A Neuron-Specific Histone Demethylase Complex Fine Tunes Neurodevelopment

by

Robert Porter

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Genetics and Genomics)
in the University of Michigan
2020

Doctoral Committee:

      Associate Professor Shigeki Iwase, Chair
      Professor Yali Dou
      Assistant Professor Jacob Kitzman
      Professor Ivan Maillard
      Professor Donna Martin

Robert S. Porter

rsporter@umich.edu

ORCID iD: 0000-0003-1627-4877

Dedication

To my family who have endlessly supported me and cultivated in me the values of creativity, curiosity, and persistence.

Acknowledgements

I am so grateful to be a part of the University of Michigan community and especially part of the Medical Scientist Training Program. To be a member of both the Rackham Graduate School and the Medical School has been an amazing privilege, and I have grown immensely in both environments. Thank you to the leadership of the MSTP for endless support, wisdom, and perspective including Dr. Ronald Koenig, Dr. Katy Keegan, Dr. Donna Martin, Ellen Elkin, Justine Hein, Hilkka Ketola, Laurie Koivupalo, Gretchen Aland, and Elizabeth Bowman. Thank you to the Human Genetics department for providing such a rich and supportive training environment including Dr. Sally Camper, Dr. David Burke, Dr. Anthony Antonellis, Karen

by your ideas and passion for science. Our weekly meetings have been a source of an immense amount of learning. I know I will never be able to pay back the mentorship and training you gave me, but I will certainly pay it forward as I continue in my career.

Thank you to all my friends, near and far, who have supported me through this challenging journey. Without you all, I wouldn't have been able to work through the challenges, and you have all made my life rich and full of laughter.

Finally, thank you to my family, especially my parents, Sharon and Bill Porter, and my sisters, Lauren and Lizzie Porter. You all have been a source of strength, encouragement, and revitalization during the ups and downs of this long double degree program. Thank you for always believing in me, for supporting me endlessly, and for loving me unconditionally. I owe all of this to you.

Table of Contents

List of Tables

List of Figures

Abstract

Cell-type specific transcriptional programs are thought to be defined by cell-type specific transcription factors (TFs) that bind to specific DNA sequences. In contrast, chromatin regulators are ubiquitously expressed and exert the same intrinsic activity but rely on cell-type specific TFs to engage specific genomic loci. However, mutations in chromatin regulators have emerged as a major driver of neurodevelopmental disorders, such as autism spectrum disorder and intellectual disability, challenging the ubiquitous action of chromatin regulators. Why does the disruption of widely-expressed chromatin factors lead to cognitive dysfunction?

This thesis work addresses this question through the investigation of neuron-specific isoforms of chromatin regulators. By surveying literature and publically available data, we identified 76 chromatin regulators that undergo neuron-specific microexon splicing, which led us to hypothesize that neuron-specific chromatin regulators themselves can exert unique intrinsic activity to establish the neuron chromatin landscape. We characterized one specific human disease example through RNA-Seq and follow-up validation how haploinsufficiency of PHF21A in Potocki Shaffer Syndrome patient-derived cells leads to a deficiency in the transcriptional response to stimulus.

To test this hypothesis, we investigated a neuron-specific histone demethylase complex where the histone H3K4 demethylase LSD1 and the accompanying histone reader PHF21A both undergo neuron-specific microexon inclusion. Interestingly, neuronal microexon inclusion in

LSD1 and PHF21A alters enzymatic and nucleosome-binding domains, respectively. We expressed and reconstituted neuronal and canonical LSD1/PHF21A/CoREST complexes and found that the neuronal complex exhibits reduced binding to nucleosomes and reduced H3K4me demethylation capability. To test the impact of PHF21A in neurodevelopment, we performed RNA- and ChIP-Seq as well as confocal microscopy to evaluate synapse formation in *Phf21a* knockout and rescue mouse models. *Phf21a*-KO mice have a defect in synapse number and accompanying transcriptomic changes in neuronal genes. However, animals with aberrant expression of the canonical-PHF21A isoform in neurons show an abnormally elevated synapse number. Our results demonstrate that microexon inclusion in the LSD1/PHF21A complex leads to a biochemical dampening of complex function impacting the neuronal histone methylome and thereby fine-tunes gene expression for proper synapse formation. This work illuminates how chromatin regulators can have neuron-specific forms with distinct activity to shape the neuronal chromatin landscape.

Chapter 1 - Introduction

Overview of Thesis

Multi-cellular organisms contain diverse cell types but yet each cell contains the same genome. Cell-type-specific gene expression is thought to be established by cell-type-specific transcription factors which in turn recruit ubiquitously-expressed chromatin regulators. Recent studies of the genetic basis of neurodevelopmental disorders such as intellectual disability and autism spectrum disorder have revealed an important association of chromatin regulators to neurodevelopmental disease. However, if chromatin regulators are widely-expressed, why does perturbation of genes encoding chromatin regulators lead to cognitive phenotypes?

This thesis work addresses this question through the investigation of neuron-specific isoforms of chromatin regulators. Neuron-specific isoforms of chromatin regulators can lead to novel functions that are vulnerable to disruption in neurodevelopmental disease. My work focuses on a histone reader protein, PHF21A, which is associated with Potocki Shaffer Syndrome (OMIM: 608325) and is an autism-risk gene. Potocki Shaffer Syndrome is an extremely rare disease, and previous literature is mostly limited to clinical case reports. PHF21A works in complex with the well-studied histone H3 lysine 4 (H3K4) demethylase, LSD1 (also known as KDM1A). Interestingly, both PHF21A and LSD1 undergo neuron-specific splicing.

In Chapter 1, I introduce the literature around neuron-specific chromatin regulation in the brain. In Chapter 2, I undertook the first molecular characterization of Potocki Shaffer Syndrome

through a RNA-Seq analysis of Potocki Shaffer Syndrome patient-derived samples and investigate the molecular pathology of these cells. In Chapter 3, I describe experiments that sought to establish a *Phf21a*-constituitve knock mouse model as a model for Potocki Shaffer Syndrome. In Chapter 4, I investigate how a neuron-specific splicing of histone demethylase complex leads to a biochemically distinct function and influences the formation of the neuronal transcriptome and consequently synapse formation. Finally, in Chapter 5, I discuss the broader impact of this work and suggest future experiments.

Introduction to chromatin regulators in the brain

One of the long-standing questions in genetics is how cells achieve cell-type-specific gene expression. DNA-binding transcription factors (TFs) have been determined to be the major driving force in establishing cell-type specific transcriptomes. Master TFs are often only expressed in specific cell types and bind to their cognate DNA sequence at promoters and enhancers, thereby activating or repressing gene expression in a cell-type-specific manner (Deplancke et al., 2016). In multicellular organisms, higher order structure of DNA is achieved by chromatin compaction with nucleosomes, DNA wrapped around the four core histones, as a central means for compaction (Luger et al., 1997). Nucleosomes are generally refractory to the actions of RNA polymerase II (Li and Reinberg, 2011), but a variety of chromatin modifiers can be recruited to alter the underlying chromatin structure. Writer enzymes can place chromatin modifications, eraser enzymes can remove chromatin modifications, and chromatin readers can recognize specific modifications and enact further changes. Additionally, chromatin-remodeling complexes are responsible for the movement and displacement of nucleosomes from particular genomic regions.

Unlike cell-type specific TFs, chromatin modifiers and remodelers tend to be ubiquitously expressed. More recent work has begun to reveal exceptions to this rule, whereby chromatin regulators can also exhibit cell-type specific action. For example, germ-cell-specific assembly of the preinitiation complexes (Goodrich and Tjian, 2010), neuron-specific micro-RNA circuitries (Yoo et al., 2011), and neuron-specific ATP-dependent chromatin-remodeling complexes (Staahl and Crabtree, 2013) have been shown to contribute to cell-type specific transcription. While these mechanisms all rely on the cell-type-restricted presence of transcriptional regulators, recent evidence indicates that alternative splicing of ubiquitously-expressed factors can contribute to cell-type specific transcription, in particular, within neurons.

In this introduction, we discuss current views on how alternative splicing contributes to complexity of the brain, its link to neurodevelopmental disorders, and how neuron-specific splicing events can influence the roles of transcriptional machineries. A growing amount of literature has begun to support the idea that compromised function of the neuronal isoforms of transcriptional regulators may underlie multiple neurodevelopmental disorders.

Alternative Splicing in the Brain

Alternative splicing generates multiple proteins from a single pre-mRNA by including and/or excluding alternative exons, thereby diversifying cellular proteomes. In complex organisms, such as humans, alternative splicing events are estimated to occur in 92-94% of genes (Wang et al., 2008). Throughout vertebrate evolution, alternative splicing programs are notably most complex in the nervous system (Barbosa-Morais et al., 2012; Chen and Manley, 2009; Merkin et al., 2012;

Yeo et al., 2004), suggesting that alternative splicing contributes to the complexity of brain anatomy, development, and function. Not only does the brain have a higher number of alternative splicing events relative to other tissues (Pan et al., 2008; Xu et al., 2002; Yeo et al., 2004), but conservation of the brain-specific alternative splicing program is especially prominent through vertebrate evolution suggesting functionality of spliced products (Barbosa-Morais et al., 2012; Merkin et al., 2012). Recent work has highlighted the neocortex, the center for higher-order cognitive processes, as a hotspot of alternative splicing events that influence cortical development, layering, and cell fate (Belgard et al., 2011; McKee et al., 2005; Zhang et al., 2016; Zhang et al., 2014). As will be discussed in detail below, dysregulation of this alternative splicing program leads to neurological disease (Licatalosi and Darnell, 2006).

Mechanisms and Biological Roles of Neuron-Specific Alternative Splicing Factors

Alternative splicing is coordinated by *cis*-acting RNA elements and *trans*-acting RNA binding proteins that regulate intron excision. The spliceosome is the major molecular machinery, which controls intron excision and determines which pre-mRNA sequences are to be included or excluded from the mature mRNA. The core spliceosome is a large RNA-protein complex and involves the five subunits defined by the five RNA components, U1, U2, U4, U5, and U6, and the associated small ribonucleoproteins (RNPs). A large number of auxiliary proteins help the spliceosome recognize splice sites (Chen and Manley, 2009; Li et al., 2007; Wahl et al., 2009). While most spliceosome components are constitutively expressed, tissue-specific RNA-binding proteins direct spliceosome machinery to specific splice sites to generate tissue-specific splicing patterns. Neuron-specific alternative splicing is one such example controlled by the coordinate actions of many brain-specific RNA-binding proteins. Several recent review articles have

comprehensively discussed the mechanisms of actions and roles in brain development of these splicing regulators (Lara-Pezzi et al., 2016; Raj and Blencowe, 2015; Vuong et al., 2016). Below, we provide a brief summary of the biological roles of key factors that are crucial in generating unique splicing patterns within neurons and also highlight the recent discovery of microexons. We highlight five key splicing factors, nSR100, NOVA, RBFOX family members, PTB, and Hu/ELAV family members, which have been well characterized. It should be noted that other factors including SAM68 family members, TDP-43, and MBNL, also contribute to neuron-specific alternative splicing as reviewed by others (Iijima et al., 2016; Raj and Blencowe, 2015; Yap and Makeyev, 2013).

Brain-Specific Spliceosome Recruiting Factor, nSR100

Neural-specific SR-related protein of 100 kDa, nSR100, was identified as a vertebrate and tissue-specific Serine/Arginine-repeat region containing splicing factor that activates inclusion of a large number of brain-specific exons (Calarco et al., 2009; Raj et al., 2014). nSR100 recognizes pyrimidine-rich motifs flanking the 3' splice site and binds specifically with U2-RNP components to assist in early-acting spliceosome assembly (Raj et al., 2014).

Expression of nSR100 increases upon neuronal maturation (Irimia et al., 2014). In mammalian cell culture and zebrafish models, nSR100 is required for neurogenesis and neuronal differentiation (Calarco et al., 2009; Raj et al., 2014). An nSR100 haploinsufficient mouse model has impaired neurite outgrowth, altered neuronal excitability and synaptic transmission, and behavioral abnormalities that resemble autism spectrum disorder (Quesnel-Vallières et al., 2016; Quesnel-Vallieres et al., 2015).

Position-Dependent Splicing Regulators

NOVA

Neurooncologic ventral antigen (NOVA) was the first described splicing factor that is responsible for neuron-specific exon content (Buckanovich et al., 1996; Jensen et al., 2000; Yang et al., 1998). NOVA was initially identified as an antigen produced in tumor tissues that leads to an autoimmune neurological disorder, paraneoplastic opsoclonus myoclonus ataxia (POMA) (Buckanovich et al., 1993; Luque et al., 1991). An initial survey of NOVA-target RNAs identified 34 transcripts regulated by NOVA in mice, but recent high-throughput methods suggest the regulatory network of NOVA may include as many as 700 gene transcripts (Ule et al., 2003; Zhang et al., 2010).

Compared to nSR100, NOVA plays more diverse roles in mRNA regulation. NOVA appears to control both alternative splicing (Ule et al., 2005) and selection of polyadenylation sites to generate brain-specific 3'-UTR of mRNAs through binding of YCAY clusters, which influences both U2 and U1 recruitment (Licatalosi et al., 2008). Interestingly, binding of NOVA near 5' splice sites promotes exon inclusion through U2 recruitment; however, binding of NOVA near 3' splice sites promotes exon skipping through inhibition of U1 binding (Licatalosi et al., 2008; Ule et al., 2006). The distinct actions at 5' and 3' splice sites are referred to as position-dependent control of splicing.

NOVA is expressed specifically in neurons (Buckanovich et al., 1996; Jensen et al., 2000; Yang et al., 1998), and NOVA targets transcripts encoding synaptic proteins that are important for

synaptic plasticity (Ule et al., 2003; Ule et al., 2005). In human and mouse, the *NOVA1* and *NOVA2* genes encode highly homologous proteins, and mouse reverse genetics has provided insights into their interplay. *Nova1*-null mice exhibit progressive motor dysfunction, brain stem and spinal cord neuronal apoptosis, and death 1-2 weeks after birth (Jensen et al., 2000). *Nova2*-null mice display a specific deficit in long-term potentiation of slow inhibitory postsynaptic current in hippocampal CA1 neurons (Huang et al., 2005). *Nova1/Nova2*-double null mice are born, but are completely paralyzed and die shortly after birth (Ruggiu et al., 2009). These mouse models and human genetics studies establish pivotal roles of NOVA in plasticity and development of both central and peripheral nervous systems.

RBFOX

The RNA-binding protein FOX paralogs (RBFOX1, 2, and 3) are another major set of splicing factors that increase in expression during neuronal development and promote neuronal exon inclusion. RBFOX specifically recognizes UGCAUG motifs, which are found at both 5'- and 3'-regions of introns. Similar to NOVA, RBFOX exerts position-dependent splicing control (Auweter et al., 2006; Zhang et al., 2008). RBFOX binding in 3' splice site regions inhibits exon inclusion, whereas binding in 5' splice site regions enhances exon inclusion. Such context-specific function suggests the combinatorial involvement of other splicing regulators to select for the inclusion of neuron-specific exons (Zhang et al., 2008).

Several lines of evidence have indicated important roles of RBFOX family proteins in neuronal development and function. Expression of RBFOX1 was downregulated in post-mortem brains from autistic individuals, and RBFOX1 downregulation was associated with splicing

dysregulation of genes relevant to synaptogenesis (Voineagu et al., 2011). Another study found specific regulation of a calcium channel alternative exon that alters the electrophysiological properties of this channel activation in neurons (Tang et al., 2009). Genome-wide mapping of protein-RNA interaction sites revealed that RBFOX1, 2, and 3 directly control splicing of genes that are up-regulated during brain development and whose dysregulation has been linked to autism (Weyn-Vanhentenryck et al., 2014).

Negative Regulator of Exon Inclusion, PTB

While NOVA, nSR100, and RBFOX primarily promote inclusion of alternative exons (Calarco et al., 2009; Ule et al., 2006; Zhang et al., 2008), Polypyrimidine Tract-Binding protein 1 (PTB or PTBP1) is a well-known negative regulator of exon inclusion. PTB binds CU-rich regions causing a looping-out of the RNA, which prevents assembly of the spliceosome (Oberstrass et al., 2005). Furthermore, PTB suppresses expression of its neuron-specific paralog neural-PTB (nPTB, PTBP2) by excluding an exon within nPTB, whereby the absence of this exon leads to a frameshift and degradation of nPTB mRNA by nonsense-mediated decay (Boutz et al., 2007). This inter-isoform suppression mechanism defines undifferentiated neuro-progenitors.

During neuronal differentiation, a canonical PTB is post-transcriptionally repressed, in part by decreased expression of the transcription factor REST (which will be discussed below) and subsequent increased expression of the neuronal microRNA, miR-124, which targets canonical PTB and in turn reduces its protein level (Yoo et al., 2011). This miR-124-mediated regulatory switch relieves suppression of nPTB. Expression of nPTB in turn initiates a neuronal program of alternative splicing events which is required for the differentiation of progenitors to mature

8

neurons (Boutz et al., 2007; Makeyev et al., 2007). Repression of PTB results in trans-differentiation of a variety of cell types to the neural lineage, illuminating this post-transcriptional regulatory circuit as a master key for neuronal cell fate (Xue et al., 2013; Xue et al., 2016).

Dual Roles in Alternative Splicing and Polyadenylation, Hu/ELAV

The Hu/ELAV family of splicing factors was identified in a similar way as NOVA as the autoimmune target of a paraneoplastic neurological syndrome (Szabo et al., 1991). The family consists of four proteins in mammals (HuA, HuB, HuC, and HuD; HuA is known as HuR in humans). HuB/C/D are exclusively expressed in neurons with the exception that HuB is also present in germ cells (Okano and Darnell, 1997). ELAV (Embryonic Lethal Abnormal Visual system) is the *Drosophila* homologue of Hu and analogously controls alternative splicing in nervous system development in the fly. Initially, Hu proteins were thought to bind the 3'-UTR of mRNAs and affect their cytoplasmic stability and thus the extent of translation (Jain et al., 1997). In this context, Hu proteins bind to AU-rich elements at the 3'-UTR of mRNAs and thereby stabilize them (Wang and Tanaka Hall, 2001). However, later studies revealed that Hu plays a role in alternative splicing of neuronal mRNAs including the calcitonin/CGRP transcript (Zhou et al., 2011; Zhu et al., 2006). Hu proteins compete with positive splicing factors TIA-1/TIAR leading to the interference of U1/U6 snRNP binding and thereby promote differential exon inclusion or alternative polyadenylation (Zhou et al., 2011; Zhu et al., 2008). Given that the 3'-UTR plays unique roles in mRNA stability, sub-cellular localization, and translation, Hu proteins represent a unique regulatory mechanism that may potentially coordinate mRNA metabolism and proteome diversity in neurons.

The dual roles of Hu in RNA regulation has been linked to both neuronal differentiation and plasticity. *HuD*-null mice exhibited hind limb clasping, which is associated motor and sensory neuron defects in the cortex and basal ganglia. The *HuD*-null brains were characterized with a reduced number of cortical neurons despite normal numbers of neural stem cells, suggesting that HuD is crucial for neuronal differentiation (Akamatsu et al., 2005). Various *in vivo* studies confirmed the role of Hu proteins in alternative splicing and alternative polyadenylation for genes implicated in neuronal function and disease such as *Bdnf* and *Nf1*. The differential 3'-UTR generated by Hu-mediated alternative polyadenylation has been shown to stabilize mRNAs in dendrites for local protein synthesis, implicating its roles in synaptic plasticity (Allen et al., 2013; Bronicki and Jasmin, 2013; Zhou et al., 2011).

Cooperation between Neuron-Specific Splicing Factors

Some of these neuronal splicing factors act cooperatively during neuronal development. nSR100 increases expression of nPTB and works cooperatively with nPTB to overcome PTB-mediated repression during neuronal differentiation and consequently promote neural-specific exon splicing (Calarco et al., 2009; Raj et al., 2014). Antagonistic interplay between PTB and RBFOX1 appears to be crucial for the progenitor-to-neuron transition (Zhang et al., 2016). Interestingly, several studies have also shown that recruitment of splicing machinery, such as Hu and PTB, during gene transcription can lead to altered local histone modifications that would reinforce the same pattern of exon inclusion in future rounds of transcription (Luco et al., 2010; Zhou et al., 2011). Identifying novel regulators of the neural splicing network and their genetic

associations with neurological disorders will provide further insights into the core role of splicing in neurodevelopment and brain function.

Microexons Are Enriched in the Brain Alternative Splicing Network

Recent work has illuminated the unique complexity of splicing in the brain. By comparing RNA-Seq data across diverse tissues in mouse and human, Irimia et al recently reported an inverse correlation between alternative exon size and their brain enrichment, i.e. as exons decrease in length, their enrichment in the brain increases (Irimia et al., 2014). The authors identified more than 200 neuron-specific "microexons", ranging between 3-27 nucleotides. Strikingly, many microexons display a "switch-like" inclusion or exclusion during neuronal maturation, when neurons are beginning to form synapses. The neuron-specific splicing factor nSR100 appears to be a key regulator of microexon inclusion (Irimia et al., 2014). Another study also identified a neural-program of microexon splicing and defined microexons as exons fewer than 51 nucleotides (Li et al., 2015). In contrast to the regulation by nSR100 identified by Irimia and colleagues, Li and colleagues found that intronic sequences near microexons contained RNA motifs of RBFOX and PTB proteins (Li et al., 2015). Microexons often encode domains that mediate protein-protein interactions such that target proteins gain novel binding partners and/or altered binding affinity (Buljan et al., 2012; Irimia et al., 2014). A variety of proteins harbor microexons, such as cytoskeletal proteins, ion channels, and signaling molecules, suggesting that microexons modulate a broad range of cellular processes in neuronal maturation and synaptogenesis. Intriguingly, a significant number of genes that have been implicated in autism

spectrum disorder contain microexons, suggesting that neuron-specific splicing events underlie the pathogenesis of autism (Irimia et al., 2014).

Vulnerability of the Brain to Transcriptional Dysregulation

Genetic association studies of neurodevelopmental disorders, including autism, schizophrenia, and intellectual disability syndromes, have identified numerous mutations in transcriptional regulators (De Rubeis et al., 2014; Iossifov et al., 2014; McCarthy et al., 2014; Najmabadi et al., 2011; Ronan et al., 2013). Neurodevelopmental disorders affect 1%–8% of the population and impose the leading health-care cost in the developed world (Ropers, 2010). The mutated transcriptional regulators comprise a large fraction of nuclear proteins such as DNA-binding transcription factors (TFs), histone-modification enzymes (which "write" or "erase" post-translational modifications), and their cognate "reader" proteins (De Rubeis et al., 2014; Iossifov et al., 2014; Ronan et al., 2013). In most cases, how mutations in these transcriptional regulators lead to neurodevelopmental disorders is not well understood.

Transcriptional regulation is fundamental to survival and function of all cell types, so why do mutations in transcriptional regulators influence particularly cognitive phenotypes? It can be argued that the complexity of the central nervous system requires even finer transcriptional control than other tissues; therefore, the impact of hypomorphic mutations is more strongly manifested in the brain than other tissues. An alternative but not mutually-exclusive possibility is that there might be unique characteristics of gene regulation in neurons, which confer vulnerability of the brain to mutations in transcriptional regulators.

Neuronal Isoforms of Transcription Regulators

Potential vulnerability of the brain to transcriptional and splicing dysregulations prompted us to examine chromatin regulating genes whose transcripts undergo neuron-specific splicing events. Zhang, et al, published the first brain cell-type specific transcriptomic database, which included alternative splicing events (Zhang et al., 2014). We compared their list of neuron-specific alternative splicing events to a published list of "EpiFactors" that compiles known and putative chromatin regulators (Medvedeva et al., 2015). This intersection revealed 115 chromatin regulators that exhibit neuronal alternative splicing events of a variety of types (Table 1-1). Since some of the specific genes that are discussed below are not included in this list, there may be others that were not captured by this analysis.

We also intersected the lists of microexons (Irimia et al., 2014) and EpiFactors (Medvedeva et al., 2015) and found that 76 transcriptional regulators contain neuron-specific microexons (Table 1-2). When considering the two tables together, 161 genes are represented (22% of all EpiFactors) as either containing microexons or a neuron-specific splicing event. 30 genes are represented in both tables (bolded in both tables), which likely represent EpiFactors that have neuron-specific and microexon splicing events.

Altogether, the large number of genes described in these tables suggests an important role for splicing in the neuron-specific chromatin landscape and transcriptome. Though the impact of microexon inclusion/exclusion on functions of these proteins remains largely undetermined,

some studies have begun to reveal marked impact of such splicing events in transcriptional regulation in the brain. Below, we summarize the roles of alternative splicing in DNA-binding TFs and chromatin modification writers and erasers and their implications for neurodevelopmental disorders.

DNA-Binding TFs and Neuronal Isoforms

As discussed earlier, the cell-type restricted presence of TFs can be a determinant for cell-type specific gene expression during stem cell differentiation. TFs are able to bind cognate DNA sequences and subsequently dictate target specificity of chromatin regulators which do not generally carry sequence-specific DNA binding domains. This paradigm allows for cell-type specific transcriptional programs. Although the "EpiFactor" database does not cover TFs, this group of transcriptional regulators also appears to be spliced in a brain-specific manner. The examples below represent unique cases, where TFs are expressed ubiquitously yet adopt neuron-specific forms. These examples highlight alternative splicing as a core mechanism of neuron-specific gene regulation and its crucial role in neuronal differentiation, maturation, and function.

REST/NRSF

RE-1 Silencing Transcription factor (REST, aka NRSF) was originally identified as a TF which is largely restricted to non-neuronal tissues and represses neuronal genes in non-neuronal cell types (Chong et al., 1995). REST binds to RE-1 sequences that lie in the promoter regions of neuronal genes and recruits a variety of co-repressor complexes (Bruce et al., 2004; Chen et al., 1998; Chong et al., 1995; Schoenherr and Anderson, 1995). Within non-neuronal cells, REST

directly represses nSR100 expression, and nSR100 overexpression results in REST-target gene de-repression (Raj et al., 2011). *REST* is, in fact, expressed to some extent in neurons but its function is repressed via a neuron-specific alternative splicing event that is controlled by nSR100. nSR100 acts to promote inclusion of a microexon in REST, generating the neuron-specific "REST4" isoform (Palm et al., 1999; Raj et al., 2011) (Figure 1-1 shows positions of neuronal alternative splicing events). Interestingly, the 16-nucleotide microexon carries a premature stop codon, and the resultant REST4, which lacks DNA-binding capability, acts as a dominant negative protein which sequesters full-length REST in nonfunctional hetero-oligomers (Shimojo et al., 1999). While overall REST expression decreases in neurons, alternative splicing by nSR100 acts as a fail-safe mechanism to ensure complete loss of REST function in neurons. These findings highlight the antagonistic molecular circuitry between the transcriptional repressor, REST, and splicing activator, nSR100, to generate and maintain the identity of neurons and non-neuronal cells (Figure 1-1).

SnoN

SnoN (aka SKI-like proto-oncogene, SKIL) was identified as a TF responsible for axon growth in cerebellar granule neurons, in part by promoting transcription of cytoskeletal regulators (Ikeuchi et al., 2009; Stegmuller et al., 2006). The *SnoN* gene is alternatively spliced into SnoN1 and SnoN2, where SnoN2 has an alternative splice site in exon 3 leading to a 46 amino acid deletion relative to the SnoN1 isoform (Pelzer et al., 1996). This alternative exon is near the SMAD-binding domain, with which SMAD negatively regulates SnoN function (Stroschein et al., 1999; Wan et al., 2001) (Fig 1-1). Whereas SnoN1 promotes axon branching and inhibits migration in the cerebellar cortex, SnoN2 represses branching and promotes migration (Huynh et

al., 2011). Only the SnoN1 isoform (and not SnoN2) is capable of forming a complex with

FOXO1, another TF, which rises in expression during neuronal maturation. Unlike the cell-type

specific roles of REST and REST4, both SnoN isoforms operate in neurons, but their functions

are restricted to specific layers of the cerebellum — SnoN1 is highly expressed in the inner

granular layer, whereas SnoN2 is expressed in the molecular layer of the cerebellum.

Furthermore, SnoN1, but not SnoN2, is crucial for repression of the X-linked lissencephaly gene,

DCX, in cerebellar granule neurons, providing a potential link between SnoN-mediated neural

network assembly and neurodevelopment (Huynh et al., 2011).

MEF2

Mads box transcription enhancer factor 2 (MEF2) is a TF family encoded by four *MEF2* genes

(*MEF2A, MEF2B, MEF2C,* and *MEF2D*) and is involved in nervous system development by

promoting neuronal-activity dependent transcription and negatively regulating the number of

excitatory synapses (Ataman et al., 2016; Flavell et al., 2006; Flavell et al., 2008; Lyons et al.,

2012). MEF2 family TFs undergo complex patterns of alternative splicing that involve multiple

exon inclusions/exclusions within the single *MEF2* gene. Some alternative splicing events

modulate binding affinity of MEF2 to other TFs, such as NeuroD, MRF4, and MASH1 (Janson

et al., 2001). Altered expression of specific MEF2 isoforms has been associated with myotonic

dystrophy and other neuromuscular disorders (Bachinski et al., 2010). Furthermore, three MEF2

family members (MEF2A, MEF2C, and MEF2D) have a conserved 24-nucleotide microexon

that is expressed solely in striated-muscle and neuronal cells, particularly in the cerebral cortex

(Leifer et al., 1993). This microexon is incorporated specifically during myocyte differentiation

or neuronal differentiation (Zhu et al., 2005), and in the case of neuronal differentiation, the

splicing factor nSR100 is responsible for its inclusion (Irimia et al., 2014). This microexon is adjacent to the MEF2 transcriptional activating domains, and the microexon-containing protein product is a much more potent activator of MEF2 target genes as shown through reporter assays (Zhu et al., 2005) (Fig 1-1). Mutations within or microdeletions encompassing *MEF2C* have been found in an intellectual disability syndrome associated with epilepsy, muscular hypotonia, and cerebral malformations (Le Meur et al., 2010; Zweier et al., 2010). A mouse model of *Mef2c* loss exhibited reduced neuronal differentiation and severe autism-like behavioral deficits (Li et al., 2008), reinforcing the pivotal roles of MEF2-mediated transcriptional control in brain development.

Histone Modifying Enzymes with Neuronal Isoforms

Once TFs occupy target genomic loci, they recruit histone modification enzymes to either relax or compact higher-order chromatin structure, thereby modulating accessibility of RNA polymerase machinery. Genome-wide transcriptome approaches revealed that many histone modifiers are targeted by neuron-specific splicing machineries (Tables 1-1 and 1-2). The examples below illustrate how neuron-specific splicing can have profound influences on biochemical functions of histone modifiers and potentially chromatin landscapes within neurons.

LSD1

Lysine Specific Demethylase 1 (LSD1, aka KDM1A) removes mono- (me1) and di-methylation (me2) specifically from Histone 3 Lysine 4 (H3K4), which are hallmarks of regulatory regions of transcriptionally-engaged genes (Heintzman et al., 2007; Jenuwein and Allis, 2001). LSD1 was

originally identified as a component of the CoREST corepressor complex, which is operated by REST to repress neuronal genes in non-neuronal cells (Shi et al., 2004; Shi et al., 2005).

Several groups have reported that LSD1 has a neuronal isoform (LSD1-n), which includes a 4-amino acid alternative microexon in its catalytic amine oxidase domain (Fig. 1-1). LSD1-n appears to be important for neurite morphogenesis, synaptogenesis, and proper transcriptional response to neuronal depolarization and is transcriptionally upregulated relative to canonical LSD1 as neurons begin to mature (Laurent et al., 2015; Rusconi et al., 2015; Wang et al., 2015; Zibetti et al., 2010). However, the biochemical function of the LSD1-n isoform has been debated. The first study reporting this isoform found that LSD1-n acts like the canonical protein to remove H3K4 mono- or di- methylation marks (H3K4me1/2) from a histone peptide with similar efficiency (Zibetti et al., 2010). This first study also carried out X-ray crystallography on the histone-interacting segment of LSD1-n and did not find an altered structure compared to the canonical isoform. Another study found that LSD1-n associates with a nuclear factor, Svil, which changes its substrate specificity from H3K4me1/2 to H3K9me1/2, a repressive histone modification (Laurent et al., 2015). A third group reported that LSD1-n demethylates H4K20me1/2, another mark associated with transcriptionally-repressed regions (Wang et al., 2015). These conflicting data suggest that other unknown regulatory proteins, genomic contexts, or timing in neuron maturation may, in concert, determine the substrate specificity of LSD1-n. The threonine at position 369 of LSD1-n, which is located within the microexon, can be phosphorylated, leading to conformational changes and disassembly of the LSD1/CoREST complex (Toffolo et al., 2014). Thus, the LSD1 microexon might control dynamics of complex assembly instead of, or in addition to, modulating substrate specificity.

Heterozygous missense mutations in *LSD1* have been recently implicated in an unnamed neurodevelopmental disorder with clinical features such as developmental delays, craniofacial and palate abnormalities, thinning of the corpus callosum, and hypotonia (Chong et al., 2016; Tunovic et al., 2014). Interestingly, these features resemble those of Kabuki syndrome, which is primarily associated with haploinsufficiency of MLL2/4 (KMT2D), an H3K4 methyltransferase (Ng et al., 2010), which catalyzes the writer reaction reciprocal to the LSD1 eraser reaction. *LSD1* mutations fall into the catalytic domain, and they interfere with H3K4 demethylation activity (Pilotto et al., 2016). Although the known *LSD1* missense mutations are not located in or around the alternative microexon, it is highly conceivable that these mutations also affect the enzymatic activity of LSD1-n. Given the roles of LSD1-n in neurite morphogenesis, synaptogenesis, and normal excitability, the reported mutations in *LSD1-n* could alter these parameters in patients, thereby leading to their cognitive dysfunction.

PHF21A

PHD Finger protein 21A (PHF21A, also known as BHC80) is a histone reader protein that acts in complex with LSD1 (Hakimi et al., 2002; Shi et al., 2005). Its PHD finger domain was the first discovered histone reader domain that recognizes the absence of methylation on Histone H3 Lysine 4. Compared to other PHD finger containing proteins that have aromatic residues that allow for recognition of methyl-modified histone lysines, the PHD finger of PHF21A contains several neutral and acidic residues that change its specificity to H3K4me0 (Lan et al., 2007).

We recently identified a 23 nucleotide neuron-specific microexon in PHF21A that is just 5' to the sequence encoding the PHD finger domain. Splicing of this neuron-specific microexon is mutually exclusive with the canonical exon (Fig 1-1). The discovery and characterization of this neuronal microexon will be described in Chapter 4 and is a major focus of this thesis work. Interestingly, a group studying a form of neuroendocrine prostate cancer (which aberrantly expresses the splicing factor nSR100) also identified the neuronal PHF21A isoform as a splicing product that contributes to cancer progression (Li et al., 2017). Another study analyzed the role of LSD1 in HIV transactivation found preferential association of the canonical isoform of PHF21A with LSD1 at the HIV promoter upon viral reactivation in J-Lat-A2 Jurkat immortalized T-cells (Liu et al., 2018). However, neither of these studies characterized the biochemical function of the two PHF21A isoforms on chromatin.

PHF21A is associated with Potocki-Shaffer Syndrome (PSS), a rare congenital disorder that is associated with microdeletions in chromosomal region 11p11.2. Patients with PSS are characterized by intellectual disability, craniofacial abnormalities and two bone phenotypes: parietal foramina and bone exostoses. The originally described PSS microdeletion encompassed a 2.1 Mb segment resulting in the heterozygous loss of 13 genes, including PHF21A (Potocki and Shaffer, 1996). Subsequent studies characterized patients with smaller deletions and were able to associate the two bone phenotypes to the genes *EXT2* and *ALX4*, and *PHF21A* as the gene associated specifically with intellectual disability and craniofacial abnormalities (Kim et al., 2012; Labonne et al., 2015; Mavrogiannis et al., 2001; McCool et al., 2017; Stickens et al., 1996; Wakui et al., 2005; Wu et al., 2000).

G9A

While LSD1 was discovered to remove active chromatin marks, G9A (aka EHMT2 and KMT1C) is a histone lysine methyltransferase that deposits H3K9me2, which is enriched in repressive chromatin environments (Tachibana et al., 2008). Despite the enrichment in repressive chromatin, a recent study reported that H3K9me2 can promote transcription within constitutive heterochromatin (Jih et al., 2017). A mouse model of G9a inactivation in forebrain neurons or the hippocampus led to behavioral abnormalities including learning impairment and decreased exploration (Gupta-Agarwal et al., 2012; Sampath et al., 2007; Schaefer et al., 2009). In the *G9a*-forebrain null mouse model, non-neuronal genes were de-repressed in neurons leading the authors to hypothesize that this irregular expression underlies the learning and memory impairment and other behavioral abnormalities (Schaefer et al., 2009). Although there have been no reported human cases of intellectual disability associated with G9A disruption, haploinsufficiency of the closely related paralog EHMT1 (aka GLP or KMT1D) appears to be responsible for the neurodevelopmental condition, Kleefstra syndrome (Kleefstra et al., 2006). G9A and EHMT1 form a stable heteromeric complex in many cell types (Tachibana et al., 2008), including post-mitotic neurons (Benevento et al., 2016), raising the possibility that genetic alterations of *G9A* may be linked to undiagnosed neurodevelopmental conditions.

G9A has a 33 amino acid alternative exon 10 (E10) (Brown et al., 2014). Recently, the G9A E10 isoform was identified to be upregulated upon neuronal differentiation of the mouse neural crest derived cell line, N2A (Fiszbein et al., 2016). E10 inclusion did not affect methyltransferase enzyme activity *in vitro*, but rather increased nuclear localization of the E10-containing G9A (Brown et al., 2014; Fiszbein et al., 2016). Thus, Fiszbein, et al hypothesized that the inclusion

of E10 promoted nuclear localization by better exposing a nearby nuclear localization signal. Specific siRNA-targeted knockdown of the G9A E10 isoform abolished neurite outgrowth, phenocopying the pan-isoform knockdown of G9A. Importantly, G9A E10 is not uniquely present in differentiated neurons; other non-neuronal cell types, including stimulated lymphocytes and differentiated mammary gland cells, also display increased transcription of the E10 isoform (Fiszbein et al., 2016; Martinez et al., 2012). Nonetheless, the switch-like inclusion of E10 in G9A in neurons is reminiscent of the LSD1/LSD1-n dynamics during neuronal differentiation and maturation (Zibetti et al., 2010). It is tempting to speculate that functional modulation of a histone methyl writer for an inactive mark (H3K9me) and eraser for an active mark (H3K4me) coordinates to relax or compact chromatin structure for normal development of neurocircuitries.

TAF1

TAF1 is a histone acetyltransferase component of the TFIID transcriptional initiation complex (Jacobson et al., 2000; Mizzen et al., 1996; Ruppert et al., 1993). TFIID directs RNA Polymerase II to transcription start sites. Neuronal TAF1 (N-TAF1) includes a microexon encoding 2 amino acids (A-K) close to one of the two bromodomains (Fig. 1-1) (Ito et al., 2016). Bromodomains are acetyl-histone reader modules; however, the functional impact of this insertion remains unknown. N-TAF1 is expressed as neurons begin to mature (Jambaldorj et al., 2012). Knockdown of N-TAF1 in neuroblastoma cells leads to decreased expression of genes related to vesicular transport, synapse function, and dopamine metabolism, implicating is role in synapse development and dopaminergic neurotransmission (Herzfeld et al., 2013).

A retrotransposon insertion at the *TAF1* locus has been linked to X-linked Dystonia

Parkinsonism (XLDP), which is characterized by severe torsion dystonia followed by

parkinsonism (Haberhausen et al., 1995). Interestingly, expression of the N-TAF1 isoform is

specifically reduced in patient cells suggesting that the retrotransposon could have disrupted a

neuron specific *cis*-element (Makino et al., 2007). Another study reported nine families with

point mutations in *TAF1* that led to X-linked intellectual disability along with facial

dysmorphologies, generalized hypotonia, and other variable neurological features (O'Rawe et al.,

2015). O'Rawe et al also identified two families with duplications of *TAF1* with phenotypic

features overlapping those in the individuals with *TAF1* point mutations, but it is unclear how

duplication of *TAF1* could mechanistically lead to similar consequences. With the exception of

XLDP, these genetic lesions alter both canonical and neuron-specific TAF1 isoforms. It would

be important to test whether N-TAF1 is specifically responsible for cognitive deficits by

knocking out N-TAF1 in mice.

