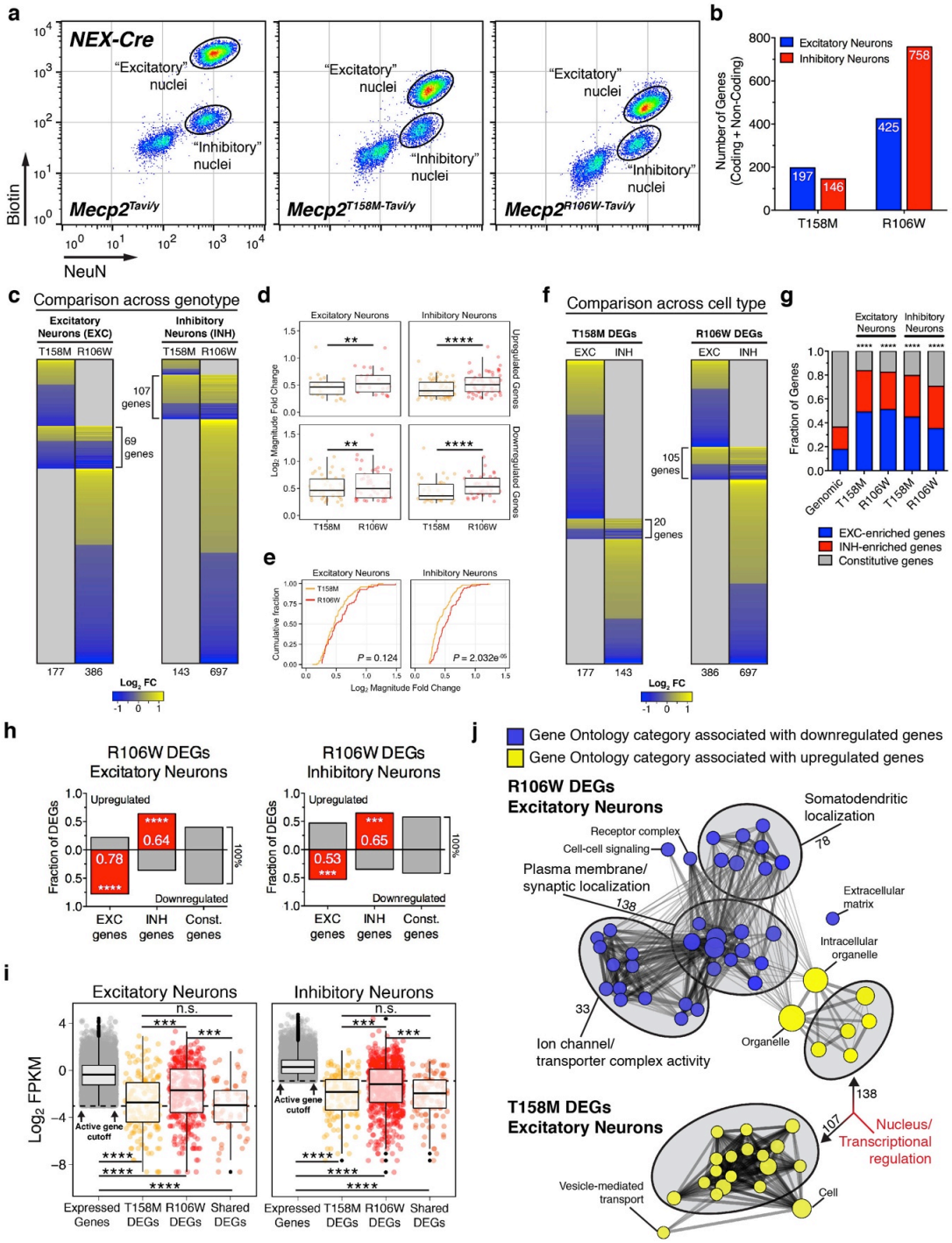


Figure 3 T158M and R106W differentially expressed genes at 6 weeks of age **(a)** FACS isolation of cortical excitatory and inhibitory neuronal nuclei from TAVI, T158M, or R106W male mice at 6 weeks of age. **(b)** Total number of protein coding and non-coding

differentially expressed genes (DEGs) identified in excitatory or inhibitory neurons of *Mecp2*-mutant mice. **(c)** Heatmap displaying $\log_2$ fold changes among protein-coding DEGs in excitatory and inhibitory neurons of *Mecp2*-mutant mice, compared across genotypes. Excitatory DEGs n_{shared} = 69 genes, Hypergeometric $P = 3.15e^{-77}$. Inhibitory DEGs n_{shared} = 107 genes, Hypergeometric $P = 5.33e^{-134}$. **(d)** Boxplots comparing $\log_2$ median fold changes among overlapping DEGs between T158M and R106W neurons in (c) (One-tailed Wilcoxon Signed Rank). **(e)** Cumulative distribution function of the absolute $\log_2$ fold change among overlapping DEGs between T158M and R106W neurons in (c) (One-tailed Two-sample Kolmogorov-Smirnov). **(f)** Heatmap displaying $\log_2$ fold changes among protein-coding DEGs in excitatory and inhibitory neurons of *Mecp2*-mutant mice, compared across cell types. **(g)** Distribution of constitutive, EXC- or INH-enriched genes among T158M and R106W protein-coding DEGs, compared against genomic distribution (Chi-square Goodness-of-Fit). **(h)** Bar plot summarizing R106W DEGs, partitioned by cell type-enriched or constitutive genes, which are preferentially upregulated or downregulated. Red indicates statistical significance (One-tailed Fisher's Exact Test). **(i)** Boxplots comparing the $\log_2$ FPKM distribution of actively expressed genes against T158M, R106W, and shared DEGs for each cell type (Pairwise Wilcoxon Rank Sum *P* displayed). **(j)** Enrichment map of pre-ranked Gene Set Enrichment Analysis (GSEA) functional network associations. Data represents DEGs from R106W (top) and T158M (bottom) excitatory neurons (*P-value* < 0.01, *Q-value* < 0.1). Nodes denote functional categories, colored by Normalized Enrichment Score (NES). Line weight denotes extent of gene overlap between connected nodes. **P* < 0.5, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001, n.s. = not significant. See also **Supplementary Figure 4**.

Figure 4

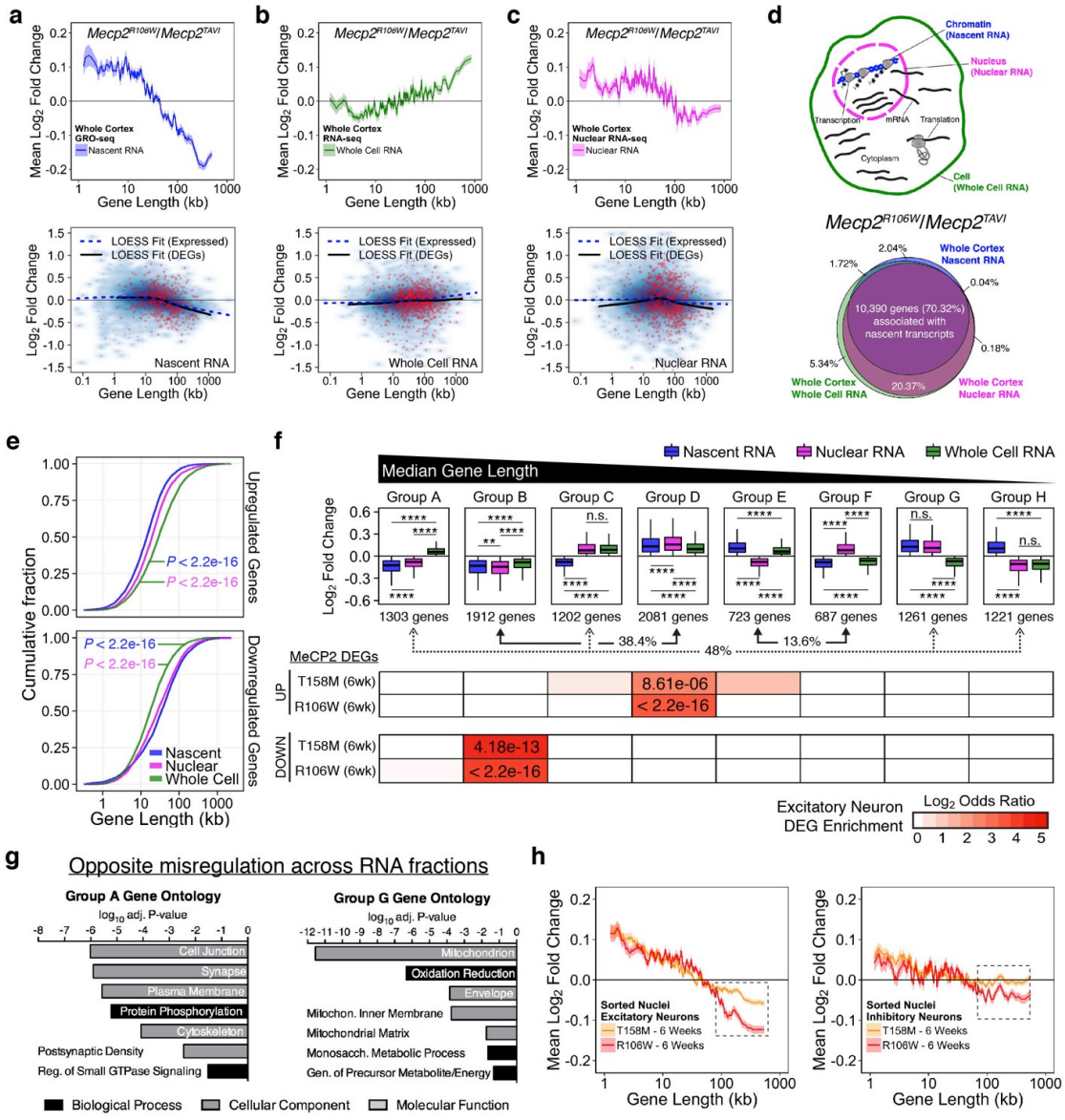


Figure 4 Genome-wide length-dependent transcriptional changes in RTT mutant mice **(a)** Genome-wide $\log_2$ fold changes in R106W mice (n = 2) compared to TAVI mice (n = 2) at 6 weeks of age using GRO-seq. *Top*, Lines represent mean fold change in expression for genes binned according to gene length (200 gene bins, 40 gene step) as described in [23]. Ribbon represents SEM of genes in each bin. *Bottom*, Smoothed scatterplot depicting LOESS correlation between gene length and $\log_2$ fold change for all individual protein-

coding genes detected in GROseq. Genes in red highlight R106W DEGs identified from sorted excitatory and inhibitory neuronal nuclei. **(b,c)** Same as in (a), but using total RNA-seq analysis of whole cell (b) or nuclear (c) cortical RNA isolated from the same biological samples. **(d)** *Top*, Diagram of RNA distribution across subcellular compartments. *Bottom,* Area proportional Venn diagram comparing overlap in gene expression changes between nuclear RNA, whole cell RNA, and nascent RNA. **(e)** Cumulative distribution function of gene lengths for all upregulated and downregulated protein-coding genes among nascent, nuclear, and whole cell RNA fractions (n = 10,390 genes, Kolmogorov-Smirnov). **(f)** *Top,* Boxplots depicting median $\log_2$ fold changes in R106W mice between nascent, nuclear, and whole cell RNA fractions, classified by the direction of gene misregulation (n = 10, 390 genes, Pairwise Wilcoxon Rank Sum *P* displayed). Gene groups are sorted by median gene length. Arrows highlight gene groups with similar changes (38.4% of genes), opposite changes (48%), or incongruent changes (13.6%) across subcellular fractions. *Bottom,* Heatmap displaying statistical enrichment of 6-week T158M and R106W DEGs in excitatory neurons among gene groups (One-tailed Fisher's Exact Test). **(g)** DAVID Gene ontology terms (Benjamini $P < 0.01$, FDR < 0.05) for Group A and Group G sets of genes defined in (f). **(h)** Mean $\log_2$ fold change in 6-week R106W (red; n = 4) and T158M (orange, n = 4) sorted excitatory neurons (*left*) and inhibitory neurons (*right*) using GRO-seq filtered genes. See also **Supplementary Figure 5**.

Figure 5

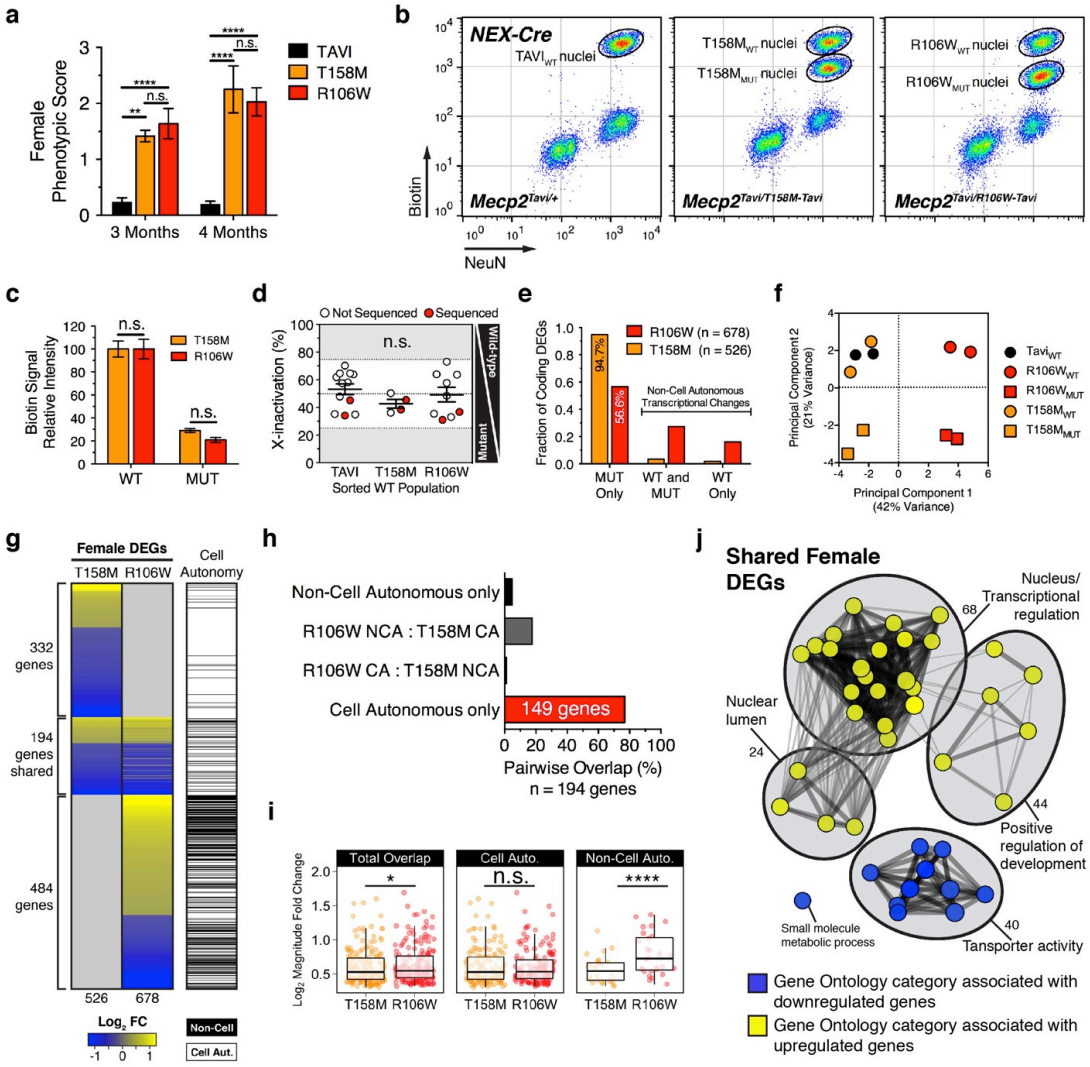
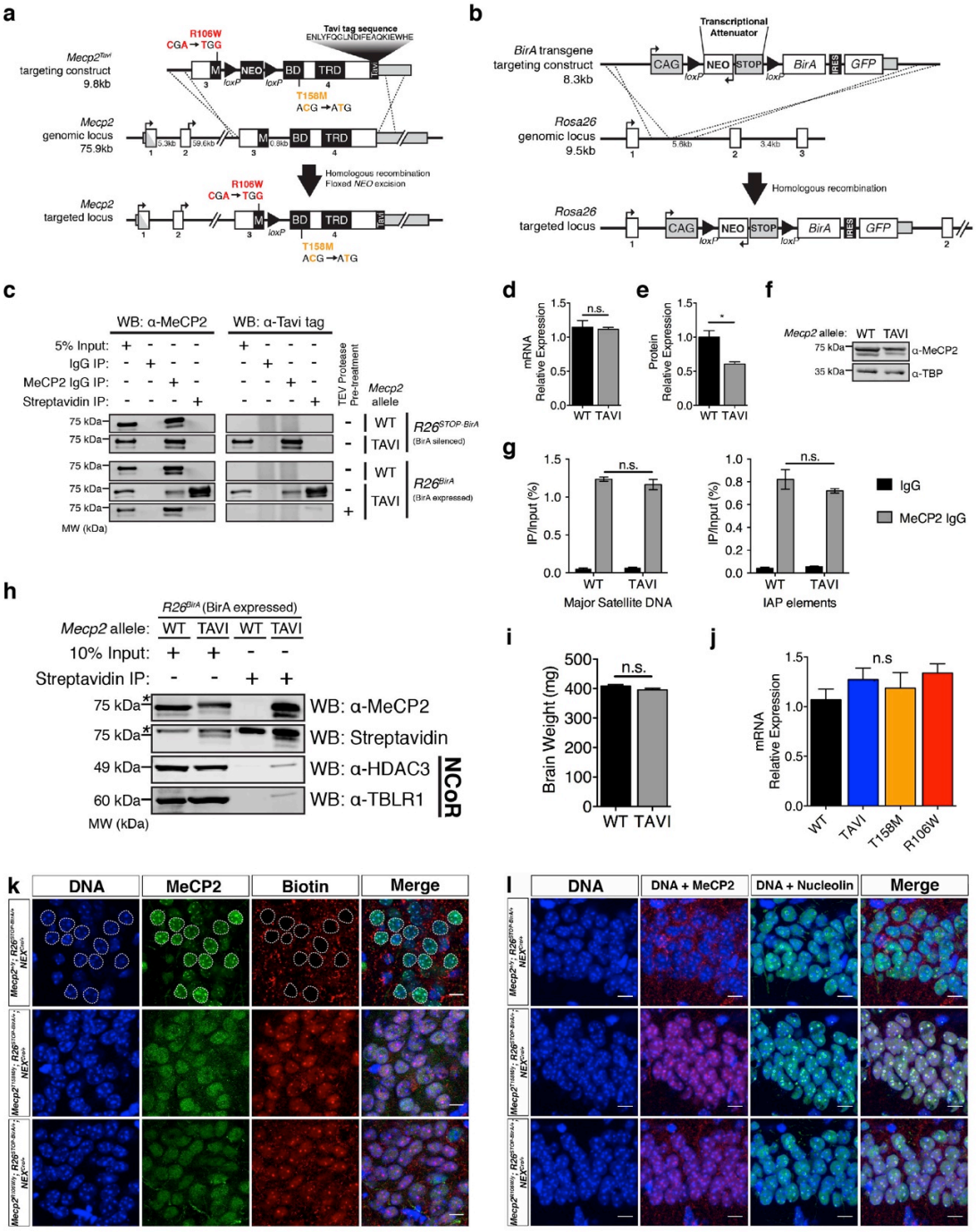


Fig. 5. T158M and R106W differentially expressed genes in mosaic female mice (a) RTT-like phenotypic score in TAVI (n = 12), T158M (n = 4), and R106W (n = 9) female mice (Two-way ANOVA). Data depicts mean ± SEM. **(b)** FACS isolation of cortical mosaic excitatory neuronal nuclei from heterozygous TAVI, T158M, or R106W female mice. **(c)** Biotin signal intensity from FACS-isolated populations depicted in (a) (n_{T158M} = 4, n_{R106W} = 9, Two-way ANOVA). Data depicts mean ± SEM. **(d)** X-inactivation ratios among cortical excitatory neurons in female mice, displayed as a percentage of the

FACS-sorted WT population (n_{TAVI} = 12, n_{T158M} = 4, n_{R106W} = 9, One-way ANOVA). Data points in red highlight samples used for RNA-seq. Data depicts mean ± SEM. **(e)** Bar graph showing the cell and non-cell autonomous distribution of total protein-coding DEGs identified from T158M and R106W female mice. **(f)** Principal component analysis of WT and MUT cell populations isolated from TAVI, T158M, and R106W female mice. **(g)** Heatmap displaying $\log_2$ fold changes among the total number of protein-coding DEGs detected in both WT and MUT populations from T158M or R106W female mice. Note genes that overlap across genotype (n = 194). **(h)** Proportion of cell autonomous and non-cell autonomous genes that overlap between T158M and R106W female excitatory neurons (One-tailed Fisher's Exact Test). **(i)** Boxplots comparing absolute $\log_2$ fold change between cell autonomous and non-cell autonomous shared DEGs (n = 185) between T158M and R106W female mice (One-tailed Wilcoxon Signed Rank). **(j)** Enrichment map of pre-ranked GSEA functional network associations (*P-value* < 0.01, *Q-value* < 0.1). Data represents DEGs that overlap between T158M and R106W mice (n = 185). Nodes denote functional categories, colored by NES. Line weight denotes extent of gene overlap between connected nodes. **P* < 0.5, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001, n.s. = not significant. See also **Supplementary Figure 6**

Supplementary Information

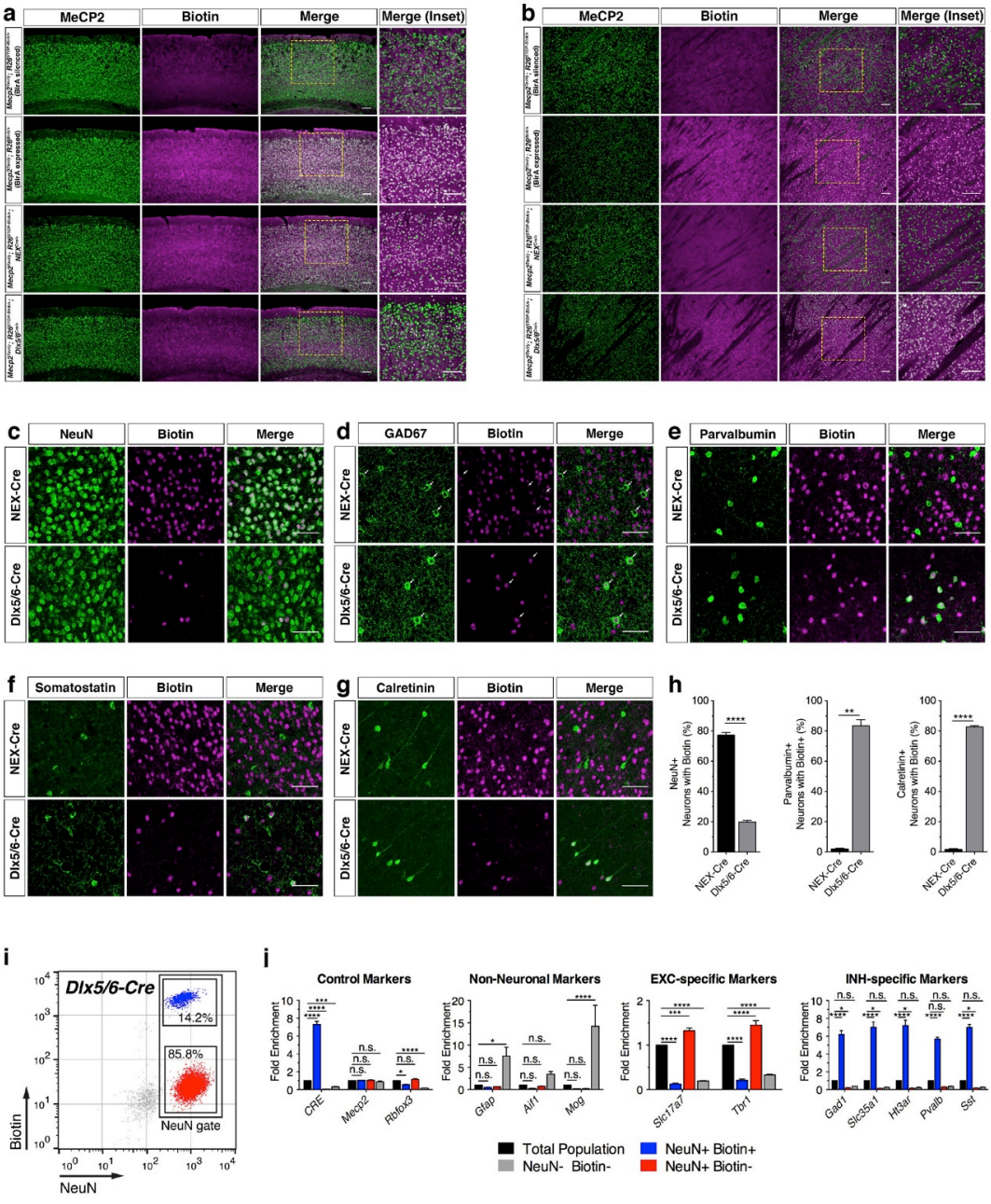
Supplementary Figure 1



Supplementary Figure 1. Design and characterization of MeCP2-Tavi and BirA mice **(a)** Schematic of targeting strategy for generating Tavi-tagged knock-in mice at the

endogenous *Mecp2* gene locus. MBD, Methyl-CpG Binding Domain; TRD, Transcriptional Repression Domain; NEO, Neomycin cassette. **(b)** Schematic of targeting strategy for generating conditional BirA transgenic mice at the *Rosa26* gene locus. **(c)** Western blot showing the conditions in which the Tavi tag is biotinylated and accessible for biochemical pulldown using whole brain nuclear extracts. Pulldowns were performed using streptavidin or an antibody against MeCP2 C-terminus. Blot is probed with antibodies against MeCP2 N-terminus and Tavi tag. **(d)** Relative MeCP2 mRNA expression levels between wild-type untagged and tagged mice at 20 weeks of age, normalized to GAPDH mRNA (n = 3, Two-tailed Unpaired Student t-test $P = 0.7882$). **(e)** Relative MeCP2 protein expression levels between wild-type untagged and tagged mice at 20 weeks of age, normalized to TBP (n = 3, Two-tailed Unpaired Student t-test $P = 0.0203$). **(f)** Western blot comparing MeCP2 protein expression levels between wild-type untagged and tagged mice at 20 weeks of age. Blot is probed with antibodies against MeCP2 C-terminus and TBP. **(g)** ChIP-PCR signal of MeCP2 binding at highly methylated repetitive elements using antibodies against MeCP2 C-terminus or normal rabbit IgG ($n_{replicates} = 3$, Two-way ANOVA, Major Sattelite [Antibody-Genotype Interaction, $F = 1.313$, $P = 0.2850$; Antibody, $F = 933.5$, $P < 0.0001$; Genotype, $F = 0.5824$, $P = 0.4673$], IAP [Antibody-Genotype Interaction, $F = 1.672$, $P = 0.2320$; Antibody, $F = 262.6$, $P < 0.0001$; Genotype, $F = 0.9161$, $P = 0.3665$]). **(h)** Streptavidin-mediated pulldown of biotinylated MeCP2-Tavi and interacting NCoR co-repressor components (HDAC3, TBLR1). Blot is probed with streptavidin for biotin detection and antibodies against MeCP2 C-terminus, HDAC3, and TBLR1. **(i)** Brain weight from untagged (WT) and tagged (TAVI) male mice at 20 weeks of age ($n_{WT} = 21$, $n_{TAVI} = 23$, Two-tailed Unpaired Student t-test $P = 0.1096$). **(j)** Relative MeCP2 mRNA expression levels between wild-type untagged and tagged mice at 6 weeks of age (n=3-4 per

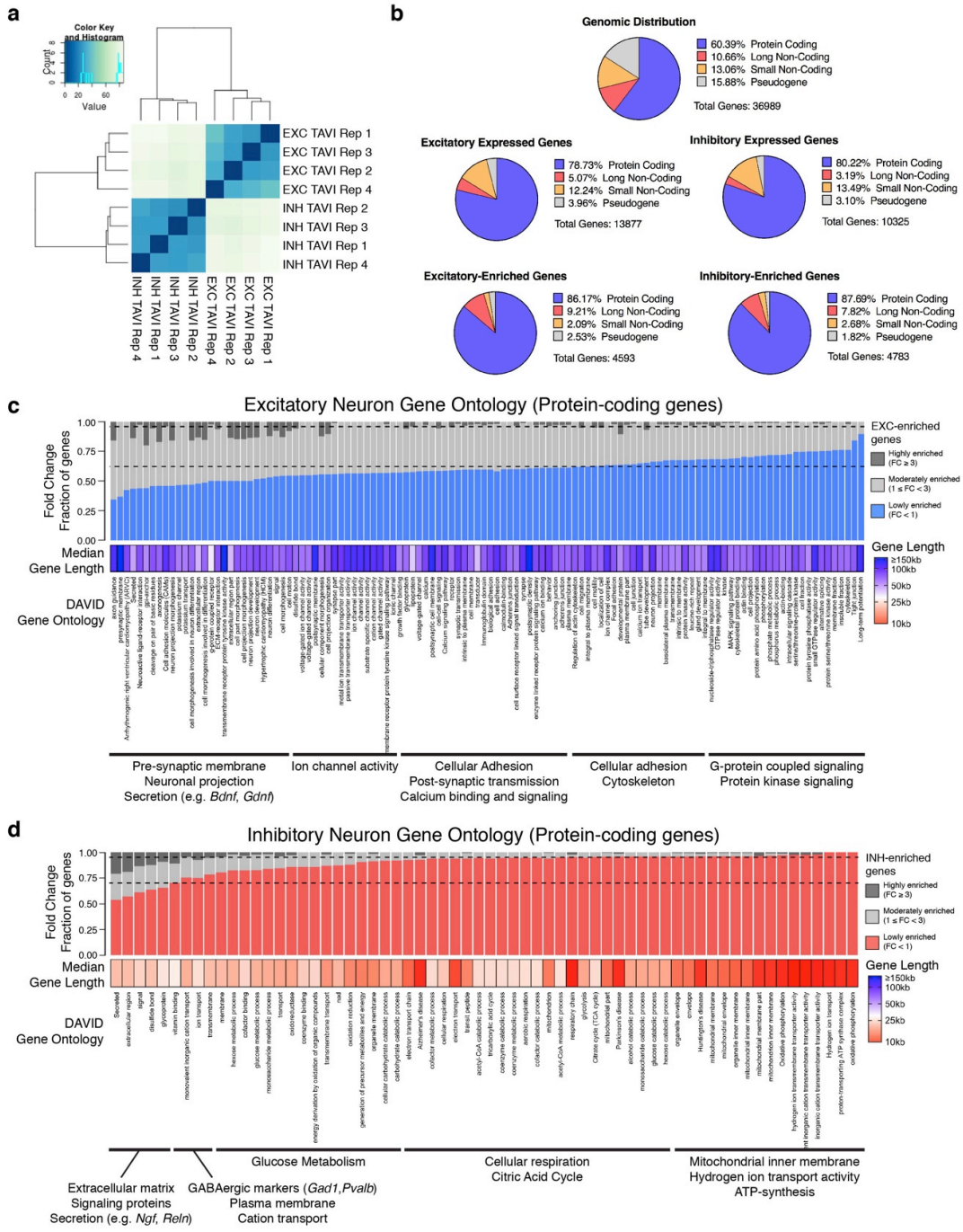
genotype, One-way ANOVA [F = 0.9238, P = 0.4721]). **(k)** Representative images showing immunofluorescent detection of biotinylated mutant MeCP2 in hippocampal tissues of untagged (WT) and tagged (T158M, R106W) male mice at 6 weeks of age. Tissue is probed with streptavidin for biotin detection and an antibody against MeCP2 C-terminus. Scale bars represent 10μm. **(l)** Immunofluorescent spatial colocalization of biotinylated mutant MeCP2 and nucleoli in hippocampal tissue of untagged (WT) and tagged (T158M, R106W) male mice at 6 weeks of age. Tissue is probed with streptavidin for biotin detection and an antibody against nucleolin. Scale bars represent 10μm. *P < 0.5, **P < 0.01, ***P < 0.001, ****P < 0.0001, n.s. = not significant; all pooled data depicts mean ± SEM. See also **Fig. 1**.



Supplementary Figure 2

Validation of NEX-Cre and Dlx5/6-Cre mouse lines **(a,b)** Representative images showing immunofluorescent detection of biotinylated MeCP2-Tavi protein in Cre-specified neuronal populations of the mouse cortex (a) and striatum (b). Probed using streptavidin for biotin detection and an antibody against the MeCP2 C-terminus. Scale

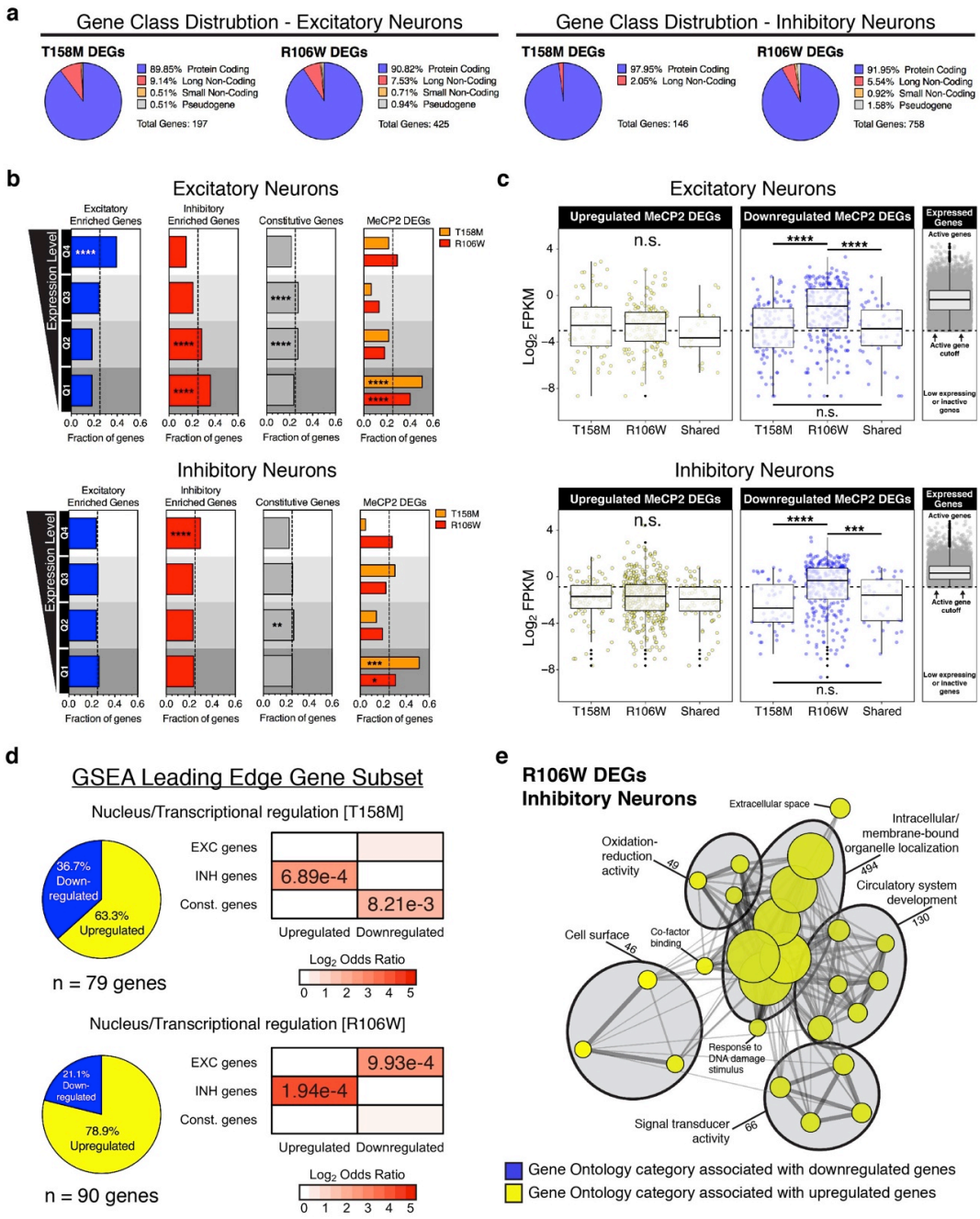
bars represent 100μm. **(c,d,e,f,g)** Representative confocal images depicting Cre-mediated biotinylation in NeuN+ (c), GAD67+ (d), Parvalbumin+ (e), Somatostatin+ (f), and Calretinin+ (g) neurons. Scale bars represent 50μm. **(h)** Quantification of Cre-mediated biotinylation in NeuN+, Parvalbumin+, or Calretinin+ neurons (n = 3, Two-tailed Unpaired Student t-test $P_{NeuN} < 0.0001$, $P_{Calretinin} < 0.0001$, $P_{Parvalbumin} = 0.0023$). **(i)** FACS analysis of labeled cortical nuclei populations. Data shown is representative of three independent experiments using Dlx5/6-Cre mice. Percentages indicate the mean distribution of neurons that are biotin+ (inhibitory; 14.2% ± 0.4) or biotin- (excitatory; 85.8% ± 0.4). **(j)** RT-PCR validation of FACS-isolated populations depicted in (i) ($n_{replicates} = 3$, Two-way ANOVA, Control [Cell Type-Gene Interaction, F = 353.8, $P < 0.0001$; Cell Type, F = 337.4, $P < 0.0001$; Gene, F = 203.2, $P < 0.0001$], Non-Neuronal [Cell Type-Gene Interaction, F = 3.535, $P = 0.0119$; Cell Type, F = 20.98, $P < 0.0001$; Gene, F = 2.973, $P = 0.0702$], EXC-specific [Cell Type-Gene Interaction, F = 0.9698, $P = 0.4312$; Cell Type, F = 351.1, $P < 0.0001$; Gene, F = 7.491, $P = 0.0146$], INH-spectific [Cell Type-Gene Interaction, F = 2.251, $P = 0.0274$; Cell Type, F = 901.3, $P < 0.0001$; Gene, F = 1.345, $P = 0.2703$]; Dunnett's multiple comparisons correction applied). $^{*}P < 0.5$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, n.s. = not significant; all pooled data depicts mean ± SEM. See also **Fig. 2**.



Supplementary Figure 3

Functional characterization of excitatory and inhibitory neurons **(a)** Unsupervised hierarchical clustering and heatmap showing correlations between excitatory and inhibitory neuronal nuclei from 6-week TAVI mice. **(b)** Distribution of protein coding

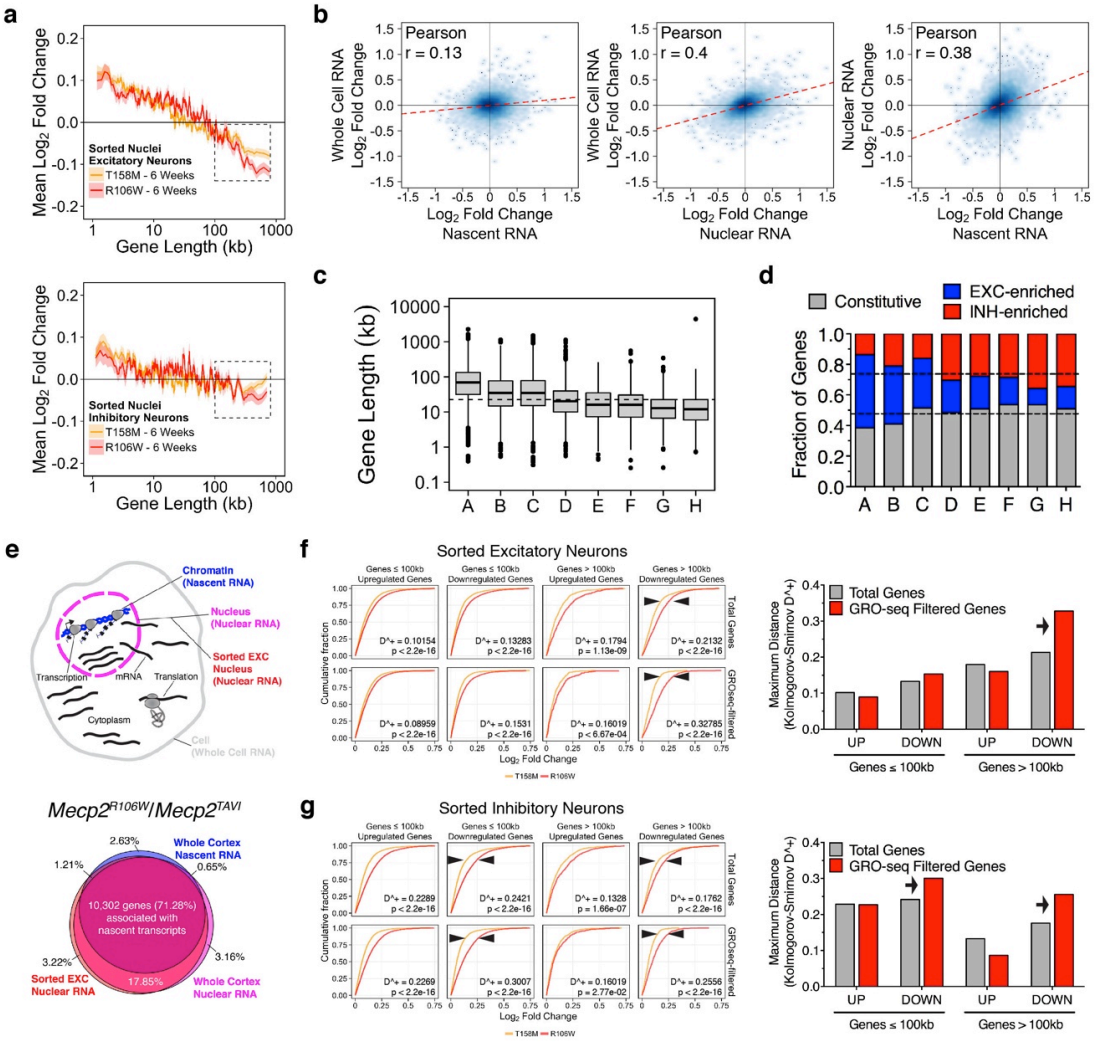
and non-coding genes across genomic, actively expressed, or cell type-enriched genes in each cell type. **(c)** DAVID gene ontology terms (Benjamini $P < 0.01$, FDR < 0.05) for EXC-enriched genes and their median gene lengths, all sorted by degree of fold change (low, moderate, high enrichment) relative to inhibitory neurons. Dotted line indicates the null proportion of low (62.3%), moderate (33.6%), and high (4.1%) enrichment among total EXC-enriched protein-coding genes (n = 3,968). **(d)** DAVID gene ontology terms (Benjamini $P < 0.01$, FDR < 0.05) for INH-enriched genes and their median gene lengths, all sorted by degree of fold change (low, moderate, high) relative to excitatory neurons. Dotted line indicates the null proportion of low (70.0%), moderate (25%), and high (5%) enrichment among total INH-enriched protein-coding genes (n = 4,194). See also **Fig. 2**.



