2017-08-08

Regulation of Pluripotency and Differentiation by Chromatin Remodeling Factors

Ly-Sha Ee

University of Massachusetts Medical School

REGULATION OF PLURIPOTENCY AND DIFFERENTIATION BY CHROMATIN REMODELING FACTORS

A Dissertation Presented

By

LY-SHA EE

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 8, 2017

INTERDISCIPLINARY GRADUATE PROGRAM

REGULATION OF PLURIPOTENCY AND DIFFERENTIATION BY CHROMATIN REMODELING FACTORS

A Dissertation Presented
By

LY-SHA EE

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

_____
Thomas Fazzio, Ph.D., Thesis Advisor

_____
Craig Peterson, Ph.D., Member of Committee

_____
Oliver Rando, M.D. Ph.D., Member of Committee

_____
Chinmay Trivedi, M.D. Ph.D., Member of Committee

_____
Toshio Tsukiyama, Ph.D., External Committee Member

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

_____
Paul Kaufman, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school.

_____
Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

Interdisciplinary Graduate Program

August 8, 2017

ACKNOWLEDGEMENTS

I first have to thank Tom Fazzio for having me in his lab, mentoring me, and for always being excited to talk about science and everything else in life. Tom taught me how to clone, how to give a good talk, and how to interact with strangers. Most importantly, Tom taught me how to find solutions and not give up when research was going wrong. Thanks to all past and present members of the Fazzio and Benanti labs for their advice and support both in science and elsewhere, and for making the lab a happy place to work. Thanks to Sarah Hainer for teaching me bioinformatics, and to Kurtis McCannell for his work on the MBD3 project, running, and for having something nice to say whenever I'm in a bad mood.

Thanks to my committee members Paul Kaufman, Craig Peterson, Ollie Rando, Chinmay Trivedi, and Toshi Tsukiyama who provided much valuable scientific insight and career guidance over the years. I also thank my collaborators- Rod Hardy for teaching me the basics of reprogramming that have become essential for my postdoc, and Feixia Chu and her lab for their contributions to this thesis work.

I could not have come to grad school without the support of Angelika Amon and her lab. Angelika gave me a chance when I knew nothing about working in a lab and taught me essential skills to pursue independent research.

Finally, thanks to Alex La Hurreau for food, love, and unlike everyone in science, letting me be right even when I am wrong.

ABSTRACT

Central to the control of virtually all cellular activity is the regulation of gene expression. In eukaryotes, this regulation is greatly influenced by chromatin structure, which is itself regulated by numerous chromatin-remodeling complexes. These are typically large protein complexes with interchangeable subunits that allow for highly specialized functions in different cell types. Moreover, additional specificity can be gained through complexes formed from different subunit isoforms. Histone modifications also regulate chromatin by recruiting remodeling complexes to particular genomic regions.

In this thesis we characterize MBD3C, an isoform of the Nucleosome Remodeling and Deacetylase (NuRD) complex subunit MBD3. MBD3 is essential for pluripotency and development, but MBD3C appears to be expressed only in embryonic stem cells (ESCs), and whether it forms a distinct NuRD complex, how its expression is regulated, and its precise function(s) remain unknown. We show that MBD3C forms a complete NuRD complex that functions redundantly with the other MBD3 isoforms in ESC gene regulation. Furthermore, MBD3C binds the SET/MLL complex subunit WDR5 through a conserved motif within its unique N-terminal region, and this interaction is necessary for the regulation of >2,000 ESC genes. Together, these findings indicate that ESCs can utilize isoforms of the same protein to achieve similar functions through diverse mechanisms.

The second part of this thesis focuses on the role of the histone modification H3.3K56ac in pluripotency and differentiation. Although H3K56ac is well-studied in yeast, in mammalian cells it is far less abundant and its functions

are largely unknown. Our data indicate that the H3.3K56R mutant is largely normal for ESC maintenance and loss of pluripotency markers during differentiation, but H3.3K56ac is necessary for proper lineage commitment. Ongoing studies will characterize the H3.3K56Q phospho-mimetic mutant during differentiation, and examine H3.3K56ac function at lineage-specific genes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND NOMENCLATURE

ESCs	Embryonic stem cells
NuRD	Nucleosome Remodeling and Deacetylase
MBD	Methyl-CpG binding domain
H3K56ac	Histone H3 acetylated at lysine 56
EC	Embryonal carcinoma
EB	Embryoid body
LIF	Leukemia inhibitory factor
lncRNA	Long noncoding RNA
iPSC	Induced pluripotent stem cell
PHD	Plant homeodomain
PRC2	Polycomb repressive complex 2
BAF	BRG1-associated factor
TSA	Trichostatin A
MEF	Mouse embryonic fibroblast
SET/MLL	SuVar3-9, enhancer of zeste, and trithorax/Mixed lineage leukemia
Trxg	Trithorax group
Pcg	Polycomb group
MSL	Male-specific lethal
NSL	Nonspecific lethal
H3K4me3	Histone H3 trimethylated on lysine 4

H3K27me3	Histone H3 trimethylated on lysine 27
5mc	Cytosine methylated at the 5’ position
5hmc	Cytosine hydroxymethylated at the 5’ position
H3F	6x Histidine, 3x FLAG tag
coIP	Coimmunoprecipitation
CGI	CpG island
ChIP-Seq	Chromatin immunoprecipitation followed by genome-wide sequencing
RNA-Seq	Genome-wide RNA sequencing
TSS	Transcription start site
CRISPR	Clustered regularly interspersed short palindromic repeats
RA	Retinoic acid
AP	Alkaline phosphatase

CHAPTER I: INTRODUCTION

The work in this dissertation focuses on two features of chromatin and the mechanisms by which they contribute to embryonic stem cell (ESC) biology. Much of ESC maintenance and function is centered on the ability to regulate gene expression, which in turn is directly linked to chromatin and the protein complexes that modify chromatin structure. Two defining properties of ESCs are self-renewal, the ability to infinitely and stably give rise to more ESCs, and pluripotency, the capacity to differentiate and form any of the cell types in the adult organism. Activation or repression of pluripotency-related genes instructs ESCs to continue to self renew or begin the differentiation process. The DNA encoding all genes is packaged with histone proteins into chromatin, and expression of a particular gene depends largely on accessibility of the gene's DNA to RNA polymerase and the transcriptional machinery. Chapter II of this thesis examines an ESC-specific isoform of the MBD3 subunit of the Nucleosome Remodeling and Deacetylase (NuRD) complex, MBD3C. MBD3C is the smallest of three MBD3 isoforms present in ESCs, and was previously largely uncharacterized. My work establishes MBD3C as a member of a complete NuRD complex, which can functionally compensate for the larger Mbd3 isoforms. The second part of this thesis focuses on a histone modification, acetylated histone H3 lysine 56 (H3K56ac). This project originated from the observation that H3K56ac localizes to many of the same genomic regions (specifically gene promoters) as two of the ESC "master regulators" OCT4 and SOX2. Although the regulation, deposition, and functions of H3K56ac have been widely studied in

yeast, far less is known about H3K56ac in mammals. How H3K56ac might interact with pluripotency factors and its role in chromatin regulation during differentiation is still unknown. The study of H3K56ac is further complicated in mammalian systems by the presence of histone H3 variants, which could be differentially acetylated during a particular cell state or developmental stage. Chapter III describes ongoing experiments to elucidate the roles of H3K56ac in ESC pluripotency and differentiation.

Embryonic stem cells

Embryonic stem cells were first isolated from mouse blastocysts in 1981 (Evans and Kaufman, 1981; Martin, 1981). The two defining features of ESCs are self-renewal and pluripotency. Self-renewal is the ability to divide indefinitely while retaining the same undifferentiated state, while pluripotency is defined as the capacity to generate any somatic cell type in the organism. The blastocyst containing ESCs forms after a series of divisions of the zygote at around embryonic day 3 (E3). It is made up of the trophectoderm, which generates extra-embryonic tissues, and the inner cell mass from which the pluripotent ESCs are derived. The ESCs are not totipotent, as they generally do not differentiate into the trophectoderm lineage, but are otherwise able to form all somatic cell types, including the cells of the germline.

The origins of ESC biology can be traced back to studies of embryonal carcinoma (EC) cells, a type of tumor cell also capable of self-renewal. Like ESCs, EC cells can give rise to multiple differentiated cell types in culture and

mice injected with single EC cells grew tumors containing multiple cell lineages (called teratocarcinomas). It was then shown that EC cells undergo differentiation *in vitro* via embryoid body (EB) formation similar to cells isolated from early mouse embryos (Martin and Evans, 1975). EBs are spherical aggregates of cells that can differentiate to form cells of each of the three germ layers (ectoderm, mesoderm, and endoderm), and EB assays in ESCs are thought to replicate developmental conditions *in vivo* and are widely used to date for differentiation studies. Furthermore, like EC cells the early mouse embryonic cells were pluripotent and could also give rise to teratocarcinomas when grafted elsewhere (reviewed in (Martello and Smith, 2014)). It was these cells, when expanded in culture without becoming cancerous that were labeled embryonic stem cells. Importantly, it was noted that ESCs had to be cultured on feeder cells to remain in the undifferentiated state. It was then shown that a cytokine secreted by the feeder cells, leukemia inhibitory factor (LIF) was the agent necessary to prevent ESC differentiation in culture (Smith et al., 1988; Williams et al., 1988). Through a series of phosphorylation events, LIF activates the transcription factor STAT3, which in turn upregulates pluripotency network transcription factors such as MYC and KLF4 (Cartwright et al., 2005; Hall et al., 2009; Matsuda et al., 1999; Niwa et al., 1998)

Among ESC transcription factors, OCT3/4 (henceforth referred to as OCT4), SOX2, and NANOG are commonly known as “core” factors that maintain ESC pluripotency in conjunction with a large network of proteins that includes both transcription factors and chromatin modifying enzymes (Orkin and

Hochedlinger, 2011; Young, 2011). OCT4 is expressed in the early embryo, and expression is confined to the inner cell mass after blastocyst formation, with OCT4 null embryo cells differentiating largely along the trophectoderm lineage (Nichols et al., 1998; Niwa et al., 2000). When LIF is withdrawn from ESCs in culture, the expression of OCT4 quickly declines, and OCT4 is absent from differentiated cell types. Similar to *Oct4* null ESCs, *Sox2* null ESCs are not pluripotent and differentiate to trophectoderm-like cells (Masui et al., 2007). SOX2 acts as a cofactor alongside OCT4 at enhancers to regulate expression of various pluripotency genes (Nishimoto et al., 1999; Yuan et al., 1995) as well as the *Oct4* and *Sox2* expression themselves (Okumura-Nakanishi et al., 2005; Tomioka et al., 2002). A third factor essential for pluripotency, NANOG is also expressed only in undifferentiated cells. Overexpression of *Nanog* allows ESCs to self-renew and remain pluripotent in the absence of LIF (Chambers et al., 2003; Mitsui et al., 2003), although *Nanog* appears to function independently of the STAT3 pathway (Chambers and Smith, 2004), and *Nanog* null ESCs are capable of continued self-renewal (Chambers et al., 2007). OCT4, SOX2, and NANOG cooperate in a positive feedback loop to maintain their own expression in pluripotent cells as well as the expression of known ESC regulators such as LIF signaling pathway or microRNA genes (Chen et al., 2008b; Marson et al., 2008; Young, 2011). The core pluripotency network can be expanded to include chromatin regulatory complexes that can localize to the same gene regulatory elements as OCT4, SOX2 or NANOG and are thought to help promote the more “open” chromatin structure at these sites characteristic of ESCs (Orkin and

Hochedlinger, 2011). Mechanistically, it has recently been proposed that OCT4 can act as a “pioneer” factor (Zaret and Carroll, 2011) through recruitment of the SWI/SNF complex to inaccessible regulatory sites. Nucleosome remodeling by SWI/SNF at said sites subsequently allows for binding and regulation by the remaining core transcription factors (King and Klose, 2017). Other chromatin regulatory complexes in both ATP-dependent remodeler and histone-modifying families have been shown to interact with OCT4, SOX2 or NANOG (Ang et al., 2011; Liang et al., 2008), and the pluripotency network has further expanded to include long noncoding RNAs (lncRNAs) (Guttman et al., 2011).