Methyl DNA Reader with Neuronal Isoforms

In addition to histone modifications, methyl moieties placed on DNA, including CpG

methylation and a variety of non-CpG methylations, also play important roles in transcriptional

regulation in higher eukaryotes (Ambrosi et al., 2017; Edwards et al., 2017). Uniqueness of the

brain in methyl DNA regulation was illustrated by the discovery of hydroxymethylation

originally in the cerebellum (Kriaucionis and Heintz, 2009). Later investigations confirmed a

much higher level of this oxidized form of CpG methylation in the brain compared to other

tissues (Richa and Sinha, 2014). While roles of methyl-DNA regulation in neurodevelopment

and plasticity have been intensively studied recently (Bayraktar and Kreutz, 2017; Jang et al., 2017), the impact of neuron-specific splicing on methyl-DNA regulation remains largely unexplored. An exception is MeCP2, a multifunctional protein canonically known for its epigenetic silencing function through binding to methylated CpG sites (Lewis et al., 1992).

MeCP2

X-linked Methyl CpG Binding Protein 2 (MeCP2) is one of the most well-characterized chromatin regulators in the brain (Pohodich and Zoghbi, 2015). Heterozygous disruption of MeCP2 is responsible for Rett Syndrome, the progressive neurodevelopmental disorder which primarily affects young females (Amir et al., 1999). In contrast, duplication of *MeCP2* in males leads to *MeCP2* duplication syndrome, characterized by feeding difficulties, poor or absent speech, and muscle stiffness (Ramocki et al., 2010). Thus, precise regulation of MeCP2 dose appears crucial for normal brain function. MeCP2 has two isoforms, E1 and E2, that differ at the N-terminus of the protein (Kriaucionis and Bird, 2004; Mnatzakanian et al., 2004) (Fig. 1-1). The E1 isoform is expressed at a much higher level than E2 in postnatal neurons (Dragich et al., 2007; Zachariah et al., 2012). DNA CpG methylation levels in the promoter and first intron of *MeCP2* correlate with the expression of the two isoforms, suggesting the isoforms regulate themselves through DNA methylation (Olson et al., 2014). The alternate N-terminal exon lies near the Methyl-cytosine-Binding Domain (MBD), and therefore may influence methyl-CpG binding capabilities, but this hypothesis has yet to be tested.

The pan-*Mecp2*-null mouse is an established model of Rett Syndrome (Chen et al., 2001; Guy et al., 2001). The *Mecp2e2*-specific knockout mice did not display Rett syndrome-related

symptoms (Itoh et al., 2012). On the other hand, *Mecp2e1* deletion in mice recapitulated the neurological deficits observed in Rett Syndrome, suggesting that haploinsufficiency of MeCP2E1, but not E2, accounts for Rett Syndrome (Yasui et al., 2014). High levels of MeCP2E2, but not E1, in the mouse brain have been shown to be neurotoxic through functional inhibition and physical interaction with the transcription factor FOXG1, another Rett-associated gene product (Dastidar et al., 2012). Thus, in contrast to the strong implication of MeCP2E1 loss in Rett Syndrome, MeCP2E2 might be the causative agent for the MeCP2 duplication syndrome.

Perspectives

Neurodevelopmental disorders are not unique in their association with mutations in transcriptional regulators. Many chromatin regulators, including histones (Schwartzentruber et al., 2012), are somatically mutated in a variety of cancers (Allis and Jenuwein, 2016). In the context of cancers, recurrent gain-of-function mutations or translocations in transcriptional machinery, in particular, can provide proliferative advantages (Deng et al., 2013; Morgan and Shilatifard, 2015). However, many germline loss-of-function mutations of transcriptional regulators, which often have functional orthologues with redundant molecular function in the genome, appear to dominate the genetic landscape for neurodevelopmental conditions. It remains unclear why cognitive functions are particularly susceptible to mutations in genes encoding transcriptional regulators.

A number of outstanding questions remain in this field. Although many neuronal isoforms have been identified in transcriptional regulators, the impact of alternative splicing in the context of

protein complexes is not understood. As discussed earlier, the transcription factor REST and the histone demethylase LSD1 have both been shown to be present as specific isoforms within neurons. These two proteins are known to act together in a complex, canonically repressing neuron-specific genes in non-neuronal cells. However, it is unknown how the two neuronal isoforms, REST4 and LSD1-n, act together in complex, if at all. It is possible that different combinations of canonical and neuronal isoforms could influence transcription differently. Some initial studies have used genome-wide interrogation methods, such as ChIP-Seq, to ask whether neuronal isoforms have unique targets within the genome (Laurent et al., 2015; Wang et al., 2015). However, the data published so far have been contradictory across studies and have not revealed a clear pattern. An important next step is to harness experimental approaches, by which isoforms with subtle sequence differences can be separately analyzed in specific cell-types. These would include genomic-wide approaches on flow-sorted cell types and CRISPR-mediated destruction of neuron-specific exons.

A recurrent observation is the alternative splicing "switch" in maturation processes of post-mitotic neurons rather than earlier developmental processes such as proliferation of neural progenitor cells or cell-type specification. When progenitors are dividing, transcription might primarily satisfy a high demand for generating cellular materials. As cells cease to divide, transcriptional regulations are necessary to allow neurons to respond to extra-cellular cues including synaptic inputs. Given that a transcriptional response to synaptic inputs is required for synaptic plasticity, which is a basis of cognitive function, an important future direction is to determine how neuron-specific splicing events contribute to synaptic plasticity via transcriptional responses. The neuron-specific gene regulatory machineries encompassing splicing and

transcription factors might become prime drug targets for cognitive disorders for which we

currently have no therapies.

Tables
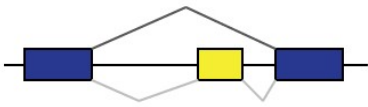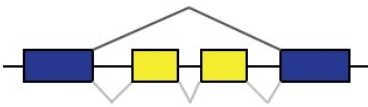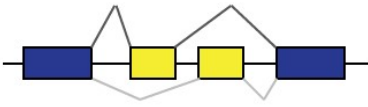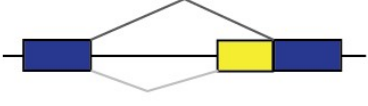
Splicing Event	Genes
Exon Skipping	*Actl6b, Adnp, Arid4b, Arntl, Arrb1, Ash2l, Atf2, Banp, Baz2b, Bptf, Brd2, Brd8, Brpf3, Carm1, Chd2, Chd5, Cit, Dpf2, Ep400, Epc2, Exosc1, Eya3, Ezh2, Gatad2a, Gse1, Gtf2i, Gtf3c4, Hdac5, Hp1bp3, Huwe1, Ino80e, Map3k7, Maz, Mecp2, Mllt1, Morf4l2, Mta1, Mtf2, Ncoa1, Ncoa2, Ncor1, Nsd1, Pbrm1, Pcgf6, Phc1, Phf14, Phf20l1, Phf21a, Pkn1, Prkaa1, Prr14, Rps6ka3, Rsf1, Scmh1, Senp1, Setd5, Sirt2, Smarca2, Smarca4, Smarcc2, Smarce1, Spen, Tlk1, Trim33, Ubn1, Uhrf2, Usp16, Vrk1, Whsc1, Yeats2, Ywhaz, Zzz3*
Tandem Exon Skipping	*Bptf, Dot1l, Hdac5, Nap1l1, Pbrm1, Phf20, Rps6ka3, Scmh1, Setd3, Smarca2, Smarce1, Vrk1, Zzz3*
Mutually Exclusive Exon Usage	*Phf21a, Setd5*
Intron Retention	*Aebp2, Asxl1, Brd9, Brms1, Brpf1, Chd4, Chd8, Exosc9, Hcfc1, Hdac7, Hdac10, Hirip3, Lrwd1, Mta2, Nfrkb, Phf1, Prkcd, Prr14, Sf3b1, Sirt7, Tle2, Trim28, Wdr77*
Alternative 3' Splice Site	*Abpp1, Brd9, Chd3, Chd4, Cit, Ehmt1, Eya1, Ezh1, Gtf2i, Mbd6, Mllt10, Ncor1, Ncor2, Nfrkb, Ogt, Prdm4, Ssrp1, Taf1, Tle4, Trim33*
Alternative 5' Splice Site	*Bptf, Chd3, Huwe1, Kat2a, Ncor1, Ncor2, Prmt1, Smarcb1, Trrap*

Table 1-1 Chromatin Regulators that Undergo Neuron-Specific Splicing

Genes that have neuron specific alternative splicing events as published by Zhang, *et al* (Zhang et al., 2014) was compared to the Epifactors list by Medvedeva, *et al* (Medvedeva et al., 2015) to produce the above list of epigenetic factors that have neuron-specific alternative splicing events. Some genes are listed twice, as they had multiple types of alternative splicing events.

Class	Subclass	Genes
Chromatin remodeling /cofactor		*ACTL6A, BPTF, CHD1L, CHD9, EP400, HP1BP3, INO80C, MAPKAPK3, MLLT1, MTA1, PHF19, PSIP1, RAD54B, SETD6, SMARCA1, SMARCC2, SMARCD2, SP100, SRCAP, UBR5*
Histone modification	Writer/Cofactor	*APBB1, ARRB1, ATXN7, AURKC, BRPF3, CARM1, CIT, DDB2, EP400, EZH2, LAS1L, MBD1, MEAF6, NAT10, OGT, PAXIP1, PHF19, PRDM2, PRKCB, PRMT1, PRMT5, RPS6KA3, SETD6, SUV420H2, TAF1, TEX10, TRRAP, UBR5, WHSC1*
	Reader	*BAZ2B, BRD8, CECR2, L3MBTL3, PBRM1, PHF20L1, PHF21A, ZCWPW1*
	Eraser	*HDAC1, HDAC10, HDAC3, HDAC6, HDAC7, KDM1A, KDM2B, KDM5B, MORF4L2, NCOR2, SIRT7, SMARCA1, SRCAP, USP16, USP21, USP49, ZMYND8*
Histone chaperone		*CHRAC1*
RNA Modification		*DND1, EXOSC4, MOV10*
Transcription Factor		*E2F6, GTF2I, MBD1*
Polycomb Group Protein		*SFMBT1, EZH2*

Table 1-2 Chromatin Regulators that Contain Microexons

Human microexon list as published by Irimia, *et al* (Irimia et al., 2014) was compared to the Epifactors list by Medvedeva, *et al* (Medvedeva et al., 2015) to produce the above list of epigenetic factors that contain microexons. Some genes are listed twice, as they fell into multiple categories.
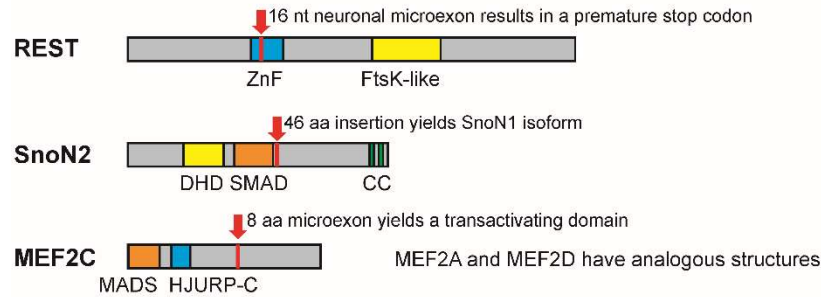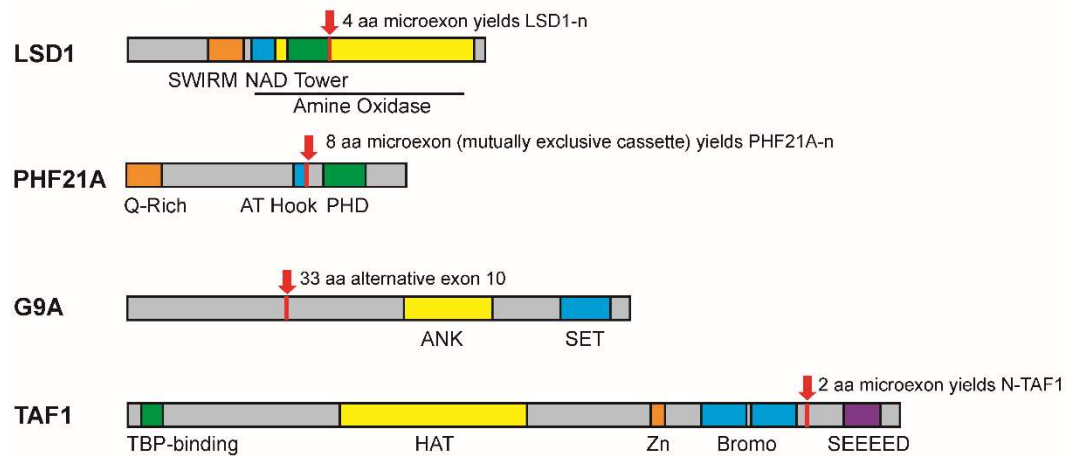
Figures
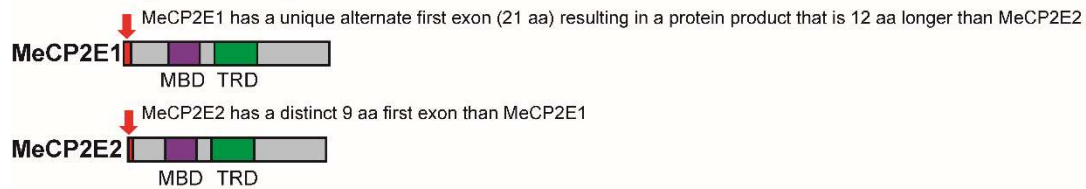
Figure 1-1 Domain organization with neuron-specific alternative exons of TFs and chromatin regulators.

Length of protein and functional domains are drawn to scale, relative to each other. ZnF: Zinc finger; DHD: Dach Homology Domain; SMAD: SMAD (SMAD refers to homologs of both the *C. elegans* SMA protein, for small body size, and the *Drosophila* MAD protein, for mothers against decaptentaplegic) -binding domain; CC: Coiled Coil domain; HJURP-C: Holliday Junction Regulator Protein family C-terminal repeat; NAD: NAD-binding domain; ANK: Ankyrin repeats; Bromo: Bromodomain; HAT: Histone Acetyl-Transferase domain; SEEEED: A conserved sequence with reference to a serine-rich region of AP3B1, a clathrin-adaptor complex; MBD: Methyl-CpG binding domain; TRD: Transcriptional Repression Domain.

Chapter 2 - Transcriptome Analysis Revealed Impaired cAMP Responsiveness in PHF21A-Deficient Human Cells

Introduction

Recent genome-wide studies that have sought the genetic basis for neurodevelopmental disorders, such as intellectual disability and autism, have implicated a large number of histone methylation regulating genes (De Rubeis et al., 2014; Iossifov et al., 2014). Histone H3 Lysine 4 methylation (H3K4me) is a histone modification associated with areas of open chromatin and is one of the most extensively regulated histone modifications in higher eukaryotes, by seven writer enzymes, six eraser enzymes, and a number of reader proteins that recognize this modification and recruit effectors (Vallianatos and Iwase, 2015; Zhou et al., 2016). Mutation in 8 out of these 13 H3K4me writers and erasers and multiple H3K4me readers leads to neurodevelopmental disorders (Vallianatos and Iwase, 2015), indicating that correct dynamic regulation of histone H3K4 methylation is critical for proper brain development and cognitive function. However, little is known about the molecular mechanisms that underlie the dynamics of histone methylation and how their function contributes to proper neurodevelopment.

PHF21A is a histone-binding protein that is associated with Potocki-Shaffer Syndrome (PSS, OMIM: 601224). PSS is a rare, congenital disorder resulting from a deletion in chromosomal region 11p11.2 (Potocki and Shaffer, 1996). PSS is characterized by intellectual disability, craniofacial abnormalities, and two bone phenotypes: multiple exostoses and parietal foramina. The original genetic lesion identified in PSS was a 2.1 Mb microdeletion, which leads to the

heterozygous loss of 13 genes (Potocki and Shaffer, 1996). Within this chromosomal region, *EXT2* and *ALX4* have been identified as the genes responsible for the bone phenotypes of PSS (Mavrogiannis et al., 2001; Stickens et al., 1996; Wakui et al., 2005; Wu et al., 2000). *PHF21A,* however, has been specifically associated to the intellectual disability and craniofacial abnormality phenotypes, since patients with genetic alterations only in *PHF21A* do not exhibit the characteristic bone malformations (Kim et al., 2012; Labonne et al., 2015; McCool et al., 2017). Although the genetic evidence linking *PHF21A* in intellectual disability and craniofacial abnormalities is compelling, the molecular mechanism by which *PHF21A* loss leads to these phenotypes has not been previously determined.

The *PHF21A* gene encodes a histone-binding protein that recognizes the absence of post-translational modifications (i.e. the lack of methylation) on histone 3 lysine 4 (H3K4me0) through its PHD finger domain (Lan et al., 2007). PHF21A is a component of the Lysine Specific Demethylase 1, Corepressor of REST (LSD1-CoREST) complex. LSD1 (also known as KDM1A) demethylates mono- or di-methylated histone 3 lysine 4 (H3K4me1/2) to repress gene transcription (Shi et al., 2004; Shi et al., 2005). PHF21A therefore binds to the reaction product of LSD1-mediated H3K4 demethylation. The LSD1-CoREST corepressor complex is recruited to the neuron-restrictive silencer element (RE-1, or NRSE) via REST and is important for mediating repression of neuron-specific genes in non-neuronal cells (Bruce et al., 2004; Hakimi et al., 2002). Previous work has shown that loss of *PHF21A* leads to the de-repression of REST target genes in non-neuronal cells (Klajn et al., 2009; Lan et al., 2007). *PHF21A* is expressed ubiquitously, but expression is highest in the brain and the testes, implicating specialized roles of PHF21A in these two tissues (Iwase et al., 2004). A mouse model of *Phf21a* homozygous loss

led to neonatal lethality due to a defect in suckling (Iwase et al., 2006); however, structural and/or cytoarchitectural abnormalities have yet to be identified in the brain. It remains elusive if PHF21A plays any roles outside the repression of neuron-specific genes in non-neuronal cells.

In this study, we performed RNA-Sequencing (RNA-Seq) of *PHF21A*-deficient patient-derived cells to probe the molecular dysfunction associated with heterozygous loss of *PHF21A* in an unbiased manner. Our bioinformatic analyses and reporter assays identified cAMP signaling as an impaired molecular pathway in the *PHF21A*-deficient patient cells, thereby providing insights into the cellular role of PHF21A and how *PHF21A* loss may contribute to cognitive defects.

Experimental Procedures

Patient-Derived Cell Lines

Patient blood samples were collected from the individuals as described previously (Kim et al., 2012; Labonne et al., 2015). Lymphocytes were harvested and then transformed by Epstein-Barr Virus into lymphoblastoid cell lines as described previously (Nishimoto et al., 2014). Lymphoblast cell lines were maintained in RPMI medium 1640 (Gibco) containing 10% FBS, 1x GlutaMax (Gibco), and 1% penicillin and streptomycin (Gibco).

RNA-Sequencing

RNA was isolated from lymphoblast cell lines in technical duplicates, where each cell line was collected from two separate culture dishes, using the RNA purification kit from Life

Technologies. Poly-A mRNA was separated from total RNA with the NEB Magnetic mRNA

Isolation Kit. Libraries were prepared using Direct Ligation of Adapters to First-strand cDNA as

previously described (Agarwal et al., 2015). Multiplexed libraries were pooled in approximately

equimolar ratios and were purified from a 1.8% TBE-agarose gel. The libraries were sequenced

to a length of single-end 50 bases using an Illumina HiSeq 2000 according to standard

procedures. Reads were mapped to the human genomes (hg19) using bowtie2, allowing up to 2

mismatches and only uniquely mapped reads were analyzed further. Aligned reads of technical

replicates from each individual were merged using samtools for all subsequent analyses. Merged

files were then converted to bigwig files for visualization in the Integrated Genome Viewer.

Merged files were also analyzed for differential expression using DESeq between sex-matched

patient and controls (Anders and Huber, 2010). Genes called as significantly misregulated were

calculated by creating a merged differential expression file that averaged the fold change and

calculated an average p-value using Fisher's method (p_average = p1*p2*[1-log(p1*p2)]).

Genes that were up-regulated in one comparison but down-regulated in the other comparison

were excluded. Pathway analysis was run using LR-Path (Sartor et al., 2009). Network analysis

was carried out by separating down- and up-regulated GO terms and using REVIGO (Reduce

and Visualize Gene Ontology) (Supek et al., 2011), and network representation was generated

using Cytoscape (Cline et al., 2007).

shRNA-mediated PHF21A knockdown

Lentiviral packaging plasmids and shRNA constructs were transfected into 293T cells using

Transit293 transfection reagent. shRNA plasmids, including a scramble and three *PHF21A*

targeting shRNAs (V2LHS_135304, V2LHS_135309, V2LHS_162572) were obtained from the

pGIPZ microRNA-adapted shRNA library (GE Life Sciences). Viral supernatants were collected and concentrated using LentiX (ClonTech). Control male lymphoblast cells were transduced with the lentiviruses containing scramble or *PHF21A* shRNA constructs, and cell lines were then selected with 1μg/ml puromycin for 4 days. RNAs were harvested from cells using PureLink RNA Mini Kit from Ambion. cDNAs were prepared by RevertAid RT Reverse Transcription Kit (Thermo Scientific) and then were analyzed by qPCR (Applied Biosystems 7500 Instrument). The oligonucleotide sequences used for qPCR are available upon request.

Luciferase Assays

Three CRE sequences (both the consensus sequence: TGACGTCA, as well as a mutated sequence: TGATATCA) were tandemly inserted upstream of the HSV-TK promoter in a pGL3-based Luciferase plasmid using Gibson assembly. 10,000 HEK293T cells in a 96-well dish were transfected by Lipofectamine 2000 (Thermo Fisher) with 100 ng of the CRE-TK-Luciferase constructs, 1 ng of a CMV-Renilla construct, and 100 ng of either scramble or *PHF21A* shRNA. Two days later, cells were exposed to 30 μM Forskolin or an equal volume of DMSO for 8 hours. Then, cells were harvested for Luciferase analysis using the Promega Dual Luciferase Assay System. The ratio between luciferase and renilla expression was normalized to the empty plasmid (TK-Luciferase only), and then Luciferase expression was reported relative to the DMSO, scramble shRNA condition.

Stimulation of patient derived cells

The two patient derived lymphoblast cell lines, as described above, and one control lymphoblast cell line were incubated with 30 μM Forskolin or an equal volume of DMSO for 0 minutes, 30

minutes, 2 hours, or 6 hours. Then, cells were harvested for RNA extraction and qRT-PCR as described above. mRNA expression was reported relative to the untreated cells (0 minutes), and mRNA expression from forskolin treated cells was normalized to DMSO treated cells.

Results

Transcriptome analysis of patient derived cells with PHF21A alterations

To interrogate the genome-wide gene expression changes in *PHF21A* deficiency, we performed RNA-Seq analysis on two patients with *PHF21A* alterations and two unaffected controls. The male patient, DGDP262, was recently published as one of the smallest microdeletion cases of PSS-related developmental delay (Labonne et al., 2015). The female patient, MCN1762, has a balanced translocation with the breakpoint within the *PHF21A* gene leading to a truncation of 13 out of 18 *PHF21A* coding exons (Kim et al., 2012) (Figure 2-1A). Lymphoblasts were derived from these two *PHF21A* haploinsufficient patients and two unaffected, unrelated individuals. Then, cDNA libraries of poly-adenylated mRNAs were prepared and subjected to high-throughput sequencing. At least 17 million uniquely-mapped reads were obtained per sample and inter-replicate variability was low (Figure 2-1B) indicating the sufficient coverage and reproducibility of sequencing data.

Consistent with the heterozygous *PHF21A* alterations, 58.4% and 49.7% of *PHF21A* expression was observed in the male and female patient cells compared to the control cells (Figure 2-1A and 2-1C). The male patient cells have a ~234 kb microdeletion that encompasses five genes, including *PHF21A* (Labonne et al., 2015). Our RNA-Seq data confirmed this finding with an

approximately 50% reduction of mRNA levels of genes mapped to this microdeletion region (Figure 2-1D). The female patient cells contain a translocation in the fifth intron of *PHF21A* (Chr. 11) with Chromosome 1 (t(1:11)(p21.1; p11.2)) (Kim et al., 2012). The translocation breakpoint on Chromosome 1 is located within a gene desert, mostly composed of LINE-1 elements, with the nearest gene 635 kb downstream (*PRMT6)*. We did find ectopic transcripts generated from the Chr. 1 translocation breakpoint, which is likely attributed to read through of RNA polymerase II from the translocated *PHF21A* promoter (Figure 2-1E). Our data corroborate the previously-reported genetic lesions in the patient cells and found local and patient-specific changes in transcripts, which are associated with each genetic alteration.

We next sought to identify commonly misregulated genes in the two patient lymphoblast cell lines. Differentially expressed genes (DE genes) were determined by DESeq (Anders and Huber, 2010) for each patient lymphoblast line compared to the sex-matched control. We found 1,885 genes (7% of 26,463 total annotated genes) that were misregulated in both of the two patient cell lines with a p-value < 0.05 (Figure 2-2A).

Given that PHF21A is a component of the LSD1-CoREST corepressor complex, we speculated that PHF21A haploinsufficiency may lead to up-regulation of gene transcription. However, we found that roughly an equal number of genes are up- (49.4%) and down-regulated (50.6%) in patient lymphoblasts (Figure 2-2B). The magnitude of misregulation, judged by median fold change, of up-regulation (log2-fold change: 0.715) was slightly higher than the magnitude of down-regulation (log2-fold change: -0.549, Figure 2-2C). DE-genes were distributed uniformly throughout lowly- and highly- expressed genes (Figure 2-2D). We then asked if the published set

of 2,172 REST-target genes (Bruce et al., 2004) are significantly misregulated in the patient lymphoblasts. We found that although REST-target genes were over-represented in the set of misregulated genes (chi-square test, p-value = 1.06×10^{-5}), these REST-target genes were equally distributed between up- and down-regulated (Figure 2-2E) similar to the total DE-genes. Overall these data suggest that PHF21A has bidirectional roles in maintaining expression levels of REST-target genes and a large number of non-REST targets.

Pathway analysis suggests down-regulation of cAMP-signaling genes

To obtain insights into biological processes influenced by *PHF21A* haploinsufficiency, we applied our RNA-Seq data to LR-Path, a gene set enrichment ontology program that takes into account statistical significance and direction of differential expression in the entire RNA-Seq data set (Sartor et al., 2009). Interestingly, although we analyzed lymphoblasts, the most significantly downregulated pathways were relevant to neural development and function, such as "cerebral cortex neuron differentiation", "chemical synaptic transmission, postsynaptic", "visual learning", and "cortical actin cytoskeletal organization" (Table 2-1 and Figure 2-3A-B). At a FDR<0.05, many more pathways were found to be significantly downregulated than up-regulated (Figure 2-3A-B). LR-Path provides signature genes that have been well-characterized in their roles in a given biological process, thereby contributing to the significant enrichment. An example of signature down-regulated genes in our analysis are Amyloid Precursor Protein (*APP*) and Neuropilin and Tolloid-like 1 (*NETO1*) which have been shown to play important roles in NMDA-receptor trafficking required for synaptic plasticity and learning (Cousins et al., 2013;

Ng et al., 2009). *APP* and *NETO1* were down-regulated 4-fold and 37-fold on average, respectively, in the patient cells compared to the control cells (Figure 2-2A).

It is noteworthy that two cyclic AMP- (cAMP) related pathways were among the top down-regulated categories because cAMP signaling plays pivotal roles in broad physiological processes including learning and memory (Kandel, 2001; West et al., 2002). cAMP signaling elicited by external stimuli ultimately leads to phosphorylation of the transcription factor, CREB, and the expression of CREB target genes that carry out cellular responses to these stimuli (Mayr and Montminy, 2001; Zhang et al., 2005). In the patient cells, key genes, whose products mediate cAMP signaling, were down-regulated. For example, the down-regulated genes *CRTC3* (CREB regulated transcription coactivator 3) and *FOSL1* (Fos-like 1 subunit) have well characterized roles as a CREB co-activator (Conkright et al., 2003) and a CREB-responsive inducible gene (Benito and Barco, 2015), respectively. Thus, we next sought to determine if the CREB target genes, which are down-stream targets of cAMP signaling, are misregulated in our RNA-Seq data sets. We first obtained 4,084 putative CREB target genes that were identified in a previous study based upon the presence of evolutionarily-conserved CRE sequences in proximal gene promoters (Zhang et al., 2005). We found that the putative CREB target genes were over-represented in our set of DE-genes in patient cells (chi-square test, p-value = 2.96 x 10^{-10}). Slightly more genes were down-regulated (53.2%) than up-regulated (46.8%) in the patient lymphoblasts (Figure 2-3C). These results suggest that *PHF21A* haploinsufficiency leads to reduced expression of genes that mediate cAMP signaling, yet the ultimate consequence in the expression of CREB-target genes can be bidirectional in unstimulated lymphoblasts.

PHF21A knockdown in lymphoblasts recapitulates the gene expression changes in the patient-derived cells

We next sought to confirm that the key gene expression changes we observed in our RNA-Seq analysis were due to reduced expression of *PHF21A* rather than any other confounding factors such as inter-person variation. To this end, we quantified expression of selected DE-genes upon shRNA-mediated knockdown of *PHF21A* in the control lymphoblast cells derived from an unaffected individual. Male control lymphoblast cells were transduced with lentivirus carrying either scramble (sc) shRNA or one of three independent shRNAs against *PHF21A*. After establishing cell lines that stably express the shRNAs, we harvested RNA and performed qRT-PCR analysis. We first validated efficient knockdown, yielding 8%-38% of *PHF21A* expression compared to the sc-transduced cells. We found that several of the most significantly misregulated genes in the patient cells show similar changes, including *APP* and *OXTR*, upon *PHF21A* RNAi by the three independent shRNAs (Figure 2-4A). We also analyzed a number of genes that mediate cAMP-signaling, namely *NETO1, CRTC3,* and *FOSL1,* and found that they were all down-regulated upon *PHF21A* knockdown (Figure 2-4A). Finally, we chose two REST target genes, one that was down-regulated in the patient RNA-Seq data (*SCN3A*) and one that was up-regulated in the RNA-Seq data (*MAP1B*). In line with previous work on the effect of *PHF21A* knockdown in cancer cell lines and non-neuronal tissues (Iwase et al., 2006; Klajn et al., 2009; Lan et al., 2007), both genes were up-regulated upon *PHF21A* RNAi (Fig. 4B). While patient lymphoblasts lack half *PHF21A* expression constitutively, the RNAi knockdown depletes *PHF21A* in a relatively short time window; therefore, down-regulation of some REST-targets, including *SCN3A,* in the patient lymphocytes may reflect an indirect consequence of long-term

PHF21A deficiency throughout development. These data provide support to the role of PHF21A in promoting expression of cAMP-signaling genes and also suggest that some gene misregulation in the patient lymphoblasts could be the result of indirect effects.

PHF21A is required for the optimal transcriptional response mediated by cAMP-signaling

The down-regulation of cAMP-signaling mediators (Table 2-1) and misregulation of CREB-target genes in patient cells (Figure 2-3C) prompted us to test whether the induction of the cAMP-mediated gene transcription is altered by *PHF21A* deficiency. To do this, we designed a reporter plasmid with three cyclic-AMP responsive elements (CRE: TGACGTCA) upstream of the firefly *luciferase* gene, which is linked to the HSV *thymidine kinase* (TK) promoter. As a negative control, we mutated two critical nucleotides in the CRE sequence (mCRE: TGATATCA) (Tinti et al., 1997). We transfected these luciferase constructs and scramble or *PHF21A* shRNAs into HEK293T cells and then elicited cAMP signaling by treatment of forskolin, a cAMP analog (Zhang et al., 2005). In this luciferase reporter assay, *PHF21A* knockdown did not change the basal activity of the CRE-Luciferase construct (Figure 2-5A). However, upon forskolin stimulation, cells with two independent *PHF21A* shRNAs were not able to induce expression of CRE-luciferase as highly as the scramble shRNA-treated cells ($p < 0.05$, ANOVA, Figure 2-5A), and another shRNA showed a similar trend (shRNA-3). The construct with the mutated CRE sequence did not respond to forskolin nor did it change upon *PHF21A* knockdown, demonstrating that this effect is cAMP/CREB-dependent (Figure 2-5B). These results demonstrate that PHF21A is required for the transcriptional response to cAMP-mediated signaling.

Our RNA-Seq analysis of *PHF21A*-deficient patient cells captured a snapshot of the steady-state transcriptome of these cells. Given our luciferase data following forskolin stimulation, we evaluated the response kinetics to cAMP signaling in the patient cells. We exposed control and *PHF21A*-deficient patient cells to forskolin, collected cells at different time points, and measured mRNA levels of two immediate early genes (IEGs) by qRT-PCR. *c-FOS* is a classic IEG that exhibits rapid transcriptional induction following stimulation with a peak response between 30-60 minutes (Sheng and Greenberg, 1990), whereas *FOSB* is a related gene that shows a delayed response (Kovacs, 1998). Whereas control cells displayed the peak of *c-FOS* expression 30 minutes after treatment, patient cells were unable to increase transcription of *c-FOS* until two hours (Figure 2-5C). The delayed response gene, *FOSB,* rose and fell in expression gradually in the control cells, but only began to rise in mRNA level at six hours in the patient cells (Figure 2-5D). These data suggest that PHF21A is required to elicit a rapid transcriptional response to cAMP signaling.

Discussion

Genetic evidence has associated *PHF21A* with the pathogenesis of the intellectual disability and craniofacial abnormalities in Potocki-Shaffer Syndrome (PSS), but previous work has not studied the mechanism or molecular pathogenesis underlying these phenotypes. The present study is the first to describe the molecular dysfunction associated with heterozygous loss of the histone-binding protein, PHF21A, which is implicated in PSS-related cognitive deficit. Through an RNA-Seq study of *PHF21A*-deficient patient-derived cells, we found that pathways relevant to brain development and learning, including the cAMP-signaling pathway, were downregulated.

Moreover, we demonstrated that *PHF21A* is required for full induction of a CRE-Luciferase reporter gene and that *PHF21A*-deficient cells exhibit a delayed transcriptional response to cAMP signaling. Altogether, these results suggest that *PHF21A*-deficient cells are unable to mount the proper response to external stimuli.

cAMP-mediated signaling is one of the primary pathways utilized by cells in response to extracellular stimuli, such as depolarization of neurons by afferent sensory inputs (Kandel, 2001). The experiments carried out in this study used non-neuronal cells, but it is tempting to speculate that these findings may extend to neurons. Recent literature has implicated activity-dependent signaling in the pathogenesis of neurodevelopmental disorders (NDDs) (Chahrour et al., 2012; Cohen et al., 2011; Ebert and Greenberg, 2013; Morrow et al., 2008); a common feature of NDDs may therefore involve an inability for neurons to respond to sensory inputs and form the proper neural networks during brain development. cAMP signaling is a major pathway in neurons that establishes memory, which involves long-term potentiation of synaptic efficacy (Benito and Barco, 2015; Bourtchuladze et al., 1994; Kandel, 2001; West et al., 2002). Down-regulation of cAMP-mediated signaling pathways could partially explain reduced expression of genes relevant to neuronal signaling, such as the CREB-target gene, *NETO1*. Our data suggest that PHF21A is required for maintaining expression of cAMP signaling molecules, thereby ensuring the proper response to extracellular stimuli.

LSD1, a histone demethylase, which physically associates with PHF21A and generates a binding substrate for PHF21A has been found to directly interact with CREB and other transcription factors that are necessary for neuronal activity-response pathways, such as SRF (Rusconi et al.,

2016a; Wang et al., 2015). PHF21A, together with LSD1, may therefore be directly engaged in transcription of neuronal activity-dependent genes. Importantly, mutations in LSD1 have also been reported in a rare NDD. Loss-of-function missense mutations of *LSD1* lead to a Kabuki-like syndrome characterized by intellectual disability, thinning of the corpus callosum, hypotonia, and craniofacial dysmorphisms (Chong et al., 2016; Pilotto et al., 2016; Tunovic et al., 2014). Thus, future investigations should address whether cognitive deficits associated with *PHF21A* and *LSD1* share impaired response to neuronal activity via cAMP signaling as a core mechanism.

There is controversy in the literature regarding the role of *PHF21A* as a coactivator or corepressor of transcription. *PHF21A* knockdown in cells causes de-repression of REST target genes, in line with the role of PHF21A as part of the repressive LSD1-CoREST complex (Klajn et al., 2009; Lan et al., 2007). On the other hand, biochemical studies have shown that PHF21A inhibits LSD1-mediated demethylation of H3K4me2 *in vitro*, suggesting that PHF21A may dampen the repressive function of the LSD1-CoREST complex (Shi et al., 2005). A previous genetic screen in *C. elegans* identified CoREST and LSD1 orthologues, *spr-1* and *spr-5,* respectively, as suppressors of the developmental defects observed in the *sel-12*/presenilin mutant worms (Eimer et al., 2002; Jarriualt and Greenwald, 2002). In humans, mutations in the presenilin genes, *PSEN1* and *PSEN2,* encoding intramembrane proteases, and their substrate, amyloid precursor protein (*APP*) gene are the major genetic basis for the severe cognitive decline in familial Alzheimer's disease (Brunkan and Goate, 2005). The LSD1-CoREST complex may therefore play evolutionarily-conserved roles in repressing presenilin/APP pathway genes, thereby contributing to normal cognitive development and function. Contrary to the repressive roles of LSD1 and CoREST on presenilin expression, our patient RNA-Seq data showed

44

significant down-regulation of both *APP* and the human presenilin, *PSEN2* upon reduction of PHF21A level. Of note, invertebrates, including *C. elegans,* do not appear to carry an orthologue of *PHF21A* (Lakowski et al., 2006), while *PHF21A* orthologues are present broadly among vertebrates. These observations lead us to speculate that PHF21A may have evolved in vertebrates as a new component of the LSD1-CoREST complex to negatively tune the action of that complex, thereby promoting transcription of specific genes.

In conclusion, this work identified reduced expression of cAMP signaling genes as a molecular signature in PSS-related developmental delay and provided insights into the etiology of this condition. Modulation of cAMP-signaling could be a novel therapeutic target for the cognitive disorders stemming from dysfunction of PHF21A and LSD1, and more broadly, dynamic regulation of H3K4 methylation.

Tables

GO-Biological Processes Pathway	P-Value	Direction of misregulation	Signature Genes
Visual Learning	7.49×10^{-9}	Down	APP, NETO1
Visual Behavior	1.43×10^{-8}	Down	APP, HOXA1, NETO1
Myelin Maintenance	2.89×10^{-8}	Down	MYRF, EPB41L3
cAMP-Mediated Signaling	7.52×10^{-8}	Down	PCLO, CRTC3, RIMS2
Associative Learning	1.38×10^{-7}	Down	APP, NETO1
Cortical Actin Cytoskeletal Organization	2.75×10^{-7}	Down	EPB41L3, FMNL2
Regulation of cAMP-Mediated Signaling	3.09×10^{-7}	Down	CRTC3, PEXL5, RGS2
Positive Regulation of G2/M Transition of Mitotic Cell Cycle	4.25×10^{-7}	Down	APP, CCND1
Protein Localization to Synapse	6.20×10^{-7}	Down	NETO1
Learning	6.28×10^{-7}	Down	FOSL1, NETO1, APP
Endoplasmic Reticulum Calcium Ion Homeostasis	1.05×10^{-6}	Down	PSEN2, CAMK2D, APP
Suckling Behavior	1.13×10^{-6}	Down	OXTR, APP, PHF21A
Cyclic-Nucleotide-Mediated Signaling	1.37×10^{-6}	Down	CRTC3, RIMS2, PEX5L
Regulation of Voltage-Gated Calcium Channel Activity	1.54×10^{-6}	Up	AHNAK

Table 2-1 Most Significantly Misregulated Gene Ontology Pathways (Biological Processes)

LR-Path Analysis of *PHF21A*-deficient patient RNA-Seq compared to control samples (called by Gene Ontology-Biological Processes terms). Shown are the top 14 most significantly misregulated pathways, ranked by most significant p-value, as reported by LR-Path.