Supplementary Figure 4

Transcriptional features characterize MeCP2 DEGs **(a)** Distribution of protein coding and non-coding genes among T158M or R106W DEGs in each cell type. **(b)** Bar plot showing significant enrichment of MeCP2 DEGs across binned actively expressed genes (One-tailed Fisher's Exact Test). Dotted line represents null distribution. Q4 represents

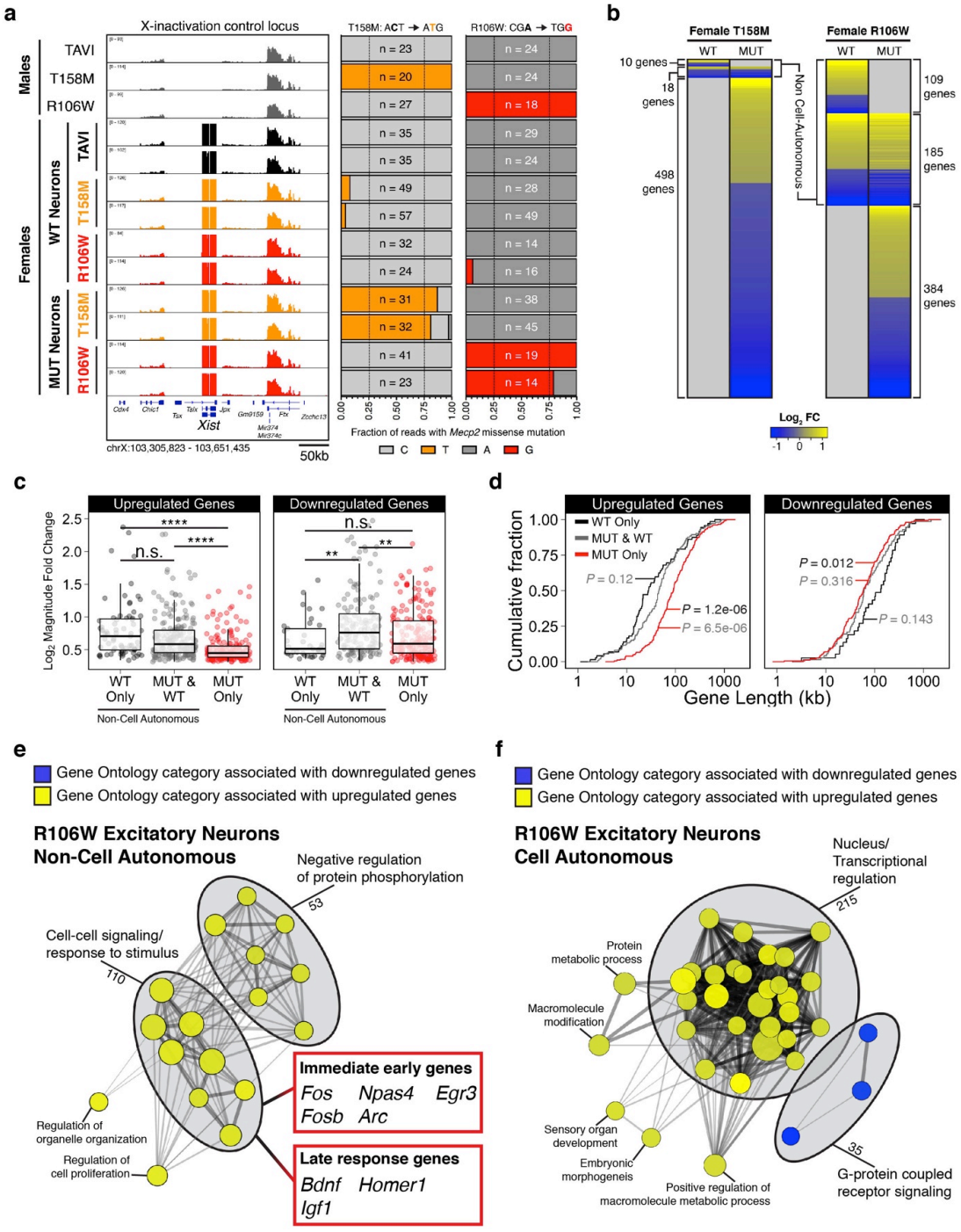
top 25% of actively expressed genes (high expression), Q1 represents bottom 25% of actively expressed genes (low expression). **(c)** Boxplots comparing $\log_2$ FPKM distributions between T158M, R106W, and shared DEGs, partitioned by upregulated and downregulated genes (Two-tailed Kruskal-Wallis Rank Sum, Excitatory $P_{\text{Upregulated}}$ = 0.4739 [χ^2 = 1.4937, df = 2], Excitatory $P_{\text{Downregulated}}$ = 6.276e-10 [χ^2 = 42.378, df = 2], Inhibitory $P_{\text{Upregulated}}$ = 0.3037 [χ^2 = 2.3831, df = 2], Inhibitory $P_{\text{Downregulated}}$ = 1.162e-08 [χ^2 = 36.542, df = 2]; Pairwise Wilcoxon Rank Sum *P* displayed). Additional boxplot shows $\log_2$ FPKM distribution of actively expressed genes for each cell type. **(d)** *Left*, Pie chart depicting the proportion of leading edge genes that are upregulated and downregulated among nodes related to transcriptional regulation (see Figure 3J). *Right*, Heat map depicting the preference of cell type-enriched genes among leading edge genes (One-tailed Fisher's Exact Test). **(e)** Enrichment map of pre-ranked GSEA functional network associations (*P-value* < 0.01, *Q-value* < 0.1). Data represents DEGs from R106W inhibitory neurons. Nodes denote functional categories, colored by NES. Line weight denotes extent of gene overlap between connected nodes. **P* < 0.5, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001, n.s. = not significant. See also **Fig. 3**.



Supplementary Figure 5

Extended Data for Figures 4 **(a)** Mean $\log_2$ fold change in 6-week R106W (red; n = 4) and T158M (orange, n = 4) sorted excitatory neurons (*top*) and inhibitory neurons (*bottom*); not filtered for GRO-seq detected genes. **(b)** Smoothed scatterplot comparing $\log_2$ fold changes between nascent and whole cell RNA fractions (*left*), nuclear and whole cell RNA fractions (*center*), and nascent and nuclear RNA fractions (*right*) in R106W mutant mice (n = 10,390 genes). **(c)** Boxplots depicting gene length for corresponding gene groups in **Figure 4F**, sorted by median gene lengths. Dotted line depicts median gene length for all 10,390 genes. **(d)** Barplot depicting fraction of protein-coding DEGs

classified as constitutive, EXC- or INH-enriched genes within the gene groups defined in **Figure 4F**. Dotted line represents null distribution among all 10,390 genes. **(e)** *Top*, Diagram of RNA distribution across subcellular compartments. *Bottom,* Area proportional Venn diagram comparing overlap in gene expression changes between sorted excitatory nuclear RNA, whole cell RNA, and nascent RNA. **(f)** *Left,* Genome-wide Kolmogorov-Smirnov maximum distance (D^+) between cumulative distributions of $\log_2$ fold changes between 6-week T158M and R106W sorted excitatory neurons. D^+ is summarized in bar plots (*right*). **(g)** Same as in (f), but using $\log_2$ fold changes between t-week T158M and R106W sorted inhibitory neurons. $^{*}P < 0.5$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, n.s. = not significant. See also **Fig. 4**.



Supplementary Figure 6

Cell and non-cell autonomous gene expression changes in RTT are functionally distinct

(a) *Left*, Browser snapshot of X-inactivation control genomic locus in both WT and MUT

neurons of TAVI, T158M, and R106W female mice at 18 weeks of age. *Right*, Percentage of reads detected with knock-in wild-type or mutant allele for each sorted population. **(b)** Heatmap displaying $\log_2$ fold changes among protein-coding DEGs in WT and MUT populations of T158M or R106W female excitatory neurons. **(c)** Boxplots comparing median $\log_2$ fold changes among cell and non-cell autonomous genes (Two-tailed Kruskal-Wallis Rank Sum, $P_{\text{Upregulated}}$ = 3.483e-14 [χ^2 = 61.976, df = 2], $P_{\text{Downregulated}}$ = 0.0015 [χ^2 = 12.983, df = 2]; Pairwise Wilcoxon Rank Sum *P* displayed). **(d)** Cumulative distribution function comparing gene lengths among cell and non-cell autonomous genes (Two-tailed Kruskal-Wallis Rank Sum, $P_{\text{Upregulated}}$ = 1.246e-08 [χ^2 = 36.401, df = 2], $P_{\text{Downregulated}}$ = 0.015 [χ^2 = 8.4161, df = 2]; Pairwise Wilcoxon Rank Sum *P* displayed). **(e)** Enrichment map of pre-ranked GSEA functional network associations (*P-value* < 0.01, *Q-value* < 0.1). Data represents non-cell autonomous DEGs in R106W female excitatory neurons. Nodes denote functional categories, colored by NES. Line weight denotes extent of gene overlap between connected nodes. **(f)** Same as in (e), but using data from cell autonomous DEGs in R106W female excitatory neurons. **P* < 0.5, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001, n.s. = not significant. See also **Fig.**

CHAPTER 3

A Rett Syndrome-associated mutation in *Mecp2* alters the epigenomic and chromatin architectural landscape of excitatory neurons

One Sentence Summary

Given the transcriptomic changes we observed in Chapter 2, this chapter evaluates epigenomic and chromatin architectural changes in *Mecp2* mutant mice, providing insight into why DEGs are differentially susceptible to dysregulation in RTT and positing MeCP2 as a key player in global maintenance of the methylome and chromatin architecture for the preservation of neuronal gene expression.

Results

DNA methylation changes associated with R106W Mice

Numerous studies investigating the genome-wide binding patterns of MeCP2 via chromatin immunoprecipitation followed by sequencing (ChIP-seq) in various mouse brain regions have found that MeCP2 binds broadly across the genome (Chen et al., 2015; Cohen et al., 2011; Gabel et al., 2015; Skene et al., 2010). Although broadly found, MeCP2 is preferentially enriched at methylated regions of the genome, correlating with both 5mCG (Chen et al., 2015; Cohen et al., 2011; Gabel et al., 2015; Skene et al., 2010) and 5mCH dinucleotide density (Chen et al., 2015; Gabel et al., 2015). Given that MeCP2 binds to methylated DNA and that DNA methylation patterns are cell type-specific, we sought to profile MeCP2 binding in a neuronal cell type-specific manner. Therefore, we investigated the cell type-specific binding pattern of MeCP2 with the use of a *Mecp2*$^{WT\text{-}Tavi/y}$;*R26*$^{cBirA/+}$;*NEX*$^{Cre/+}$ mouse line (subsequently referred to as WT), in which an affinity tag (TEV protease cleavage site and a biotinylation consensus motif, termed Tavi for *T*EV and *Avi*din-binding) on MeCP2 is biotinylated in a Cre-dependent manner in

forebrain excitatory neurons. With the use of streptavidin conjugated magnetic beads, we carried out MeCP2 ChIP-seq in cortical excitatory neurons of 6-week mice. We found that MeCP2 binds globally across the genome (Figure 1a), which is in agreement with previous reports (Chen et al., 2015; Cohen et al., 2011; Gabel et al., 2015; Skene et al., 2010). Also in line with published work, we find that MeCP2 is highly enriched at major satellite DNA and previously established high-affinity sites and lowly enriched at known low-affinity sites (Chen et al., 2015) (Figure S1A).

Given this global binding pattern and the large body of literature supporting that MeCP2 binds to methylated DNA, we next sought to determine the global effect of the R106W mutation on DNA methylation since this mutation in the methyl-CpG-binding domain (MBD) of MeCP2 is known to abolish its ability to bind to methylated DNA. A very simplistic model would predict that MeCP2 perseveres the methylation status of the cytosines to which it binds, and therefore, in the presence of a mutation like R106W that abolishes MeCP2's ability to bind to DNA, unbound cytosines could be subjected to hydroxymethylation and/or demethylation. Studies on this subject are few in number, with one group finding that levels of 5hmC anticorrelated with levels of MeCP2 in mouse cerebellar tissue (loss of MeCP2 resulted in an increase in 5hmC and overexpression of MeCP2 lead to an decrease in 5hmC) (Szulwach et al., 2011) and another group finding correlation between MeCP2 presence and 5hmC, with the absence of MeCP2 correlating with a reduction in 5hmC in mouse granule cells of the cerebellum (Mellén et al., 2012a). In addition to the conflicting conclusions of these findings, neither of these studies examined 5mC and 5hmC simultaneously, and therefore, it remains to be determined whether 5mC levels are altered, as well as the interplay between 5mC and 5hmC, in RTT. Therefore, with the use of streptavidin conjugated to a fluorophore, we applied a fluorescence-activated cell sorting (FACS)-based approach to isolate cortical excitatory nuclei in 6-week WT and $Mecp2^{R106W\text{-}Tavi/y}$;$R26^{cBirA/+}$;$NEX^{Cre/+}$ (subsequently referred to as R106W) mice. Whole genome bisulfite sequencing (WGBS) and Tet-assisted bisulfite sequencing (TAB-seq) was carried out on these nuclei to obtain genome-wide, base pair resolution profiles of 5mC and 5hmC, as integration of these two approaches enables distinction between the two. The average genomic sequencing coverage for WGBS and TAB-seq were ~14-fold (average cytosine coverage ~10 fold) (Figure S1B) and our methylation data is highly correlated (r = 0.94) with previously published WGBS in WT

excitatory neurons (Figure S1C). We find that the majority of 5mC and 5hmC are in the CG context in similar percentages to previously published work (Lister et al., 2013; Mo et al., 2015) (Figure S1D).

We found significant DNA methylation changes in R106W, with a global decrease in 5mC and increase in 5hmC in both the CG and CH contexts in R106W (Figure 1B-C & Figure S1E). Additionally, a significantly higher fraction of cytosines in the CH context are altered than those in the CG context for both 5mC and 5hmC (Figure S1F), suggesting that methylation in the CH context is more susceptible to demethylation in RTT. Given this global reduction in 5mC and increase in 5hmC, we next investigated whether the same cytosines that are losing 5mC are gaining 5hmC. We found that while 18% of cytosines with a decrease in 5mC are gaining 5hmC, the majority (55%) of cytosines with a reduction in 5mC are completely unmethylated (neither 5mC nor 5hmC) (Figure 1D). These findings support the simplistic model that loss of MeCP2 binding leads to a reduction in methylation and suggest that MeCP2 binding at CH is particularly important for the preservation of methylation in this context.

Next, we sought to determine whether the reduction in 5mC in R106W leads to an alteration in genome wide methylation states. Therefore, we employed MethylSeekR (Burger et al., 2013) to segment 5mC in the CG context into fully methylated regions (FMRs; average methylation is greater than 50%; associated with gene repression), low-methylated regions (LMRs; average methylation of 30%; which are associated with gene regulatory regions), unmethylated regions (UMRs; average methylation of 5.7%; which are associated with CpG islands), and partially methylated domains (PMDs; regions with disordered methylation)(Burger et al., 2013; Lister et al., 2009). In agreement with previous reports(Avrahami et al., 2015), FMRs comprise the majority of the genome, with ~2.1 Gbp found in this state (Figure S1G). In contrast to previous reports(Lister et al., 2013; Mo et al., 2015), PMDs are present in our data (~5.8Mbp), which is likely due to the methylation states in this current study being demarcated based solely on 5mC, rather than evaluating both 5mC and 5hmC simultaneously (Figure S1G). In R106W, in comparison to WT, there is an increase in the number of basepairs in LMRs, UMRs, and PMDs, and a decrease in the number of basepairs in FMRs (Figure S1G). This finding suggests that methylated regions are shifting to a lower methylated state and/or a more

disordered methylation state, and thus, implicates MeCP2 as necessary for the maintenance of FMRs, which are associated with repression.

Following our observation of global hypomethylation at the single cytosine resolution in R106W, we next sought to determine whether stretches of DNA have different methylation patterns in RTT. To do so, we defined differentially methylated regions (DMRs; consisting methylation changes in both CG and CH contexts combined) between WT and R106W. We identified 35,962 5mC DMRs (3,207 hyper-5mC-DMRs with a median length of 248 bp; and 32,755 hypo-5mC-DMRs with a median length of 319bp) and 8,925 5hmC DMRs (935 hyper-5hmC-DMRs with a median length of 246bp and 7,990 hypo-5hmC-DMRs with a median length of 246bp) (Figure 1E), which is in agreement with the genome wide changes we found at the single cytosine level. Next, we investigated whether differentially expressed genes (DEGs) in excitatory cortical neurons of R106W mice at 6 weeks of age are significantly associated with methylation changes. In previous work, we employed our FACS-based approach to profile the transcriptome of cortical excitatory nuclei of 6-week WT and R106W mice, leading to the identification of 425 DEGs. We found that the number of these DEGs that intragenically overlap with hypo-5mC-DMRs, but not any of the other DMR states, is significantly greater than chance (chance being 1,000 permutations of randomly selected intragenic regions that are number and length matched to DEGs) (Figure 1F). Furthermore, similar percentages of upregulated and downregulated DEGs are enriched for each of the DMR states (Figure S1H), suggesting that a loss of methylation is not simply a reflection of the directionality of gene expression changes in the mutant since loss of methylation is usually associated with gene upregulation, but not downregulation.

Following the simplistic model that predicts loss of MeCP2 binding would lead to demethylation of unbound cytosines, we reasoned that DEGs could preferentially be subjected to hypomethylation if they are enriched with MeCP2 occupancy in the WT setting. Therefore, we plotted the average MeCP2 read count in excitatory neurons in the gene bodies, plus 100kb upstream and downstream, of upregulated DEGs, downregulated DEGs, and all actively expressed genes in cortical excitatory neurons (Figure 1G). We found that both upregulated and downregulated DEGs are intragenically enriched for MeCP2 binding in comparison to all actively expressed genes. Additionally, upregulated DEGs have more MeCP2 enrichment than downregulated DEGs (Figure 1G).

Notably, these patterns are not observed when the average read count for the Input is plotted in a similar manner (Figure S1I), supporting that this enrichment is specific to MeCP2. These results suggest that MeCP2's ability to bind to methylated DNA preserves the methylation status of cytosines, and when MeCP2's ability to bind to methylated DNA is compromised, such as in the R106W mutant, methylated cytosines are subjected to hydroxymethylation and/or demethylation. Furthermore, these data suggest that DEGs are preferentially affected in RTT due to their enrichment of MeCP2 binding in the WT setting.

Given the enrichment of MeCP2 occupancy at DEGs in the WT setting and the reduction of methylation at DEGs in the mutant, we next sought to determine the relationship between DEGs and 5mC levels in the WT setting. To do so, we plotted average methylation levels intragenically, plus 1kb upstream and downstream, across upregulated DEGs, downregulated DEGs, and all actively expressed genes in cortical excitatory neurons (Figure 1H-I). While upregulated DEGs have higher levels of intragenic 5mCG and 5mCH methylation than actively expressed genes in the WT setting, downregulated DEGs do not. Additionally, upregulated DEGs have higher levels of intragenic 5mCG and 5mCH than downregulated DEGs, which is in agreement with upregulated DEGs having a greater enrichment of MeCP2 than downregulated DEGs (Figure 1G). These findings suggest that high average levels of 5mC, either in the 5mCG or 5mCH context, isn't the sole feature in determining MeCP2 enrichment, since both upregulated and downregulated DEGs are enriched with MeCP2 binding, but only upregulated genes have higher average levels of 5mCH and 5mCG than actively expressed genes in excitatory neurons.

Chromatin Features of Rett Syndrome-associated differentially expressed genes

Our finding that average intragenic levels of 5mCG and 5mCH are not the sole determinant of MeCP2 binding led us to determine common chromatin features of DEGs to gain insight into why these genes are enriched for MeCP2 binding in WT and preferentially susceptible to misregulation in Rett Syndrome. With the use of a Hidden Markov Model, ChromHMM (http://compbio.mit.edu/ChromHMM/) (Ernst and Kellis,

2012), multiple chromatin datasets were integrated to define chromatin states in the adult male mouse cortex. Chromatin immunoprecipitation sequencing (ChIP-seq) datasets from six histone marks were used, three of which were generated from 6-15-week mouse cortical tissue by our lab, H3K9me3 (associated with heterochromatin), H3K4me3 (associated with promoter regions), and H3K79me2 (associated with gene bodies), and three are publicly accessible datasets from 8-11-week mouse excitatory neurons of the cortex, H3K27ac (associated with promoter and enhancer activation), H3K4me1 (associated with enhancers), and H3K27me3 (associated with polycomb regions)(Mo et al., 2015). Additionally, we used publicly accessible CCCTF-binding factor (CTCF) ChIP-seq (Encode: ENCFF001LIJ) and Polr2a ChIP-seq (Encode: ENCFF001LIV) data from 8-week mouse cortical tissue. We also assessed open chromatin regions in excitatory neurons of the cortex from 6-week old mice (using our *Mecp2*$^{WT\text{-}Tavi/y}$;*R26*$^{cBirA/+}$;*NEX*$^{Cre/+}$ mouse line) with the assay for transposase accessible chromatin using sequencing (ATAC-seq), and we found that our ATAC-seq data highly correlates with previously published ATAC-seq data in excitatory neurons (r = 0.88) (Figure S2A) and that our biological replicates are well correlated (r = 0.986) (Figure S2B). Finally, we incorporated our MeCP2 ChIP-seq data from excitatory cortical neurons of 6-week with the use of our *Mecp2*$^{WT\text{-}Tavi/y}$;*R26*$^{cBirA/+}$;*NEX*$^{Cre/+}$ mouse line. Therefore, these data sets allow for the demarcation of chromatin states in cortical neurons, primarily excitatory neurons due to 85% of neurons in the cortex consisting of excitatory neurons and the fact that some of these datasets specifically assessed excitatory neurons, of 6-15-week male mice.

Using these various epigenomic datasets, ChromHMM identified 11 independent chromatin states (Figure 2A). We found that the number of DEGs intragenically enriched for heterochromatin, active enhancers, and CTCF chromatin states is significantly greater than chance (chance being random sampling (1,000 times) of genes that are number and length matched to DEGs) (Figure 2B). When the relative enrichments are compared across upregulated and downregulated DEGs, it was found that three states are differentially enriched: the polycomb and active enhancer chromatin states are found in a significantly higher percentage of downregulated DEGs and the active promoter state is found in a significantly higher percentage of upregulated DEGs (Figure 2C). Given that DEGs display intragenic enrichment for heterochromatin states,

we next sought to determine whether our DEGs are associated with lamin associated domains (LADs) since H3K9me 2/3 is enriched along the entirety of LADs(Guelen et al., 2008; Harr et al., 2015; Wen et al., 2009). To do so, we used publicly available data of lamin B1 associated domains from mouse neural precursor cells (NPCs) since this is the closest data available to mouse cortical tissue (Peric-Hupkes et al., 2010) and found that DEGs are not associated with LADs (Figure S2C). These data suggest that DEGs are associated with a unique chromatin environment, which may explain the susceptibility of these genes to misregulation in RTT. Furthermore, the differential enrichment of particular chromatin states at upregulated versus downregulated genes provides insight into the directionality of gene misregulation in RTT.

Given that DEGs are significantly intragenically associated with hypo-5mC-DMRs, we next sought to determine whether particular intragenic chromatin states are enriched for this DMR state. To do so, we isolated all chromatin state regions that overlap with DEGs and determined whether particular states significantly overlap with 5mC-hypo-DMRs. Heterochromatin, polycomb, and gene body states that are intragenically located within DEGs are significantly associated with hypo-5mC-DMRs (Figure 2D). We next determined which chromatin states are significantly associated with hypo-5mC-DMRs genome-wide and found similar states to be enriched: heterochromatin, polycomb, gene body, and low-signal 1 states (Figure S2D). Taken together, these data suggest that repressive states (heterochromatin and polycomb) are globally demethylated in the mutant and that the enrichment of heterochromatin within DEGs can partially explain their differential vulnerability to misregulation. Additionally, the enrichment of gene bodies with hypo-5mC-DMRs is suggestive of a potential global effect on genes.

Given our finding that DEGs are associated with active enhancer and CTCF chromatin states, we next sought to determine whether particular transcription factors (TFs) bind in or near DEGs (intragenic region plus 100kb upstream and downstream). To do so, we performed TF footprinting analysis (Piper et al., 2015; 2013) in our ATAC-seq peak regions of excitatory neurons and subsequently used HOMER(Heinz et al., 2010) to identify TF motifs significantly associated with these footprints. Using this approach, we identified 19 TFs that are associated with footprints in ATAC-seq peak regions of WT cortical excitatory neurons (Figure S2E). When evaluating these TF

footprints for significant enrichment with DEGs, we found that DEGs are significantly associated with CTCF, NeuroD2, and POU2F2 (Figure S2E). Furthermore, when the relative enrichments of these TFs are compared across upregulated and downregulated DEGs, we found that the ATF1 footprint is associated with a significantly higher percentage of upregulated DEGs than downregulated DEGs (Figure S2F). In addition to enrichment intragenically plus 100kb upstream and downstream, CTCF was also enriched intragenically plus 500kb upstream and downstream (Figure 2E). These findings suggest that certain TFs and architectural proteins, such as CTCF, are important in the regulation of DEGs, and furthermore, that particular TFs regulate upregulated versus downregulated DEGs.

CTCF Binding Site Alterations in R106W Mice

Given our findings that DEGs are significantly enriched with particular TF and architectural protein footprints, the CTCF chromatin state, and the active enhancer state, we next sought to investigate whether open regulatory regions are altered in the R106W mice compared to WT. To do so, we carried out ATAC-seq in cortical excitatory neurons of 6-week R106W male mice and compared the ATAC-seq profiles between WT and R106W. The ATAC-seq read coverage for WT and R106W were comparable (Figure S3A), and similar to the WT biological replicates, the R106W ATAC-seq biological replicates were highly correlated ($r = 0.985$) (Figure S3B). The number and genomic distribution of ATAC-seq peak regions are similar across WT and R106W (Figure S3C). When the genomic locations of ATAC-seq peak regions were statistically tested for differences in Tn5 transposase integrations (the enzyme used in ATAC-seq to fragment and insert adapters into open regions of the genome) between WT and R106W, only 5 regions were found to be significantly different (0.005% of all ATAC-seq peak regions). These findings suggest that the location of open regulatory regions does not change in R106W mice, which is in agreement with the subtle gene expression changes observed in the R106W mice.

Although the genomic location of open regulatory regions did not change in the mutant, it is possible that the recruitment and/or binding affinity of TFs is subtly

affected within these open regulatory regions, and that such a difference cannot be detected by simply comparing Tn5 transposase integrations at ATAC-seq peak regions. In order to assess this, we performed differential ATAC-seq footprinting with the use of Wellington-bootstrap (Piper et al., 2013; 2015) and found thousands of differential ATAC-seq footprints: 31,907 sites that are WT specific (sites that are lost in R106W; referred to as "R106W lost footprint sites") and 36,474 sites that are R106W specific (sites that are gained in R106W; referred to as "R106W gained footprint sites"). To increase the stringency of our differential footprints, we subsequently only evaluated footprints with a differential footprint score of 2 or higher, which equated to 3,459 R106W gained footprint sites and 2,471 R106W lost footprint sites. Visualizing average Tn5 transposase integrations across all R106W gained or lost footprint sites confirms that these are differential footprint sites, given the large difference in average Tn5 transposase integrations across the differential sites in WT and R106W (Figures 3A and Figure S3D). We then used HOMER(Heinz et al., 2010) to identify TF motifs significantly associated with these differential footprints. We identified 15 TF motifs that are significantly associated with R106W gained footprint sites and 10 TF motifs that are significantly associated with R106W lost footprint sites (Figure S3E-F). Since CTCF was the most significantly enriched TF motif, we focused subsequent analyses on R106W CTCF gained footprint sites. Visualizing Tn5 transposase integrations at CTCF differential footprint sites by either evaluating Tn5 transposase insertions at each of the R106W CTCF gained footprint sites independently (Figure 3B) or by averaging integrations across all R106W CTCF gained footprint sites (Figure 3C) confirmed that these footprints are more enriched in R106W than WT.

To corroborate that the R106W CTCF gained sites are enriched for CTCF binding in R106W in comparison to WT, we next performed CTCF ChIP-seq in cortical tissue from WT and R106W male mice at 6 weeks of age. When the average CTCF ChIP-seq read counts at the R106W CTCF gained footprint sites are compared between WT and R106W, CTCF read count is greater at these sites in R106W than WT (Figure 3D). This supports that these R106W CTCF gained footprint sites are enriched for CTCF binding in R106W compared to WT. One plausible explanation for the gain of CTCF binding at these regions in R106W could be that these sites are losing methylation in R106W, and therefore, enabling the binding of the methylation-sensitive CTCF at these sites. To test

this, we compared the average 5mC methylation levels at each cytosine from the center of the R106W CTCF gained footprint sites plus 1kb upstream and downstream in WT and R106W and found a significant reduction in average methylation levels (p=0.0002501, Fisher's exact test) in R106W. And following our simple model that predicts that MeCP2 binding preserves the methylation status of cytosines to which it binds, we next investigated whether MeCP2 binding is enriched at these R106W CTCF gained footprint sites in the WT setting. Therefore, we analyzed the average MeCP2 ChIP-seq read count at the R106W CTCF gained footprint sites plus 2kb upstream and downstream and found that MeCP2 is enriched at these sites (Figure3E). These findings suggest that the loss of MeCP2 binding at these regions leads to demethylation and/or accessibility of these sites to CTCF binding.

We next evaluated the genomic distribution of R106W CTCF gained sites and found that they are located on every chromosome, in both intergenic and intragenic regions (Figure 3F). We subsequently sought to determine whether DEGs are enriched for R106W CTCF gained sites. We found that a higher percentage of DEGs are associated with R106W CTCF gained sites than chance (Figure 3G). Additionally, DEGs overlap with more R106W CTCF gained sites than chance (Figure 3G). However, when upregulated and downregulated DEGs were evaluated independently, a significant percentage of downregulated genes were associated with R106W CTCF gained sites than chance, but not upregulated DEGs (Figure 3H and Figure S3G). These data suggest that the R106W mutation leads to a loss of methylation at sites that are bound by MeCP2 in the WT setting, which therefore leads to CTCF binding at a portion of these sites. Furthermore, downregulated DEGs, but not upregulated DEGs, are enriched for CTCF gained sites, providing insight into the directionality of gene misregulation in R106W mice.

Nuclear Architectural Alterations in R106W mice

The observation of genome-wide R106W CTCF gained sites, along with the known importance of CTCF in genome organization(Ong and Corces, 2014), led us to next examine whether there are detectable nuclear architectural changes in RTT. To do so, we isolated nuclei from WT and R106W mouse cortical tissue and visualized various

genomic regions with the use of Oligopaint FISH probes, which allow for highly specific and efficient genomic targeting (Beliveau et al., 2012). We began by investigating the extent to which global alterations occur in the two large-scale chromatin organizational components, heterochromatin and euchromatin. Heterochromatin is a compact conformational state with low levels of transcription, and euchromatin is an open, transcriptionally active conformational state (Becker et al., 2016).

We first investigated the euchromatic conformational state by designing primary Oligopaint FISH probes that span the entire euchromatic region of chromosome 9 (Chr9) and subsequently using secondary oligo probes conjugated to a fluorophore for visualization of euchromatic Chr9 (Figure 4A). Chr9 was chosen because of its intermediary size (12th largest chromosome out of 21 chromosomes). We also used streptavidin conjugated to a fluorophore to visualize and restrict analysis to one cortical cell type, excitatory neurons (Figure 4A). We first measured the volume of the nucleus to determine whether our FISH protocol preserves the known phenotype of reduced nuclear volume in RTT (Mazumder et al., 2008b), and we found that the volume of R106W nuclei is significant reduced compared to WT (Figure 4B). Chr9 volume is also significantly reduced in R106W nuclei (Figure 4C), and when Chr9 volume is normalized to nuclear volume, the effect is nullified (Figure 4D). This suggests that Chr9 is condensing in R106W nuclei and that the reduction in Chr9 volume is proportional to the reduction in nuclear volume. This general relationship between chromatin condensation and nuclear size is supported by previous work (Linhoff et al., 2015; Mazumder et al., 2008a; Shen et al., 1996). Additionally, we observed an elongation of Chr9 (ratio of longest axis/shortest axis) in R106W (Figure S4A). Finally, we investigated the nuclear location of Chr9 by segmenting the nucleus into 5 regions of equal volume from the periphery to center (Figure S4B) and computing the fraction of Chr9 found in each nuclear segment. There was found to be a small, but significant, shift of Chr9 towards the periphery of the nucleus (Figure S4C). Taken together, these data support Chr9 condensation, as well as slight changes in Chr9 shape and nuclear positioning, in RTT.

Given these euchromatic observations, we next investigated the extent of heterochromatic changes in R106W nuclei. A major component of heterochromatin is repetitive DNA sequences, existing primarily in two forms: pericentric major satellites

and centromeric minor satellites. Therefore, we visualized heterochromatin by designing primary Oligopaint FISH probes that specifically target mouse major or minor satellites, as well as designing secondary oligo probes that uniquely bind to these two regions with different conjugated fluorophores (Figure 4E). Similar to euchromatic Chr9 volume, both major and minor satellite volume (both total and average volume per foci) was significantly reduced in R106W (Figure 4F-G and S4D-E) However, differing from euchromatin, when major and minor total volume was normalized to nuclear volume, the reduction in volume remains significant (Figure 4H and S4). This suggests that the reduction in volume for major and minor satellites is greater than the reduction in nuclear volume. We also observed a significant reduction in major and minor satellite number (Figure 4I and S4G), which might be explained by the significant reduction in average distance between major and minor satellites (Figure 4J and S4H), suggesting that similar organizational compartments are merging together. There is also an increase in merging across major and minor satellites compartments, given the significant increase in major and minor overlap (Figure 4K). Finally, there was found to be a slight, but significant, shift in major and minor satellite positioning toward the center of the nucleus (Figure S4I-J). Therefore, these findings support condensational and organizational defects in heterochromatin in RTT.

Given these alterations in euchromatin and heterochromatin, we next sought to determine whether we could detect organizational changes at the gene level, comparing a DEG to a gene of similar length (referred to as Control) whose expression does not significantly change in R106W excitatory neurons. We chose to examine two genes on Chr9 given our prior findings of condensation of the euchromatin regions of this chromosome: *Bmper* was chosen as the DEG (length: 262,127 bp and fold change (FC): 1.3) and *Ncam1* was chosen as the Control (length: 294,562bp). Oligopaint probes were designed to span the entire intragenic regions of these genes (Figure 5A). Differing from what was observed for the euchromatic region of Chr9 or heterochromatin, both the DEG (*Bmper*) and Control (*Ncam1*) volume increased in R106W (Figure S5A-B) and this increase remained significant after normalizing for nuclear size (Figure 5B-C). Next, given our bioinformatics findings that DEGs intragenically overlap with the heterochromatic chromatin state, we next determined the distance from the center of each gene to the center of the nearest major satellite. In agreement with our

bioinformatics data, we find that the DEG (*Bmper*) is closer to a major satellite than the control gene (*Ncam1*) (Figure 5D and Figure S5C) in the WT setting; however, this difference was not observed in the mutant setting (Figure S5D-E). Therefore, we reasoned that the distance of at least one of the genes to its nearest major satellite is perhaps changing in the RTT. When the distance of each gene to its nearest major satellite was compared between WT and R106W, we found that although the raw distance does not change for either gene (Figure S5F-G), the normalized distance of the DEG (*Bmper*) to its nearest major satellite significantly increases (Figure 5E and S5H). Taken together, these data suggest that gene bodies, for both DEGs and Controls alike, are decondensing in RTT, and that the proximity of a DEG to major satellite could be a contributing factor to misregulation in RTT. Additionally, we garnered insight into the directionality of gene misregulation in RTT, since this upregulated DEG shifts significantly away from repressive heterochromatin, which could contribute to transcriptional upregulation.

Next, we sought to investigate another DEG and Control gene to determine whether similar differences are observed. This time we chose a downregulated DEG, *Sgcd* (length: 1,092,717bp, RPKM: 4.9, and FC: 0.79), and a Control gene of similar length and expression levels, Gpc6 (length: 1,054,233 and RPKM: 6.8). This downregulated DEG is also associated with a R106W CTCF gained footprint site. Similar to what was observed above for the other DEG and Control pair, both genes increased in volume in RTT (Figure S5I-J) and this remained significant after normalizing for nuclear size (Figure 5F-G). Another characteristic similar to the other gene pair, the DEG (*Sgcd*) is closer to a major satellite than the Control gene (*Gpc6*) in WT (Figure 5H and S5K), and in R106W, the distances are similar between the DEG and Control (Figure S5L-M). However, differing from what was observed previously, when measuring the distance of a gene to its nearest major satellite across WT and R106W, the downregulated DEG distance remains unchanged (Figure 5I and S5N) and the Control gene distance increases (Figure S5O), although this effect is lost when normalized to nuclear volume (Figure S5P). Collectively, this corroborates the gene body decondensation, for both DEGs and Control genes alike, in RTT, and supports that DEGs are located in closer proximity to major satellites than Controls. Moreover, the downregulated DEG stays in similar proximity to its nearest major satellite, which is in opposition to the upregulated

DEG shifting significantly away from its nearest major satellite, implicating this proximity to a repressive state as a plausible explanation the directionality of gene misregulation in RTT.

To further substantiate the observed increase in nuclear volume of gene bodies of both DEGs and Controls in RTT, we used Oligopaint probes to intragenically labeled 11 DEGs of varying lengths, expression levels, and FC, as well as 11 matched Controls with similar lengths and expression levels (Figure S6A). Using this approach, we again find an increase in volume for both DEGs and Control genes in R106W nuclei (Figure 6A-B and Figure S6B-C). We also evaluated the nuclear location of DEGs and Controls across WT and R106W and found a slight, but significant, shift of both DEGS and Controls towards the center of the nucleus (Figure S6D-E), suggestive of global organizational alterations in RTT. When the nuclear location of DEGs was compared to that of Controls, it was found that DEGs are significantly located more centrally than Controls in both WT and R106W (Figure 6D and Figure S6F), which is in agreement with DEGs being in closer proximity to major satellites than Controls since the majority of major satellites are located centrally.

Finally, given our findings of decondensation of both DEGs and Controls by visualizing the entire gene body, we next sought to determine whether this increase in volume could also be found by measuring the distances between different regions of the gene body, since an increase in volume should correlate with an increase in distance between different regions of the gene. To do so, we designed Oligopaint probes that target 100kb regions at the 5' end, Middle (Mid), and 3' end of a downregulated DEG (*Kcnip4*; length: 1,135,403bp, RPKM: 33.7, FC: 0.76) and Control Gene (*Tenm2*; length: 1,229,39bp and RPKM: 29.5) (Figure 6E). This downregulated DEG is associated with a R106W CTCF gained footprint site. To gain insight into the conformation of these genes, we measured the distances across all combinations of intragenic locations (5' to Mid, Mid to 3', and 5' to 3') and compared the 5' to 3' distance to that of the sum of 5' to Mid plus Mid to 3' (referred to as 5' to 3' Summed) and found that the 5' to 3' distance is shorter than the 5' to 3' Summed distance (Figure S6G), suggesting that both of these genes are not in a linear conformation, but rather a loop-like conformation. First, we compared the sum of all distances (5' to Mid plus Mid to 3' plus 5' to 3') across WT and R016W and found that this summed distance is greater in R016W for both the DEG

(*Kcnip4*) and Control (*Tenm2*) (Figure S6H), which confirms what we previously found with the volume expansion of two pairs of genes. Next, we measured the distances between each of the intragenic regions in both the DEG (*Kcnip4*) and Control (*Tenm2*) across WT and R106W. The distance between 5' to Mid remained unchanged for both genes (Figure 6F) and both genes demonstrate a significant increase in 5' to 3' distance (Figure 6G) across WT and R106W. However, the DEG (*Kcnip4*), but not Control (*Tenm2*), displayed an increase Mid to 3' distance (Figure 6H) across WT and R106W. Given this DEG-specific increase in the Mid to 3' region of the gene, we next evaluated the distances between Mid to 3' in the WT setting between the DEG (*Kcnip4*) and Control (*Tenm2*) and found that the distance between the Mid to 3' of the DEG is the only distance significantly shorter in the DEG (Figure S6I), suggesting that perhaps this is a unique feature of DEGs and that disruption of this tight chromatin unit might be contributing to gene misregulation. This is particularly intriguing given that the R106W CTCF gained site is located nearest to the 3' end of DEG, suggesting that this gain in CTCF binding could possibly be contributing to the alteration in chromatin compaction.

Materials and Methods

ChIP-seq library construction and mapping

MeCP2, H3K4me3, and CTCF ChIP-seq were generated from 3 cortices of 6-week male mice (*Mecp2*$^{WT\text{-}Tavi/y}$;*R26*$^{cBirA/+}$;*NEX*$^{Cre/+}$ or *Mecp2*$^{R106W\text{-}Tavi/y}$;*R26*$^{cBirA/+}$;*NEX*$^{Cre/+}$). H3K79me2 and H3K9me3 ChIP-seq were generated from 3 cortices of 15-week old male mice. 50-200ug of chromatin was used per IP, which was performed as previously described in Chapter 2 of this dissertation. Libraries were generated using NEB enzymatic reagents (End Repair, 5' Adenylation, Adapter Ligation) with Illumina Truseq adapters according to Illumina ChIP-seq prep kit manufacturer instructions. Libraries were sequenced on a HiSeq 2500 (50SE). The ChIP-seq FASTQ files were mapped to the mouse mm10 genome with bowtie using "-v 2 -m 1". Only uniquely mapped reads were included for further analysis.

MeCP2 ChIP-RT-PCR

The following primers were used:

1) Major Satellite: forward 5′-catccacttgacgacttgaaaa-3′, reverse, 5′-gaggtccttcagtgtgcattt-3′
2) high-affinity site 1: forward 5′-atgttccagggatgatgaactt-3′, reverse 5′-gctacgcagtaga-gacatgctg-3′
3) high-affinity site 2: forward 5′-atgtagacgtacacgcccttg-3′, reverse 5′-agttctgactgtcctggaactg-3′
4) low-affinity site 1: forward 5′- ccagctactcgcggctttacgg-3′, reverse 5′-cctcccgccctgcttatccagt-3′
5) low- affinity site 2: forward 5′-ggaagtgaggacagaaatggac-3′, reverse 5′-attggctcataccaaaccagag-3′

TAB-seq and WGBS library construction, mapping, and analysis

Genomic DNA was isolated from nuclei of cortical excitatory neurons using an AllPrep DNA/RNA Mini kit (Qiagen cat. #80204) and subsequently treated with RNase (Roche cat. # 11119915001). TAB-seq and WGBS libraries were generated from 400ng and 300ng of DNA, respectively, as previously reported (Yu et al., 2012a). Trim Galore was used to remove the adaptor contamination from the raw sequencing reads using "-stringency 2 --length 36". The trimmed reads were mapped to mm10 genome by Bismark[41] using parameters "-n 2 -l 40". Clonal reads were excluded to avoid PCR artifacts. Methylation calling was performed by Binomial distribution followed by the Benjamini-Hochberg correction. The mC level for a cytosine was calculated as the WGBS methylation levels subtracted the hmC level in that cytosine.

MethylSeekR

Identification of UMRs, LMRs, and PMDs was carried out as previously described (Burger et al., 2013). Unique to our analysis was the demarcation of states based solely on 5mC, rather than evaluating both 5mC and 5hmC combined given that we performed both WGBS and TAB-seq, allowing for the distinction between the two methylation marks. Our α distribution had a distinct long tail, and therefore, we identified and masked PMDs genome-wide. To identify FMRs, all UMRs, LMRs, and PMDs were excluded from the genome using bedtools subtract, and the remaining sequences were taken to be FMRs.