Differentiated or somatic cells can be reprogrammed to return to the pluripotent state by ectopic expression of the core pluripotency factors OCT4, SOX2, and KLF4 along with C-MYC (Takahashi and Yamanaka, 2006). These induced pluripotent stem cells (iPSCs) reactivate their endogenous pluripotency factors and their formation is further marked by methylation and deactivation of somatic genes, demethylation of other pluripotency genes, and the re-establishment of activating histone marks and other hallmarks of ESC-like chromatin structure (reviewed in (Apostolou and Hochedlinger, 2013); and see below). Although the reprogramming process is lengthy and many cell types do not form iPSCs with high efficiency, recent studies have successfully identified barriers to reprogramming and optimized reprogramming conditions in various contexts (Bar-Nur et al., 2014; Rais et al., 2014; Vidal et al., 2014). As they are a renewable source of pluripotent cells, iPSCs have greatly facilitated the study of

regenerative biology and are a potentially useful tool for treatments of cancer and developmental diseases.

Chromatin

Chromatin is the material within which the eukaryotic cell's genetic information is packaged. It is composed primarily of DNA and histone proteins tightly packed into chromosomes in the nucleus of the cell. The nucleosome is the basic unit of chromatin, formed by 147 base pairs of DNA wrapped around an octamer of conserved histone proteins (two dimers of H2A and H2B and a tetramer of H3 and H4) (Kornberg, 1974; Luger et al., 1997). Although the primary function of chromatin is to organize and retain a cell's entire DNA within the nucleus, the structure of chromatin is highly dynamic and closely regulated to accommodate cellular activities such as DNA replication, repair, transcription, and mitosis. Specifically, the densest regions of chromatin will be generally inaccessible to DNA replication, repair, or transcription factors while the chromatin at replication forks or sites containing active genes will be more loosely packed. Nucleosomes and the linker DNA connecting them can be seen in electron micrographs as "beads on a string", and numerous genome-wide mapping studies have examined nucleosome occupancy and positioning within diverse eukaroyotes. Although the existence of higher order chromatin loops and interaction domains *in vivo* has been widely postulated, the details of higher order chromatin structure are only beginning to be known. The advent of 3C (and its variants 4C, 5C, HI-C), ChIA-PET, more recently Micro-C-based technologies

(Dekker et al., 2002; Dostie et al., 2006; Fullwood et al., 2009; Hsieh et al., 2015; Lieberman-Aiden et al., 2009; Zhao et al., 2006) have provided more insight into both short and long-range chromatin interactions. At the level of individual nucleosomes, variations in nucleosome architecture can manifest through post-translational modifications of histones, enrichment or depletion of different histone variants, and nucleosome remodeling catalyzed by ATP-dependent protein complexes.

ATP-dependent nucleosome remodelers

Rearrangements of nucleosomes along chromatin are often necessary to allow transcription and other protein factors access to key regulatory regions throughout the genome. This task is commonly performed by remodeling complexes, which can function by sliding nucleosomes along DNA, removing nucleosomes, or removing and replacing different histone variants within nucleosomes (Clapier and Cairns, 2009). Nucleosome remodelers are generally classified into four families (SWI/SNF, ISWI, INO80, and CHD); remodelers in all families possess a catalytic ATPase domain homologous to the DEAD/H family of helicases (Hargreaves and Crabtree, 2011). This ATPase was first identified in yeast in two separate genetic screens: one for genes that prevented mating type switching (*swi* mutants), and a screen for mutants inhibiting growth on sucrose (*snf* mutants) (Neigeborn and Carlson, 1984; Stern et al., 1984). The ATPase gene was identified in both screens and the ATPase thus became known as SWI2/SNF2. The families each harbor different classes of domains (e.g.

bromodomain on SWI/SNF, helicase/SANT domain on SWI/SNF and INO80, chromodomain on CHD, and SANT-SLIDE domain on ISWI), which aid in recognition of other chromatin proteins or binding modules generated by post-translational modifications.

Chromatin remodelers play numerous roles in ESC pluripotency and development. In mammals chromatin remodelers are often large protein complexes with interchangeable subunits specific to particular cell lineages or developmental stages (reviewed in (Ho and Crabtree, 2010)). The ATPase BRG1, a member of the mammalian SWI/SNF complex BAF (BRG1-associated factor) is required for early embryonic development (Bultman et al., 2000) and is linked to the network of pluripotency transcription factors in ESCs (Young, 2011) where BRG1 depletion leads to loss of self-renewal and misregulation of pluripotency genes (Fazzio et al., 2008; Ho et al., 2009; Kidder et al., 2009). SWI/SNF complexes are also essential during the formation of different cell lineages in later developmental stages. For example, BRG1 is thought to bind enhancers to activate mesoderm and cardiomyocyte-specific genes in ESCs (Alexander et al., 2015) and mutants of BRG1 or its homolog *Brahma* in mice and zebrafish exhibit severe cardiac defects (reviewed in (Hota and Bruneau, 2016)). Furthermore, tissue-specific BAF subunits are required for proliferation of neural progenitors (Lessard et al., 2007) and BRG1 is important in both neuronal and muscle cell differentiation (la Serna et al., 2001; Weider et al., 2012). ISWI complexes have similar roles as SWI/SNF in early embryonic development, with null mutants of the ISWI ATPase *Snf2h* exhibiting embryonic lethality (Stopka

and Skoultchi, 2003), and ISWI complexes are required for ectoderm, endoderm, and erythropoietic development and gene regulation (Landry et al., 2011; 2008; Yip et al., 2012). INO80 complexes also regulate ESC self-renewal, pluripotency, and differentiation. In yeast and mammals the INO80 subfamily members SWR1, p400, or SRCAP function primarily in deposition of the H2A.Z histone variant into nucleosomes (Htz1 in yeast), which is important for regulation of gene expression and differentiation (Creyghton et al., 2008; Krogan et al., 2003; Mizuguchi et al., 2004; Ruhl et al., 2006), while the INO80 subfamily removes H2A.Z from nucleosomes and replaces it with canonical H2A-H2B dimers (Papamichos-Chronakis et al., 2011). In ESCs INO80 and another SWR1-related complex, TIP60-p400 complex are both implicated in ESC-specific gene regulation and self-renewal. TIP60-p400 has both H2A.Z dimer exchange and histone acetyltransferase activity and represses developmental genes while INO80 activates pluripotency genes through RNA Polymerase II and Mediator recruitment (Fazzio et al., 2008; Wang et al., 2014). Finally, the functions of the CHD family in ESCs (specifically the CHD3 and CHD4 subfamily) are a major focus of this work and will be discussed in subsequent sections. CHD1 is required both during pre-implantation development to maintain euchromatin structure and pluripotency in ESCs (Gaspar-Maia et al., 2009) and later during blood cell development in the upregulation of hematopoietic-specific genes (Koh et al., 2015).The remaining members, CHD2 and CHD5-9 have diverse functions in multiple developmental stages (reviewed in Ho and Crabtree, 2010, and Hota and Bruneau, 2016). Together, it is apparent that ATP-dependent chromatin

remodelers have adapted to function in many different contexts by utilization of both unique domains that recognize particular chromatin landmarks and accessory subunits that are particular to certain cell or tissue types.

Histone modifications

Alongside the information encoded in the sequence of DNA, histone modifications provide additional means for the regulation of chromatin. The first histone modifications identified were acetylation and methylation (ALLFREY et al., 1964) and the list of modifications has expanded to include ubiquitin, SUMO, phosphate, crotonyl, and other groups which are post-translationally added to histone residues. Advances in ChIP-Seq and mass spectrometry have allowed for the continuous identification and genome-wide mapping of further modifications to date (reviewed in (Kouzarides, 2007; Lawrence et al., 2016)). Histones can be simultaneously modified at many different residues, with some modifications present in multiple groups on the same amino acid (lysine and arginine methylation are probably the most well known examples), lending a great deal of potential complexity to studies of biological function. The ever increasing number of modifications and possible combinations of modifications led some researchers to propose the existence of a “histone code” (Strahl and Allis, 2000) which hypothesizes that distinct combinations of modifications are recognized by specific regulatory proteins which accordingly trigger a specific outcome. The validity of the histone code hypothesis has been strongly questioned (reviewed in (Rando, 2012)). Although the most well-studied

modifications are situated on the N-terminal histone tails, modifications have also been identified within the globular histone domains (Lawrence et al., 2016; Tessarz and Kouzarides, 2014). The functions of one such modification, acetylation of histone H3 lysine 56 in ESCs will be discussed in detail in Chapter III of this thesis (see also below). Two commonly known roles for histone modifications are 1) the regulation of higher order chromatin structure, and 2) the creation or elimination of binding sites for chromatin remodeling enzymes. An example of 1) is H4K16ac, which was shown to inhibit higher order chromatin folding and compaction (Shogren-Knaak et al., 2006). Countless instances of 2) exist; notable ones include H3K4 methylation and H3K9 methylation. H3K4 methylation is catalyzed and bound by the trithorax group (Trxg) proteins (Wysocka et al., 2005) and has also been shown to inhibit binding of the NuRD complex (Zegerman et al., 2002). H3K9 methylation is associated with gene repression and can trigger heterochromatin formation at centromeres and other repetitive regions of the genome when bound by HP1 (Bannister et al., 2001; Lachner et al., 2001). Several families of chromatin modifiers have been identified that both catalyze and remove histone methylation, acetylation and other modifications, and the characteristic domains of chromatin remodelers (described in the previous section) have similarly evolved to recognize specific modifications (e.g. SWI/SNF family bromodomains bind acetylated histones, and CHD family chromodomains and PHD domains recognize methylated histones), implicating histone modifications in recruiting regulatory complexes to their target genes and in processes such as transcription, replication, and DNA repair

(Kouzarides, 2007; Lawrence et al., 2016). While a particular modification can be associated with both active and repressed genes in different contexts, acetylation is generally correlated with active genes, as are some methyl marks such as promoter-proximal H3K4me3 and H3K36me3 in gene bodies. Deacetylation and other methyl marks such as H3K9me3 and H3K27me3 are correlated with silenced genes. It is important to note however, that while a particular modification is often mapped to a large subset of active or repressed genes, there is often little evidence that that modification directly activates or silences said genes.

Lysine 56 of histone H3 can be acetylated or methylated and associated with both transcriptional activation and repression (Hyland et al., 2005; Jack et al., 2013; Xu et al., 2005). H3K56ac is catalyzed by the RTT109 histone acetyltransferase and ASF1 chaperone in yeast (Driscoll et al., 2007; Han et al., 2007; Schneider et al., 2006; Tsubota et al., 2007) and by p300/CBP and GCN5 in *Drosophila* and mammals (Das et al., 2009; Tjeertes et al., 2009). Located in the globular histone fold domain of H3 (see Chapter III, Figure 3.1), H3K56 is thought to form water-mediated contacts between the nucleosome and the DNA (Luger et al., 1997; Masumoto et al., 2005); acetylation of H3K56 is believed to enhance nucleosomal unwrapping and “breathing”, allowing for increased DNA accessibility and remodeling by SWI/SNF complexes (Neumann et al., 2009). H3K56ac is required to recruit SWI/SNF to activate histone genes in vivo (Xu et al., 2005) and also for chromatin assembly both during DNA replication by the CAF-1 chaperone and in response to DNA damage (Chen et al., 2008a; Li et al.,

2008; Masumoto et al., 2005). More recently in yeast, it has been shown that a K56 hyperacetylation mimic, H3K56Q promotes H2A-H2A.Z dimer exchange by the SWR-C chromatin remodeler, resulting in decreased promoter-proximal H2A.Z (Watanabe et al., 2013). H3K56ac and H2A.Z both mark promoter-proximal nucleosomes with high turnover rate (Kaplan et al., 2008; Raisner et al., 2005; Rufiange et al., 2007) and are thus associated with transcriptional activation. How these histone marks regulate gene expression remains largely undescribed, although loss of H3K56ac was recently shown to cause decrease in RNA polymerase II occupancy at sites of both coding and noncoding transcription (Rege et al., 2015).

H3K56ac has been detected in only very small amounts in mammalian cells (1% of histone H3 vs. ~30% in yeast) and its functions in mammals are still mostly unknown. In ESCs H3K56ac is thought to be important for both pluripotency and differentiation through its interactions with OCT4, SOX2, and NANOG at pluripotency genes, and through its enrichment at developmental regulators such as the *HOX* genes during differentiation (Tan et al., 2013; Xie et al., 2009). Deletion of *Sirt6,* a NAD-dependent deacetylase that targets H3K56, leads to de-repression of OCT4, SOX2, and NANOG during differentiation, and mis-expression of markers from all three germ layers (Etchegaray et al., 2015).