Figures
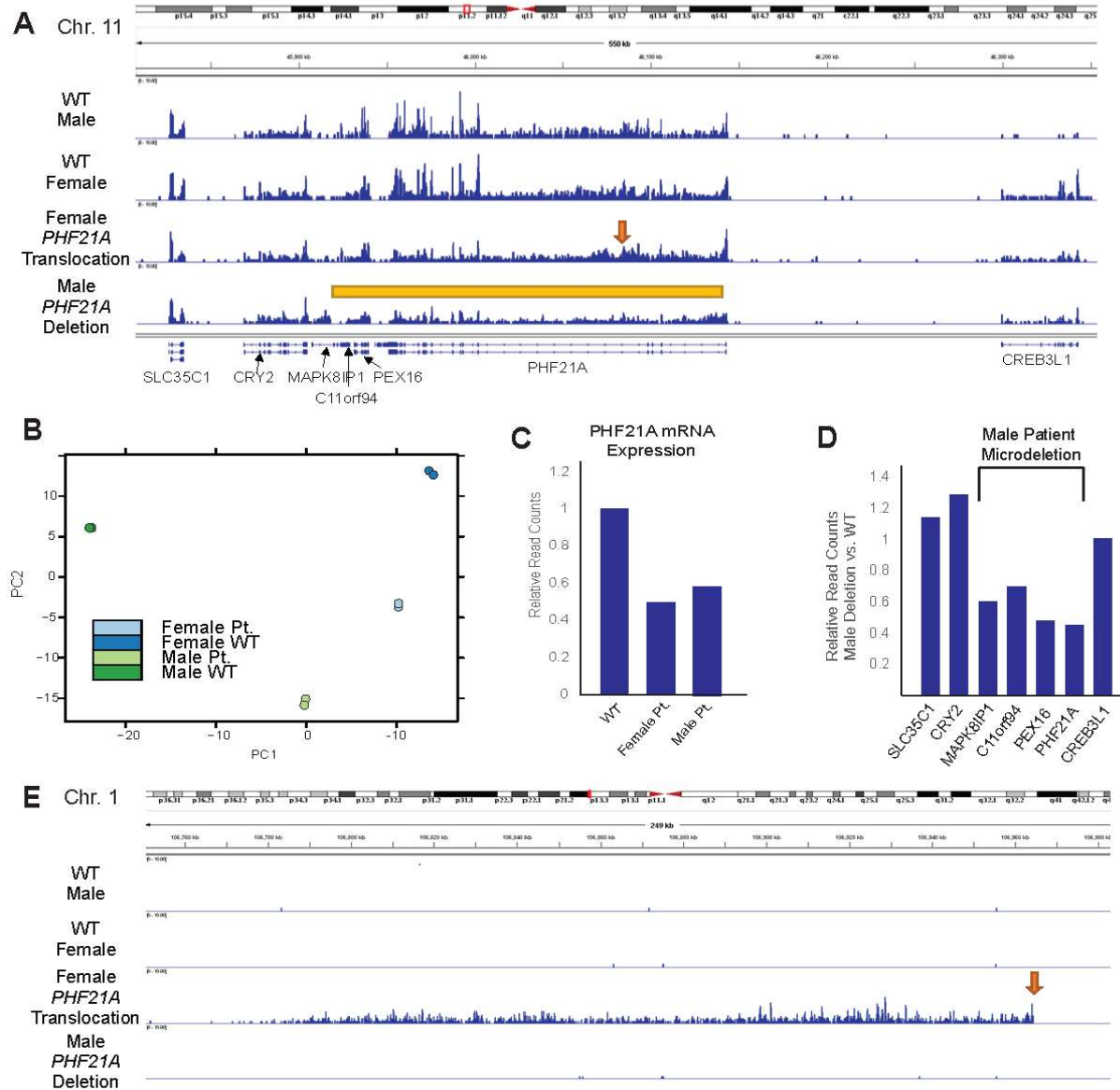
Figure 2-1 RNA-Seq analysis of PHF21A-deficient patient-derived lymphoblasts.

A. Genome browser shot of the *PHF21A* locus shows decreased read density from the male patient (heterozygous deletion) and the female patient (translocation) compared to the unaffected individuals. The gold bar represents the microdeletion in the male patient. Arrow: translocation breakpoint. **B.** PCA plot showing the clustering of each of the samples with technical duplicates along two principle components. Each technical duplicate clusters tightly with the other. Male vs. female and WT vs. *PHF21A*-deficient also cluster together in the plot. **C.** Relative gene read counts of *PHF21A* calculated by DESeq. **D.** Relative gene read counts of genes in the microdeletion region for the male patient calculated by DESeq. **E.** The gene desert region on Chromosome 1 is the recipient region of the translocation found in the female patient. Ectopic transcripts are detected likely due to the *PHF21A* promoter being fused to this region.
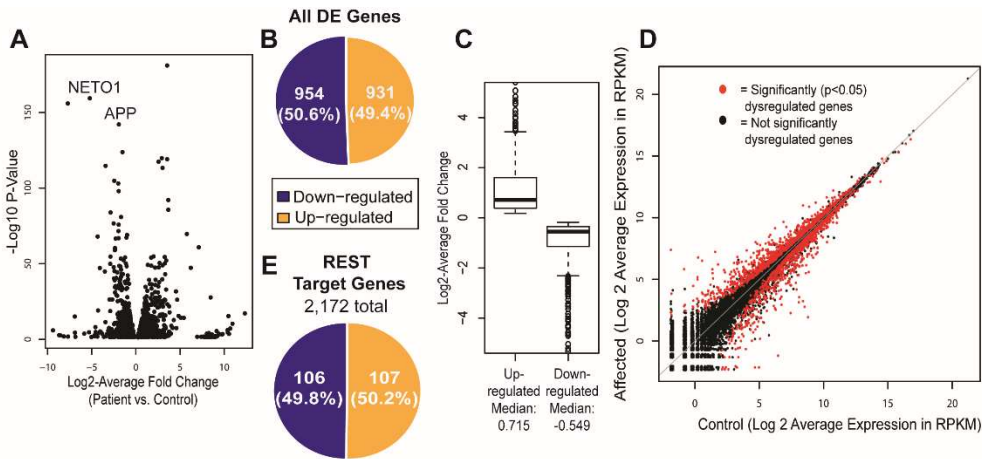
Figure 2-2 Commonly misregulated genes in PHF21A-deficient cells.

A. DESeq reveals 1,885 misregulated genes (significantly misregulated genes defined as a p-value < 0.05, after determining average non-adjusted p-value through Fisher's method). Volcano plot profiles –log10 p-value and log2 fold change of gene expression between *PHF21A*-deficient vs. control samples. **B.** and **C.** A similar number of genes (B) are up-regulated and down-regulated with similar magnitude of change (C). In the box plot, whiskers represent 1.5 times the interquartile range (IQR) with the median. **D.** Scatter plot of gene expression changes in the *PHF21A*-deficient cells. Average log2 read counts per gene of the two *PHF21A*-deficient samples and the two unaffected samples are plotted on the Y- and X-axes respectively. Differentially expressed genes with a p-value <0.05 by DESeq are in red. **E.** 213 REST target genes are misregulated with a similar number up- and down-regulated.
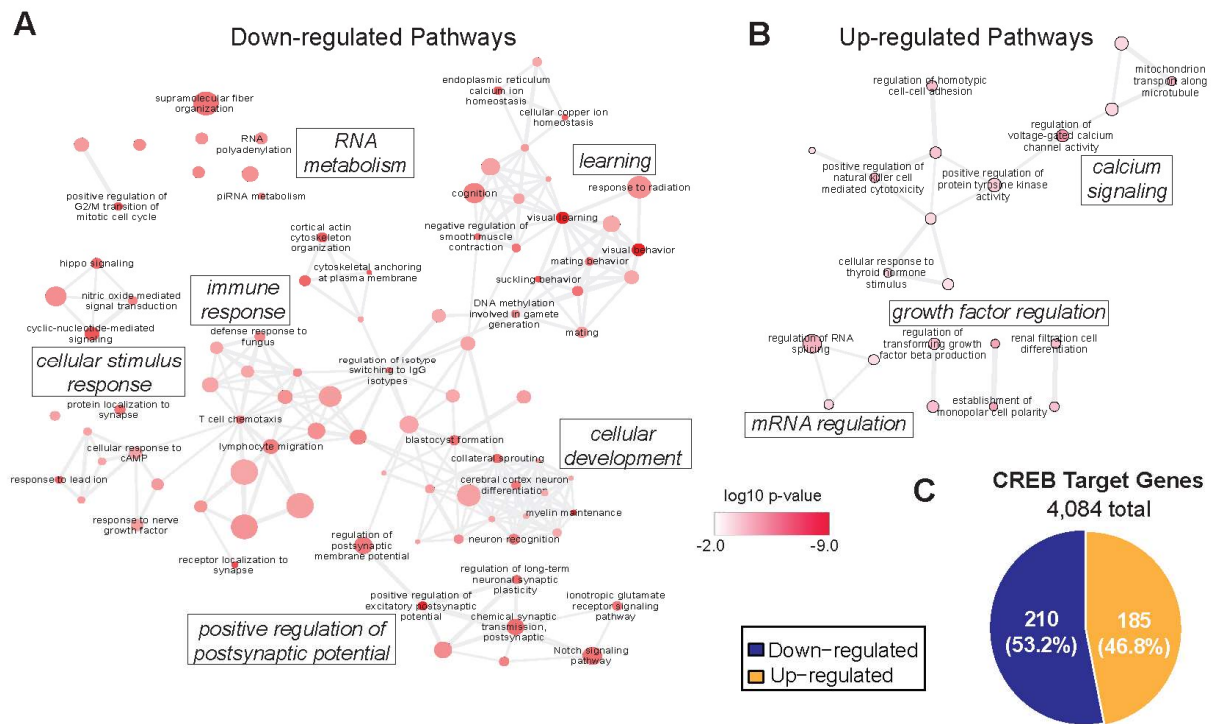
Figure 2-3 Pathway analysis shows down-regulation of processes important to neurodevelopment and function.

A and B. Cytoscape network representations (Cline et al., 2007) of significantly down-regulated (A) and up-regulated (B) pathways (FDR<0.05) and their connections based upon commonly-shared signature genes. The significance of the GO term misregulation is denoted by color. The size of each circle represents the number of genes in the ontology term (e.g. a large circle denotes a GO term with many genes, whereas a small circle denotes a more specific GO term with fewer genes). The larger terms in boxes represent main clusters identified by REVIGO (Supek et al., 2011). **C.** 395 CREB target genes are misregulated with slightly more of them being down-regulated.
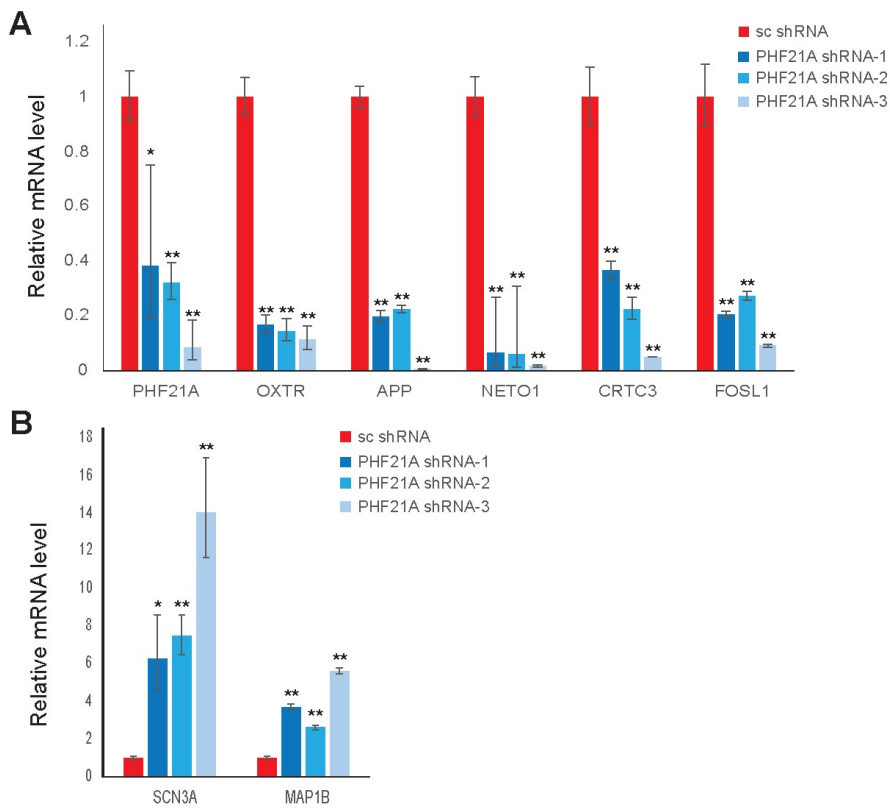
Figure 2-4 qRT-PCR analysis of PHF21A shRNA knockdown in lymphoblasts.

A. *PHF21A* shRNA leads to downregulation of *APP* and *OXTR,* two genes shown to be down-regulated in the RNA-Seq analysis. Several genes relevant to cAMP signaling are also downregulated upon *PHF21A* knockdown using three independent shRNAs. **B.** Expression of two REST-target genes, *SCN3A* and *MAP1B*, upon *PHF21A* knockdown. Error bars represent ±SEM of a technical triplicate (N=3) of qRT-PCR analysis. *p-value < 0.05, **p-value < 0.01. Significance was calculated using ANOVA with Tukey post-hoc analysis between the scramble shRNA samples and each *PHF21A* shRNA sample.
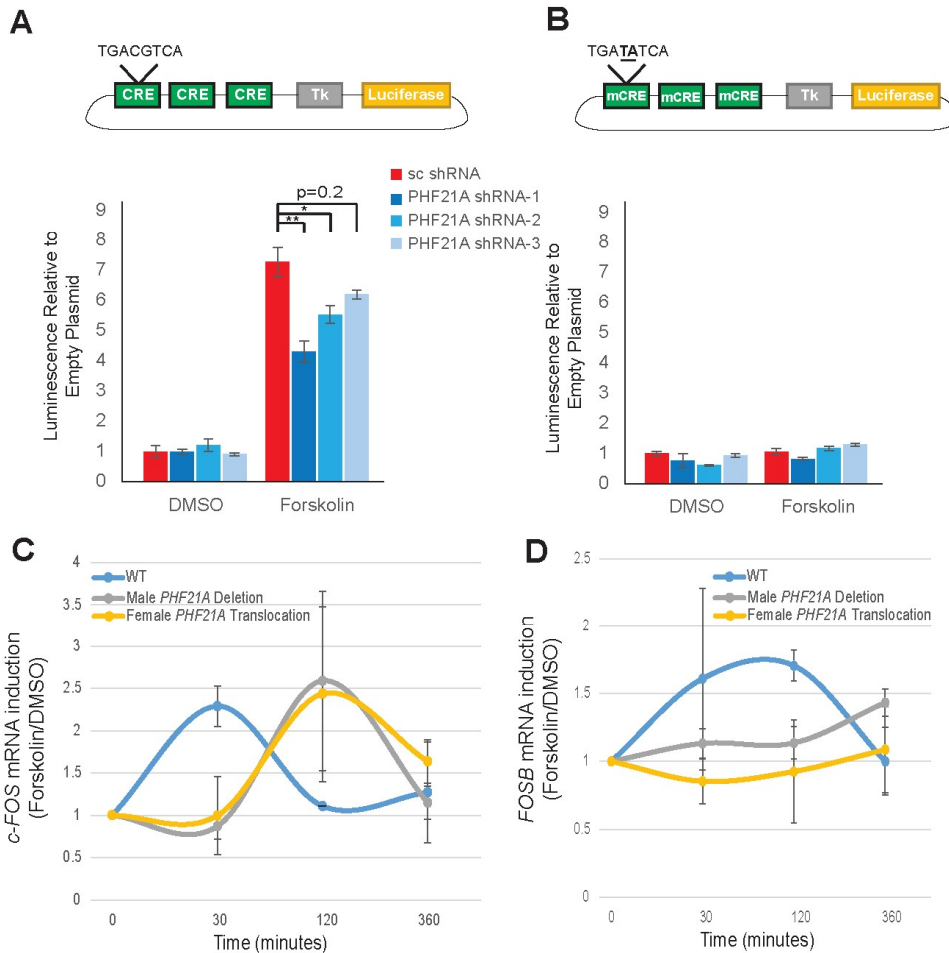
Figure 2-5 Luciferase reporter assays and transcription of IEGs following forskolin stimulation demonstrate PHF21A's roles in the transcriptional response to the cAMP-mediated signaling pathway.

A. HEK293T cells were co-transfected with a luciferase construct with CRE sequences inserted upstream of the HSV-TK promoter and expression plasmids for *PHF21A*- or scramble shRNAs. After 8 hours of 30 μM forskolin treatment, the luciferase activity was lower upon *PHF21A* knockdown compared to scramble shRNA (*p-value < 0.05; **p-value < 0.01 as calculated using ANOVA with Tukey post-hoc analysis between the scramble shRNA samples and each *PHF21A* shRNA sample). **B.** The same experiment was performed using the mutated (mCRE) reporter. Forskolin treatment nor *PHF21A* shRNA knockdown induced increased expression of the mutated CRE-luciferase construct. Error bars represent ±SEM of a biological triplicate (N=3) obtained from cells grown in independent wells. **C and D.** Patient-derived lymphoblastoid cells were stimulated with forskolin or DMSO for 30 minutes, 2 hours, or 6 hours and *c-FOS* (C) and *FOSB* (D) mRNA levels were quantified by qRT-PCR. Shown is mRNA induction by forskolin normalized to DMSO relative to unstimulated cells (zero hour time point). Error bars represent the range of data from two independent experiments.

51

Notes and Acknowledgements

Ethics Approval

This study was approved by the Institutional Review Board of Augusta University. All contributors declare no financial conflicts of interest.

Accession Numbers

Raw and processed sequence data files are available on the Gene Expression Omnibus (GEO) under accession GSE94587.

Chapter 3 - Mouse Model of Potocki Shaffer Syndrome

Introduction

Potocki Shaffer Syndrome is a rare genetic disease caused by the haploinsufficiency of PHF21A. To better understand PHF21A and its role in both neurodevelopment and how loss of PHF21A can cause Potocki Shaffer Syndrome, we sought a mammalian model of PHF21A deficiency. A constitutive germline *Phf21a*-knockout mouse was previously created (Iwase et al., 2006). This mouse model has a neomycin cassette incorporated into the seventh exon leading to a premature stop codon and complete loss of Phf21a protein. The initial study of this mouse model identified few phenotypes relevant to Potocki Shaffer Syndrome. *Phf21a*-null animals die neonatally and are notable for an inability to suckle milk. *Phf21a*-null mice do not have any obvious defects in brain structure, neuronal migration, or apoptosis. *Phf21a*-null neurons also have normal axonal branching and elongation. A survey of other organ structure and histology, including lungs, liver, and kidneys appeared normal in *Phf21a*-null mice as well.

As a consequence of the neonatal phenotype, much of the previous analysis was done right at P0; however, recent evidence has implicated the role of synapse maturation in neurodevelopmental disorders such as intellectual disability and autism spectrum disorder (De Rubeis et al., 2014; Iossifov et al., 2014; Nelson and Valakh, 2015; Rubenstein and Merzenich, 2003; Satterstrom et al., 2020). Thus, we sought to investigate post-natal phenotypes at time points after synapse maturation had already occurred. One option to address this question is analyze *Phf21a*-heterozygous animals that are viable into adulthood. A second option is to harvest embryonic

neurons and allow them to mature in culture. In the next chapter, we'll discuss experiments of synapse formation in culture.

In this present chapter, we explore how mature *Phf21a*-deficient neurons respond to stimuli as well as a number of behavioral paradigms between WT and *Phf21a*-heterozygous adult mice. In the previous chapter, we uncovered an intriguing phenotype of delayed induction of immediate early gene induction upon stimulation in *PHF21A*-deficient lymphoblasts. Recent literature has demonstrated the role of LSD1 and particularly the neuronal isoform of LSD1 (LSD1-n) as important for the recruitment of stimulus-responsive transcription factors and consequent expression of immediate early genes (Rusconi et al., 2016a; Rusconi et al., 2015; Wang et al., 2015). We hypothesized that Phf21a loss in our mouse model could lead to similar effects, and we further hypothesized that we may see different effects in neurons compared to non-neuronal cells based on expression of different isoforms of *Phf21a*.

Methods

Forskolin Stimulation of Cortical Neuron Cultures

Timed pregnant female mice were sacrificed at day E16.5 for dissection and culturing of cortical and hippocampal neurons. Briefly, cortices and hippocampi were microdissected, treated with 2.5% Trypsin (Invitrogen 15090), quenched by FBS, and then treated with 1% DNase (Sigma DN-25) to dissociate brain tissue into neurons. Cells were counted and then plated in Neurobasal Media (Gibco 21103049) with 1x B27 (Gibco 17504-044), 0.5 mM GlutaMax (Gibco 35050061), 25 uM beta-mercaptoethanol (Gibco 21985), and 1% Penicillin/Streptomycin (Gibco 15140122). Before plating, plates were coated with PDL (Sigma P7886) overnight at 37C and

then washed with ddH2O three times. Neuron culture cells were fed every 3-4 days by replacing half of the above media with fresh media and were grown until day *in vitro* 14. Then, cells were treated with 30 uM forskolin or DMSO as a control and incubated for 0, 0.5, 2, or 6 hours. Then, RNA was harvested and purified using the PureLink RNA Mini Kit (Thermo 12183018A) and converted cDNA by RevertAid Reverse Transcriptase (Thermo K1621) according to the manufacturer's instructions. qRT-PCR using Power SYBR Green PCR Master Mix (Thermo 4367659) was used to determine the relative level of immediate early gene expression relative to a Tbp housekeeping control gene. The following primers were used: Fos: F: 5'-TACTACCATTCCCCAGCCGA-3', R: 5'-GCTGTCACCGTGGGGATAAA-3'; FosB F: 5'-CGACTTCAGGCGGAAACTGA-3', R: 5'-TTCGTAGGGGATCTTGCAGC-3'.

Behavioral Tests

All of the behavioral tests were performed by the Michigan Mouse Metabolic Phenotyping Core, except for the Contextual Fear Condition test, which was done in collaboration with Dr. Natalie Tronson's lab. Procedures are described briefly below. Unless otherwise indicated, all behavioral tests were performed on adult (~4 months old) male mice with 12 animals from each genotype.

Open Field Test

Mice were placed in an open field and the total distance traveled along with the time spent in the outer, intermediate, or center parts of the arena were measured. The trial lasted 30 minutes, and two trials were performed on different days. This is a test of anxiety and motor activity.

Elevated Plus Maze Test

In this test, mice are placed on an elevated platform. Two arms of the platform are enclosed, whereas the other two arms of the platform are open. The time spent in the open arms is measured. This is a test of anxiety and motor activity.

Y-Maze Test

In this test, mice are able to explore three different arms of a "Y-Maze." Alternations indicate that mice enter a different arm from the previous one they entered. This is a test of spatial learning and memory.

Morris Water Maze

Each mouse was given 4 trials per day for 4 days to find the hidden platform with a maximum time of 60 sec before gently pushing the mouse to the platform. The time in sec for each mouse was averaged for the 4 trials on each day. The mice then underwent a probe trial on the fifth day consisting of a 60 sec swim with the platform removed from the tank. This is a learning and memory test.

Novel Object Recognition Test

In the training phase, mice are placed in a cage with two objects. The testing phase is 24 hours later, and mice then return to the same cage with one object changed to a novel object. The time spent with the familiar and novel objects is measured. In the second testing phase, the object is

moved to a different part of the cage and the time spent with the familiar and novel objects is measured again. This is a short term learning and memory test.

Contextual Fear Conditioning

Mice were allowed to habituate to the housing room for 1 week. Then, they were moved to the context A, allowed to explore for 3 minutes and then received a 0.8 mA shock during the last two seconds. The next day, around the same time, animals were placed back into the Context A for another 3 minutes. Locomotion and freezing activity were measured. This is a test for fear-associated learning and memory. For this test, we used 5 animals each of male and female *Phf21a*-WT and *Phf21a*-heterozygous animals.

Skeletal Staining

E18.5 embryonic mouse heads were skinned and then fixed in 95% ethanol for 24 hours. Fresh 95% was added and incubated for an additional 24 hours. Then, heads were moved to acetone for three days. Next, heads were moved to dye solution (3 mg Alizarin Red, 9 mg Alcian Blue in 42 ml 70% ethanol plus 3 ml acetic acid) for 4 days with agitation. Heads were then rinsed in distilled H2O for two hours to rinse off the excess dye. Then, heads were incubated in 1% KOH for 1-2 days to clear. Then, the heads were incubated in increasing concentrations of glycerol (20%, 50%, 80%, and 100%) each for 5-7 days. Stained skulls were imaged using a high-resolution scanner.

Results

Immediate Early Gene Induction in Phf21a-deficient neurons

To follow up on previous results showing a defect in immediate early gene expression in *PHF21A*-deficient patient-derived lymphoblasts (Porter et al., 2017b), we repeated this experiment in cells from our *Phf21a*-KO mouse model. We grew *Phf21a*-WT, *Phf21a*-heterozygous, and *Phf21a*-KO cortical neurons until maturity at day *in vitro* 14 and then stimulated with 30 uM forskolin or DMSO as a control. Forskolin is a potent stimulator of adenylate cyclase which causes increased levels of intracellular cAMP (Alasbahi and Melzig, 2012). We next harvest RNA, reverse transcribed to cDNA, and measured the relative amount of immediate early gene expression by qRT-PCR.

We interestingly observed a decrease in Fos and FosB expression at base line in *Phf21a*-heterozygous and *Phf21a*-KO neurons (Figure 3-1). This decrease could represent a difference in network activity or synapse number. However, upon stimulation, there was no difference in the induction of Fos or FosB across genotypes (Figure 3-1). Additionally, we observed that even after 6 hours of treatment, Fos and FosB levels did not decline back to baseline.

This experiment was conducted in the same way that it was conducted with the patient-derived lymphoblasts; however, the results of this experiment revealed differences between a suspension culture of lymphoblasts and cortical neurons. Given the lack of return to baseline of Fos and FosB expression after 6 hours of stimulation in neurons, we hypothesize that 30 uM forskolin was an inappropriately high concentration of the drug. Likely, these neurons were over-stimulated, and the persistently high expression level of Fos and FosB reflects the continued

hyper-excitation of these cultured neurons. In the future, it would be interesting to repeat this experiment using lower concentrations of Forskolin. This would permit observation of the rise and return to baseline of immediate early gene expression. Additionally, a lower dose of forskolin could permit detection of differences between *Phf21a*-genotypes. The baseline difference suggests that the network activity of *Phf21a*-deficient cultures is different, but in the present experiment we were unable to detect a difference upon stimulation.

Another important caveat of this experiment is that Forskolin is soluble in DMSO, and therefore DMSO is the control reagent used. However, cells also exhibit induction of immediate early gene expression in response to DMSO treatment. We reasoned that a different cAMP-inducing agent would be a superior drug in this experiment to eliminate background stimulation from DMSO. We performed several experiments using a cAMP analog, Dibutyryl cAMP (Sigma D0627), which is a water soluble drug that can permeate the cell membrane. This drug would allow water to be the control. However, in several trials using 0.5 or 1 mM dibutyryl cAMP, we were unable to detect induction of Fos or FosB immediate early gene expression (data not shown). We also performed several trials of both Forskolin and dibutyryl cAMP treatment in *Phf21a*-WT and *Phf21a*-KO MEFs, but we were unable to detect a difference between drug-treated and control samples (data not shown). Further optimization of this experiment is required.

Behavioral Analysis of Phf21a-deficient mice

In line with the intellectual disability phenotype seen in Potocki Shaffer Syndrome patients, we would anticipate cognitive and behavioral changes in a mouse model of *Phf21a* loss. Previous mouse models of similar neurodevelopmental disorders have shown such phenotypes as impaired

learning and memory, either reduced or increased anxiety and stress phenotypes, and altered aggression or social dominance (Iwase et al., 2016; Rusconi et al., 2016a; Silverman et al., 2010). Although the *Phf21a*-null mice die neonatally, *Phf21a*-heterozygous animals survive and we were able to compare *Phf21a*-WT and *Phf21a*-heterozygous animals in behavioral paradigms. This experimental strategy also mirrors Potocki Shaffer Syndrome patients who exhibit heterozygous loss of PHF21A.

We began with the Elevated Plus Maze and Open Field tests which both test motor activity and anxiety in mice. In the Elevated Plus Maze, total distance moved and velocity were equal across WT and *Phf21a*-heterozygous animals indicating the motor activity is unchanged upon *Phf21a*-deficiency (Figure 3-2). Time spent in the open arms of the elevated plus maze is associated with decreased anxiety in mice. The number of entries into the open arm as well as the amount of time spent in the open arms was equal between WT and *Phf21a*-heterozygous animals (Figure 3-2). In the Open Field Test, mice tend to stay along the walls of the field. Mice with less anxiety will spend more time in the intermediate and center zones of the field. In measuring the amount of distance covered in each zone in 5 minute increments, there was a statistically significant increase in the amount of time that *Phf21a*-heterozygous animals spend in the intermediate and center zones (Figure 3-3). There was a no statistically-significant change in the total distance traveled in *Phf21a*-heterozygous animals compared to WT animals; however, there did appear to be a non-significant trend toward *Phf21a*-heterozygous animals having more total distance moved (Figure 3-3). These data suggest that in some assays, but not others, *Phf21a*-heterozygous animals appear less anxious than their WT littermates.

Next, we performed several tests of learning and memory. The Y-maze test probes learning and memory by measuring how often mice 'alternate' through the arms of the maze, which indicates they remember where they had previously been as they navigate the maze. We found no difference in the percent of alternations between *Phf21a*-heterozygous and WT animals (Figure 3-4). The Morris Water maze includes four days of training where mice learn that the platform is in the southeast quadrant of the tank. *Phf21a*-heterozygous and WT animals exhibited an equal learning curve in swimming toward this quadrant (Figure 3-5). On the fifth day, the platform is removed and time spent in the correct (southeast) quadrant is measured as an additional metric of learning and memory. Again, there was no difference between *Phf21a*-heterozygous and WT animals in this test (Figure 3-6). Finally, in the novel object recognition test, we found that neither the WT nor the *Phf21a*-heterozygous animals exhibited a preference for the novel object (Figure 3-7). The test administrator noted that the mice did not interact with the objects very much in this assay which accounts for a poor test result since not even the WT animals exhibited a preference for the novel object. Given that the animals did not interact with the novel object to a great extent, we also calculated the discrimination index. The discrimination index is defined as the frequency of time exploring the novel object divided by the total exploration time. This metric normalizes for the fact that some mice in this assay did not explore the cage very much. In this metric, there was a trend toward the *Phf21a*-heterozygous animals having a lower (more negative) discrimination index (Figure 3-8). Interestingly, a more negative discrimination index could indicate an avoidance or fear of a novel object(Vogel-Ciernia and Wood, 2014). In total, these three learning and memory tests indicate that there is no learning or memory impairment in *Phf21a*-heterozygous animals.

Finally, we performed a Contextual Fear Condition test on both male and female WT and *Phf21a*-heterozygous animals. We again found no differences in locomotion between animals before administration of a shock (Figure 3-9). On the test day, we did not identify any differences in freezing behavior between WT and *Phf21a*-heterozygous animals in either males or females (Figure 3-10) indicating that there is not a difference in fear-associated learning and memory. However, the test administrator noticed a profound hyperactivity on the test day of the *Phf21a*-heterozygous animals (Figure 3-11). There were several instances of *Phf21a*-heterozygous animals jumping around the cage on the test day (video not shown). It would be interesting to investigate whether this phenotype is reproducible or associated with some change in anxiety in the mice.

Skeletal Staining of Phf21a-deficient mouse skulls

Although the majority of this work focuses on the neurodevelopmental disorder that is associated with Potocki Shaffer Syndrome, another major phenotype associated with Potocki Shaffer Syndrome is craniofacial abnormalities. Specifically, humans with Potocki Shaffer Syndrome exhibit microcephaly, brachycephaly, mid-facial hypoplasia, downturned mouth, narrow nose, hypoplastic mandible, and eye abnormalities including severe myopia, nystagmus, and strabismus (Hamanaka et al., 2018; Kim et al., 2012). One study administered a Phf21a morpholino to zebrafish and found that these fish exhibited small heads, a defect in lower jaw growth, and distorted Meckel's and palatoquadrate cartilage development (Kim et al., 2012).

Our *Phf21a*-KO mouse model exhibits a neonatal lethality phenotype that is associated with a lack of milk in the neonatal bodies indicating that KO mice are unable to suckle for an unknown

reason (Iwase et al., 2006). The original study did not identify a reason for this inability to suckle, but we hypothesized that a craniofacial abnormality could contribute to this suckling defect. We sacrificed E18.5 animals and use Alizarin Red bone staining and Alcian Blue cartilage staining to visualize bone structure and development. We imaged the skulls from the inferior, superior, and lateral aspects (Figure 3-12). We first quantified basic metrics of the skulls including the rostral-caudal length, mandibular length, inter-occipital distance, and inter-squamosal distance. We found that none of these metrics changed across *Phf21a*-WT, *Phf21a*-heterozygous, and *Phf21a*-KO animals, with the exception of a single *Phf21a*-KO animal that had a complete lack of jaw development (Figure 3-13).

We next looked for fusion of the palatine bone, which is expected to be complete by E15.5 in mouse development (Bush and Jiang, 2012). We found failure of palate fusion at E18.5 in 1/4 *Phf21a*-WT animals, 1/7 *Phf21a*-heterozygous animals, and 3/5 *Phf21a*-null animals (Figure 3-14). Although we identified increased penetrance of a palatal fusion phenotype in *Phf21a*-null animals, this is likely not the sole cause of the suckling failure observed since the phenotype is not 100% penetrant. Nonetheless, more animals should be analyzed for skeletal phenotypes as there does appear to be a difference between genotypes. Thus, our *Phf21a*-KO animal may be a means to further analyze craniofacial abnormalities as a model for Potocki Shaffer Syndrome.

Discussion

In this work, we explored several phenotypes relevant to Potocki Shaffer Syndrome pathology in our constitutive *Phf21a* knock out mouse model. This work yielded some interesting preliminary results that merit follow up. Stimulation of *Phf21a*-deficient mature cortical neurons in culture

resulted in equal induction of immediate early genes such as *Fos* and *FosB*. This result is difficult to interpret because the data suggest that the concentration of forskolin used was too high to detect subtle differences in immediate early gene induction. We did identify, however, a reproducible difference in the baseline levels of these immediate early genes between genotypes. This baseline difference prompted us to analyze formation of synapses which will be described in the next chapter. Next steps for this experiment would include optimizing this protocol for MEFs and testing different concentrations of forskolin in neuron stimulation. Such experiments would be important to understand the relevance of the phenotype seen in lymphoblasts as well as to understand how PHF21A interacts with LSD1 in its regulation of immediate early genes (Porter et al., 2017b; Rusconi et al., 2016b).

The behavioral testing we performed comparing *Phf21a*-heterozygous and WT adult animals revealed no impairment in learning and memory. Some tests such as the Open Field Test and Contextual Fear Condition showed an interesting trend of decreased anxiety in the *Phf21a*-heterozygous animals, but it is also difficult to know if this trend is due to a difference in activity of the animals. There are several caveats to these experiments. One is that the animals used for this study were the initial F0 generation produced by *in vitro* fertilization using *Phf21a*-heterozgous sperm. Animals generated through IVF have been reported to have behavioral alterations including anxiety and impaired spatial learning (Ecker et al., 2004). A superior strategy would have been to use F1 animals, especially F1 animals crossed with another *Mus musculus* strain to remove deleterious effects from inbred C57/B6 animals. Therefore, it is difficult to conclude whether the constitutive *Phf21a*-knock out animals are a model for behavioral abnormalities in Potocki Shaffer Syndrome.

At this point, we still do not understand the mechanism of neonatal lethality of *Phf21a*-null mice. The lack of milk in the abdomens of P0 *Phf21a*-null mice strongly suggests that an ability to suckle milk contributes to their neonatal death. However, a suckling defect could originate from structural, neuroendocrine, behavioral, or other abnormalities. We found that an increased number of *Phf21a*-null animals have failure of palatal bone fusion at E18.5. However, given the incomplete penetrance of this phenotype, this abnormality is likely not the sole contributor to neonatal lethality of *Phf21a*-null mice. Nonetheless, given the human craniofacial phenotype and evidence of jaw and palatoquadrate cartilage malformation in zebrafish treated with a Phf21a morpholino (Kim et al., 2012), this experiment in mouse should be repeated to increase the sample size and more carefully measure various craniofacial bone and cartilage structures.

As will be described in the following chapter, the two isoforms of PHF21A are mutually exclusively expressed such that only neuronal PHF21A (PHF21A-n) is expressed in neurons. Future work will also include careful dissection of the contribution of PHF21A-n to neurodevelopment. The compatibility of the two isoforms in neurons vs non-neuronal cells is another crucial question: what are the transcriptomic and physiological consequences of canonical PHF21A (PHF21A-c) expression in neurons?

The scope of this work allowed for some preliminary analyses of mature neuron function in *Phf21a*-deficient animals. The next chapter describes synapse formation and gene expression changes in *Phf21a*-deficient and WT animals. However, it would be informative to use the *Phf21a*-knockout mouse to test other phenotypes that have been analyzed in other models of

neurodevelopmental disorders caused by chromatin regulators, including LSD1. Such experiments include more careful behavioral tests including stress paradigms such as novelty suppressed feeding and autism-associated paradigms such as marble burying. More sophisticated microscopic methods such as Golgi staining or neuronal lineage tracing methods would refine our understanding of how PHF21A contributes to specific aspects of brain development. Mouse genetic tools such as Cre-mediated excision of PHF21A would allow us to probe the function of PHF21A in specific cell-types or at specific times in development. Finally, electrophysiological measurement of Phf21a-deficient neurons would give insight to the contribution of PHF21A to neuron function.

Figures
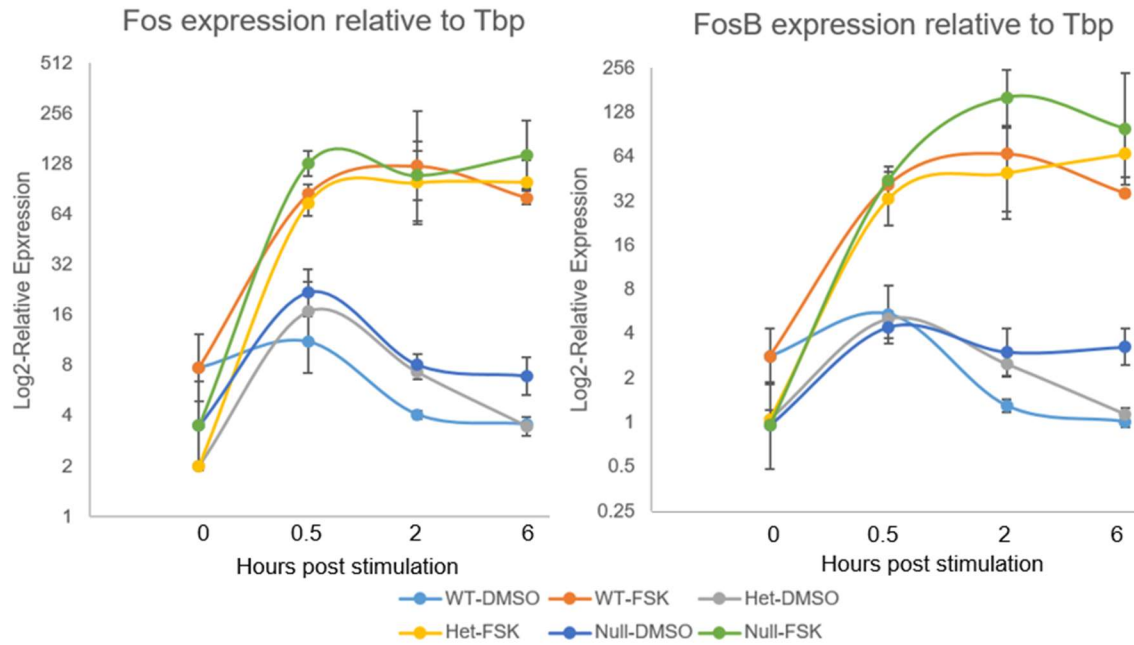
Figure 3-1 Forskolin Stimulation in Phf21a-deficient neurons.