ATAC-seq library construction

Five hundred thousand cell type-specific nuclei were sorted and subsequently pelleted as previously described (Jiang et al., 2008). Briefly, 1.6 mL of nuclei in PBS were mixed with 372 μL of pelleting buffer (360 μL of 2M Sucrose, 10 μL of 1M $CaCl_2$, and 2 μL of 3M $Mg(Ac)_2$) by gentle inversion, incubated on ice for 15 min, and centrifuged for 15 min at 5,000 rpm at 4°C. Nuclei were resuspended in 60 μL of 1x TD Buffer (Illumina Cat #FC-121-1030) and counted using a hemocytometer. 50,000 nuclei were brought up to 47.5 μL with 1x TD buffer and transposed with 2.5 μL of transposase (Illumina Cat #FC-121-1030) for 30 min at 37°C. ATAC-seq libraries were constructed exactly as laid out previously (Buenrostro et al., 2013), with a few exceptions. First, only 2.0 μL of 25 μM customized Nextera PCR Primers were used during PCR amplification. Second, the additional number of PCR cycles for each library was kept to 3-4 cycles (step 6 under III. PCR Amplification). Third, the final amplified library was purified using AMPure XP beads (Beckman A63881) to remove larger fragments (> 800 bp) by using 0.5X beads to remove larger fragments, removing the supernatant, adding 1.6X beads to the supernatant, and subsequently following the AMPure XP bead (Beckman A63881) instructions for DNA isolation. Libraries were analyzed on a Bioanalyzer (Agilent) and sequenced on an Illumina HiSeq (100 PE).

ATAC-seq mapping

The mate1 and mate2 files of the paired-end sequencing were mapped separately to the mm10 genome by bowtie[38] using "-v 2 -m 1". If both mates of a read pair were mapped to the opposite strands of the same chromosome and the distance in-between was less than 2 kb, the read pair was included for further analysis.

Random Sampling for determining significant enrichment

Bedtoools shuffle with the –incl flag was used to select number- and length-matched regions that were then intersected with features of interest. Carrying this out 1,000 times led to the generation of a random distribution from which a population mean and standard deviation was obtained for the generation of a z-score that was then converted into a two-tailed p-value.

Footprint and differential footprint analysis

ATAC-seq Fastq files were merged across biological replicates and Epinomics' ATAC-Seq platform (https://open.epigenomics.co/) was used for adapter trimming, alignment, filtering (removal of duplicates and mitochondrial reads). Footprinting was performed using the pyDNase (http://pythonhosted.org/pyDNase/) (Piper et al., 2013; 2015) script wellington_footprints.py with default parameters and the–A flag to indicate that the data come from ATAC-seq. Footprints with a p-value cutoff of e^{-20} were subsequently used for analysis. Footprints were extended 10bp upstream and downstream, and Homer (Heinz et al., 2010) was used for *de novo* TF motif discovery using findMotifsGenome.pl with the flag –size given. For differential footprint analysis, the pyDNase script wellington_bootstrap.py was used to determine WT-specific andR106W-specific footprints with default parameters and the –A flag. To create a background file for motif enrichment, WT and R106W ATAC-seq peaks that did not contain either a WT-specific or R106W-specific footprint were subjected to footprinting

using the pyDNase script wellington_footprints.py, and footprints with a p-value cutoff of e^{-20} were subsequently used for analysis. All footprints were extended 10bp upstream and downstream prior to *de novo* TF motif discovery with Homer with the background files specified using findMotifsGenome.pl with the flags –h, -size given, and -bg. To determine the location of TFs of interest, Homer's annotatePeaks.pl script was used with the flag –size given. For the generation of plots that display the average Tn5 integrations around a set of regions, the pyDNase script dnase_average_profile.py was used with the –A flag. For the generation of the heatmaps displaying the raw data, the pyDNase script dnase_to_javatreeview.py was used along with the –A flag and –o flag, which orders the differential footprint from most differential at the top to least differential at the bottom. JavaTreeView was used for heatmap visualization.

ChromHMM Modeling

Modeling was followed exactly as laid out in the ChromHMM manual (http://compbio.mit.edu/ChromHMM/) (Ernst and Kellis, 2012), with the specification of 11 states.

RP30M, RPKM, and FPKM

The number of sequencing reads or read-pairs mapped to the whole gene bodies (GRO-seq, ChIP-seq, ATAC-seq, and total RNA-seq) or the exonic regions (mRNA-seq) of each gene were counted by in house Perl programs. RP30M, read counts per 30 million uniquely mapped reads or read-pairs. For RPKM and FPKM, edgeR was used to normalize the read counts to obtains the count per million (CPM) values, and then the CPM values were further normalized by the length of whole gene (GRO-seq, ChIP-seq, ATAC-seq, and total RNA-seq) or the total length of all exons (mRNA-seq) to obtain the values of RPKM (reads per kilobase per million mapped reads, single-end sequencing data) and FPKM (fragments per kilobase per million mapped read-pairs, paired-end

sequencing data). A gene was defined as an expressed gene if its GRO-seq RPKM > 0.5 and as a non-expressed gene if its GRO-seq RPKM < 0.01.

Metagene analysis (heatmaps and line plots)

The total numbers of uniquely mapped reads or read-pairs for all sequencing data were normalized to 30 million. The signal represents the RP30M per base pair or per bin. For bin analysis, the gene body, upstream, and downstream regions were divided into 1000 equal bins for each gene, respectively, and the number of sequencing reads for each bin was calculated and normalized. For the heatmaps, the normalized bin signals were used to generate the heatmaps. For the line plots, the median value of the normalized bin signals of a given bin of that gene group was used to generate the plots. Heatmaps were generated by heatmap.2 function in the gplots R package.

Oligopaint probe design and synthesis

Probes were designed as previously described (Beliveau et al., 2012), with one exception: bowtie2 was used instead of OligoArray. CustomArray was used for Oligopaint library synthesis as previously described (Beliveau et al., 2012). Oligopaint probe synthesis was carried out as previously described (Beliveau et al., 2015).

Oligopaint FISH

A single cortex per biological sample was homogenized, and the homogenate was treated with 1% formaldehyde for 5 minutes and subsequently quenched with glycine. Nuclei were then isolated as previously described in Chapter 2 and resuspended in PBS. Nuclei in PBS were then placed on Polysine slides (Thermo Scientific™, cat. #6776215) for 30 minutes and subsequently fixed onto the slides with 4% PFA for 5 minutes,

followed by rinsing with PBS for 10 minutes. Nuclei on slides were stored in PBS at 4°C until use. FISH was carried out as previously described (Beliveau et al., 2015; 2012). Briefly, slides in PBS were warmed to room temperature (RT). Slides were then incubated in 2X SCCT for 5 min at RT, 50% formamide in 2X SCCT for 5 min at RT, 50% formamide in 2X SCCT for 2.5 minutes at 92°C, and 50% formamide in 2X SSCT for 20 minutes at 60°C. Nuclei were then hybridized with 50pmol of Oligopaint probe in a hybridization mixture consisting of 2X SSCT, 50% formamide, 10% (w/v) dextran sulfate, and 10mg of RNase. A coverslip and rubber cement were used to seal the hybridization mixture. Slides were then placed on a 92°C heat block for 2.5 min to enable denaturation, followed by overnight hybridization in a humidified chamber at 42°C. Slides were then washed in 2X SSCT for 15 minutes at 60°C followed by two 10 min 2X SCCT RT washes. Nuclei were then hybridized with 10pmol of secondary oligos conjugated to fluorophores in hybridization mixture, sealed with a coverslip and rubber cement, and hybridized for 2 hours at RT. Slides were then washed in 2X SSCT for 15 minutes at 60°C followed by two 10 min 2X SCCT RT washes. Slides were then stained with streptavidin conjugated to a fluorophore (1:1000) for 2 hr at RT in 2X SCCT. Slides were then stained for 5 minutes with DAPI (1:5000) and washed for 10 min in 2X SCCT at RT twice. Slides were then mounted and imaged.

Oligopaint analysis

Images were analyzed using TANGO (Tools for Analysis of Nuclear Genome Organization) (Ollion et al., 2013). Downstream data analysis was performed only on excitatory neurons, which was determined by measuring the intensity of streptavidin staining.

Figures

Figure 1

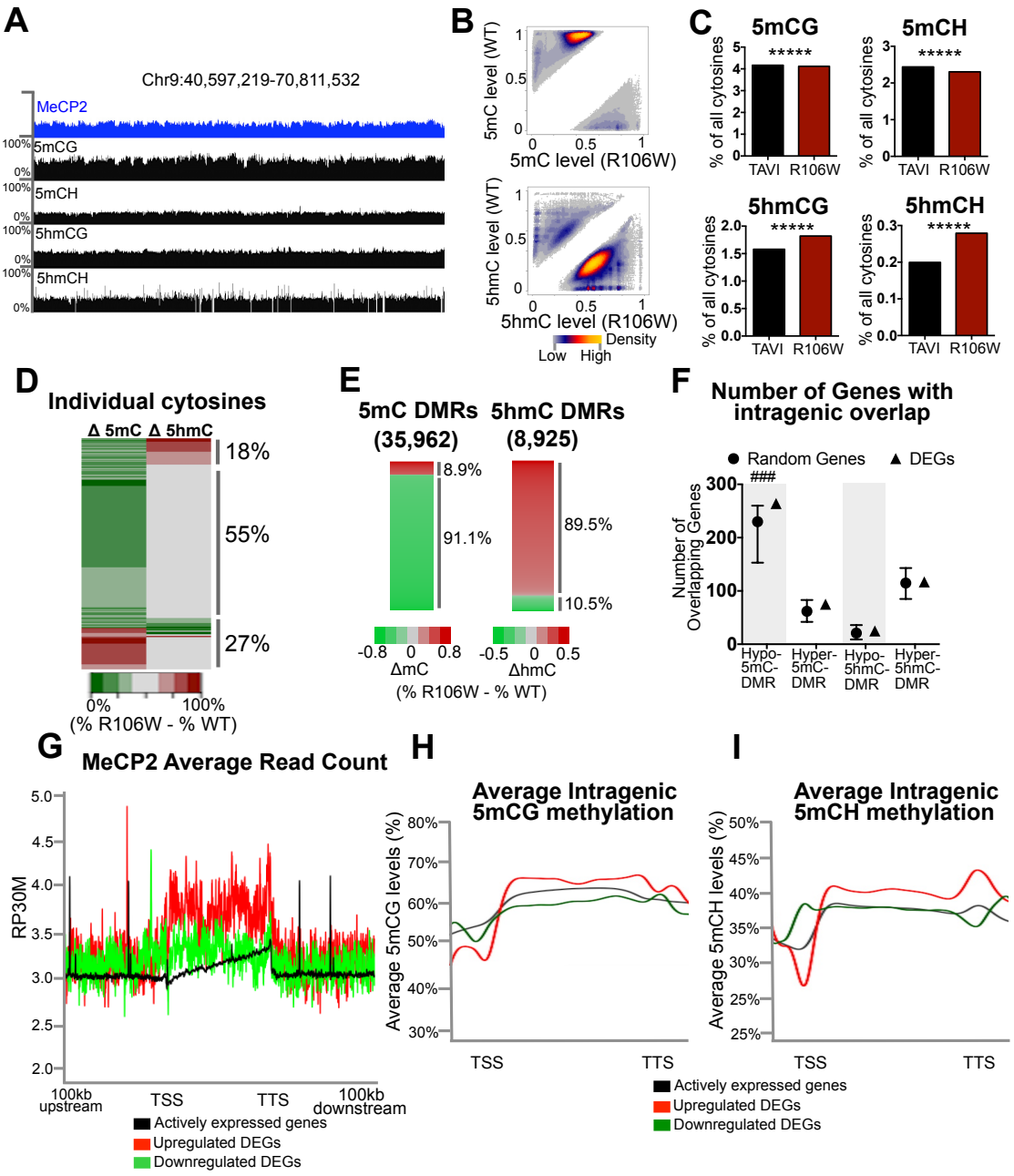


Figure 1. Global Hypomethylation is significantly association with DEGs in R106W mice. (A) Browser representation of MeCP2 binding and the four major forms of methylation in cortical excitatory neurons across a region of Chr9. For the DNA methylation tracks, the height of each track represents % methylation (0-100%). **(B)**

Top: Heat-scatterplot representation of 5mC differentially methylated cytosines in WT and R106W mice. Bottom: Heat-scatterplot representation of 5hmC differentially methylated cytosines in WT and R106W mice. Each dot represents a single differentially methylated cytosine. **(C)** Comparison of the percentage of all cytosines in the CG (left) or CH (right) context that are 5mC (top) or 5hmC (bottom) in WT versus R106W excitatory cortical neurons (*****p < 2.2e-16, Fisher's exact test). **(D)** Heatmap representation of differentially methylated cytosines in R106W mice. Each line represents a single differentially methylated cytosine. **(E)** Heatmaps of differentially methylated regions (DMRs) in R106W. **(F)** Number of DEGs vs Random Genes (random sampling (1,000 times) of genes that are number and length matched to DEGs) that overlap with each of the 4 types of DMRs. The mean, maximum, and minimum numbers of overlapping genes are displayed for the Random Genes (###p <0.005, two-tailed test). **(G)** MeCP2 read count from our MeCP2 ChIP-seq in excitatory neurons in the gene bodies, plus 100kb upstream and downstream, of upregulated DEGs, downregulated DEGs, and all actively expressed genes in cortical excitatory neurons. The gene body, upstream, and downstream regions were divided into 1000 equal bins for each gene and the number of sequencing reads for each bin was calculated and normalized. RP30M: read counts per 30 million. **(H)** 5mCG or **(I)** 5mCH gene body methylation levels across upregulated DEGs, downregulated DEGs, and all actively expressed genes in WT cortical excitatory neurons. The gene body region was segmented in 1000 equally sized bins for each gene and the average cytosine methylation level for each bin was calculated and normalized.

Figure 2

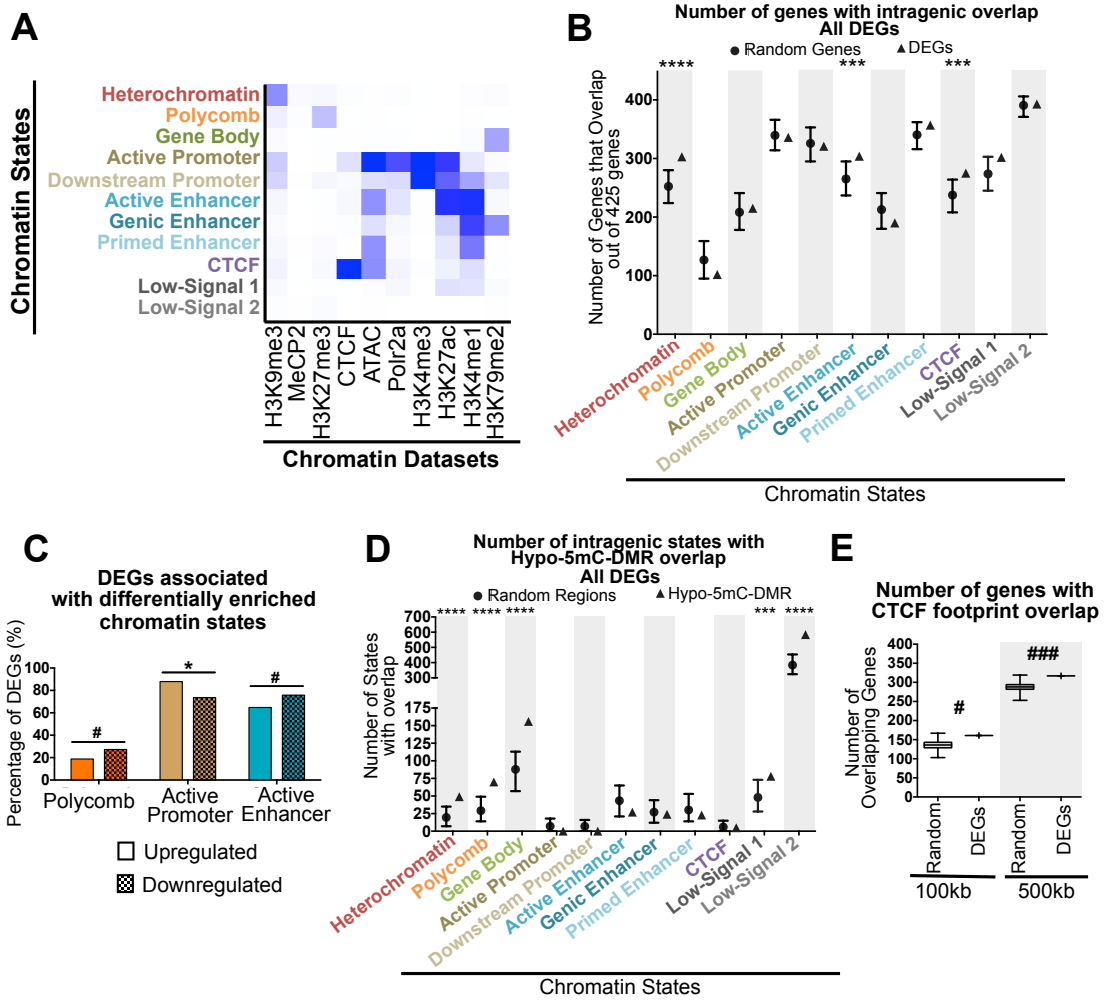


Figure 2. Differentially expressed genes in R106W are associated with particular chromatin states. (A) Hidden Markov Modeling of 10 chromatin datasets let to the demarcation of 11 re-occurring combinatorial patterns that were used to define chromatin states. **(B)** Number of DEGs vs Random Genes (random sampling (1,000 times) of genes that are number and length matched to DEGs) that overlap with each of the 11 states. The mean, maximum, and minimum numbers of overlapping genes are displayed for the Random Genes ($^{***}p<0.00005$ & $^{****}p<0.00001$, two-tailed test). **(C)** The percentage of upregulated DEGs and downregulated DEGs associated with chromatin states that significantly differed in enrichment between the two DEG states ($^{\#}p<0.05$ & $^{*}p<0.0005$, Fisher's exact test). **(E)** Number of DEGs vs Random Genes (random sampling (1,000 times) of genes that are number and length matched to DEGs)

that overlap with the CTCF site either intragenically plus 100kB upstream and downstream or intragenically plus 500kb upstream and downstream ($^{\#}p<0.05$ & $^{\#\#\#}p<0.005$, two-tailed test).

Figure 3

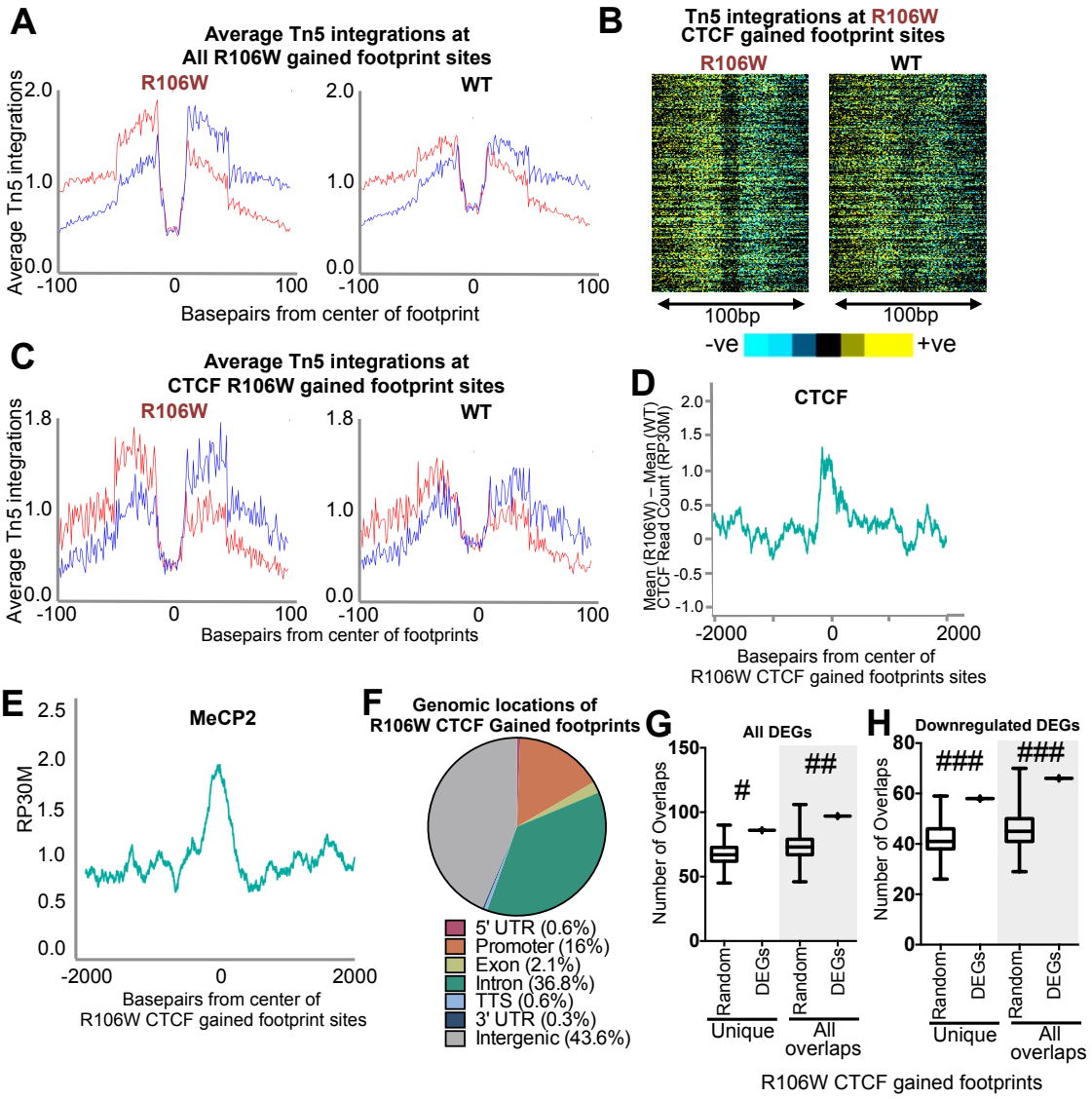


Figure 3. Alterations in CTCF Binding in R106W Mice (A) Average Tn5 insertion profiles at R106W gained footprint sites plus 100bp upstream and downstream in R106W vs WT (red: positive strand cuts; blue: negative strand cuts). **(B)** Heatmap of the Tn5 insertions at R106W CTCF gained footprint sites in R106W and WT ordered from most differential at the top to least differential at the bottom. Yellow indicates a greater number of positive strand cuts compared to negative strand cuts per nucleotide position and blue indicates a greater number of negative strand cuts. **(C)** Average Tn5 insertion profiles at R106W CTCF gained footprint sites plus 100bp upstream and downstream in

R106W vs WT (red: positive strand cuts; blue: negative strand cuts). **(D)** Mean difference in CTCF ChIP-seq read count in R106W and WT at R106W CTCF gained footprint sites plus 2kb upstream and downstream. RP30M: read counts per 30 million. **(E)** MeCP2 mean read count at R106W CTCF gained footprint sites plus 2kb upstream and downstream. RP30M: read counts per 30 million. **(F)** Genomic distribution of R106W CTCF gained footprint sites. **(G)** Number of unique overlaps or total overlaps of DEGs or **(H)** downregulated DEGs vs Random Genes (random sampling (1,000 times) of genes that are number and length matched to DEGs) with R106W CTCF gained footprint sites (gene bodies plus 500kb upstream and downstream). The mean, maximum, and minimum numbers of overlapping genes are displayed for the Random Genes ($^{\#}p<0.05$, $^{\#\#}p<0.01$, & $^{\#\#\#}p<0.005$, two-tailed test).

Figure 4

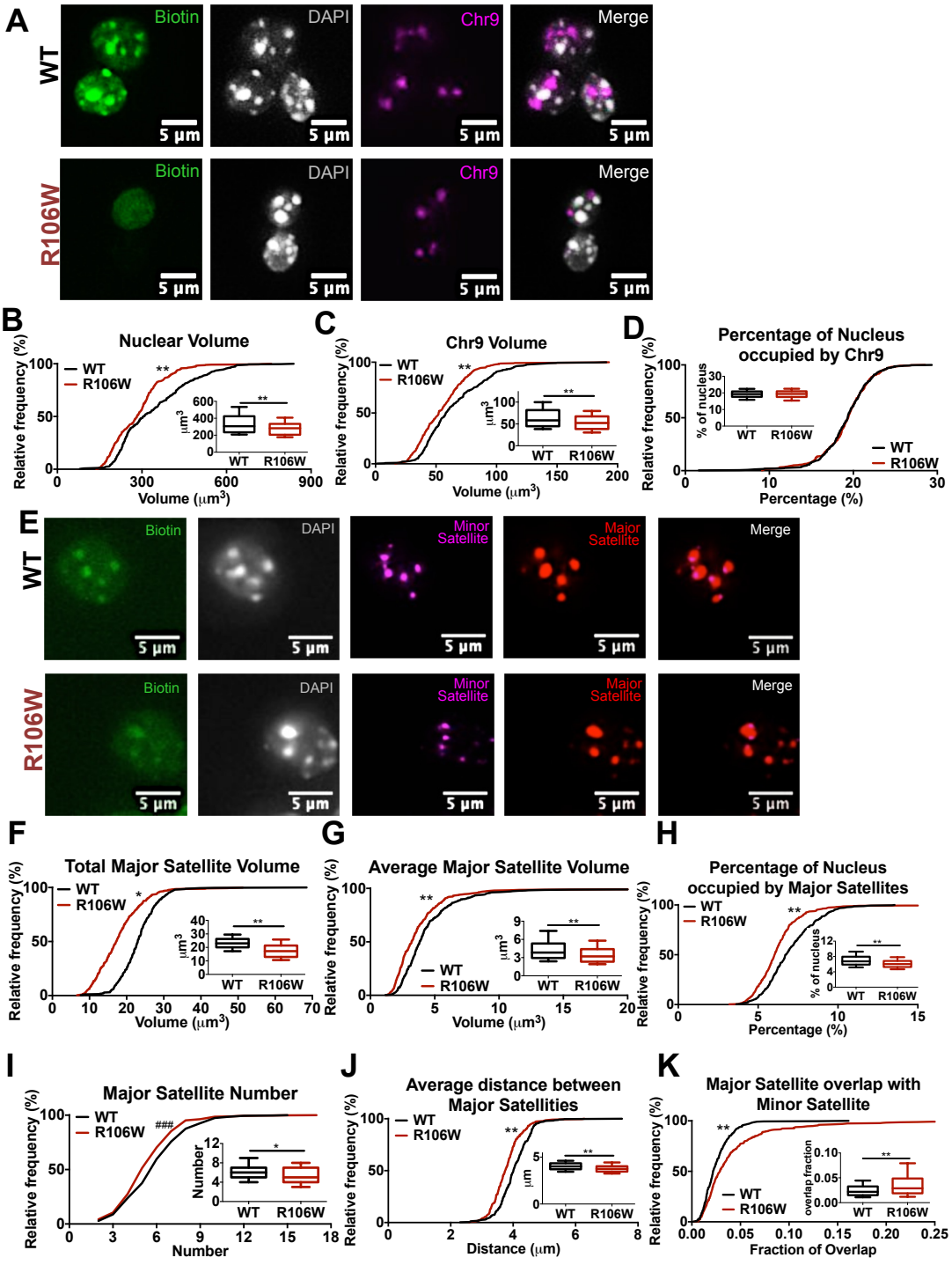


Figure 4. Heterochromatic and euchromatic condensation in R106W mice.

(A) Fluorescence images of DAPI, biotin (labeling biotinylated MeCP2 in excitatory neurons), and Chr9 in WT and R106W nuclei. Cumulative frequency distribution and

boxplot of **(B)** nuclear volume**, (C)** total Chr9 volume (per nucleus sum), **(D)** and the percentage of the nucleus occupied by Chr9 ((Chr9 volume divided by nuclear volume) x 100) (n=3 WT animals and n=3 R106W animals with 126 nuclei per animal). **(E)** Fluorescence images of DAPI, biotin, major satellites, and minor satellites in WT and R106W nuclei. Cumulative frequency distribution and boxplot of **(F)** total major satellite volume (per nucleus sum), **(G)** average major satellite volume (average volume of a single major satellite foci), **(H)** percentage of the nucleus occupied by major satellites ((total major satellite volume divided by nuclear volume) x 100), **(I)** major satellite number per nucleus, **(J)** average distance between major satellites, **(K)** fraction of major satellite volume that overlaps with minor satellites (n=3 WT animals and n=3 R106W animals with 118 nuclei per animal). Boxplots extend from the 25^{th} to 75^{th} percentiles with a line at the median and whiskers range from the 10^{th} to 90^{th} percentiles. $^{\#\#\#}p<0.005$, $^{*}p<0.0005$, $^{**}p<0.0001$. Significance marks in the frequency distribution area indicate significance via the Kolmogorov-Smirnov test and significance marks in the boxplots indicate significance via the Mann-Whitney test.

Figure 5

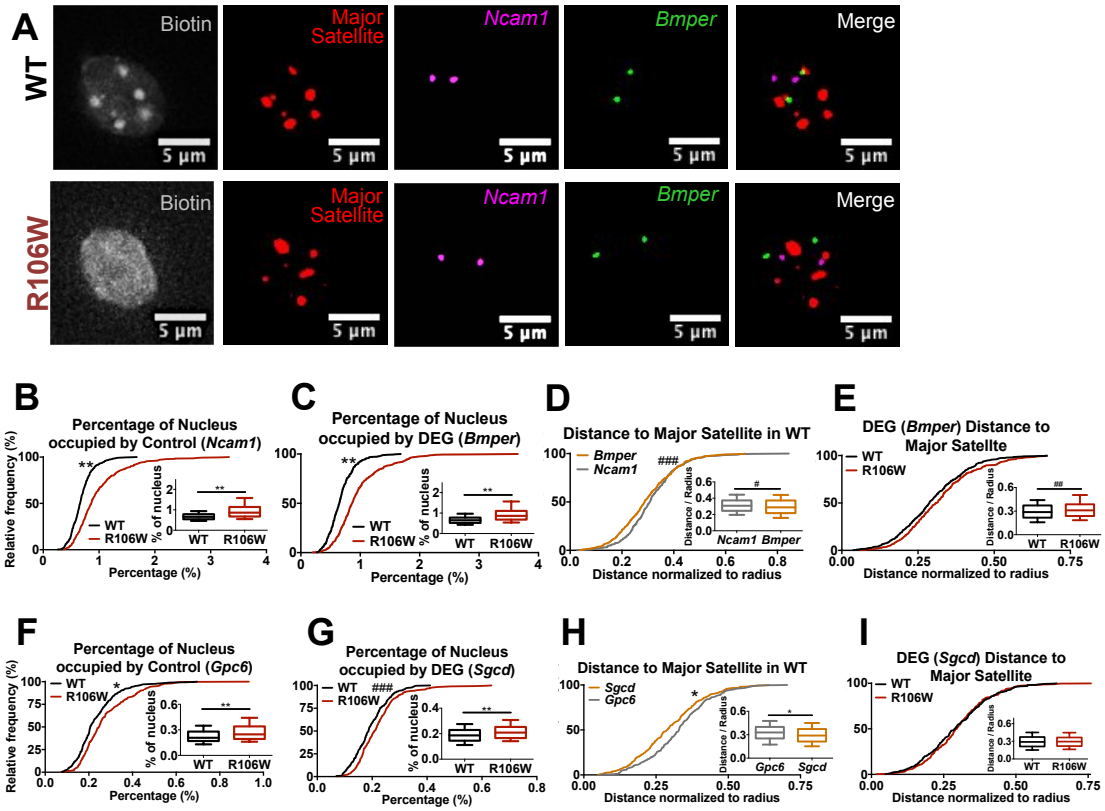


Figure 5. Gene body decondensation in R106W mice. (A) Fluorescence images of DAPI, biotin (labeling biotinylated MeCP2 in excitatory neurons), DEG (*Bmper*), and Control (Ncam1) in WT and R106W nuclei. Cumulative frequency distribution and boxplot of **(B)** percentage of the nucleus occupied by Control (*Ncam1*) ((total Control volume divided by nuclear volume) x 100), **(C)** percentage of the nucleus occupied by DEG (*Bmper*) ((total DEG volume divided by nuclear volume) x 100), **(D)** distance of DEG (*Bmper*) or Control (*Ncam1*) from nearest major satellite in WT, and **(E)** DEG (*Bmper*) distance to nearest major satellite (n=3 WT animals and n=3 R106W animals with 125 nuclei per animal; $^{\#}p<0.05$; $^{\#\#}p<0.01$; $^{\#\#\#}p<0.005$; $^{**}p<0.0001$; significance marks in the frequency distribution area indicate significance via the Kolmogorov-Smirnov test and significance marks in the boxplots indicate significance via the Mann-Whitney test). Cumulative frequency distribution and boxplot of **(F)** percentage of the nucleus occupied by Control (*Gpc6*) ((total Control volume divided by nuclear volume) x 100), **(G)** percentage of the nucleus occupied by DEG (*Sgcd*) ((total DEG volume divided by nuclear volume) x 100), **(H)** distance of DEG (*Sgcd*) or Control (*Gpc6*) from nearest

major satellite in WT, and **(I)** DEG (*Sgcd*) distance to nearest major satellite (n=3 WT animals and n=3 R106W animals with 100 nuclei per animal; $^{\#\#}p<0.01$; $^{*}p<0.0005$; $^{**}p<0.0001$; significance marks in the frequency distribution area indicate significance via the Kolmogorov-Smirnov test and significance marks in the boxplots indicate significance via the Mann-Whitney test). Boxplots extend from the 25th to 75th percentiles with a line at the median and whiskers range from the 10th to 90th percentiles.

Figure 6

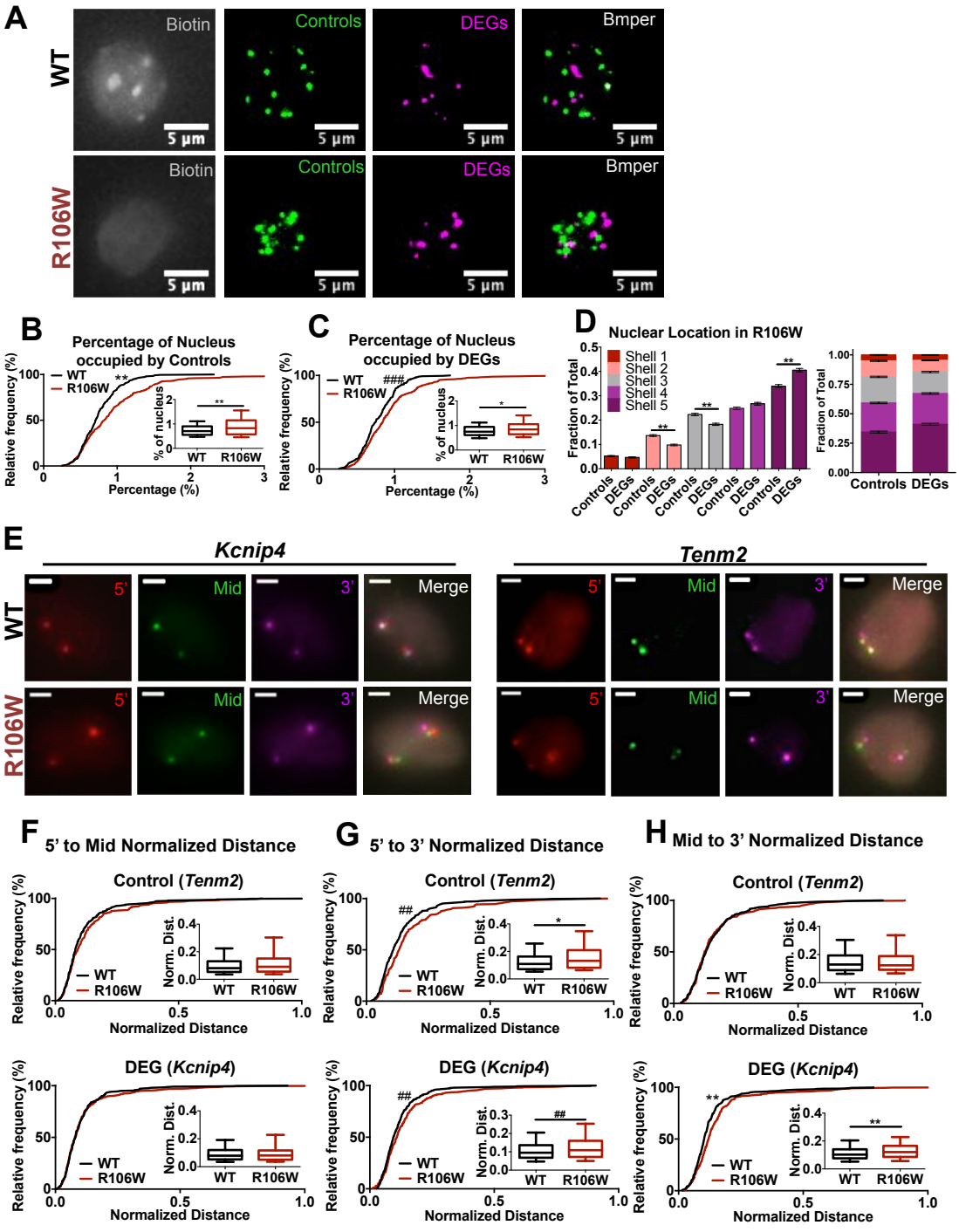
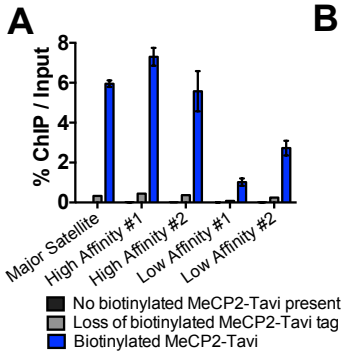


Figure 6. Further confirmation of gene body decondensation in R106W. (A) Fluorescence images of DAPI, biotin (labeling biotinylated MeCP2 in excitatory neurons), DEGs, and Controls in WT and R106W nuclei. Cumulative frequency

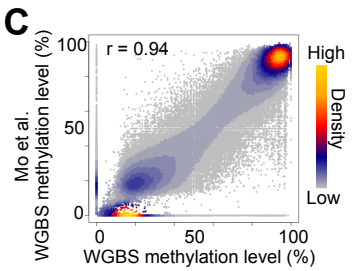
distribution and boxplot of **(B)** percentage of the nucleus occupied by Controls ((total volume of Controls divided by nuclear volume) x 100), **(C)** percentage of the nucleus occupied by DEGs ((total volume of DEGs divided by nuclear volume) x 100) (n=2 WT animals and n=3 R106W animals with 125 nuclei per animal; $^{###}p<0.005$; $^{*}p<0.0005$; $^{**}p<0.0001$; significance marks in the frequency distribution area indicate significance via the Kolmogorov-Smirnov test and significance marks in the boxplots indicate significance via the Mann-Whitney test). **(D)** Comparison of the nuclear location of Controls vs DEGs in WT across all 5 Shells ($^{**}p<0.0001$; Mann-Whitney test). **(E)** Fluorescence images of DAPI and biotin (labeling biotinylated MeCP2 in excitatory neurons), along with the 5' end, Middle, and 3' end of a DEG (*Kcnip4*) or Control (*Tenm2*) WT and R106W nuclei. Cumulative frequency distribution and boxplot of **(F)** the normalized distance between the 5' to Mid gene regions (distance divided by nuclear radius) of a DEG (*Kcnip4*) or Control (*Tenm2*), **(G)** the normalized distance between the 5' to 3' gene regions (distance divided by nuclear radius) of a DEG (*Kcnip4*) or Control (*Tenm2*), and **(H)** the normalized distance between the Mid' to 3' gene regions (distance divided by nuclear radius) of a DEG (*Kcnip4*) or Control (*Tenm2*) (n=3 WT animals and n=3 R106W animals with 100 nuclei per animal; $^{##}p<0.01$; $^{*}p<0.0005$; $^{**}p<0.0001$; significance marks in the frequency distribution area indicate significance via the Kolmogorov-Smirnov test and significance marks in the boxplots indicate significance via the Mann-Whitney test). Boxplots extend from the 25^{th} to 75^{th} percentiles with a line at the median and whiskers range from 10^{th} to 90^{th} percentiles.

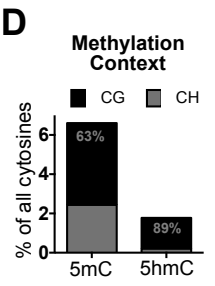
Supplemental Figure 1



B

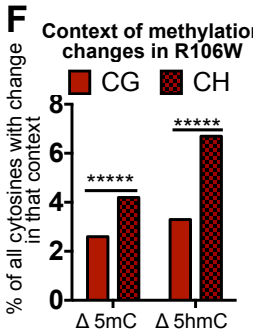
	Genotype	Platform	Uniquely mapped reads	Genomic cover-age	No. of profiled cytosines	Cytosine coverage (mean)
WT	WGBS	Paired end	354,206,959	14.41	910,767,892	10.26
		Single end	69,910,972			
	TAB-seq	Paired end	355,482,726	14.84	892,356,502	9.34
		Single end	90,399,255			
R106W	WGBS	Paired end	342,226,844	14.09	919,794,624	9.71
		Single end	76,664,971			
	TAB-seq	Paired end	362,161,658	14.95	897,101,884	9.61
		Single end	82,948,394			





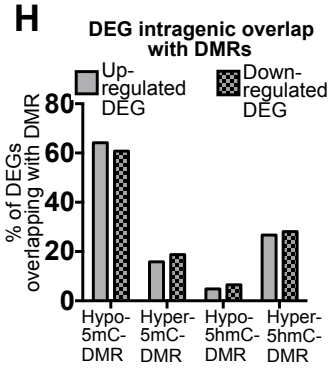
E

	Genotype	CG context	Number of methylated sites	Number of cytosine sites that change in mutant (↑, Hypermethylation) (↓, Hypomethylation) (% , percentage of cytosines with change)
5mC	WT	CG	25,097,058	N/A
		CH	14,679,722	N/A
	R106W	CG	24,937,160	↑ 121,394 (0.5%)
				↓ 538,459 (2.2%)
		CH	13,945,529	↑ 101,469 (0.7%)
				↓ 479,618 (3.4%)
5hmC	WT	CG	8,974,474	N/A
		CH	1,134,378	N/A
	R106W	CG	10,542,560	↑ 270,875 (2.6%)
				↓ 80,668 (0.8%)
		CH	1,615,932	↑ 74,582 (4.6%)
				↓ 34,172 (2.1%)



G

	Methylation State	Number of Basepairs
WT	FMR	2,114,052,792
	LMR	23,498,024
	UMR	8,366,590
	PMD	580,648,498
R106W	FMR	2,095,492,757
	LMR	25,894,156
	UMR	8,376,457
	PMD	597,113,098



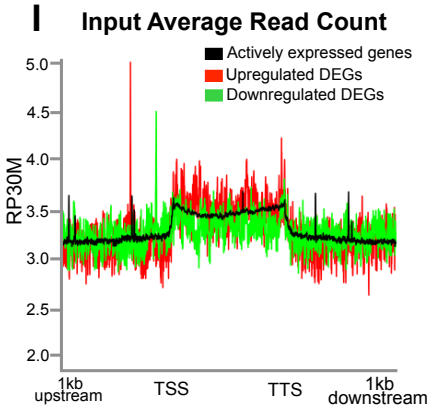


Figure S1. (A) RT-PCR of DNA pulled down by streptavidin in cortical tissue of mice with no biotinylated (*R26*$^{cBirA/+}$;*NEX*$^{Cre/+}$), with loss of the biotinylated MeCP2-Tavi tag (*Mecp2*$^{WT\text{-}Tavi/y}$;*R26*$^{cBirA/+}$;*NEX*$^{Cre/+}$; lysates are treated with Tev protease prior to streptavidin mediated pulldown, which leads to the cleavage of the biotinylated Tavi tag

on MeCP2, or Biotinylated MeCP2-Tavi (*Mecp2*$^{WT\text{-}Tavi/y}$;*R26*$^{cBirA/+}$;*NEX*$^{Cre/+}$) using primers for major satellite DNA and previously established high-affinity sites and lowly enriched at known low-affinity sites (Chen et al., 2015). **(B)** WGBS and TAB-seq sequencing information. **(C)** Density map comparing the methylation status of individual cytosines between previously published data and our own data (MeCP2^{WT}) from glutamatergic cortical neurons. r, Pearson correlation. **(D)** Percentage of all cytosines that are methylated in each of the 4 major contexts of DNA methylation in WT. Percentages within the bars represent the percentage of 5mC or 5hmC in the CG context. **(E)** Summary of the number of methylated cytosines and differentially methylated cytosines in various contexts. **(F)** Context of methylation changes in R106W (*****p < 2.2e-16, Fisher's exact test). **(G)** Summary of the number of basepairs in FMRs, LMRs, UMRs, and PMDs. **(H)** The percentage of upregulated DEGs and downregulated DEGs that overlap with each of the 4 types of DMRs. **(I)** Input read count in the gene bodies, plus 100kb upstream and downstream, of upregulated DEGs, downregulated DEGs, and all actively expressed genes in cortical excitatory neurons. The gene body, upstream, and downstream regions were divided into 1000 equal bins for each gene and the number of sequencing reads for each bin was calculated and normalized. RP30M, read counts per 30 million.