Histone variants and chaperones

Because the packaging of cellular DNA is essential in all cells, histones are among the most highly conserved proteins in eukaryotes. In mice and

humans the canonical histone genes are arranged in large, multicopy clusters distributed onto three chromosomes (Marzluff et al., 2002). The gene clusters are transcribed in a cell cycle-dependent manner during S phase, and histone transcripts possess a unique structure without introns and terminating in a stem-loop structure in place of the typical poly A tail (Dominski and Marzluff, 1999) that allow for large quantities of histones to be synthesized when DNA is replicated. Although canonical histones make up the majority of the nucleosomes in the genome, a number of variants exist for every histone (H4 variants are only known to exist in a few lower eukaryotes) each with a unique gene sequence(s) and specific genomic localization(s). The histone variants are encoded by one or two genes separately from the histone clusters, and are expressed and deposited on chromatin throughout the cell cycle. While histone variants can be removed or exchanged from nucleosomes by ATP-dependent remodelers, histone chaperones function to assemble newly synthesized histones into nucleosomes. Chaperones can be specific to different histone variants; most notably the Chromatin Assembly Factor 1 (CAF-1) chaperone complex incorporates the core histone H3 variants H3.1 and H3.2 into chromatin following DNA replication, while the H3.3 variant is deposited by the HIRA, DAXX/ATRX, and p400 complexes independently of replication (Drané et al., 2010; Goldberg et al., 2010; Lewis et al., 2010; Pradhan et al., 2016; Tagami et al., 2004). H3.3 differs from H3.1 and H3.2 by four or five amino acids (see Chapter III, Figure 3.1A) and the recognition of H3.3 by separate chaperones is specified by the changed amino acids in the histone fold domain (Ahmad and Henikoff, 2002). In addition to H3.3,

other H3 variants exist in mammals including the centromere- and testis-specific variants CENP-A and H3t (reviewed in (Szenker et al., 2011)).

H3.3 has traditionally been associated with actively transcribed genes and enhancers and is enriched for “active” histone modifications (Goldberg et al., 2010; Hake et al., 2006; McKittrick et al., 2004; Wirbelauer et al., 2005). H3.3 also co-localizes with H2A.Z on nucleosomes at many promoters and enhancers, such nucleosomes are believed to be less stable than nucleosomes containing canonical histones and thus more conducive to creating open chromatin structure and activating transcription (Chen et al., 2013a; Jin and Felsenfeld, 2007; Jin et al., 2009). Surprisingly, H3.3 is also associated with gene repression. It is enriched at heterochromatic regions and in ESCs is involved in silencing of endogenous retroviral elements (Elsässer et al., 2015; Goldberg et al., 2010). Furthermore, PRC2-mediated H3K27me3 is reduced in H3.3-depleted ESCs at developmental genes, resulting in increased expression of trophectoderm and other lineage markers (Banaszynski et al., 2013).

Unique chromatin structure of embryonic stem cells

ESCs must precisely regulate the expression of different genes in order to retain both the capacity for self-renewal and the ability to differentiate. Depending on differentiation signals received, pluripotent ESCs are able to upregulate specific subsets of genes corresponding to any lineage. Thus, the structure and features of ESC chromatin largely differ from that of most somatic cell types. Most notably, ESCs exhibit a dynamic, “open” chromatin structure, which can be

visualized under the microscope by reduced staining for the heterochromatin markers HP1 and H3K9me3 as compared with differentiated cells (Meshorer and Misteli, 2006). In contrast to euchromatin, heterochromatin consists of highly condensed nucleosomes and its formation often results in gene silencing. As such, global transcription levels are elevated in ESCs (Efroni et al., 2008) and ESC knockdown of the chromatin remodeler *Chd1* has been shown to trigger heterochromatin formation and reduced pluripotency (Gaspar-Maia et al., 2009). ESC chromatin is also enriched for “active” histone modifications such as H3K4me3 and H3K9ac. Another hallmark of pluripotent chromatin is a subset of gene promoters marked with both an active (H3K4me3) and a repressive (H3K27me3) histone modification by the Trxg and PRC2 complexes respectively. These “bivalent” genes are lowly expressed in undifferentiated cells or thought to be “poised” for activation in response to developmental signaling (Azuara et al., 2006; Bernstein et al., 2006). During differentiation one of the marks is typically lost from the promoter while the other becomes enriched depending on whether the gene is expressed or silenced. Consistent with their poised state, bivalent genes tend to have low levels of DNA methylation, another repressive epigenetic mark, with he CpG islands of germline and Polycomb genes becoming increasingly methylated as ESCs differentiate and commit to specific lineages (Mohn et al., 2008).

DNA methylation

DNA methylation primarily occurs at the 5’ position of cytosine bases that immediately precede a guanine (CpG dinucleotides). Methylated cytosine (5mc) is conserved in many organisms ranging from fungal to plant and mammalian species and has broad functions in transcriptional control, genomic imprinting, and silencing of transposons (Goll and Bestor, 2005). Methylation in mammals is catalyzed by the DNA methyltransferases DNMT3A, DNMT3B, and DNMT1; an additional DNMT family member DNMT3L modulates methylation levels in ESCs and germ cells (Bourc'his et al., 2001; Neri et al., 2013). DNMT3A and 3B are “de novo” methyltransferases that establish methylation patterns during development and are required for embryonic viability in mouse (Okano et al., 1999), while DNMT1 largely methylates hemimethylated CpGs during DNA replication to maintain established methylation patterns. Although most mammalian CpGs are methylated, regions containing dense CpG clusters ranging from a few hundred to ~1000 base pairs are found at transcription start sites (TSSs) and are generally hypomethylated (Deaton and Bird, 2011). These CpG islands (CGIs) mark most “housekeeping” genes and tend to be enriched for H3K4me3 and other histone modifications associated with active transcription (Thomson et al., 2010). In ESCs, pluripotency genes silenced upon differentiation become methylated while concurrently losing active and acquiring repressive histone modifications, thus enabling repressive heterochromatin formation at the gene promoters (reviewed in (Smith and Meissner, 2013)). Loss of DNA methyltransferases and methylation does not affect ESC self-renewal (Tsumura et al., 2006), although *Dnmt3a* and *Dnmt3b* double null cells are unable to

differentiate to most germ layers (Jackson et al., 2004). ESCs with decreased methylation resulting from *Dnmt1* null mutations appear to upregulate trophectodermal markers (Ng et al., 2008) suggesting that DNA methylation functions to prevent otherwise pluripotent ESCs in the inner cell mass from forming extraembryonic tissues. Through its association with heterochromatin and repressive histone modifications, DNA methylation is linked to gene silencing. The primary mechanisms by which methylation could repress transcription are 1) 5mc at promoter regions could physically block transcription factors or RNA polymerase II from binding; 2) methylation could alter nucleosome occupancy or positioning at promoters and/or gene bodies to create repressive chromatin structure; or 3) methylation can recruit methyl-binding domain (MBD) proteins which themselves recruit repressive protein complexes. The MBD proteins with the exception of MBD3 bind methylated DNA (Klose and Bird, 2006) and are linked to transcriptional co-repressors (Jones et al., 1998; Kondo et al., 2005; Nan et al., 1998; Ng et al., 1999; Sarraf and Stancheva, 2004; Zhang et al., 1999). Most of these co-repressors are histone deacetylases, although MBD1 associates with the H3K9 methylase SETDB1 to repress transcription during DNA replication (Sarraf and Stancheva, 2004). Functions of the MBD2 and MBD3 proteins will be described in more detail in the following sections.

DNA methylation can be lost either through failure of DNMT1 to restore methylation during replication or through active demethylation by TET family enzymes, which oxidize 5mc to 5-hydroxymethylcytosine (5hmc), 5-

formylcytosine (5fc), and 5-carboxycytosine (5cac) (He et al., 2011; Ito et al., 2010; 2011; Tahiliani et al., 2009). The oxidized intermediates are then converted back to unmodified cytosines by the base excision repair pathway (Bhutani et al., 2011; Cortellino et al., 2011)). Although it is possible that 5hmc is merely an intermediate in DNA demethylation, it is very likely that 5hmc and the TET proteins have important cellular functions. The TET1 protein was identified as a fusion partner of the methyltransferase MLL in acute myeloid leukemia (Lorsbach et al., 2003; Ono et al., 2002) and TET2 is frequently mutated in leukemia patients (Cimmino et al., 2011; Wu and Zhang, 2011). In ESCs, TET1 and TET2 KD ESCs have self-renewal defects and similarly to *Dnmt1* mutants can differentiate into trophectoderm (Ito et al., 2010; Koh et al., 2011)), although *Tet1* null mice are still viable (Dawlaty et al., 2011). Although current evidence indicates that TET1 and TET2 are regulated by OCT4 and the core pluripotency factors (Koh et al., 2011; Wu et al., 2013), and our lab has identified 5hmc as a possible recruitment module for MBD3 and the NuRD co-repressor complex (Yildirim et al., 2011) see below) future studies are necessary to more precisely determine the roles of TET proteins and 5hmc in ESCs and development.

The Nucleosome Remodeling and Deacetylase Complex

The NuRD complex (originally known as the Mi-2 complex) was isolated and characterized by four separate groups in 1998 (Tong et al., 1998; Wade et al., 1998; Xue et al., 1998; Zhang et al., 1998). The first three groups sought to understand how chromatin-remodeling complexes might regulate transcription and to that end were biochemically characterizing HeLa cell histone deacetylase

or SWI2/SNF2 remodeler complexes. The remaining group isolated the related deacetylase activity from *Xenopus* egg extracts following an observation that deacetylation was important for normal *Xenopus* development. Each of the studies described a complex containing parallel ATP-dependent nucleosome remodeling activity and histone modification (deacetylase) activity, a property unique to NuRD. The complex is highly conserved in metazoans and contains at least one subunit from each of five major protein families (Figure 1.1). Several groups have reported additional proteins that co-purify with NuRD in different tissue and cell types (reviewed in (McDonel et al., 2009)). Because deacetylation is traditionally associated with transcriptional repression, NuRD is primarily known as a repressive complex, and the ATP-dependent remodeling activity is thought to block RNA polymerase II and transcription factors at promoters by increasing nucleosome occupancy (Denslow and Wade, 2007; Hainer and Fazzio, 2015; Yildirim et al., 2011). In addition to transcriptional control, NuRD activity is also associated with higher order chromatin assembly, maintenance of genome stability, hematopoietic stem cell differentiation, various human cancers, and aging (Lai and Wade, 2011; Pegoraro et al., 2009; Yoshida et al., 2008).

Overview of NuRD subunits

In support of the idea that multiple different NuRD complexes and sub-complexes are adapted for highly specific functions, the NuRD complex is highly modular, with many subunits interchangeable, mutually exclusive, or present only in certain cell types. The enzymatic subunits of the canonical NuRD complex are

an ATP-dependent remodeler (CHD3 or CHD4, also known as MI2-α and MI2-β), and two histone deacetylases (HDAC1 and HDAC2). Other subunits are the metastasis-associated proteins MTA1, MTA2, or MTA3, WD40 repeat proteins RBBP4 and RBBP7 (also known as RBAP48 and RBAP46), the zinc finger proteins p66α and p66β (also known as GATAD2A and GATAD2b), and the methyl binding domain proteins MBD2 or MBD3. The exact stoichiometries of each subunit are still unknown, but a recent study using applied mass spectrometry analysis suggests that there is a single ATP remodeling subunit (either CHD3 or CHD4) and a single deacetylase (HDAC1 or HDAC2) (Smits et al., 2013). The study also proposes three subunits of MTA1, 2, or 3; six subunits of RBBP4 and/or RBBP7; one MBD3; two subunits of p66α or p66β (Figure 1.1), and two of an additional DOC1 subunit. Using structural studies, Schwabe and colleagues propose an alternative stoichiometry where an MTA homodimer binds two HDAC and four RBBP4/7 subunits (Millard et al., 2016; 2013).

Several subunits have been identified as members of related chromatin modifying complexes and sub-complexes such as CoREST, NODE, and MeCP1 (Liang et al., 2008; Ng et al., 1999; You et al., 2001), and NuRD's deacetylase activity can function independently of the chromatin remodeling subunit in vitro (Low et al., 2016), suggesting that NuRD's enzymatic activites can act either separately, together within the complex, and/or in conjunction with other chromatin remodelers. Numerous peripheral subunits are present in functionally specialized NuRD complexes. One such example is LSD1, an H3K4 and H3K9 demethylase which is co-enriched with NuRD at enhancers in ESCs (Whyte et

al., 2012) and binds the MTA subunits of NuRD in breast cancer cells (Wang et al., 2009b). LSD1 and NuRD-mediated enhancer activity is essential for proper ESC differentiation, while Wang et al. propose that LSD1 inhibits breast cancer metastasis. A second NuRD-interacting protein is FOG1, which interacts with NuRD in erythroid progenitors and is important for hematopoietic differentiation (Gao et al., 2010; Hong et al., 2005). A third such example is BCL6, a transcriptional repressor that functions with the NuRD MTA3 subunit to control B lymphocyte cell fate (Fujita et al., 2004). The canonical NuRD subunits are described below, with emphasis placed on their function in ESCs.