Day *in vitro* 14 cortical neurons from *Phf21a*-WT, –heterozygous (Het), and –null animals were stimulated with 30 uM Forskolin or an equal volume of DMSO. RNA was harvested after 0, 0.5, 2, or 6 hours of stimulation. Fos and FosB expression was measured by qRT-PCR relative to Tbp expression. Error bars represent the average of three biological replicates.
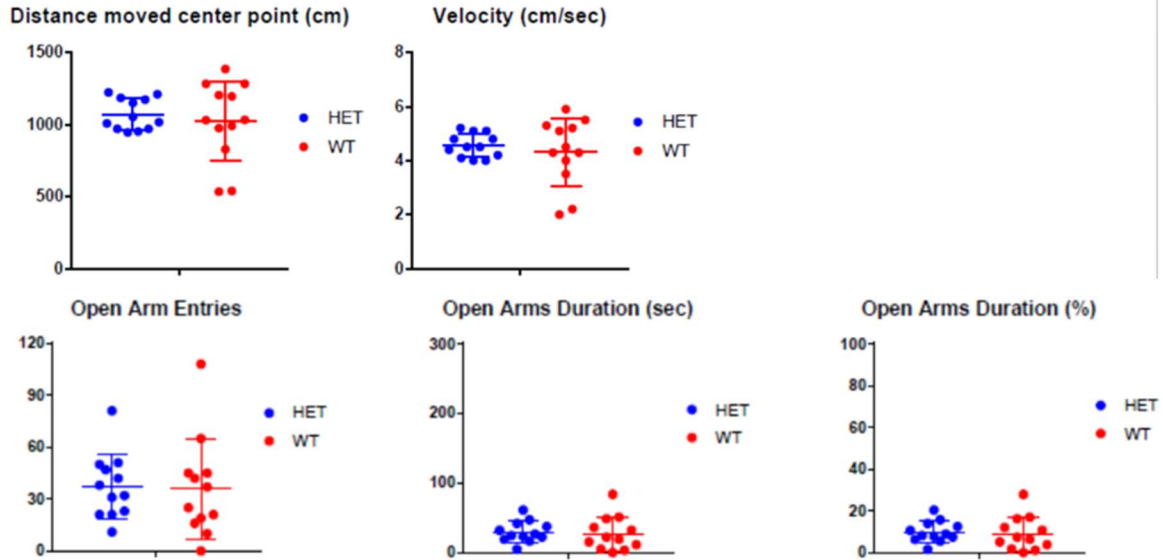
Figure 3-2 Elevated Plus Maze.

12 *Phf21a*-heterozygous and WT animals were placed in an elevated plus maze and the above metrics were measured over a period of 300 seconds.
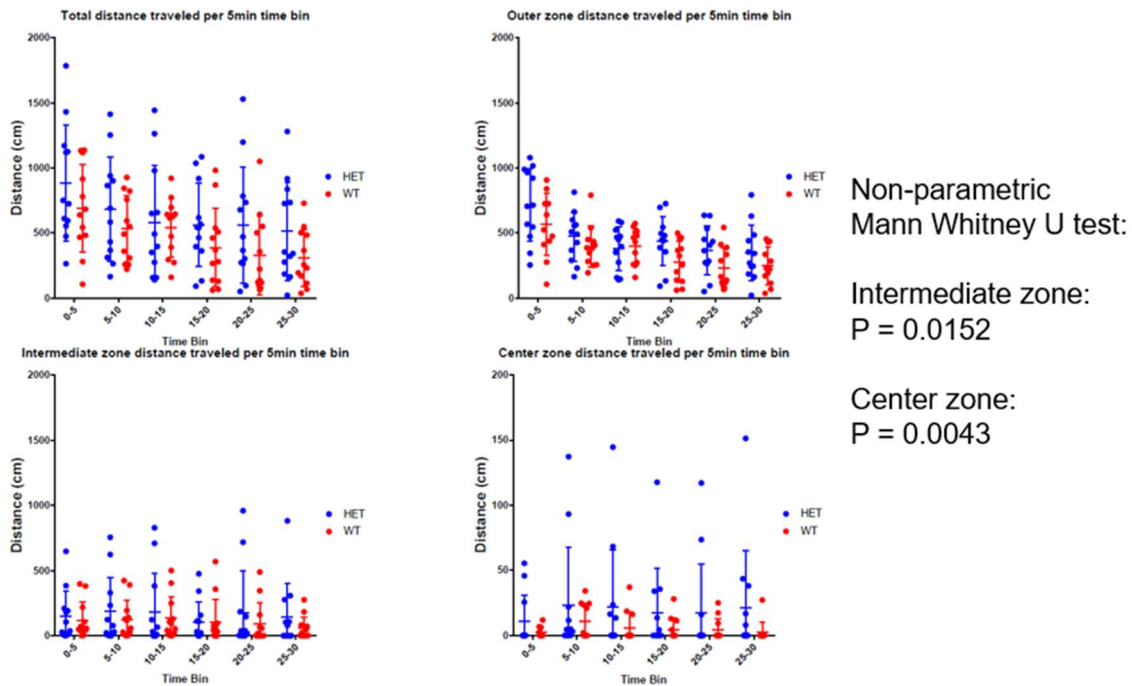
Figure 3-3 Open Field Test.

12 *Phf21a*-heterozygous and WT animals were placed in an Open Field for 30 minutes. Distance traveled in the outer, intermediate, and center zones along with total distance is plotted above in 5 minute increments.
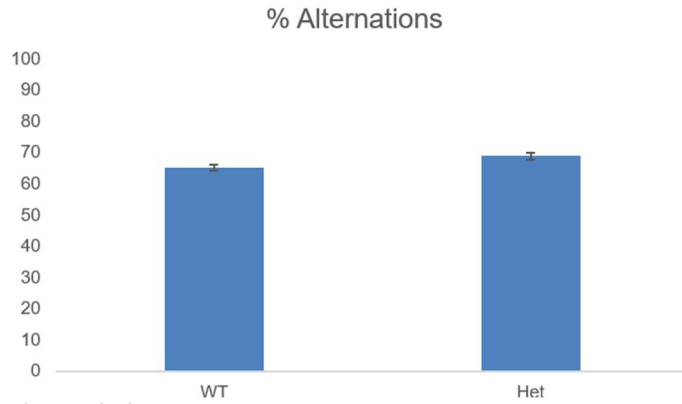
% Alternations

Figure 3-4 Y Maze.

12 *Phf21a*-heterozygous and WT animals were placed in a Y Maze and the percent of alternations was measured.
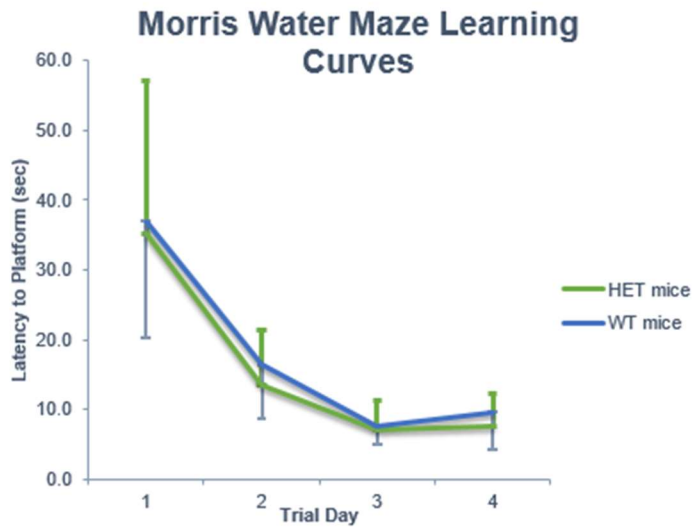
Figure 3-5 Morris Water Maze Learning Curve.

12 *Phf21a*-heterozygous and WT animals were placed in a water maze for 60 seconds at which point they were nudged toward a hidden platform in the Southeast quadrant each day for four days. The time taken for the mouse to find the hidden platform is plotted above across the four trial days.
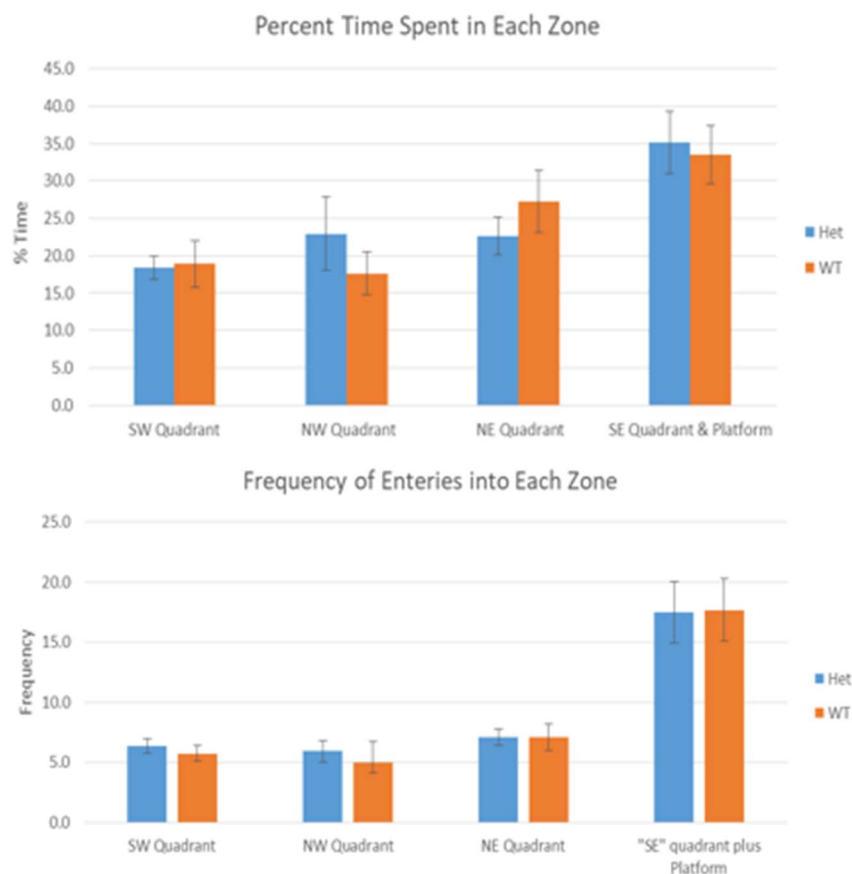
Figure 3-6 Morris Water Maze Test.

12 *Phf21a*-heterozygous and WT animals were placed in the Morris Water Maze on the fifth day after four trial days. The platform in the Southeast quadrant was removed on the fifth day, and the percent of time in each quadrant along with the frequency of entries into each quadrant was measured.
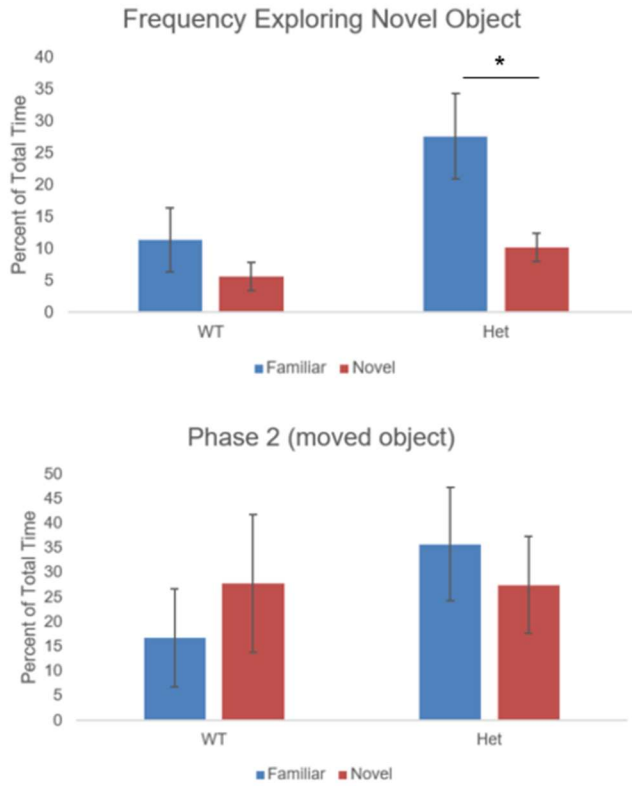
Figure 3-7 Novel Object Recognition Test.

12 *Phf21a*-heterozygous and WT animals were given a trial period to explore two objects. In Phase 1, 24 hours later, they were placed in the same cage with a novel object and the time spent with the novel and familiar object was measured. In Phase 2, they were placed in the same cage, but this time one of the objects was moved to a different corner of the cage and the time spent with the novel and familiar objects was measured.
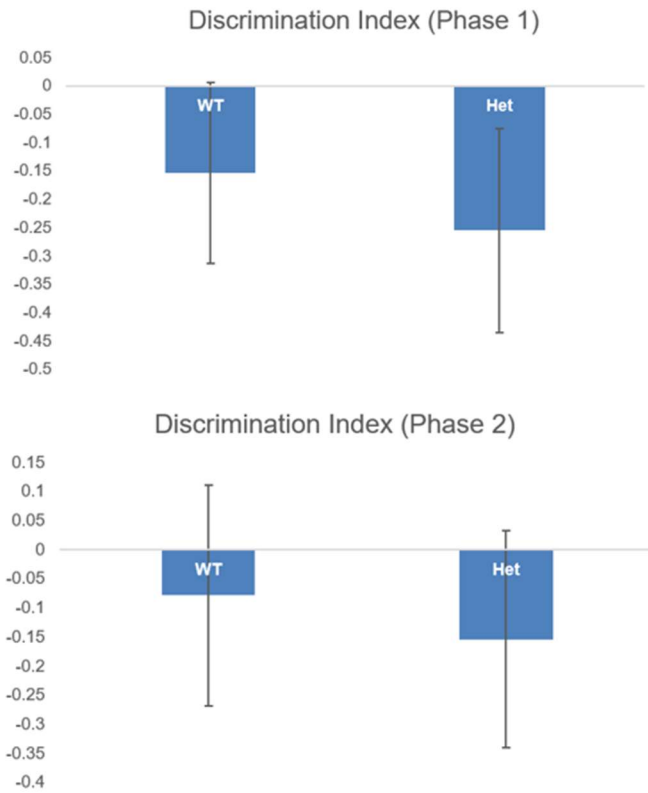
Figure 3-8 Novel Object Recognition Discrimination Index.

Using the same data as Figure 3-7, the discrimination index was calculated. The discrimination index is defined as the frequency of time exploring the novel object divided by the total exploration time and is plotted above.
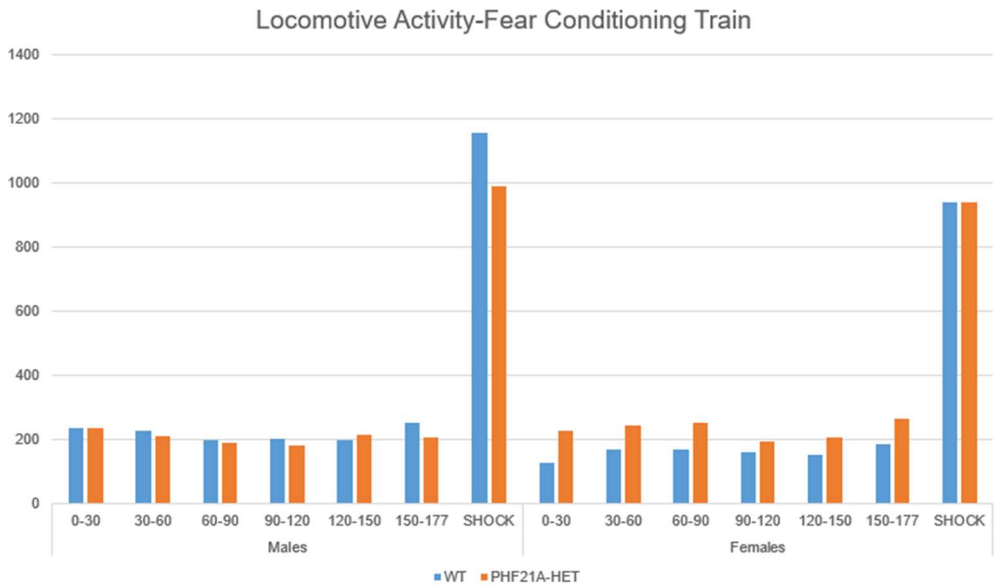
Figure 3-9 Contextual Fear Condition Training Day Locomotion.

5 *Phf21a*-heterozygous and WT Male and Female mice were placed in a novel context for 180 seconds and allowed to explore. In the last three seconds, a 0.8 mA shock was administered. Locomotion was measured.
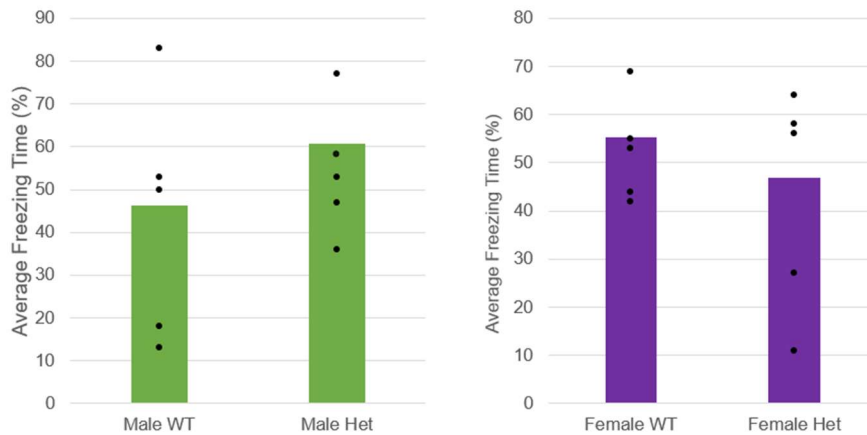
Figure 3-10 Contextual Fear Condition Test Day Freezing Behavior.

24 hours after the training day, the same mice were placed in the same context, and freezing behavior was measured.
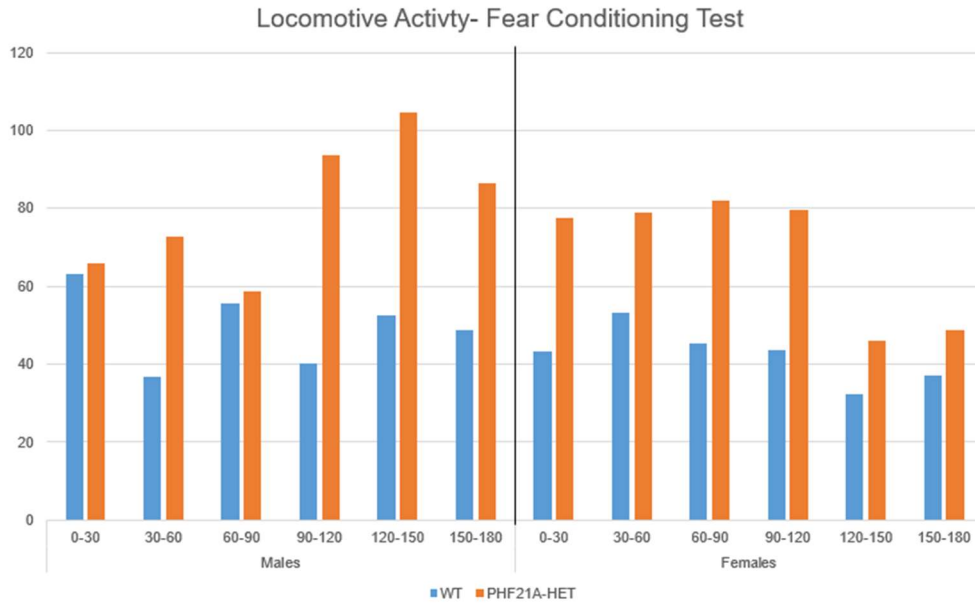
Figure 3-11 Contextual Fear Conditioning Test Day Locomotion.

On the test day, as described in the previous figure, locomotive activity was measured.

Lateral view

Inferior view   Superior view
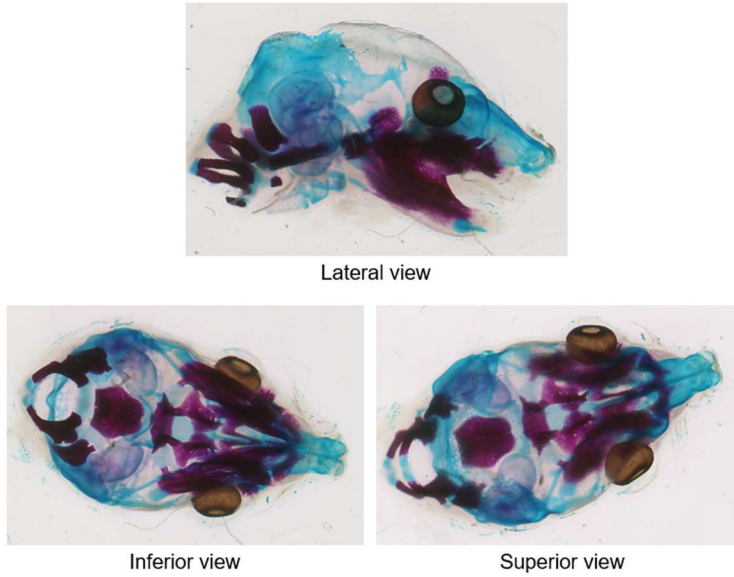
Figure 3-12 Alizarin Red (Bone) and Alcian Blue (Cartilage) Staining of E18.5 Mouse Heads.

Phf21a-null, -heterozygous, and –WT E18.5 mouse heads underwent Alizarin Red and Alcian Blue staining (see Methods). Images were taken from the lateral, inferior, and superior view as demonstrated above.
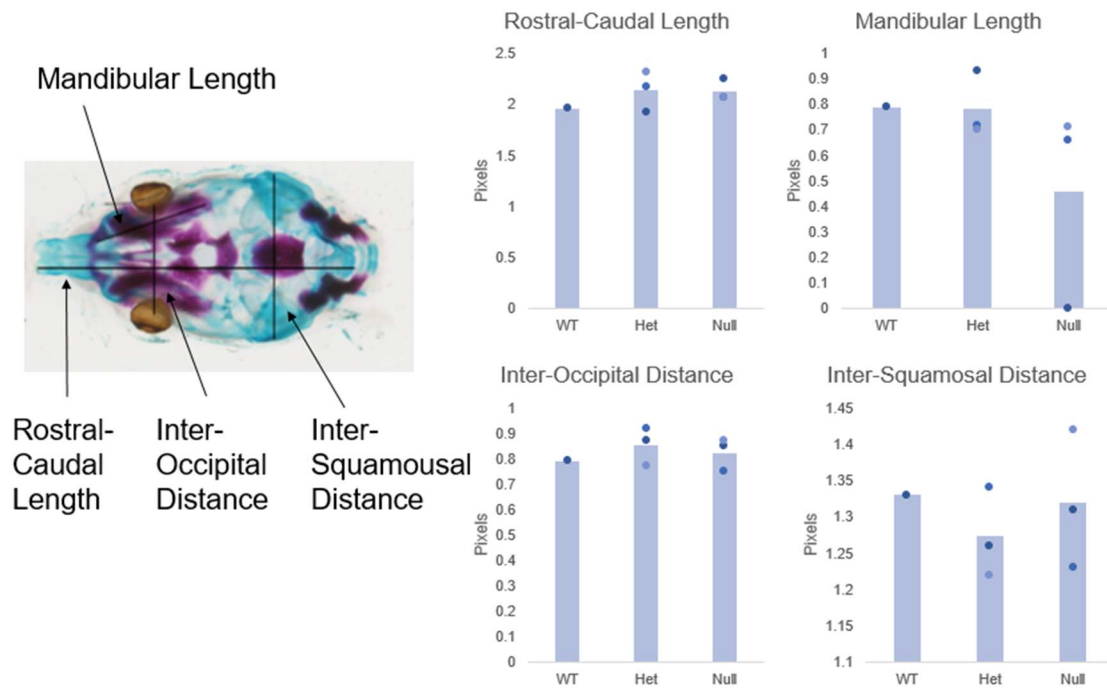
Figure 3-13 Skull Measurements.

Rostral-caudal length, mandibular length, inter-occipital distance, and inter-squamosal distance was measured for one litter of bone and cartilage stained heads. One *Phf21a*-null animal had complete malformation of its jaw as can be seen in the outlier above.

Pterygoid
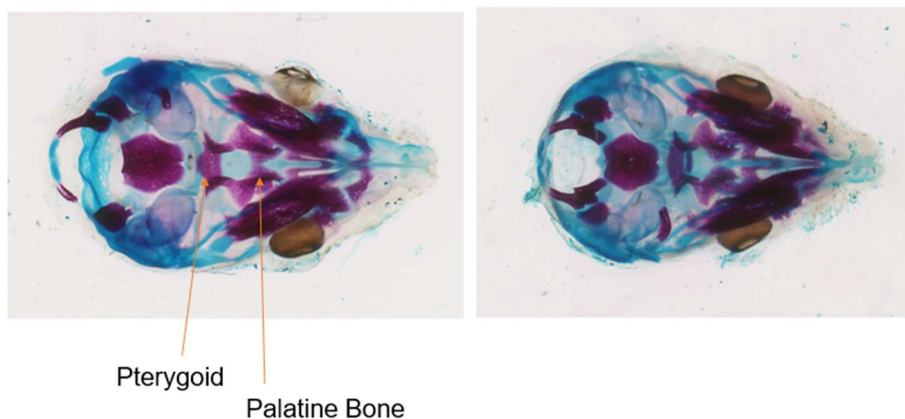
Palatine Bone

	WT	Het	Null
Palate Fused	3	6	2
Palate not Fused	1	1	3
Total (2 litters)	4	7	5

Figure 3-14 Palatal Fusion Phenotype.

Some animals showed failure of fusion of the palatine bone (as can be seen in the image on the right above). The summary of this phenotype over two litters is shown in the table below.

Notes and Acknowledgements

Chapter 4 - Neuron-Specific Microexon Inclusion in a Histone Demethylase Complex Fine Tunes Neurodevelopment

Introduction

Cell-type specific transcriptional programs are thought to be defined by cell-type specific transcription factors (TFs) that bind to specific DNA sequences. In turn, these TFs recruit chromatin regulators that lay down epigenomic marks to reinforce cell-type specific transcriptional identity (Mayran and Drouin, 2018; Zaret, 2018). These chromatin regulators differ from the TFs in that they are ubiquitously expressed and are thought to exert the same intrinsic activity. Thus, the genomic locus-specific action of chromatin regulators is directed by cell-type specific TFs. Challenging this notion is emerging evidence that mutations in chromatin regulators are a major driver for neurodevelopmental disorders (NDDs), such as Intellectual Disability and Autism Spectrum Disorder (De Rubeis et al., 2014; Iossifov et al., 2014; Satterstrom et al., 2020). How could disruption of ubiquitously-expressed chromatin regulators selectively impair cognitive function?

In this study, we hypothesize that chromatin regulators and in particular, histone methylation regulators, exhibit distinct action in neurons relative to other cell types. The disruption of these distinct functions in neurons could be a source of cognitive vulnerability in NDDs. Recent advances in cell-specific transcriptomics revealed a potential mechanism for novel functioning of chromatin regulators in neurons through microexon splicing. Microexons are short exons fewer than 27 nucleotides that are more abundant in neurons compared to other cell types (Irimia et al.,

2014; Li et al., 2015). Neuron-specific splicing factors, such as nSR100, contribute to the enrichment of microexons in neurons. The expression of these splicing factors increases as neurons become terminally differentiated, and the genes undergoing neuron-specific microexon splicing as well as the splicing factors themselves are enriched with autism risk genes. Interestingly, microexon splicing is predicted to alter protein-protein interactions (Buljan et al., 2012; Irimia et al., 2014). Microexon inclusion is conserved across vertebrate evolution.

We identified a pattern of neuron-specific microexon inclusion in chromatin regulators. We predict that these alterations could change the function of chromatin regulators such that they act uniquely in neurons compared to other tissues. In this study, we focus on a particular example whereby multiple members of the CoREST histone demethylase complex undergo neuron-specific microexon inclusion. In this complex, the histone demethylase LSD1 and the associated histone reader PHF21A undergo neuron specific microexon splicing, and we hypothesized that this neuronal complex functions to shape the histone methylation landscape in neurons. Importantly, genetic alterations in both genes are implicated in NDDs. Mutations in LSD1 lead to a Kabuki-like condition that is characterized by global developmental delay, hypotonia, and characteristic facial features (Chong et al., 2016; Pilotto et al., 2016; Tunovic et al., 2014). Microdeletion of PHF21A leads to Potocki Shaffer Syndrome which is characterized by intellectual disability and craniofacial abnormalities (Hamanaka et al., 2018; Kim et al., 2012; Potocki and Shaffer, 1996). We characterize how neuronal splicing of LSD1 and PHF21A changes complex action upon nucleosomes and how the neuronal complex functions to shape histone methylation and the transcriptome distinctly in neurons.

Methods

Constructs and Lentiviral Vectors

Mammalian expression vectors for PHF21A and LSD1 were cloned into first by Gibson Cloning into a pENTR vector. Subsequently, these entry plasmids were recombined into mammalian destination plasmids using LR Recombination into pcDNA3.1 (Addgene #52535) and pLIX-402 (Addgene #55169). For bacterial protein expression, cDNA constructs were cloned into pGEX-4T1 (Addgene #2876) using Gibson Assembly. For insect cell protein expression, cDNA constructs were cloned into a pFastBac Dual Expression Vector (Thermo Fisher) using Ligation Independent Cloning (Doyle, 2005). AT Hook mutations were introduced into the above plasmids using site-directed mutagenesis.

Protein Expression and Purification

PHF21A PHD/AT Hook fragments cloned into pGEX-4T1 constructs were transformed into BL21-DE3 bacteria for bacterial protein expression. 10 ml liquid cultures were grown overnight at 37C in LB media and then transferred into 1 ml LB media the next morning. Growth was monitored until OD600 = 0.6, then 0.25 mM IPTG was added to induce expression and cells were incubated at 15C for 16 hours. Then, cells were harvested and the cell pellets were frozen at -80C. Cells were lysed in Wash Buffer (20 mM HEPES-KOH at pH 7.2, 150 mM KCl, 0.05% NP-40, 10% glycerol) with protease inhibitors (Sigma S8830) and then sonicated for five minutes at 50% amplitude. Then, the soluble fraction was recovered after centrifugation at max speed for 15 minutes. The soluble fraction was bound to a glutathione column (Thermo 16101) for two hours. The column was washed five times by wash buffer and then eluted with wash buffer with 10 mM reduced glutathione at pH 7.5. Eluted proteins were then dialyzed against

fresh wash buffer and concurrently the tag was cleaved by thrombin. To remove free GST, the sample was run through a second glutathione column as well as a Ni-NTA column (Qiagen 30210) to remove thrombin. Purified protein concentration was assessed with a Bradford Assay, and stocks were aliquoted and stored at -80C with 20% glycerol added.

To reconstitute PHF21A/LSD1/CoREST complexes, we used the Bac-to-Bac Baculovirus Expression System (Thermo 10359016) to grow baculoviruses and express protein in Hi5 insect cells. Following harvest of Hi5 cells, cells were lysed by homogenization in High Salt Wash Buffer (30 mM Tris-HCl at pH 7.5, 500 mM NaCl, 0.05% beta-mercaptoethanol) with protease inhibitors added. Lysates were passed twice through a 26G needle, and then the soluble fraction was recovered following centrifugation at max speed for two hours. The soluble fraction was incubated with an amylose column (NEB E8021S) for two hours, washed five times with High Salt Wash Buffer, and then eluted with High Salt Wash Buffer with 10 mM maltose. Elution fractions were dialyzed against High Salt Wash Buffer and were concurrently treated with TEV protease to cleave the tag. Then, purified complexes were subject to FPLC to identify the fraction that exhibited stoichiometric binding of the three complex components.

Recombinant nucleosomes were expressed and purified as described in Lee, et al, 2015 (Lee et al., 2015). We also used recombinant nucleosomes with a synthetically added di-methylation at Lysine 4 (H3K4me2) as well as other histone marks (H3K9me2, H3K27me2, H3K36me2, and H4K20me2) from EpiCypher. We also purified nucleosomes from HEK293T cells with a stably-expressing Strep-H2A transgene. Cells were incubated in hypotonic lysis buffer (20 mM HEPES-KOH at pH 7.2, 20 mM KCl, 10% glycerol, and 0.05% NP-40) with protease inhibitors

for 10 minutes at 4C. Then, chromatin was digested with Micrococcal Nuclease (NEB) in 20 mM HEPES-KOH at pH 7.2, 340 mM KCl, 10% glycerol, 0.05% NP-40, and 1mM CaCl2 for 2 hours. The reaction was quenched with 4 mM EGTA and centrifuged at max speed for 10 minutes to recover the soluble chromatin fraction. The soluble fraction was incubated with a Strep-tactin column (Qiagen 1057979) for 3 hours, washed five times with wash buffer (20 mM HEPES-KOH at pH 7.2, 150 mM KCl, 10% glycerol, 0.05% NP-40), and eluted with wash buffer with 50 mM desthiobiotin (IBA 2-1000-002) at pH 8.0.

Electrophoretic Mobility Shift Assays

100 nM DNA or 1 ug of purified nucleosomes were incubated with an increasing concentration of PHF21A-fragment proteins in EMSA Buffer (10 mM Tris-HCl at pH 7.5, 50 mM KCl, 0.5 mM MgCl2, 0.1 mM EDTA, 5% glycerol, and 0.1 mg/ml BSA) for 30 minutes at 25C. Samples were loaded into a 6% Acrylamide, 0.5X TBE PAGE gel, electrophoresed, stained by Ethidium Bromide, and visualized.

Binding Assays

For binding assays with histone peptides, 1 ug of histone peptides (Active Motif) were incubated with 1 ug of purified protein in binding buffer (50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 0.05% NP-40) and rotated overnight. The next day, Strep-avidin agarose beads (Thermo 20349) were incubated with the binding reaction for 1 hour at 4C. Beads were washed five times with binding buffer and then resuspended for SDS-PAGE and Western Blot analysis using GST antibody (Millipore 05-782) at 1:1000 concentration to detect GST-PHF21A binding.

Histone Demethylation Assays

1 ug of recombinant (Epicypher) or cellular nucleosomes (see protein purification above) were incubated with 1, 3, or 9 ug of LSD1-CoREST and/or PHF21A in histone demethylation buffer (50 mM Tris-HCl at pH 8.0, 50 mM KCl, 0.5% BSA, 5% glycerol) and incubated for 4 hours at 37C. Then, samples went through SDS-PAGE and quantitative fluorescent Western Blot analysis. The following antibodies were used: H4 (Abcam ab31830), H3 (Abcam ab24834), H3K4me1 (Abcam ab8895), H3K4me2 (Abcam ab7766), H3K4me3 (Abcam ab8580), H3K9me1 (Epicypher 13-0029), H3K9me2 (Abcam ab194680), H3K9me3 (Abcam ab8898), H3K27me1 (Abcam ab194688), H3K27me2 (Active Motif 39245), H3K27me3 (Active Motif 39155), H3K36me1 (Abcam ab9048), H3K36me2 (Abcam ab9049), H3K36me3 (Active Motif 61101), H4K20me1 (Abcam ab9051), H4K20me2 (Abcam ab9052), H4K20me3 (Abcam 9053). LiCOR Image Studio software was used to quantify histone modification band level relative to a histone H3 or H4 loading control. Each Western Blot was reproduced with at least two independent experiments, and the average quantification was reported.

Mouse models

Two different mouse models were used in this study. The first is a constitutive germline Phf21a knock out mouse model that is described in Iwase, et al, 2006 (Iwase et al., 2006). In this mouse model, all isoform of Phf21a are lost, and Phf21a-null animals die neonatally. The second mouse model is a germline Phf21a-N knockout mouse model. This model was generated using the University of Michigan Transgenic Animal Core through CRISPR genome editing targeting the Phf21a-N microexon. CRISPR-Cas9 constructs were microinjected into mouse eggs, and founder

animals were screened for exon deletion using primers: Forward 5'-

ACAGACGCCCAGCACCTTTAG-3'; Reverse: 5'-GTAAGGGCTCCAAACCCCAG-3'.

Cortical Neuron Culture

Timed pregnant female mice were sacrificed at day E16.5 for dissection and culturing of cortical

and hippocampal neurons. Briefly, cortices and hippocampi were microdissected, treated with

2.5% Trypsin (Invitrogen 15090), quenched by FBS, and then treated with 1% DNase (Sigma

DN-25) to dissociate brain tissue into neurons. Cells were counted and then plated in Neurobasal

Media (Gibco 21103049) with 1x B27 (Gibco 17504-044), 0.5 mM GlutaMax (Gibco

35050061), 25 uM beta-mercaptoethanol (Gibco 21985), and 1% Penicillin/Streptomycin (Gibco

15140122). Before plating, plates were coated with PDL (reagent) overnight at 37C and then

washed with ddH2O three times. Neuron culture cells were fed every 3-4 days by replacing half

of the above media with fresh media.

Synapse Assays

Neuron cultures were grown on PDL-coated coverslips until day in vitro 14. Then, cells were

washed in artificial CSF (145 mM NaCl, 3 mM KCl, 1 mM CaCl2, 1 mM MgCl2-6H2O, 8 mM

dextrose, 10 mM HEPES), fixed in paraformaldehyde (Thermo 28908), quenched by 0.01 M

glycine, washed in PBS, blocked in 10% BSA in PBS, and then incubated with primary

antibodies overnight at 4C. For excitatory staining, we used PSD95 (NeuroMab K28/43, 1:500)

and vGlut (Synaptic Systems 135303, 1:1000). For inhibitory staining, we used gephyrin

(Synaptic Systems 147111, 1:500) and vGat (Synaptic Systems 131004, 1:1000). Slides were

washed three times in 3% BSA in PBS, incubated with 1:1000 secondary antibodies along with

1:1000 DAPI, washed three more times, and then mounted on slides using ProLong Gold

Antifade Mountant (Invitrogen P36930). Slides were blinded to genotype and imaged using a

Nikon A1 inverted confocal microscope. Images were taken in parts of the slide where the

number of neuron cell bodies was equal. Laser levels, microscope settings, and subsequent

thresholding were kept equal across all samples. ImageJ was used to separate channels and

threshold the images (Schneider et al., 2012). Then NeuronJ plugin was next used to trace

dendrite segments (Meijering et al., 2004). Finally, the SynapCountJ plugin was used to assess

colocalization of pre- and post-synaptic puncta along the defined dendrites, and average synapse

density per 100 pixels was reported (Mata et al., 2017).

Mouse Embryonic Fibroblast Cultures

E16.5 mouse carcasses were eviscerated, minced, and incubated with 0.25% trypsin (Gibco

25200072) for 30 minutes. Cells were quenched and then plated in DMEM (Gibco 11995065)

with 10% FBS, 1% Pencillin/Streptomycin (Gibco 15140122), and 1x GlutaMax (Gibco

35050061). Cells were allowed to grow out for 3-5 days, were passaged once, and then were

harvested for all experiments described below. For the RNA-Seq experiment with Doxycycline-

inducible rescue expression constructs, we established immortalized stable cell lines. We found

that primary MEFs were not able to incorporate a lentiviral-transduced transgene and survive

puromycin selection. Thus, we immortalized MEFs by allowing cells to grow for three months in

culture, passaging as necessary(Todaro and Green, 1963). Once immortalized MEFs were

growing rapidly, lentiviral constructs of pLIX-PHF21A-n and –c were transduced. Two days

later, 2ug/ml puromycin was applied, and cells were allowed to grow for two more passages

(about 8 days). Then, 2.5 ug/ml Doxycycline was applied for two days, and cells were harvested for RNA-Seq and protein analysis.

Sub-cellular Fractionation

1.5 million E16.5 neurons and first passage MEFs (see above) were fractionated using the Subcellular Protein Fractionation Kit for Cultured Cells (Thermo 78840). Following recovery of each fraction, samples were subject to SDS-PAGE and Western Blot analysis using antibodies against LSD1 (Abcam ab17721, 1:1000) or PHF21A (homemade antibody, 1:1000).