Supplemental Figure 2

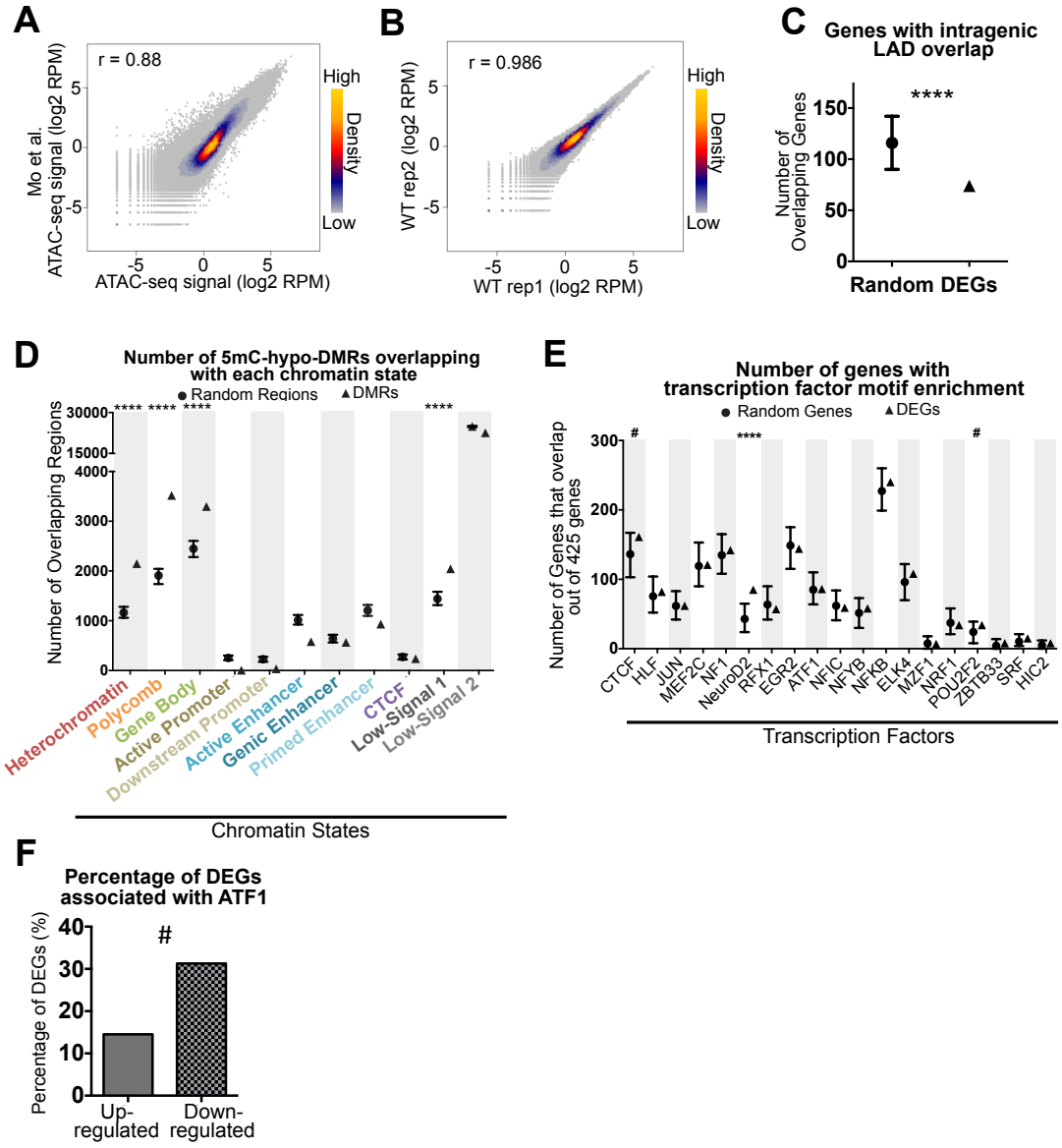


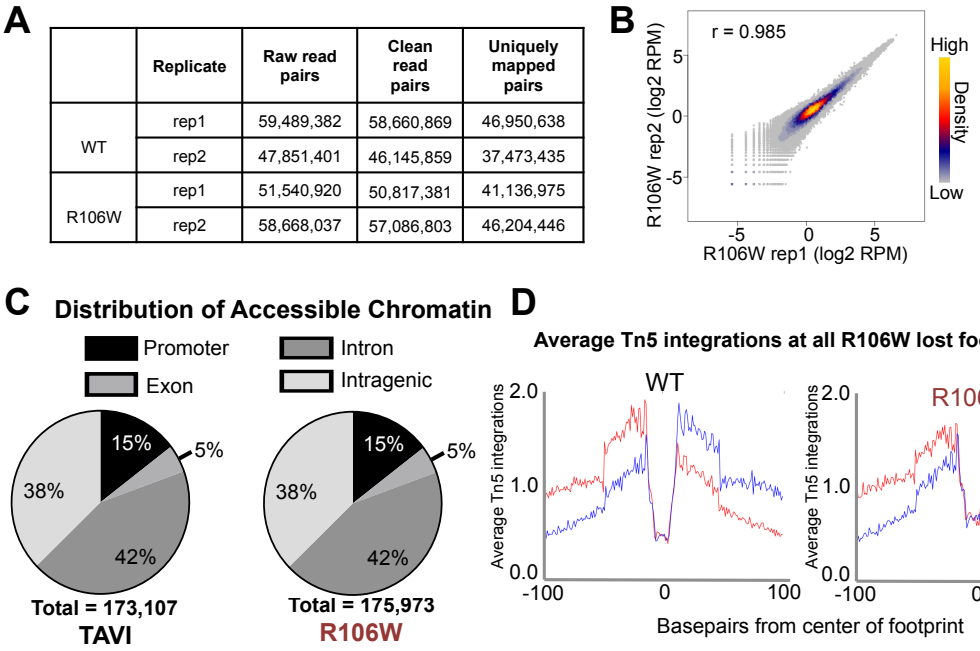
Figure S2. (A) Density map comparing accessible genomic regions (via ATAC-seq) between previously published data and our own data in glutamatergic neurons. r, Pearson correlation; RPM, reads per million. **(B)** Density map comparing accessible genomic regions between the two ATAC-seq biological replicates in excitatory neurons. r, Pearson correlation; RPM, reads per million. **(C)** Number of DEGs vs Random Genes (random sampling (1,000 times) of genes that are number and length matched to DEGs) that overlap with the lamin associated domains (LADs) (****p<0.00001, two-tailed test).

(D) Number of 5mC-hypo-DMRs vs Random Regions (random sampling (1,000 times) of regions throughout the genome that are number and length matched to 5mC-hypo-DMRs) ($^{****}p<0.00001$, two-tailed test). **(E)** Number of DEGs vs Random Genes (random sampling (1,000 times) of genes that are number and length matched to DEGs) that overlap with the each of the 19 TF footprint sites intragenically plus 100kB upstream and downstream ($^{****}p<0.00001$ & $^{\#}p<0.05$, two-tailed test). **(F)** The percentage of upregulated DEGs and downregulated DEGs associated with the TF footprint sites that significantly differed in enrichment between the two DEG states ($^{\#}p<0.05$).

Supplemental Figure 3

A

	Replicate	Raw read pairs	Clean read pairs	Uniquely mapped pairs
WT	rep1	59,489,382	58,660,869	46,950,638
	rep2	47,851,401	46,145,859	37,473,435
R106W	rep1	51,540,920	50,817,381	41,136,975
	rep2	58,668,037	57,086,803	46,204,446



E

R106W gained footprint sites		
TF	p-value	% of Targets
CTCF	1e-28	9.65%
EWSR1	1e-22	5.61%
MEF2C	1e-21	13.89%
ONECUT2	1e-19	6.53%
TCF3	1e-17	3.95%
THAP1	1e-17	19.02%
TEAD1	1e-16	5.30%
SOX5	1e-15	6.10%
ZFP105	1e-15	3.24%
NFIA	1e-15	5.81%
TBP	1e-14	3.04%
SOX4	1e-13	2.89%
GABPA	1e-12	2.03%
RFX3	1e-12	3.95%
MEIS3	1e-12	2.55%

F

R106W lost footprint sites		
TF	p-value	% of Targets
TBP	1e-19	6.30%
ZBTB18	1e-17	4.35%
RFX5	1e-16	9.71%
ESRRA	1e-15	5.77%
TCF3	1e-15	4.63%
NF1	1e-15	14.91%
SRF	1e-14	4.83%
CTCF	1e-13	3.78%
LIN54	1e-12	3.37%
NFKB	1e-12	6.74%

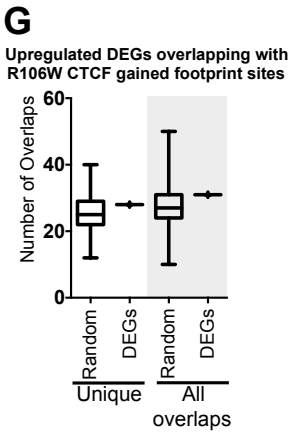


Figure S3. (A) ATAC-seq sequencing information. **(B)** Density map comparing accessible genomic regions between the two ATAC-seq biological replicates in excitatory neurons. r, Pearson correlation; RPM, reads per million. **(C)** Genomic distribution of ATAC-seq peaks in WT and R106W. **(D)** Average Tn5 insertion profiles at R106W lost footprint sites plus 100bp upstream and downstream in WT vs R106W (red: positive

strand cuts; blue: negative strand cuts). **(E-F)** Summary tables of the TF motifs enriched at R106W lost gained and lost footprint sites. **(G)** Number of unique overlaps or total overlaps of upregulated DEGs vs Random Genes (random sampling (1,000 times) of genes that are number and length matched to DEGs) with R106W CTCF gained footprint sites (gene bodies plus 500kb upstream and downstream). The mean, maximum, and minimum numbers of overlapping genes are displayed for the Random Genes.

Supplemental Figure 4

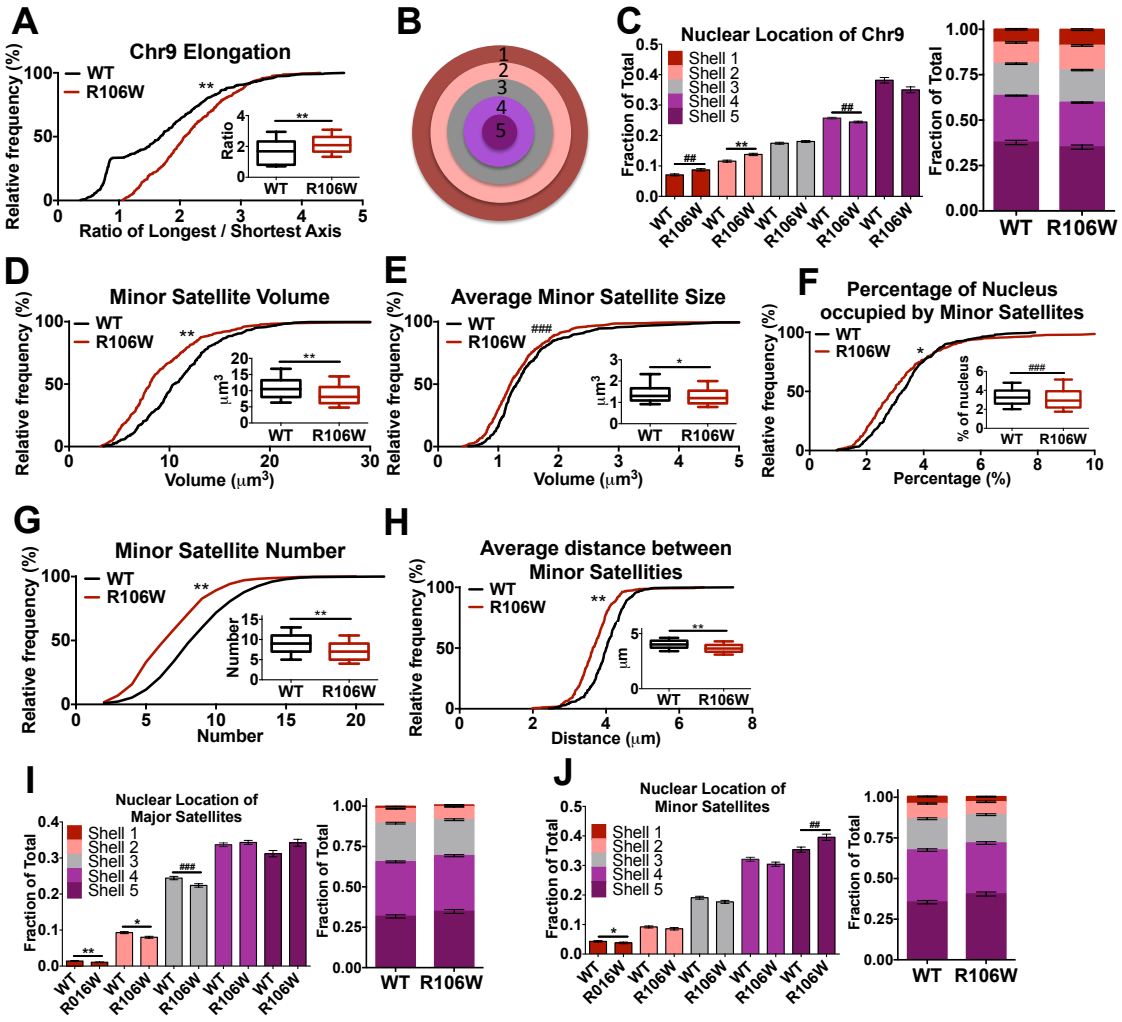


Figure S4. (A) Cumulative frequency distribution and boxplot of Chr9 elongation (n=3 WT animals and n=3 R106W animals with 126 nuclei per animal; $^{**}p<0.0001$; significance marks in the frequency distribution area indicate significance via the Kolmogorov-Smirnov test and significance marks in the boxplots indicate significance via the Mann-Whitney test). **(B)** Schematic demonstrating shell designations for nuclear location analysis. **(C)** Comparison of Chr9 nuclear location in WT vs R106W across all 5 shells (n=3 WT animals and n=3 R106W animals with 126 nuclei per animal; $^{\#\#}p<0.01$; $^{**}p<0.0001$, Mann-Whitney test). Cumulative frequency distribution and boxplot of **(D)** total minor satellite volume (per nucleus sum), **(E)** average minor satellite volume (average volume of a single major satellite foci), **(F)** percentage of the

nucleus occupied by minor satellites ((total major satellite volume divided by nuclear volume) x 100), **(G)** minor satellite number per nucleus, and **(H)** average distance between minor satellites (n=3 WT animals and n=3 R106W animals with 118 nuclei per animal; ###p<0.005; *p<0.0005; **p<0.0001; significance marks in the frequency distribution area indicate significance via the Kolmogorov-Smirnov test and significance marks in the boxplots indicate significance via the Mann-Whitney test). Comparison of **(I)** major satellite or **(J)** minor satellite location in WT and R106W across all 5 shells (n=3 WT animals and n=3 R106W animals with 118 nuclei per animal; ##p<0.01; ###p<0.005; *p<0.0005; **p<0.0001, Mann-Whitney test). Boxplots extend from the 25th to 75th percentiles with a line at the median and whiskers range from the 10th to 90th percentiles.

Supplemental Figure 5

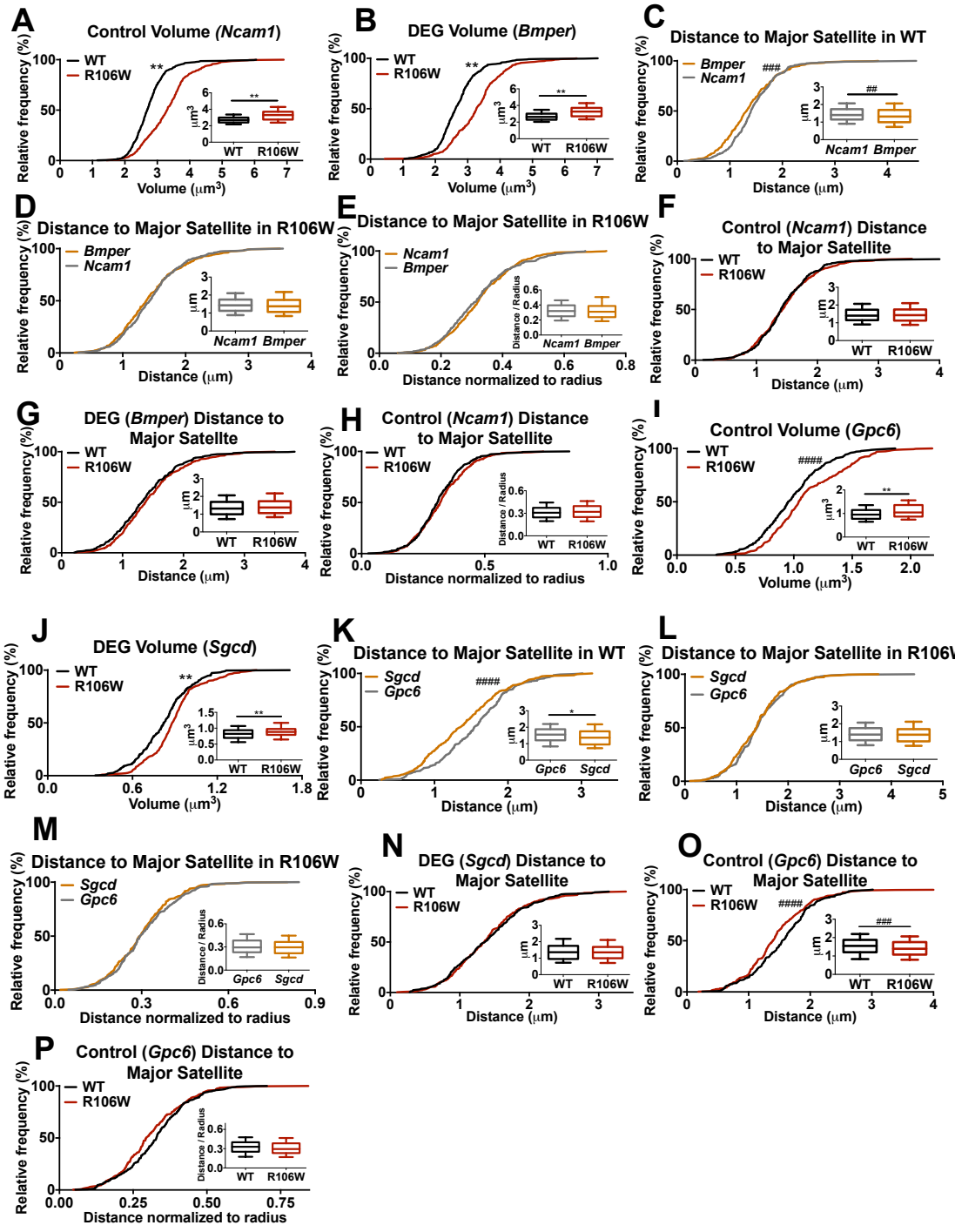


Figure S5. Cumulative frequency distribution and boxplot of **(A)** Control (*Ncam1*) volume, **(B)** DEG (*Bmper*) volume, **(C)** distance of DEG (*Bmper*) or Control (*Ncam1*) to nearest major satellite in WT, **(D)** raw distance of DEG (*Bmper*) or Control (*Ncam1*) to

nearest major satellite in R106W, **(E)** normalized distance of DEG (*Bmper*) or Control (*Ncam1*) to nearest major satellite in R106W, **(F)** Control (*Ncam1*) raw distance to nearest major satellite, **(G)** DEG (*Bmper*) raw distance to nearest major satellite, **(H)** Control (*Ncam1*) normalized distance to nearest major satellite (n=3 WT animals and n=3 R106W animals with 125 nuclei per animal; $^{\#\#}p<0.01$; $^{\#\#\#}p<0.005$; $^{**}p<0.0001$). Cumulative frequency distribution and boxplot of **(I)** Control (*Gpc6*) volume, **(J)** DEG (*Sgcd*) volume, **(K)** distance of DEG (*Sgcd*) or Control (*Gpc6*) to nearest major satellite in WT, **(L)** raw distance of DEG (*Sgcd*) or Control (*Gpc6*) to nearest major satellite in R106W, **(M)** normalized distance of DEG (*Sgcd*) or Control (*Ncam1*) to nearest major satellite in R106W, **(N)** DEG (*Sgcd*) raw distance to nearest major satellite, **(O)** Control (*Gpc6*) raw distance to nearest major satellite, **(H)** Control (*Gpc6*) normalized distance to nearest major satellite (n=3 WT animals and n=3 R106W animals with 100 nuclei per animal; $^{\#\#\#}p<0.005$; $^{\#\#\#\#}p<0.001$; $^{*}p<0.0005$ $^{**}p<0.0001$; significance marks in the frequency distribution area indicate significance via the Kolmogorov-Smirnov test and significance marks in the boxplots indicate significance via the Mann-Whitney test). Boxplots extend from the 25th to 75th percentiles with a line at the median and whiskers range from 10th to 90th percentiles.

Supplemental Figure 6

A

DEG				Matched Control Gene		
Gene	logFC	Length	RPKM	Gene	Length	RPKM
Il1rapl2	0.76	1276339	6.32	*Lrrc4c*	1313498	8.23
Sgcd	0.79	1092717	4.92	*Gpc6*	1054233	6.83
Nckap5	1.27	917180	0.93	*Dpyd*	870790	2.83
Cntnap5c	0.61	640773	5.76	*Pcdh11x*	620361	8.35
Ghr	0.80	265733	2.87	*Srgap1*	266325	1.40
Bmper	1.38	262127	5.65	*Socs6*	262258	5.27
Nxn	0.66	141943	0.31	*Scml2*	141021	0.79
Ankrd42	0.83	55420	13.06	*Tcp11l1*	54882	13.11
Steap2	1.49	29748	0.76	*Nbeal2*	29373	1.35
Scn3b	0.78	22403	20.84	*Clns1a*	22246	18.80
Sertad1	1.74	3407	2.72	*Lpcat2b*	3491	1.69

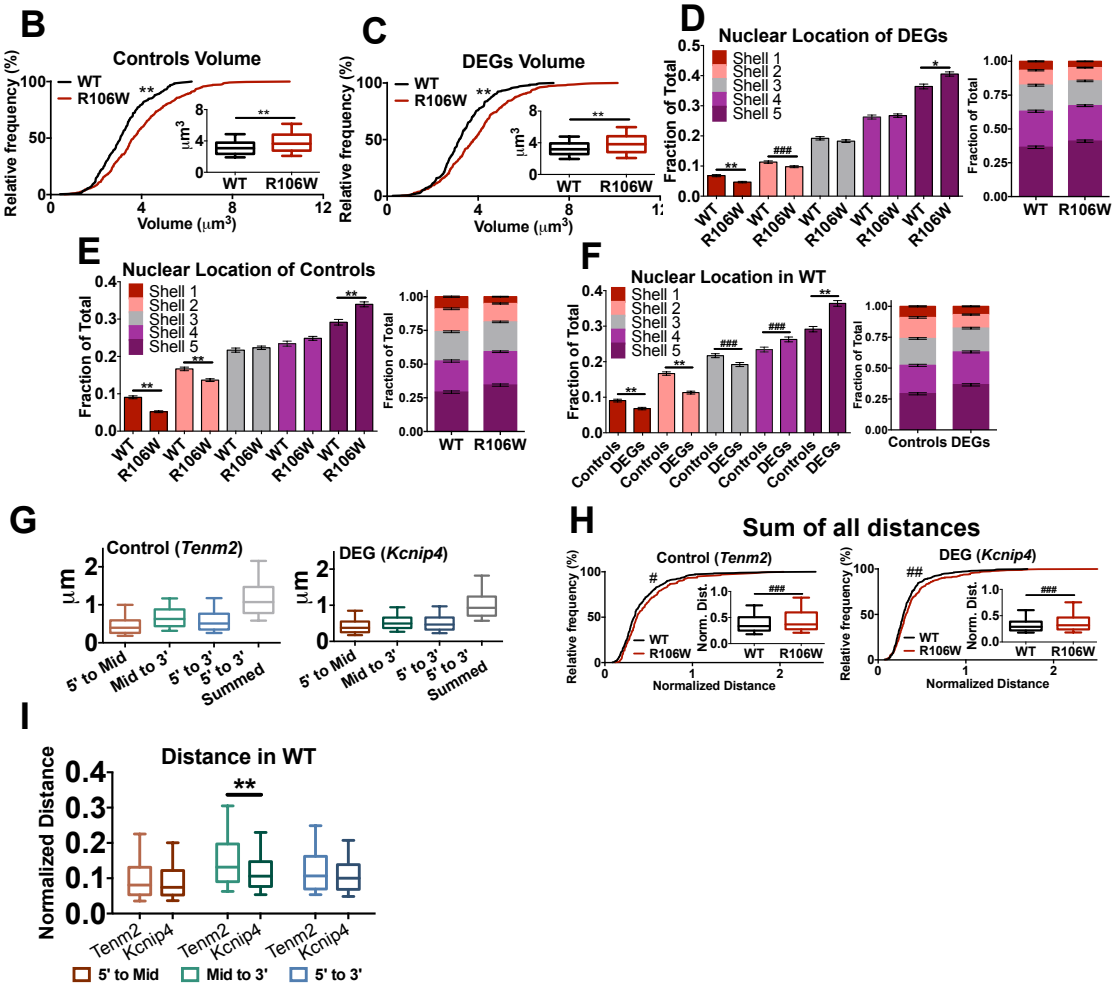


Figure S6. (A) Table summarizing the features of DEGs and Controls used. Cumulative frequency distribution and boxplot of **(B)** Controls volume and **(C)** DEGs volume (n=2

WT animals and n=3 R106W animals with 125 nuclei per animal; $^{**}p<0.0001$; significance marks in the frequency distribution area indicate significance via the Kolmogorov-Smirnov test and significance marks in the boxplots indicate significance via the Mann-Whitney test). Comparison of the nuclear location of **(D)** DEGs in WT vs R106W across all 5 shells, **(E)** Controls in WT vs R106W across all 5 shells, **(F)** Controls vs DEGs in WT across all 5 shells ($^{\#\#\#}p<0.005$; $^{*}p<0.0005$; $^{**}p<0.0001$; Mann-Whitney test). **(G)** Distances between different locations of the genes. 5' to 3' summed is the sum of the 5' to Mid distance and the Mid to 3' distance (n= 3 WT animals with 100 nuclei per animal for both genes). **(H)** Sum of the normalized distances between all pairwise combinations of locations (5' to Mid, Mid to 3', and 5' to 3') (n=3 WT animals and n=3 R106W animals with 100 nuclei per animal; $^{\#}p<0.05$; $^{\#\#}p<0.01$; $^{\#\#\#}p<0.0005$; significance marks in the frequency distribution area indicate significance via the Kolmogorov-Smirnov test and significance marks in the boxplots indicate significance via the Mann-Whitney test). **(I)** Comparison of distances the WT setting. Boxplots extend from the 25th to 75th percentiles with a line at the median and whiskers range from 10th to 90th percentiles.

CHAPTER 4

Summary / Discussion

Since the discovery of MeCP2, twenty-five years of research on this protein has led to vast insight into the complex, multifaceted function of this protein, generating numerous hypothesized molecular models, such as, a transcriptional repressor, a transcriptional activator, a chromatin organizer, a regulator of alternative splicing, and a miRNA processor (Lyst and Bird, 2015). Given that MeCP2 is a methyl-CpG-binding protein and that DNA methylation patterns are cell type-specific (Mo et al., 2015), we hypothesized that MeCP2 influences gene expression in a cell type-specific manner, and that cellular heterogeneity must first be resolved in order to clearly delineate transcriptional and epigenomic alterations for the understanding of how MeCP2 affects transcription.

With the employment of our cell type-specific approach, we found that differentially expressed genes (DEGs) in *Mecp2* mutant mice are neuronal type-dependent, given that the vast majority of misregulated genes in excitatory or inhibitory neurons are exclusive to each cell type. In agreement with various studies (Ben-Shachar et al., 2009; Chahrour et al., 2008; Tudor et al., 2002), gene expression differences in R106W mice were found to be very subtle, with 70% of DEGs exhibiting less than a 50% alteration in expression levels. Additionally, the numbers of DEGs that are upregulated and downregulated are very similar, supporting neither an activator nor repressor model for MeCP2. Furthermore, we find that DEGs are long and lowly expressed. Focusing on one cell type, excitatory cortical neurons, we find that DEGs appear to be regulated by MeCP2, rather than the DEGs being a byproduct of indirect effects, since cell type-specific ChIP-seq of MeCP2 demonstrates that DEGS are intragenically enriched with MeCP2 when compared to genes that are not misregulated in RTT. Our finding that both upregulated and downregulated DEGs are enriched with 5mC-hypomethylated differentially methylated regions (DMRs) suggests that the loss of MeCP2 binding at DEGs in *Mecp2* mutant mice leads to the demethylation of unbound cytosines within DEG gene bodies, and thus, contributes to the dysregulation of these genes. Given these findings, we next sought to investigate whether DEGs are enriched with 5mC (in both the CG or CH context) in order to gain an understanding of the epigenomic features that lead

to the recruit of MeCP2. We found that gene bodies of upregulated, but not downregulated, DEGs contain higher levels of 5mCG and 5mCH than genes that are not misregulated in RTT. Notably, our findings on upregulated DEGs are in agreement with recently published work that have found that MeCP2 enrichment at upregulated genes correlates with high levels of 5mCH (Chen et al., 2015; Gabel et al., 2015). This implies that the binding of MeCP2 at upregulated DEGs in the WT setting leads to gene repression (given the upregulation in the mutant context), suggesting that MeCP2 is a repressor. However, the regulatory role of MeCP2 at downregulated DEGs is currently unclear, with our study finding that downregulated DEGs are associated with MeCP2 enrichment, but not with high levels of 5mCH; one previously published study finding that downregulated DEGs are associated with both MeCP2 binding and high levels of 5mCH (Chen et al., 2015); and another study finding that downregulated DEGs are neither associated with MeCP2 binding nor high levels of 5mCH (Gabel et al., 2015). One plausible explanation for this discrepancy is that different brain regions/cell types were assessed in each of these studies, with our study evaluating excitatory neurons of the cortex, one study using whole cortical/cerebellar homogenate (Gabel et al., 2015), and the other study using the hypothalamus (Chen et al., 2015). These findings also raise the possibility that MeCP2 function is chromatin-feature dependent: the binding of MeCP2 at regions with high levels of methylation could lead to gene repression, whereas the binding of MeCP2 at regions with low levels of methylation (and possibly in conjunction with other chromatin features, like histone modifications or chromatin accessibility) could lead to gene activation.

Since average intragenic levels of 5mCG and 5mCH cannot fully explain MeCP2 recruitment to DEGs, we next sought to determine common chromatin features of DEGs in the WT context to gain insight into why these genes are enriched for MeCP2 binding in WT and preferentially susceptible to dysregulation in RTT. The observation that DEGs are lowly expressed is supported by our finding that DEGs are intragenically enriched for the heterochromatic state (H3K9me3) based on ChIP-seq data and that DEGs are in closer proximity to major satellites (a major component of heterochromatin) than control genes via Oligopaint FISH. These findings are in line with the fact that MeCP2 is highly concentrated at pericentromeric heterochromatin (Lewis et al., 1992), suggesting that MeCP2 could play a role in the regulation of lowly expressed genes.

Two other chromatin features were found to be significantly associated with the gene bodies of DEGs in the WT context: CTCF and active enhancer chromatin states. As a general feature, previous studies have found significant coenrichment of CTCF with enhancers (Chepelev et al., 2012; Sanyal et al., 2012). CTCF enrichment at DEGs was also corroborated via transcription factor footprinting analysis in open chromatin regions. Given that CTCF is an architectural protein important for bridging together regulatory sequences, it is possible that DEGs, especially given their length, require a particular interconnected topology that aids in enhancer recruitment for the maintenance of gene expression. And furthermore, it is possible that MeCP2 plays a role in facilitating this specialized topology for the maintenance of appropriate gene expression levels. Additionally, we found that certain chromatin features are differentially enriched in upregulated versus downregulated DEGs, which aids in understanding the directionality of gene dysregulation in RTT. A significantly higher percentage of upregulated DEGs than downregulated DEGs are associated with the active promoter state, suggesting that upregulated DEGs are in an environment that facilitates gene expression. Contrastingly, a significantly higher percentage of downregulated DEGs than upregulated DEGs are associated with the polycomb state, suggesting that downregulated DEGs are in an environment associated with gene repression. Taken together, these findings suggest that chromatin environment plays a role in determining whether a gene will be upregulated or downregulated upon loss of MeCP2 binding. Additionally, a significantly higher percentage of downregulated DEGs are associated with active enhancers, which is in opposition to a model predicting that chromatin state dictates gene expression changes observed upon loss of MeCP2 function, however, an explanation for this apparent discrepancy will be offered below.

Given the finding that CTCF and active enhancer chromatin states are enriched at DEGs, we next surveyed open chromatin regions via ATAC-seq and found that the gross location and number of regulatory regions remain similar in R106W mice. However, a more stringent analysis that evaluated the transcription factor (TF) footprint patterns and location within open regulatory regions led to the discovery of thousands of footprints (both lost and gained) in R106W mice. This suggests that the recruitment and/or binding affinity of TFs is subtly affected within open regulatory regions, which is in alignment with the subtlety of gene expression changes found in RTT (here, and (Ben-

Shachar et al., 2009; Chahrour et al., 2008; Tudor et al., 2002)). CTCF was found to be the most significantly enriched TF motif within gained footprint sites (with gained meaning more enriched in R106W than WT), and we subsequently confirmed that these sites are enriched with CTCF binding in R106W compared to WT via CTCF ChIP-seq. We found that these sites are enriched for MeCP2 binding in WT mice and that these sites are significantly associated with a loss of methylation in R106W. Taken together, these data suggest that the binding of MeCP2 in the WT context can prevent the recruitment of CTCF, either by acting as a physical hindrance to CTCF binding and/or by preserving the methylation status at these loci that would prevent methylation-sensitive CTCF from binding (Maurano et al., 2015). CTCF gained sites are widespread across chromosomes, occurring both intergenically and intragenically. We additionally found that CTCF sites that are gained in R106W are significantly enriched in/around downregulated DEGs, but not upregulated DEGs. This finding suggests that downregulated DEGs are potentially more susceptible to changes in CTCF binding, and consequently, changes in chromatin organization. And in light of our previous unexplained finding that in comparison to upregulated DEGs, downregulated DEGs are enriched with active enhancers, these findings implicate the gain of CTCF binding sites in/around downregulated DEGs as potentially leading to alterations in the topology required for the preservation of active enhancer interactions that are necessary for the maintenance of gene expression, and thus, as a possible factor leading to transcriptional downregulation.

Given the widespread gain of CTCF binding in R106W, we next investigated whether these changes in CTCF correlate with changes in chromatin architecture by employing Oligopaint FISH. On the global scale, we observed that both euchromatin and heterochromatin (major satellites) are reduced in volume in RTT, which is suggestive of chromatin condensation and that heterochromatin (major satellites) are slightly more affected than euchromatin, which is in align with the enrichment of MeCP2 at pericentromeric heterochromatin in WT. In addition to a reduction in major satellite volume, major satellites are located slightly more centrally and reduced in number, which may be partially explained by the observed reduction in the average distance between major satellites, suggesting that major satellites are clustering together in RTT. This finding might also aid in understanding the molecular and cellular phenotypes observed in RTT, given that pericentromeric heterochromatin undergo significant

changes over neuronal development (Martou and De Boni, 2000; Solovei et al., 2004). In neurons, heterochromatin number decreases from postnatal day 0 to 3 and then increases to adult levels by postnatal day 15, and heterochromatin nuclear location is predominately established by postnatal day 5-6 (Martou and De Boni, 2000; Solovei et al., 2004). The timing of these heterochromatin alterations coincide with the period of synaptogenesis in mice (2-4 weeks of age) (De Felipe, 1997), suggesting that this reorganization is important for neuronal function and that the organizational alterations observed in R106W mice could be a contributing factor in the neuroplasticity defects observed in RTT (Chao et al., 2007; Dani and Nelson, 2009; Dani et al., 2005; Goffin et al., 2012). In support of this, we find that DEGs are enriched with genes associated with synaptic morphology and function (post-synaptic membrane proteins, including Na^{2+}, K^{+}, Ca^{2+} and Cl^{-} channels, synaptic scaffolding proteins, and inotropic glutamate receptors).

Given these large-scale changes in nuclear architecture, we next investigated whether DEGs are also susceptible to changes in chromatin organization. Therefore, we employed Oligopaint FISH to label the entire intragenic regions of DEGs vs Control genes (genes without expression changes in R106W mice). Two differences were observed between DEGs and Controls: (1) DEGs are in closer proximity to major satellites in the WT setting and (2) DEGs are located more centrally than controls in both the WT and R106W setting, with the second finding possibly being linked to the first finding since major satellites are predominately centrally located. Our finding that DEGs are intragenically associated with the heterochromatin state based on ChIP data corroborates this Oligopaint observation. However, the difference between DEG and Control proximity to major satellites is lost in the *Mecp2* mutant setting, suggesting that heterochromatin proximity alone cannot account for gene misregulation in RTT. Therefore, perhaps it's a combination of all/some of the unique chromatin and gene features of DEGs that render them differentially susceptible. Or perhaps proximity to major satellites is only an effective repression mechanism for DEGs, but not Controls, due to the genomic context in which DEGs are found (Jost et al., 2015). Interestingly, it was found that upregulated DEGs slightly, but significantly, shift away from major satellites in the R106W, while downregulated DEGs remain in similar proximity. This shift away from the repressive, heterochromatic state may partly explain the

upregulation of some DEGs. Given that downregulated DEGs are significantly associated with CTCF gained sites, it is possible that this gain of binding creates a particular, perhaps inflexible, conformational state that prevents DEGs from shifting away from major satellites.

Differing from what was found for major satellites and euchromatin, we found that DEGs (both upregulated and downregulated) and Control genes alike increase in volume, which is suggestive of decondensation. To explore this more stringently, we labeled the 5' end, Middle (Mid), and 3' end of a DEG or control gene and measured the distances across all combinations. At this level of scrutiny, we found that although both the DEG and Control expand from 5' to 3', the DEG preferentially expands from Mid to 3' as well. This suggests that, although both DEGs and controls are decondensing, DEGs are either decondensing to a greater extent and that particular regions of the genes are expanding to a greater extent. It is possible that the condensation or the conformational state of the Mid to 3' region of DEGs is particularly important for gene regulation, with circumstantial evidence coming from our observation that the Mid to 3' distance of the DEG is the only distance that is significantly shorter than the Control in the WT setting. These findings complement recently published work in the RTT field that topoisomerase (an enzyme that decondenses chromatin) inhibition leads to the rectification of the expression of RTT-associated upregulated DEGs in cell culture (Gabel et al., 2015), suggesting that upregulated DEGs are decondensed. The decondensation of DEGs is also in agreement with our methylation data of 5mC-hypomethylation being enriched at DEGs.

Taken together, this work implicates MeCP2 as a global, context-dependent modulator of gene transcription, rather than a gene-specific activator or repressor, and is supportive of MeCP2's role as an architectural protein. Additionally, this work provides insight into the chromatin features of DEGs that may account for their differential susceptible to dysregulation in RTT and posits MeCP2 as a key player in global maintenance of the methylome and chromatin architecture for the preservation of neuronal gene expression. In support of this model, although we focused our analyses on CTCF, we observed thousands of additional alterations in TF footprinting, which are associated with various TFs, suggesting that the sum of all these subtle changes in TF binding could be contributing to the subtle transcriptional alterations observed. And

therefore, yet another model of MeCP2 function is emerging—a synergistic model that treats many of the proposed molecular models as an integrative piece of MeCP2's function rather than treating them as mutually exclusive entities. It is possible that MeCP2 is predominately an architectural protein important for regulating chromatin structures that aid in the activation or repression of genes necessary for long-lasting plasticity in response to neuronal activity. And therefore, in this model, the transcriptional activator or repressor effects that have been attributed to MeCP2 would be secondary to its architectural role.

APPENDIX PART 1

Distinct cellular and molecular environments support aging-related DNA methylation changes in the substantia nigra

One Sentence Summary

This story, which found that distinct cellular and molecular environments account for aging-associated methylation changes in the SN, is published in *Epigenomics*.

Abstract

We aimed to investigate the relationship between brain region-specific changes in global DNA methylation over aging and cellular and molecular environments. We measured two major forms of DNA methylation and analyzed *Dnmt*, *Tet*, and metabolite levels in the striatum and substantia nigra (SN) over aging in healthy male mice. The ratio of 5-hydroxymethylcytosine to 5-methylcytosine increases over aging in the SN, and 5-hydroxymethylcytosine increases preferentially in dopaminergic neurons. Additionally, this age-dependent alteration in methylation correlates with a reduction in the ratio of α-ketoglutarate to succinate in the SN. Distinct cellular and molecular environments account for aging-associated methylation changes in the SN, implicating this epigenetic mechanism in the susceptibility of this brain region to age-related cell loss.

Introduction

Neurons live longer than nearly all other cell types in the body. In the absence of a degenerative disease, the lifespan of most neurons is only limited by the longevity of the organism. However, some brain regions and neuronal types are more vulnerable to age-related cell death and degenerative diseases. One susceptible neuronal population is dopaminergic neurons of the substantia nigra, with a 5-8% loss per decade of life in

humans during normal aging (FEARNLEY and LEES, 1991; Jagust, 2013). This neuronal population is also susceptible to the age-related neurodegeneration in Parkinson's disease (Goedert et al., 2013). The molecular mechanisms underlying the enhanced vulnerability of dopaminergic neurons of the SN to age-related cell loss and degenerative disease remains unclear.

Currently, the leading molecular predictor of chronological age is the pattern of DNA methylation of the C5 position of cytosine at CpG dinucleotide sites, which outperforms all other biomarkers, such as cell cycle inhibitor $p16^{INK4A}$ expression and telomere length attrition rate (Benayoun et al., 2015; Sanders and Newman, 2013; Tsygankov et al., 2009). Quantitative models based solely on methylation patterns are able to predict an individual's age with an average error of only 5.2 years (Benayoun et al., 2015; Bocklandt et al., 2011; Day et al., 2013; Hannum et al., 2013; Maegawa et al., 2010). DNA methylation is an epigenetic mechanism catalyzed by DNA methyltransferases (DNMTs) that allows for the adaptability of gene expression in response to developmental or environmental factors; it plays an essential role in various biological functions such as regulation of gene transcription and establishment and maintenance of cellular identity (Jaenisch and Bird, 2003). Historically, DNA methylation was thought to be a stable, repressive covalent modification in mammalian cells, existing predominantly as 5-methylcytosine (5mC). However, in 2009, two seminal papers confirmed the existence of another cytosine modification, 5-hydroxymethylcytosine (5hmC), which is formed from the oxidation of 5mC by the ten-eleven translocation (TET) family of proteins and is generally thought to correlate with gene activation (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). It has since been discovered that 5hmC can be converted to unmodified cytosine (He et al., 2011; Ito et al., 2011). Thus, environmental influences can lead to changes in DNA methylation levels and patterns.