The **CHD3** and **CHD4** proteins are ATP-dependent nucleosome remodeling subunits. Also known as MI2-α and MI2-β, they were first identified in dermatomyositis patients as autoantigens (Seelig et al., 1995). Like all CHD family members, CHD3 and 4 contain the SWI/SNF ATPase/helicase domain, and two tandem-terminal chromodomains, which can bind methylated H3 or DNA and are required for nucleosome remodeling (Clapier and Cairns, 2009). In addition to their chromodomains and ATPase domain, CHD3 and 4 also each contain two plant homeodomain (PHD) fingers, which also bind methylated H3 tails (Musselman et al., 2012). In ESCs, CHD4 KD leads to loss of self-renewal capability, decreased proliferation, and increased embryoid body (EB) differentiation as measured by downregulation of the pluripotency markers OCT4 and TBX3 and upregulation of germ layer marker genes (Zhao et al., 2017). Recently, a third CHD family member, **CHD5** has also been co-purified with the canonical NuRD subunits (Kolla et al., 2015; Nitarska et al., 2016; Quan et al.,

2014). CHD5 is required along with CHD3 and CHD4 at distinct stages of neuronal differentiation and brain development, and the CHD subunits form mutually exclusive NuRD complexes that specifically regulate genes during the different developmental stages (Egan et al., 2013; Nitarska et al., 2016). CHD5 is not highly expressed in most tissue types outside of the brain and the testis, where it is also necessary for spermatogenesis (Li et al., 2014; Zhuang et al., 2014).

The **HDAC1** and **HDAC2** subunits are Class I lysine deacetylases and homologs of the yeast RPD3 deacetylase. In ESCs, HDAC1 binds promoters of pluripotency genes and cells treated with the HDAC inhibitor trichostatin A (TSA) exhibited a flattened morphology consistent with differentiation (Kidder and Palmer, 2012). Interestingly, the expression of a subset of pluripotency genes decreased after TSA treatment, suggesting that HDAC1 could be linked to gene activation as well as repression. Other studies have provided additional evidence for a role for HDACs in transcriptional activation (Zupkovitz et al., 2006).

The role of the **MTA1**, **MTA2**, and **MTA3** subunits within the NuRD complex and in ESCs is largely unknown. MTA1 and 2 have been widely characterized in cancer metastases, particularly in breast cancer where high MTA1 expression levels have been linked to tumorigenesis (reviewed in Lai and Wade, 2011). Each MTA family member contains four conserved domains: bromo-adjacent homology (BAH), SANT, EGL-27 and MTA1 homology (ELM), and GATA zinc finger. The specific function(s) of each domain are also unknown, but are hypothesized to be involved in recognition and binding to other NuRD

subunits and mediating interactions between the complex and other proteins. In support of this hypothesis, a study by Kumar and colleagues reported that methylated MTA1 is involved in NuRD assembly and is also required for CHD4 nucleosome remodeling *in vitro* (Nair et al., 2013).

The **RBBP4** and **RBBP7** subunits (also named RBAP48 and 46 for their respective molecular weights) are members of the WD40 repeat protein family. WD40 proteins exhibit a characteristic seven-bladed β-sheet propeller structure with each blade approximately 40 amino acids in length and flanked by N- and C-terminal helical tails. RBBP4 and RBBP7 were first identified and named for their interaction with the retinoblastoma tumor suppressor protein and later shown to bind histones. One or both of RBBP4 or RBBP7 have been co-purified with several additional complexes such as the Polycomb Repressive Complex 2 (PRC2), and Chromatin Assembly Factor 1 (CAF1), and Nucleosome Remodeling Factor (NURF) complexes, and the RBBP proteins are thought to be important both as structural subunits and in recruiting chromatin modifiers to their genomic targets through their histone binding properties.

The **MBD3** and **MBD2** subunits are mutually exclusive within the NuRD complex (Le Guezennec et al., 2006). As the biology of MBD3/NuRD is central to the work presented in Chapter II of this thesis, the origins of MBD3 and its known functions in ESCs will be discussed in more detail in the following section. Identified alongside MBD3 and MBD4, MBD2 contains a domain (the MBD), which binds methylated CpG dinucleotides on DNA (Hendrich and Bird, 1998) and a C-terminal transcription repression domain. As such, MBD2 is thought to

function in recruiting NuRD to its target genes as a “reader” of methylated DNA. MBD3 is the only MBD protein that does not bind methylated DNA, and contains a glutamate repeat domain of unknown function at its C-terminus. Despite always associating with separate NuRD complexes, MBD2 and MBD3 localize interdependently to many of the same genomic regions (Günther et al., 2013; Hainer et al., 2016). However, *Mbd2* and *Mbd3* exhibit very different phenotypes *in vivo*- *Mbd3* null mice are embryonic lethal at 8.5 d.p.c., while *Mbd2*-null mice survive to adulthood, though *Mbd2* null females exhibited abnormalities in maternal behavior resulting in smaller offspring (Hendrich et al., 2001). Both MBD2 and MBD3 bind the **p66α** and **p66β** subunits (Brackertz et al., 2002). Also known as GATAD2a and GATAD2b, these proteins each contain a conserved GATA zinc finger domain that can bind histone tails, although binding appears to be inhibited if histone tails are acetylated (Brackertz et al., 2006). p66α and p66β also interact with the HDAC1 and RBBP7 subunits in a sumolyation-dependent manner, indicating that p66α and p66β play both structural and recruitment roles within NuRD (Gong et al., 2006).

It is clear from the above overview that there is still much to be discovered about the biology of the NuRD complex. While there is a general consensus on the identities of the canonical subunits, the list of peripheral proteins interacting with NuRD continues to expand, one of which is a main focus of this thesis. Further studies will be necessary to determine how each subunit functions both within and outside the context of a NuRD complex or subcomplex, and whether these functions are universal or specific to particular cell or tissue types. It is well

established that NuRD impacts both chromatin structure and transcription in many cell types, and much present work aims to elucidate how NuRD is recruited to different target genes, the precise mechanism(s) by which the complex effects gene repression, and how crosstalk between NuRD and other chromatin regulatory complexes affects these processes. The following chapter will focus on a single subunit, MBD3 and its role in NuRD activity in ESCs.

Functions of MBD3/NuRD in embryonic stem cells and development

The first discovered methyl-DNA binding proteins were MECP1 and MECP2 (Lewis et al., 1992; Meehan et al., 1989). The biology of MECP2 became of particular interest when mutations in the *MeCP2* gene were found to cause Rett Syndrome, an X-linked neurodevelopmental disease that is lethal in males (Amir et al., 1999). Hendrich, Bird, and colleagues subsequently identified *Mbd3* and its counterparts *Mbd1*, *2*, and *4* while searching an EST (expressed sequence tag) database for additional methyl binding domain proteins (Cross et al., 1997; Hendrich and Bird, 1998). Two MBD3 isoforms, MBD3A and MBD3B were originally identified with a third, MBD3C, described later in mouse (Hendrich and Bird, 1998; Kaji et al., 2006). The MBD is truncated in MBD3B at amino acids 5-36 and is completely absent in MBD3C (see Chapter II).

Unlike MECP2 and the other Mbd family members, MBD3 does not bind methylated DNA (Hendrich and Bird, 1998; Zhang et al., 1999), despite its sequence being ~70% similar to that of MBD2. Analysis of the MBD3A MBD reveals that two amino acids present in mice and humans, His-30 and Phe-34,

are replaced by Lys and Tyr respectively in most non-mammalian species (Hendrich and Tweedie, 2003). Although MBD3's MBD does not bind methylcytosine (5mc), data have shown that mutating His-30 and Phe-34 back to Lys and Tyr restores the ability of MBD3 to bind 5mc (Saito and Ishikawa, 2002). Mbd3 appears to bind more selectively to hydroxymethylated DNA and unmethylated DNA than methylated DNA *in vitro* (Mellén et al., 2012; Yildirim et al., 2011) and MBD3 and 5'hydroxymethylcytosine (5hmc) map to overlapping regions of the genome in ESCs (Wu et al., 2011; Yildirim et al., 2011). Cytosine bases methylated at the 5' position (5mc) are converted to 5hmc by the hydroxylase TET1 (Tahiliani et al., 2009) and *Tet1* KD ESCs exhibit phenotypes similar to *Mbd3* KD ESCs (Ito et al., 2010; Koh et al., 2011). Although our lab's *in vitro* binding data has been contested by data from other groups (Hashimoto et al., 2012; Spruijt et al., 2013), we observed that genome-wide MBD3 localization is lost in *Tet1* KD cells and *Tet1* catalytic-inactive mutant cells (Hainer et al., 2016; Yildirim et al., 2011), indicating that TET1 and 5hmc are important for MBD3/NuRD to bind its target genes in ESCs.

In an RNAi screen to identify chromatin regulators important for murine ESC self-renewal, our lab identified MBD3 as important for maintenance of the ESC state (Fazzio et al., 2008). It was further observed that nucleosome occupancy is decreased and RNA Polymerase II recruitment is increased at MBD3 target genes in *Mbd3* KD ESCs (Yildirim et al., 2011), indicating that MBD3 is a transcriptional repressor. Additionally, *Mbd3* null mice are nonviable (Hendrich et al., 2001) and ESCs derived from *Mbd3* null mouse embryos are

capable of self-renewal in culture in the absence of leukemia inhibitory factor (LIF) and are compromised in pluripotency, with differentiation skewed toward the trophectoderm lineage (Kaji et al., 2006; 2007; Zhu et al., 2009) MBD3 was subsequently shown to be important for differentiation and development through silencing of a subset of pluripotency genes (*Zfp42*, *Tbx3*, *Klf4*, and *Klf5*) and the expression of these genes persists after removal of LIF in *Mbd3* KD ESCs (Reynolds et al., 2012). From this study, it was hypothesized that MBD3/NuRD could function to oppose the LIF/STAT3 pathway, which activates the same subset of pluripotency genes. In support of this hypothesis, our lab found that MBD3 interacts with and localizes to several hundred of the same genes as the esBAF catalytic subunit BRG1. BRG1 is required for STAT3 to localize to many of its target genes (Ho et al., 2011) and MBD3 and BRG1 were observed to oppositely regulate expression of their shared target genes (Yildirim et al., 2011). Furthermore, MBD3 is essential for NuRD complex assembly (Kaji et al., 2006) and NuRD target genes exhibit increased H3K27 acetylation and decreased H3K27 trimethlyation in *Mbd3* null ESCs (Reynolds et al., 2011), suggesting that MBD3 regulates pluripotency genes through NuRD-mediated deacetylation and recruitment of the H3K27 methyltransferase Polycomb Repressive Complex 2 (PRC2). Taken together, the abovementioned studies point towards a model where MBD3/NuRD works to maintain equilibrium in ESC target gene expression alongside other chromatin modifiers which in turn allows ESCs to flexibly coordinate the self-renewal or differentiation processes (Hu and Wade, 2012). Although there are some discrepancies in ChIP binding profiles as to exactly

where MBD3/NuRD maps in the ESC genome it appears that like PRC2, MBD3/NuRD is a largely repressive complex and functions in opposition to gene activating complexes like esBAF and STAT3.

The role of MBD3/NuRD in iPSC reprogramming is controversial. Two groups provided initial evidence that *Mbd3* deletion greatly enhances reprogramming efficiency (Luo et al., 2013; Rais et al., 2014) with Rais et al. reporting successful reprogramming of nearly 100% of induced cells to pluripotency. This result was subsequently challenged by work from Hendrich, Silva, and colleagues who reported that *Mbd3* is required for both initial and later stage reprogramming of neural stem cells and as well as reprogramming of epiblast stem cells to naïve pluripotency (Santos et al., 2014). To date the discrepancies between these data remain largely unresolved (Bertone et al., 2015; Zviran et al., 2015), although differing reprogramming reagents/contexts and/or experimental conditions are likely contributors to the contradictory results.