RNA-Seq Library Preparation

RNA was assessed for quality using the TapeStation (Agilent, Santa Clara, CA). Samples were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (catalog number E7760L) Ribo depletion Module NEBNext rRNA Human/Mouse/Rat (catalog number E6310X) and NEBNext Multiplex Oligos for Illumina Unique dual (catalog number E6440L) (NEB, Ipswitch, MA). Where 375ng of total RNA was ribosomal depleted using rRNA Depletion module. The rRNA-depleted RNA is then fragmented 10 minutes determined by RIN (RNA Integrity Number) of input RNA as per protocol, and copied into first strand cDNA using reverse transcriptase and dUTP mix. Samples undergo end repair and dA-Tailing step followed by ligation of NEBNext adapters. The products are purified and enriched by PCR to create the final cDNA library. Final libraries were checked for quality and quantity by TapeStation (Agilent) and qPCR using Kapa's library quantification kit for Illumina Sequencing platforms (catalog # KK4835) (Kapa Biosystems, Wilmington MA). The samples were pooled and

sequenced on the Illumina NovaSeq S4 Paired-end 150bp, according to manufacturer's recommended protocols.

RNA-Seq Analysis

Read quality was determined using FastQC (v0.11.8) (available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were mapped to the mm10 genome (Gencode) using STAR (v2.6.1) (Dobin et al., 2013). A counts file was generated by FeatureCounts (SubRead v1.5.0) (Liao et al., 2014). BAM files were converted to bigwigs using deepTools2 (v.3.2.0) (Ramirez et al., 2016). The computational workflow was managed by Snakemake (v.5.4.0) (Koster and Rahmann, 2012), and the summary statistics were visualized by multiQC (Ewels et al., 2016). DESeq2 (v.1.24.0) was used for differential gene expression analysis (Love et al., 2014). A q-value of <0.1 was used to identify differentially expressed genes. Gene ontology analysis was performed using LR-Path using the RNA-Enrich function (Sartor et al., 2009).

Results

Neuron-Specific Microexon Splicing of Chromatin Regulators

Recent studies of the genetic basis of neurodevelopmental disorders, such as autism spectrum disorder and intellectual disability, have uncovered that in addition to brain-specific genes (such as genes important for synapse development and function as well as transcription factors important for different brain cell type fates), a number of chromatin regulators are associated

with NDDs (De Rubeis et al., 2014; Iossifov et al., 2014; Satterstrom et al., 2020). We asked why the brain is so sensitive to the disruption of ubiquitously expressed chromatin regulators.

Given the association of ubiquitously-expressed chromatin regulatory genes in the genetics of NDDs, we hypothesized that chromatin regulators could have unique functions in the brain compared to other tissues. A number of recent studies have identified a unique program of microexon splicing that is enriched in the brain (Irimia et al., 2014; Li et al., 2015). We mined publically available data to see whether chromatin regulators undergo neuron-specific microexon inclusion. We identified 76 chromatin factors that undergo neuron-specific microexon splicing which suggests a potential source of neuron-specific variants of these regulators (Porter et al., 2017a). This led us to hypothesize that distinct forms of these chromatin regulators could play a role in shaping the neuronal epigenome distinctly from other cells. Interestingly, we found multiple members of the same chromatin-regulating complex on the list. Both LSD1 and PHF21A undergo neuron-specific microexon splicing as members of the CoREST complex, and we hypothesized that this neuron-specific demethylase complex regulated histone methylation distinctly in neurons.

Identification of a Neuronal Histone Demethylase Complex
Several recent studies have described the neuronal isoform of LSD1 (LSD1-n). The LSD1 neuronal microexon is a twelve nucleotide insertion within the catalytic amine oxidase domain of LSD1 (Figure 4-1A. Whereas canonical LSD1 (LSD1-c) is a known Histone H3 Lysine 4 mono- and di-methyl (H3K4me1/2) histone demethylase, initial reports disputed the function of LSD1-n. While the first report claimed that LSD1-n reduced its enzymatic activity on H3K4me (Zibetti

89

et al., 2010), subsequent reports claimed that LSD1-n instead changed its substrate specificity to either H3K9me through cooperation with SVIL (Laurent et al., 2015) or H4K20me (Wang et al., 2015). LSD1-c is broadly expressed through the body, whereas neurons express both LSD1-c and LSD1-n (Zibetti et al., 2010). LSD1-n has been shown to be physiologically important for neurogenesis and proper neuronal response to excitation through recruitment of transcription factors that induce expression of immediate early genes (Rusconi et al., 2015; Zibetti et al., 2010).

With a PHF21A antibody we generated, we found that whereas most tissues express a ~94 kDa protein, we saw the appearance of a second lower ~83 kDa band in brain tissue (Figure 4-1B). By separating neurons and astrocytes in culture, we were able to distinguish the two bands into the larger canonical PHF21A (PHF21A-c) which is broadly expressed, including in non-neuronal brain cells, and a smaller neuronal PHF21A (PHF21A-n) (Figure 4-1C). In contrast to LSD1-n, PHF21A-n is exclusively expressed in neurons. The microexon inclusion in PHF21A-n is a 23 nucleotide exon just upstream of the histone-binding PHD finger domain (Figure 4-1A). Interestingly, this splicing event disrupts a predicted AT hook domain that exists in PHF21A-c. Expression of the neuronal PHF21A microexon is specific to neurons as seen in publically-available brain cell-type specific mRNA-Seq (Zhang et al., 2014) (Figure 4-1D). We harvested mouse embryonic brain RNA at different stages of development to investigate the expression dynamics by RT-PCR. Expression of both LSD1-n and PHF21a-n appears to begin around E13.5 which coincides with the beginning of neuronal differentiation (Figure 4-1E).

PHF21A-n splicing disrupts a functional AT Hook

To investigate the biochemical impact on the neuronal microexon inclusion in PHF21A-n, we

expressed and purified PHF21A fragments that include the PHD domain and the upstream

alternatively spliced region (Figure 4-2A). We confirmed the known activity of PHF21A as a

H3K4me0 reader protein. Both PHF21A-c and PHF21A-n bind to unmodified histone H3

peptides, and this binding is ablated by the addition of H3K4 methylation (Figure 4-2B). Given

that the neuronal microexon splicing event disrupts a predicted AT Hook domain, we next tested

DNA binding through an Electrophoretic Mobility Shift Assay (EMSA). We incubated Widom

601 DNA with the PHF21A fragments and found that only PHF21A-c was able to bind to DNA

(Figure 4-2C). We next incubated the PHF21A fragments with recombinant nucleosomes and

found that while the PHF21A-PHD only and PHF21A-n fragments weakly bound to

nucleosomes by EMSA, PHF21A-c strongly bound nucleosomes (Figure 4-2D). To evaluate the

necessity of the AT hook for nucleosome binding, we mutated three core residues of the AT

Hook (GRP) to the residues found in PHF21A-n (ANE) and found ablation of binding.

Additionally, we expressed the AT hook and PHD fragments of PHF21A-c separately and found

that the AT Hook alone had much stronger affinity than the PHD finger (Figure 4-2E. We

confirmed these findings in nucleosomes purified from 293T cells which represent a population

of nucleosomes with heterogeneous DNA sequences and histone modifications (Figure 4-3A, B).

Since the GST tag is known to oligomerize (Figure 4-3C), we cleaved the tag and repeated all

EMSAs to confirm results with cleaved PHF21A fragments (Figure 4-4A, B). Cleaved PHF21A

fragments were less stable but recapitulated binding results (Figure 4-4C-F).

In total, these data show that neuronal microexon inclusion in PHF21A-n changes its affinity to nucleosomes and that the AT Hook domain is the dominant nucleosome-binding domain of PHF21A-c.

Differential chromatin functions of the canonical and neuronal LSD1/PHF21A/CoREST complexes

To investigate the biochemical actions of the canonical and neuronal complexes, we used a baculovirus/insect cell protein expression system. We expressed canonical and neuronal MBP-LSD1/ CoREST and also MBP-PHF21A/LSD1/CoREST. Using size exclusion chromatography, we isolated the stoichiometric complex (Figure 4-5). We next incubated the complexes with recombinant nucleosomes with 177 bp DNA and performed a MBP pulldown (Figure 4-5). We found that while canonical or neuronal LSD1/CoREST only weakly bound nucleosomes, the three part complex of PHF21A/LSD1/CoREST much more strongly bound nucleosomes. Additionally, the canonical PHF21A/LSD1/CoREST complex bound nucleosomes about twice as strongly as the neuronal complex (Figure 4-5). These data suggest that PHF21A, especially the canonical form that has an intact AT Hook is a more important component of how the CoREST complex contacts nucleosomes than previously appreciated.

We continued by testing the ability of the canonical and neuronal complexes to demethylate various histone marks. We incubated the complexes with recombinant nucleosomes with synthetically added di-methyl marks and chose to survey the canonical target of LSD1-c action, H3K4me2, along with four other major di-methyl histone marks: H3K9me2, H3K27me2, H3K36me2, and H4K20me2. We quantified histone marks through quantitative fluorescent

Western Blotting using histone H3 or H4 as a control. We found that while LSD1-c/CoREST only modestly removed H3K4me2, the three part PHF21A-c/LSD1-c/CoREST complex strongly removed H3K4me2 (Figure 4-6A). The LSD1-n/CoREST and PHF21A-n/LSD1-n/CoREST complex did not have appreciable demethylase activity on H3K4me2. Since this recombinant population of nucleosomes uniformly contains H3K4me2, we could also measure the activity of the complex by the appearance of H3K4me1 (Figure 4-6A). In this same way, we observed a weak activity of LSD1-c/CoREST and a very strong activity of PHF21A-c/LSD1-c/CoREST so much so that even H3K4me1 began to decrease at the highest concentration of complex added. Neither the canonical nor neuronal complexes had any appreciable activity on H3K9me2, H3K27me2, H3K36me2, or H4K20me2 (Figure 4-6B). We additionally repeated the demethylase assay on nucleosomes purified from HEK293T cells which represent a population of nucleosomes with heterogeneous DNA sequences and histone modifications. This experiment yielded the same results with stronger H3K4me2 demethylase action of PHF21A-c/LSD1-c/CoREST compared to LSD1-c/CoREST alone (Figure 4-7). In this experiment, we also found that PHF21A-c/LSD1-c/CoREST has strong demethylase activity on H3K4me1 which likely reflects the relatively high amount of H3K4me1 of the genome that can be demethylated compared to the small proportion of H3K4me2 regions throughout the genome. We furthermore did not find activity of canonical or neuronal complexes on any other histone mark (Figure 4-7).

PHF21A and LSD1 chromatin dynamics in vivo

Given the crucial role of PHF21A in the CoREST complex in nucleosome binding and facilitating LSD1-mediated H3K4me demethylation and the current lack of knowledge of the molecular function of PHF21A, we turned to two different Phf21a mouse models. The first

model is a constitutive knock out (KO) model where both isoforms are lost in all tissues (Iwase et al., 2006). In this model, Phf21a-null mice are unable to suckle milk and die neonatally, but we are able to recover all genotypes during embryonic development and are able to grow neurons in culture to maturity in vitro. To investigate the contribution specifically of PHF21A-n, we created a second mouse model where CRISPR-Cas9 was used to specifically remove the PHF21A neuronal exon (N-KO). Phf21a-n-null mice live through adulthood. While non-neuronal cells, like fibroblasts, exhibit a wild-type level of PHF21A-c, we interestingly found that neurons aberrantly express Phf21a-c in the absence of the neuronal exon (Figure 4-8A-C). Through immunofluorescent staining of neurons, we confirmed that N-KO neurons are expressing Phf21a confirming that neurons, and not other non-neuronal brain cells, are expressing Phf21a-c (Figure 4-8D). Thus, we used N-KO neurons as a system to test PHF21A isoform compatibility whereby neurons express Phf21a-c.

We harvested E16.5 cortical and hippocampal neurons along with mouse embryonic fibroblasts (MEFs) to analyze the activity of PHF21A. We began by fractionating the subcellular compartments of neurons and MEFs to assess binding of PHF21A and LSD1 to the nucleoplasm and chromatin compartments. In comparing wild type neurons and fibroblasts, we found that 80% of Phf21a-n resides in the nucleoplasm while only 10% of Phf21a-n is bound to the soluble and insoluble chromatin (Figure 4-9A). In wild type MEFs, however, only 66% of Phf21a-c resides in the nucleoplasm, whereas 33% of Phf21a-c is bound to the soluble and insoluble chromatin, in line with the increased nucleosome binding capacity of PHF21A-c compared to PHF21A-n (Figure 4-9A). When we probed for LSD1 localization we found that its chromatin binding correlated to PHF21A binding (Figure 4-9B) whereby LSD1 was more localized to

chromatin in Phf21a-c-expressing fibroblasts compared to Phf21a-n-expressing neurons. It's important to note that non-neuronal cells, like fibroblasts only express LSD1-c, but neurons express both LSD1-c and LSD1-n. Of note, in these Western Blots, given the LSD1 antibody used, we are unable to distinguish between LSD1-c and LSD1-n due to the small difference in size. Interestingly, the N-KO neurons (which aberrantly express Phf21a-c) showed an increased chromatin localization of Phf21a-c, supporting the idea that canonical and neuronal PHF21A isoforms have distinct biochemical activity and are not simply directed by neuronal transcription factors. Accordingly, N-KO neurons also exhibit greater LSD1 chromatin localization.

We next assessed H3K4 methylation in these cells. Quantitative western blotting of whole cell lysates showed only subtle differences between global H3K4 methylation in these cells (Figure 4-9C), including that N-KO neurons, which aberrantly express Phf21a-c, showed a decrease in H3K4me2 and an increase in H3K4me1. We also found that KO MEFs also exhibit a decrease in H3K4me2 and an increase in H3K4me1 which may be due to compensation by other demethylases.

The role of Phf21a-n/c in establishing the neuronal transcriptome
We continued on to investigate the impact on PHF21A isoforms in the establishment of the neuronal transcriptome through RNA-Seq. To understand the roles of PHF21A in neurons and a non-neuronal cell type, MEFs, we harvested RNA for RNA-Seq. In this experiment, we harvested E16.5 Phf21a-WT and KO cortices and grew them for 7 days in vitro. We also harvested MEFs and immortalized them by two months of passaging in culture. Upon different gene expression (DEG) analysis, we found around 100 dysregulated genes in both Phf21a-KO

95

neurons and MEFs with an equal number of genes up- and down-regulated (Figure 4-10A). However, the Phf21a-KO DEGs were different in neurons than in MEFs (Figure 4-10B) suggesting that either PHF21A-n and PHF21A-c regulate an independent set of loci or that the cellular context influences PHF21A function.

We next analyzed gene-expression at REST target genes, as defined by published REST ChIP-Seq data in ES cells (Seki et al., 2014). We categorized the direction of dysregulation of REST target genes in both KO MEFs and KO neurons and found a significantly increased number of REST target genes that were increased in Phf21a-KO MEFs and decreased in Phf21a-KO neurons, compared to the other directions of dysregulation (Figure 4-10C). In analyzing the DEGs in Phf21a-KO MEFs and neurons, we found that ontologies relevant to neurodevelopment were down-regulated in Phf21a-KO neurons and up-regulated in Phf21a-KO MEFs (Figure 4-10D). For example, down-regulated pathways in Phf21a-KO neurons included "active ion transmembrane transporter activity" and "regulation of neurotransmitter levels." "Axon guidance" and "synapse organization" were two up-regulated pathways in Phf21a-KO MEFs. These data collectively support the known role of PHF21A in the CoREST complex in MEFs whereby this complex is responsible for the down-regulation of neuronal genes in non-neuronal cells. However, these data also demonstrate that in neurons, PHF21A is responsible for the expression of neuronal genes.

We also attempted rescue of PHF21A isoform expression in Phf21a-null MEFs and neurons. We used a lentiviral construct encoding a doxycycline-inducible PHF21A expression system. We were unable to use puromycin selection in primary MEFs; however, after immortalizing MEFs,

we were able to establish stable lentiviral transduced, puromycin-resistant cell lines. We used 2.5 ug/ml doxycycline for two days to induce expression of PHF21A and then harvested RNA for RNA-Seq (Figure 4-11A). This same strategy was unsuccessful in Phf21a-null neuron cultures. In MEF rescue, we found that gene expression clustered around Phf21a-null biological replicate (i.e. animal of origin) (Figure 4-11B). This confounding factor was removed through incorporation of a covariate (Figure 4-11C); however, the rescue groups did not cluster together. A small number of genes was significantly changed upon re-expression of PHF21A-n or PHF21A-c. Most of these genes were down-regulated upon re-expression of PHF21A, in line with the function of PHF21A as a transcriptional repressor (Figure 4-11D). However, the majority of these genes had a very small base mean expression and the list of DEGs in this comparison did not overlap at all with the DEGs identified in the Phf21a-WT vs Phf21a-KO comparison (Figure 4-11D). The small magnitude of gene expression changes identified in these PHF21A rescue cells indicate that our method of rescue may have not been complete.

Nonetheless, we analyzed the change in gene expression of the Phf21a-KO MEF DEGs in both the PHF21A-n and PHF21A-c rescue cell lines relative to WT MEFs. In this analysis, we find that most Phf21a-KO MEF DEGs are unchanged in the rescue lines (e.g. near the line with slope = 1). By comparing both rescue and null data to WT, we can identify whether there is a trend toward rescue gene expression (a complete rescue would result in a slope of 0 of these genes). We find the slope of both isoform rescues is significantly less than 1 (by a 99% confidence interval), but not significantly different than each other (Figure 4-11E). Further, we defined individually rescued genes as those where the log2-fold change of Phf21a-KO vs Phf21a-WT compared to Phf21a-rescue vs Phf21a-WT is greater than 2. We find that PHF21A-n rescue

resulted in only a single Phf21a-KO MEF DEG being rescued, whereas PHF21A-c rescue was able to rescue 11 Phf21a-KO MEF DEGs (Figure 4-11E). Although PHF21A rescue expression was incomplete in this experiment, it does appear that there is a trend for PHF21A-c being more able to rescue proper gene expression than PHF21A-n.

In total, this transcriptome analysis comparing WT and *Phf21a*-KO neurons and MEFs showed that Phf21a loss in neurons surprisingly leads to a down-regulation of genes important for neurodevelopment in contrast to the known role of PHF21A-c. These results suggest that PHF21A-n functions distinctly in neurons in a manner that leads to the expression of neuronal genes important for neuron development. In analyzing the DEG and gene ontology analyses, a number of neuronal genes are down-regulated that stand out as potential mediators of neurological phenotypes such as *Grin1*, *Grin2b*, *Grm4*, *Gabra5*, *Pclo*, *Sncb*, *Syn1*, *Vamp2*, *Shank3, Snap91, Prkacb,* and *Pacsin1*. These genes encode proteins important for both glutamate and GABA receptor subunits, vesicle regulation, and cAMP signaling. Interestingly, other genes relating to calcium regulation in neurons are reliably up-regulated including voltage-gated calcium channels such as *Cacna1g*, and synaptotagmins such as *Syt2*.

Phf21a-n KO (N-KO) neurons show no changes in gene expression
The new N-KO mouse model allowed us a method to analyze how neurons change in expression when expressing the 'wrong' isoform (neurons expressing Phf21a-c). We hypothesized that differentially expressed genes in *Phf21a*-KO neurons may show opposite regulation in the N-KO neurons. In this experiment, we harvested neurons from E16.5 cortices from WT, KO, and N-KO mice, extracted RNA, and proceeded with RNA-Seq library preparation and sequencing.

Importantly, we used distinct WT littermate controls relative to both KO and N-KO animals since the two mouse colonies were generated independently. In comparing the KO E16.5 neurons, we found a similar pattern of transcription dysregulation as the DIV7 neurons (Figure 4-12A) with many common DEGs being identified. Additionally, LR Path analysis of the E16.5 KO RNA-Seq data showed similar pathways dysregulated as the DIV 7 data (Figure 4-12B). Namely, many pathways important for neurodevelopment and synapse formation including "synaptic signaling" and "monovalent inorganic cation transporter membrane transporter activity" were downregulated. However, upon differential gene expression analysis between WT and N-KO E16.5 neurons (neurons expressing Phf21a-c, Figure 4-12C) we found no difference in gene expression between the N-KO and WT animals (Figure 4-12D). These data indicate that E16.5 neurons that express Phf21a-n or Phf21a-c have no transcriptional differences. Since results of both E16.5 and DIV7 *Phf21a*-KO neurons revealed downregulation of genes and ontologies relevant to synaptogenesis, it's possible that Phf21a plays an important role in regulating these genes throughout synaptogenesis, which largely occurs after birth. Therefore, transcriptomic differences between neurons expressing Phf21a-n and Phf21a-c may not become apparent until later points in development or upon stimulation.

PHF21A-n fine tunes synapse number

Haploinsufficiency of PHF21A is associated with neurodevelopmental disorders, and it has emerged as an autism risk gene (Satterstrom et al., 2020); however, there has not yet been research uncovering the molecular mechanism of how PHF21A could lead to these phenotypes. Furthermore, since PHF21A-n expression is restricted to neurons, we hypothesized that this specific microexon inclusion is important for normal neurodevelopment. Autism studies have

pointed to the role of synapses including the balance between excitatory and inhibitory synapses in the pathology of autism (De Rubeis et al., 2014; Ebert and Greenberg, 2013; Nelson and Valakh, 2015; Rubenstein and Merzenich, 2003).

Since Phf21a-null mice die neonatally and much of synaptogenesis occurs in postnatal development, we pursued a neuron culture method to investigate how mature synapses would differ across genotypes. We dissected and cultured cortical and hippocampal neurons and allowed them to mature for 14 days in vitro, at which point synapses are mature in culture (Basarsky et al., 1994). We inferred the number of synapses by quantifying colocalization of excitatory and inhibitory pre- and post-synaptic markers through confocal microscopy (Glynn and McAllister, 2006). In Phf21a-KO animals, we observed a defect in excitatory synapse number, but no change in inhibitory synapses (Figure 4-13). Interestingly, in Phf21a-n-KO animals (i.e. neurons that express Phf21a-c), we observed a significant increase in excitatory synapse number without a change in inhibitory synapse number (Figure 4-13).

These results have led us to postulate that given that relative nucleosome-binding strength of PHF21A-c vs PHF21A-n that PHF21A-n serves as a mechanism to dampen down the action of the CoREST complex in neurons (Figure 4-14). This change in biochemical activity allows for proper gene expression of neuronal genes (as evidenced by RNA-Seq) permitting the proper balance of synapse number. Given the synapse assay was performed at DIV 14 leads us to further hypothesize that the functional difference between Phf21a-n and Phf21a-c is crucial during the post-natal period of synaptogenesis. RNA-Seq experiments of mature N-KO neurons will allow us to test this hypothesis.

Discussion

In this study, we identified a pattern of neuron-specific microexon inclusion in chromatin regulators challenging the notion that ubiquitously expressed chromatin regulators function in the same way across cell types. We explored how neuronal splicing of multiple members of the CoREST complex, the histone demethylase, LSD1, and the histone reader, PHF21A, lead to a distinct biochemical function of the complex that permits neuronal gene expression (Figure 4-14). The four amino acid inclusion in the amine oxidase domain in LSD1-n disrupts demethylation capability of H3K4me1/2. Microexon inclusion in PHF21A-n disrupts a functional AT Hook domain. The PHF21A-c AT Hook domain leads to greatly increased nucleosome affinity and this accordingly directs stronger demethylation activity of the LSD1/PHF21A/CoREST complex. The weakened neuronal LSD1/PHF21A/CoREST complex permits the expression of neuronal genes and thereby fine-tunes of synapse number. Given that expression of Phf21a-c in neurons leads to increased synapse number, we hypothesize that LSD1-n and PHF21A-n evolved in mammalian nervous system development to promote the proper balance of neuronal gene expression and thus synapse formation.

Although LSD1 is a well-studied histone demethylase which many implications in stem cell differentiation, neurogenesis, and even tumorigenesis (Sheng et al., 2018; Wang et al., 2007; Whyte et al., 2012), PHF21A has not been extensively biochemically characterized. In our biochemical assays, LSD1 and CoREST only demonstrate weak demethylation compared to the strong effect when PHF21A is added. The strong role of PHF21A in the CoREST complex has been previously not well-studied, and thus we believe the physiological demethylase action of

LSD1 requires the examination of PHF21A action upon nucleosomes. Early studies of PHF21A function found that its function is required for LSD1-mediated demethylation, but paradoxically, increased PHF21A concentration inhibited LSD1 demethylase activity (Klajn et al., 2009; Lan et al., 2007; Shi et al., 2005). This present work supports the idea that PHF21A activity enhances LSD1-mediated demethylation *in vitro* and *in vivo*. However, an interesting future experiment would be to repeat demethylation assays while adjusting the relative ratios of LSD1 and PHF21A added. Furthermore, as a H3K4me0 binding protein, it has been hypothesized that PHF21A binds to the product of LSD1-mediated demethylation. This work, however, suggests a more important role for PHF21A in this complex, potentially by recruiting LSD1, but this hypothesis should be formally tested. Furthermore, structural studies of the CoREST complex have focused on LSD1 structure upon nucleosomes (Song et al., 2020; Yang et al., 2006); however, the strong biochemical effect of PHF21A warrants structural work to understand how the PHF21A-containing complex binds nucleosomes. In particular, the coordination of the DNA-binding AT hook domain and the histone binding PHD domain of PHF21A-c is of great structural interest.

Interestingly, LSD1-n has been implicated in activity-dependent gene expression in neurons, including through coordination with stimulus-responsive transcription factors such as SRF (Rusconi et al., 2016a). Molecular characterization of Potocki Shaffer Syndrome patient-derived (*PHF21A* haploinsufficient) cells also demonstrated a defect in stimulus responsiveness, including cAMP/CREB signaling (Porter et al., 2017b). Therefore, an important future direction would include characterizing the role of PHF21A-n in response to stimuli such as neuronal activity. The role of neuronal PHF21A/LSD1 upon stimulation could differ from this present study that analyzed the complex function at baseline. RNA-Seq of mature N-KO (e.g., neurons

expressing Phf21a-c) neurons that have already undergone synaptogenesis will allow us to test whether differences between Phf21a isoforms dictate fine tuning of the transcription of genes important for synapse function.

In summary, we have described a mechanistic example by which chromatin regulators function differently in neurons compared to other cell types. This distinct regulation is important for shaping the neuronal epigenome and transcriptome and has physiological consequences for neurodevelopment. Given the large number of chromatin regulating genes that undergo neuron-specific microexon inclusion, future work will be needed to understand how these other chromatin regulators contribute to the establishment of the neuronal epigenome. A mechanistic understanding of how neuron-specific chromatin regulators function will further illuminate how cell-types establish their transcriptomic and epigenomic identities. Ultimately, these studies will help us understand the pathologic mechanism of cognitive vulnerability upon genetic disruption of chromatin regulators in neurodevelopmental disorders.

Figures
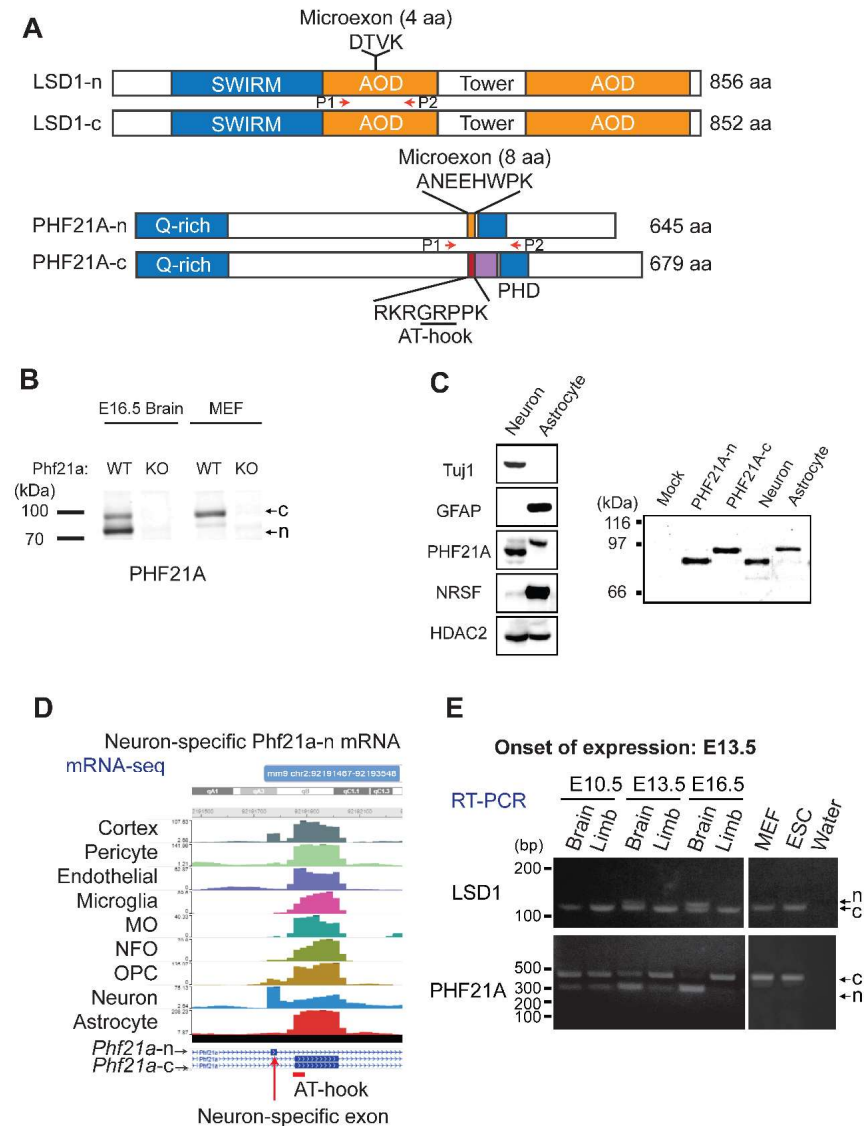
Figure 4-1 Identification of PHF21A-n at the protein and RNA level.

A. Domain structure of canonical and neuronal LSD1 (LSD1-c, LSD1-n) and PHF21A (PHF21A-c, PHF21A-n) isoforms. The LSD1-n neuronal microexon contains a four amino acid insertion in the catalytic Amine Oxidase Domain (AOD). The PHF21A-n neuronal microexon is a mutually exclusive splicing event that disrupts a functional AT Hook domain and is upstream of the Plant Homeodomain (PHD) finger. P1 and P2 indicate primers used for the semi-quantitative PCR in panel E. B. In a PHF21A Western Blot, WT E16.5 mouse brain shows two PHF21A bands while WT Mouse Embryonic Fibroblast (MEF). PHF21A KO cells are also shown since the PHF21A antibody has some non-specific bands. C. Western blot showing that the two brain bands of PHF21A separate into neurons (Tuj1-expressing) and non-neuronal cells (REST-expressing), such as astrocytes (GFAP-expressing). In the right panel, PHF21A-n and PHF21A-c mammalian expression constructs were transfected into HEK293T cells to show comparable size of each isoform to the bands probed in neuron and astrocyte samples. D. Expression of the neuronal PHF21A microexon is specific to neurons as seen in publically-available brain cell-type specific mRNA-Seq (Zhang

et al., 2014). E. RT-PCR for LSD1 and PHF21A shows appearance of the neuronal microexon at E13.5. PHF21A-n expression is mutually exclusive with PHF21A-c expression in neurons.
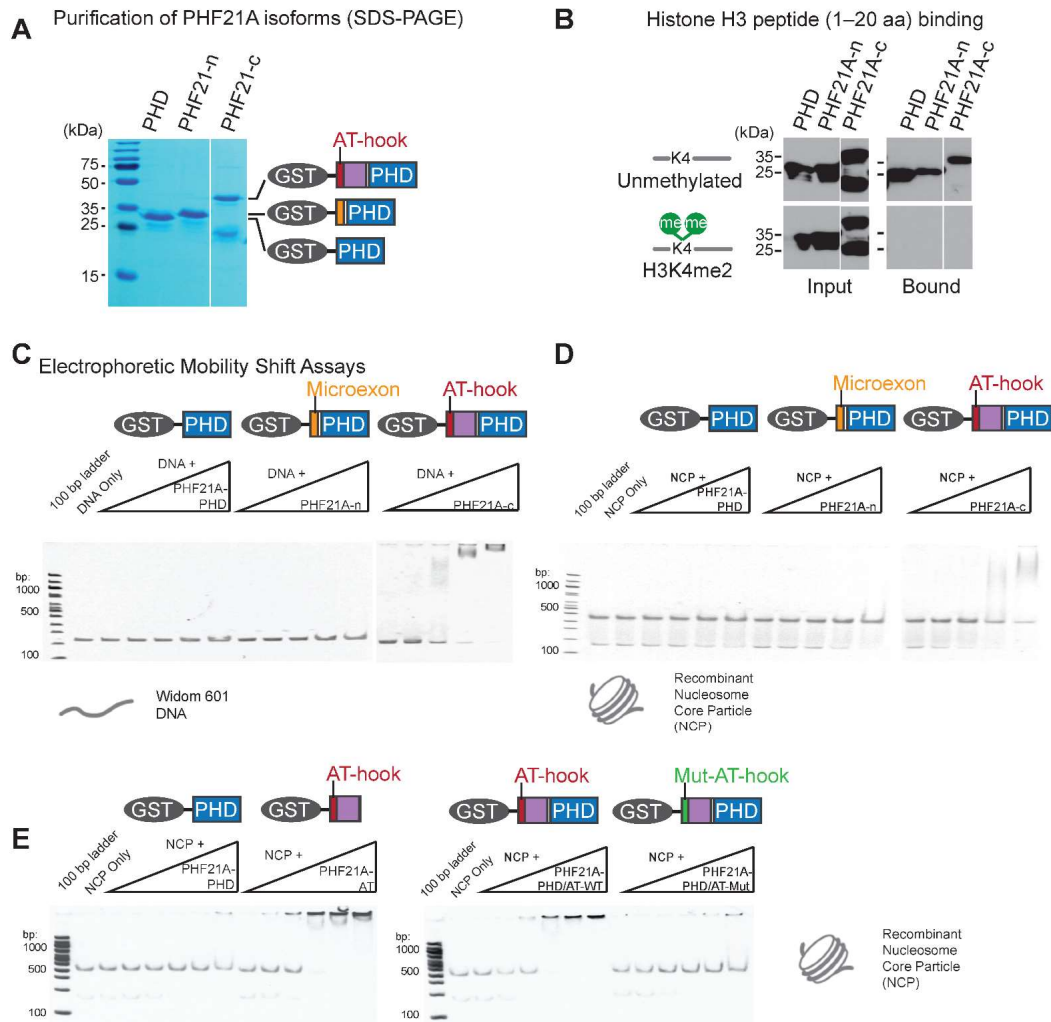
Figure 4-2 PHF21A-c contains a functional AT hook that increases affinity to nucleosomes.

A. Purification of GST-PHF21A-PHD/AT hook fragments. B. Streptavidin-tagged histone H3 (aa1-20) with or without a synthetically added di-methyl at H3K4 were incubated with GST-PHF21A fragments. Shown is GST Western Blot of input and streptavidin bound fractions. C. EMSA result assessing binding between Widom 601 DNA and PHF21A fragments. D. EMSA result assessing binding between recombinant nucleosome core particles (NCP) and PHF21A fragments. E. The PHD finger and AT hook domains were separately expressed and incubated with NCPs to assess binding. F. EMSA comparing WT PHF21A-c PHD/AT hook fragments to mutant PHF21A-c PHD/AT Hook fragments where the three core residues of the AT Hook, GRP (underlined in Figure 1A), were changed to ANE.
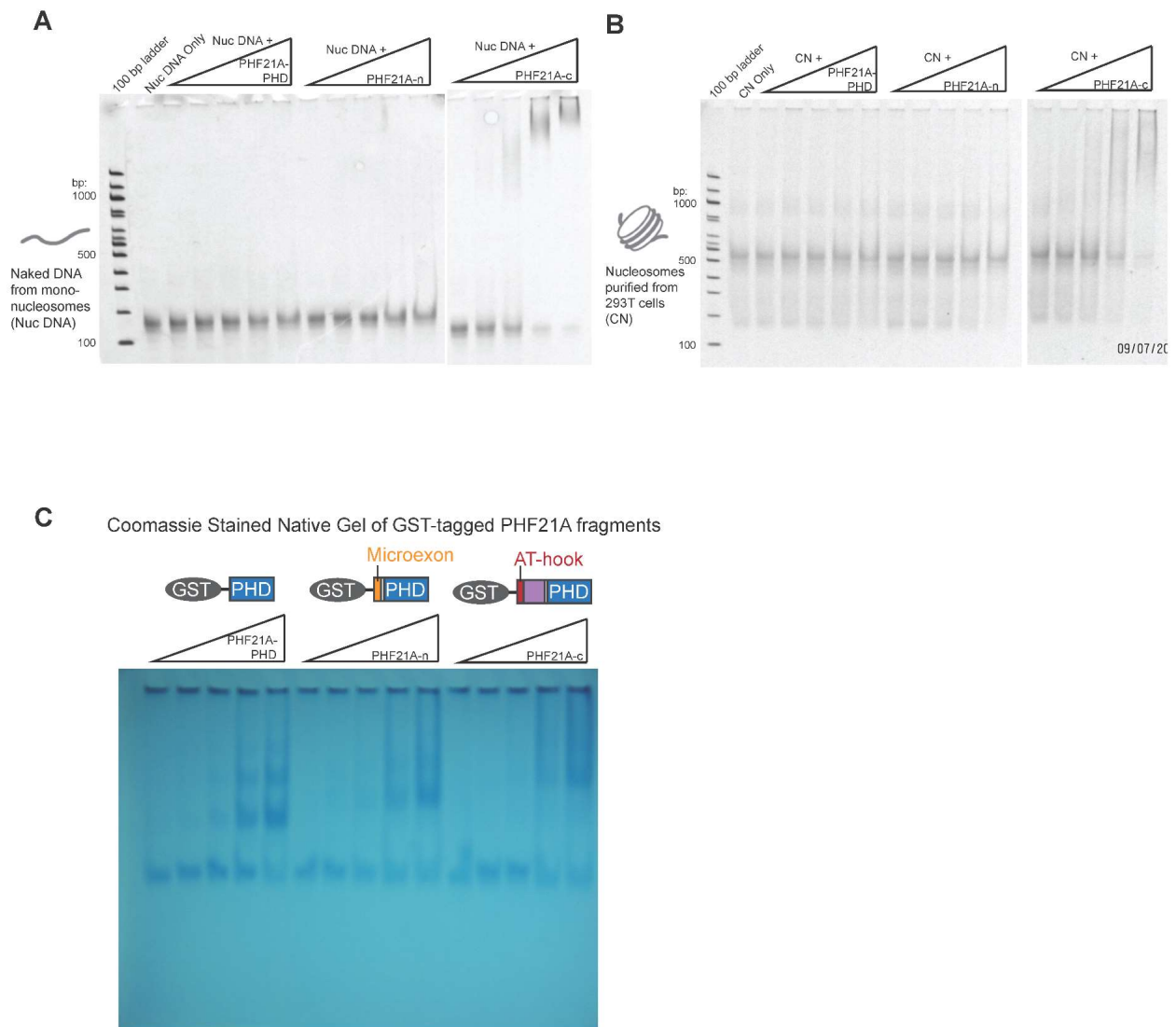
Figure 4-3 Recapitulation of PHF21A EMSA binding using nucleosomes purified from HEK293T cells.

A and B. EMSAs were repeated using nucleosomes purified from HEK293T cells along with the purified DNA fraction from the purified nucleosomes. C. Coomassie Stained Native Gel of GST-Tagged fragments with oligomers visualized.