Age-associated alterations in DNA methylation vary in absolute number and genomic loci depending on organ type (Day et al., 2013; Maegawa et al., 2010), suggesting that different tissues have different rates of changes in DNA methylation over aging. However, it is unclear whether brain regions vary in the degree of DNA methylation changes over aging, given the limited studies examining age-related changes in only a few brain regions. Early chromatography studies found that global 5mC levels in the whole brain of mice decrease with aging (Sanders and Newman, 2013; Tsygankov

et al., 2009; Wilson et al., 1987). Recent studies employing bisulfite treatment of DNA, which allows for genome-wide profiling of DNA methylation (both 5mC and 5hmC, indistinguishably), from the cortex of human and mouse over aging have generated different results, with Day and colleagues (Day et al., 2013) finding an increase in DNA methylation in human cortex using Illumina bead chips that evaluate 26,486 CpG sites and Lister et al. (Lister et al., 2013) observing stable DNA methylation levels in both the mouse or human by employing whole-genome sequencing to assess all C sites in the frontal cortex. Studies focusing on age-related changes in 5mC in the hippocampus of mice using immunohistochemistry or enzyme-linked immunosorbent assays (ELISAs) have also produced conflicting results, with one reporting an increase (Chouliaras et al., 2012b) and another finding no change (Chen et al., 2012), respectively. There are also contradictory results regarding 5hmC in the hippocampus over aging, with one study finding no change using a chemical labeling, affinity enrichment approach followed by sequencing (hMe-Seal) (Szulwach et al., 2011), while two other studies reported an increase in 5hmC via ELISA (Chen et al., 2012) and immunohistochemistry (Chouliaras et al., 2012a).

There are various factors contributing to the lack of congruity among studies. First, different methods have vastly different levels of sensitivity. Second, methods employing bisulfite treatment cannot distinguish between 5mC and 5hmC (Nestor et al., 2010), which is a necessary distinction to make given the biological difference between the two, and therefore, cannot be compared to methods that evaluate 5mC or 5hmC independently. Lastly, the underlying cellular heterogeneity might account for the discrepancies observed across brain regions. Given these incongruous findings, additional research is necessary to determine the extent to which DNA methylation and hydroxymethylation change with aging in the brain, and whether these changes are brain region-specific. It is especially important to assess age-associated changes in DNA methylation in the SN, an area hitherto unexplored, given this region's enhanced vulnerability to aging.

Since DNA methylation integrates environmental and developmental signals for the modulation of transcriptional output, we set out to determine the degree to which global levels of 5mC and 5hmC differ across two brain regions. We compare the DNA methylation status of the SN to the dorsal striatum, the region to which dopaminergic neurons of the SN project (Goedert et al., 2013), to gain insight into the susceptibility of

the SN to age-related cell death. To address whether changes observed in the SN are specific to dopaminergic neurons, we performed semiquantitative immunohistochemistry of 5hmC in tyrosine hydroxylase (TH)-positive, dopaminergic and parvalbumin (PV)-positive neurons of the SN. Finally, we focused on the regulation of 5mC and 5hmC to gain insight into the underlying mechanism contributing to the differences in DNA methylation status over aging between the striatum and SN. First, we measured the expression levels of enzymes responsible for regulating 5mC and 5hmC, and secondly, we assessed metabolites that regulate TET function. Our findings suggest that distinct cellular and molecular environments account for the different DNA methylation states observed between the striatum and SN over aging, and therefore, implicate this epigenetic mechanism in the enhanced, age-related vulnerability of this neuronal population.

Results

Age-related changes in DNA methylation are brain region-specific

We employed reversed-phase HPLC coupled with tandem mass spectrometry (LC-MS/MS and LC-MS/MS/MS), along with the inclusion of stable isotope-labeled standards, and carried out accurate measurement of 5mC and 5hmC across aging (P90, young, and P545, old) from the striatum and SN of male mice. We found that there is no statistically significant, age-dependent effect on total methylated cytosine (5mC plus 5hmC), total 5mC, or total 5hmC (Figure 1A-C). However, the ratio of 5hmC to 5mC (5hmC/5mC) shows an age-dependent change, with an effect of age ($F_{1,36}$=8.358, p=0.0065) and an interaction effect ($F_{1,36}$=4.197, p=0.0478), as well as a significant increase over aging in the SN (~2% increase, p=0.0077), but not the striatum (Figure 1D). 5mC percentage of total methylated cytosine also shows an age-related change, with an effect of age ($F_{1,36}$=8.283, p=0.0067) and an interaction effect ($F_{1,36}$=4.166, p=0.0486), along with a significant decrease over aging exclusively in the SN (~2% decrease, p=0.008) (Figure 1E). Additionally, 5hmC percentage of total methylated cytosine changes over aging, with an effect of age ($F_{1,36}$=8.283, p=0.0067) and an interaction effect ($F_{1,36}$=4.166, p=0.0486), together with an aging-related increase only in the SN (~2% increase, p=0.008) (Figure 1F).

Cell type-specific changes in 5hmC across aging in the substantia nigra

Given the brain region-specific changes in DNA methylation observed in the SN over aging, we subsequently examined the extent to which changes in DNA methylation are specific to nigral dopaminergic neurons, the cell population that is susceptible to degeneration with advanced aging. Parvalbumin (PV)-positive neurons of the SN were used as the comparison group since they are located adjacent to the dopaminergic neurons, allowing for the control of microenvironmental effects (González-Hernández and Rodríguez, 2000). Using 5hmC immunostaining, we evaluated the levels of this epigenetic mark in dopaminergic neurons via co-staining for tyrosine hydroxylase (TH, an enzyme required for dopamine synthesis) or in PV-positive neurons via costaining for PV in the SN of young and old mice (Figure 2A-B). Semiquantitative analyses revealed an increase in 5hmC immunoreactivity in nigral TH-positive, dopaminergic neurons over aging ($p=0.0003$), but not in nigral PV-positive neurons (Figure 2C).

Dnmt and Tet expression over aging in the striatum and substantia nigra

Since DNA methylation status is mediated by two families of enzymes, DNMTs and TETs, we measured mRNA levels of the family members that are highly expressed in the brain: *Dnmt1* and *Dnmt3a,* which methylate cytosine to 5mC, and *Tet2* and *Tet3*, which oxidize 5mC to 5hmC. *Dnmt1* ($p=0.0432$) and Dnmt3a ($p = 0.0185$) expression significantly increased in the striatum, but not in the SN, over aging (Figure 3A-B). *Tet2* and *Tet3* expression levels, on the other hand, did not change in either brain region over aging (Figure 3C-D).

Metabolites in the striatum and substantia nigra across aging

We next investigated the potential contribution of metabolites to the observed differences in global DNA methylation status between the striatum and SN over aging. The metabolite α-ketoglutarate (α-KG) is an obligatory cofactor for TET function (Tahiliani et al., 2009). TETs use molecular oxygen to catalyze oxidative decarboxylation of α-KG, creating a highly reactive intermediate that converts 5mC to 5hmC, as well as

generating carbon dioxide and succinate as byproducts (Kohli and Zhang, 2013). Recently, it was discovered that the intracellular ratio of α-KG to succinate (α-KG/succinate) regulates TET activity, which in turn alters DNA methylation levels (Carey et al., 2015). Therefore, we evaluated the ratio of α-KG/succinate over aging in both brain regions by using a highly quantitative method that employs gas chromatography mass spectrometry (GC-MS) and isotope-labeled standards. In addition to α-KG and succinate, we measured lactate, citrate, and fumarate to ensure that any differences observed in the α-KG/succinate ratio were not due to changes in metabolite flux through the mitochondrial tricarboxylic acid cycle. We found that the levels of all metabolites measured did not statistically differ across aging in either brain region (Figure 4A-E & G-K). However, we found that although the α-KG/succinate ratio remained unchanged across aging in the striatum (Figure 4F), it significantly decreased in the SN (Figure 4L).

Discussion

Although few in number, previous studies suggest that changes in global 5mC and 5hmC occur in the brain over the aging process (Chen et al., 2012; Chouliaras et al., 2012a; Day et al., 2013; Lister et al., 2013; Szulwach et al., 2011; Wilson et al., 1987); however, the findings have been inconsistent, likely due to the inherent differences in methods detecting 5mC and 5hmC and the brain regions assessed. Given the different biological functions of 5mC and 5hmC, and the discrepancies regarding age-associated changes of these two major forms of DNA epigenetic modifications in the brain, we used a highly sensitive LC-MS/MS and LC-MS/MS/MS method to measure 5mC and 5hmC levels from the striatum and SN at two different ages (P90, young and P545, old). We specifically chose the SN given its enhanced vulnerability to aging, and the striatum was selected as the brain region for comparison since this is the area to which dopaminergic neurons of the SN project (Goedert et al., 2013).

We found selective changes in these two epigenetic marks over aging in the SN, but not striatum. Specifically, we observed an increase in the ratio of 5hmC/5mC, which correlated with a 2% decrease in the 5mC percentage of total methylated cytosine and a

2% increase in the 5hmC percentage of total methylated cytosine. This suggests that although the absolute levels of 5mC and 5hmC do not significantly change, the percentage of sites that are converted from 5mC to 5hmC increases in the SN over aging. Although a 2% change in DNA methylation seems marginal, this could account for a change at ~1 million cytosines, given that there are ~50 million cytosines that are methylated in the mouse genome (Lister et al., 2013), and therefore, the changes observed are substantial and potentially biologically relevant. When we examined two neuronal types in the SN, we found a significant increase in 5hmC in the TH-positive, dopaminergic population, but not in neighboring PV-positive neurons. Despite the limitation of a small sample size (n=4 per group), these findings suggest that changes in global levels of 5hmC during aging are both brain region- and cell type-specific. Since 95% of the striatum is composed of GABAergic medium spiny neurons (Tepper and Bolam, 2004) and since PV-positive neurons of the SN are GABAergic, our findings suggest that GABAergic neurons maintain their DNA methylation status across aging, independent of the brain region. This study elucidates a novel age-associated molecular change in the vulnerable dopaminergic population and indicates that changes in 5hmC are contingent upon the intracellular molecular environment, rather than extracellular environment (i.e. brain region).

Therefore, we subsequently examined the molecular environment that could underlie the difference in DNA methylation status between the striatum and SN over aging and made two discoveries. First, we found an age-associated increase in both *Dnmt*s that are highly expressed in the brain, *Dnmt1* and *Dnmt3a*, in the striatum, but not the SN. It's plausible that the increase in *Dnmt*s over aging in the striatum contributes to the stability of the ratio of 5hmC/5mC, unlike the SN. Our second finding was the decrease in the α-KG/succinate ratio in the SN, but not striatum, albeit with a small sample size (n=4 for the young group and n=6 for the old group). This suggests that brain region-specific metabolic changes over aging could contribute to region-specific changes in DNA methylation. Since we are measuring steady-state levels of the metabolites, this finding supports a model in which TETs are more active in the SN of old animals given the relative depletion of the α-KG pool, which is used by TETs to oxidize 5mC to 5hmC, and an increase in succinate, which is a byproduct of TET activity. Importantly, this result is consistent with our finding of an increase in the 5hmC/5mC

ratio over aging in the SN, but not striatum.

Additional research is necessary to define a potentially causative role of the enzymatic expression and/or metabolic ratio alterations with the age-related differences in DNA methylation status between the striatum and SN. There are other regulators of TETs that were not assessed in this work that could also contribute to the brain region-specific DNA methylation differences across aging, such as iron levels, ascorbic acid levels, and calpain activity (Blaschke et al., 2013; Tahiliani et al., 2009; Wang and Zhang, 2014; Yin et al., 2013). Additionally, determining the genomic locations of these methylation changes, and consequential alterations in gene expression (Mellén et al., 2012a), will aid in elucidating how normal aging predisposes nigral dopaminergic neurons to cell death.

Conclusions

This work suggests that an increase in TET activity, as indicated by the alteration in the α-KG/succinate ratio, underlies the selective, age-dependent increase in the ratio of 5hmC/5mC in the SN. This work also supports that a brain region less susceptible to aging, like the striatum, is able to employ protective mechanisms, such as increasing *Dnmt* expression, to stabilize DNA methylation profiles over aging. Taken together, our findings suggest that distinct cellular and molecular environments support the different DNA methylation states observed between the striatum and SN over aging, and specifically within dopaminergic neurons, implicating this epigenetic mechanism in the enhanced vulnerability of this neuronal population in aging.

Summary Points

- There are age-associated changes in DNA methylation in the substantia nigra, but not striatum.
    - Aging did not significantly affect any of the DNA methylation parameters assessed in the striatum.

    - In the substantia nigra, on the other hand, the ratio of 5hmC/5mC significantly increases over aging, suggesting that although the absolute levels of 5mC and 5hmC do not significantly change, the percentage of sites that are converted from 5mC to 5hmC increases.
- There is a selective increase in 5hmC over aging in tyrosine hydroxylase-positive dopaminergic neurons, but not parvalbumin-positive neurons, of the substantia nigra.
    - Our results indicate that changes in 5hmC over aging are cell type-dependent and contingent upon the intracellular molecular environment, rather than the extracellular environment (i.e. brain region).
- There is an age-dependent increase in *Dnmt* expression in the striatum, but not substantia nigra.
    - In the substantia nigra over aging, *Tet2*, *Tet3*, *Dnmt1*, and *Dnmt3a* mRNA levels remain unchanged.
    - In the striatum, although *Tet2* and *Tet3* mRNA levels are stable across aging, *Dnmt1* and *Dnmt3a* levels increase, suggesting that an increase in *Dnmt*s contributes to the stability of the ratio of 5hmC/5mC over aging, unlike the substantia nigra.
- We found a substantia nigra-selective reduction in the α-KG to succinate ratio over aging.
    - Although the α-KG/succinate ratio remains unchanged across aging in the striatum, it significantly decreases in the substantia nigra.
    - Since we are measuring steady-state levels of the metabolites, this finding supports a model in which TETs are more active in the substantia nigra of old animals given the relative depletion of the α-KG pool, which is used by TETs to oxidize 5mC to 5hmC, and an increase in succinate, which is a byproduct of TET activity.
- Conclusion
    - Our findings suggest that distinct cellular and molecular environments account for the different DNA methylation states observed between the striatum and substantia nigra over aging, and therefore, implicate this epigenetic mechanism in the enhanced vulnerability of this neuronal population over aging.

Materials and Methods

Animal tissue

All experiments were performed according to protocols approved by the Institutional Animal Care and Use Committee of the University of Pennsylvania. All mice were on the C57BL/6 genetic background. Tissues from healthy young male animals were collected at postnatal day 90 (P90), and tissues from healthy old male animals were collected at P545. Bilateral dorsal striatum and substantia nigra samples were collected as previously described (Salvatore et al., 2012; Spijker, 2011). Briefly, a mouse brain matrix (ASI-instruments) was used to coronally section the brain. Bilateral dorsal striatum collection started rostrally when the anterior portion of the anterior commissure crosses the midline (Bregma 0.20mm) and ended caudally with the loss of the connection between the dorsal 3rd ventricle and the lateral ventricles (Bregma -0.58mm). Bilateral SN collection started rostrally when the hippocampus completely wraps around the midbrain (Bregma -2.92mm) and ended caudally with the disappearance of distinct blades of the dentate gyrus (Bregma -3.64mm).

Reverse-phase HPLC coupled with tandem mass spectrometry

LC-MS/MS and MS/MS/MS measurements of 5mC and 5hmC were performed blind to age and brain region. DNA was isolated using an AllPrep DNA/RNA Micro Kit (Qiagen) and treated with RNase (Roche). DNA (1 μg) was then treated with nuclease P1 (0.5 U) and phosphodiesterase 2 (0.00025 U) at 37°C for 48 hr in a digestion buffer (30 mM sodium acetate, 1.0 mM zinc acetate, 1 mM EHNA, pH 5.6), followed by the addition of alkaline phosphatase (0.05 U) and phosphodiesterase 1 (0.0005 U) in another buffer (50 mM Tris-HCl, pH 8.9) for 2 hr. For MS quantification of 5-hydroxymethyl-2′-deoxycytidine (5hmC) and 5-methyl-2′-deoxycytidine (5mC), 50 fmol of [1,3-$^{15}N_2$-2′-D]-5hmC and 600 fmol of [1′,2′,3′,4′,5′-D_5]-5mC were added to the digestion mixture of 50 ng genomic DNA. Enzymes were removed by chloroform extraction. The resulting aqueous layer was subjected to LC-MS/MS and MS/MS/MS analyses on an LTQ linear ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) that was equipped with an electrospray ionization interface and coupled to an Agilent 1200 capillary HPLC

(Agilent Technologies, Santa Clara, CA). Separation was carried out on a 0.5 × 250 mm Zorbax SB-C18 column (5 μm in particle size, Agilent Technologies, Santa Clara, CA) with a flow rate of 8.0 μL/min. A solution of 2 mM sodium bicarbonate (pH 7.0, solution A) and methanol (solution B) were used as mobile phases and a gradient of 5 min 0-20% B and 25 min 20-70% B was employed for the separation. MS settings were as follows: electrospray voltage, 5 kV; capillary temperature, 275°C; capillary voltages, 38 V; tube lens voltages, 60 V; sheath gas flow rate, 15 arbitrary units. Selected-ion chromatograms were plotted for monitoring the transitions of *m/z* 258→142→124 for 5hmC, *m/z* 261→144→126 for [1,3-$^{15}N_2$, 2′-D]-5hmC, *m/z* 242→126 for 5mC, and *m/z* 247→126 for [1′,2′,3′,4′,5′-D_5]-5mC. The calibration curve for 5hmC and 5mC was constructed previously (Liu et al., 2015; 2013). The numbers of moles of 2′-deoxyguanosine (dG), 5mdC, and 5hmC in each sample were calculated from the peak area ratios, the calibration curves, and the amounts of stable isotope-labeled standards added. 5mC and 5hmC were then calculated as the percentage of dG (% dG) by dividing the moles of 5mC or 5hmC, respectively, by the moles of dG. Total methylated cytosine was calculated by adding 5mC (% dG) and 5hmC (% dG) per sample, the ratio of 5hmC to 5mC was calculated by dividing 5hmC (% dG) by 5mC (% dG), 5mC percentage of total methylated cytosine was calculated by dividing 5mC (% dG) by total methylated cytosine, and 5hmC percentage of total methylated cytosine was calculated by dividing 5hmC (% dG) by total methylated cytosine. The sample size is 10 animals per group (n=10 bilateral young striatum samples, n=10 bilateral old striatum samples, n=10 bilateral young SN samples, and n=10 bilateral old SN samples). To reduce biological variation, the striatum and SN samples were taken from the same animal. Additionally, RNA was taken from these same tissue samples for the quantitative RT-PCR, and therefore, a total of 10 old and 10 young animals (20 animals total) were used for the reverse-phase HPLC coupled with tandem mass spectrometry and quantitative RT-PCR experiments.

Semiquantitative Immunohistochemistry

Mice were anesthetized with 1.25% Avertin, transcardially perfused with 4% paraformaldehyde in PBS, and postfixed with 4% paraformaldehyde in PBS overnight at 4°C. Immunohistochemistry was carried out on 20 μm free-floating sections as previously described (Liao et al., 2008).Tissue sections from the SN (see animal tissue section above for the Bregma coordinates of the SN) were incubated with a 5hmC

antibody (1:1000, Active Motif (39791)) in blocking solution overnight at 4°C. The next day, the tissues were removed from the 5hmC antibody solution and incubated with either a tyrosine hydroxylase (TH) antibody (1:1000, Abcam (ab76442)) or a parvalbumin (PV) antibody (1:1000, Millipore (mab1572)) for 1 hr at 22-24°C. Fluorescence detection of 5hmC was performed using Alexa Fluor 488 (1:1000, Invitrogen) and of TH or PV using Alexa Fluor 594 for 1 hr at 22-24°C. Images were captured and subsequently analyzed blind to age (the marked differences in neuronal morphology prevent blinding of neuronal cell types). Images were captured using a Leica confocal microscope with identical settings for laser power, detector gain amplifier offset, and pinhole diameter for each channel for TH-positive neurons across both ages, and the same for PV-positive neurons. TH-positive and PV-positive neurons were analyzed from the same biological samples, with 4 animals in the young group and 4 animals in the old group. 38 bilateral TH-positive neurons and 15 bilateral PV-positive neurons were analyzed per animal, giving a total of 152 TH-positive and 60 PV-positive neurons that were analyzed at each age. 5hmC fluorescence was measured using ImageJ as previously described (Burgess et al., 2010). Briefly, the free form tool was used to calculate the area and integrated density of 5hmC for each TH-positive or PV-positive neuron with a nucleus in focus. The freeform tool was then used to measure the mean fluorescence of three background regions in close proximity to the 5hmC measured (three regions were selected to improve accuracy). Each of the three background readings was multiplied by the area of the 5hmC tracing to give an integrated density of each of the three background readings dependent upon the size of the 5hmC tracing. These calculated integrated densities were then averaged and subtracted from the integrated density of the 5hmC tracing to get the corrected 5hmC fluorescence.

Quantitative RT-PCR

Total RNA was isolated using an AllPrep DNA/RNA Micro Kit (Qiagen) and treated with DNase (Qiagen). Total RNA (400 ng) was reverse transcribed by oligo-dT priming using SuperScriptIII reverse transcriptase (Invitrogen). Quantitative real-time PCR was performed on the resulting cDNA using Taqman Probes with primer pairs that are exon-spanning. All mRNA levels of genes of interest were normalized to *Hprt* mRNA levels. The sample size was 10 animals per group (see LC-MS/MS and MS/MS/MS methods section above for more details).

Mass spectrometry for the measurement of metabolites

Concentrations of metabolites were determined blind to age and brain region by the Metabolomics Core at The Children's Hospital of Philadelphia, using a previously described isotope dilution approach (Nissim et al., 1996; Weinberg et al., 2000). Briefly, an aliquot of the sample was spiked with a mixture of ^{13}C-labeled organic acids. GC-MS measurement of ^{13}C isotopic abundance in each sample was then performed. Concentrations of metabolites in the sample were calculated as previously described (Nissim et al., 1996). 4 animals are in the young striatum group, 4 animals are in the young SN group, 6 animals are in the old striatum group, and 6 animals are in the old SN group. The striatum and SN samples were taken from the same animal, and therefore, 10 animals in total were used for this analysis.

Statistics

Statistics were performed using Prism 6.0 (GraphPad Software) and RStudio. The D'Agostino & Pearson omnibus normality test was used to test normality ($p > 0.05$), and the F-test was used to test for equal variances ($p > 0.05$). The individual statistical tests performed for each experiment can be found in the figure legends

Figures

Figure 1

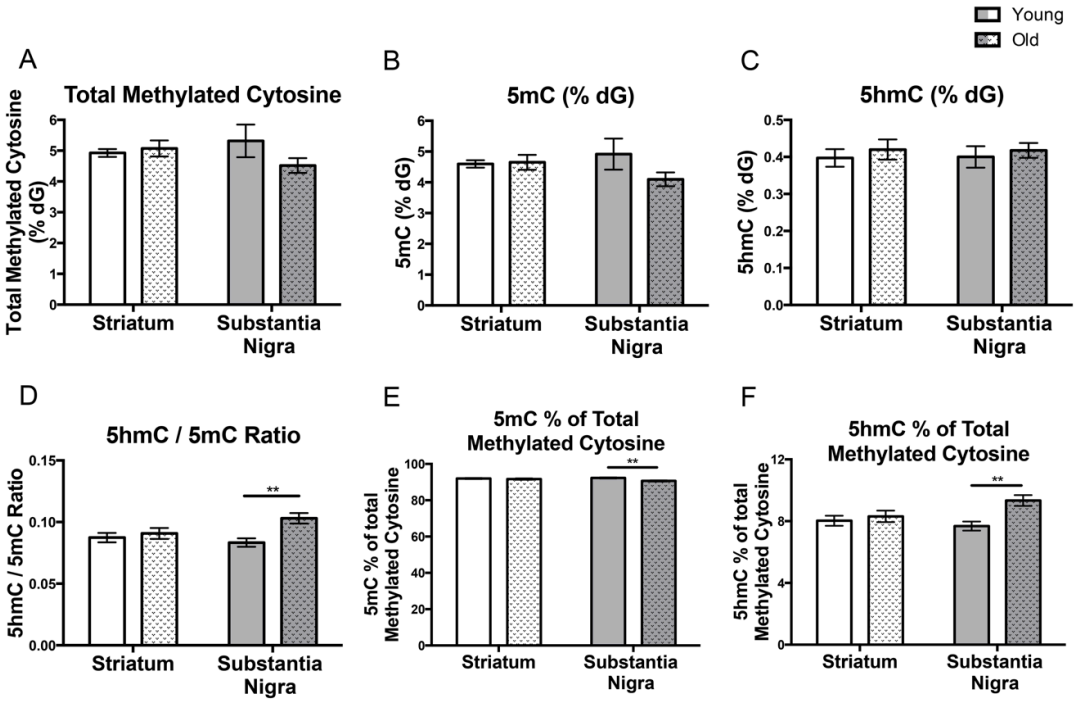


Figure 1: Age-associated changes in global DNA methylation levels over aging in the substantia nigra, but not striatum. Measurement of global total methylated cytosine (5mC plus 5hmC; Kruskal-Wallis rank sum test) (**A**), 5mC (Kruskal-Wallis rank sum test) (**B**), 5hmC (Kruskal-Wallis rank sum test) (**C**), the ratio of 5hmC to 5mC (two-way ANOVA with pairwise comparisons with Bonferroni correction, **p<0.01) (**D**), 5mC percentage of total methylated cytosine (two-way ANOVA with pairwise comparisons with Bonferroni correction, **p<0.01) (**E**), and 5hmC percentage of total methylated cytosine (two-way ANOVA with pairwise comparisons with Bonferroni correction, **p<0.01) (**F**) by LC-MS/MS and LC-MS/MS/MS across aging in the two brain regions. Solid bars indicated the young cohort and dotted bars indicate the old cohort (mean ± SEM; n = 10 per group; dG, 2′-deoxyguanosine; 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine)

Figure 2

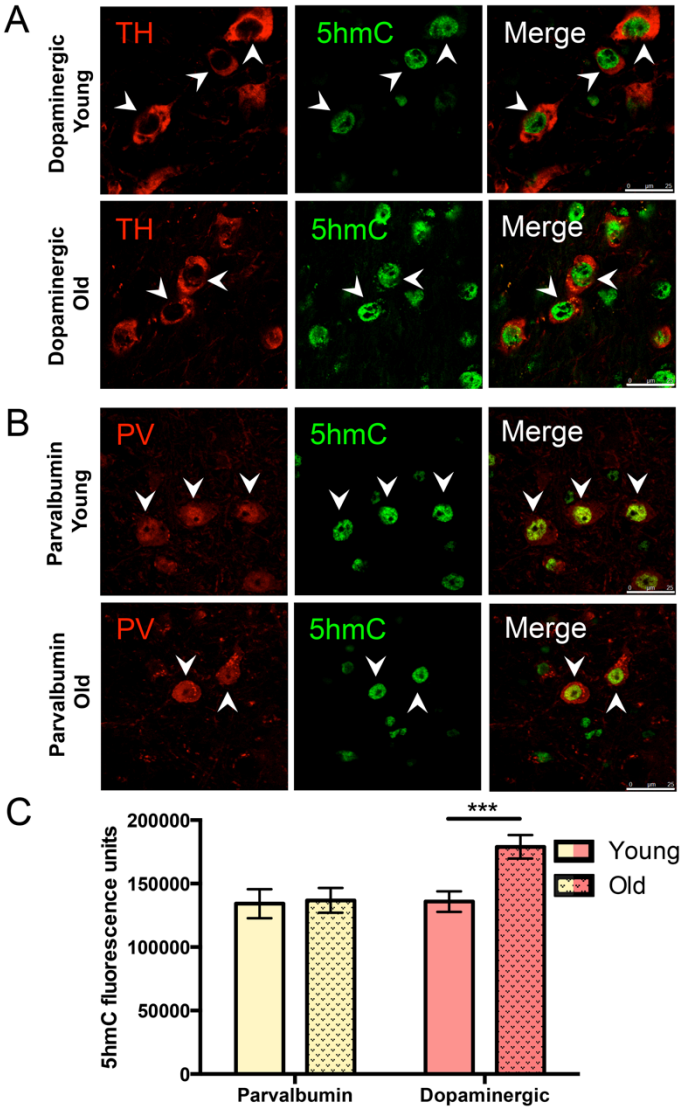


Figure 2: Selective increase in 5hmC over aging in tyrosine hydroxylase-positive dopaminergic neurons, but not parvalbumin-positive neurons, of the substantia nigra. Immunohistochemistry in substantia nigral sections of young and old mice showing 5hmC (green) in tyrosine hydroxylase (TH)-positive, dopaminergic neurons (red) (**A**) and 5hmC in parvalbumin (PV)-positive neurons (red) (**B**). Cells that were analyzed are marked with an arrow. (**C**) Quantification of 5hmC fluorescence intensity in young and old PV-positive and TH-positive, dopaminergic neurons. Solid bars indicated the young cohort and dotted bars indicate the old cohort (mean ± SEM; n=152 neurons from 4 animals (38 neurons per animal) for TH-positive,

dopaminergic neurons per age group; n=60 neurons from 4 animals (15 neurons per animal) for PV-positive neurons per age group; ***p = 0.0003; Dunn's multiple comparisons test; TH, tyrosine hydroxylase; PV, parvalbumin; 5hmC, 5-hydroxymethylcytosine).

Figure 3

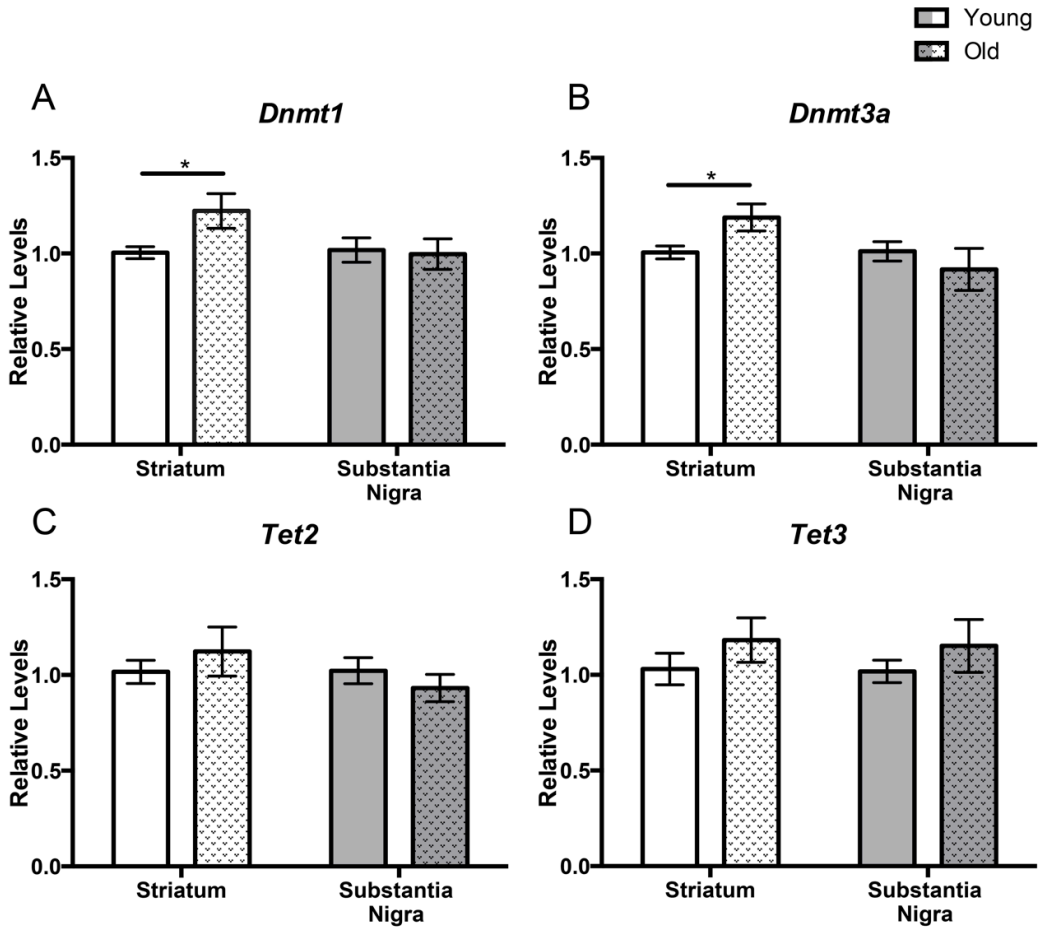


Figure 3: An age-dependent increase of *Dnmt* expression is specific to the striatum. Relative changes in mRNA levels of *Dnmt1* (striatum, two-tailed t-test with Welch's correction; SN, two-tailed t-test) (**A**), *Dnmt3a* (striatum, Mann-Whitney test; SN, two-tailed t-test with Welch's correction) (**B**), *Tet2* (two-tailed t-test) (**C**), and *Tet3* (striatum, two-tailed t-test; SN, two-tailed t-test with Welch's correction) (**D**) as measured by quantitative RT-PCR with the young group serving as the reference for each brain region. Solid bars indicated the young cohort and dotted bars indicate the old cohort (mean ± SEM; n = 10 per group; *$p < 0.05$).

Figure 4

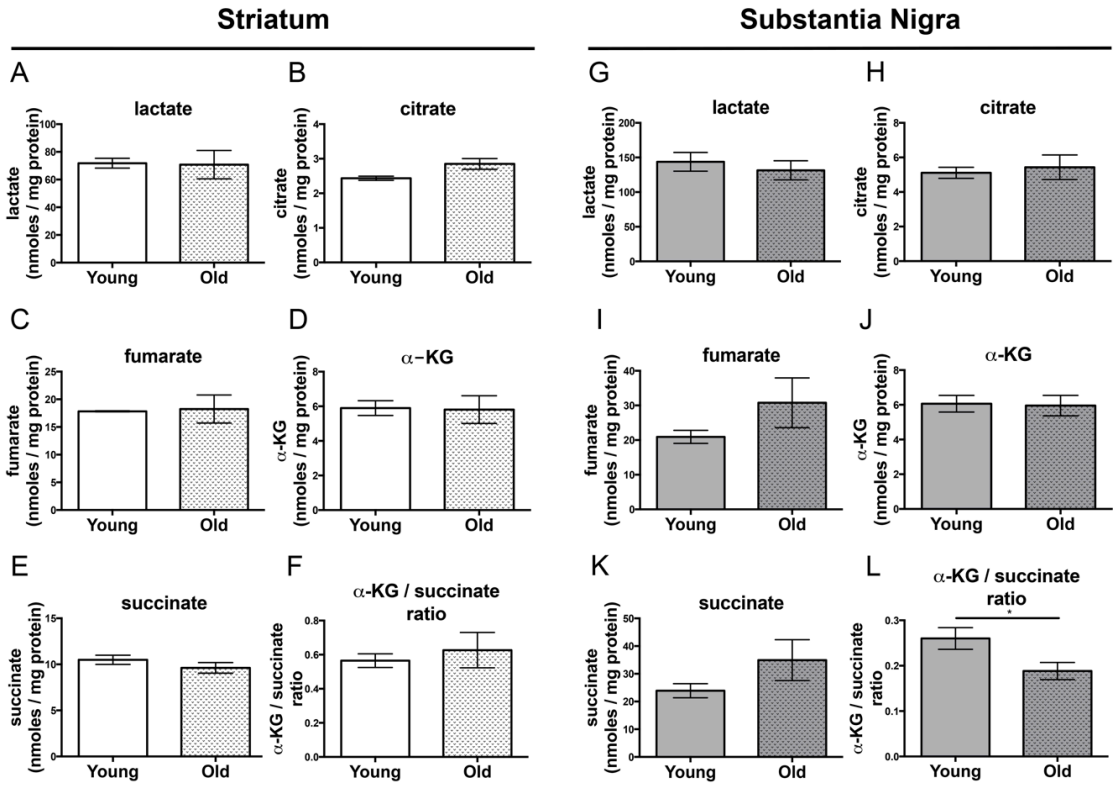


Figure 4: Decrease in the α-KG to succinate ratio in the substantia nigra, but not striatum, over aging. Quantification of lactate (two-tailed t-test) (**A & G**), citrate (two-tailed t-test) (**B & H**), fumarate (two-tailed t-test with Welch's correction) (**C & I**), α-KG (two-tailed t-test) (**D & J**), succinate (two-tailed t-test) (**E & K**), and α-KG to succinate ratio (two-tailed t-test) (**F & L**) as measured by mass spectrometry with isotope-labeled standards. Solid bars indicated the young cohort and dotted bars indicate the old cohort (mean ± SEM; n = 4 for young group and n = 6 for old group; *p < 0.05; α-KG, α-ketoglutarate)

APPENDIX PART 2

Tet and 5hmC in Neurodevelopment and the Adult Brain

One Sentence Summary

This review, which summarizes the importance of 5hmC and TETs in neuronal function, is a book chapter published in "*DNA Modifications in the Brain: Neuroepigenetic Regulation of Gene Expression.*"

Keywords

5-hydroxymethylcytosine, 5hmC, Tet, Tet1, Tet2, Tet3, neurodevelopment, adult, methylation, brain

Abstract

Since the discovery in 2009 that 5-hydroxymethylcytosine (5hmC) is remarkably high in the brain, much effort has been put into understanding the role of this epigenetic mark in neuronal function. With the advent of various approaches for the detection of 5hmC, in less than a decade, the field has gained invaluable insight into the role of 5hmC in gene regulation, neurodevelopment, neurogenesis, differentiation, and electrophysiological and circuit properties. Parallel to these advances has been the substantial gain of knowledge for the functional role of the enzymes that convert 5-methylcytosine (5mC) to 5hmC, the ten-eleven translocation family of enzymes (Tet1, Tet2, and Tet3; Tets), in neuronal processes. Various knockout/knockdown and overexpression techniques have been employed against Tets, implicating these enzymes in maintaining and establishing 5hmC patterns, hippocampal function, differentiation, and electrophysiological properties. Here we summarize the current understanding of 5hmC and Tets in the developing and adult brain.

Introduction

DNA methylation at the 5-carbon of cytosine (C) is widely distributed throughout the mammalian genome, with 5% of all C and 85% of all cytosine-phosphate-guanine dinucleotides (CpGs) being methylated (Lister et al., 2013). Such methylation plays an essential role in various biological functions such as the regulation of gene transcription, establishment and maintenance of cellular identity, imprinting, silencing of transposons and repetitive elements, and X-chromosome inactivation (Jaenisch and Bird, 2003). Historically, DNA methylation of C was thought to be a stable covalent modification, existing exclusively as 5-methylcytosine (5mC). However, this view was challenged in 2009 when two seminal papers published in parallel described another C modification, 5-hydroxymethylcytosine (5hmC), which is formed from the oxidation of 5mC (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). Tahiliani et al. also described enzymes that were able to convert 5mC to 5hmC, the ten-eleven translocation family of enzymes, or Tet enzymes. These enzymes were found to be paralogues of the base J binding proteins (JBP) from the parasite *Trypanosoma brucei*. However, instead of converting the base thymine to 5-hydroxymethyl-uracil, as JBP enzymes do, Tet enzymes convert modified cytosines to 5hmC (Tahiliani et al., 2009).

Since its rediscovery in 2009, 5hmC has added an important dimension in understanding the epigenetic regulation of neuronal function (Kriaucionis and Heintz, 2009; Penn et al., 1972; WYATT and COHEN, 1953). The particular importance of 5hmC in the brain is highlighted by the fact that while global 5mC levels are similar across different tissue types, levels of 5hmC are highly variable with the highest concentration in the central nervous system (CNS) (Globisch et al., 2010; Kriaucionis and Heintz, 2009; Münzel et al., 2010). Notably, all mature neurons in the central nervous system are post-mitotic, meaning that they no longer divide. While it was previously known that 5mC could be passively removed through cell division, the discovery of Tet enzymes meant that 5mC could be actively removed via oxidation of Tets to 5hmC, and this could occur in post-mitotic cells. Further, 5hmC can also be removed, completely reverting the base back to unmodified C. Removal of 5hmC occurs first via iterative oxidation by the Ten-eleven translocation family of proteins (Tet1, Tet2, and Tet3, collectively referred to as Tets) which convert 5hmC to 5-formylcytosine (5fC), and subsequently to 5-carboxylcytosine (5caC) (Ito et al., 2011). Finally, 5caC is converted to C by thymine

DNA-glycosylase (TDG)-mediated base excision repair (BER) (He et al., 2011) (**Figure 1**). Therefore, it has been called into question whether 5hmC is a mere transient, uninformative byproduct of DNA demethylation or a stable, purposeful epigenetic mark with biological functional significance. Throughout this chapter, we will highlight recent findings that have greatly advanced our understanding of the role of 5hmC in brain.

Global 5hmC and Tet expression throughout the adult brain

To gain a better understanding of the biological function of 5hmC, initial studies interrogated the levels of 5hmC in various tissues using isotope-based liquid chromatography-mass spectrometry (LC-MS). It has been unanimously found that although this mark is present throughout the body, 5hmC levels are markedly higher in the nervous system than other tissues. 5hmC concentrations in the central nervous system are brain region-specific: 5hmC constitutes 0.7% of cytosine (C) bases in the cortex and hypothalamus; 0.6% in the brainstem, olfactory bulb, and hippocampus; 0.5% in the spinal cord; 0.4% in the midbrain; and 0.3% in the cerebellum (Globisch et al., 2010; Kriaucionis and Heintz, 2009; Münzel et al., 2010). For comparison, tissues with the next closest levels of 5hmC, ranging from 0.15-0.17%, are the kidney, nasal epithelium, bladder, heart, muscle, and lung. In the pituitary gland, a non-neuronal structure at the base of the brain, 5hmC constitutes only 0.06% of total C, which supports that high concentrations of 5hmC are specific to neural tissues, rather than topographic location (Globisch et al., 2010). Further support of high levels of 5hmC being specific to neurons comes from a comprehensive genome-wide study that found 5hmC levels are higher in neuronal than in non-neuronal cell types within the frontal cortex (Lister et al., 2013).

It remains unknown what accounts for the high levels of global 5hmC within the brain as compared to other tissues. Since Tets are expressed at similar levels in other tissue with much lower 5hmC levels, overall expression levels of this family of proteins cannot account for the difference. It also remains to be determined what accounts for the differences in 5hmC levels across brain regions. Although the cortex, hippocampus, and cerebellum have vastly different 5hmC levels that range from 0.7-0.3% of total C, the regions have similar Tet expression levels (Szwagierczak et al., 2010). It's possible that brain region-specific levels of 5hmC are dependent upon the relative levels of Tet-

dependent cofactors, iron and alpha-ketoglutarate (a-KG) (Tahiliani et al., 2009), or other potential regulators, such as ascorbic acid, calpain, succinate, and fumarate. In embryonic stem cell (ESC) culture, supplementation of ascorbic acid decreased the levels of 5mC and increased levels of 5hmC (Blaschke et al., 2013; Wang and Zhang, 2014; Yin et al., 2013). This has unique implications for Tet activity in the brain, since the brain has the highest levels of ascorbic acid in the body, mostly due to high expression of the ascorbic acid-specific transporter SVCT2 (Lin JACS 135 2013). More research is this area is warranted to elucidate what controls the levels of 5hmC.