Functions of WDR5 in embryonic stem cells

WDR5 (WD repeat protein 5) is conserved from yeast to humans and is a member of several transcriptional co-activator complexes (Cai et al., 2010; Dou et al., 2006; 2005; Ruthenburg et al., 2006). It is most commonly known as a subunit of SET/MLL (SuVar3-9, Enhancer of zeste and Trithorax/Mixed Lineage Leukemia), a histone methyltransferase complex homologous to the yeast COMPASS complex. The SET/MLL proteins are also homologs of the *Drosophila* trithorax group (Trxg) proteins, which were first identified as positive regulators of

the *Drosophila Hox* gene cluster, and thought to function to counteract the Polycomb Group (Pcg) negative regulator proteins (reviewed in (Schuettengruber et al., 2011)). COMPASS was subsequently purified from yeast as the first H3K4 methylase (Miller et al., 2001; Roguev et al., 2001). H3K4 methylation is catalyzed by the SET1 subunit (encoded by the *MLL* gene in humans and mice) and the complex can catalyze all forms of H3K4 methylation (H3K4me1, 2, and 3). WDR5 and the other complex core subunits ASH2L and RBBP5 are essential for this catalytic activity in yeast and mammals. WDR5 binds di- and tri-methylated H3K4 and has been shown to activate transcription in reporter gene assays and to be required for *HOX* gene expression maintenance (Wysocka et al., 2005). However, although the H3K4me3 mark is widely associated with transcriptional activation as it localizes to the promoters of many active genes (Bernstein et al., 2002; Santos-Rosa et al., 2002), it does not appear to itself directly activate these genes.

In ESCs, WDR5 is required for proper self-renewal, and its expression decreases along with H3K4me3 levels during differentiation (Ang et al., 2011). Consistent with these observations, Ang et al. further showed that *Wdr5* KD MEFs did not reprogram to iPSCs as efficiently as wildtype, and that WDR5 both physically interacts with and binds to overlapping genomic regions as OCT4, likely co-regulating genes important for ESC maintenance. WDR5's activity within the SET/MLL complex is of particular importance in maintaining H3K4me3 at the aforementioned "bivalent" genes in ESCs. Another mechanism by which WDR5 maintains ESC chromatin state is through histone acetylation. WDR5 was also

purified as a member of MOF (Male absent On First), a protein complex with both H4K16 acetyltransferase and H3K4 methyltransferase activity (Dou et al., 2005). Like H3K4me3, H4K16ac is associated with transcriptional activation, and *Mof* null ESCs exhibit similar self-renewal and pluripotency defects as *Wdr5* KD ESCs (Li et al., 2012). Additionally, there are two types of MOF-containing complexes in *Drosophila* and mammals, the male-specific lethal (MSL) and non-specific lethal (NSL) and WDR5 is also a member of both (Zhao et al., 2013). MSL and NSL bind different genomic regions and have separate functions in ESCs (Chelmicki et al., 2014; Ravens et al., 2014), and the exact function(s) of WDR5 in each of these MOF complexes is still unknown.

An additional mechanism for WDR5 function in ESCs is through interaction with lncRNAs. LncRNAs are >200bp RNAs that are typically processed (contain a 5'cap and polyA tail) and are not usually translated (Guttman et al., 2013; Wang and Chang, 2011). Following up on studies characterizing interactions between WDR5 and lncRNAs (Gomez et al., 2013; Wang et al., 2011), Howard Chang and colleagues identified ~1000 RNAs binding to WDR5 in ESCs using a variant of RNA immunoprecipitation, RNA:protein immunoprecipitation in tandem (RIPiT-Seq, (Yang et al., 2014)). Through additional analysis of lncRNA binding site mutants, Yang et al. showed that lncRNA binding to WDR5 was required for stable H3K4 trimethylation and proper expression of pluripotency genes in ESCs. Additionally, WDR5 protein levels were reduced in the lncRNA binding site mutants, suggesting a model where lncRNA binding would stabilize WDR5 levels, thus allowing SET/MLL

complex assembly and activity at target genes. Taken together, these studies establish that WDR5 can regulate ESC chromatin and gene expression through multiple complexes and mechanisms, and that *Wdr5* is likely essential for proper ESC function.

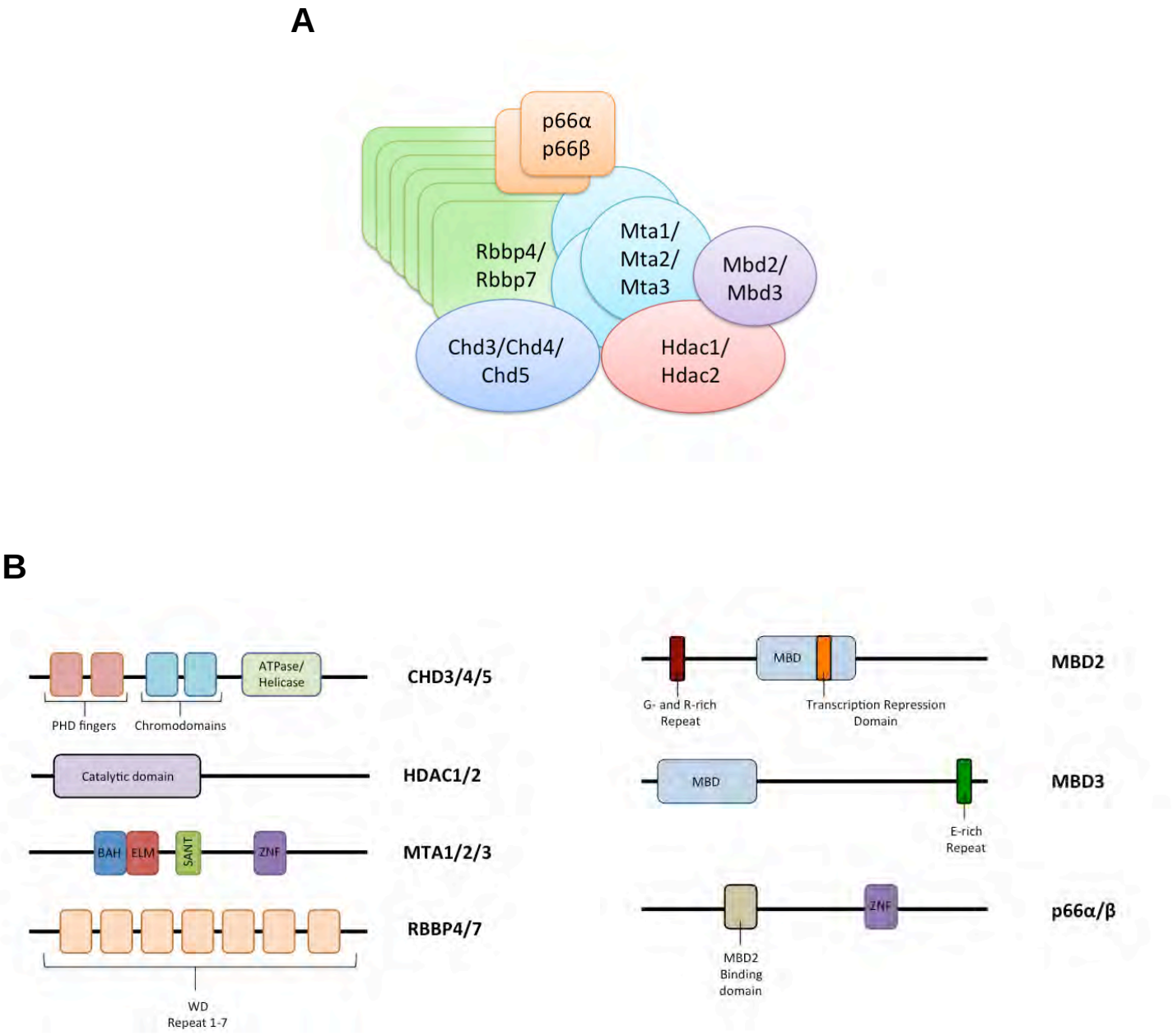


Figure 1.1: The NuRD complex

(A) Schematic of the NuRD complex showing predicted subunit stoichiometries (Smits et al., 2013). **(B)** Schematic of canonical NuRD subunit families and their known domains. See text for details.

CHAPTER II: AN ES CELL-SPECIFIC NURD COMPLEX FUNCTIONS THROUGH INTERACTION WITH WDR5

PREFACE

Chapter II is derived from an article of the same name published in *Stem Cell Reports* under a Creative Commons License:

Ee, L.S., McCannell, K.N., Tang, Y., Fernandes, N., Hardy, W.R., Green, M.R., Chu, F., and Fazzio, T.G. (2017). *Stem Cell Reports* 8, 1488-1496.

Contributions

Kurtis McCannell performed experiments in Figures 2.1D and 2.3, assisted with data collection in Figures 2.4A, C-E, and 2.5B, and constructed cell lines used in experiments in Figure 2.7A.

Mass spectrometry experiments were performed in collaboration with Feixia Chu, Yang Tang, and Nancy Fernandes (Figure 2.2D; Tables 2.1 and 2.2).

Michael Green and Rod Hardy provided the CAG overexpression vector and reprogramming reagents. Rod Hardy assisted with reprogramming experiments (Figure 2.4C-E).

All other experiments were performed and analyzed by Ly-sha Ee and designed by Ly-sha Ee and Tom Fazzio. Ly-sha Ee and Tom Fazzio wrote the manuscript with input from the other authors.

ABSTRACT

The Nucleosome Remodeling and Deacetylase (NuRD) complex is a chromatin regulatory complex that functions as a transcriptional co-repressor in metazoans. The NuRD subunit MBD3 is essential for targeting and assembly of a functional NuRD complex as well as embryonic stem cell (ESC) pluripotency. Three MBD3 isoforms (MBD3A, MBD3B, and MBD3C) are expressed in mouse. Here, we find that the MBD3C isoform contains a unique 50–amino acid N-terminal region that is necessary for MBD3C to specifically interact with the histone H3 binding protein WDR5. Domain analyses of WDR5 reveal that the H3 binding pocket is required for interaction with MBD3C. We find that while *Mbd3c* KO ESCs differentiate normally, MBD3C is redundant with the MBD3A and MBD3B isoforms in regulation of gene expression, with the unique MBD3C N-terminus required for this redundancy. Together, our data characterize a unique NuRD complex variant that functions specifically in ESCs.

INTRODUCTION

Embryonic stem cells (ESCs) are characterized by two unique properties: 1) self-renewal, the ability to indefinitely produce more stem cells and 2) pluripotency, the ability to differentiate into any cell type. To maintain their cellular identity ESCs utilize a network of core transcription factors, which bind and regulate pluripotency genes and differentiation genes in response to developmental signaling (Boyer et al., 2005; Kim et al., 2008). To achieve this flexibility in gene expression ESCs maintain an “open” chromatin structure that

contains more euchromatin than heterochromatin (Meshorer and Misteli, 2006) as well as unique patterns of histone modifications. Various protein complexes can remodel this chromatin structure (and thus regulate gene expression patterns) by modifying nucleosome positioning as well as through posttranslational modification of histones (Fazzio and Panning, 2010).

The Nucleosome Remodeling and Deacetylase (NuRD) complex is unique amongst chromatin regulators because it couples ATP-dependent nucleosome remodeling activity with histone modification (deacetylase) activity (Tong et al., 1998; Wade et al., 1998; Xue et al., 1998; Zhang et al., 1998). NuRD alters chromatin structure near its target genes by repositioning local nucleosomes to block the binding of transcriptional machinery at gene promoters, thus functioning primarily as a co-repressor (Denslow and Wade, 2007). The NuRD subunit MBD3 can also act as a transcriptional repressor, as we have observed that nucleosome occupancy is decreased and RNA Polymerase II recruitment is increased at MBD3 target genes in *Mbd3* knockdown (KD) ESCs (Yildirim et al., 2011). However, three MBD3 isoforms (MBD3A, B and C) are expressed in mouse ESCs and only MBD3A has a full-length MBD (Kaji et al., 2006). Thus, the possibility exists for formation of multiple NuRD complexes of varying subunit combinations and functional specificities. For example, each MBD3 isoform could form unique complexes (with NuRD and/or other chromatin regulators) and recruit said complexes to its genomic targets through different mechanisms. We sought to investigate this possibility in the present study.

Here, we have characterized a unique variant of the NuRD chromatin remodeling complex that harbors MBD3C, an ESC-specific isoform of MBD3, as well as the histone H3 binding protein WDR5. MBD3C is expressed almost exclusively in ESCs via an alternative CpG island (CGI)–containing promoter located in the second intron of the *Mbd3* gene. We further show that MBD3C contains a unique 50–amino acid N-terminus that is necessary for WDR5 interaction. MBD3C interacts with the WDR5 H3 binding pocket through an arginine-containing motif also utilized by MLL1 for WDR5 binding. RNA-seq analysis revealed that the three MBD3 isoforms are largely redundant for gene regulation, since knockout (KO) of all three isoforms had a more severe effect on gene expression than individual KO of *Mbd3c* or simultaneous KO of *Mbd3a* and *b*. Importantly, the WDR5-interaction domain of *Mbd3c* is critical for its gene regulatory function, suggesting that WDR5 plays critical roles in MBD3C/NuRD complex.