106
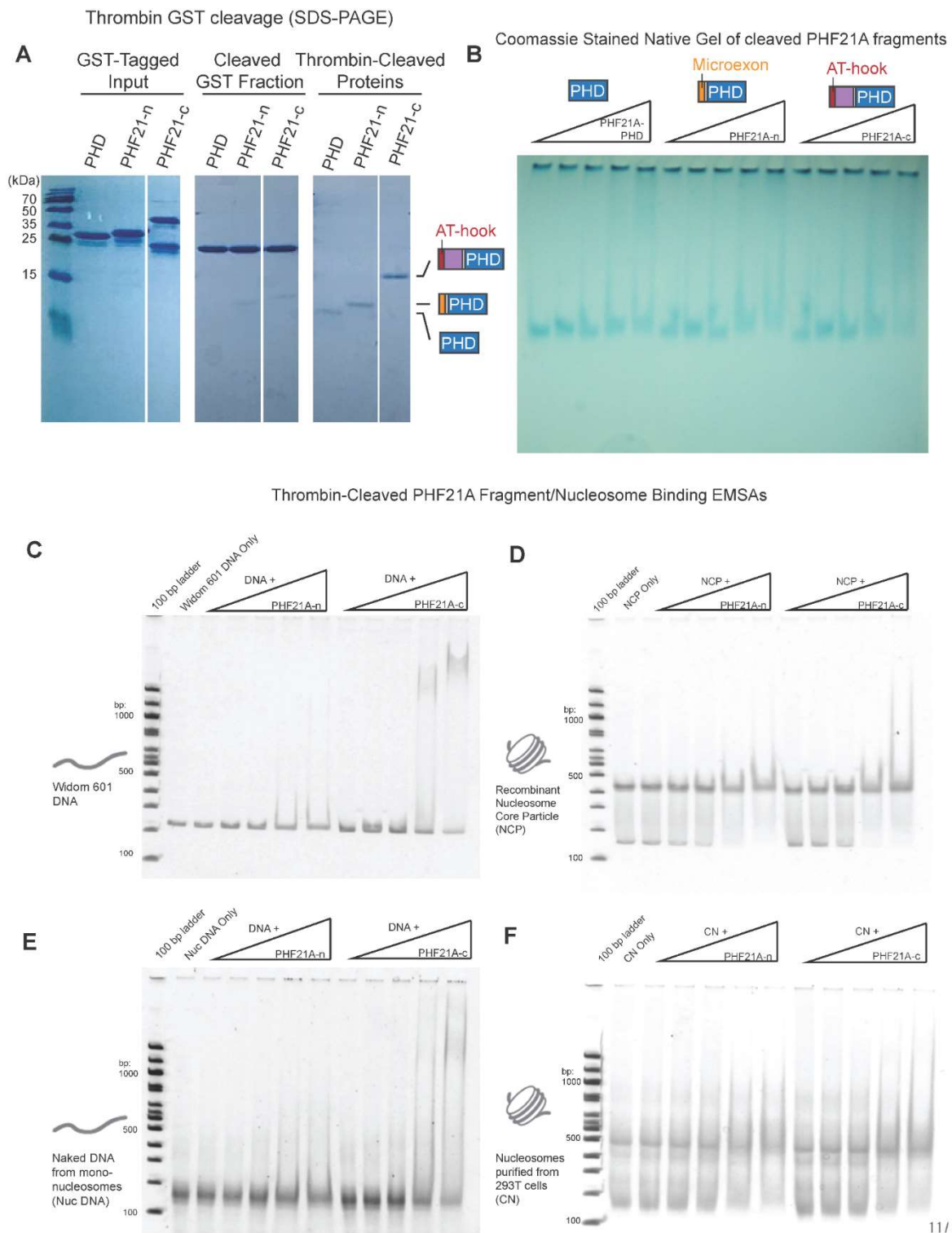
Figure 4-4 Thrombin cleavage of GST and recapitulation of EMSA results.

A. Thrombin was used to cleave GST tag from PHF21A fragments. Shown is SDS-PAGE Coomassie stained gel. B. Coomassie-stained native gel corresponding to EMSA experiment shows cleaved PHF21A fragments do not oligomerize. C-F. Recapitulation of EMSA results using cleaved PHF21A fragments.
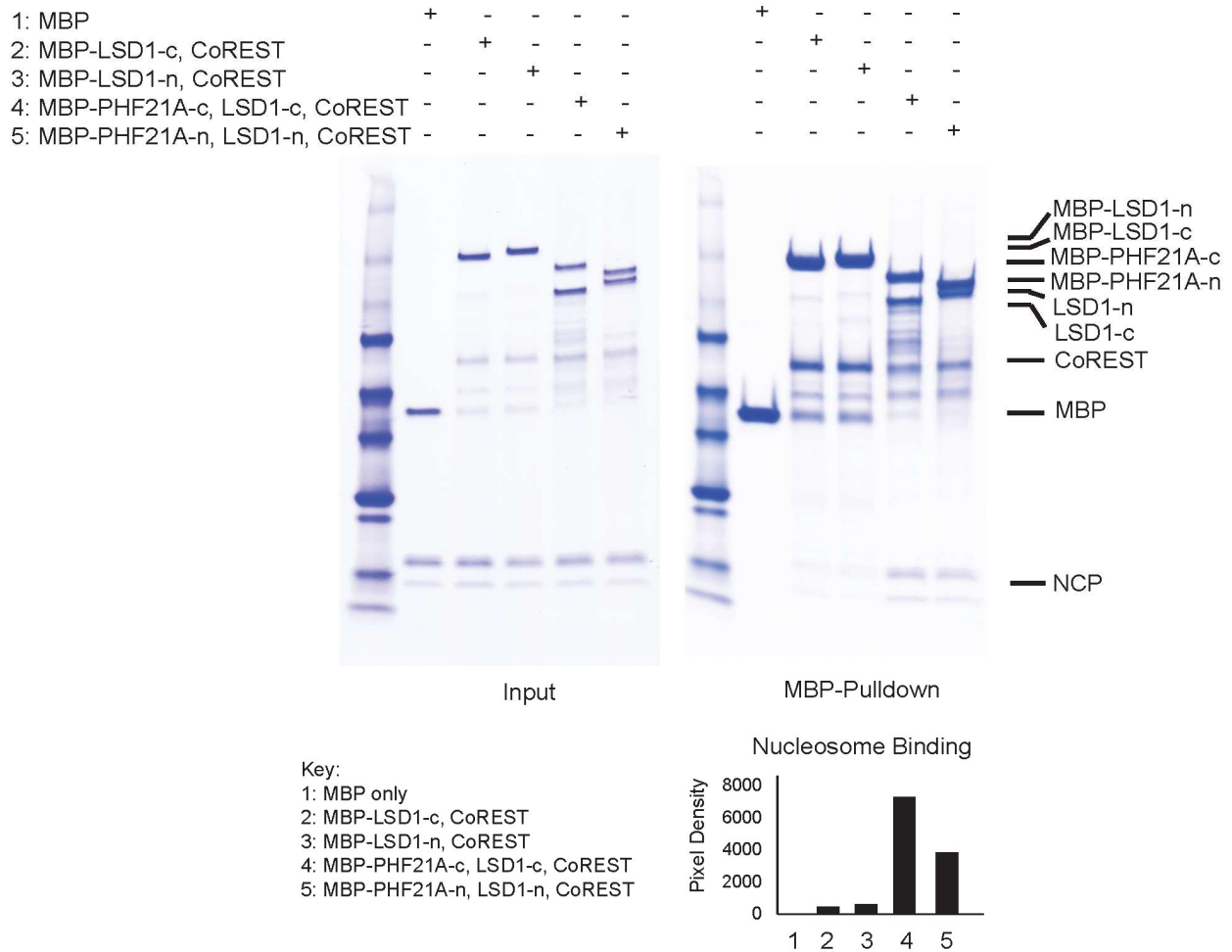
Figure 4-5 Purification of stoichiometric CoREST/PHF21A/LSD1 complexes and nucleosome binding.

Insect cells were infected by baculoviruses with (1) MBP only, (2,3) canonical or neuronal MBP-LSD1 and CoREST, or (4,5) canonical or neuronal MBP-PHF21A, LSD1, and CoREST. The input panel shows the stoichiometric purified complexes following amylose pulldown and FPLC. In this binding assay, Nucleosome Core Particles (NCP) with 177 bp DNA were incubated with purified complexes and pulled down by MBP. Complex members are precipitated and relative amounts of nucleosome pulldown are quantified in the bar graph. Quantification is the average of two trials of this pulldown.

108
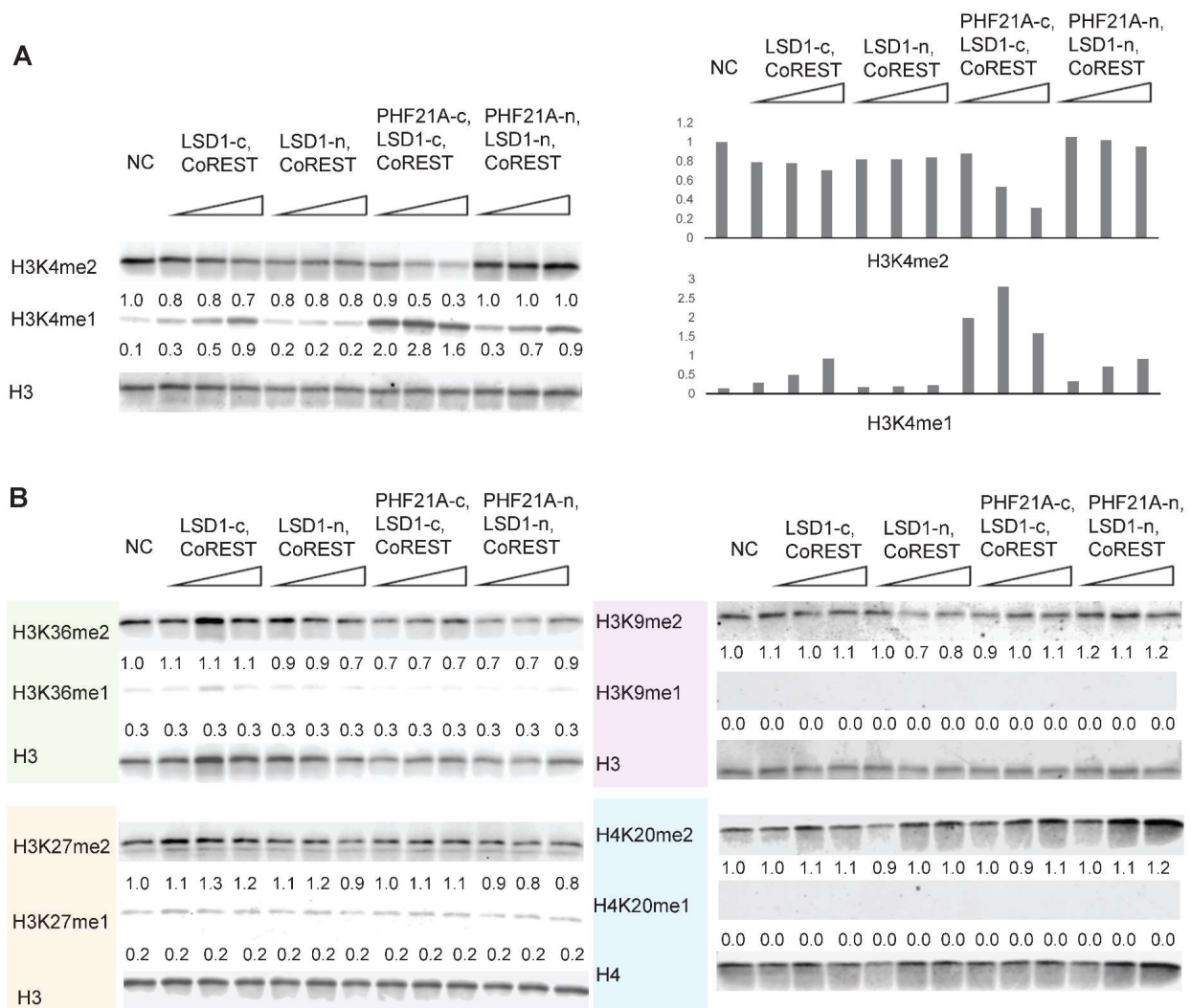
Figure 4-6 Demethylation of recombinant nucleosomes by neuronal and canonical CoREST/PHF21A/LSD1 Complexes.

A. PHF21A/LSD1/CoREST complexes were incubated with recombinant nucleosomes with a synthetically added H3K4me2 modification and then subject to SDS-PAGE followed by quantitative Western Blot. H3K4me2 quantification shows histone modification levels relative to a H3 loading control. The negative control H3K4me1 quantification is normalized to the negative control of H3K4me2, thereby representing the background mono-methyl detection in this assay. B. PHF21A/LSD1/CoREST complexes were incubated with recombinant nucleosomes with synthetically added H3K36me2, H3K9me2, H3K27me2, or H4K20me2. The respective mono- and di-methyl marks were probed by quantitative Western Blotting relative to the respective H3 or H4 control. The same antibody concentrations are used in the subsequent figure where cellular nucleosomes (with heterogeneous modifications) were subject to the same assay to prove the antibodies sensitively detect the substrate.
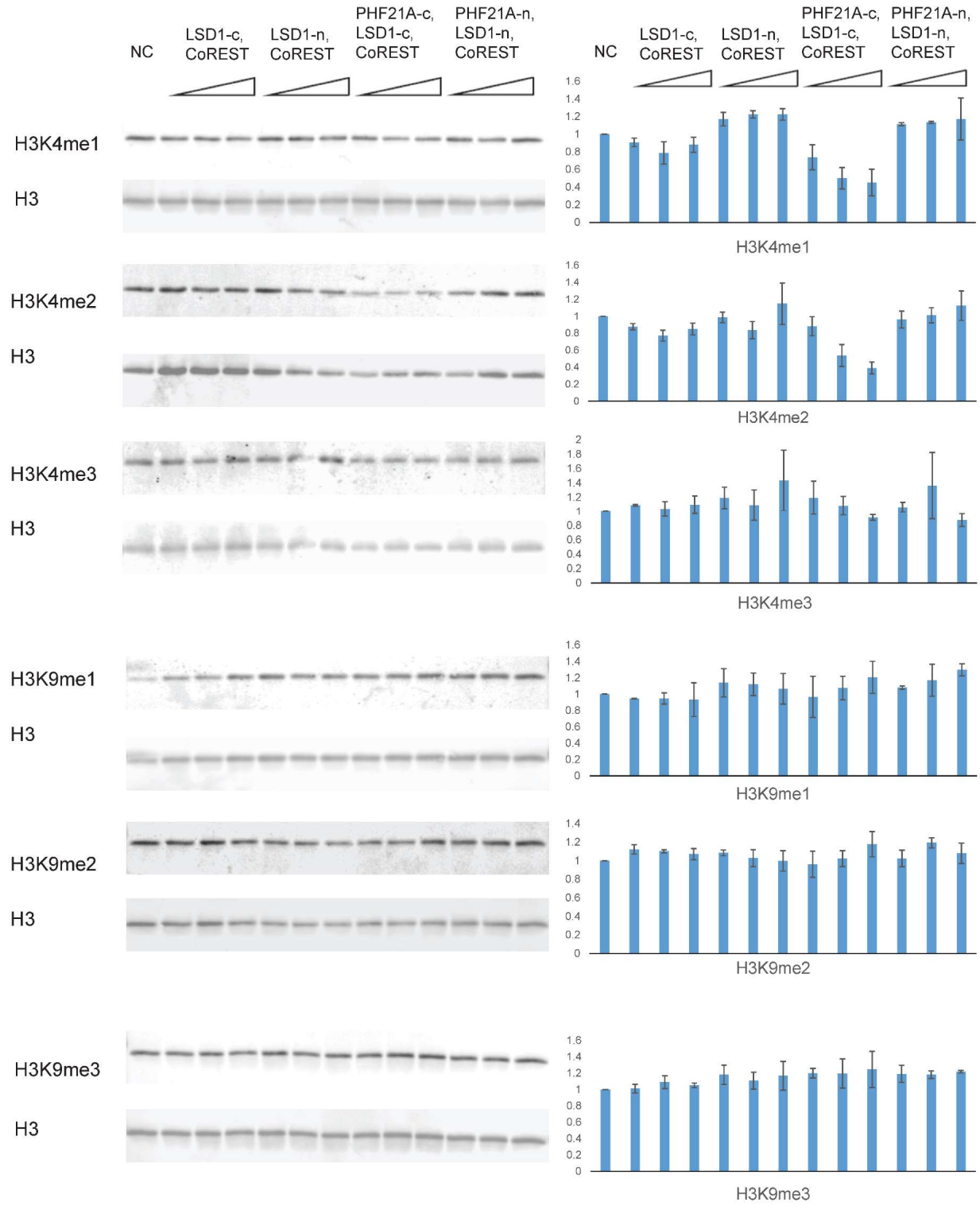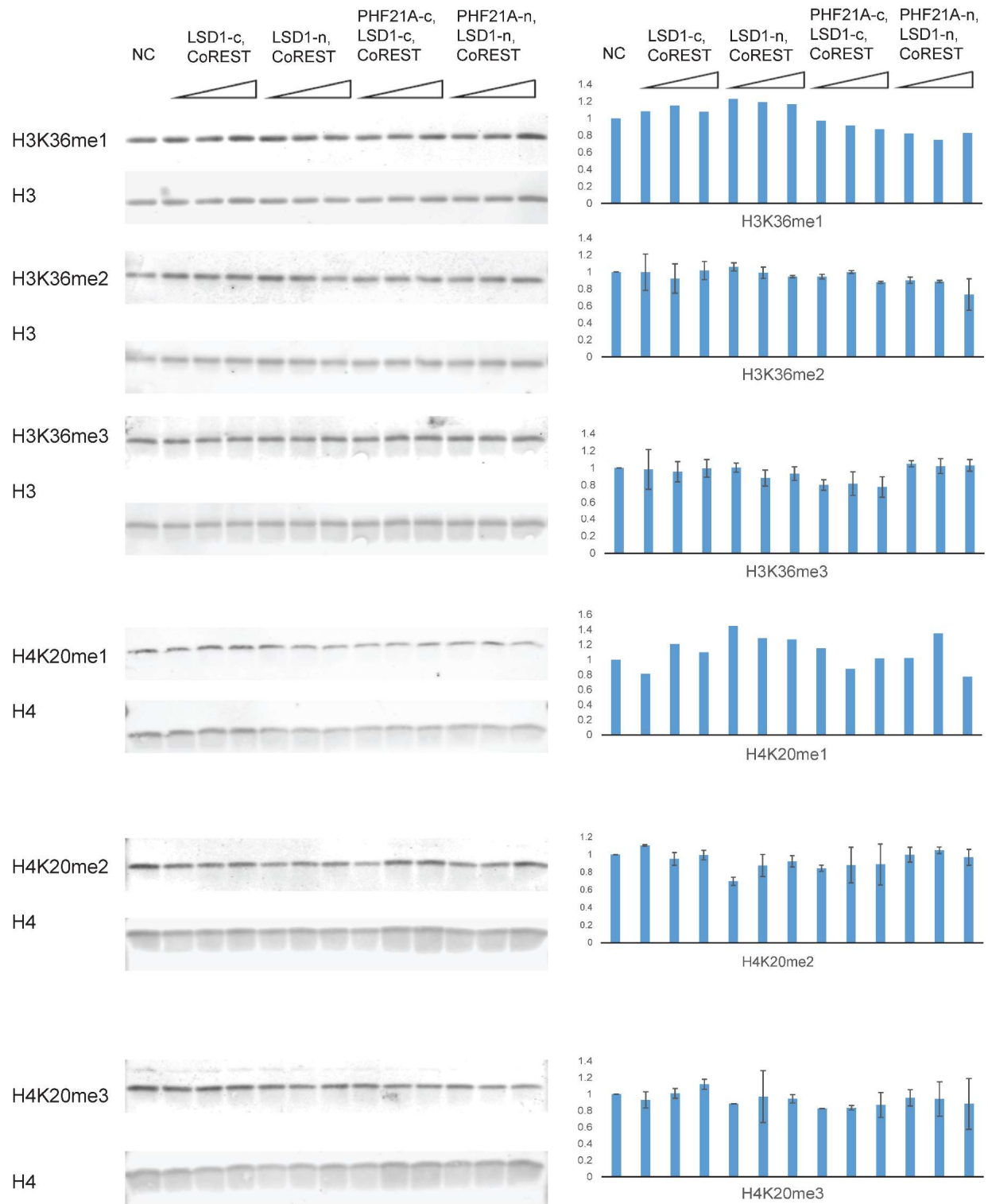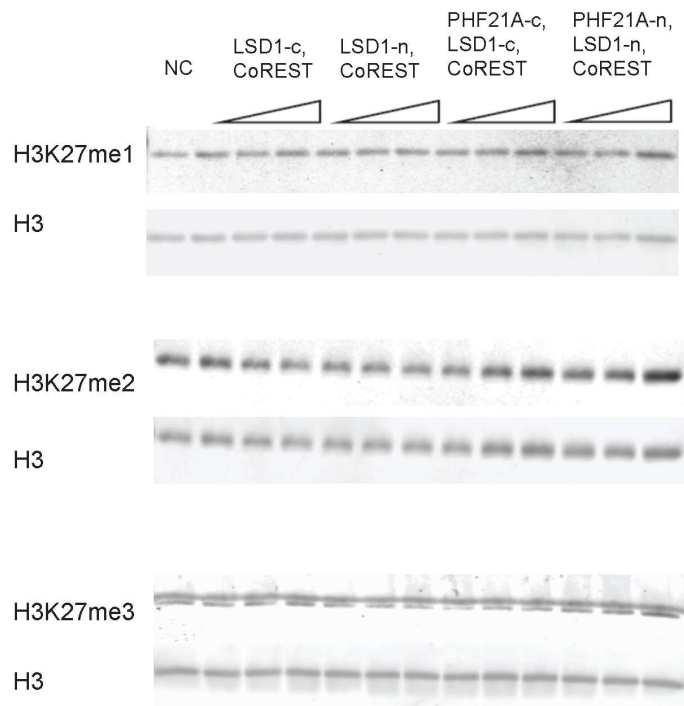
Figure 4-7 Demethylation of cellular nucleosomes by canonical and neuronal CoREST/PHF21A/LSD1 complexes.

PHF21A/LSD1/CoREST complexes were also incubated with nucleosomes that were purified from HEK293T cells. Western Blot probed for the various histone modifications shown and were quantitated relative to a H3 or H4 control. Quantifications are shown as an average of N=2 or 3 independent demethylase experiments and associated quantitative Western Blots. Error bars = SEM.
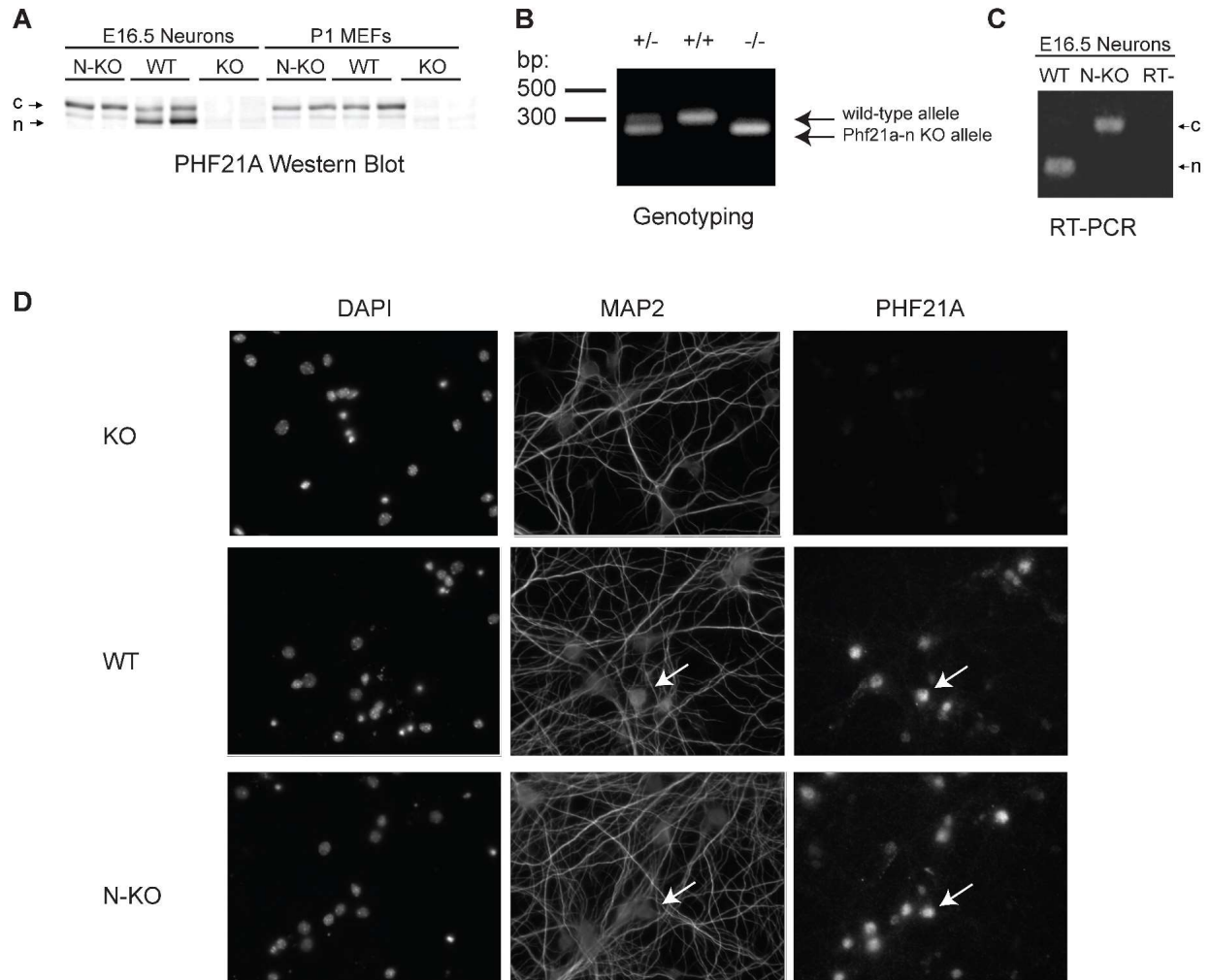
Figure 4-8 Validation of new Phf21a-n-KO (N-KO) mouse model.

A. Whole cell lysate Phf21a Western Blot for WT, KO, and NKO neurons and MEFs. B. Phf21a-N mouse genotyping (Ethidium bromide DNA gel) using primers just outside of the Phf21a-n exon. C. RT-PCR for the Phf21a alternate exon region of WT and NKO mice (the same primers as Figure 1). D. Immunofluorescent images of DIV7 KO, WT, and N-KO neurons using MAP2 and PHF21A antibodies. Arrows indicate cells that are both MAP2 and PHF21A positive indicating that both WT and N-KO neurons are expressing Phf21a.
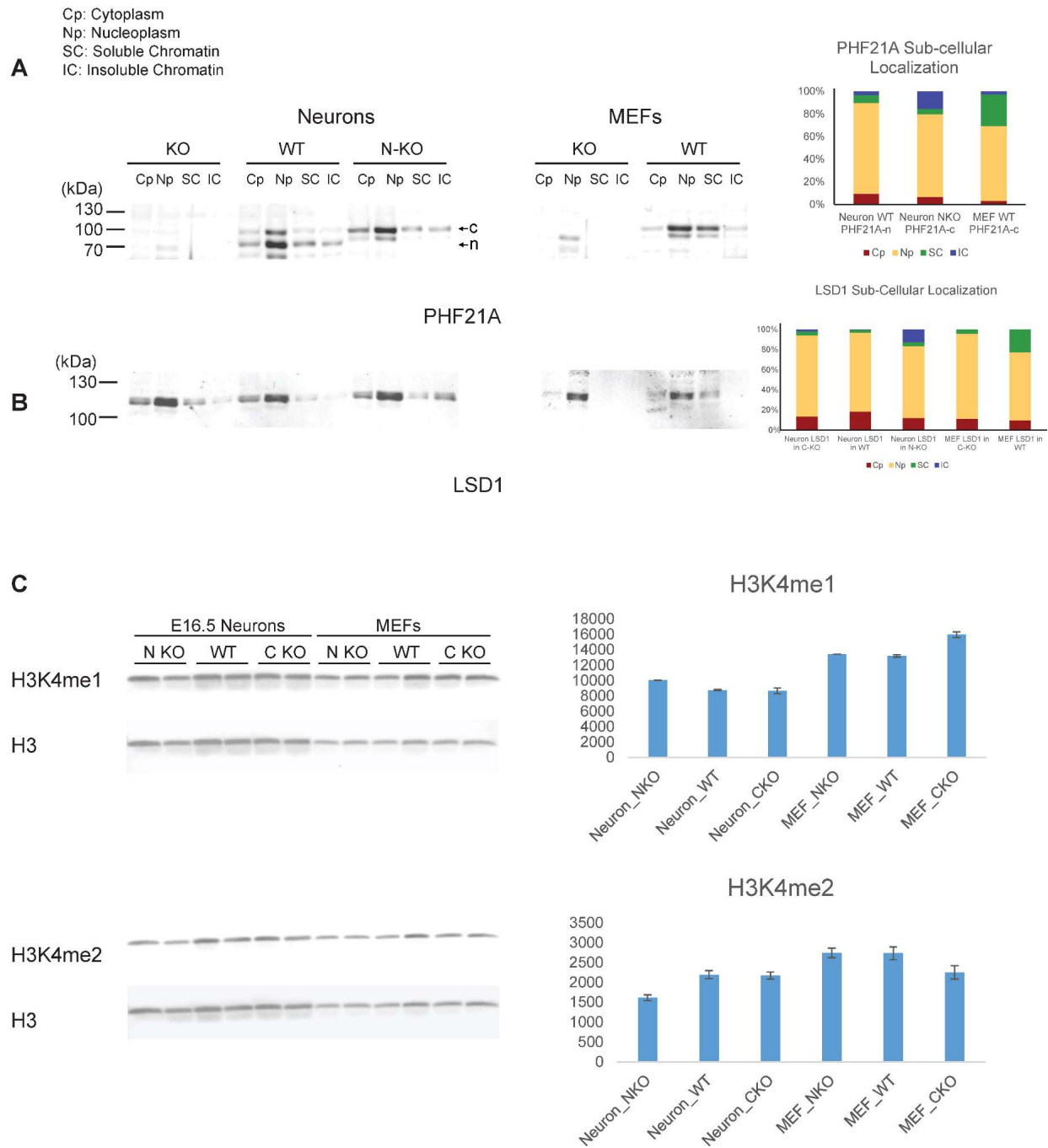
Figure 4-9 PHF21A nucleosome affinity in vivo correlates with LSD1 chromatin occupancy.

A. Cellular fractionation of E16.5 neurons and MEFs into the cytoplasm, nucleoplasm, soluble chromatin, and insoluble chromatin fractions. Shown is Phf21a Western Blot for WT, KO, and NKO neurons and MEFs. The proportion of Phf21a in each fraction is quantified. B. Cellular fractionation results for LSD1 including quantification for each fraction. C. Whole-cell lysates from E16.5 neurons and MEFs were run through SDS-PAGE,

114

and H3K4me1 and H3K4me2 was measured by quantitative Western Blot using H3 as a control. Quantification represents the average of two biological replicates (independent animals) in the blot.

A

Red dots indicates significantly (Adj p-val<0.1) dysregulated genes

Neuron (KO/WT)

MEF (KO/WT)

B

MEF DEGs
121

Neuron DEGs
107
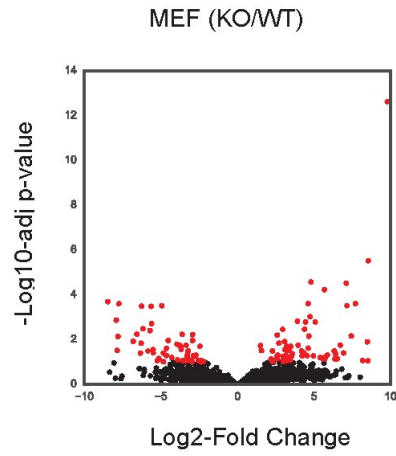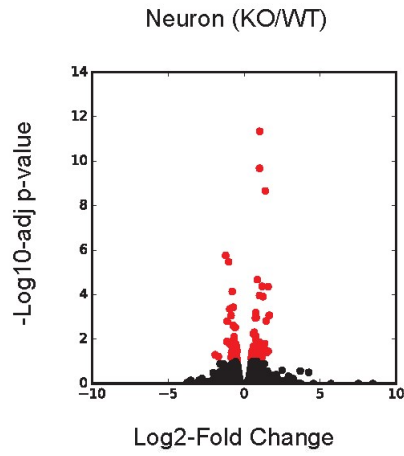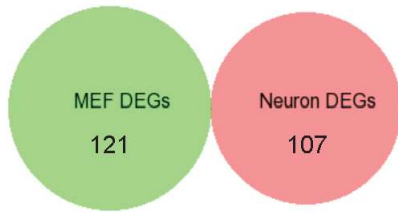
C

REST target genes

	Neuron Null/WT FC	
MEF Null/WT FC	*Neuron-Up* **MEF-Down** 45	*Neuron-Up MEF-Up* 18
	Neuron-Down MEF-Down 25	*Neuron-Down MEF-Up* 24

Chi-squared Test: p-value = 0.02

D

Neuron (Null/WT)

MEF (Null/WT)

-Log10-FDR

-Log10-FDR
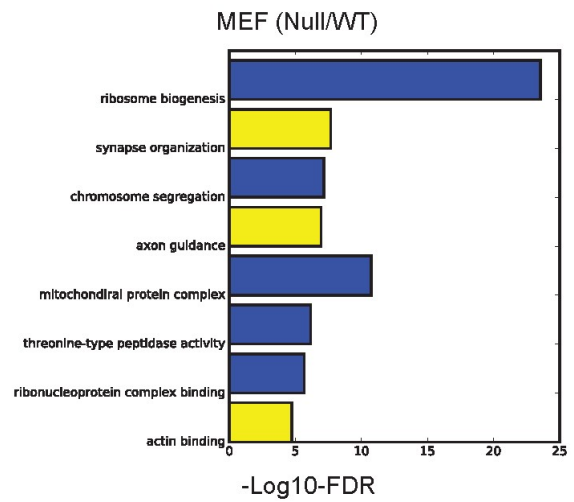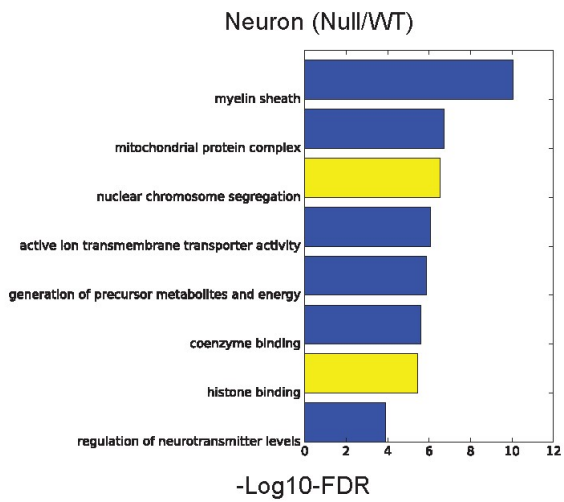
Figure 4-10 RNA-Seq shows distinct transcriptomic control for Phf21a-c vs Phf21a-n.

A. Volcano plots show change in gene expression between *Phf21a*-KO vs WT DIV7 neurons (left) and immortalized MEFs (right). Red dots indicate significantly dysregulated genes with adjusted p-value < 0.1. B. Comparing the significantly dysregulated genes between *Phf21a*-KO neurons and MEFs shows no overlap in dysregulated genes. C. Gene expression changes between *Phf21a*-KO and WT neurons and MEFs were calculated at REST-binding sites, as determined by ChIP-Seq data (Seki et al., 2014). REST-target genes were assigned to one quadrant based on their change in regulation in *Phf21a*-KO neurons and MEFs. D. The RNA-Enrich function was used in LR-Path (Sartor et al., 2009) to determine the most highly dysregulated pathways. Shown in this figure are the most highly dysregulated pathways from the Biological Processes Gene Ontology (GO-BP).

A Anti-PHF21A

B

C After covariate incorporation

D Null + PHF21A-n vs Null   Null + PHF21A-c vs Null

E PHF21A-n rescue

1 rescued gene

Slope = 0.88
[99%CI = 0.84-0.92]

PHF21A-c rescue

11 rescued genes

Slope = 0.83
[99%CI = 0.76-0.89]
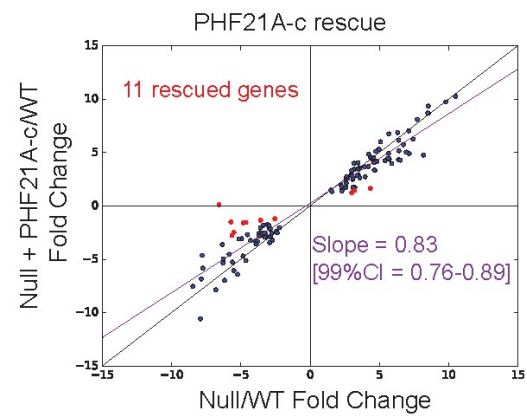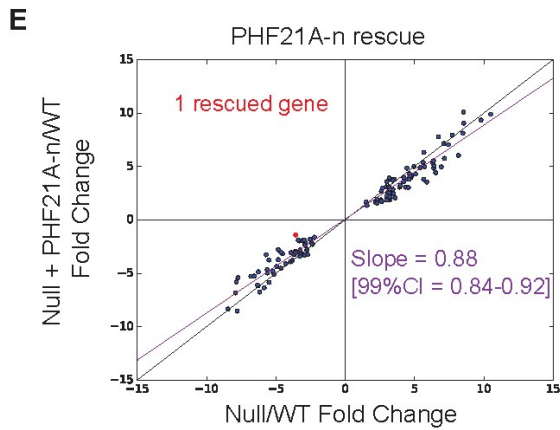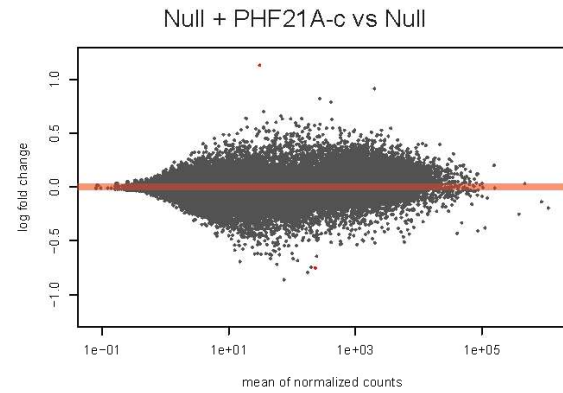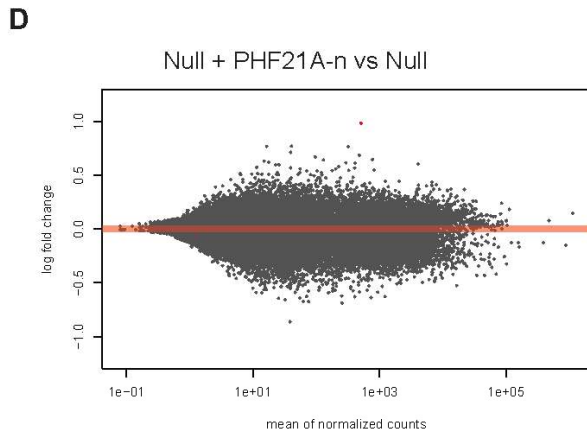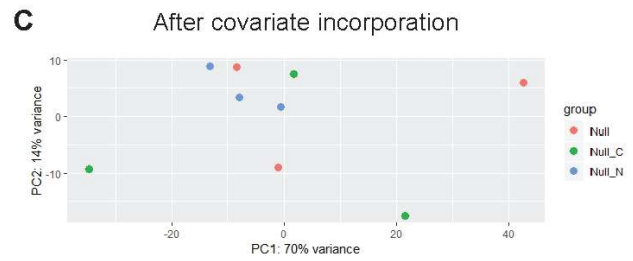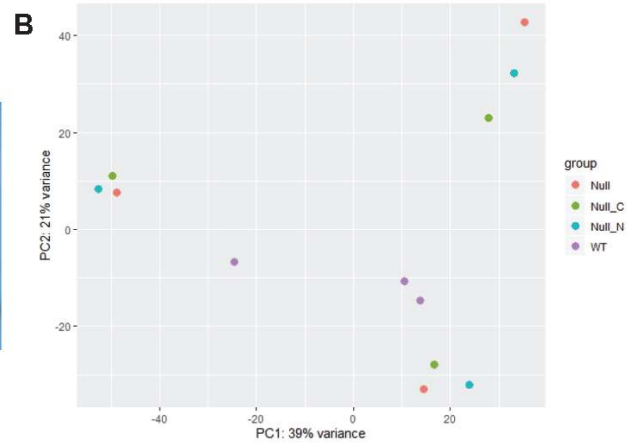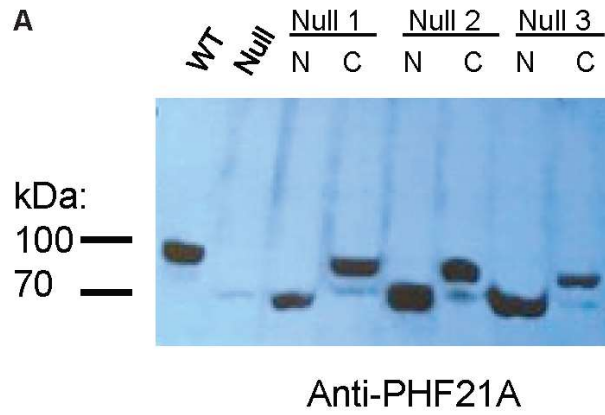
Figure 4-11 Inducible rescue of PHF21A isoforms in Phf21a-Null MEFs.