In addition to brain region-specificity, global 5hmC levels are also cell type-specific and developmentally dependent. The cell type-specificity of 5hmC was noted in the paper that rediscovered it in 2009 with the description that 5hmC constitutes 0.6% of total CpGs in Purkinje neurons, but only 0.2% of granule cells (Kriaucionis and Heintz, 2009). This has been confirmed by an independent group that found that the levels and genomic distribution of 5hmC vary across Purkinje cells, granule cells, and Bergmann glia of the cerebellum (Mellén et al., 2012b). The establishment of these cell type differences in 5hmC levels in the brain occurs between embryonic day 12.5 and 13.5, which is the time at which active specification of neurons and glial cells commences, supporting that the 5hmC is important for neuronal identity (Wheldon et al., 2014). The rapid increase in global 5hmC during neuronal differentiation and synaptogenesis also highlights the developmental importance of this epigenetic mark in the brain (Hahn et al., 2013; Song et al., 2011; Szulwach et al., 2011).

In the adult mouse (10-12 wk) brain, neuronal activity can also lead to changes in global 5hmC. When neuronal activity is increased via flurothyl induced seizures, there is a significant reduction in 5mC and 5hmC in area CA1 of adult mice at 24 hours after the occurrence of seizures (Kaas et al., 2013). This differs from the dentate granule cells in that synchronous activation of dentate granule cells in the adult mice by electroconvulsive stimulation or voluntary running does not lead to a global change in methylation, as there are similar number of CpGs that become methylated and demethylated under these conditions as assessed by methyl-sensitive cut counting (MSCC) method (Guo et al., 2011a).

These findings on the global 5hmC patterns throughout the brain and other tissues suggest that 5hmC might be especially important in neuronal function and identity. However, to truly unravel the biology of 5hmC, determining the genomic

location of this epigenetic mark and its effect on gene expression is necessary, which will be discussed in the following section.

The Genomic Distribution of 5hmC

Recent advances in sequencing have allowed for the mapping of the genomic distribution of 5hmC in the brain, providing invaluable insight into the biological function of this epigenetic mark. Various affinity- and enzyme-based methods have been developed for profiling 5hmC genome wide, with the three most commonly used approaches. First is 5-hmC selective chemical labeling (5hmC-Seal) in which 5hmC is converted to biotin-N3-5-hydroxymethyl-cytosine (Biotin-N3-5-gmC) for affinity enrichment through a two step synthesis (Song et al., 2011). Second is hydroxymethylated DNA immunoprecipitation in which 5hmC is enriched via antibodies that specifically bind to 5hmC (Jin et al., 2011). The third is Tet-assisted bisulfite sequencing (TAB-Seq) in which 5hmC is exclusively protected via glycosylation and Tet-mediated oxidation before bisulfite treatment (Yu et al., 2012b).

With the employment of these approaches, general features have emerged. Quantitatively, intragenic and global 5hmC levels are equivalent across chromosomes in both human and mouse, except for the male ChrX, which has 22% lower enrichment (Lister et al., 2013; Mellén et al., 2012b; Szulwach et al., 2011). 5hmC is predominantly found in CpGs in both the human and mouse across development (Lister et al., 2013; Wen et al., 2014). In the fetal mouse brain, 5% of CpGs and 0% of non-CpGs (CH, where H = A, C, or T) are hydroxymethylated, whereas in the adult mouse frontal cortex (6 wk), hydroxymethylation occurs at 19% of CpGs and 0.02% of CHs. This epigenetic mark is found across transcriptional end sites, intragenic regions, DNaseI-hypersensitive sites (DHSs), and enhancers (Lister et al., 2013). It is present at both poised enhancers (solely marked by distal H3K4me1) and active enhancers (marked by distal H3K4me1 and H3K27ac), with poised enhancers constituting 32.6% of all 5hmC sites and active enhancers constituting 28.6%. Major satellite and promoter regions, on the other hand, are relatively devoid of 5hmC (Wen et al., 2014). Most 5hmC (71%) is found intragenically, with a much higher concentration at exons than introns (Szulwach et al., 2011). These findings on the genomic distribution of 5hmC implicate that it may play a role in gene regulation.

Given the relatively high enrichment of 5hmC across exons, and the proposed hypothesis that methylation modulates alternative splicing (Maunakea et al., 2013), studies have evaluated the role of this epigenetic mark in splicing. 5hmC appears to play an important role in alternative exon usage in the mammalian brain, as there is a distinct pattern of methylation at exon-intron boundaries. Firstly, there is a sharp decrease in 5hmC at the 5' end of the intron at the exon-intron boundary. Secondly, across exons from 5' to 3', there is a substantial increase in 5mC levels and a less pronounced decrease in 5hmC (Khare et al., 2012b; Wen et al., 2014). Thirdly, 5hmC levels, but not 5mC levels, within 20 basepairs of the exon-intron boundary correlate with constitutively used exons relative to alternatively spliced exons. The importance of these features in alternative exon usage rather than general transcription is highlighted by the fact that first exons have much lower 5mC and 5hmC than internal exons and that exons of intron-less or single-exon genes have lower 5hmC than multiple-exon genes (Khare et al., 2012b). This feature seems to be specific to brain tissue since neither 5mC nor 5hmC correlate with exon usage in the liver. Thirdly, flanking the highly conserved "GT" splice site sequence at the 5' splicing sites (5' ss) of internal exons, at the -1 and -2 positions on the exon side and +4 and +5 positions of the intron side of the exon-intron boundary, are two prominent 5hmC peaks. 5mC, on the other hand, does not exhibit this type of pattern in the brain (Khare et al., 2012b; Wen et al., 2014). This patterning of 5hmC at the 5' ss seems to brain-specific as 5mC, rather than 5hmC, marks exon-intron boundaries in the liver (Khare et al., 2012b). Further examination of alternatively spliced (AS) exons by RNA-seq, it was found that low or no methylation flanking the 5' ss is associated with significantly more exon skipping than methylated or hydroxymethylated boundaries. This suggests that demethylation is associated with alternative splicing events, which is consistent with the idea that DNA 5hmC aids in exon recognition and inclusion (Khare et al., 2012b; Maunakea et al., 2013; Wen et al., 2014).

In addition to the correlation of 5hmC in exon usage, there is also a stronger correlation between intragenic 5hmC levels and gene expression in both main cell types of the brain, neurons and glia (Lister et al., 2013; Mellén et al., 2012b; Song et al., 2011). 5mC levels across the gene body, on the other hand, negatively correlate with gene expression (Lister et al., 2013; Mellén et al., 2012b; Wen et al., 2014). The best correlate with gene expression is the intragenic ratio of 5hmC to 5mC (5hmC/5mC). This correlation extends to the tissue-specific and cell subtype-specific level, with a relatively

high 5hmC/5mC ratio correlating with brain region-specific and cell type-specific differentially expressed transcripts (Lister et al., 2013; Mellén et al., 2012b). When the 5hmC genomic distribution and expression profiles of three different cell types of the cerebellum (Purkinje cells, granule cells, and Bergmann glia) were compared, it was found that cell type-specific transcripts have higher intragenic 5hmC/5mC levels than the other cell types (Mellén et al., 2012b). This cell-type specific patterning also hold true across neuronal differentiation as cell type-specific genes that are developmentally regulated gain intragenic 5hmC and lose intragenic 5mC across differentiation (Colquitt et al., 2013).

There is also a significant difference between 5hmC and 5mC for strand bias of expressed genes in both glia and neurons. When comparing the lowest to the highest expressed genes, there is a seven-fold bias in 5hmC enrichment on the sense strand and a five-fold bias in 5mC enrichment on the antisense strand. These findings suggest that 5hmC enrichment on the sense strand is correlated with activation. In agreement with this is the finding that 5hmC is inversely related to two repressive histone modifications, H3K27me3 and H3K9me3. On the other hand, these two histone modifications correlate with 5mC (Wen et al., 2014). Genes enriched for 5hmC in the mammalian brain relative to other tissues are synapse-related (Khare et al., 2012b). These findings highlight the importance of 5hmC in the activation of genes specific to the brain, neuronal subtypes, and neuronal function.

While 5hmC levels correlate highly with gene expression, it is unknown how this is accomplished. In other words, what are the mechanisms that allow 5hmC to influence or be influenced by gene expression? It is hypothesized that there are specific proteins that are able to bind 5hmC and influence gene expression through binding to additional protein complexes or initiating a signaling cascade(s). The first paper to inquire into what proteins can bind 5hmC used quantitative mass spectrometry (MS)-based proteomics. They used fragments of DNA that contained 5hmC to isolate interacting proteins from mouse ESCs and analyzed the resulting proteins using LC-MS/MS. They identified a large number of potential 5hmC interacting proteins, most notably Mpg and Neil3, DNA glycosylases that, like TDG, may participate in the active DNA demethylation pathway to convert 5hmC to unmodified C via BER. Interestingly, they found proteins that had previously been uncharacterized, such as Wdr76. By purifying Wdr76 they identified Wdr76-interacting proteins, including a DNA helicase that is thought to

regulate DNA methylation levels (Hells, or Lsh) and a protein that binds H3K4me3 (OCR, or Spindlin-1). Looking at adult mouse brain cells, they confirmed the interaction of 5hmC with Wdr76 and Thy28. A brain-specific 5hmC interaction was found with Thap11 or Ronin, which is highly expressed in Purkinje cells. Additionally, they found that the proteins thymocyte nuclear protein 1 (Thy28), ubiquitin-like with PHD and Ring finger domains (Uhrf1), and methyl-CpG- binding protein 2 (MeCP2) bind to both 5mC and 5hmC, although MeCP2 binds to 5mC with much higher affinity. The authors conclude that 5hmC is an active intermediate in DNA demethylation and may be involved in global epigenetic regulation (Spruijt et al., 2013).

However, the binding of MeCP2 to 5hmC is a contentious finding. Previous *in vitro* studies found that conversion of 5mC to 5hmC abolished binding of MeCP2 to oligonucleotide sequences (Valinluck et al., 2004). Another study compared the affinity of Uhrf1 and MeCP2 to modified DNA *in vitro*, and found that Uhrf1 had a similar affinity for 5mC and 5hmC, while MeCP2 had a greater affinity for 5mC, as Spruijt et al. showed (Frauer et al., 2011; Spruijt et al., 2013). Similarly, yet another independent group found that although MeCP2 is able to bind 5hmC, the affinity is 20 fold less than that of 5mC (Khrapunov et al., 2014). On the other hand, Mellén et al. reported that MeCP2 binds 5mC and 5hmC with similar affinity and that a Rett-associated mutation in MeCP2 causes the disruption of its binding preferentially to 5hmC *in vitro* (Mellén et al., 2012a). Additionally, Baubec et al. found that MeCP2 localization correlates with 5hmC in embryonic stem cells (Baubec et al., 2013). Further confounding results come from two studies addressing the affects of MeCP2 on levels of 5hmC *in vivo* (Mellén et al., 2012a; Szulwach et al., 2011). Szulwach et al. showed that decreased levels of MeCP2 correlated with higher levels of 5hmC, and overexpression of MeCP2 revealed a decrease in 5hmC in the cerebellum (Szulwach et al., 2011). On the other hand, Mellén et al. reported that loss of MeCP2 results in a small, but significant, decrease in 5hmC levels (Mellén et al., 2012a). Together, these findings suggest that further research is warranted to determine whether MeCP2 is a *bona fide* binding partner of 5hmC *in vivo*.

The Role of 5hmC in brain development

Across brain development in both the human and mouse, 5hmC levels increase, with the adult human prefrontal cortex containing tenfold more 5hmC than the fetal

brain (Lister et al., 2013; Wen et al., 2014). In the adult (6 wk mouse), high levels of 5hmC are found across enhancer, transcriptional end sites, intragenic, and DNaseI-hypersensitive (DHS) regions. In the fetus, on the other hand, enrichment of 5hmC is primarily in DHS regions and enhancer regions that are unique to fetal development (Lister et al., 2013). Many of the gene bodies that gain 5hmC enrichment in the adult stage also have 5hmC present at the fetal stage, albeit at much lower levels. This implies that the cell type-specific increase in 5hmC observed over development occurs at intragenic regions that were partially established at the fetal stage. Although 5hmC levels of the fetal brain are much lower than those of the adult brain across all genomic features, there are numerous CpGs in which hydroxymethylation is higher in the fetal brain. These Fetal > Adult hydroxymethylated CpGs are enriched at enhancers (Lister et al., 2013; Wen et al., 2014). Developmentally downregulated genes have high levels of gene body 5hmC at the fetal stage, but not at the adult stage (Lister et al., 2013). Regions that have significantly higher hydroxymethylation in the fetal frontal cortex than the adult are dormant regions poised for demethylation and activation over development in both the mouse and human (Lister et al., 2013; Wen et al., 2014). Analysis of these developmentally-dependent, differentially hydroxymethylated regions in adult Tet2-/- mice revealed that these 5hmC-poised loci are dependent upon Tet2 activity (Lister et al., 2013). These findings suggest a key role of 5hmC in brain development that is conserved across mammals.

5hmC marks both developmentally dependent (P7, 6 wk, and 1 year mice) and brain region-specific (hippocampus compared to cerebellum) loci. When brain region-specific differentially hydroxymethylated loci were evaluated across development, it was found that 5hmC marks region-specific genes during early development (at P7 or earlier) to facilitate in region-specific transcriptional programs. These tissue-specific differentially hydroxymethylated regions are enriched for a 21-nucleotide motif that might be critical for regulating specific gene expression programs. 5hmC regulated regions across development within mouse cerebellum and hippocampus revealed that some loci are stable (5hmC acquired during P7 and maintained through 1 year of age), whereas others are dynamic (5hmC is not present at all time points). When repeat elements were assessed across development, it was found that, in both mouse hippocampus and cerebellum, 5hmC enrichment increases at short interspersed nuclear elements (SINEs) and long tandem repeats (LTRs) and decreases in long interspersed

nuclear elements (LINEs) and satellites. In P7 cerebellum, but not adult, 5-hmC is associated with the transcriptional start site (TSS) of genes with low expression. This can be explained by the fact that at this stage, a significant amount of progenitors cells are present, which is in line with the finding in mouse and human ES cells in which 5hmC is enriched at the TSS of repressed, but developmentally poised, genes. At developmentally activated genes, there is an increase in intragenic 5hmC, while at developmentally repressed genes, there is only a small decrease in 5hmC across the gene body (Szulwach et al., 2011). These findings highlight the importance of 5hmC in brain development.

5hmC changes associated with neuronal differentiation

In addition to developmentally-dependent changes in 5hmC, alterations in this epigenetic mark also occur across neuronal differentiation as assessed by comparing the genomic distribution of 5hmC in neural progenitor cells (NPCs) in the subventricular zone to that of maturing neurons of the cortical plate in E15.5 mice. In NPCs and maturing neurons alike, there is an absence of 5hmC at enhancers and enrichment at promoters and gene bodies. Over differentiation there is an increase in 5hmC (but not 5mC), which is primarily intragenic. As typically found, intragenic 5hmC correlate transcription in both progenitors and mature neurons, however this association is more pronounced in maturing neurons. Genes that gain intragenic 5hmC are associated with neuronal differentiation and axonogenesis. Genes with the highest increase in 5hmC over differentiation did not show an increase in demethylation, indicating the stability of this epigenetic mark. Concomitant with the increase in intragenic 5hmC during neuronal differentiation at activated genes, is a gain of H3K4me3 at the promoter and a loss of H3K27me3 in gene bodies and promoters. Tet2 or Tet3 may play a role in neurogenesis since knockdown of their expression via electroporation of shRNAs lead to defects in the progression of differentiation (Hahn et al., 2013).

Similar changes in 5hmC over neuronal differentiation also occurs in the olfactory sensory neuron (OSN) differentiation as there is an increase in 5hmC from the horizontal basal cell (HBC) stage to the globose cell basal cell (GBC) to mature olfactory sensory neurons (mOSNs). Similar to neuronal differentiation in the ventricular zone, the increase in 5hmC across OSN differentiation is primarily intragenic and associated with developmentally regulated genes. When Tet3 is overexpressed in mOSNs via a transgenic mouse model approach, genes with modest levels of 5hmC in the wild type

(WT) mice exhibit an increase in 5hmC levels, and subsequent increase in expression, whereas gene with high levels of 5hmC exhibit a loss of 5hmC, and subsequent decrease in expression. The downregulation of the most highly expressed mOSN-specific genes, such as olfactory receptors and guidance molecules, affects glomerular formation (Colquitt et al., 2013). These findings from two neurogenic regions of the brain support the role of 5hmC in neuronal differentiation.

The Role of 5hmC in aging and neurodegeneration

In light of the importance of 5hmC in neurodevelopment, researchers have investigated the role of 5hmC in aging and neurodegenerative diseases. The two studies conducted on aging (from 6 wk to 2 yr) have found that global 5hmC levels increase in the hippocampus as aging occurs (Chen, Dzitoyeva, & Manev, 2012; Chouliaras et al., 2012). Given this finding, more research is warranted to determine whether 5hmC also increases in other brain regions, as well as the specific loci with 5hmC enrichment across aging. While there are not an expansive number of studies on neurodegeneration, several correlations have been found with 5hmC for both Huntington's disease (HD) and Alzheimer's disease (AD). In a mouse model of HD, a genome-wide decrease in 5hmC levels was found, correlating with decreases in gene expression (Wang et al., 2013). The differentially hydroxymethylated regions contained genes involved in neuronal development and survival, which could have an important affect on the neurodegeneration seen in this disease (Wang et al., 2013). Additionally, studies in AD patients have seen a decrease in 5hmC in the hippocampus, and 5hmC levels were negatively correlated with amyloid plaque load, a common marker of AD pathogenesis (Chouliaras et al., 2013; Condliffe et al., 2014). However, another study looking at preclinical stages of AD saw an increase in 5hmC in hippocampal regions of the brain (Coppieters, Dieriks, Lill, Faull, Curtis, & Dragunow, 2013; Sun et al., 2013). While preliminary, these studies point to a potential role of 5hmC in neurodegeneration through deregulation of gene expression. Additionally, 5hmC may act as a biomarker for diagnosing and determining the stage of the disease. Further research is required to understand the role 5hmC plays in neurodegeneration.

The Role of Tet enzymes in brain function

The three Tet proteins, Tet1, Tet2, and Tet3, are members of the Fe(II)/α-ketoglutarate (α-KG)-dependent dioxygenase family of enzymes. It has been proposed that the enzymes use molecular oxygen to catalyze oxidative decarboxylation of α-KG, creating a highly reactive intermediate that converts 5mC to 5hmC (**Figure 2**). This proposed mechanism is based off of other Fe(II)/α-KG-dependent dioxygenase family proteins, since a structure of mammalian Tet enzymes has not been solved. Each of the Tets contain a conserved catalytic domain (double-stranded β-helix fold (DSβH)) that contains the metal-binding residues required for the oxidation reaction (**Figure 3**) (Kohli and Zhang, 2013). Additionally, a cysteine-rich (Cys-rich) domain is found in all Tet proteins upstream of the catalytic domain and is thought to be required for activity (Iyer et al., 2009; Tahiliani et al., 2009). Tet1 and Tet3 contain a CXXC domain near the N-terminal end of the protein, which is known to bind to CpG sites (Kohli and Zhang, 2013). While each of the human Tet enzymes only share ~18-24% sequence identity (UniProt Consortium, 2015), the catalytic and Cys-rich domains are highly conserved between the three enzymes (**Figure 4**). It is hypothesized that the remaining non-conserved portions of the protein may serve as regulatory domains and convey different functionality between the three Tets.

All three Tet family members, Tet1, Tet2, and Tet3, are expressed in the brain, with Tet3 having the highest expression, followed by Tet2, and Tet1 having much lower expression levels than the other two family member (Szwagierczak et al., 2010). Knockout, loss of function, and overexpression studies have revealed diverse functions of these enzymes, and the importance of 5hmC, in neuronal function.

Tet1

Currently, Tet1 is the most well-studied Tet family member in the brain, most likely due to the fact that Tet1 was the first enzyme discovered to convert 5mC to 5hmC (Tahiliani et al., 2009). Although Tet1 expression is markedly lower than Tet2 and Tet3 in the brain, various studies have demonstrated the importance of Tet1 in neuronal function.

Tet1 whole-body knockout (KO) mice are viable and fertile without apparent health deficits, albeit a smaller body weight and litter size than WT animals (Dawlaty et al., 2011; Rudenko et al., 2013; Zhang et al., 2013). Additionally, there are no obvious brain morphological or developmental brain abnormalities (Rudenko et al., 2013; Zhang et al., 2013). In agreement with the importance of Tet1 in the generation of 5hmC, there is a small, but significant reduction of 5hmC in the brains of Tet1 KO mice, but no change in 5mC. The fact that the change is small is likely due to presence of Tet2 and Tet3, the other members of the Tet family that are endogenous expressed much higher than Tet1 in the brain. There also are no apparent deficits in synaptic connectivity as measured by Synapsin I, a marker of synaptic abundance (Rudenko et al., 2013). This compensation effect is supported by the deficits and abnormalities observed in Tet1 and Tet2 double KO (DKO) mice. DKO die perinatally, although a small percentage survive without gross abnormalities. When compared to WT mice, DKO adult mice (2.5 mo) have reduced 5hmC levels (34%) and increased 5mC levels (~5%) in the cerebrum and cerebellum. Although these are appreciable changes in methylation, a large portion of 5hmC remains intact, suggesting that Tet3 plays a critical role in its maintenance (Dawlaty et al., 2013).

Behaviorally, adult Tet1 single KO mice (4 mo) do not show deficits in locomotion, anxiety, fear memory acquisition, or depression-related behaviors. Multiple groups have observed memory deficits, however, there is not a consensus as to the specific type of memory deficit. According to one group, Tet1 KO mice have impairments in short-term memory and spatial learning, but normal long-term memory, as assessed by Morris water maze (MWM) (Zhang et al., 2013). Another group reported normal short-term memory and spatial learning, but impaired spatial memory extinction, in the MWM and classical Pavlovian fear conditioning (Rudenko et al., 2013). When Tet1 is overexpressed in the CA1 region of the hippocampus, long-term memory was affected (fear conditioning), but not locomotion, anxiety, or short-term memory. This deficit in long-term memory formation was observed for both catalytically active and inactive forms of TET1, suggesting that TET1's role in memory formation is independent of its catalytic activity. Tet1 expression, but not Tet2, Tet3, or other proteins involved in the demethylation pathway, is significantly downregulated in the dorsal CA1 of mice after fear learning (Kaas et al., 2013). These findings support that Tet1 contributes to basal neuronal 5hmC levels that are potentially important for neuronal function. The

behavioral effects of Tet1 in the brain still warrant further investigation seeing as though there are confounding results.

At the cellular and molecular level, evidence suggests that Tet1 is important in neurogenesis and hippocampal function. When Tet1 KO mice were bred with Nestin-GFP transgenic mice, the number of GFP-positive cells in the subgranular zone (SGZ) in adult mice was dramatically reduced by 45% when compared to WT animals (Zhang et al., 2013). This is different from two other non-neurogenic brain regions examined, the cingulate cortex and hippocampus CA1 (Rudenko et al., 2013). The reduction in proliferation potential of neural progenitor cells (NPCs) is likely to underlie this deficit as evidenced by a reduction in neurospheres isolated from Tet1 KO mice, the decrease in bromodeoxyuridine (BrdU, which marks dividing cells) positive neurons in Tet1-dentate gyrus-knockdown in adult mice, and the 35% decrease in BrdU positive neurons in animals in which Tet1 is specifically deleted in neural progenitors at 2 mo of age. Examination of the gene expression and methylation changes in Tet1 KO mice revealed that the decreased expression of a cohort of genes involved in neurogenesis was associated with an increase in 5mC at their promoters, suggesting that Tet1 positively regulates adult neurogenesis through the oxidation of 5mC to 5hmC at these genes (Zhang et al., 2013).

Tet1 overexpression in the dentate gryus (DG) or cornus ammonis 1 (CA1) region of the hippocampus of mice results in a dramatic increase in 5hmC and decrease in 5mC, providing evidence that Tet1 *in vivo* oxidizes 5mC to 5hmC (Guo et al., 2011b; Kaas et al., 2013). The overexpression of Tet1 in the DG led to a significant decrease in methylation at promoter IX of *Bdnf* (*Bdnf IX*) and the brain specific promoter of *Fgf1* (*Fgf1B*), and a concomitant increase in the expression of these two genes, supporting the role of Tet1 in the demethylation pathway, and subsequent gene activation (Guo et al., 2011b). Tet1 overexpression in area CA1 or DG of the hippocampus leads to the increased expression of various activity dependent genes (*Fos*, *Arc*, *Egr1*, *Homer1*, and *Nf4a2*), as well as genes downstream of the Tet mediated oxidation (*Tdg, Apobec1, Smug1, and Mbd4)* (Kaas et al., 2013). In the DG, the increased expression of these genes is dependent upon the catalytic domain TET1, as evidence by the fact that only the expression of human TET1 catalytic domain, but not expression of the catalytically inactive version of TET1; however, in the CA1 region, either the catalytic active or inactive TET1 leads to increase in expression of majority of these genes. This implies that TET1 acts via region-

dependent mechanisms (Guo et al., 2011b; Kaas et al., 2013). Furthermore, Tet1 is required for neuronal activity-induced active DNA demethylation and gene expression since short-hairpin mediated knockdown of endogenous Tet1 in the DG completely abolished electroconvulsive stimulation (ECS)-induced demethylation of *Bdnf IX* and *Fgf1B* promoters. These *in vivo* findings are in agreement with *in vitro* work that has shown that Tet1 knockdown in hippocampal neuron leads the hypermethylation of promoter IV of *Bdnf* and subsequent decreased expression from this promoter (Yu et al., 2015). Given that demethylation at these promoters is similar completely abolished after ECS with knockdown of Apobec1, this suggests that Tet1 and Apobec1 work together through oxidation-deamination to achieve active demethylation in the adult mouse brain (Guo et al., 2011b).

Loss of Tet1 causes electrophysiological deficits in the hippocampus. Tet1 KO mice have normal basal synaptic transmission and intrinsic neuronal properties, as measured by paired-pulse facilitation and presynaptic excitability, respectively. However, long-term potentiation (LTP), assessed in the Schaffer collateral-CA1 pathway, is attenuated, and long-term depression (LTD) is amplified. These *in vivo* electrophysiological findings confirm what is found *in vitro*. Overexpression of the catalytically active form of Tet1 prevents TTX-induced scaling-up, and knockdown of Tet1 leads to synaptic scaling-down that is unaltered by bicuculline treatment (Yu et al., 2015). Further analysis *in vivo* has demonstrated that alterations in metabotropic glutamate receptor (mGluR)-dependent form of LTD is not affected, therefore suggesting that a deficit in NMDAR-dependent LTD. Neuronal activity-regulated genes, including *c-Fos*, *Egr2*, *Egr4*, *Arc,* and *Npas4,* are affected in Tet1 KO mice. Analysis of the *Npas4* promoter-exon 1 region confirmed a decrease of 5hmC and an increase in 5mC, which could explain the downregulation of this group of genes. After memory extinction in Tet1 KO mice (but not after fear memory acquisition), the *Npas4* and *c-Fos* genes exhibit a decrease in 5hmC and an increase in 5mC, concomitant with a decrease in mRNA and protein expression levels in both brain regions assessed, the cortex and hippocampus. Since Tet1/Tet2/Tet3 expression does not increase during either fear memory extinction or acquisition, the activity of these proteins change, rather than absolute levels (Rudenko et al., 2013).

This body of work on Tet1 function in the brain suggests that Tet1, although expressed at much lower levels in the mammalian brain than the other Tet family

members, plays an important role in maintaining 5hmC levels, and subsequent gene-expression levels, at basal and activity-induced conditions.

Tet2

Despite it's high level of expression, Tet2 is presently the least well-studied Tet family member in the brain, but the limited studies conducted thus far have demonstrated the importance of Tet2 in brain function. There are no reported brain abnormalities or dysfunction in Tet2 knockout mice (Ko et al., 2011; Li et al., 2011). However, when Tet2 is knocked down in hippocampal neurons *in vitro*, there is an increase in miniature glutamatergic excitatory postsynaptic currents (mEPSC) amplitudes compared to controls (Yu et al., 2015). This implies that neuronal function may be impaired in the absence of Tet2.

Tet2 is also thought to play a role in the demethylation of developmentally dependent genomic loci. With the use of Tet2 knockout mice, it was found that this member of the Tet family is responsible for the oxidation of large fraction (19.7%) of CpG genomic regions that gain hydroxymethylation status over development. On the other hand, CpG regions with higher 5hmC in the adult than fetal stage are largely unaffected in Tet2 knockout mice. Across development and aging (6 wk, 10 wk, and 22 mo) in Tet2 knockout mice, there are over fourfold more hypermethylated CpG regions (14,000 CpG regions in total) than hypomethylated region, suggesting that Tet2 plays a role in the demethylation over development and aging (Lister et al., 2013). Tet2 may also play a role in neurogenesis since the knockdown of Tet2 and Tet3 via electroporation of shRNAs into the cortex lead to defects in the progression of differentiation from the subventricular zone (Hahn et al., 2013). These Tet2 findings suggest that Tet2 plays an important role in regulating developmentally dependent, differentially hydroxymethylated regions.

Tet3

Various studies on Tet3 function in the brain have confirmed that this most highly expressed brain Tet family member is essential in regulating neuronal activity. When mice undergo extinction training, there is a significant increase in Tet3 mRNA in

the cortex. Tet3 knockdown via lentiviral plasmids in the infralimbic prefrontal cortex (ILPFC) have normal fear memory acquisition but impaired fear memory extinction. Furthermore, inhibiting NMDAR activity blocked the increase in Tet3 expression associated with fear memory extinction, suggesting that the rise in Tet3 occurs via an NMDAR-mediated pathway. Fear acquisition and fear extinction result in genome-wide differences of 5hmC at locations that contain CA or CT dinucleotide repeats instead of CpGs. Additionally, there is a reduction in 5hmC at intronic and intergenic sites and an increase in 5hmC enrichment at distal promoters, 5'-UTR, 3'-UTR, exonic sequences, and DNaseI-hypersensitive regions. Gene ontology analysis revealed that 16% genes enriched for 5hmC after extinction learning are involved in synaptic signaling. When one of these genes, gephyrin, was evaluated, it was found that there was enrichment for 5hmC, co-occurring with decrease in 5mC, within one intron. Moreover, in response to extinction, there was an increase Tet3 occupancy at the gephyrin gene, as well as an increase in specificity protein 1 (*Sp1*), a transcription factor that activates gene expression by preventing the active loci from becoming methylated. The observed reduction in transient H3K9me3 and increase in H3K27ac, p300, H3K4me1, and dimethyl H3 arginine 2 (H3R2me2s), which is crucial for maintaining a euchromatic state, all support the role of Tet3 in extinction-induced gene expression changes. These changes are specific to fear extinction, and do not occur during fear acquisition. All of these changes at the gephyrin gene are blocked with the use of a Tet3 shRNA (Li et al., 2014).

Tet3 expression levels correlate with neuronal activity *in vitro* as well; an increase in synaptic transmission correlates with an increase in Tet3, but not Tet1 or Tet2. When Tet3 is knocked down from hippocampal neurons in culture, mEPSC amplitudes are significantly larger than controls, and the reciprocal effect occurs when Tet3 is overexpressed. Notably, knockdown of either Tet1 or Tet2 also increases mEPSC amplitudes, but not as drastically as Tet3 knockdown. Tet3 is also essential for the maintenance of homeostatic synaptic plasticity since knockdown of Tet3 leads to synaptic scaling-up that is unaltered by tetrodotoxin (TTX) or retinoic acid (RA) treatment; knockdown of Tet3 leads to synaptic scaling-down that occludes further alterations with bicuculline treatment; and Tet3 overexpression prevents TTX-induced synaptic scaling-up or bicuculline-induced scaling-down. Given that a similar effect on mEPSC amplitudes and synaptic scaling occurs when poly (ADP-ribose) polymerase

(PARP) or the apurinic/apyrimidinic endonuclease is inhibited, two major components of the BER pathway, these data suggest that excitatory synaptic transmission is regulated by the oxidation of DNA through Tet, followed by BER (Yu et al., 2015).

The molecular mechanism through which Tet3 causes these effects is likely through the regulation of surface GluR1 receptors. Knockdown of Tet3 leads to an increase in surface GluR1 receptors that is resistant to a further increase or decrease in surface GluR1 receptors, which occurs in control neurons treated with TTX or bicuculline. When gene expression changes were assessed in Tet3 knockdown neurons, GO term enrichment revealed expression changes of genes involved in synapse and synaptic transmission. Genes with differential expression due to TTX or bicuculline treatment in control neurons lost responsiveness in Tet3 knockdown neurons. A very similar effect was observed for bicuculline in Tet3 knockdown neurons. In Tet3 knockdown neurons, promoter IV of the brain-derived neurotrophic factor (*Bdnf*) is hypermethylated, and there is a decrease in expression from this promoter. The bicuculline-induced hypomethylation, as well as the TTX-induced hypermethylation, of *Bdnf* promoter IV are occluded in Tet3 knockdown neurons. ChIP-PCR revealed that Tet3 binds to the *Bdnf* promoter IV (Yu et al., 2015).

The importance of Tet3 in neural function is conserved across vertebrates, as knockdown of Tet3 in *Xenopus* by morpholino antisense oligonucleotide leads to marked developmental abnormalities, including malformation of the eye, small head, and early death. At the molecular level, Tet3 depletion causes a reduction in expression of master eye developmental genes (*pax6*, *rx*, and *six3*), primary neuronal markers (*ngn2* and *tubb2b*), neural crest markers (*sox9* and *snail*), and major shh signaling components (*shh* and *ptc-1*). Additionally, Tet3 chromatin immunoprecipitation (ChIP) assays confirm the binding of Tet3 to the promoters of *pax6*, *rx*, *six3*, *ptc-1*, *ptc-2*, *sox9*, and *ngn2*. Furthermore, at the promoters of some of these target genes, there was found to be a developmental increase in 5hmC and decrease in 5mC from stage 10 to 19 in *Xenopus* development, which is perturbed when Tet3 is knockdown. These findings suggest that Tet3 acts as an upstream activator of key neural developmental genes (Xu et al., 2012). Furthermore, these studies suggest that Tet3 plays an important role in brain function that is conserved across animals.

Conclusion

Despite it's recent rediscovery in 2009, major advances have been made regarding our understanding of the DNA modification 5hmC. 5hmC is now known to be an intermediate of active DNA demethylation in neurons. However, the biological role of 5hmC in the brain is still up for debate. It is known that even within defined brain regions, such as the cerebellum, there are a multitude of cell types that have vastly different functions, gene expression patterns, and 5mC and 5hmC levels. While studies have shown 5hmC distribution in different brain regions throughout development, it is still not understood what happens to 5hmC in specific cell types as an animal ages. Investigating 5hmC distribution in a heterogeneous brain region is likely to mask important cell-type specific features and reveal false patterns due to the multitude of cell types in the sample. Looking at 5hmC and Tet enzymes in specific cell types is crucial for our understanding of how DNA demethylation and 5hmC affect neural functions.

In the search for the biological role of 5hmC, it is also imperative to know what proteins bind to the modified base, and the cellular consequences of the interactions. While a variety of putative 5hmC binding partners have been identified, the potential interactions have not been confirmed *in vivo*, and the functions of 5hmC-protein interactions have not been established. Studies addressing the binding partners of 5hmC will greatly advance our understanding of 5hmC not just as a demethylation intermediate, but also as an essential epigenetic mark.

Figures

Figure 1

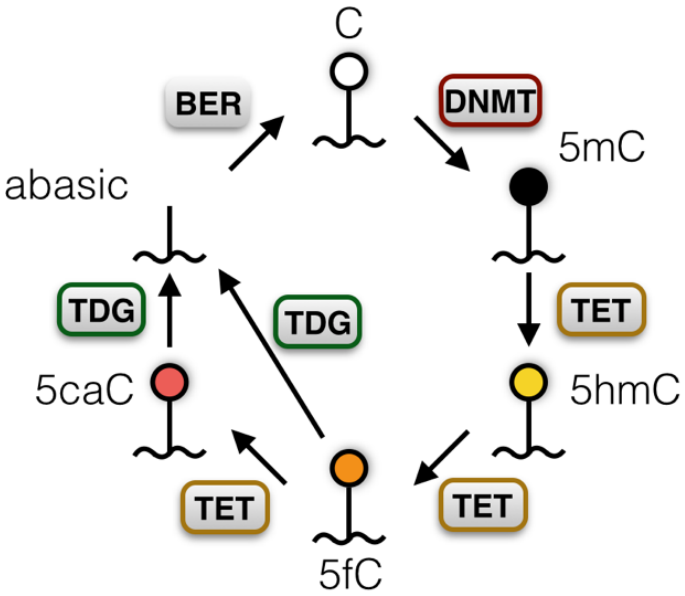


Figure 1. Cytosine modification cycle. Unmodified cytosine (C) is converted to 5-methylcytosine (5mC) by DNA methyltransferase enzymes (DNMTs) DNMT1, DNMT3a, or DNMT3b. 5mC can then be iteratively oxidized by Tet enzymes Tet1, Tet2, or Tet3, to become 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Thymine DNA glycosylase (TDG) can recognize the bases 5fC and 5caC and excise them from the DNA, leaving an abasic site. An abasic site triggers the base excision repair pathway (BER) which restores the base to cytosine.

Figure 2

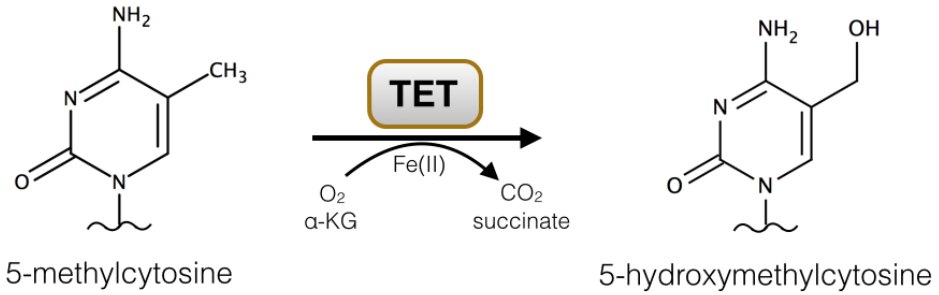


Figure 2. 5mC to 5hmC conversion by Tet enzymes. Tet enzymes oxidize the 5-methyl group of 5mC. With cofactors alpha-ketoglutarate (α-KG) and molecular oxygen, Tet oxidizes the 5-methyl carbon, adding a hydroxyl group, yielding 5-hydroxymethylcytosine (5hmC). Other byproducts of the enzymatic reaction include carbon dioxide and succinate.

Figure 3

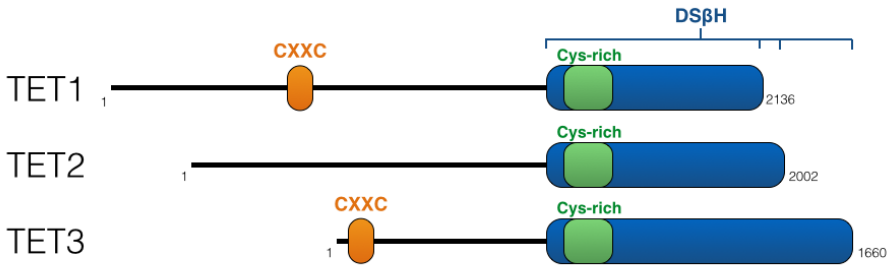


Figure 3. Schematic of Tet enzymes. Tet1, Tet2, and Tet3 all share a conserved catalytic double-stranded β-helix fold (DSβH) domain at the C-terminal end of the protein. Additionally, a conserved cysteine-rich (Cys-rich) domain is found at the N-terminal portion of the catalytic domain. Tet1 and Tet3 contain an additional CpG-binding CXXC domain at the N-terminal end of the protein.

Figure 4

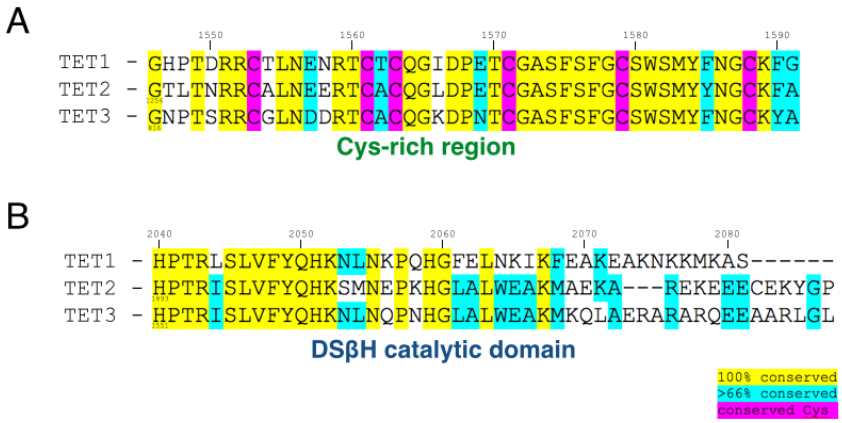


Figure 4. Sequence conservation of Tet domains. (*a*) The cysteine-rich (Cys-rich) domain is almost fully conserved between the Tet enzymes. (*b*) The catalytic double-stranded β-helix fold (DSβH) domain is also highly conserved between the three Tet enzymes, even though the proteins only share 18-24% sequence identity. The remaining non-conserved domains are hypothesized to serve as regulatory domains.

APPENDIX PART 3

Locus- and cell type-specific epigenetic switching during cellular differentiation in mammals

One Sentence Summary

This story, which found that the loss of DNA methylation is accompanied by the gain of different histone modifications in a locus- and cell type-specific manner over cellular differentiation, is published in *Frontiers in Biology*.

Abstract

BACKGROUND: Epigenomic reconfiguration, including changes in DNA methylation and histone modifications, is crucial for the differentiation of embryonic stem cells (ESCs) into somatic cells. However, the extent to which the epigenome is reconfigured and the interplay between components of the epigenome during cellular differentiation remain poorly defined. **METHODS:** We systematically analyzed and compared DNA methylation, various histone modification, and transcriptome profiles in ESCs with those of two distinct types of somatic cells from human and mouse. **RESULTS:** We found that global DNA methylation levels are lower in somatic cells compared to ESCs in both species. We also found that 80% of regions with histone modification occupancy differ between human ESCs and the two human somatic cell types. Approximately 70% of the reconfigurations in DNA methylation and histone modifications are locus- and cell type-specific. Intriguingly, the loss of DNA methylation is accompanied by the gain of different histone modifications in a locus- and cell type-specific manner. Further examination of transcriptional changes associated with epigenetic reconfiguration at promoter regions revealed an epigenetic switching for gene regulation—a transition from stable gene silencing mediated by DNA methylation in ESCs to flexible gene repression

facilitated by repressive histone modifications in somatic cells. **CONCLUSIONS:** Our findings demonstrate that the epigenome is reconfigured in a locus- and cell type-specific manner and epigenetic switching is common during cellular differentiation in both human and mouse.

Introduction

Cellular differentiation is a precisely regulated process by which hundreds of morphologically and functionally distinct cell lineages are derived from the same origin, embryonic stem cells (ESCs). Given that ESCs and somatic cell types share an identical genome, a fundamental question is how lineage-specific gene expression patterns are established and maintained. Accumulating evidence supports that two epigenetic mechanisms, DNA methylation and histone modifications, are crucial in this process (Reik, 2007).

DNA methylation at the fifth carbon position of cytosine is an abundant covalent modification that primarily occurs in CpG dinucleotides (mCpGs), with 80% of CpGs being methylated in mammals. DNA methylation affects DNA-protein interactions and therefore influences transcription in various physiological processes, such as imprinting, X-inactivation, tumorigenesis, and embryogenesis (Jaenisch and Bird, 2003). In cellular differentiation, DNA methylation plays a critical role as the loss of enzymes responsible for establishing or removing methylation in embryos or ESCs results in skewed lineage specification and transdifferentiation (Jackson et al., 2004; Koh et al., 2011; Tsumura et al., 2006). Similar to DNA methylation, covalent modifications on histone tails play an imperative role in cellular differentiation and modulate transcription by altering DNA accessibility. Altering the activity of histone modifying enzymes, such as methyltransferases and acetyltransferases, diminishes the differentiation potential of ESCs (Margueron and Reinberg, 2011). Therefore, although histone modifications and DNA methylation are required for successful differentiation, the precise function of each in determining cell fate decisions is unknown.