RESULTS AND DISCUSSION

MBD3C/NuRD co-purifies with WDR5

To identify proteins co-purifying with MBD3 in ESCs, we used a cell line in which one copy of endogenous MBD3 is fused to a C-terminal 6xHis-3xFLAG tag (*Mbd3-H3F*; (Yildirim et al., 2011), allowing for affinity purification of MBD3A, B, and C simultaneously (Figures 2.1A and 2.1B). LC-MS/MS of purified MBD3 complexes identified all canonical NuRD subunits, several of which were subsequently confirmed by western blot (Figure 2.1C and Table 2.1).

Consistent with recent mass spectrometry analyses of NuRD components (Bode et al., 2016), we detected an interaction between MBD3 and the SET/MLL complex component WDR5 (Figures 2.1C and 2.1D; Table 2.1). The MLL complex is a histone methyltransferase that catalyzes methylation of H3K4, a mark found at transcriptionally active genes (Bernstein et al., 2002; Santos-Rosa et al., 2002). WDR5 binds the histone H3 tail *in vitro* and is essential for H3K4 trimethylation and MLL complex formation (Couture et al., 2006; Dou et al., 2006; Ruthenburg et al., 2006; Schuetz et al., 2006; Wysocka et al., 2005). We did not observe any other MLL subunits co-purifying with MBD3 (Figure 2.1C and Table 2.1), suggesting WDR5 interacts with MBD3/NuRD independently of MLL complex. To validate these data, we performed co-immunoprecipitation (co-IP) assays. Interestingly, WDR5 IP pulled down MBD3C, but not the more abundant isoforms, MBD3A and B (Figure 2.1E). These data suggest that WDR5 interacts specifically with this smallest and least characterized isoform of MBD3.

To further investigate the composition of the MBD3C/NuRD complex, we generated an ESC line expressing *Mbd3c-H3F* from a viral vector, such that only the MBD3C isoform is epitope-tagged. To this end, we first performed 5’ rapid amplification of cDNA ends (5’RACE) to obtain the *Mbd3c* coding sequence. We found that MBD3C is translated from a start codon within intron 2 of the *Mbd3* gene, consistent with a recent report (Santos et al., 2014). Thus, MBD3C lacks the entire MBD and contains a unique 50–amino acid N-terminus (Figure 2.2A). MBD3C-H3F complexes were affinity purified (Figure 2.2B) and analyzed by LC-MS/MS. As expected, WDR5 co-purifies with MBD3C-H3F, but not MBD3A-H3F

(Figures 2.2C and 2.2D; Table 2.2). Importantly, we found that WDR5 interaction was disrupted by deletion of the unique MBD3C N-terminus (MBD3CΔN; Figures 2.2C and 2.2D), demonstrating that this domain is necessary for WDR5 binding. Co-IP experiments confirmed these results (Figure 2.2E). Furthermore, we observed that MBD3C-H3F, MBD3A-H3F, and MBD3CΔN-H3F all co-purify with the canonical NuRD subunits (Figure 2.2C; 2.2D; Table 2.2). Together with data showing that WDR5 also co-purifies with NuRD subunits (Figure 2.1D; (Bode et al., 2016) and that MBD3C co-fractionates exclusively with NuRD subunit MTA1 (Figure 2.2F) these data demonstrate that MBD3C assembles into a canonical NuRD complex that also includes WDR5. Although the MBD3 MBD was previously shown to directly interact with NuRD subunits HDAC1 and MTA2 *in vitro* (Saito and Ishikawa, 2002), our findings suggest that HDAC1 and MTA2 can also associate with the NuRD complex by MBD-independent mechanisms *in vivo*. Additionally, while the unique MBD3C N-terminus is required for interaction with WDR5, it is dispensable for interaction with the other known NuRD subunits (Figure 2.2D and Table 2.2).

Figure 2.1 The NuRD subunit MBD3C co-purifies with WDR5

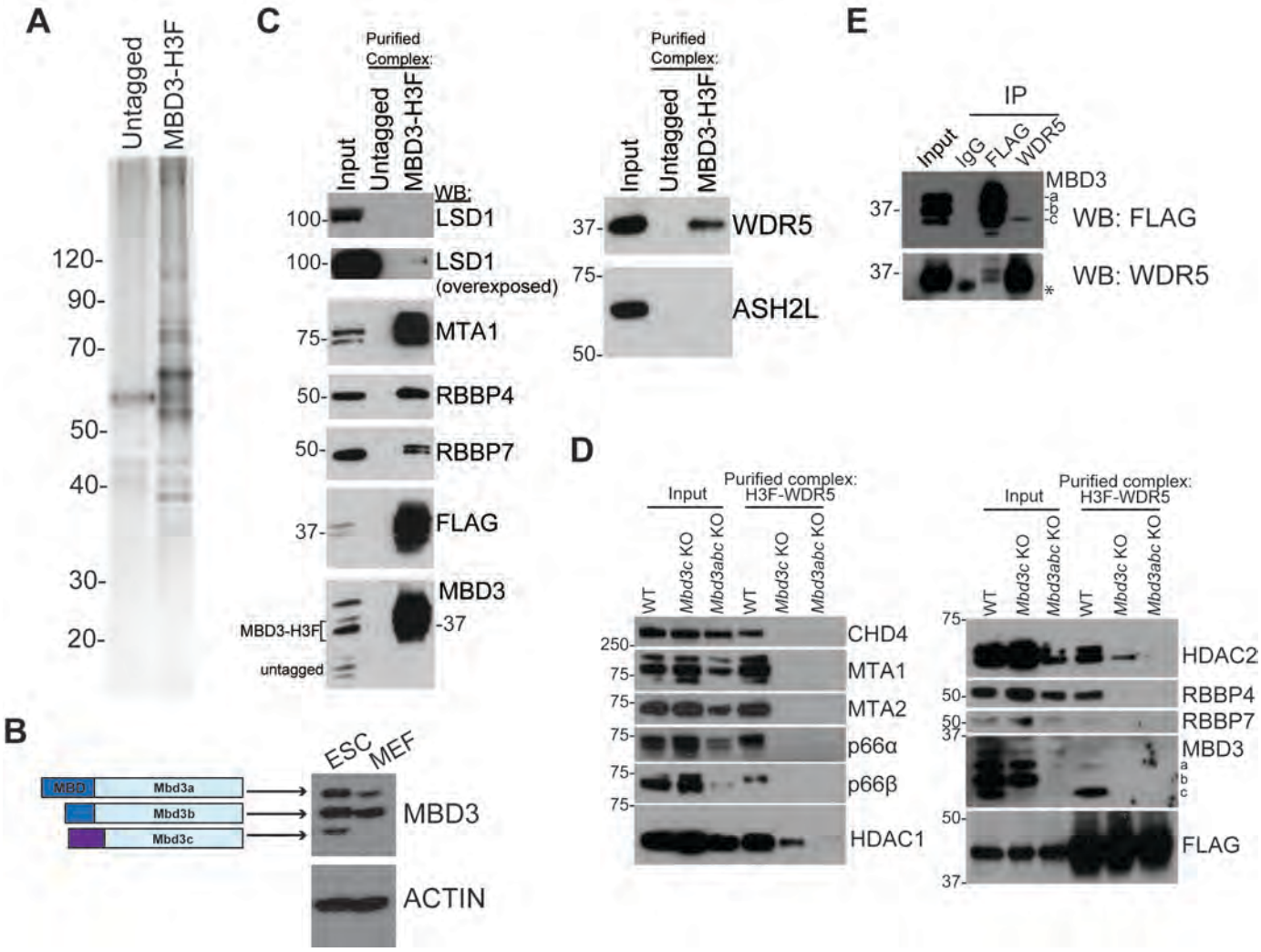


Figure 2.1: The NuRD subunit MBD3 co-purifies with WDR5
(A) Silver stain of MBD3-H3F complex. **(B)** Left: Schematic of the three MBD3 isoforms in ESCs. MBD3A contains a methyl-binding domain (MBD) that is truncated in MBD3B and absent in MBD3C. The purple box signifies the unique Mbd3 N-terminal 50 amino acids. Right: Western blot of Mbd3 in ESCs and MEFs. Actin serves as a loading control. **(C)** Western blot of purified complex from (A) showing interaction of MBD3 with NuRD subunits (left) and with WDR5 or MLL subunit ASH2L (right). **(D)** Western blots of NuRD subunits from purified H3F-WDR5 complexes in WT, *Mbd3c* KO or *Mbd3abc* KO ESCs. **(E)** Western blots for MBD3-H3F or WDR5 upon IP of each. Asterisks (*), IgG.

Table 2.1: Proteins identified in LC-MS/MS of endogenous *Mbd3-H3F* ESCs

Protein Name	Description	# of peptides
Chd4	Chromodomain-helicase-DNA-binding protein 4	252
Mta2	Metastasis-associated protein MTA2	324
Mta3	Metastasis-associated protein MTA3	250
Mta1	Metastasis-associated protein MTA1	206
Gatad2a/p66-alpha	Transcriptional repressor p66 alpha	351
Gatad2b/p66-beta	Transcriptional repressor p66-beta	341
Mthfd1l	Monofunctional C1-tetrahydrofolate synthase, mitochondrial	82
Lima1	LIM domain and actin-binding protein 1	102
Actb	Actin, cytoplasmic 1	93
Actg1	Actin, cytoplasmic 2	93
Acta2	Actin, aortic smooth muscle	61
Actg2	Actin, gamma-enteric smooth muscle	61
Actc1	Actin, alpha cardiac muscle 1	61
Acta1	Actin, alpha skeletal muscle	61
Hspa8	Heat shock cognate 71 kDa protein	56
Hspa5	78 kDa glucose-regulated protein	18
Hspa9	Stress-70 protein, mitochondrial	12
Mbd3	Methyl-CpG-binding domain protein 3	162
Mbd2	Methyl-CpG-binding domain protein 2	28
Tubb2b	Tubulin beta-2B chain	67
Tubb5	Tubulin beta-5 chain	77
Tubb4b	Tubulin beta-4B chain	73
Tubb4a	Tubulin beta-4A chain	54
Tubb3	Tubulin beta-3 chain	47
Tubb6	Tubulin beta-6 chain	33
Stk38	Serine/threonine-protein kinase 38	116
Stk38l	Serine/threonine-protein kinase 38-like	16
Trim28	Transcription intermediary factor 1-beta	37
Myo1c	Unconventional myosin-Ic	34
Myo1f	Unconventional myosin-If	7
Hsp90ab1	Heat shock protein HSP 90-beta	32
Hsp90aa1	Heat shock protein HSP 90-alpha	31
Hsp90b1	Endoplasmin	4
Hdac1	Histone deacetylase 1	132
Hdac2	Histone deacetylase 2	99
Tuba1b	Tubulin alpha-1B chain	62
Tuba1c	Tubulin alpha-1C chain	61
Tuba1a	Tubulin alpha-1A chain	59

Tuba4a	Tubulin alpha-4A chain	49
Tuba3a	Tubulin alpha-3 chain	45
Tuba8a	Tubulin alpha-8 chain	31
Rbbp4/RbAp48	Retinoblastoma binding protein 4	120
Rbbp7/RbAp46	Retinoblastoma binding protein 7	91
Prpf3	U4/U6 small nuclear ribonucleoprotein Prp3	31
Sall4	Sal-like protein 4	54
Sall1	Sal-like protein 1	26
Sall3	Sal-like protein 3	12
Actn4	Alpha-actinin-4	23
Actn1	Alpha-actinin-1	13
Actn2	Alpha-actinin-2	4
Myh10	Myosin-10	21
Myh9	Myosin-9	6
L1td1	LINE-1 type transposase domain-containing protein 1	13
Tmod3	Tropomodulin-3	23
Mccc1	Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial	17
Ppm1b	Protein phosphatase 1B	19
Msh2	DNA mismatch repair protein Msh2	15
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase	16
Atp5a1	ATP synthase subunit alpha, mitochondrial	10
Myo1e	Unconventional myosin-Ie	20
Myo1f	Unconventional myosin-If	7
Setx	Probable helicase senataxin	20
Ruvbl1	RuvB-like 1	9
Wdr5	WD repeat-containing protein 5	15
Tcp1	T-complex protein 1 subunit alpha	9
Hdac6	Histone deacetylase 6	11
Cct8	T-complex protein 1 subunit theta	9
Rest	RE1-silencing transcription factor	12
Kpna3	Importin subunit alpha-3	11
Kpna4	Importin subunit alpha-4	7
Mccc2	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	9
Cct5	T-complex protein 1 subunit epsilon	7
Cct6a	T-complex protein 1 subunit zeta	8
Kpna2	Importin subunit alpha-2	6
Hnrnph1	Heterogeneous nuclear ribonucleoprotein H	6
Prpf4	U4/U6 small nuclear ribonucleoprotein Prp4	9
Ruvbl2	RuvB-like 2	5
Thrap3	Thyroid hormone receptor-associated protein 3	12
Myo1b	Unconventional myosin-Ib	6