A. PHF21A Western Blot showing rescue of PHF21A expression through Doxycycline-inducible construct. Shown are three *Phf21a*-Null MEF cell lines with rescued expression of PHF21A-n (N) or PHF21A-c (C). B. Principal Component Analysis of RNA-Seq data from WT, *Phf21a*-Null, *Phf21a*-Null + PHF21A-n (Null_N) rescue, and *Phf21a*-Null + PHF21A-c (Null_C). Samples are segregating according to the animal of origin of each *Phf21a*-Null MEF sample. C. We incorporated a covariate for *Phf21a*-Null animal of origin and plotted the principal component analysis again. D. MA Plots showing rescue of gene expression in PHF21A-n rescue vs. null (left) and PHF21A-c rescue vs. null (right). Red points indicate significantly changed gene expression at an adjusted p-value < 0.1 (16 in PHF21A-n rescue; 16 in PHF21A-c rescue). E. Plot showing *Phf21a*-Null vs WT DEGs with rescue fold change on the y-axis and Null vs. WT on x-axis. We defined rescued genes as those whose absolute value of fold change difference between rescue/WT and null/WT to be greater than 2. Using this criteria, PHF21A-n rescues 1 gene, and PHF21A-c rescues 11 genes.
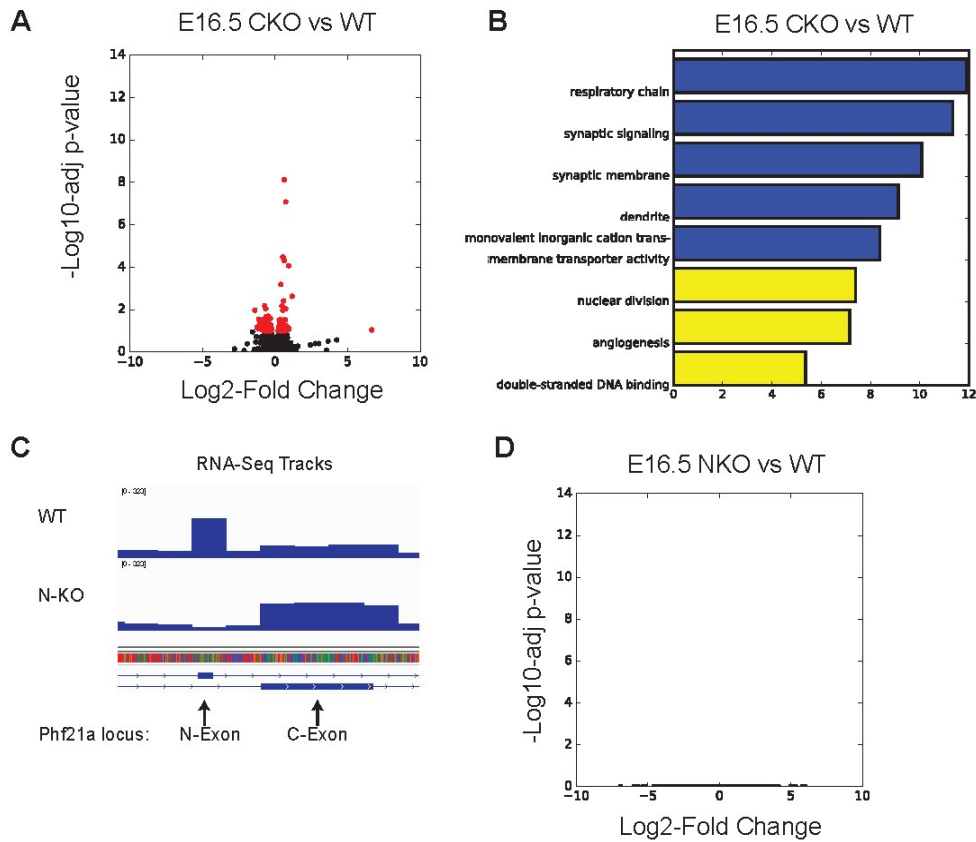
Figure 4-12 Phf21a-N KO results in no gene expression changes at E16.5.

A. Volcano Plot shows changes in gene expression between *Phf21a*-KO and WT E16.5 neurons, recapitulation the previous results with DIV 7 neurons. Red dots indicate significantly dysregulated genes at a q-value < 0.1. B. Volcano plot of N-KO vs WT E16.5 neurons shows no significantly dysregulated genes. C. Ontology analysis for *Phf21a*-KO vs WT E16.5 neurons reveals dysregulation of similar pathways as seen in the DIV7 neurons.
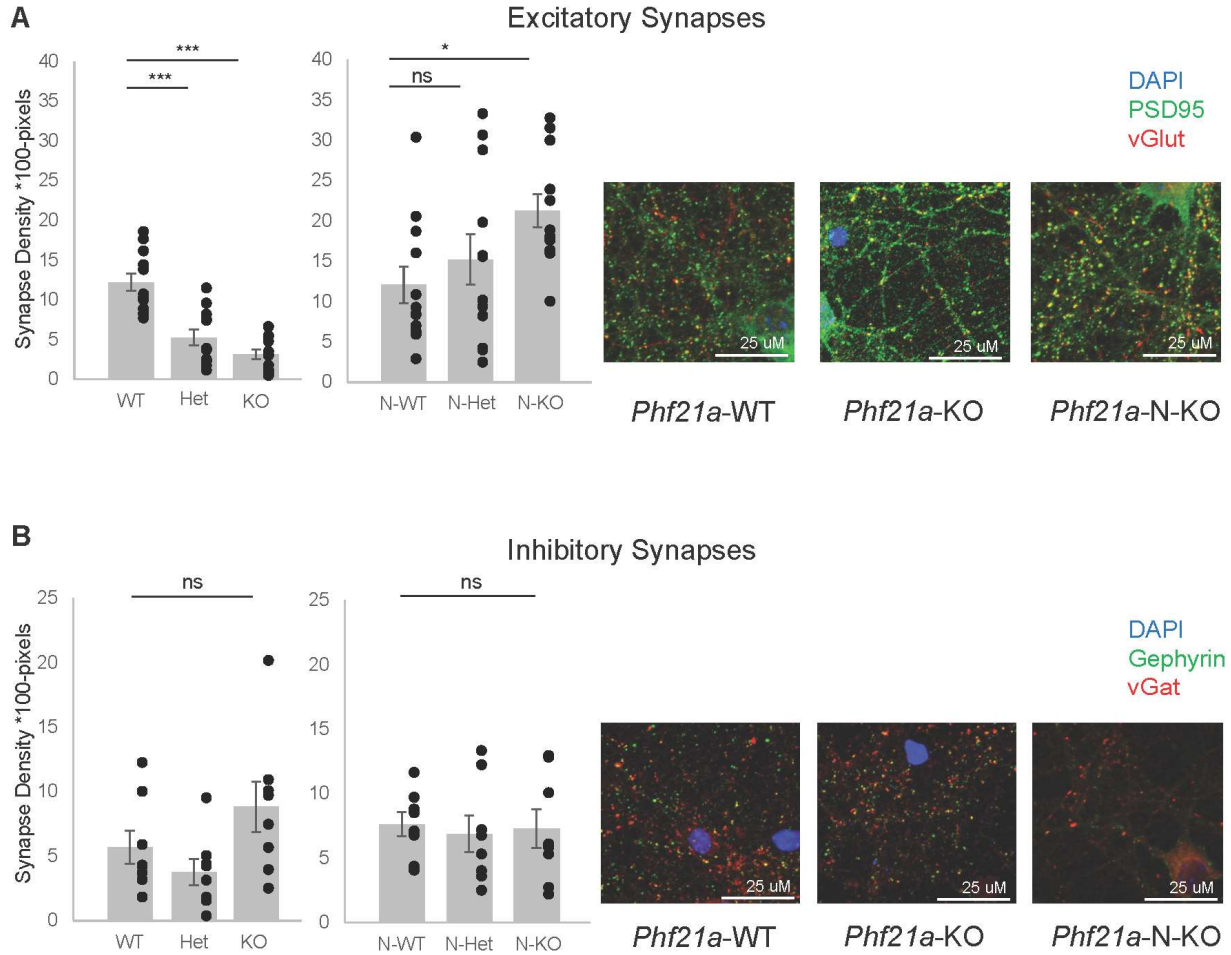
Figure 4-13 PHF21A-n fine tunes excitatory synapse number.

A. Confocal imaging of DIV14 cortical neuron cultures were stained for markers of excitatory synapses (pre-synaptic: vGlut, post-synaptic: PSD95) for all genotypes of *Phf21a*-KO and *Phf21a*-N-KO. Co-localization of pre- and post-synaptic markers were quantified in ImageJ to infer the number of synapses. Quantification of synapse density (per 100-pixels) is shown on the left. B. The same analysis was performed for markers of inhibitory synapses (pre-synaptic: vGat, post-synaptic: Gephyrin). Error bars represent SEM. Statistical significance was calculated using Student's t-test comparing to WT samples. ***: p-value < 0.001; *: p-value < 0.05; ns = not significant, p-value > 0.05.
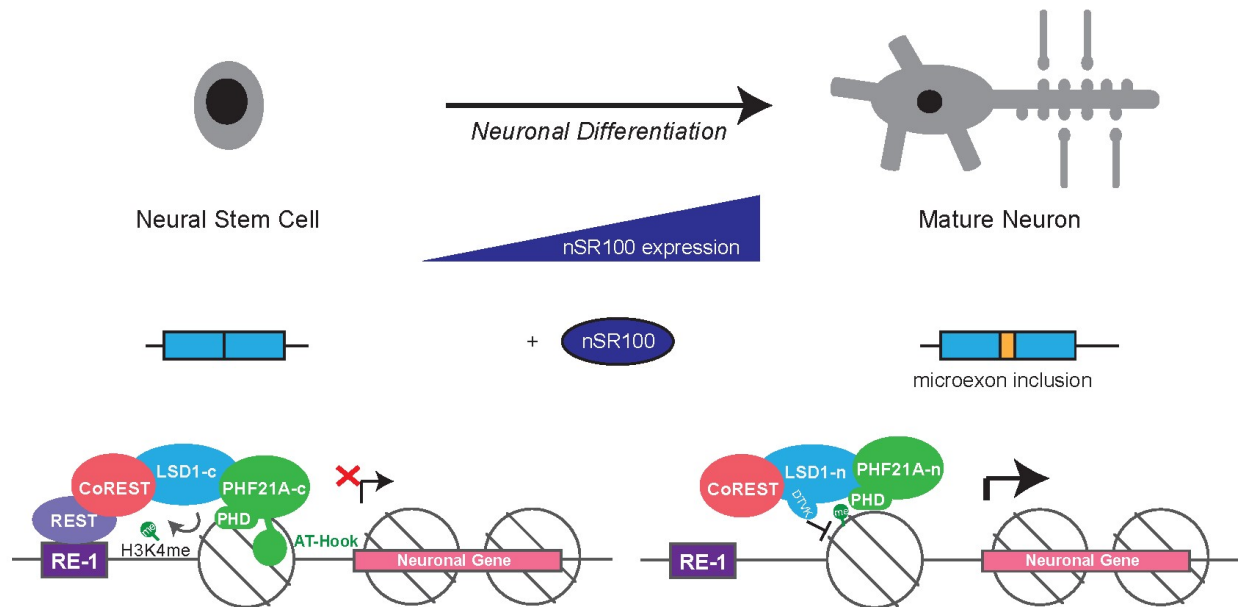
Figure 4-14 Model. Neuronal microexon inclusion of CoREST complex dampens down function during neurodevelopment.

As mature neurons differentiate from neural stem cells, the expression of neuronal splicing factors such as nSR100 increase in expression. This leads to a pattern of microexon inclusion in chromatin regulators and other genes. In the CoREST complex, neuronal microexon inclusion in PHF21A and LSD1 leads to a decreased engagement of nucleosomes and a reduced demethylase enzymatic activity, respectively. This biochemical change in complex function permits higher gene expression for neuronal genes important for neurodevelopment.

Notes and Acknowledgements

Chapter 5 - Conclusions and Future Directions

Individual cell types must selectively activate and repress loci throughout the genome to establish cell identity. Cell-type specific transcription factors, particularly pioneer factors, are thought to be the master regulators of cell fate and the establishment of the appropriate transcriptome (Donaghey et al., 2018; Iwafuchi-Doi and Zaret, 2014). Chromatin regulators are broadly expressed and are thought to have the same intrinsic biochemical activity. Locus-specific action of chromatin regulators has been thought to be directed by cell-specific transcription factors.

Genome-wide studies have identified three major categories of genes that contribute to the pathogenesis of neurodevelopmental disorders such as autism spectrum disorder and intellectual disability: 1) regulators of synapse structure and function, 2) transcription factors that are important for brain development, and 3) chromatin regulators. Unlike the first two categories, chromatin regulators stand out as a group that are ubiquitously expressed and whose function has been thought to be invariant across tissues. These chromatin regulators include proteins that place DNA or histone modifications (writers), proteins that remove these modifications (erasers), and proteins that recognize modifications (readers). Mutations in many writers, readers, and erasers including H3K4 methylation regulators lead to neurodevelopmental disorders with phenotypic overlap (Vallianatos and Iwase, 2015). Given the universally important and genome-wide impact of these proteins, why is that disruption of these factors preferentially leads to

cognitive dysfunction as opposed to dysfunction in another tissue or loss of viability? One could argue that the brain is an organ that is more sensitive to subtle perturbations in gene expression, but my thesis work instead sought a biological explanation for this selective vulnerability.

In my initial review of chromatin regulators in the brain, I identified several chromatin regulators of interest that undergo neuron-specific microexon inclusion. Namely, I identified that multiple members of the CoREST complex including the histone demethylase, LSD1, and the histone reader protein, PHF21A undergo neuron-specific microexon splicing. While there have been recent studies that describe the function of LSD1-n, we identified the presence of PHF21A-n and hypothesized that LSD1-n and PHF21A-n coordinate during neuronal differentiation and maturation to contribute to the establishment of the neuronal histone methylome and consequently the neuronal transcriptome. By reviewing publically available data of microexon splicing, I identified a pattern of chromatin regulators that undergo neuron-specific microexon splicing. Therefore, I hypothesized that this form of splicing could be a basis of how neurons establish a distinct epigenome compared to other tissues. Distinct functioning of these chromatin regulators could be selectively disturbed in neurodevelopmental disorders leading to the cognitive phenotypes seen upon disruption of chromatin regulating genes.

Previous work has also identified epigenomic mechanisms that demonstrate the unique nature of neuronal chromatin compared to other cell types. The most apparent example is the predominance of the hydroxy-methylation DNA modification. Whereas 5-methylcytosine is a common DNA modification found throughout the genome that is associated with gene repression (Bird, 2002), 5-hydroxymethylcytosine, a stable intermediate in the DNA de-methylation

pathway, accumulates to much higher levels in neurons relative to other cell types (Kriaucionis and Heintz, 2009). Furthermore, recent work has found that 5-methyladenosine, another DNA modification, accumulates to much higher level in postnatal neurons and functions to fine-tune cell-type-specific transcription of genes in coordination with the methyl-DNA-binding protein, MeCP2 (He and Ecker, 2015; Stroud et al., 2017). Furthermore, alterations in patterns of DNA methylation have been associated with a wide variety of neuropsychiatric diseases (Kuehner et al., 2019; Rizzardi et al., 2019). The most striking example of the importance of DNA methylation in brain pathology is the disruption of MeCP2 in Rett Syndrome (Lavery and Zoghbi, 2019; Pohodich and Zoghbi, 2015).

Recent work has also shown the role of novel histone modifications to be crucial to neuronal function. One study identified that the neurotransmitter, serotonin, can function as a post translational modification on Histone H3 Glutamine 5 (H3Q5Ser) (Farrelly et al., 2019). This novel histone modification can interact with H3K4me3 to potentiate the action of TFIID-mediated initiation of transcription. Although serotonin acts in other parts of the body, this example demonstrates how neurotransmitter function can uniquely influence the epigenome and transcriptome in neurons.

My thesis work focused on a new aspect of the uniqueness of neuronal chromatin in exploring the neuron-specific microexon-splicing of chromatin regulators. My thesis work demonstrated how splicing of chromatin regulators, particularly histone methylation regulators, leads to distinct function with important consequences in neurodevelopment. In Chapter 1, I reviewed the existing literature on neuron-specific splicing of chromatin regulators. This review included

relevant examples such as MeCP2 along with histone methylation regulators such as LSD1 and G9A.

As my thesis focused specifically on the histone reader protein, PHF21A, I devoted much time to the characterization of PHF21A function in neurodevelopment and in the human disease, Potocki Shaffer Syndrome. Most of the previously published literature about Potocki Shaffer Syndrome was clinical case reports with limited mechanistic work in model organisms.

In Chapter 2, I took advantage of the availability of Potocki-Shaffer Syndrome patient-derived lymphoblasts from a collaborator, Dr. Hyung-Goo Kim. We used RNA-Seq on these samples to do the first molecular pathological characterization of this disease. In doing so, we identified a defective stimulus-response pathway that we were able to test independently through reporter assays using a PHF21A shRNA knockdown strategy. These data collectively demonstrated that PHF21A plays a role in cAMP-mediated gene induction.

To better understand the role of PHF21A in neurodevelopment and in human disease, we next turned to mouse models of Phf21a-deficiency. Our first mouse model is a constitutive *Phf21a* germline knockout that results in complete loss of protein. In Chapter 3, we continued to work on the stimulus-induced gene expression phenotype we observed in the patient-derived lymphoblasts. We were unable to replicate this finding in mouse *Phf21a*-deficient neurons, but this assay should undergo further troubleshooting. We collaborated with both Dr. Natalie Tronson and the Michigan Metabolism, Bariatric Surgery, and Behavior Core to perform a battery of behavioral tests on *Phf21a*-heterozygous adult mice that led to inconclusive results.

Finally, we identified a partially penetrant phenotype of palatal bone fusion in *Phf21a*-null embryos that merits follow up.

Our second mouse model was a germline *Phf21a-N* knockout mouse whereby loss of the neuronal microexon led to expression of Phf21a-c in neurons. Using these two mouse models, we undertook a study of the role of each PHF21A isoform in neurodevelopment. In Chapter 4, we used these mouse models to show Phf21a-c and Phf21a-n appear to regulate distinct sets of genes through RNA-Seq. Additionally, we found that whereas total loss of Phf21a led to a defect of excitatory synapse formation, neuronal expression of Phf21a-c led to an aberrant increase in excitatory synapse number. This aberrant increase is striking as increased balance between excitatory and inhibitory synapses is a hallmark of autism spectrum disorder (Nelson and Valakh, 2015).

In Chapter 4, we also characterized the biochemical function of the neuronal CoREST/LSD1/PHF21A complex. We found that neuronal microexon inclusion in LSD1 leads to an ablation of H3K4 demethylation ability. We further found that neuronal microexon inclusion in PHF21A leads to the loss of an AT Hook domain leading to greatly decreased affinity to nucleosomes. Whereas PHF21A-c was found to greatly increase demethylation capability of LSD1-c, it appears that LSD1-n and PHF21A-n is a much weaker demethylase complex. These biochemical data combined with the results from the mouse models of the role of PHF21A in neurodevelopment have led us to hypothesize that neuronal microexon inclusion in PHF21A and LSD1 functions to dampen down the function of the complex. This weakened complex permits the expression of neuronal genes and fine tunes the synapse number.

This characterization of the biochemical and cellular consequences of neuronal splicing of the CoREST complex demonstrates how chromatin regulators can function distinctly in specific cell types. In the establishment of cell-type-specific epigenomes, organisms have evolved strategies to manipulate the chromatin landscape in ways independent of transcription factor function. Distinct functions of canonical and neuronal LSD1 and PHF21A lead to a unique shaping of the neuronal chromatin landscape.

In this thesis work, I generated multiple independent RNA-Seq data sets. I found in Chapter 4 that there was no overlap in differentially expressed genes (DEGs, q-value <0.1) between the neuron and MEF KO vs WT data sets. However, I next compared the human lymphoblast RNA-Seq with the mouse RNA-Seq DEGs. I converted human DEGs to their mouse orthologues and identified 8 DEG overlaps with the neuron data set and 4 overlaps with the MEF data set (Table 5-1). The neuron data set above includes both DEGs from the E16.5 brain and DIV7 neuron data. The degree of overlap between the two neuron data sets is very high – in the above list, genes are DEGs in both data sets or are close to the q < 0.1 cutoff.

Although this list of common DEGs is relatively small and underpowered to detect pathway differences, a couple features stood out when the list was run through EnrichR (Chen et al., 2013). REST targets, as defined by ChIP Enrichment Analysis (ChEA) (Lachmann et al., 2010) are significantly dysregulated in the above list with a q-value = 0.005. SMAD4 is the second-most enriched transcription factor on this list with a q-value = 0.2. Protein-protein interaction analysis predicted GRIN1 (q=0.03) and GRIN2B (q=0.05) as major hub proteins. Glutamate

receptors, including these two, also were shown to be downregulated in the neuron RNA-Seq analysis, further suggesting that receptor regulation at the synapse is regulated transcriptionally by PHF21A. Ontology analyses are underpowered to detect terms above the False Discovery Rate cutoff; however, the top ontologies that are enriched include cAMP-dependent protein kinase activity ($p = 0.003$), regulation of Wnt signaling ($p = 0.002$), filopodia ($p = 0.0005$), and clathrin vesicle coat ($p = 0.004$). Together, these genes and pathways confirm REST target dysregulation across all of the RNA-Seq data sets. Furthermore, these gene sets point to common abnormalities in pathways important for synaptic function, including receptor function as well as the second messenger pathways that are responsible for the transcriptional response to neuronal signal transduction.

This work has been influenced by the available human genetics knowledge of PHF21A. Interestingly, all patients so far described with Potocki Shaffer Syndrome have microdeletions within the *PHF21A* locus on chromosome 11. The gnomAD browser (Genome Aggregation Database, the successor to the Exome Aggregation Consortium (ExAC) database; pre-print available: Karczewski, et al, 2019. BioRxiv: https://doi.org/10.1101/531210) for the PHF21A gene shows exceptionally high constraint metrics. In other words, there are many fewer missense and predicted loss of function variants observed compared to expected variants indicating this gene is intolerant to variation in the population. A collaborator recently identified the first individual with a missense mutation c.1285G>A in *PHF21A* leading to Potocki Shaffer Syndrome (Fig 5-1). Interestingly, this mutation lies in the last nucleotide of the common exon upstream of the neuronal and canonical exons of *PHF21A*. Importantly, this mutation also changes the amino acid sequence of one the core AT Hook motif amino acids. We are currently

undertaking experiments to predict whether this mutation affects splicing of the PHF21A-c and PHF21A-n transcripts. Furthermore, we are cloning this mutation into a bacterial expression system to express the mutant PHF21A AT hook domains to assess DNA and nucleosome binding changes. Our collaborator is working to obtain and immortalize patient-derived cells to perform a more complete molecular phenotyping of this patient mutation.

Future work must be done to more completely understand the roles of PHF21A and LSD1 neuronal microexon splicing. First, although our work analyzing the two different mouse models at a synaptic and transcriptomic level provided information about the individual roles of PHF21A-n and PHF21A-c, further phenotypic characterization must be done on the *Phf21a-n* KO mouse. As this mouse has recently been generated, many interesting questions can be addressed. For example, *Phf21a-n*-null mice live until adulthood, and a careful examination of the structure and function of the brain should be undertaken. As the mice have grown, many of the cages have been flagged for eye abnormalities, which could be indicative of some underlying brain pathology.

Furthermore, I continue to be intrigued by the result described in Chapter 2 whereby PHF21A-deficient lymphoblasts exhibit a temporal delay in immediate early gene induction. The efficient transcription of immediate early genes following a stimulus involves rapid changes to the chromatin structure of the associated genomic loci, and it would be interesting to better understand the mechanism of how PHF21A plays a role in this process. Recent studies on the function of LSD1-n have also focused on the recruitment of stimulus-responsive transcription factors such as SRF (Rusconi et al., 2016a). Given the described difference in nucleosome

131

affinity, it would be especially interesting to dissect the differences between PHF21A-c and PHF21A-n action in the induction of immediate early genes.

Our biochemical results showing great enhancement of LSD1-c activity by PHF21A-c raises the question of the mechanism of CoREST complex function. The known role for PHF21A was as a H3K4me0 binding protein, which is the theoretical product of the LSD1-c enzymatic reaction. How exactly do the DNA-binding AT Hook domain and the PHD finger coordinate to contact nucleosomes? Structural work visualizing the canonical and neuronal CoREST/LSD1/PHF21A complex on nucleosomes would shed light on this mechanism. Does PHF21A bind first to the DNA and then recruit LSD1? Or does PHF21A bind to the reaction product and allow for coordination of the adjacent demethylation reaction? Well-controlled biochemical and genomic studies should be undertaken to address these possibilities.

In this study, I focused on the action of PHF21A and LSD1 within the CoREST complex. Indeed, pulldown of PHF21A-c yielded co-immunoprecipitation of the CoREST complex members (Iwase et al., 2004). LSD1-c is a known member of both the CoREST and NuRD complexes (Wang et al., 2009). However, a major question remains in that PHF21A-n and LSD1-n may interact with different proteins in neurons. One group identified that LSD1-n interacts with SVIL in order to mediate demethylation of H3K9me in human neuroblastoma cells (Laurent et al., 2015). SVIL is a known androgen receptor (AR)-interactor, and these data fit into literature of LSD1 interacting with AR in the removal of repressive H3K9 methylation marks (Metzger et al., 2005). Immunoprecipitation followed by Mass Spectrometry (IP-MS) comparing the interacting partners of PHF21A-c vs PHF21A-n would elucidate whether PHF21A-n

participates in the action of a distinct complex in neurons. Another way to understand the differences between PHF21A-c and PHF21A-n action would be through genomic localization studies. I attempted PHF21A ChIP-Seq several times but was unable to detect any signal enrichment over the input. This failure may be a result of the new antibody, so further troubleshooting possibly with a different antibody could yield better results. A potential alternate approach would be analysis of LSD1 genome localization in our various Phf21a mouse models in both neurons and fibroblasts.

Finally, we believe that the example of neuronal microexon splicing of the CoREST complex is only the beginning of understanding how cell-type specific chromatin regulators contribute to shaping a cell-type specific epigenomic landscape. *In silico* tools may be able to predict disruptions in protein-protein networks caused by microexon inclusion. Approaches similar to the one we undertook in this work could also shed light on other chromatin regulating complexes that differ between tissues. A mechanistic understanding of how neuron-specific chromatin regulators function will further illuminate how cell-types establish their transcriptomic and epigenomic identities. Ultimately, these studies will help us understand the pathologic mechanism of selective vulnerability in the brain upon genetic disruption of chromatin regulators in neurodevelopmental disorders.

Tables

Gene	Gene Info	Dysregulation in *PHF21A*-deficient lymphoblasts	Dysregulation in *Phf21a*-null neurons
Prkacb	protein kinase cAMP-activated catalytic subunit beta	Down	Down
Glt1d1	glycosyltransferase 1 domain containing 1	Down	Down
Igf2bp1	insulin like growth factor 2 mRNA binding protein 1	Up	Up
Kdm7a	lysine demethylase 7A	Up	Up
Myo1b	myosin IB	Up	Up
Antxr2	ANTXR cell adhesion molecule 2	Up	Up
Gap43	growth associated protein 43 (associated with neuronal growth)	Up	Down
Ccnjl	cyclin J like	Down	Up

Gene	Gene Info	Dysregulation in *PHF21A*-deficient lymphoblasts	Dysregulation in *Phf21a*-null MEFs
Tbkbp1	TBK1 binding protein 1	Down	Down
Bmp2	bone morphogenetic protein 2	Down	Down
Cxxc4	CXXC finger protein 4	Down	Up
Ncald	neurocalcin delta	Down	Up

Table 5-1 Common differentially expressed genes across RNA-Seq experiments

Differentially Expressed Genes (q-value < 0.1) that are common to the *PHF21A*-haploinsufficient lymphoblast and *Phf21a*-null mouse (neuron and Mouse Embryonic Fibroblast) RNA-Seq data sets. Human genes from the lymphoblast RNA-Seq were converted to their mouse orthologues and overlap was determined with the neuron and MEF RNA-Seq data.

Figures
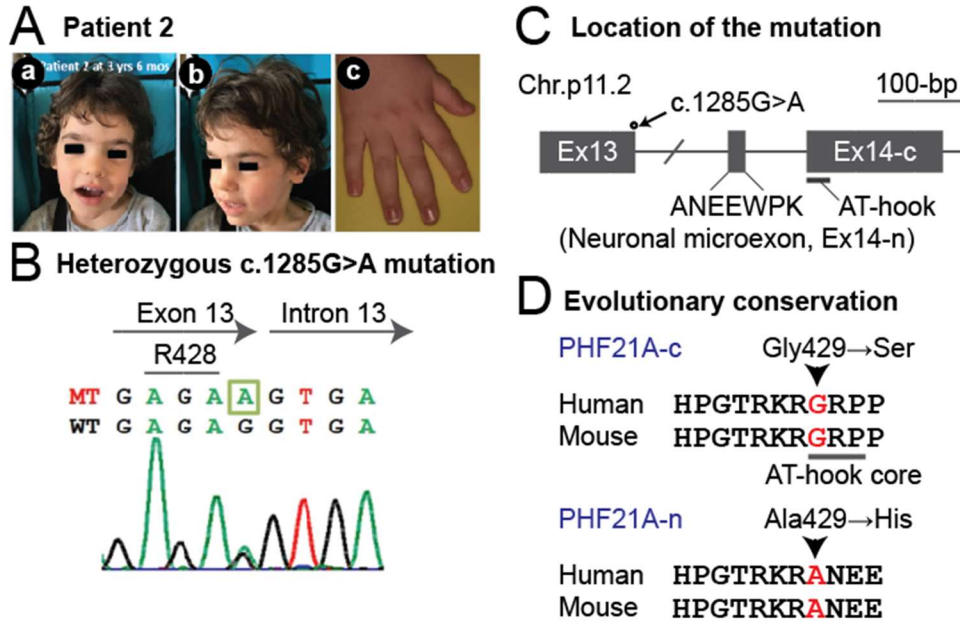
Figure 5-1 Novel PHF21A Missense Patient Mutation in AT Hook Domain.

A, B, and C. The first missense mutation in *PHF21A* has been identified in an individual with Potocki Shaffer Syndrome. This heterozygous mutation, c. 1285G>A, lies in the last nucleotide of exon 13, upstream of the genomic region of the alternatively spliced neuronal and canonical exons. D. This residue is conserved across mammals in both PHF21A-c and PHF21A-n. The codon where this mutation lies spans an intron such that the predicted amino acid change is Gly429Ser in PHF21A-c and Ala429His in PHF21A-n.

Bibliography

Agarwal, S., Macfarlan, T.S., Sartor, M.A., and Iwase, S. (2015). Sequencing of first-strand cDNA library reveals full-length transcriptomes. Nat Commun *6*, 6002.

Akamatsu, W., Fujihara, H., Mitsuhashi, T., Yano, M., Shibata, S., Hayakawa, Y., Okano, H.J., Sakakibara, S., Takano, H., Takano, T.*, et al.* (2005). The RNA-binding protein HuD regulates neuronal cell identity and maturation. Proc Natl Acad Sci U S A *102*, 4625-4630.

Alasbahi, R.H., and Melzig, M.F. (2012). Forskolin and derivatives as tools for studying the role of cAMP. Die Pharmazie *67*, 5-13.

Allen, M., Bird, C., Feng, W., Liu, G., Li, W., Perrone-Bizzozero, N.I., and Feng, Y. (2013). HuD promotes BDNF expression in brain neurons via selective stabilization of the BDNF long 3'UTR mRNA. PloS one *8*, e55718.

Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nature reviews Genetics *17*, 487-500.

Ambrosi, C., Manzo, M., and Baubec, T. (2017). Dynamics and Context-Dependent Roles of DNA Methylation. J Mol Biol *429*, 1459-1475.

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. Nature genetics *23*, 185-188.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.

Ataman, B., Boulting, G.L., Harmin, D.A., Yang, M.G., Baker-Salisbury, M., Yap, E.L., Malik, A.N., Mei, K., Rubin, A.A., Spiegel, I*., et al.* (2016). Evolution of Osteocrin as an activity-regulated factor in the primate brain. Nature *539*, 242-247.

Auweter, S.D., Fasan, R., Reymond, L., Underwood, J.G., Black, D.L., Pitsch, S., and Allain, F.H. (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. The EMBO journal *25*, 163-173.

Bachinski, L.L., Sirito, M., Bohme, M., Baggerly, K.A., Udd, B., and Krahe, R. (2010). Altered MEF2 isoforms in myotonic dystrophy and other neuromuscular disorders. Muscle & nerve *42*, 856-863.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R*., et al.* (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science (New York, NY) *338*, 1587-1593.

Basarsky, T.A., Parpura, V., and Haydon, P.G. (1994). Hippocampal synaptogenesis in cell culture: developmental time course of synapse formation, calcium influx, and synaptic protein distribution. The Journal of neuroscience : the official journal of the Society for Neuroscience *14*, 6402-6411.

Bayraktar, G., and Kreutz, M.R. (2017). Neuronal DNA Methyltransferases: Epigenetic Mediators between Synaptic Activity and Gene Expression? Neuroscientist, 1073858417707457.

Belgard, T.G., Marques, A.C., Oliver, P.L., Abaan, H.O., Sirey, T.M., Hoerder-Suabedissen, A.,
Garcia-Moreno, F., Molnar, Z., Margulies, E.H., and Ponting, C.P. (2011). A
transcriptomic atlas of mouse neocortical layers. Neuron *71*, 605-616.

Benevento, M., Iacono, G., Selten, M., Ba, W., Oudakker, A., Frega, M., Keller, J., Mancini, R.,
Lewerissa, E., Kleefstra, T.*, et al.* (2016). Histone Methylation by the Kleefstra
Syndrome Protein EHMT1 Mediates Homeostatic Synaptic Scaling. Neuron *91*, 341-355.

Benito, E., and Barco, A. (2015). The neuronal activity-driven transcriptome. Mol Neurobiol *51*,
1071-1088.

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

Bourtchuladze, R., Frenguelli, B., Blendy, J., Cioffi, D., Schutz, G., and Silva, A.J. (1994).
Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive
element-binding protein. Cell *79*, 59-68.

Boutz, P.L., Stoilov, P., Li, Q., Lin, C.H., Chawla, G., Ostrow, K., Shiue, L., Ares, M., Jr., and
Black, D.L. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-
binding proteins reprograms alternative splicing in developing neurons. Genes &
development *21*, 1636-1652.

Bronicki, L.M., and Jasmin, B.J. (2013). Emerging complexity of the HuD/ELAVl4 gene;
implications for neuronal development, function, and dysfunction. Rna *19*, 1019-1037.

Brown, S.E., Campbell, R.D., and Sanderson, C.M. (2014). Novel NG36/G9a gene products
encoded within the human and mouse MHC class III regions. Mammalian Genome *12*,
916-924.

Bruce, A.W., Donaldson, I.J., Wood, I.C., Yerbury, S.A., Sadowski, M.I., Chapman, M., Gottgens, B., and Buckley, N.J. (2004). Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. Proc Natl Acad Sci U S A *101*, 10458-10463.

Brunkan, A.L., and Goate, A.M. (2005). Presenilin function and gamma-secretase activity. J Neurochem *93*, 769-792.

Buckanovich, R.J., Posner, J.B., and Darnell, R.B. (1993). Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. Neuron *11*, 657-672.

Buckanovich, R.J., Yang, Y.Y., and Darnell, R.B. (1996). The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. The Journal of neuroscience : the official journal of the Society for Neuroscience *16*, 1114-1122.

Buljan, M., Chalancon, G., Eustermann, S., Wagner, G.P., Fuxreiter, M., Bateman, A., and Babu, M.M. (2012). Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Mol Cell *46*, 871-883.

Bush, J.O., and Jiang, R. (2012). Palatogenesis: morphogenetic and molecular mechanisms of secondary palate development. Development *139*, 231-243.

Calarco, J.A., Superina, S., O'Hanlon, D., Gabut, M., Raj, B., Pan, Q., Skalska, U., Clarke, L., Gelinas, D., van der Kooy, D*., et al.* (2009). Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. Cell *138*, 898-910.

Chahrour, M.H., Yu, T.W., Lim, E.T., Ataman, B., Coulter, M.E., Hill, R.S., Stevens, C.R., Schubert, C.R., Collaboration, A.A.S., Greenberg, M.E*., et al.* (2012). Whole-exome

sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. PLoS Genet *8*, e1002635.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC bioinformatics *14*, 128.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nature reviews Molecular cell biology *10*, 741-754.

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. Nature genetics *27*, 327-331.

Chen, Z.F., Paquette, A.J., and Anderson, D.J. (1998). NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. Nature genetics *20*, 136-142.

Chong, J.A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altshuller, Y.M., Frohman, M.A., Kraner, S.D., and Mandel, G. (1995). REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. Cell *80*, 949-957.

Chong, J.X., Yu, J.H., Lorentzen, P., Park, K.M., Jamal, S.M., Tabor, H.K., Rauch, A., Saenz, M.S., Boltshauser, E., Patterson, K.E.*, et al.* (2016). Gene discovery for Mendelian conditions via social networking: de novo variants in KDM1A cause developmental delay and distinctive facial features. Genet Med *18*, 788-795.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B*., et al.* (2007). Integration of biological networks and gene expression data using Cytoscape. Nature protocols *2*, 2366-2382.

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.

Conkright, M.D., Canettieri, G., Screaton, R., Guzman, E., Miraglia, L., Hogenesch, J.B., and Montminy, M. (2003). TORCs: Transducers of Regulated CREB Activity. Molecular Cell *12*, 413-423.

Cousins, S.L., Innocent, N., and Stephenson, F.A. (2013). Neto1 associates with the NMDA receptor/amyloid precursor protein complex. J Neurochem *126*, 554-564.

Dastidar, S.G., Bardai, F.H., Ma, C., Price, V., Rawat, V., Verma, P., Narayanan, V., and D'Mello, S.R. (2012). Isoform-specific toxicity of Mecp2 in postmitotic neurons: suppression of neurotoxicity by FoxG1. The Journal of neuroscience : the official journal of the Society for Neuroscience *32*, 2846-2855.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S*., et al.* (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209-215.

Deng, B., Melnik, S., and Cook, P.R. (2013). Transcription factories, chromatin loops, and the dysregulation of gene expression in malignancy. Semin Cancer Biol *23*, 65-71.

Deplancke, B., Alpern, D., and Gardeux, V. (2016). The Genetics of Transcription Factor DNA Binding Variation. Cell *166*, 538-554.

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.

Donaghey, J., Thakurela, S., Charlton, J., Chen, J.S., Smith, Z.D., Gu, H., Pop, R., Clement, K., Stamenova, E.K., Karnik, R.*, et al.* (2018). Genetic determinants and epigenetic effects of pioneer-factor occupancy. Nature genetics *50*, 250-258.

Doyle, S.A. (2005). High-throughput cloning for proteomics research. Methods in molecular biology (Clifton, NJ) *310*, 107-113.

Dragich, J.M., Kim, Y.H., Arnold, A.P., and Schanen, N.C. (2007). Differential distribution of the MeCP2 splice variants in the postnatal mouse brain. J Comp Neurol *501*, 526-542.

Ebert, D.H., and Greenberg, M.E. (2013). Activity-dependent neuronal signalling and autism spectrum disorder. Nature *493*, 327-337.

Ecker, D.J., Stein, P., Xu, Z., Williams, C.J., Kopf, G.S., Bilker, W.B., Abel, T., and Schultz, R.M. (2004). Long-term effects of culture of preimplantation mouse embryos on behavior. Proc Natl Acad Sci U S A *101*, 1595-1600.

Edwards, J.R., Yarychkivska, O., Boulard, M., and Bestor, T.H. (2017). DNA methylation and DNA methyltransferases. Epigenetics Chromatin *10*, 23.

Eimer, S., Lakowski, B., Donhauser, R., and Baumeister, R. (2002). Loss of spr-5 bypasses the requirement for the C.elegans presenilin sel-12 by derepressing hop-1. The EMBO journal *21*, 5787-5796.

Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047-3048.