Investigating the role of these epigenetic mechanisms has recently become possible with the construction of single base-pair resolution profiles of DNA methylation

(methylomes) and genome-wide histone modification profiles in multiple cell types, tissues, and organisms (Lister and Ecker, 2009; Rivera and Ren, 2013). Comparison of methylomes across cellular differentiation has revealed both *de novo* methylation (Reik, 2007; Smith and Meissner, 2013) and *de novo* demethylation (Hon et al., 2012; Lister et al., 2009; Xie et al., 2013; Ziller et al., 2013). These findings raise the possibility that mCpG is more dynamic than previously thought. Indeed, literature supports that mCpG undergoes a demethylation process mediated by the TET family of proteins that oxidize mCpG to hydroxymethyl-CpG (Tahiliani et al., 2009). The acquisition of methylomes and histone modification profiles from multiple cell types also provides an opportunity to examine the interplay between these two epigenetic mechanisms. Increasing evidence supports that the loss of DNA methylation is coupled to the gain of histone modifications, such as H3K27me3 and H3K9me3 in tumorigenesis (Hon et al., 2012); H3K27me3, H3K9me3, and H3K4me1 in differentiation (Gifford et al., 2013; Hawkins et al., 2010); and H3K27ac and H3K4me1 in development (Lister et al., 2013). However, the interplay between these two epigenetic mechanisms has only been observed for a limited number of histone modifications, leaving the large scope of DNA-histone modification interactions to be explored.

In this study, we systematically integrated DNA CpG methylation, various histone modification, and transcriptome profiles in ESCs and two somatic cell types in both human and mouse. With the use of our novel unbiased, enrichment based statistical approach, we found that both DNA methylation and histone modifications are considerably altered during cellular differentiation. Approximately 70% of the reconfigurations in DNA methylation and histone modifications are cell type-dependent and locus-specific. Additionally, integrative analyses of these epigenomic reconfigurations revealed a widespread locus-specific switching from DNA methylation to histone modifications during cellular differentiation in both species. Furthermore, when these epigenetic changes at promoter regions were correlated with gene transcription, we found that the switching at promoter regions implicates a transition from stable gene silencing mediated by DNA methylation to flexible gene repression facilitated by histone modifications. Our study provides a comprehensive insight into the role of epigenomic reconfiguration in cellular differentiation.

Materials and methods

Bioinformatics analyses

All analyses in this study were processed in the MAC terminal window. Perl was used to do in-house programming. R (http://www.r-project.org/) and Bioconductor libraries (http://www.bioconductor.org) were used to do all statistical analyses. Integrative Genomics Viewer (IGV, version 2.2.4) (Robinson et al., 2011; Thorvaldsdottir et al., 2013) was used as the local genome browser in this study. The first replicate of each histone modification (ChIP-seq) and gene expression (mRNA-seq) data set was used in the browser representation for all the figures (Table S1). Samtools version 0.1.18 (Li et al., 2009) was used to convert sam files to bam files and to sort the bam files by name. Sorted bam files were converted to tdf files by igvtools version 2.2.1 using parameters "-z 5 -w 25 -e 250" for ChIP-seq files and "-z 5 -w 25 -e 250 --strands read" for strand-specific mRNA-seq files. The track scales for ChIP-seq and mRNA-seq were normalized to the total number of uniquely mapped reads.
ChIP-seq_scale = (TDF_scale / Total_number_uniquely_mapped_reads) * 10,000,000
mRNA-seq_scale = (TDF_scale / Total_number_uniquely_mapped_reads) * 20,000,000

Obtaining the raw data and FASTQ files

The SRA raw data files for whole-genome bisulfite sequencing (WGBS), ChIP-seq, and mRNA-seq were downloaded from NCBI Sequence Read Archive database (http://www.ncbi.nlm.nih.gov/Traces/sra) under the accession numbers listed in Table S1. The SRA Toolkit version 2.1.7 (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view =software) was used to obtain the FASTQ files from the SRA files. The command used was "fastq-dump".

Genomic sequences and annotations

The reference genomic sequences used in this study were hg19 for the human and mm9 for the mouse. The genomic annotations, human GRCh37 and mouse NCBIM37 that were based on Ensembl release 66, were downloaded from Illumina iGenome database (http://cufflinks.cbcb.umd.edu/igenomes.html). The distal 5′ end of a gene was defined as the transcription start site (TSS). The exonic region was defined as all exons of a gene. The region from the 500 bp downstream of TSS to the distal 3′ end of a gene was defined as the gene body. Intergenic regions were downloaded from UCSC Genome Browser using Table Browser function (http://genome.ucsc.edu/cgi-bin/hgTables?command=start). Intergenic regions with at least 1 kb in length were included in subsequent analyses.

WGBS data mapping and methylation calling

The FASTQ raw data files were first processed by Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to remove adaptor contamination, to trim lower-quality sequences, and to discard sequences shorter than 36 nucleotides. The Trim Galore parameters used were "-s 2 -length 36". Bismark v0.7.4 (Krueger and Andrews, 2011) was used to map the WGBS data to the hg19 and the mm9 genomes. We used "-n 2 -l 40" and default parameters in Bismark. We excluded reads that had three or more methylated cytosines in non-CpG contexts or had cytosine in the reads but thymine in the corresponding location of the reference genomes. For mapped reads from a library of the same PCR reaction with identical 5′ sites, the one with highest average phred quality score was kept and the others were discarded to get rid of possible PCR amplification artifacts. Libraries from multiple PCR reactions and from different replicates for each cell line were merged together. SNPs between C57BL/6 and 129 were downloaded from the Mouse Genome Project of the Sanger Institute (ftp://ftp-mouse.sanger.ac.uk/REL-1003/SNPs/20100310-all-snps.tab.gz). For the mouse methylomes, CpGs that overlapped with known SNPs were excluded from subsequent analyses. The methylation status of both strands of a CpG pair were merged together.

CpGs with 10 or more read coverage were included in subsequent analyses. The methylation level for each CpG was calculated as follows:

$$mCpG = \frac{R_{C+} + R_{C-}}{(R_{C+} + R_{T+}) + (R_{C-} + R_{T-})}$$

where

R_{C+} is the number of reads with C mapped to the sense strand of a CpG,

R_{T+} is the number of reads with T mapped to the sense strand of a CpG,

R_{C-} is the number of reads with C mapped to the antisense strand of a CpG,

R_{T-} is the number of reads with T mapped to the antisense strand of a CpG.

Genome-wide and locus-specific comparisons of DNA methylation levels

For genome-wide analysis, the two-tailed Fisher's exact test was used to compare the methylomes between two cell lines at single-CpG resolution. The p-value for each CpG was calculated as follows:

p-value

$$= \frac{\binom{C_{ESC}+C_{Somatic}}{C_{ESC}}\binom{T_{ESC}+T_{Somatic}}{T_{ESC}}}{\binom{C_{ESC}+T_{ESC}+C_{Somatic}+T_{Somatic}}{C_{ESC}+T_{ESC}}} = \frac{(C_{ESC} + C_{Somatic})!(T_{ESC} + T_{Somatic})!(C_{ESC} + T_{ESC})!(C_{Somatic} + T_{Somatic})!}{C_{ESC}!C_{Somatic}!T_{ESC}!T_{Somatic}!(C_{ESC} + T_{ESC} + C_{Somatic} + T_{Somatic})!}$$

where

$\binom{n}{k}$ is the binomial coefficient,

! is the factorial operator,

C_{ESC} is the number of methylated reads that covered this CpG in ESCs,

T_{ESC} is the number of non-methylated reads that covered this CpG in ESCs,

$C_{Somatic}$ is the number of methylated reads that covered this CpG in somatic cells,

$T_{Somatic}$ is the number of non-methylated reads that covered this CpG in somatic cells.

Raw p-values were adjusted using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995) to control the False Discovery Rate (FDR). CpGs with an adjusted p-value (FDR) less than 0.05 were defined as differentially methylated CpGs.

For locus-specific analysis, we developed an enrichment-based statistical approach to compare and identify differentially methylated genomic regions (Figure 2A). We defined promoter regions using stringent criteria (Figure S2A) and analyzed them in parallel.

To determine the distribution of CpGs and mCpGs at promoter regions (Figure 5A), the region 5 kb upstream and 1 kb downstream of TSS for each gene was dissected into 600 bins with 10 bp in length. For each bin, we calculated the total number of genomic CpGs, the ratio between the total number of mCpGs versus the total number of genomic CpGs, the ratio between the total number of diff-mCpGs and the total number of mCpGs, respectively.

ChIP-seq data mapping, peak calling, and sliding-window analysis

Bowtie version 0.12.7 (http://bowtie-bio.sourceforge.net/index.shtml) (Langmead et al., 2009) was used to map the FASTQ files of ChIP-seq data to the genomes. "-v 2 -m 1 -p 10 " and default parameters were used. The mapping results for each replicate were merged together.

FindPeaks in Homer version 4.2 (http://biowhat.ucsd.edu/homer/ngs/) (Heinz et al., 2010) was used for peak calling using "-style histone" and default parameters. If the region 1 kb upstream and 1 kb downstream of the TSS for a gene overlapped with H3K27me3 or H3K9me3 peaks, we defined this gene as marked by that histone modification.

All replicates for each histone modification in each cell type were merged together. The total number of uniquely mapped reads was normalized to 40 million. For sliding window analysis, the human and mouse genomes were partitioned into 1 kb sized windows, and the number of normalized reads for each window was calculated. The fold enrichment score for each histone modification at each window between ESCs and somatic cells was calculated as follows:

$$fold\ enrichment = \frac{R_{Somatic} + 1}{R_{ESC} + 1}$$

where $R_{Somatic}$ is the number of normalized reads in this window for somatic cells and R_{ESC} is the number of normalized reads in this window for ESCs.

Based on the fold enrichment score for each window, we classified all windows into three categories: those in which histone modifications are gained, lost, or remained. The gain windows were defined as those that have higher read numbers in somatic cells than in ESCs (fold enrichment score > 1), whereas the lose windows were defined as those that have lower read numbers in somatic cells than in ESCs (fold enrichment score < 1). We then obtained the top 2000 gain windows and the top 2000 lose windows based on their fold enrichment scores, and termed them as the highly dynamic regions. Only those dynamic regions with greater than or equal to 2 fold enrichment or depletion were included in subsequent analyses. The DNA methylation change associated with each highly dynamic region was calculated as follows:

$$methylation\ change = \frac{\sum_{i=1}^{n} S_i - \sum_{i=1}^{n} E_i}{n}$$

where *n* is the total number of CpGs with 10 or more read coverage in both cells in this region, S_i is the methylation level of the number *i* CpG in somatic cells, and E_i is the methylation level of the number *i* CpG in ESCs.

mRNA-seq data mapping and differential expression analysis

The mRNA-seq data for mouse frontal cortex were mapped to mouse NCBIM37/mm9 genomes by TopHat version 2.0.6 (Trapnell et al., 2009) using parameters "--bowtie1 --color --quals -p 5 -g 1 --no-coverage-search --no-novel-juncs --library-type fr-secondstrand -G UCSC_mm9.gtf mm9_c". The other mRNA-seq data in FASTQ files were mapped to human GRCh37/hg19 or mouse NCBIM37/mm9 genomes by STAR for MAC version 2.2.1d (Dobin et al., 2013). We used STAR parameters "--

outFilterMultimapNmax 1 --outFilterMismatchNmax 3" to only report uniquely mapped reads with no more than three mismatches.

The number of mapped reads in the exonic region of each gene was calculated using an in-house Perl program. All replicates for each cell type were calculated separately. Reads that overlapped with the exonic regions of multiple genes were excluded. The edgeR Bioconductor package (Robinson et al., 2010) was used to perform the differential expression analysis. The classic edgeR method, which is negative binomial model based exact test, was used. For differential expression analysis, we only included genes with a count per million (CPM) larger than 1 in at least two of the data sets. Genes with adjusted p-values less than 0.05 were defined as differentially expressed. Gene Ontology enrichment analysis was performed using DAVID v6.7 (http://david.abcc.ncifcrf.gov/) (Huang da et al., 2009a, b).

Results

Global DNA hypomethylation in somatic cells

Given the essential role of DNA methylation in cellular differentiation, we first examined global DNA methylation in terminally differentiated somatic cells compared to ESCs using raw data from the NCBI GEO database (Hon et al., 2012; Lister et al., 2009; Mann et al., 2013; Stadler et al., 2011; Xie et al., 2012). Thus, we compared the methylomes from human H1 ESCs (hESC) and two distinct types of human somatic cells, IMR90 fetal lung fibroblasts (hFB) and mammary epithelial cells (hEP)(Figure S1A). To evaluate DNA methylation changes in a different species, we also compared the methylomes of mouse ESCs (mESC), mouse primary dermal fibroblasts (mFB), and mouse frontal cortex tissue (mFC)(Figure S1A).

We found that global mCpG levels are significantly lower in somatic cells compared to ESCs (Figure 1). In human, 93% of the differentially methylated CpG sites (diff-mCpGs) exhibit a decrease in methylation in hFB compared to hESC and 89% of diff-mCpGs show a decrease in hEP (Figure 1A-B, two-tailed Fisher's exact test,

Benjamini and Hochberg (B-H) correction). Global DNA hypomethylation was also found in mouse somatic cells (Figure 1A-B). Moreover, we found that hypomethylated diff-mCpGs are evenly distributed across all autosomes in both species, indicating widespread genome-wide DNA hypomethylation in somatic cells compared to ESCs (Figure S1C). Of note, global mCpG levels are not significantly different across somatic cell types (Figure S1B), suggesting that hypomethylation is found upon lineage commitment, rather than across cell types.

DNA hypomethylation is locus- and cell type-specific

Next, we investigated whether changes in DNA methylation occur at particular genomic features: intergenic regions, gene body, and promoter regions. To do so, we developed an enrichment based statistical approach for identifying genomic regions in which the ratio of diff-mCpGs is statistically higher than the ratio expected by chance (Figure 2A). Thus, we were able to identify genomic regions that are differentially methylated between different cell types.

Consistent with our finding of global hypomethylation, a significantly larger portion of the differentially methylated regions is hypomethylated in somatic cells compared to ESCs in both human and mouse (Figure S2A-C, FDR<0.0001, Hypergeometric test, B-H correction). For example, >90% of differentially methylated promoter, gene body, and intergenic regions are hypomethylated in hFB compared to hESC (Figure S2A-C). Notably, approximately 60-80% of the identified differentially methylated genomic regions are unique to that specific somatic cell type regardless of being in the promoter, gene body, or intergenic region (Figure 2B; Figure S2D). Genomic feature-specific alterations of DNA methylation are illustrated in the promoter regions of cell type-specific genes, including *LHX8*, *SH3TC1, COL6A3,* and *FOXI2* (Figure 2C). Only 30% of the genomic regions identified in hFB and hEP share the same changes in DNA methylation when compared to hESC (Figure 2B), such as the promoter regions of *TMEM173* and *TRIM4* (Figure 2C). Furthermore, we found that the changes in DNA methylation are strictly confined in a locus-specific manner, as surrounding regions do not show similar alterations (Figure 2C; Figure S2E-F). Together, these results

demonstrate that the decrease in DNA methylation across cellular differentiation is lineage-specific and restricted to particular loci.

Locus- and cell type-specific alterations of histone modifications in human

Modification of histone tails, such as methylation, acetylation, ubiquitination or phosphorylation, is an evolutionarily conserved epigenetic mechanism crucial for gene regulation in many biological processes (Berger, 2007; Jenuwein and Allis, 2001). Characteristic histone modification features include H3K9me3 and H3K27me3 marks in transcriptionally silent loci and H3K4me3 and H3K27ac in transcriptionally active promoters (Turner, 2007). Moreover, genes that show transient transcriptional changes in response to environmental stimuli versus genes that are persistently transcribed are marked by distinct types of histone modifications (Metivier et al., 2003). Given the importance of histone modifications in regulating cellular processes, we next analyzed the extent to which histone modifications are altered in somatic cells compared to ESCs. We acquired all publicly available ChIP-seq data sets for histone modifications in hESC (26 data sets in total), hFB (26 data sets), and hEP (11 data sets) and compared the enrichment profiles for each modification across cell types.

We found that 70-80% of the enriched sites for each of the 26 histone modifications are altered in hFB compared to hESC and 60-90% of the 11 histone modifications are altered between hEP and hESC, indicating a widespread epigenetic reorganization in human somatic cells (Figure 3A). Comparison of histone modification profiles between the two somatic cell types revealed that alterations of histone modifications are predominately cell type-specific (Figure 3B). For example, 70% of the identified H2A.Z sites that change during cellular differentiation are unique to each somatic cell type (Figure 3B). Consistent with previous findings (Hawkins et al., 2010), we also found that repressive histone modifications, such as H3K27me3 and H3K9me3, are markedly enriched, including an increase in the total number of sites (Figure 3A) as well as a significant expansion of occupancy for most of the enriched sites across cellular differentiation (Figure 3 C-D, $p<2.2\times10^{-16}$, Student's t-test). These results show that

regions of histone modification occupancy change dramatically in a cell type-specific manner across cellular differentiation.

Epigenetic switching from DNA methylation to histone modifications

We next examined whether changes in histone modifications are associated with changes in DNA methylation across differentiation. To this end, we focused on genomic regions in somatic cells that exhibited the most dynamic changes in histone modifications relative to hESC. For each histone modification, these regions were defined as those with an enrichment score ≥ 2, and we limited our analyses to the top 4,000 dynamic regions (2,000 gain and 2,000 loss).

We found that the majority of regions with increased histone modifications in somatic cells also exhibited decreased DNA methylation levels compared to hESC (Green boxes in Figure 4A-B). This correlation occurred for 85% of the histone modifications in hFB and 73% of modifications in hEP (Green boxes in Figure 4A-B). Moreover, we observed that many of the identified epigenetic switching from DNA methylation to histone modifications during cellular differentiation are specific to each cell type (Gray regions in Figure 4D; Gray regions in Figure S3C). For example, two sites in the *RIN2* gene locus with specific loss of DNA methylation in hEP, but not in hFB, are accompanied by an increase in histone modifications in hEP, but not in hFB (Gray regions in Figure 4D). Similarly, when we analyzed the relationship between DNA methylation and histone modifications in mouse mESC and mFC, we found that gains of H3K4me3 and H3K27ac are associated with a decrease in DNA methylation, while the loss of H3K4me3 and H3K27ac are coupled with an increase in DNA methylation (Figure 4C). This epigenetic switching is not found between regions exhibiting decreased histone modifications and DNA methylation for the majority of regions (Figure S3A-B). Together, these data demonstrate a cell type-dependent and locus-specific switching from DNA methylation to histone modifications over cellular differentiation in both species.

Epigenetic switching at promoters is associated with distinct gene regulation

We next set out to determine how switching from DNA methylation to histone modifications during cellular differentiation affects gene expression. We focused on promoter regions since changes in DNA methylation and histone modifications at promoter regions are correlated with changes in transcription relative to other genomic features (Ball et al., 2009). We found that proximal promoters, regions 0.5kb upstream and downstream of transcriptional start sites (TSSs), have a higher CpG density and are more dynamically methylated than distal promoter regions, defined as regions from 0.5-5kb upstream and 0.5-1kb downstream of TSSs (Figure 5A; Figure S4A). Given this finding, we identified differentially methylated proximal promoters over differentiation with the use of our enrichment-based statistical approach (Figure 2A). Compared to ESCs, we found that the majority of differentially methylated promoters were hypomethylated: 88% of the promoters in hFB, 82% of the promoters in hEP, 77% of the promoters in mFB, and 65% of the promoters in mFC. Therefore, we focused on hypomethylated promoters for subsequent analyses. Interestingly, gene ontology (GO) analysis of genes with hypomethylated proximal promoters revealed significant enrichment in processes characteristic of each somatic cell type as well as pathways responsive to environmental signals (Figure S4B, Fisher′s exact test, B-H correction). For example, many of the significant GO terms for hFB were related to cell adhesion, which is an essential characteristic of hFB, and injury response, a pathway highly dependent on external cues (Figure S4B). These findings suggest that hypomethylation at promoter regions in somatic cells is important for defining cellular identity and rendering them responsive to environmental cues.

Examination of transcriptional changes of genes with hypomethylated promoters revealed that 25% of genes showed an increase in expression in hFB compared to hESC (Red regions in Figure 5B). This is consistent with the prevailing model that a decrease in DNA methylation at promoters correlates with an increase in gene expression (Ball et al., 2009). However, 65% of genes identified in hFB did not exhibit changes in transcription despite a significant decrease in DNA methylation at their promoters (Black regions in Figure 5B, FDR<0.0001, Hypergeometric test, B-H correction), with 78% of these genes are transcriptionally silent in both hESC and hFB (RPKM<1). Alternative promoter usage (Maniatis and Reed, 2002) may account for some of the

genes in which promoter hypomethylation is associated with a decrease in gene expression in hFB (Green regions in Figure 5B; Figure S5A). Similar percentages of genes with hypomethylated promoters show an increase (34%), no change (50%), or decrease (16%) in gene expression in hEP compared to hESC (Figure 5B). In mice, promoter hypomethylation had similar effects on gene expression with the majority of genes with hypomethylated promoters either showing an increase in gene expression, 47% in mFB and 67% in mFC, or no change in gene expression, 46% in mFB and 26% in mFC, while only a small percentage exhibited a decrease in gene expression, 7% in mFB and 7% in mFC)(Figure 5B). These data suggest that promoter hypomethylation during cellular differentiation is not only important for increasing lineage specific transcripts, but also for flexible gene repression facilitated by repressive histone modifications in somatic cells.

To gain an insight into the mechanism by which genes with hypomethylated promoters remain transcriptionally silent in human somatic cells, we examined the profiles of two repressive histone modifications, H3K27me3 and H3K9me3, at the promoter regions of these silent genes in hFB and hEP. We found that, 77% of these genes in hFB and 45% in hEP are marked by one or both of H3K27me3 and H3K9me3 at their promoter regions (Figure 5C-D; Figure S5B). Furthermore, these repressive histone modifications at the promoter regions are either established *de novo* during cellular differentiation, such as at the promoter regions of *KBTBD5*, *ZPBP2*, and *VMO1* (Figure 5D), or inherited from hESC, such as at the promoter region of *LHX8* (Figure 5D). Given that DNA methylation is known to be a stable repressive mechanism (Bird, 2002; Jones, 2012), the epigenetic switching at the promoter regions of these genes during cellular differentiation may allow them to be dynamically expressed in response to external cues in somatic cells (Figure 5E). Together, these results uncover a potential role for epigenomic reconfiguration in facilitating gene regulation, by which genes stably silenced by DNA methylation in ESCs switch to flexible repression mediated by repressive histone marks in somatic cells.

Discussion

Two longstanding questions in cellular differentiation are how ESCs maintain pluripotency and how they give rise to a multitude of diverse types of somatic cells.

Given that ESCs and somatic cells share identical genomic sequences, the establishment and maintenance of each somatic cell identity has been attributed to epigenetic modifications. This has led to an appealing model in which a flexible epigenome in ESCs allows for a multitude of genes to be transiently activated or repressed to support pluripotency.

In support of this model, studies have found an increase in DNA methylation at the critically important pluripotency gene *Oct4* (Ben-Shushan et al., 1993; Deb-Rinker et al., 2005). Additionally, the promoters of key developmental regulators are enriched with histone modifications, such as H3K27me3 or H3K4me3, rather than more stable DNA methylation in ESCs compared to somatic cells (Azuara et al., 2006; Bernstein et al., 2006; Pan et al., 2007). A limitation of these studies is the biased focus on loci involved in pluripotency and development. Therefore, to fully understand the role of the epigenome in differentiation, we took advantage of publicly available data sets from ESCs and distinct types of somatic cells in both human and mouse and performed an unbiased, genome-wide analysis of DNA methylation, histone modifications, and transcription across differentiation. Our data provides evidence against the prevailing model in which ESCs have a more flexible epigenome than somatic cells and instead supports that somatic cells have a more adaptable epigenetic landscape.

By analyzing two components of the epigenome, DNA methylation and histone modifications, across cellular differentiation in two somatic cell types in human and mouse, we found extensive, cell type-specific changes. Global analysis of DNA methylation revealed that millions of CpGs are differentially methylated across differentiation, with 90% of these differentially methylated CpGs being hypomethylated. Additionally, we found that 70% of these differentially methylated CpGs across differentiation are unique to each somatic cell type, which is in agreement with a recent finding of tissue-specific methylation patterns in mouse (Hon et al., 2013). Strikingly, we found that genomic regions that are hypomethylated in somatic cells are accompanied by a gain of 22 different histone modifications in a cell type-specific manner. This suggests that the switching from DNA methylation to histone modifications is an important epigenetic event for differentiation and lineage specification.

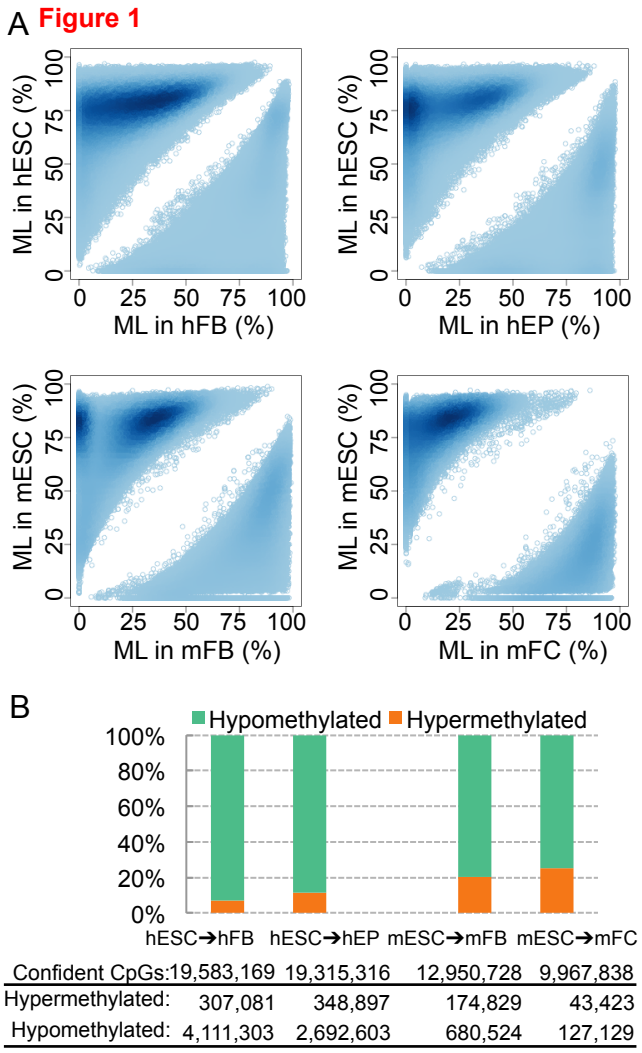
The association between DNA hypomethylation and the gain of histone modifications has been suggested for only a few histone modifications, such as H3K27me3, H3K9me3,

and H3K4me1 (Gifford et al., 2013; Hawkins et al., 2010; Lister et al., 2009). Our results have identified 18 new histone modifications that correlate with hypomethylation. We found that genes repressed in both ESCs and somatic cells are held silent in ESCs by DNA methylation, a more permanent epigenetic modification, whereas in somatic cells genes are repressed by histone modifications, a more transient modification. This switching from DNA methylation to histone modifications at repressed genes across differentiation supports that the epigenome is more flexible in terminally differentiated cells than in ESCs.

Our finding that epigenomic reconfiguration occurs in a cell type-dependent and locus-specific manner raises the importance of profiling epigenomic information in homogeneous cell populations. Although cellular heterogeneity has been difficult to address, especially for complex tissues such as the brain, advanced technology in combination with newly designed approaches should make this feasible in the near future. The data processing pipelines developed in this study can allow us to address the extent of reconfiguration, and interplay with histone modifications, including recently discovered hmCpG and non-mCpG**,** once genomic data for these modifications in different cell types become available.

Figures

Figure 1



	hESC→hFB	hESC→hEP	mESC→mFB	mESC→mFC
Confident CpGs:	19,583,169	19,315,316	12,950,728	9,967,838
Hypermethylated:	307,081	348,897	174,829	43,423
Hypomethylated:	4,111,303	2,692,603	680,524	127,129

Figure 1 Global DNA hypomethylation in somatic cells in both human and mouse. (A) Methylation levels (ML) of differentially methylated CpGs (diff-mCpGs) between ESCs and two somatic cell types in human and mouse. Each circle represents a CpG. The deeper the shade of blue, the higher the point density (number of CpGs). **(B)** The percentage of diff-mCpGs (FDR<0.05, two-tailed Fisher's exact test, B-H correction) that are hypomethylated or hypermethylated during cellular differentiation. Confident CpGs are those with 10 or more read coverage in both ESCs and somatic cells.

Figure 2

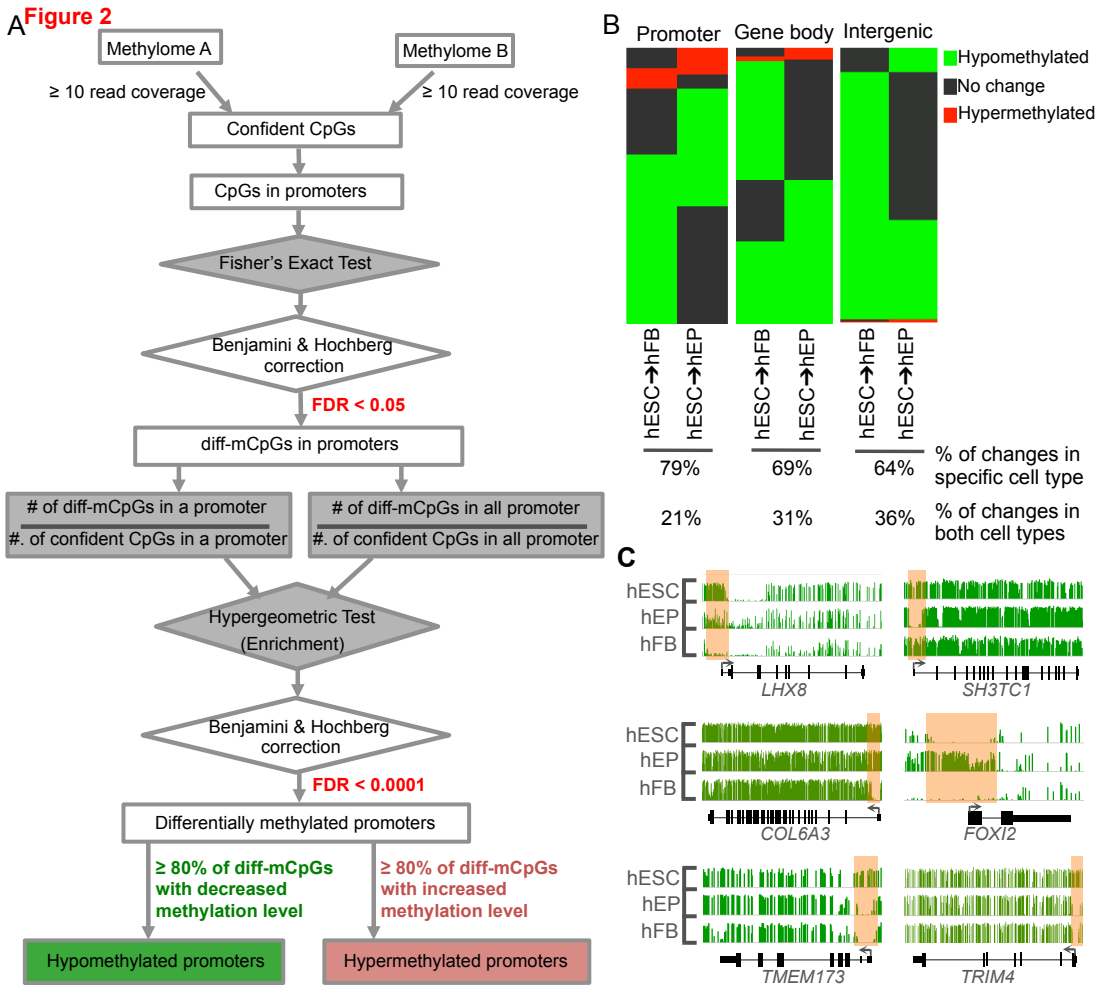


Figure 2 DNA hypomethylation during cellular differentiation is locus- and cell type-specific. (A) A schematic of the enrichment-based statistical approach used to identify differentially methylated genomic regions (e.g. promoters). Confident CpG, CpGs with 10 or more read coverage in both data sets; diff-mCpGs, differentially methylated CpGs. **(B)** Heat maps of differentially methylated genomic regions between hESC and the two human somatic cell types after clustering analysis. **(C)** Browser representation of methylation profiles of the genes with changes in methylation levels at their promoter regions during cellular differentiation. Each green vertical bar represents a CpG, and the height of the bar represents its methylation level (from 0 to 100%). Grey arrows indicate the transcriptional orientation of each gene. Regions marked by orange bars indicate the locus- and cell type-specific DNA methylation changes.

Figure 3

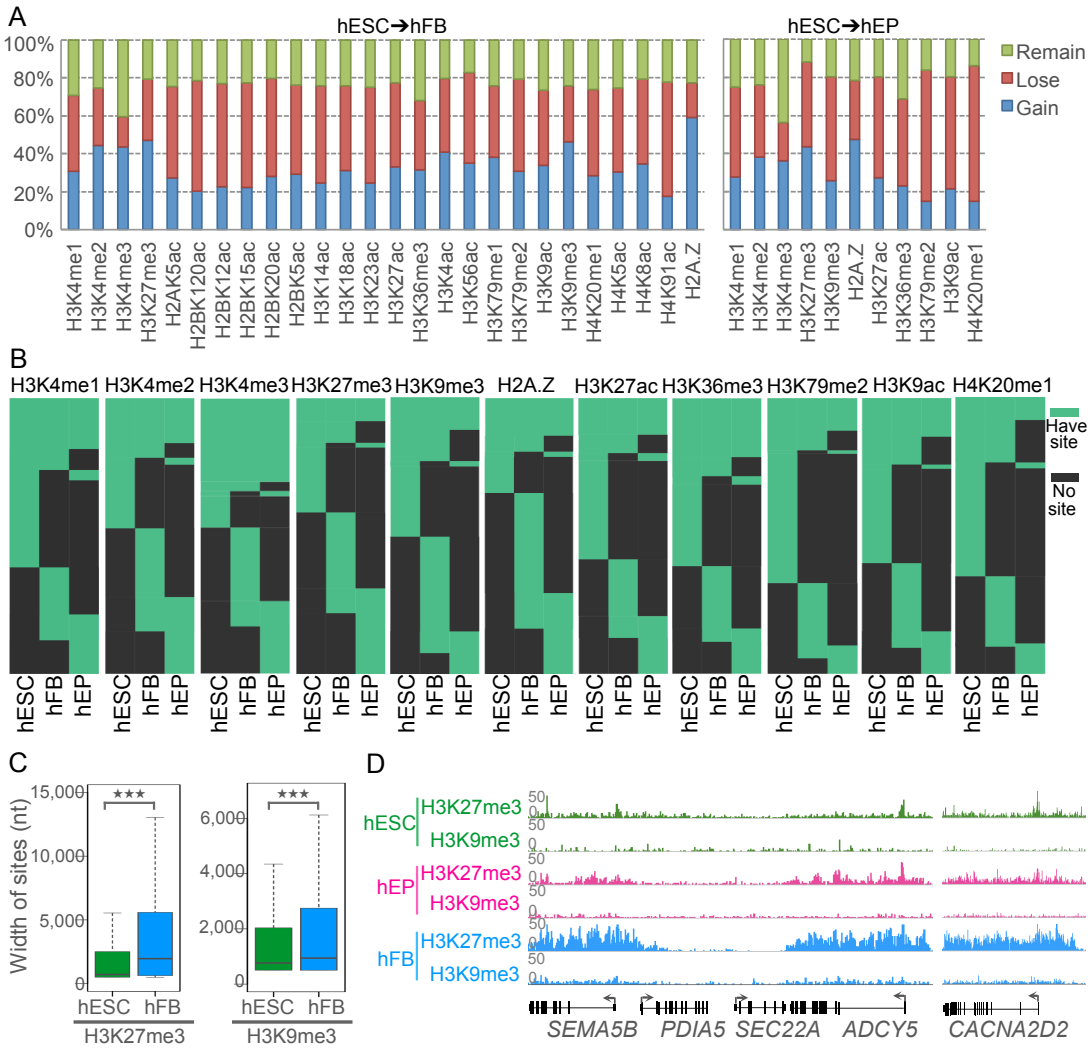


Figure 3 Changes in histone modifications during cellular differentiation are locus- and cell type-specific. (A) The percentage of enriched sites for each histone modification that were gained (gain), lost (lose), and remained (remain) during differentiation of hESC to hFB and hEP. **(B)** Heat maps of changes of enriched sites for 11 histone modifications during cellular differentiation in human after clustering analysis. **(C)** The width of H3K27me3 and H3K9me3 enriched sites in hESC and hFB. ★ ★ ★, p-value < 2.2×10^{-16}, Student's t-test. **(D)** Browser representation of two regions with H3K27me3 and H3K9me3 expansion during cellular differentiation in human.

Figure 4

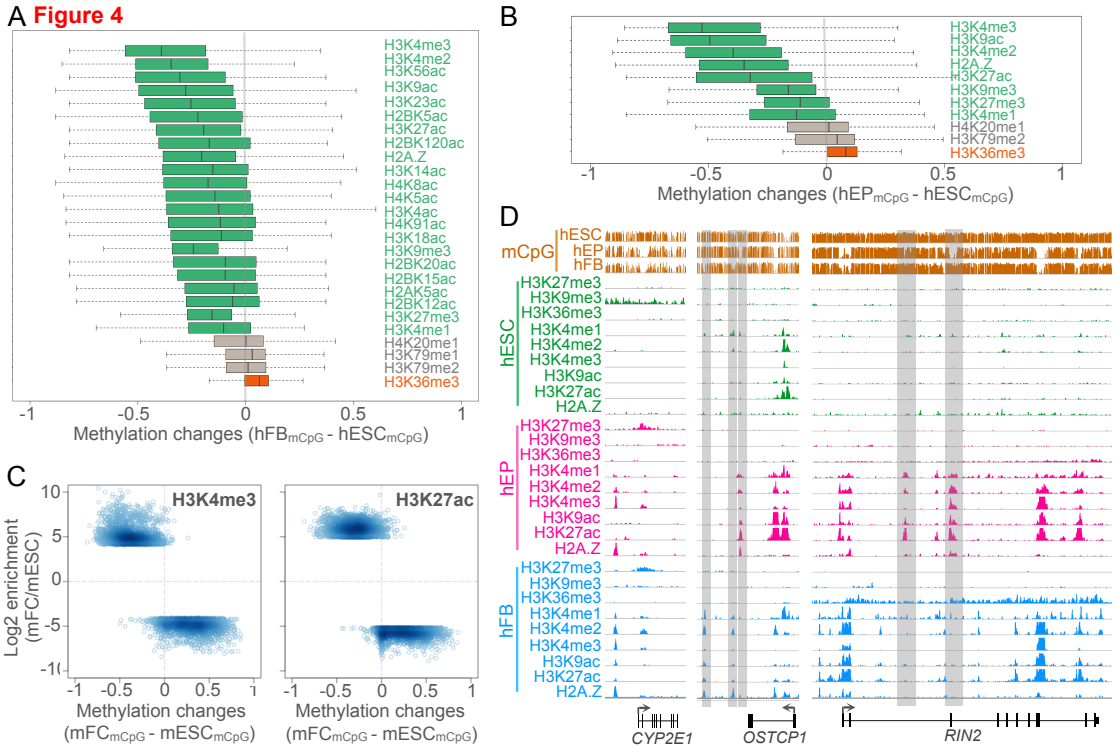


Figure 4 A switching from DNA methylation to histone modifications during cellular differentiation. (A) DNA methylation changes in the highly dynamic regions that gain histone modifications during the differentiation of hESC to hFB. Green regions, histone modifications that are associated with a decrease in DNA methylation; Gray regions, histone modifications that are not associated with a change in DNA methylation; Red regions, histone modifications that are associated with an increase in DNA methylation. **(B)** DNA methylation changes in the highly dynamic gain regions during the differentiation of hESC to hEP. **(C)** DNA methylation changes in the highly dynamic regions that gain or lose H3K4me3 and H3K27ac during the differentiation of mESC to mFC. Enrichment score for each region was calculated by dividing the number of normalized reads aligned in a particular region in mFC by that in mESC. **(D)** Browser representation of three regions that exhibit switching from DNA methylation to histone modifications during cellular differentiation in human. Regions marked by grey bars indicate the cell type-specific switching. The track scale for all histone modifications is from 0 to 50 normalized reads.

Figure 5

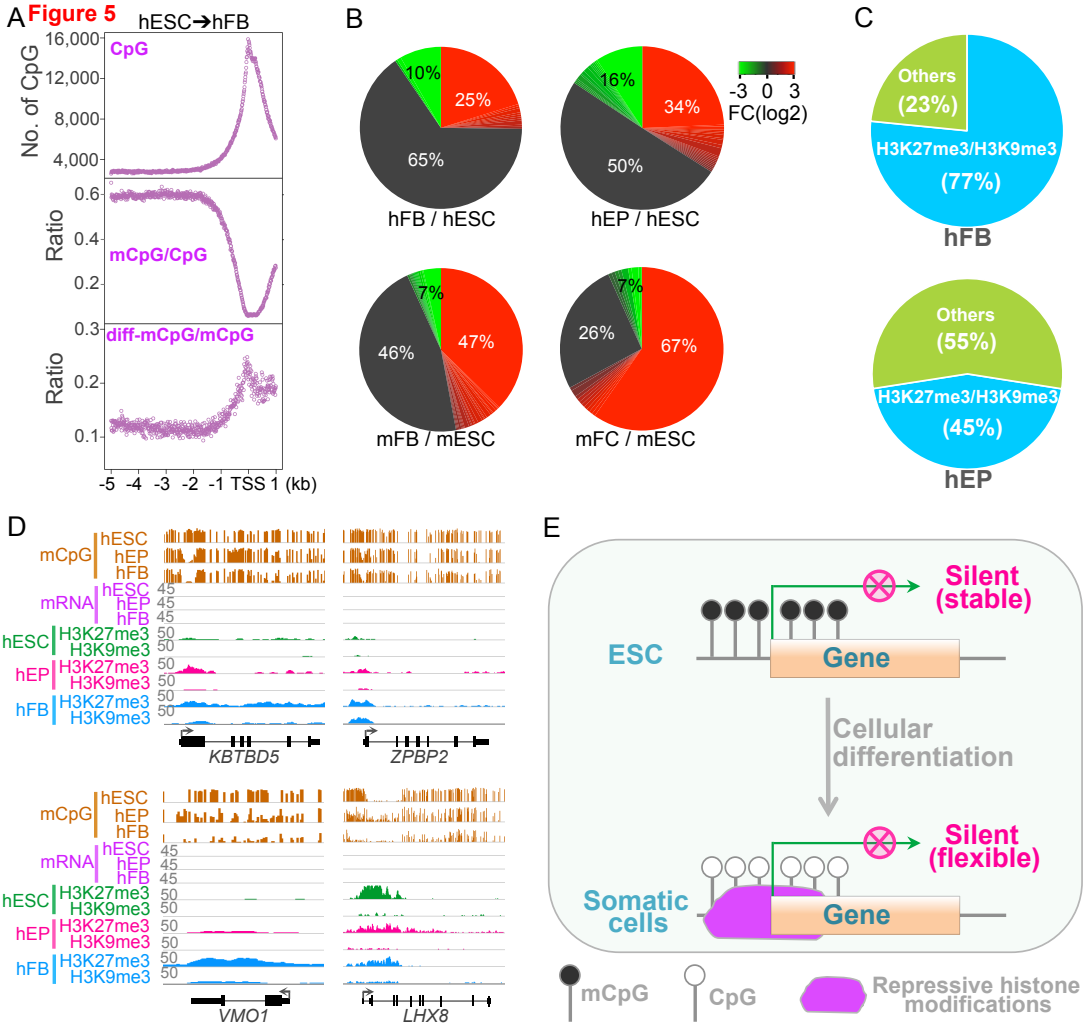


Figure 5 Switching from DNA methylation to repressive histone modifications at promoter regions uncovers a change of gene regulation from ESCs to somatic cells. (A) Scatter plots of the number of genomic CpGs, the ratio of mCpG, and the ratio of diif-mCpG in 10 bp bins within the region 5 kb upstream and 1 kb downstream of known transcription start site (TSS). **(B)** Heat map-pie charts of the expression changes of the genes with hypomethylated promoters. FC, fold change, gene expression levels in somatic cell types divided by that in ESCs. **(C)** Pie charts of percentages of genes, which are hypomethylated at the promoter regions but with no expression changes during cellular differentiation, marked by H3K27me3 or H3K9me3 or both in somatic cells. **(D)** Browser representation of four genes with epigenomic

reconfiguration at their promoter regions. These genes were silent in both hESC and somatic cell types. The track scale for all mRNA-seq is from 0 to 45 normalized reads. The track scale for all histone modifications is from 0 to 50 normalized reads. **(E)** A model depicting gene regulation changes during cellular differentiation.