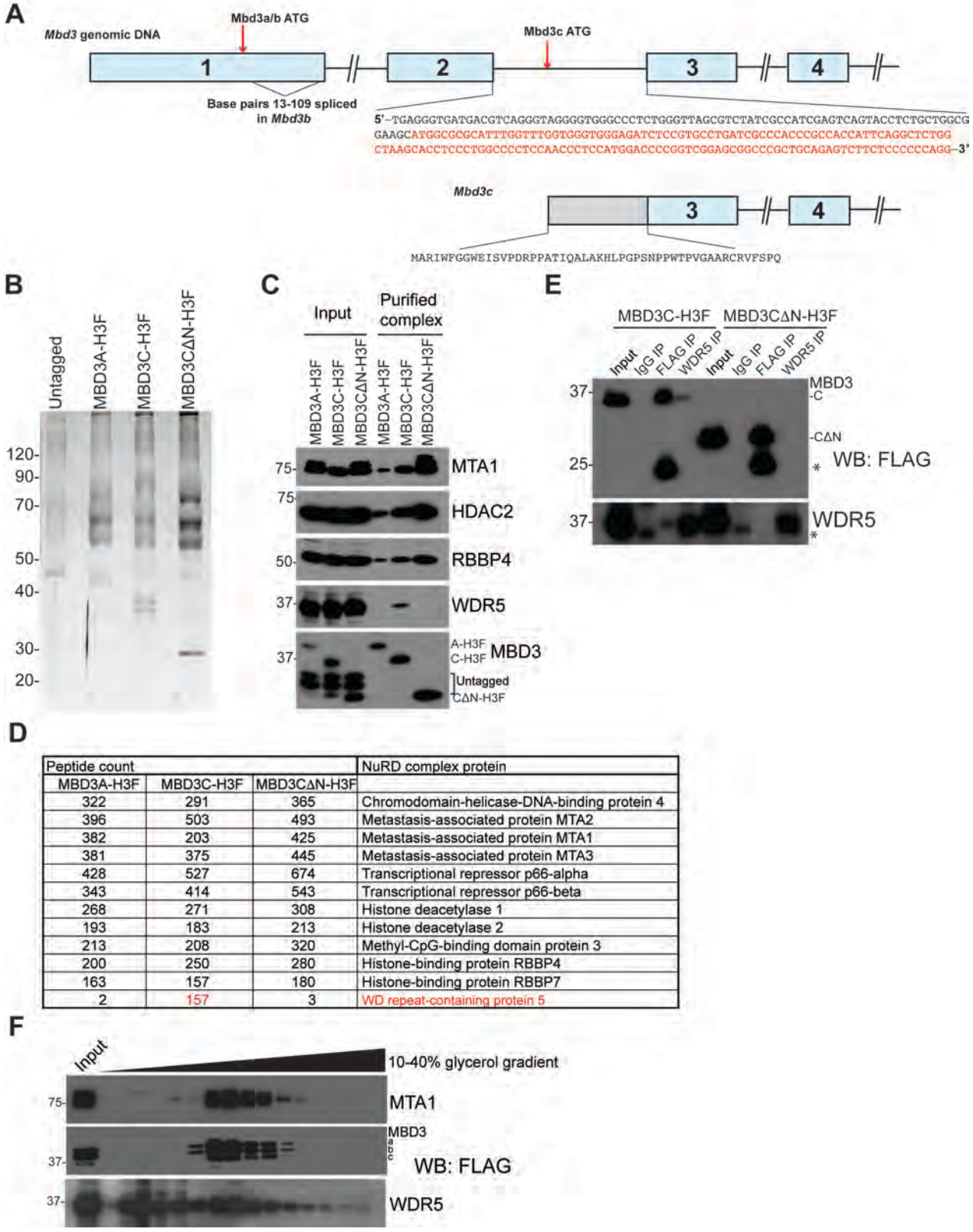
Ogt	UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit	6
Kpnb1	Importin subunit beta-1	6
Pabpc1	Polyadenylate-binding protein 1	6
Gart	Trifunctional purine biosynthetic protein adenosine-3	5
Cct3	T-complex protein 1 subunit gamma	7
Hspd1	60 kDa heat shock protein, mitochondrial	5
Capzb	F-actin-capping protein subunit beta	5
Cct2	T-complex protein 1 subunit beta	4
Mcm3	DNA replication licensing factor MCM3	4
Hnrnpu	Heterogeneous nuclear ribonucleoprotein U	4
Ppp2r1a	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	5
Adsl	Adenylosuccinate lyase	3
D1Pas1	Putative ATP-dependent RNA helicase Pl10	3
Ivns1abp	Influenza virus NS1A-binding protein homolog	3
Myo1d	Unconventional myosin-Id	5
Capza2	F-actin-capping protein subunit alpha-2	4
Capza1	F-actin-capping protein subunit alpha-1	4
Mapk1	Mitogen-activated protein kinase 1	5
Eef2	Elongation factor 2	5
Col1a1	Collagen alpha-1(I) chain	3
Canx	Calnexin	3
Ddx5	Probable ATP-dependent RNA helicase DDX5	3
Bclaf1	Bcl-2-associated transcription factor 1	7
Vim	Vimentin	3
Ppp2r2a	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	3
Ipo5	Importin-5	3
Mcm5	DNA replication licensing factor MCM5	2
Eef1a1	Elongation factor 1-alpha 1	6
Rplp0	60S acidic ribosomal protein P0	2
Psmd1	26S proteasome non-ATPase regulatory subunit 1	8
Prmt5	Protein arginine N-methyltransferase 5	3
Rps3a	40S ribosomal protein S3a	3
Tra2b	Transformer-2 protein homolog beta	1
Eif3b	Eukaryotic translation initiation factor 3 subunit B	7
Cse1l	Exportin-2	2
Caprin1	Caprin-1	2
Gfap	Glial fibrillary acidic protein	11
Top2a	DNA topoisomerase 2-alpha	1
Nup107	Nuclear pore complex protein 107	2
Eif3c	Eukaryotic translation initiation factor 3 subunit C	2

Kpna2	Importin subunit alpha-1	1
Dhx15	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	1
Otud4	OTU domain-containing protein 4	1
Nedd4	E3 ubiquitin-protein ligase NEDD4	1
Akap8	A-kinase anchor protein 8	2
Smarcc1	SWI/SNF complex subunit SMARCC1	1
Glud1	Glutamate dehydrogenase 1, mitochondrial	2
Psmd11	26S proteasome non-ATPase regulatory subunit 11	1
Hnrnpc	Heterogeneous nuclear ribonucleoproteins C1/C2	2
Ddx39A	ATP-dependent RNA helicase DDX39A	1
Prpsap2	Phosphoribosyl pyrophosphate synthase-associated protein 2	1
Rps2	40S ribosomal protein S2	1
Tufm	Elongation factor Tu, mitochondrial	1
Psmc3	26S protease regulatory subunit 6A	3
G3bp1	Ras GTPase-activating protein-binding protein 1	3
Mcm7	DNA replication licensing factor MCM7	2
Psmd2	26S proteasome non-ATPase regulatory subunit 2	3
Tdh	L-threonine 3-dehydrogenase, mitochondrial	1
Hnrnpl	Heterogeneous nuclear ribonucleoprotein L	1
Rbm10	RNA-binding protein 10	2
Jak1	Tyrosine-protein kinase JAK1	1
Hist1h4	Histone H4	1
Npm1	Nucleophosmin	1
Matr3	Matrin-3	1
Hnrnpm	Heterogeneous nuclear ribonucleoprotein M	1
H2afz	Histone H2A.Z	1
Dnajb11	DnaJ homolog subfamily B member 11	1
Rpsa	40S ribosomal protein SA	1
Tubg1	Tubulin gamma-1 chain	1
Asap2	Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 2	15
Hnrnpk	Heterogeneous nuclear ribonucleoprotein K	1
Psmd13	26S proteasome non-ATPase regulatory subunit 13	1
H3f3c	Histone H3.3C	1

Figure 2.2: MBD3c contains a unique 50-amino acid N-terminus that is required for interaction with WDR5

(A) Schematic of the N-terminal *Mbd3c* DNA and MBD3C protein sequences determined by 5'RACE. Exons are indicated by numbered blue boxes, introns by connecting black lines. The sequence of *Mbd3* intron 2 is shown, with the *Mbd3c* N-terminus in red. The amino acid sequence of the MBD3C N-terminus derived from intron 2 is shown (represented by the gray box in the *Mbd3c* gene). **(B)** Silver stain of MBD3 complex expressing individually H3F-tagged MBD3 isoforms. **(C)** Western blot of purified complexes from (B) showing interaction with WDR5 and NuRD subunits. **(D)** Table of peptide counts from mass spec analysis of individually FLAG-tagged MBD3 isoforms. **(E)** WDR5 IP or FLAG IP from individually tagged MBD3C-H3F and MBD3CΔN-H3F ESCs. **(F)** Glycerol gradient analysis of *Mbd3-H3F* nuclear extracts.

Figure 2.2 MBD3C contains a unique 50-amino acid N-terminus that is required for interaction with WDR5



Peptide count			NuRD complex protein
MBD3A-H3F	MBD3C-H3F	MBD3CΔN-H3F	
322	291	365	Chromodomain-helicase-DNA-binding protein 4
396	503	493	Metastasis-associated protein MTA2
382	203	425	Metastasis-associated protein MTA1
381	375	445	Metastasis-associated protein MTA3
428	527	674	Transcriptional repressor p66-alpha
343	414	543	Transcriptional repressor p66-beta
268	271	308	Histone deacetylase 1
193	183	213	Histone deacetylase 2
213	208	320	Methyl-CpG-binding domain protein 3
200	250	280	Histone-binding protein RBBP4
163	157	180	Histone-binding protein RBBP7
2	157	3	WD repeat-containing protein 5

Table 2.2: Proteins identified in LC-MS/MS of *Mbd3a-H3F*, *Mbd3c-H3F*, and *Mbd3cΔN-H3F* ESCs

Protein Name	Description	# of peptides Mbd3a-H3F	# of peptides Mbd3c-H3F	# of peptides Mbd3cΔN-H3F
Chd4	Chromodomain-helicase-DNA-binding protein 4	322	291	365
Setx	Probable helicase senataxin	135	100	106
Dnaja1	DnaJ homolog subfamily A member 1	9	15	28
Mta2	Metastasis-associated protein MTA2	396	503	493
Mta1	Metastasis-associated protein MTA1	382	203	425
Mta3	Metastasis-associated protein MTA3	381	375	445
Mthfd1l	Monofunctional C1-tetrahydrofolate synthase, mitochondrial	130	286	218
Gatad2a/p66-alpha	Transcriptional repressor p66 alpha	428	527	674
Gatad2b/p66-beta	Transcriptional repressor p66-beta	343	414	543
Hspa8	Heat shock cognate 71 kDa protein	114	169	181
Hspa9	Stress-70 protein, mitochondrial	34	32	68
Hspa5	78 kDa glucose-regulated protein	51	61	64
Stk38	Serine/threonine-protein kinase 38	313	371	306
Stk38l	Serine/threonine-protein kinase 38-like	67	115	92
Prpf3	U4/U6 small nuclear ribonucleoprotein Prp3	192	30	108
Hdac1	Histone deacetylase 1	268	271	308
Hdac2	Histone deacetylase 2	193	183	213
Actb	Actin, cytoplasmic 1	193	263	220
Brg1	Transcription activator BRG1	57	22	64
Mbd3	Methyl-CpG-binding domain protein 3	213	208	320
Hsp90ab1	Heat shock protein HSP 90-beta	81	53	76
Hsp90aa1	Heat shock protein HSP 90-alpha	67	44	64
Vim	Vimentin	28	83	26
Pabpc1	Polyadenylate-binding protein 1	31	46	31
Wdr5	WD repeat-containing protein 5	2	157	3
Ruvbl2	RuvB-like 2	22	51	46
Mccc2	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	24	41	28
Rps3a	40S ribosomal protein S3a	47	49	47
Tuba1b	Tubulin alpha-1B chain	64	162	126
Tcp1	T-complex protein 1 subunit alpha	15	25	19
Mccc1	Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial	20	53	56