Farrelly, L.A., Thompson, R.E., Zhao, S., Lepack, A.E., Lyu, Y., Bhanu, N.V., Zhang, B., Loh, Y.E., Ramakrishnan, A., Vadodaria, K.C*., et al.* (2019). Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. Nature *567*, 535-539.

Fiszbein, A., Giono, L.E., Quaglino, A., Berardino, B.G., Sigaut, L., von Bilderling, C., Schor, I.E., Steinberg, J.H., Rossi, M., Pietrasanta, L.I*., et al.* (2016). Alternative Splicing of G9a Regulates Neuronal Differentiation. Cell Rep *14*, 2797-2808.

Flavell, S.W., Cowan, C.W., Kim, T.K., Greer, P.L., Lin, Y., Paradis, S., Griffith, E.C., Hu, L.S., Chen, C., and Greenberg, M.E. (2006). Activity-dependent regulation of MEF2 transcription factors suppresses excitatory synapse number. Science (New York, NY) *311*, 1008-1012.

Flavell, S.W., Kim, T.K., Gray, J.M., Harmin, D.A., Hemberg, M., Hong, E.J., Markenscoff-Papadimitriou, E., Bear, D.M., and Greenberg, M.E. (2008). Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. Neuron *60*, 1022-1038.

Glynn, M.W., and McAllister, A.K. (2006). Immunocytochemistry and quantification of protein colocalization in cultured neurons. Nature protocols *1*, 1287-1296.

Goodrich, J.A., and Tjian, R. (2010). Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nature reviews Genetics *11*, 549-558.

Gupta-Agarwal, S., Franklin, A.V., Deramus, T., Wheelock, M., Davis, R.L., McMahon, L.L., and Lubin, F.D. (2012). G9a/GLP histone lysine dimethyltransferase complex activity in the hippocampus and the entorhinal cortex is required for gene activation and silencing

during memory consolidation. The Journal of neuroscience : the official journal of the Society for Neuroscience *32*, 5440-5453.

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 *27*, 322-326.

Haberhausen, G., Schmitt, I., Kohler, A., Peters, U., Rider, S., Chelly, J., Terwilliger, J.D., Monaco, A.P., and Muller, U. (1995). Assignment of the dystonia-parkinsonism syndrome locus, DYT3, to a small region within a 1.8-Mb YAC contig of Xq13.1. Am J Hum Genet *57*, 644-650.

Hakimi, M.A., Bochar, D.A., Chenoweth, J., Lane, W.S., Mandel, G., and Shiekhattar, R. (2002). A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. Proc Natl Acad Sci U S A *99*, 7420-7425.

Hamanaka, K., Sugawara, Y., Shimoji, T., Nordtveit, T.I., Kato, M., Nakashima, M., Saitsu, H., Suzuki, T., Yamakawa, K., Aukrust, I.*, et al.* (2018). De novo truncating variants in PHF21A cause intellectual disability and craniofacial anomalies. Eur J Hum Genet.

He, Y., and Ecker, J.R. (2015). Non-CG Methylation in the Human Genome. Annu Rev Genomics Hum Genet *16*, 55-77.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature genetics *39*, 311-318.

Herzfeld, T., Nolte, D., Grznarova, M., Hofmann, A., Schultze, J.L., and Muller, U. (2013). X-linked dystonia parkinsonism syndrome (XDP, lubag): disease-specific sequence change

DSC3 in TAF1/DYT3 affects genes in vesicular transport and dopamine metabolism. Hum Mol Genet *22*, 941-951.

Huang, C.S., Shi, S.H., Ule, J., Ruggiu, M., Barker, L.A., Darnell, R.B., Jan, Y.N., and Jan, L.Y. (2005). Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition. Cell *123*, 105-118.

Huynh, M.A., Ikeuchi, Y., Netherton, S., de la Torre-Ubieta, L., Kanadia, R., Stegmuller, J., Cepko, C., Bonni, S., and Bonni, A. (2011). An isoform-specific SnoN1-FOXO1 repressor complex controls neuronal morphogenesis and positioning in the mammalian brain. Neuron *69*, 930-944.

Iijima, T., Hidaka, C., and Iijima, Y. (2016). Spatio-temporal regulations and functions of neuronal alternative RNA splicing in developing and adult brains. Neurosci Res *109*, 1-8.

Ikeuchi, Y., Stegmuller, J., Netherton, S., Huynh, M.A., Masu, M., Frank, D., Bonni, S., and Bonni, A. (2009). A SnoN-Ccd1 pathway promotes axonal morphogenesis in the mammalian brain. The Journal of neuroscience : the official journal of the Society for Neuroscience *29*, 4312-4321.

Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E.*, et al.* (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216-221.

Irimia, M., Weatheritt, R.J., Ellis, J.D., Parikshak, N.N., Gonatopoulos-Pournatzis, T., Babor, M., Quesnel-Vallieres, M., Tapial, J., Raj, B., O'Hanlon, D.*, et al.* (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell *159*, 1511-1523.

Ito, N., Hendriks, W.T., Dhakal, J., Vaine, C.A., Liu, C., Shin, D., Shin, K., Wakabayashi-Ito, N., Dy, M., Multhaupt-Buell, T.*, et al.* (2016). Decreased N-TAF1 expression in X-linked dystonia-parkinsonism patient-specific neural stem cells. Dis Model Mech *9*, 451-462.

Itoh, M., Tahimic, C.G., Ide, S., Otsuki, A., Sasaoka, T., Noguchi, S., Oshimura, M., Goto, Y., and Kurimasa, A. (2012). Methyl CpG-binding protein isoform MeCP2_e2 is dispensable for Rett syndrome phenotypes but essential for embryo viability and placenta development. The Journal of biological chemistry *287*, 13859-13867.

Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. Genes & development *28*, 2679-2692.

Iwase, S., Brookes, E., Agarwal, S., Badeaux, A.I., Ito, H., Vallianatos, C.N., Tomassy, G.S., Kasza, T., Lin, G., Thompson, A.*, et al.* (2016). A Mouse Model of X-linked Intellectual Disability Associated with Impaired Removal of Histone Methylation. Cell Rep.

Iwase, S., Januma, A., Miyamoto, K., Shono, N., Honda, A., Yanagisawa, J., and Baba, T. (2004). Characterization of BHC80 in BRAF-HDAC complex, involved in neuron-specific gene repression. Biochemical and biophysical research communications *322*, 601-608.

Iwase, S., Shono, N., Honda, A., Nakanishi, T., Kashiwabara, S., Takahashi, S., and Baba, T. (2006). A component of BRAF-HDAC complex, BHC80, is required for neonatal survival in mice. FEBS letters *580*, 3129-3135.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science (New York, NY) *288*, 1422-1425.

Jain, R.G., Andrews, L.G., McGowan, K.M., Pekala, P.H., and Keene, J.D. (1997). Ectopic expression of Hel-N1, an RNA-binding protein, increases glucose transporter (GLUT1) expression in 3T3-L1 adipocytes. Molecular and cellular biology *17*, 954-962.

Jambaldorj, J., Makino, S., Munkhbat, B., and Tamiya, G. (2012). Sustained expression of a neuron-specific isoform of the Taf1 gene in development stages and aging in mice. Biochemical and biophysical research communications *425*, 273-277.

Jang, H.S., Shin, W.J., Lee, J.E., and Do, J.T. (2017). CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. Genes (Basel) *8*.

Janson, C.G., Chen, Y., Li, Y., and Leifer, D. (2001). Functional regulatory regions of human transcription factor MEF2C. Brain research Molecular brain research *97*, 70-82.

Jarriualt, S., and Greenwald, I. (2002). Suppressors of the egg-laying defective phenotype of sel-12 presenilin mutants implicate the CoREST corepressor complex in LIN-12/Notch signaling in C. elegans. Genes & development *16*, 2713-2728.

Jensen, K.B., Dredge, B.K., Stefani, G., Zhong, R., Buckanovich, R.J., Okano, H.J., Yang, Y.Y., and Darnell, R.B. (2000). Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. Neuron *25*, 359-371.

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

Jih, G., Iglesias, N., Currie, M.A., Bhanu, N.V., Paulo, J.A., Gygi, S.P., Garcia, B.A., and Moazed, D. (2017). Unique roles for histone H3K9me states in RNAi and heritable silencing of transcription. Nature *547*, 463-467.

Kandel, E.R. (2001). The molecular biology of memory storage: a dialogue between genes and synapses. Science (New York, NY) *294*, 1030-1038.

Kim, H.G., Kim, H.T., Leach, N.T., Lan, F., Ullmann, R., Silahtaroglu, A., Kurth, I., Nowka, A., Seong, I.S., Shen, Y*., et al.* (2012). Translocations disrupting PHF21A in the Potocki-Shaffer-syndrome region are associated with intellectual disability and craniofacial anomalies. Am J Hum Genet *91*, 56-72.

Klajn, A., Ferrai, C., Stucchi, L., Prada, I., Podini, P., Baba, T., Rocchi, M., Meldolesi, J., and D'Alessandro, R. (2009). The rest repression of the neurosecretory phenotype is negatively modulated by BHC80, a protein of the BRAF/HDAC complex. The Journal of neuroscience : the official journal of the Society for Neuroscience *29*, 6296-6307.

Kleefstra, T., Brunner, H.G., Amiel, J., Oudakker, A.R., Nillesen, W.M., Magee, A., Genevieve, D., Cormier-Daire, V., van Esch, H., Fryns, J.P*., et al.* (2006). Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. Am J Hum Genet *79*, 370-377.

Koster, J., and Rahmann, S. (2012). Snakemake--a scalable bioinformatics workflow engine. Bioinformatics *28*, 2520-2522.

Kovacs, K.J. (1998). c-Fos as a transcription factor: a stressful (re)view from a functional map. Neurochemistry international *33*, 287-297.

Kriaucionis, S., and Bird, A. (2004). The major form of MeCP2 has a novel N-terminus generated by alternative splicing. Nucleic Acids Res *32*, 1818-1823.

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

Kuehner, J.N., Bruggeman, E.C., Wen, Z., and Yao, B. (2019). Epigenetic Regulations in Neuropsychiatric Disorders. Front Genet *10*, 268.

Labonne, J.D., Vogt, J., Reali, L., Kong, I.K., Layman, L.C., and Kim, H.G. (2015). A microdeletion encompassing PHF21A in an individual with global developmental delay and craniofacial anomalies. Am J Med Genet A *167*, 3011-3018.

Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., and Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics *26*, 2438-2444.

Lakowski, B., Roelens, I., and Jacob, S. (2006). CoREST-like complexes regulate chromatin modification and neuronal gene expression. Journal of molecular neuroscience : MN *29*, 227-239.

Lan, F., Collins, R.E., De Cegli, R., Alpatov, R., Horton, J.R., Shi, X., Gozani, O., Cheng, X., and Shi, Y. (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. Nature *448*, 718-722.

Lara-Pezzi, E., Desco, M., Gatto, A., and Gomez-Gaviro, M.V. (2016). Neurogenesis: Regulation by Alternative Splicing and Related Posttranscriptional Processes. Neuroscientist.

Laurent, B., Ruitu, L., Murn, J., Hempel, K., Ferrao, R., Xiang, Y., Liu, S., Garcia, B.A., Wu, H., Wu, F.*, et al.* (2015). A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. Mol Cell *57*, 957-970.

Lavery, L.A., and Zoghbi, H.Y. (2019). The distinct methylation landscape of maturing neurons and its role in Rett syndrome pathogenesis. Curr Opin Neurobiol *59*, 180-188.

Le Meur, N., Holder-Espinasse, M., Jaillard, S., Goldenberg, A., Joriot, S., Amati-Bonneau, P., Guichet, A., Barth, M., Charollais, A., Journel, H.*, et al.* (2010). MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is

responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. J Med Genet *47*, 22-29.

Lee, Y.T., Gibbons, G., Lee, S.Y., Nikolovska-Coleska, Z., and Dou, Y. (2015). One-pot refolding of core histones from bacterial inclusion bodies allows rapid reconstitution of histone octamer. Protein Expr Purif *110*, 89-94.

Leifer, D., Krainc, D., Yu, Y.T., McDermott, J., Breitbart, R.E., Heng, J., Neve, R.L., Kosofsky, B., Nadal-Ginard, B., and Lipton, S.A. (1993). MEF2C, a MADS/MEF2-family transcription factor expressed in a laminar distribution in cerebral cortex. Proc Natl Acad Sci U S A *90*, 1546-1550.

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

Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Curr Opin Genet Dev *21*, 175-186.

Li, H., Radford, J.C., Ragusa, M.J., Shea, K.L., McKercher, S.R., Zaremba, J.D., Soussou, W., Nie, Z., Kang, Y.J., Nakanishi, N.*, et al.* (2008). Transcription factor MEF2C influences neural stem/progenitor cell differentiation and maturation in vivo. Proc Natl Acad Sci U S A *105*, 9397-9402.

Li, Q., Lee, J.A., and Black, D.L. (2007). Neuronal regulation of alternative pre-mRNA splicing. Nature reviews Neuroscience *8*, 819-831.

Li, Y., Donmez, N., Sahinalp, C., Xie, N., Wang, Y., Xue, H., Mo, F., Beltran, H., Gleave, M., Wang, Y.*, et al.* (2017). SRRM4 Drives Neuroendocrine Transdifferentiation of Prostate Adenocarcinoma Under Androgen Receptor Pathway Inhibition. Eur Urol *71*, 68-78.

Li, Y.I., Sanchez-Pulido, L., Haerty, W., and Ponting, C.P. (2015). RBFOX and PTBP1 proteins
regulate the alternative splicing of micro-exons in human brain transcripts. Genome
research *25*, 1-13.

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

Licatalosi, D.D., and Darnell, R.B. (2006). Splicing regulation in neurologic disease. Neuron *52*,
93-101.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer,
A.C., Blume, J.E., Wang, X*., et al.* (2008). HITS-CLIP yields genome-wide insights into
brain alternative RNA processing. Nature *456*, 464-469.

Liu, Y., Zhou, D., Qi, D., Feng, J., Liu, Z., Hu, Y., Shen, W., Liu, C., and Kong, X. (2018).
Lysine-specific demethylase 1 cooperates with BRAF-histone deacetylase complex 80 to
enhance HIV-1 Tat-mediated transactivation. Virus Genes *54*, 662-671.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
dispersion for RNA-seq data with DESeq2. Genome biology *15*, 550.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010).
Regulation of alternative splicing by histone modifications. Science (New York, NY)
327, 996-1000.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal
structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luque, F.A., Furneaux, H.M., Ferziger, R., Rosenblum, M.K., Wray, S.H., Schold, S.C., Jr.,
Glantz, M.J., Jaeckle, K.A., Biran, H., Lesser, M*., et al.* (1991). Anti-Ri: an antibody

associated with paraneoplastic opsoclonus and breast cancer. Annals of neurology *29*, 241-251.

Lyons, M.R., Schwarz, C.M., and West, A.E. (2012). Members of the myocyte enhancer factor 2 transcription factor family differentially regulate Bdnf transcription in response to neuronal depolarization. The Journal of neuroscience : the official journal of the Society for Neuroscience *32*, 12780-12785.

Makeyev, E.V., Zhang, J., Carrasco, M.A., and Maniatis, T. (2007). The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. Mol Cell *27*, 435-448.

Makino, S., Kaji, R., Ando, S., Tomizawa, M., Yasuno, K., Goto, S., Matsumoto, S., Tabuena, M.D., Maranon, E., Dantes, M.*, et al.* (2007). Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. Am J Hum Genet *80*, 393-406.

Martinez, N.M., Pan, Q., Cole, B.S., Yarosh, C.A., Babcock, G.A., Heyd, F., Zhu, W., Ajith, S., Blencowe, B.J., and Lynch, K.W. (2012). Alternative splicing networks regulated by signaling in human T cells. RNA *18*, 1029-1040.

Mata, G., Cuesto, G.n., Heras, J., Morales, M., Romero, A., and Rubio, J. (2017). SynapCountJ: A Validated Tool for Analyzing Synaptic Densities in Neurons (Cham: Springer International Publishing).

Mavrogiannis, L.A., Antonopoulou, I., Baxova, A., Kutilek, S., Kim, C.A., Sugayama, S.M., Salamanca, A., Wall, S.A., Morriss-Kay, G.M., and Wilkie, A.O. (2001). Haploinsufficiency of the human homeobox gene ALX4 causes skull ossification defects. Nature genetics *27*, 17-18.

Mayr, B., and Montminy, M. (2001). Transcriptional regulation by the phosphorylation-dependent factor CREB. Nature reviews Molecular cell biology *2*, 599-609.

Mayran, A., and Drouin, J. (2018). Pioneer transcription factors shape the epigenetic landscape. The Journal of biological chemistry *293*, 13795-13804.

McCarthy, S.E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., Berstein, Y., Mistry, M., Pavlidis, P., Solomon, R., Ghiban, E*., et al.* (2014). De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. Mol Psychiatry *19*, 652-658.

McCool, C., Spinks-Franklin, A., Noroski, L.M., and Potocki, L. (2017). Potocki-Shaffer syndrome in a child without intellectual disability-the role of PHF21A in cognitive function. Am J Med Genet A.

McKee, A.E., Minet, E., Stern, C., Riahi, S., Stiles, C.D., and Silver, P.A. (2005). A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. BMC Dev Biol *5*, 14.

Medvedeva, Y.A., Lennartsson, A., Ehsani, R., Kulakovskiy, I.V., Vorontsov, I.E., Panahandeh, P., Khimulya, G., Kasukawa, T., Consortium, F., and Drablos, F. (2015). EpiFactors: a comprehensive database of human epigenetic factors and complexes. Database (Oxford) *2015*, bav067.

Meijering, E., Jacob, M., Sarria, J.C., Steiner, P., Hirling, H., and Unser, M. (2004). Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. Cytometry Part A : the journal of the International Society for Analytical Cytology *58*, 167-176.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and

    isoform regulation in Mammalian tissues. Science (New York, NY) *338*, 1593-1599.

Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T.,

    Buettner, R., and Schule, R. (2005). LSD1 demethylates repressive histone marks to

    promote androgen-receptor-dependent transcription. Nature *437*, 436-439.

Mizzen, C.A., Yang, X.-J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T.,

    Workman, J., Wang, L., Berger, S.L., Kouzarides, T*., et al.* (1996). The TAFII250

    Subunit of TFIID Has Histone Acetyltransferase Activity. Cell *87*, 1261-1270.

Mnatzakanian, G.N., Lohi, H., Munteanu, I., Alfred, S.E., Yamada, T., MacLeod, P.J., Jones,

    J.R., Scherer, S.W., Schanen, N.C., Friez, M.J*., et al.* (2004). A previously unidentified

    MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome.

    Nature genetics *36*, 339-341.

Morgan, M.A., and Shilatifard, A. (2015). Chromatin signatures of cancer. Genes &

    development *29*, 238-249.

Morrow, E.M., Yoo, S.Y., Flavell, S.W., Kim, T.K., Lin, Y., Hill, R.S., Mukaddes, N.M.,

    Balkhy, S., Gascon, G., Hashmi, A*., et al.* (2008). Identifying autism loci and genes by

    tracing recent shared ancestry. Science (New York, NY) *321*, 218-223.

Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., Chen, W., Hosseini, M.,

    Behjati, F., Haas, S., Jamali, P*., et al.* (2011). Deep sequencing reveals 50 novel genes for

    recessive cognitive disorders. Nature *478*, 57-63.

Nelson, S.B., and Valakh, V. (2015). Excitatory/Inhibitory Balance and Circuit Homeostasis in

    Autism Spectrum Disorders. Neuron *87*, 684-698.

Ng, D., Pitcher, G.M., Szilard, R.K., Sertie, A., Kanisek, M., Clapcote, S.J., Lipina, T., Kalia, L.V., Joo, D., McKerlie, C.*, et al.* (2009). Neto1 is a novel CUB-domain NMDA receptor-interacting protein required for synaptic plasticity and learning. PLoS Biol *7*, e41.

Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C.*, et al.* (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature genetics *42*, 790-793.

Nishimoto, H.K., Ha, K., Jones, J.R., Dwivedi, A., Cho, H.M., Layman, L.C., and Kim, H.G. (2014). The historical Coffin-Lowry syndrome family revisited: identification of two novel mutations of RPS6KA3 in three male patients. Am J Med Genet A *164A*, 2172-2179.

O'Rawe, J.A., Wu, Y., Dorfel, M.J., Rope, A.F., Au, P.Y., Parboosingh, J.S., Moon, S., Kousi, M., Kosma, K., Smith, C.S.*, et al.* (2015). TAF1 Variants Are Associated with Dysmorphic Features, Intellectual Disability, and Neurological Manifestations. Am J Hum Genet *97*, 922-932.

Oberstrass, F.C., Auweter, S.D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L.*, et al.* (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. Science (New York, NY) *309*, 2054-2057.

Okano, H.J., and Darnell, R.B. (1997). A hierarchy of Hu RNA binding proteins in developing and adult neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience *17*, 3024-3037.

Olson, C.O., Zachariah, R.M., Ezeonwuka, C.D., Liyanage, V.R., and Rastegar, M. (2014). Brain region-specific expression of MeCP2 isoforms correlates with DNA methylation within Mecp2 regulatory elements. PloS one *9*, e90645.

Palm, K., Metsis, M., and Timmusk, T. (1999). Neuron-specific splicing of zinc finger transcription factor REST/NRSF/XBR is frequent in neuroblastomas and conserved in human, mouse and rat. Brain research Molecular brain research *72*, 30-39.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature genetics *40*, 1413-1415.

Pelzer, T., Lyons, G.E., Kim, S., and Moreadith, R.W. (1996). Cloning and characterization of the murine homolog of the sno proto-oncogene reveals a novel splice variant. Developmental dynamics : an official publication of the American Association of Anatomists *205*, 114-125.

Pilotto, S., Speranzini, V., Marabelli, C., Rusconi, F., Toffolo, E., Grillo, B., Battaglioli, E., and Mattevi, A. (2016). LSD1/KDM1A mutations associated to a newly described form of intellectual disability impair demethylase activity and binding to transcription factors. Hum Mol Genet *25*, 2578-2587.

Pohodich, A.E., and Zoghbi, H.Y. (2015). Rett syndrome: disruption of epigenetic control of postnatal neurological functions. Hum Mol Genet *24*, R10-16.

Porter, R.S., Jaamour, F., and Iwase, S. (2017a). Neuron-specific alternative splicing of transcriptional machineries: Implications for neurodevelopmental disorders. Molecular and Cellular Neuroscience.

Porter, R.S., Murata-Nakamura, Y., Nagasu, H., Kim, H.G., and Iwase, S. (2017b). Transcriptome Analysis Revealed Impaired cAMP Responsiveness in PHF21A-Deficient Human Cells. Neuroscience.

Potocki, L., and Shaffer, L.G. (1996). Interstitial deletion of 11(p11.2p12): a newly described contiguous gene deletion syndrome involving the gene for hereditary multiple exostoses (EXT2). American journal of medical genetics *62*, 319-325.

Quesnel-Vallières, M., Dargaei, Z., Irimia, M., Gonatopoulos-Pournatzis, T., Ip, J.Y., Wu, M., Sterne-Weiler, T., Nakagawa, S., Woodin, M.A., Blencowe, B.J.*, et al.* (2016). Misregulation of an Activity-Dependent Splicing Network as a Common Mechanism Underlying Autism Spectrum Disorders. Molecular Cell *64*, 1023-1034.

Quesnel-Vallieres, M., Irimia, M., Cordes, S.P., and Blencowe, B.J. (2015). Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. Genes & development *29*, 746-759.

Raj, B., and Blencowe, B.J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. Neuron *87*, 14-27.

Raj, B., Irimia, M., Braunschweig, U., Sterne-Weiler, T., O'Hanlon, D., Lin, Z.Y., Chen, G.I., Easton, L.E., Ule, J., Gingras, A.C.*, et al.* (2014). A global regulatory mechanism for activating an exon network required for neurogenesis. Mol Cell *56*, 90-103.

Raj, B., O'Hanlon, D., Vessey, J.P., Pan, Q., Ray, D., Buckley, N.J., Miller, F.D., and Blencowe, B.J. (2011). Cross-regulation between an alternative splicing activator and a transcription repressor controls neurogenesis. Mol Cell *43*, 843-850.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.

Ramocki, M.B., Tavyev, Y.J., and Peters, S.U. (2010). The MECP2 duplication syndrome. Am J Med Genet A *152A*, 1079-1088.

Richa, R., and Sinha, R.P. (2014). Hydroxymethylation of DNA: an epigenetic marker. EXCLI journal *13*, 592-610.

Rizzardi, L.F., Hickey, P.F., Rodriguez DiBlasi, V., Tryggvadottir, R., Callahan, C.M., Idrizi, A., Hansen, K.D., and Feinberg, A.P. (2019). Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability. Nat Neurosci *22*, 307-316.

Ronan, J.L., Wu, W., and Crabtree, G.R. (2013). From neural development to cognition: unexpected roles for chromatin. Nature reviews Genetics *14*, 347-359.

Ropers, H.H. (2010). Genetics of early onset cognitive impairment. Annu Rev Genomics Hum Genet *11*, 161-187.

Rubenstein, J.L., and Merzenich, M.M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. Genes, brain, and behavior *2*, 255-267.

Ruggiu, M., Herbst, R., Kim, N., Jevsek, M., Fak, J.J., Mann, M.A., Fischbach, G., Burden, S.J., and Darnell, R.B. (2009). Rescuing Z+ agrin splicing in Nova null mice restores synapse formation and unmasks a physiologic defect in motor neuron firing. Proc Natl Acad Sci U S A *106*, 3513-3518.

Ruppert, S., Wang, E.H., and Tjian, R. (1993). Cloning and expression of human TAFII250: a TBP-associated factor implicated in cell-cycle regulation. Nature *362*, 175-179.

Rusconi, F., Grillo, B., Ponzoni, L., Bassani, S., Toffolo, E., Paganini, L., Mallei, A., Braida, D.,
Passafaro, M., Popoli, M*., et al.* (2016a). LSD1 modulates stress-evoked transcription of
immediate early genes and emotional behavior. Proc Natl Acad Sci U S A *113*, 3651-
3656.

Rusconi, F., Grillo, B., Toffolo, E., Mattevi, A., and Battaglioli, E. (2016b). NeuroLSD1:
Splicing-Generated Epigenetic Enhancer of Neuroplasticity. Trends Neurosci.

Rusconi, F., Paganini, L., Braida, D., Ponzoni, L., Toffolo, E., Maroli, A., Landsberger, N.,
Bedogni, F., Turco, E., Pattini, L*., et al.* (2015). LSD1 Neurospecific Alternative Splicing
Controls Neuronal Excitability in Mouse Models of Epilepsy. Cereb Cortex *25*, 2729-
2740.

Sampath, S.C., Marazzi, I., Yap, K.L., Sampath, S.C., Krutchinsky, A.N., Mecklenbrauker, I.,
Viale, A., Rudensky, E., Zhou, M.M., Chait, B.T*., et al.* (2007). Methylation of a histone
mimic within the histone methyltransferase G9a regulates protein complex assembly.
Mol Cell *27*, 596-608.

Sartor, M.A., Leikauf, G.D., and Medvedovic, M. (2009). LRpath: a logistic regression approach
for identifying enriched biological groups in gene expression data. Bioinformatics *25*,
211-217.

Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., Peng, M.,
Collins, R., Grove, J., Klei, L*., et al.* (2020). Large-Scale Exome Sequencing Study
Implicates Both Developmental and Functional Changes in the Neurobiology of Autism.
Cell *180*, 568-584 e523.

Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A., and Greengard, P. (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic suppressor complex. Neuron *64*, 678-691.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature methods *9*, 671-675.

Schoenherr, C.J., and Anderson, D.J. (1995). The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. Science (New York, NY) *267*, 1360-1363.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A., Tonjes, M.*, et al.* (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226-231.

Seki, M., Masaki, H., Arauchi, T., Nakauchi, H., Sugano, S., and Suzuki, Y. (2014). A comparison of the rest complex binding patterns in embryonic stem cells and epiblast stem cells. PloS one *9*, e95374.

Sheng, M., and Greenberg, M.E. (1990). The regulation and function of c-fos and other immediate early genes in the nervous system. Neuron *4*, 477-485.

Sheng, W., LaFleur, M.W., Nguyen, T.H., Chen, S., Chakravarthy, A., Conway, J.R., Li, Y., Chen, H., Yang, H., Hsu, P.H.*, et al.* (2018). LSD1 Ablation Stimulates Anti-tumor Immunity and Enables Checkpoint Blockade. Cell *174*, 549-563 e519.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Shi, Y.J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005). Regulation of LSD1 histone demethylase activity by its associated factors. Mol Cell *19*, 857-864.

Shimojo, M., Paquette, A.J., Anderson, D.J., and Hersh, L.B. (1999). Protein kinase A regulates cholinergic gene expression in PC12 cells: REST4 silences the silencing activity of neuron-restrictive silencer factor/REST. Molecular and cellular biology *19*, 6788-6795.

Silverman, J.L., Yang, M., Lord, C., and Crawley, J.N. (2010). Behavioural phenotyping assays for mouse models of autism. Nature reviews Neuroscience *11*, 490-502.

Song, Y., Dagil, L., Fairall, L., Robertson, N., Wu, M., Ragan, T.J., Savva, C.G., Saleh, A., Morone, N., Kunze, M.B.A*., et al.* (2020). Mechanism of Crosstalk between the LSD1 Demethylase and HDAC1 Deacetylase in the CoREST Complex. Cell Reports *30*, 2699-2711.e2698.

Staahl, B.T., and Crabtree, G.R. (2013). Creating a neural specific chromatin landscape by npBAF and nBAF complexes. Curr Opin Neurobiol *23*, 903-913.

Stegmuller, J., Konishi, Y., Huynh, M.A., Yuan, Z., Dibacco, S., and Bonni, A. (2006). Cell-intrinsic regulation of axonal morphogenesis by the Cdh1-APC target SnoN. Neuron *50*, 389-400.

Stickens, D., Clines, G., Burbee, D., Ramos, P., Thomas, S., Hogue, D., Hecht, J.T., Lovett, M., and Evans, G.A. (1996). The EXT2 multiple exostoses gene defines a family of putative tumour suppressor genes. Nature genetics *14*, 25-32.

Stroschein, S.L., Wang, W., Zhou, S., Zhou, Q., and Luo, K. (1999). Negative feedback regulation of TGF-beta signaling by the SnoN oncoprotein. Science (New York, NY) *286*, 771-774.

Stroud, H., Su, S.C., Hrvatin, S., Greben, A.W., Renthal, W., Boxer, L.D., Nagy, M.A., Hochbaum, D.R., Kinde, B., Gabel, H.W.*, et al.* (2017). Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic States. Cell *171*, 1151-1164 e1116.

Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PloS one *6*, e21800.

Szabo, A., Dalmau, J., Manley, G., Rosenfeld, M., Wong, E., Henson, J., Posner, J.B., and Furneaux, H.M. (1991). HuD, a paraneoplastic encephalomyelitis antigen, contains RNA-binding domains and is homologous to Elav and Sex-lethal. Cell *67*, 325-333.

Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H., and Shinkai, Y. (2008). G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. The EMBO journal *27*, 2681-2690.

Tang, Z.Z., Zheng, S., Nikolic, J., and Black, D.L. (2009). Developmental control of CaV1.2 L-type calcium channel splicing by Fox proteins. Molecular and cellular biology *29*, 4757-4765.

Tinti, C., Yang, C., Seo, H., Conti, B., Kim, C., Joh, T.H., and Kim, K.S. (1997). Structure/function relationship of the cAMP response element in tyrosine hydroxylase gene transcription. The Journal of biological chemistry *272*, 19158-19164.

Todaro, G.J., and Green, H. (1963). Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. J Cell Biol *17*, 299-313.

Toffolo, E., Rusconi, F., Paganini, L., Tortorici, M., Pilotto, S., Heise, C., Verpelli, C., Tedeschi, G., Maffioli, E., Sala, C.*, et al.* (2014). Phosphorylation of neuronal Lysine-Specific Demethylase 1LSD1/KDM1A impairs transcriptional repression by regulating interaction with CoREST and histone deacetylases HDAC1/2. J Neurochem *128*, 603-616.

Tunovic, S., Barkovich, J., Sherr, E.H., and Slavotinek, A.M. (2014). De novo ANKRD11 and KDM1A gene mutations in a male with features of KBG syndrome and Kabuki syndrome. Am J Med Genet A *164A*, 1744-1749.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science (New York, NY) *302*, 1212-1215.

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J., and Darnell, R.B. (2006). An RNA map predicting Nova-dependent splicing regulation. Nature *444*, 580-586.

Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M*., et al.* (2005). Nova regulates brain-specific splicing to shape the synapse. Nature genetics *37*, 844-852.

Vallianatos, C.N., and Iwase, S. (2015). Disrupted intricacy of histone H3K4 methylation in neurodevelopmental disorders. Epigenomics *7*, 503-519.

Vogel-Ciernia, A., and Wood, M.A. (2014). Examining object location and object recognition memory in mice. Current protocols in neuroscience *69*, 8.31.31-17.

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

Vuong, C.K., Black, D.L., and Zheng, S. (2016). The neurogenetics of alternative splicing. Nature reviews Neuroscience *17*, 265-281.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wakui, K., Gregato, G., Ballif, B.C., Glotzbach, C.D., Bailey, K.A., Kuo, P.L., Sue, W.C., Sheffield, L.J., Irons, M., Gomez, E.G.*, et al.* (2005). Construction of a natural panel of 11p11.2 deletions and further delineation of the critical region involved in Potocki-Shaffer syndrome. Eur J Hum Genet *13*, 528-540.

Wan, Y., Liu, X., and Kirschner, M.W. (2001). The anaphase-promoting complex mediates TGF-beta signaling by targeting SnoN for destruction. Mol Cell *8*, 1027-1039.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476.

Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G.G., Krones, A., Ohgi, K.A., Zhu, P., Garcia-Bassets, I.*, et al.* (2007). Opposing LSD1 complexes function in developmental gene activation and repression programmes. Nature *446*, 882-887.

Wang, J., Telese, F., Tan, Y., Li, W., Jin, C., He, X., Basnet, H., Ma, Q., Merkurjev, D., Zhu, X.*, et al.* (2015). LSD1n is an H4K20 demethylase regulating memory formation via transcriptional elongation control. Nat Neurosci *18*, 1256-1264.

Wang, X., and Tanaka Hall, T.M. (2001). Structural basis for recognition of AU-rich element RNA by the HuD protein. Nature structural biology *8*, 141-145.

Wang, Y., Zhang, H., Chen, Y., Sun, Y., Yang, F., Yu, W., Liang, J., Sun, L., Yang, X., Shi, L.*, et al.* (2009). LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. Cell *138*, 660-672.

West, A.E., Griffith, E.C., and Greenberg, M.E. (2002). Regulation of transcription factors by neuronal activity. Nature reviews Neuroscience *3*, 921-931.

Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q.*, et al.* (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell Rep *6*, 1139-1152.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221-225.

Wu, Y.Q., Badano, J.L., McCaskill, C., Vogel, H., Potocki, L., and Shaffer, L.G. (2000). Haploinsufficiency of ALX4 as a potential cause of parietal foramina in the 11p11.2 contiguous gene-deletion syndrome. Am J Hum Genet *67*, 1327-1332.

Xu, Q., Modrek, B., and Lee, C. (2002). Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. Nucleic Acids Res *30*, 3754-3766.

Xue, Y., Ouyang, K., Huang, J., Zhou, Y., Ouyang, H., Li, H., Wang, G., Wu, Q., Wei, C., Bi, Y.*, et al.* (2013). Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. Cell *152*, 82-96.

Xue, Y., Qian, H., Hu, J., Zhou, B., Zhou, Y., Hu, X., Karakhanyan, A., Pang, Z., and Fu, X.D. (2016). Sequential regulatory loops as key gatekeepers for neuronal reprogramming in human cells. Nat Neurosci *19*, 807-815.

Yang, M., Gocke, C.B., Luo, X., Borek, D., Tomchick, D.R., Machius, M., Otwinowski, Z., and Yu, H. (2006). Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. Mol Cell *23*, 377-387.

Yang, Y.Y., Yin, G.L., and Darnell, R.B. (1998). The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. Proc Natl Acad Sci U S A *95*, 13254-13259.

Yap, K., and Makeyev, E.V. (2013). Regulation of gene expression in mammalian nervous system through alternative pre-mRNA splicing coupled with RNA quality control mechanisms. Mol Cell Neurosci *56*, 420-428.

Yasui, D.H., Gonzales, M.L., Aflatooni, J.O., Crary, F.K., Hu, D.J., Gavino, B.J., Golub, M.S., Vincent, J.B., Carolyn Schanen, N., Olson, C.O.*, et al.* (2014). Mice with an isoform-ablating Mecp2 exon 1 mutation recapitulate the neurologic deficits of Rett syndrome. Hum Mol Genet *23*, 2447-2458.

Yeo, G., Holste, D., Kreiman, G., and Burge, C.B. (2004). Variation in alternative splicing across human tissues. Genome biology *5*, R74.

Yoo, A.S., Sun, A.X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., Lee-Messer, C., Dolmetsch, R.E., Tsien, R.W., and Crabtree, G.R. (2011). MicroRNA-mediated conversion of human fibroblasts to neurons. Nature *476*, 228-231.

Zachariah, R.M., Olson, C.O., Ezeonwuka, C., and Rastegar, M. (2012). Novel MeCP2 isoform-specific antibody reveals the endogenous MeCP2E1 expression in murine brain, primary neurons and astrocytes. PloS one *7*, e49763.

Zaret, K.S. (2018). Pioneering the chromatin landscape. Nature genetics *50*, 167-169.

Zhang, C., Frias, M.A., Mele, A., Ruggiu, M., Eom, T., Marney, C.B., Wang, H., Licatalosi, D.D., Fak, J.J., and Darnell, R.B. (2010). Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. Science (New York, NY) *329*, 439-443.

Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A.R., and Zhang, M.Q. (2008). Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. Genes & development *22*, 2550-2563.

Zhang, X., Chen, M.H., Wu, X., Kodani, A., Fan, J., Doan, R., Ozawa, M., Ma, J., Yoshida, N., Reiter, J.F.*, et al.* (2016). Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. Cell *166*, 1147-1162 e1115.

Zhang, X., Odom, D.T., Koo, S.H., Conkright, M.D., Canettieri, G., Best, J., Chen, H., Jenner, R., Herbolsheimer, E., Jacobsen, E.*, et al.* (2005). Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. Proc Natl Acad Sci U S A *102*, 4459-4464.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N.*, et al.* (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience *34*, 11929-11947.

Zhou, B., Wang, J., Lee, S.Y., Xiong, J., Bhanu, N., Guo, Q., Ma, P., Sun, Y., Rao, R.C., Garcia, B.A.*, et al.* (2016). PRDM16 Suppresses MLL1r Leukemia via Intrinsic Histone Methyltransferase Activity. Mol Cell.

Zhou, H.L., Hinman, M.N., Barron, V.A., Geng, C., Zhou, G., Luo, G., Siegel, R.E., and Lou, H. (2011). Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. Proc Natl Acad Sci U S A *108*, E627-635.

Zhu, B., Ramachandran, B., and Gulick, T. (2005). Alternative pre-mRNA splicing governs expression of a conserved acidic transactivation domain in myocyte enhancer factor 2 factors of striated muscle and brain. The Journal of biological chemistry *280*, 28749-28760.

Zhu, H., Hasman, R.A., Barron, V.A., Luo, G., and Lou, H. (2006). A nuclear function of Hu proteins as neuron-specific alternative RNA processing regulators. Mol Biol Cell *17*, 5105-5114.

Zhu, H., Hinman, M.N., Hasman, R.A., Mehta, P., and Lou, H. (2008). Regulation of neuron-specific alternative splicing of neurofibromatosis type 1 pre-mRNA. Molecular and cellular biology *28*, 1240-1251.

Zibetti, C., Adamo, A., Binda, C., Forneris, F., Toffolo, E., Verpelli, C., Ginelli, E., Mattevi, A., Sala, C., and Battaglioli, E. (2010). Alternative splicing of the histone demethylase LSD1/KDM1 contributes to the modulation of neurite morphogenesis in the mammalian nervous system. The Journal of neuroscience : the official journal of the Society for Neuroscience *30*, 2521-2532.

Zweier, M., Gregor, A., Zweier, C., Engels, H., Sticht, H., Wohlleber, E., Bijlsma, E.K., Holder, S.E., Zenker, M., Rossier, E.*, et al.* (2010). Mutations in MEF2C from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish MECP2 and CDKL5 expression. Hum Mutat *31*, 722-733.