Supplemental Materials

Figure S1. Global DNA methylation levels decrease during cellular differentiation in both human and mouse. **(A)** Distributions of CpG methylation levels in hESC, hFB, hEP, mESC, mFB, and mFC. M, million. **(B)** The percentage of diff-mCpGs that are hypomethylated or hypermethylated between somatic cells. **(C)** The percentage of diff-mCpGs that are hypomethylated or hypermethylated on each autosome during cellular differentiation.

Figure S2. The decrease in DNA methylation levels during cellular differentiation is locus- and cell type-specific. **(A)** Different promoter definitions used in this study and the percentage of differentially methylated promoter regions that are hypomethylated or hypermethylated during cellular differentiation. Red rectangle indicates the transcription start site (TSS). **(B)** The percentage of differentially methylated gene body regions that are hypomethylated or hypermethylated during cellular differentiation. **(C)** The percentage of differentially methylated intergenic regions that are hypomethylated or hypermethylated during cellular differentiation. **(D)** Heat maps of differentially methylated genomic regions between mESC and the two mouse somatic cell types after clustering analysis. **(E,F)** Browser representation of methylation profiles of the genes with changes in methylation levels at their promoter regions during cellular differentiation. Each green vertical bar represents a CpG, and the height of the bar represents its methylation level (from 0 to 100%). Red arrow indicates the transcriptional orientation of each gene.

Figure S3. A switch from DNA methylation to histone modifications during cellular differentiation in human. **(A)** DNA methylation changes in the highly dynamic lose regions that lose histone modifications during the differentiation of hESC to hFB. **(B)** DNA methylation changes in the highly dynamic lose regions during the differentiation of hESC to hEP. **(C)** Browser representation of regions that switch from DNA methylation to histone modifications during cellular differentiation in human. The track scale for all histone modifications is from 0 to 50 normalized reads.

Figure S4. Epigenomic reconfiguration at promoter regions. **(A)** Scatter plots of the number of genomic CpGs, the ratio of mCpG, and the ratio of diff-mCpG in 10 bp bins within the regions of 5 kb upstream and 1 kb downstream of known transcription start

site (TSS). **(B)** Top 10 enriched GO terms of genes with hypomethylated promoters. ★ indicates GO terms with Benjamini-corrected p-values < 0.05.

Figure S5. Epigenomic reconfiguration at promoter regions. **(A)** Browser representation of six genes with alternative promoter usage. Red arrow indicates the transcriptional orientation of each gene. The track scale for all mRNA-seq is from 0 to 45 normalized reads. **(B)** Browser representation of twelve genes with epigenomic reconfiguration at their promoter regions during cellular differentiation. The track scale for all mRNA-seq was from 0 to 45. The track scale for all histone modifications is from 0 to 50 normalized reads.

References (Specific to Appendix 3)

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., *et al.* (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532-538.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., LeProust, E.M., Park, I.H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol *27*, 361-368.

Ben-Shushan, E., Pikarsky, E., Klar, A., and Bergman, Y. (1993). Extinction of Oct-3/4 gene expression in embryonal carcinoma x fibroblast somatic cell hybrids is accompanied by changes in the methylation status, chromatin structure, and transcriptional activity of the Oct-3/4 upstream region. Mol Cell Biol *13*, 891-901.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met *57*, 289-300.

Berger, S.L. (2007). The complex language of chromatin regulation during transcription. Nature *447*, 407-412.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev *16*, 6-21.

Deb-Rinker, P., Ly, D., Jezierski, A., Sikorska, M., and Walker, P.R. (2005). Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. J Biol Chem *280*, 6257-6260.

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

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., *et al.* (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell *153*, 1149-1163.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., *et al.* (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell *6*, 479-491.

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

Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., *et al.* (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res *22*, 246-258.

Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat Genet *45*, 1198-1206.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol Cell Biol *24*, 8862-8871.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet *33 Suppl*, 245-254.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet *13*, 484-492.

Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G.*, et al.* (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell *8*, 200-213.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571-1572.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Lister, R., and Ecker, J.R. (2009). Finding the fifth base: genome-wide sequencing of cytosine methylation. Genome Res *19*, 959-966.

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

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.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499-506.

Mann, I.K., Chatterjee, R., Zhao, J., He, X., Weirauch, M.T., Hughes, T.R., and Vinson, C. (2013). CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. Genome Res *23*, 988-997.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343-349.

Metivier, R., Penot, G., Hubner, M.R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. Cell *115*, 751-763.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell *1*, 299-312.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425-432.

Rivera, C.M., and Ren, B. (2013). Mapping human epigenomes. Cell *155*, 39-55.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

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

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat Rev Genet *14*, 204-220.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D.*, et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490-495.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform *14*, 178-192.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R.*, et al.* (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells *11*, 805-814.

Turner, B.M. (2007). Defining an epigenetic code. Nat Cell Biol *9*, 2-6.

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

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D.*, et al.* (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell *153*, 1134-1148.

Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E.*, et al.* (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477-481.

BIBLIOGRAPHY

Agarwal, N., Becker, A., Jost, K.L., Haase, S., Thakur, B.K., Brero, A., Hardt, T., Kudo, S., Leonhardt, H., and Cardoso, M.C. (2011). MeCP2 Rett mutations affect large scale chromatin organization. Hum. Mol. Genet. *20*, 4187–4195.

Agarwal, N., Hardt, T., Brero, A., Nowak, D., Rothbauer, U., Becker, A., Leonhardt, H., and Cardoso, M.C. (2007). MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. Nucleic Acids Res. *35*, 5402–5408.

Allan, J. (1984). Higher order structure in a short repeat length chromatin. J. Cell Biol. *98*, 1320–1327.

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

Armstrong, D., Dunn, J.K., Antalffy, B., and Trivedi, R. (1995). Selective Dendritic Alterations in the Cortex of Rett Syndrome. Journal of Neuropathology and Experimental Neurology *54*, 195–201.

Avrahami, D., Li, C., Zhang, J., Schug, J., Avrahami, R., Rao, S., Stadler, M.B., Burger, L., Schübeler, D., Glaser, B., et al. (2015). Aging-Dependent Demethylation of Regulatory Elements Correlates with Chromatin State and Improved β Cell Function. Cell Metab. *22*, 619–632.

Baker, S.A., Chen, L., Wilkins, A.D., Yu, P., Lichtarge, O., and Zoghbi, H.Y. (2013). An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. Cell *152*, 984–996.

Ballas, N., Lioy, D.T., Grunseich, C., and Mandel, G. (2009). Non–cell autonomous influence of MeCP2-deficient glia on neuronal dendritic morphology. Nat. Neurosci. *12*, 311–317.

Balmer, D., Goldstine, J., Rao, Y.M., and LaSalle, J.M. (2002). Elevated methyl-CpG-binding protein 2 expression is acquired during postnatal human brain development and is correlated with alternative polyadenylation. J. Mol. Med. *81*, 61–68.

Baubec, T., Ivanek, R., Lienert, F., and Schübeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell *153*, 480–492.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends in Genetics.

Becker, P.B., Ruppert, S., and Schütz, G. (1987). Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. Cell *51*, 435–443.

Belichenko, N.P., Belichenko, P.V., and Mobley, W.C. (2009a). Evidence for both neuronal cell autonomous and nonautonomous effects of methyl-CpG-binding protein 2 in the cerebral cortex of female mice with Mecp2 mutation. Neurobiol. Dis.

Belichenko, P.V., Oldfors, A., Hagberg, B., and Dahlström, A. (1994). Rett syndrome: 3-D confocal microscopy of cortical pyramidal dendrites and afferents. NeuroReport *5*, 1509.

Belichenko, P.V., Wright, E.E., Belichenko, N.P., Masliah, E., Li, H.H., Mobley, W.C., and Francke, U. (2009b). Widespread changes in dendritic and axonal morphology in Mecp2-mutant mouse models of rett syndrome: Evidence for disruption of neuronal networks. Journal of Comparative Neurology *514*, 240–258.

Beliveau, B.J., Boettiger, A.N., and Avendaño, M.S. (2015). Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes : Nature Communications. Nature.

Beliveau, B.J., Joyce, E.F., Apostolopoulos, N., Yilmaz, F., Fonseka, C.Y., McCole, R.B., Chang, Y., Li, J.B., Senaratne, T.N., Williams, B.R., et al. (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. Proceedings of the ….

Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C.A., and Zoghbi, H.Y. (2009). Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. Hum. Mol. Genet. *18*, 2431–2442.

Benayoun, B.A., Pollina, E.A., and Brunet, A. (2015). Epigenetic regulation of ageing: linking environmental inputs to genomic stability. Nat. Rev. Mol. Cell Biol. *16*, 593–610.

Bestor, T.H., Xu, G.-L., Bourc'his, D., Hsieh, C.-L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene : Abstract : Nature. Nature *402*, 187–191.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev.

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

Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martínez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature *500*, 222–226.

Blitzer, R.D. (1998). Gating of CaMKII by cAMP-Regulated Protein Phosphatase Activity During LTP. Science *280*, 1940–1943.

Bocklandt, S., Lin, W., Sehl, M.E., Sánchez, F.J., Sinsheimer, J.S., Horvath, S., and Vilain, E. (2011). Epigenetic predictor of age. PLoS ONE *6*, e14821.

Brero, A., Easwaran, H.P., Nowak, D., Grunewald, I., Cremer, T., Leonhardt, H., and

Cardoso, M.C. (2005). Methyl CpG–binding proteins induce large-scale chromatin reorganization during terminal differentiation. J. Cell Biol. *169*, 733–743.

Brooks, P.J., Marietta, C., and Goldman, D. (1996). DNA mismatch repair and DNA methylation in adult brain neurons. J. Neurosci. *16*, 939–945.

Brown, G.P., Blitzer, R.D., Connor, J.H., Wong, T., Shenolikar, S., Iyengar, R., and Landau, E.M. (2000). Long-Term Potentiation Induced by θ Frequency Stimulation Is Regulated by a Protein Phosphatase-1-Operated Gate. J. Neurosci. *20*, 7880–7887.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Burger, L., Gaidatzis, D., Schubeler, D., and Stadler, M.B. (2013). Identification of active regulatory regions from DNA methylation data. Nucleic Acids Res. *41*, e155–e155.

Burgess, A., Vigneron, S., Brioudes, E., Labbé, J.-C., Lorca, T., and Castro, A. (2010). Loss of human Greatwall results in G2 arrest and multiple mitotic defects due to deregulation of the cyclin B-Cdc2/PP2A balance. Proc. Natl. Acad. Sci. U.S.a. *107*, 12564–12569.

Carey, B.W., Finley, L.W.S., Cross, J.R., Allis, C.D., and Thompson, C.B. (2015). Intracellular α-ketoglutarate maintains the pluripotency of embryonic stem cells. Nature *518*, 413–416.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. *10*, 295–304.

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T.C., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a Key Contributor to Neurological Disease, Activates and Represses Transcription. Science *320*, 1224–1229.

Chang, Q., Khare, G., Dani, V., Nelson, S., and Jaenisch, R. (2006). The Disease Progression of Mecp2 Mutant Mice Is Affected by the Level of BDNF Expression. Neuron.

Chao, H.-T., Zoghbi, H.Y., and Rosenmund, C. (2007). MeCP2 Controls Excitatory Synaptic Strength by Regulating Glutamatergic Synapse Number. Neuron *56*, 58–65.

Chen, H., Dzitoyeva, S., and Manev, H. (2012). Effect of aging on 5-hydroxymethylcytosine in the mouse hippocampus. Restor. Neurol. Neurosci. *30*, 237–245.

Chen, L., Chen, K., Lavery, L.A., Baker, S.A., Shaw, C.A., Li, W., and Zoghbi, H.Y. (2015). MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. Proc. Natl. Acad. Sci. U.S.a. *112*, 5509–5514.

Chen, R.Z., Akbarian, S., Tudor, M., and Jaenisch, R. (2001). Deficiency of methyl-CpG

binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nat. Genet. *27*, 327–331.

Chen, W.G. (2003). Derepression of BDNF Transcription Involves Calcium-Dependent Phosphorylation of MeCP2. Science *302*, 885–889.

Chepelev, I., Wei, G., Wangsa, D., Tang, Q., and Zhao, K. (2012). Cell Research - Abstract of article: Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. Cell Res.

Chouliaras, L., van den Hove, D., and Kenis, G. (2012a). Age-related increase in levels of 5-hydroxymethylcytosine in mouse hippocampus is prevented by caloric restriction. Current Alzheimer ….

Chouliaras, L., van den Hove, D.L.A., Kenis, G., Keitel, S., Hof, P.R., van Os, J., Steinbusch, H.W.M., Schmitz, C., and Rutten, B.P.F. (2012b). Prevention of age-related changes in hippocampal levels of 5-methylcytidine by caloric restriction. Neurobiol. Aging *33*, 1672–1681.

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

Colquitt, B.M., Allen, W.E., Barnea, G., and Lomvardas, S. (2013). Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. Proc. Natl. Acad. Sci. U.S.a. *110*, 14682–14687.

Dani, V.S., and Nelson, S.B. (2009). Intact Long-Term Potentiation but Reduced Connectivity between Neocortical Layer 5 Pyramidal Neurons in a Mouse Model of Rett Syndrome. J. Neurosci. *29*, 11263–11270.

Dani, V.S., Chang, Q., and Maffei, A. (2005). Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett Syndrome.

Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M., et al. (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev. Cell *24*, 310–323.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.-C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.-W., Page, D.C., et al. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell *9*, 166–175.

Day, K., Waite, L.L., Thalacker-Mercer, A., West, A., Bamman, M.M., Brooks, J.D., Myers, R.M., and Absher, D. (2013). Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. Genome Biol. *14*, R102.

De Felipe, J. (1997). Inhibitory synaptogenesis in mouse somatosensory cortex. Cerebral Cortex *7*, 619–634.

Delépine, C., Meziane, H., Nectoux, J., Opitz, M., Smith, A.B., Ballatore, C., Saillour, Y., Bennaceur-Griscelli, A., Chang, Q., Williams, E.C., et al. (2016). Altered microtubule dynamics and vesicular transport in mouse and human MeCP2-deficient astrocytes. Hum. Mol. Genet. *25*, 146–157.

Ebert, D.H., Gabel, H.W., Robinson, N.D., Kastan, N.R., Hu, L.S., Cohen, S., Navarro, A.J., Lyst, M.J., Ekiert, R., Bird, A.P., et al. (2013). Activity-dependent phosphorylation of MECP2 threonine 308 regulates interaction with NcoR. Nature.

Ellison, K.A., Fill, C.P., Terwilliger, J., DeGennaro, L.J., Martin-Gallardo, A., Anvret, M., Percy, A.K., Ott, J., and Zoghbi, H. (1992). Examination of X chromosome markers in Rett syndrome: exclusion mapping with a novel variation on multilocus linkage analysis. Am. J. Hum. Genet. *50*, 278–287.

Ernst, J., and Kellis, M. (2012). correspondence. Nat. Methods *9*, 215–216.

Fan, G., Beard, C., Chen, R.Z., Csankovszki, G., Sun, Y., Siniaia, M., Biniszkiewicz, D., Bates, B., Lee, P.P., Kuhn, R., et al. (2001). DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. J. Neurosci. *21*, 788–797.

FEARNLEY, J.M., and LEES, A.J. (1991). AGEING AND PARKINSON'S DISEASE: SUBSTANTIA NIGRA REGIONAL SELECTIVITY. Brain *114*, 2283–2301.

Feng, J., Chang, H., Li, E., and Fan, G. (2005). Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. J. Neurosci. Res. *79*, 734–746.

Feng, J., Zhou, Y., Campbell, S.L., Le, T., Li, E., Sweatt, J.D., Silva, A.J., and Fan, G. (2010). Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nat. Neurosci. *13*, 423–430.

Filion, G.J.P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., and Defossez, P.A. (2005). A Family of Human Zinc Finger Proteins That Bind Methylated DNA and Repress Transcription. Mol. Cell. Biol. *26*, 169–181.

Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M.C., Antes, I., and Leonhardt, H. (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. PLoS ONE *6*, e21306.

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

Georgel, P.T., Horowitz-Scherer, R.A., Adkins, N., Woodcock, C.L., Wade, P.A., and Hansen, J.C. (2003). Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. J. Biol. Chem. *278*, 32181–32188.

Ghosh, R.P., Horowitz-Scherer, R.A., Nikitina, T., Shlyakhtenko, L.S., and Woodcock, C.L. (2010). MeCP2 Binds Cooperatively to Its Substrate and Competes with Histone H1 for Chromatin Binding Sites. Mol. Cell. Biol. *30*, 4656–4670.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS ONE *5*, e15367.

Goedert, M., Spillantini, M.G., Del Tredici, K., and Braak, H. (2013). 100 years of Lewy pathology.

Goffin, D., Allen, M., Zhang, L., Amorim, M., Wang, I.-T.J., Reyes, A.-R.S., Mercado-Berton, A., Ong, C., Cohen, S., Hu, L., et al. (2012). Rett syndrome mutation MeCP2 T158A disrupts DNA binding, protein stability and ERP responses. Nat. Neurosci. *15*, 274–283.

Golshani, P., Hutnick, L., Schweizer, F., and Fan, G. (2007). Conditional Dnmt1 deletion in dorsal forebrain disrupts development of somatosensory barrel cortex and thalamocortical long-term potentiation. Thl *3*, 227.

Gonzales, M.L., Adams, S., Dunaway, K.W., and LaSalle, J.M. (2012). Phosphorylation of distinct sites in MeCP2 modifies cofactor associations and the dynamics of transcriptional regulation. Mol. Cell. Biol. *32*, 2894–2903.

González-Hernández, T., and Rodríguez, M. (2000). Compartmental organization and chemical profile of dopaminergic and GABAergic neurons in the substantia nigra of the rat. J. Comp. Neurol. *421*, 107–135.

Goto, K., Numata, M., Komura, J.I., Ono, T., and Bestor, T.H. (1994). Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice - ScienceDirect. Differentiation.

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription : Abstract : Nature. Nature *389*, 349–352.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Guo, J.U., Ma, D.K., Mo, H., Ball, M.P., Jang, M.-H., Bonaguidi, M.A., Balazer, J.A., Eaves, H.L., Xie, B., Ford, E., et al. (2011a). Neuronal activity modifies the DNA methylation landscape in the adult brain. Nat. Neurosci. *14*, 1345–1351.

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., et al. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat. Neurosci. *17*, 215–222.

Guo, J.U., Su, Y., Zhong, C., Ming, G.-L., and Song, H. (2011b). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423–434.

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

Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.-G., Jiang, Y., Pfeifer, G.P., et al. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. Cell Rep *3*, 291–300.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.-B., Gao, Y., et al. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol. Cell *49*, 359–367.

Harr, J.C., Luperchio, T.R., Wong, X., Cohen, E., Wheelan, S.J., and Reddy, K.L. (2015). Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. J. Cell Biol. *208*, 33–52.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. *40*, 4841–4849.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Hendrich, B., and Bird, A. (1998). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. Mol. Cell. Biol. *18*, 6538–6547.

Horike, S.-I., Cai, S., Miyano, M., Cheng, J.-F., and Kohwi-Shigematsu, T. (2004). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat. Genet. *37*, 31–40.

Hutnick, L.K., Golshani, P., Namihira, M., Xue, Z., Matynia, A., Yang, X.W., Silva, A.J., Schweizer, F.E., and Fan, G. (2009). DNA hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation. Hum. Mol. Genet. *18*, 2875–2888.

Ishibashi, T., Thambirajah, A.A., and Ausio, J. (2008). MeCP2 preferentially binds to methylated linker DNA in the absence of the terminal tail of histone H3 and independently of histone acetylation - Ishibashi - 2008 - FEBS Letters - Wiley Online Library. FEBS Lett.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell Cycle *8*, 1698–1710.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. *33 Suppl*, 245–254.

Jagust, W. (2013). Vulnerable neural systems and the borderland of brain aging and neurodegeneration. Neuron *77*, 219–234.

Jeffrey L Neul, Kaufmann, W.E., Glaze, D.G., Christodoulou, J., Clarke, A.J., Bahi-Buisson, N., Leonard, H., Bailey, M.E.S., Schanen, N.C., Zappella, M., et al. (2010). Rett syndrome: Revised diagnostic criteria and nomenclature. Ann Neurol. *68*, 944–950.

Jiang, Y., Matevossian, A., Huang, H.-S., Straubhaar, J., and Akbarian, S. (2008). Isolation of neuronal chromatin from brain tissue. BMC Neuroscience 2008 9:1 *9*, 42.

Jin, S.-G., Wu, X., Li, A.X., and Pfeifer, G.P. (2011). Genomic mapping of 5-hydroxymethylcytosine in the human brain. Nucleic Acids Res. *39*, 5015–5024.

Jones, P.L., Jan Veenstra, G.C., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription

. Nat. Genet. *19*, 187–191.

Jost, K.L., Bertulat, B., Rapp, A., Brero, A., Hardt, T., Domaing, P., Gösele, C., Schulz, H., Hübner, N., and Cardoso, M.C. (2015). Gene repositioning within the cell nucleus is not random and is determined by its genomic neighborhood. Epigenetics & Chromatin 2015 8:1 *8*, 36.

Kaas, G.A., Zhong, C., Eason, D.E., Ross, D.L., Vachhani, R.V., Ming, G.-L., King, J.R., Song, H., and Sweatt, J.D. (2013). TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. Neuron *79*, 1086–1093.

Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M., et al. (2012a). nsmb.2372. Nat. Struct. Mol. Biol. *19*, 1037–1043.

Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M., et al. (2012b). 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. Nat. Struct. Mol. Biol. *19*, 1037–1043.

Khrapunov, S., Warren, C., Cheng, H., Berko, E.R., Greally, J.M., and Brenowitz, M. (2014). Unusual characteristics of the DNA binding domain of epigenetic regulatory protein MeCP2 determine its binding specificity. Biochemistry *53*, 3379–3391.

Kinde, B., Gabel, H.W., Gilbert, C.S., Griffith, E.C., and Greenberg, M.E. (2015). Reading the unique DNA methylation landscape of the brain: Non-CpG methylation,

hydroxymethylation, and MeCP2. Proceedings of the National Academy of Sciences *112*, 6800–6806.

Klein, C.J., Botuyan, M.-V., Wu, Y., Ward, C.J., Nicholson, G.A., Hammans, S., Hojo, K., Yamanishi, H., Karpf, A.R., Wallace, D.C., et al. (2011). Mutations in. Nature Publishing Group *43*, 595–600.

Klose, R.J., Sarraf, S.A., Schmiedeberg, L., and McDermott, S.M. (2005). DNA Binding Selectivity of MeCP2 Due to a Requirement for A/T Sequences Adjacent to Methyl-CpG: Molecular Cell. Mol. Cell.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends in Biochemical Sciences *31*, 89–97.

Ko, M., Bandukwala, H.S., An, J., Lamperti, E.D., Thompson, E.C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S.B., and Rao, A. (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proc. Natl. Acad. Sci. U.S.a. *108*, 14566–14571.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472–479.

Kokura, K., Kaul, S.C., Wadhwa, R., Nomura, T., Khan, M.M., Shinagawa, T., Yasukawa, T., Colmenares, C., and Ishii, S. (2001). The Ski Protein Family Is Required for MeCP2-mediated Transcriptional Repression. Journal of Biological Chemistry *276*, 34115–34121.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929–930.

Laget, S., Joulie, M., Le Masson, F., Sasai, N., Christians, E., Pradhan, S., Roberts, R.J., and Defossez, P.-A. (2010). The Human Proteins MBD5 and MBD6 Associate with Heterochromatin but They Do Not Bind Methylated DNA. PLoS ONE *5*, e11982.

Lagger, S., Connelly, J.C., Schweikert, G., Webb, S., Selfridge, J., Ramsahoye, B.H., Yu, M., DeSousa, D., Seiser, C., He, C., et al. (2016). Domains of methylated CAC and CG target MeCP2 to tune transcription in the brain. bioRxiv 087577.

Lehnertz, B., Ueda, Y., Derijck, A., and Braunschweig, U. (2003). Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. Current Biology.

Lewis, J.D., Meehan, R.R., Henzel, W.J., and Maurer-Fogy, I. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to Methylated DNA. Cell.

Li, X., Wei, W., Zhao, Q.-Y., Widagdo, J., Baker-Andresen, D., Flavell, C.R., D'Alessio, A., Zhang, Y., and Bredy, T.W. (2014). Neocortical Tet3-mediated accumulation of 5-hydroxymethylcytosine promotes rapid behavioral adaptation. Proc. Natl. Acad. Sci. U.S.a. *111*, 7120–7125.

Li, Y., Wang, H., Muffat, J., Cheng, A.W., and Orlando, D.A. (2013). Global Transcriptional and Translational Repression in Human-Embryonic-Stem-Cell-Derived Rett Syndrome Neurons. Cell Stem Cell.

Li, Z., Cai, X., Cai, C.-L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.-C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood *118*, 4509–4518.

Liao, W.-L., Tsai, H.-C., Wang, H.-F., Chang, J., Lu, K.-M., Wu, H.-L., Lee, Y.-C., Tsai, T.-F., Takahashi, H., Wagner, M., et al. (2008). Modular patterning of structure and function of the striatum by retinoid receptor signaling. Proc. Natl. Acad. Sci. U.S.a. *105*, 6765–6770.

Linhoff, M.W., Garg, S.K., and Mandel, G. (2015). A High-Resolution Imaging Approach to Investigate Chromatin Architecture in Complex Tissues. Cell.

Lioy, D.T., Garg, S.K., Monaghan, C.E., Raber, J., Foust, K.D., Kaspar, B.K., Hirrlinger, P.G., Kirchhoff, F., Bissonnette, J.M., Ballas, N., et al. (2011). A role for glia in the progression of Rett's syndrome. Nature *475*, 497–500.

Lisman, J., Yasuda, R., and Raghavachari, S. (2012). Mechanisms of CaMKII action in long-term potentiation. Nat. Rev. Neurosci.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global Epigenomic Reconfiguration During Mammalian Brain Development. Science.

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., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Liu, S., Jiang, J., Li, L., Amato, N.J., Wang, Z., and Wang, Y. (2015). Arsenite Targets the Zinc Finger Domains of Tet Proteins and Inhibits Tet-Mediated Oxidation of 5-Methylcytosine. Environ. Sci. Technol. *49*, 11923–11931.

Liu, S., Wang, J., Su, Y., Guerrero, C., Zeng, Y., Mitra, D., Brooks, P.J., Fisher, D.E., Song, H., and Wang, Y. (2013). Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. Nucleic Acids Res. *41*, 6421–6429.

Lubin, F.D., Roth, T.L., and Sweatt, J.D. (2008). Epigenetic Regulation of bdnf Gene Transcription in the Consolidation of Fear Memory. J. Neurosci. *28*, 10576–10586.

LYNCH, M.A. (2004). Long-Term Potentiation and Memory. Physiological Reviews *84*, 87–136.

Lyst, M.J., and Bird, A. (2015). Rett syndrome: a complex disorder with simple roots. Nat. Rev. Genet. *16*, 261–275.

Lyst, M.J., Ekiert, R., Ebert, D.H., Merusi, C., Nowak, J., Selfridge, J., Guy, J., Kastan,

N.R., Robinson, N.D., de Lima Alves, F., et al. (2013). Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. Nat. Neurosci. *16*, 898–902.

Maegawa, S., Hinkal, G., Kim, H.S., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L.A., and Issa, J.-P.J. (2010). Widespread and tissue specific age-related DNA methylation changes in mice. Genome Res. *20*, 332–340.

Maezawa, I., Swanberg, S., Harvey, D., LaSalle, J.M., and Jin, L.W. (2009). Rett Syndrome Astrocytes Are Abnormal and Spread MeCP2 Deficiency through Gap Junctions. J. Neurosci. *29*, 5051–5061.

Malenka, R.C., and Bear, M.F. (2004). LTP and LTD. Neuron *44*, 5–21.

Marchetto, M.C.N., Carromeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H., and Muotri, A.R. (2010). A Model for Neural Development and Treatment of Rett Syndrome Using Human Induced Pluripotent Stem Cells. Cell *143*, 527–539.

Martinowich, K. (2003). DNA Methylation-Related Chromatin Remodeling in Activity-Dependent Bdnf Gene Regulation. Science *302*, 890–893.

Martou, G., and De Boni, U. (2000). Nuclear Topology of Murine, Cerebellar Purkinje Neurons: Changes as a Function of Development. Exp. Cell Res. *256*, 131–139.

Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res. *23*, 1256–1269.

Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep *12*, 1184–1195.

Mazumder, A., Roopa, T., Basu, A., and Mahadevan, L. (2008a). Dynamics of Chromatin Decondensation Reveals the Structural Integrity of a Mechanically Prestressed Nucleus. Biophysical Journal.

Mazumder, A., Roopa, T., Basu, A., Mahadevan, L., and Shivashankar, G.V. (2008b). Dynamics of Chromatin Decondensation Reveals the Structural Integrity of a Mechanically Prestressed Nucleus. Biophysical Journal *95*, 3028–3035.

McClung, C.A., and Nestler, E.J. (2007). Neuroplasticity Mediated by Altered Gene Expression. Neuropsychopharmacology *33*, 3–17.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*, 499–507.

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012a). MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. Cell *151*, 1417–1430.

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012b). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell *151*, 1417–1430.

Miller, C.A., and Sweatt, J.D. (2007). Covalent modification of DNA regulates memory formation. Neuron *53*, 857–869.

Mo, A., Mukamel, E.A., Davis, F.P., Luo, C., and Henry, G.L. (2015). Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron.

Morris, M.J., Adachi, M., Na, E.S., and Monteggia, L.M. (2014). Selective role for DNMT3a in learning and memory. Neurobiology of Learning and ….

Morris, M.J., Na, E.S., Autry, A.E., and Monteggia, L.M. (2016). Impact of DNMT1 and DNMT3a forebrain knockout on depressive- and anxiety like behavior in mice. Neurobiology of Learning and ….

Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmiller, V., Michalakis, S., Müller, M., Biel, M., and Carell, T. (2010). Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew. Chem. Int. Ed. Engl. *49*, 5375–5377.

Nan, X., Campoy, F.J., and Bird, A. (1997). MeCP2 Is a Transcriptional Repressor with Abundant Binding Sites in Genomic Chromatin. Cell.

Nan, X., Hou, J., Maclean, A., Nasir, J., Lafuente, M.J., Shu, X., Kriaucionis, S., and Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. Proceedings of the National Academy of Sciences *104*, 2709–2714.

Nan, X., Meehan, R.R., and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res.

Nestor, C., Ruzov, A., Meehan, R., and Dunican, D. (2010). Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. BioTechniques *48*, 317–319.

Nguyen, M.V.C., Du, F., Felice, C.A., Shan, X., Nigam, A., Mandel, G., Robinson, J.K., and Ballas, N. (2012). MeCP2 is critical for maintaining mature neuronal networks and global brain anatomy during late stages of postnatal brain development and in the mature adult brain. J. Neurosci. *32*, 10021–10034.

Nguyen, S., Meletis, K., Fu, D., Jhaveri, S., and Jaenisch, R. (2007). Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. Dev. Dyn. *236*, 1663–1676.

Nikitina, T., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., Grigoryev, S.A., and Woodcock, C.L. (2007a). MeCP2-Chromatin Interactions Include the Formation of Chromatosome-like Structures and Are Altered in Mutations Causing Rett Syndrome. Journal of Biological Chemistry *282*, 28237–28245.

Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. (2007b). Multiple Modes of Interaction between the Methylated DNA Binding Protein MeCP2 and Chromatin. Mol. Cell. Biol. *27*, 864–877.

Nissim, I., Yudkoff, M., and Brosnan, J.T. (1996). Regulation of [15N]urea synthesis from [5-15N]glutamine. Role of pH, hormones, and pyruvate. J. Biol. Chem. *271*, 31234–31242.

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

Ollion, J., Cochennec, J., Loll, F., Escude, C., and Boudier, T. (2013). TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. Bioinformatics *29*, 1840–1841.

Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. *15*, 234–246.

PEARSON, E.C., BATES, D.L., PROSPERO, T.D., and THOMAS, J.O. (1984). Neuronal nuclei and glial nuclei from mammalian cerebral cortex. The FEBS Journal *144*, 353–360.

Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K., and Yura, R. (1972). The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. Biochem. J. *126*, 781–790.

Percy, A.K., Neul, J.L., Glaze, D.G., Motil, K.J., Skinner, S.A., Khwaja, O., Lee, H.-S., Lane, J.B., Barrish, J.O., Annese, F., et al. (2010). Rett syndrome diagnostic criteria: lessons from the Natural History Study. Ann Neurol. *68*, 951–955.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. Mol. Cell *38*, 603–613.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194–1211.

Piper, J., Assi, S.A., and Cauchy, P. (2015). Wellington-bootstrap: differential DNase-seq footprinting identifies cell-type determining transcription factors | BMC Genomics | Full Text. Bmc ….

Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., and Ott, S. (2013). Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Res. *41*, e201–e201.

Prokhortchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev. *15*, 1613–1618.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc. Natl. Acad. Sci. U.S.a. *97*, 5237–5242.

Roloff, T.C., Ropers, H.H., and Nuber, U.A. (2003). Comparative study of methyl-CpG-binding domain proteins. BMC Genomics *4*, 1.

Rudenko, A., Dawlaty, M.M., Seo, J., Cheng, A.W., Meng, J., Le, T., Faull, K.F., Jaenisch, R., and Tsai, L.-H. (2013). Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. Neuron *79*, 1109–1122.

Salvatore, M.F., Pruett, B.S., Dempsey, C., and Fields, V. (2012). Comprehensive profiling of dopamine regulation in substantia nigra and ventral tegmental area. J Vis Exp.

Sanders, J.L., and Newman, A.B. (2013). Telomere length in epidemiology: a biomarker of aging, age-related disease, both, or neither? Epidemiol Rev *35*, 112–131.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Schanen, C., and Francke, U. (1998). A severely affected male born into a Rett syndrome kindred supports X-linked inheritance and allows extension of the exclusion map. Am. J. Hum. Genet. *63*, 267–269.

Schanen, N.C., Dahle, E.J., Capozzoli, F., Holm, V.A., Zoghbi, H.Y., and Francke, U. (1997). A new Rett syndrome family consistent with X-linked inheritance expands the X chromosome exclusion map. Am. J. Hum. Genet. *61*, 634–641.

Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N., Nery, J.R., Urich, M.A., Chen, H., et al. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. Nature *523*, 212–216.

Shahbazian, M.D. (2002). Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. Hum. Mol. Genet. *11*, 115–124.

Shen, X., Yu, L., and Weir, J.W. (1996). Linker histories are not essential and affect chromatin condensation in vivo - ScienceDirect (...).

Singleton, M.K., Gonzales, M.L., Leung, K.N., Yasui, D.H., Schroeder, D.I., Dunaway, K., and LaSalle, J.M. (2011). MeCP2 is required for global heterochromatic and nucleolar changes during activity-dependent neuronal maturation. Neurobiol. Dis. *43*, 190–200.

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

Solovei, I., Grandi, N., Knoth, R., Volk, B., and Cremer, T. (2004). Positional changes of pericentromeric heterochromatin and nucleoli in postmitotic Purkinje cells during

murine cerebellum development. Cytogenet Genome Res *105*, 302–310.

Song, C.-X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X., et al. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat. Biotechnol. *29*, 68–72.

Spijker, S. (2011). Dissection of Rodent Brain Regions. In Neuroproteomics, (Totowa, NJ: Humana Press), pp. 13–26.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell *152*, 1146–1159.

Stancheva, I., Collins, A.L., Van den Veyver, I.B., Zoghbi, H., and Meehan, R.R. (2003). A Mutant Form of MeCP2 Protein Associated with Human Rett Syndrome Cannot Be Displaced from Methylated DNA by Notch in Xenopus Embryos. Mol. Cell *12*, 425–435.

Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., et al. (2011). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat. Neurosci. *14*, 1607–1616.

Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F., and Leonhardt, H. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. *38*, e181.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

Tai, D.J.C., Liu, Y.C., Hsu, W.L., Ma, Y.L., Cheng, S.J., Liu, S.Y., and Lee, E.H.Y. (2016). MeCP2 SUMOylation rescues Mecp2-mutant-induced behavioural deficits in a mouse model of Rett syndrome. Nat Commun *7*, 10552.

Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Aksglaede, L., et al. (2014). Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. Nat. Genet. *46*, 385–388.

Tepper, J.M., and Bolam, J.P. (2004). Functional diversity and specificity of neostriatal interneurons. Curr. Opin. Neurobiol. *14*, 685–692.

Tsygankov, D., Liu, Y., Sanoff, H.K., Sharpless, N.E., and Elston, T.C. (2009). A quantitative model for age-dependent expression of the p16INK4a tumor suppressor. Proc. Natl. Acad. Sci. U.S.a. *106*, 16562–16567.

Tudor, M., Akbarian, S., and Chen, R.Z. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain.

UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res.

43, D204–D212.

Unoki, M., Nishidate, T., and Nakamura, Y. (2004). ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. Oncogene *23*, 7601–7610.

Valinluck, V., Tsai, H.-H., Rogstad, D.K., Burdzy, A., Bird, A., and Sowers, L.C. (2004). Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Res. *32*, 4100–4108.

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E., et al. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. Genome ….

Wang, I.-T.J., Reyes, A.-R.S., and Zhou, Z. (2013). Neuronal morphology in MeCP2 mouse models is intrinsically variable and depends on age, cell type, and Mecp2 mutation. Neurobiol. Dis. *58*, 3–12.

Wang, Y., and Zhang, Y. (2014). Regulation of TET Protein Stability by Calpains. Cell Rep *6*, 278–284.

Weinberg, J.M., Venkatachalam, M.A., Roeser, N.F., and Nissim, I. (2000). Mitochondrial dysfunction during hypoxia/reoxygenation and its correction by anaerobic metabolism of citric acid cycle intermediates. Proc. Natl. Acad. Sci. U.S.a. *97*, 2826–2831.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat. Genet. *41*, 246–250.

Wen, L., Li, X., Yan, L., Tan, Y., Li, R., Zhao, Y., Wang, Y., Xie, J., Zhang, Y., Song, C., et al. (2014). Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. Genome Biol. *15*, R49.

Wheldon, L.M., Abakir, A., Ferjentsik, Z., Dudnakova, T., Strohbuecker, S., Christie, D., Dai, N., Guan, S., Foster, J.M., Corrêa, I.R., et al. (2014). Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. Cell Rep *7*, 1353–1361.

Wilson, V.L., Smith, R.A., Ma, S., and Cutler, R.G. (1987). Genomic 5-methyldeoxycytidine decreases with age. J. Biol. Chem. *262*, 9948–9951.

Wood, L., and Shepherd, G. (2010). Synaptic circuit abnormalities of motor-frontal layer 2/3 pyramidal neurons in a mutant mouse model of Rett syndrome. Neurobiol. Dis.

Wood, L., Gray, N.W., Zhou, Z., Greenberg, M.E., and Shepherd, G.M.G. (2009). Synaptic Circuit Abnormalities of Motor-Frontal Layer 2/3 Pyramidal Neurons in an RNA Interference Model of Methyl-CpG-Binding Protein 2 Deficiency. J. Neurosci. *29*, 12440–12448.

WYATT, G.R., and COHEN, S.S. (1953). The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. Biochem. J. *55*, 774–782.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., and Eubanks, J. (2012). Base-Resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. Cell.

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., et al. (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell *151*, 1200–1213.

Yasui, D.H., Peddada, S., and Bieda, M.C. (2007). Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes.

Yazdani, M., Deogracias, R., Guy, J., Poot, R.A., Bird, A., and Barde, Y.A. (2012). Disease Modeling Using Embryonic Stem Cells: MeCP2 Regulates Nuclear Size and RNA Synthesis in Neurons. Stem Cells *30*, 2128–2139.

Yin, R., Mao, S.-Q., Zhao, B., Chong, Z., Yang, Y., Zhao, C., Zhang, D., Huang, H., Gao, J., Li, Z., et al. (2013). Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. J. Am. Chem. Soc. *135*, 10396–10403.

Ying, S.-W., Futter, M., Rosenblum, K., Webber, M.J., Hunt, S.P., Bliss, T.V.P., and Bramham, C.R. (2002). Brain-derived neurotrophic factor induces long-term potentiation in intact adult hippocampus: requirement for ERK activation coupled to CREB and upregulation of Arc synthesis. J. Neurosci. *22*, 1532–1540.

Yu, H., Su, Y., Shin, J., Zhong, C., Guo, J.U., Weng, Y.-L., Gao, F., Geschwind, D.H., Coppola, G., Ming, G.-L., et al. (2015). Tet3 regulates synaptic transmission and homeostatic plasticity via DNA oxidation and repair. Nat. Neurosci. *18*, 836–843.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012a). Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. Cell *149*, 1368–1380.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012b). Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. Cell *149*, 1368–1380.

Zhang, R.-R., Cui, Q.-Y., Murai, K., Lim, Y.C., Smith, Z.D., Jin, S., Ye, P., Rosa, L., Lee, Y.K., Wu, H.-P., et al. (2013). Tet1 regulates adult hippocampal neurogenesis and cognition. Cell Stem Cell *13*, 237–245.

Zhao, Y.-T., Goffin, D., Johnson, B.S., and Zhou, Z. (2013). Loss of MeCP2 function is associated with distinct gene expression changes in the striatum. Neurobiol. Dis. *59*, 257–266.

Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.-N., Ho, H.-Y.H., Schmidt, L., Chen, W.G., Lin, Y., Savner, E., Griffith, E.C., et al. (2006). Brain-Specific Phosphorylation of MeCP2

Regulates Activity-Dependent Bdnf Transcription, Dendritic Growth, and Spine Maturation. Neuron *52*, 255–269.

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477–481.

Ziller, M.J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T., et al. (2011). Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. PLoS Genet. *7*, e1002389.

Zoghbi, H.Y., Percy, A.K., Schultz, R.J., and Fill, C. (1990). Patterns of X chromosome inactivation in the rett syndrome. Brain and Development.