Tmod3	Tropomodulin-3	34	64	75
Hnrnpc	Heterogeneous nuclear ribonucleoproteins C1/C2	20	48	44
Mybbp1a	Myb-binding protein 1A	11	42	28
Ncl	Nucleolin	3	39	17
Ruvbl1	RuvB-like 1	16	43	33
Sall4	Sal-like protein 4	9	56	14
Sall1	Sal-like protein 1	8	18	13
Prpf4	U4/U6 small nuclear ribonucleoprotein Prp4	60	15	56
Rpl7	60S ribosomal protein L7	31	49	32
Rbbp4/RbAp48	Retinoblastoma binding protein 4	200	250	280
Rbbp7/RbAp46	Retinoblastoma binding protein 7	163	157	180
Cct7	T-complex protein 1 subunit eta	20	25	22
Rbm10	RNA-binding protein 10		27	19
Tubb5	Tubulin beta-5 chain	61	97	104
Flii	Protein flightless-1 homolog	1	47	3
Rpl7a	60S ribosomal protein L7a	23	53	31
Pkm	Pyruvate kinase isozymes M1/M2	18	31	21
Actl6a	Actin-like protein 6A	30	11	40
Myo1c	Myosin-Ic	22	50	19
Rps4x	40S ribosomal protein S4, X isoform	20	48	53
Rpl6	60S ribosomal protein L6	33	83	37
Prpf31	U4/U6 small nuclear ribonucleoprotein Prp31	25	16	50
Rpl4	60S ribosomal protein L4	17	65	29
Jak1	Tyrosine-protein kinase JAK1	20	32	39
Rps3a	40S ribosomal protein S3	23	31	40
Eif4a3	Eukaryotic initiation factor 4A-III	14	22	30
Eif4a1	Eukaryotic initiation factor 4A-I	3	5	14
Cct8	T-complex protein 1 subunit theta	13	17	20
Hdac6	Histone deacetylase 6	26	47	34
Atp5a1	ATP synthase subunit alpha, mitochondrial	20	22	16
Csnk2a2	Casein kinase II subunit alpha	7	20	25
Lrrfip2	Leucine-rich repeat flightless-interacting protein 2	3	48	9
Cct2	T-complex protein 1 subunit beta	22	18	23
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase	42	64	89
Psmc5	26S protease regulatory subunit 8	5	13	23
Psmc1	26S protease regulatory subunit 4		9	2
Trim28	Transcription intermediary factor 1-beta	3	39	25
Cct6a	T-complex protein 1 subunit	21	31	23

	zeta			
Thrap3	Thyroid hormone receptor-associated protein 3	48	30	52
Bclaf1	Bcl-2-associated transcription factor 1	21	8	33
Coro1c	Coronin-1C		48	
Rpl3	60S ribosomal protein L3	25	74	45
Atp5b	ATP synthase subunit beta, mitochondrial	16	14	11
Hnrnpf	Heterogeneous nuclear ribonucleoprotein F	11	52	38
Hnrnph1	Heterogeneous nuclear ribonucleoprotein H	14	31	28
Hnrnph2	Heterogeneous nuclear ribonucleoprotein H2	12	20	18
Actn4	Alpha-actinin-4	1	40	16
Ddx3x	ATP-dependent RNA helicase DDX3X	5	21	22
Ddx5	Probable ATP-dependent RNA helicase DDX5		10	8
Ddx17	Probable ATP-dependent RNA helicase DDX17		4	7
Phb2	Prohibitin-2	22	5	30
Spin1	Spindlin-1	26	30	29
Ddb1	DNA damage-binding protein 1		19	6
Tra2b	Transformer-2 protein homolog beta	34	17	38
Tra2a	Transformer-2 protein homolog alpha	18	9	17
Tufm	Elongation factor Tu, mitochondrial	2	10	24
Psmc4	26S protease regulatory subunit 6B		7	16
Rps2	40S ribosomal protein S2	20	48	58
Ppm1b	Protein phosphatase 1B	20	39	51
Ppm1a	Protein phosphatase 1A	5	8	8
Rplp0	60S acidic ribosomal protein P0	16	48	28
Psmd3	26S proteasome non-ATPase regulatory subunit 3	4	17	9
Capza1	F-actin-capping protein subunit alpha-1	17	19	15
Capza2	F-actin-capping protein subunit alpha-2	15	15	12
Dars	Aspartyl-tRNA synthetase, cytoplasmic	8	14	7
Cct3	T-complex protein 1 subunit gamma	14	21	19
Ybx1	Nuclease-sensitive element-binding protein 1	6	16	15
Ybx3	DNA-binding protein A	2	13	7
Hspd1	60 kDa heat shock protein, mitochondrial	6	15	14
Lmnb1	Lamin-B1		18	3
Psmd11	26S proteasome non-ATPase regulatory subunit 11	3	16	22
Wdr1	WD repeat-containing protein 1		20	
Ssb	Lupus La protein homolog	8	17	16

Myh10	Myosin-10	5	17	15
Myh9	Myosin-9	9	6	8
Cct5	T-complex protein 1 subunit epsilon	4	14	16
Eftud2	116 kDa U5 small nuclear ribonucleoprotein component		15	9
Kpna3	Importin subunit alpha-3	22	34	32
Eef1a1	Elongation factor 1-alpha 1	30	46	62
Cct4	T-complex protein 1 subunit delta	17	20	15
Prpf19	Pre-mRNA-processing factor 19	10	17	12
Serbp1	Plasminogen activator inhibitor 1 RNA-binding protein	4	13	8
Eif3e	Eukaryotic translation initiation factor 3 subunit E	2	9	10
Sf3b3	Splicing factor 3B subunit 3		17	2
Hnrnpu	Heterogeneous nuclear ribonucleoprotein U	8	38	18
Dnaja2	DnaJ homolog subfamily A member 2	5	9	20
Lima1	LIM domain and actin-binding protein 1		32	3
Dnajb11	DnaJ homolog subfamily B member 11	5	7	11
Dhx15	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	5	16	10
Ivns1abp	Influenza virus NS1A-binding protein homolog	10	21	14
G3bp2	Ras GTPase-activating protein-binding protein 2	7	19	13
G3bp1	Ras GTPase-activating protein-binding protein 1	4	8	8
Hnrnpk	Heterogeneous nuclear ribonucleoprotein K	2	23	8
Hnrnpl	Heterogeneous nuclear ribonucleoprotein L	3	11	10
Sptan1	Spectrin alpha chain, brain		16	2
Igf2bp1	Insulin-like growth factor 2 mRNA-binding protein 1	1	20	3
Rpl8	60S ribosomal protein L8	24	55	27
Hist1h1c	Histone H1.2	22	8	20
Aifm1	Apoptosis-inducing factor 1, mitochondrial	2	11	7
Nudc	Nuclear migration protein nudC	6	14	18
Prmt5	Protein arginine N-methyltransferase 5	8	13	14
Eif3f	Eukaryotic translation initiation factor 3 subunit F		8	10
Cdc5l	Cell division cycle 5-related protein		11	3
Cdk1	Cyclin-dependent kinase 1	5	16	8
Ilf2	Interleukin enhancer-binding factor 2	7	12	14
Elavl1	ELAV-like protein 1	12	13	21
Hnrnpm	Heterogeneous nuclear ribonucleoprotein M	1	18	3
Rps6	40S ribosomal protein S6	5	30	17

Psmd1	26S proteasome non-ATPase regulatory subunit 1	2	14	9
Capzb	F-actin-capping protein subunit beta	12	12	15
Psmd2	26S proteasome non-ATPase regulatory subunit 2	17	8	12
Polr1c	DNA-directed RNA polymerases I and III subunit RPAC1	2	7	8
Slc25a5	ADP/ATP translocase 2	7	31	28
Twf1	Twinfilin-1		12	4
Atp5c1	ATP synthase subunit gamma, mitochondrial	8	10	12
Puf60	Poly(U)-binding-splicing factor PUF60		11	3
Kpna2	Importin subunit alpha-2	11	17	16
Dnaja3	DnaJ homolog subfamily A member 3, mitochondrial	2	6	11
Rpsa	40S ribosomal protein SA	3	6	9
Sptbn1	Spectrin beta chain, brain 1		8	
Kpnb1	Importin subunit beta-1	4	9	11
Dnajb6	DnaJ homolog subfamily B member 6	4	6	10
Eif3i	Eukaryotic translation initiation factor 3 subunit I	5	13	11
Ssrp1	FACT complex subunit SSRP1	9	4	9
Sf3b1	Splicing factor 3B subunit 1	1	14	2
Otud4	OTU domain-containing protein 4	8	4	
Srsf7	Serine/arginine-rich splicing factor 7	18	11	30
Ppp2r1a	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform		9	
Mov10	Putative helicase MOV-10		9	1
Pgam5	Serine/threonine-protein phosphatase PGAM5, mitochondrial	4	13	11
Rnps1	RNA-binding protein with serine-rich domain 1	6		11
Eif3c	Eukaryotic translation initiation factor 3 subunit C		8	9
Ilf3	Interleukin enhancer-binding factor 3		9	8
Myo1b	Myosin-Ib	4	8	
Tdh	L-threonine 3-dehydrogenase, mitochondrial	8	2	22
Nono	Non-POU domain-containing octamer-binding protein	3	13	6
Sfpq	Splicing factor, proline- and glutamine-rich		15	12
Npm1	Nucleophosmin	2	13	10
Sun2	SUN domain-containing protein 2	9	3	6
Rpl11	60S ribosomal protein L11	14	15	27
Mga	MAX gene-associated protein		6	
U2af2	Splicing factor U2AF 65 kDa subunit		13	6

Eif3b	Eukaryotic translation initiation factor 3 subunit B		12	7
C1qbp	Complement component 1 Q subcomponent-binding protein, mitochondrial	2		12
Prdx1	Peroxiredoxin-1	17	29	42
Gsn	Gelsolin	11	8	4
Rps8	40S ribosomal protein S8	13	49	23
Psmc3	26S protease regulatory subunit 6A	2	6	11
Rpl23	60S ribosomal protein L23	4	26	23
Ap2m1	AP-2 complex subunit mu			8
Ctbp2	C-terminal-binding protein 2		7	11
Rpl13	60S ribosomal protein L13	3	44	14
Fbl	rRNA 2'-O-methyltransferase fibrillarin	8	9	14
Srsf1	Serine/arginine-rich splicing factor 1	9	8	11
Wdr77	Methylosome protein 50	3	4	11
Rps9	40S ribosomal protein S9	13	30	24
Rpl14	60S ribosomal protein L14	1	25	5
Stmn2	Stathmin-2	4	14	16
Rps23	40S ribosomal protein S23	7	26	20
Rps26	40S ribosomal protein S26		19	7
Rpl21	60S ribosomal protein L21		22	9
Rpl18	60S ribosomal protein L18		13	7
Ubb	Polyubiquitin-B	5	16	12
Rpl27a	60S ribosomal protein L27a	6	12	5
Rps11	40S ribosomal protein S11	12	19	21
Alb	Serum albumin	11	13	12
Hist1h3a	Histone H3.1	6	11	10
Rpl34	60S ribosomal protein L34		8	
B3galt5	Beta-1,3-galactosyltransferase 5		10	

The WDR5 histone H3 binding pocket is required for interaction with MBD3C

To gain insight into the functions of the MBD3C-WDR5 interaction, we dissected the domains within MBD3C and WDR5 important for their interaction. WDR5 contains two binding surfaces on opposite sides of the protein, one that binds the histone H3 N-terminal tail or the SET/MLL complex subunit MLL1, and another that binds both the SET/MLL subunit RBBP5 (Figure 2.3A; (Avdic et al., 2011; Odho et al., 2010; Patel et al., 2008; Song and Kingston, 2008) and long noncoding RNAs (Yang et al., 2014). To test whether MBD3C interacts with either WDR5 binding surface, we performed co-IPs in 293T cells co-transfected with vectors expressing MBD3C-H3F and FLAG-tagged point mutants from both binding surfaces of WDR5: D107A on the H3K4/MLL1 binding surface and F266A, K250A, and R181A on the RBBP5/RNA binding surface (Yang et al., 2014). We found that the F266A, K250A, and R181A mutants of WDR5 co-IP with antibodies recognizing endogenous MBD3 (Figure 2.3B). In contrast, the D107A mutant was absent from MBD3 immunoprecipitates, suggesting that MBD3C binds near or within the WDR5 H3K4/MLL1 binding pocket.

To extend these findings, we generated a series of truncation mutants of the 50–amino acid MBD3C N-terminus to pinpoint the residues necessary for binding (Figure 2.3C). Deletion of the first 40 amino acids of H3F-tagged MBD3C did not disrupt the interaction with V5-tagged WDR5 (Figure 2.3D). However, upon deletion of amino acids 41-50 of MBD3C, interaction with WDR5 was completely lost (Figure 2.3E), as we observed for MBD3C mutants lacking amino acids 1-50 (Figures 2.2C, E and Figure 2.3D). Furthermore, an N-terminal fusion

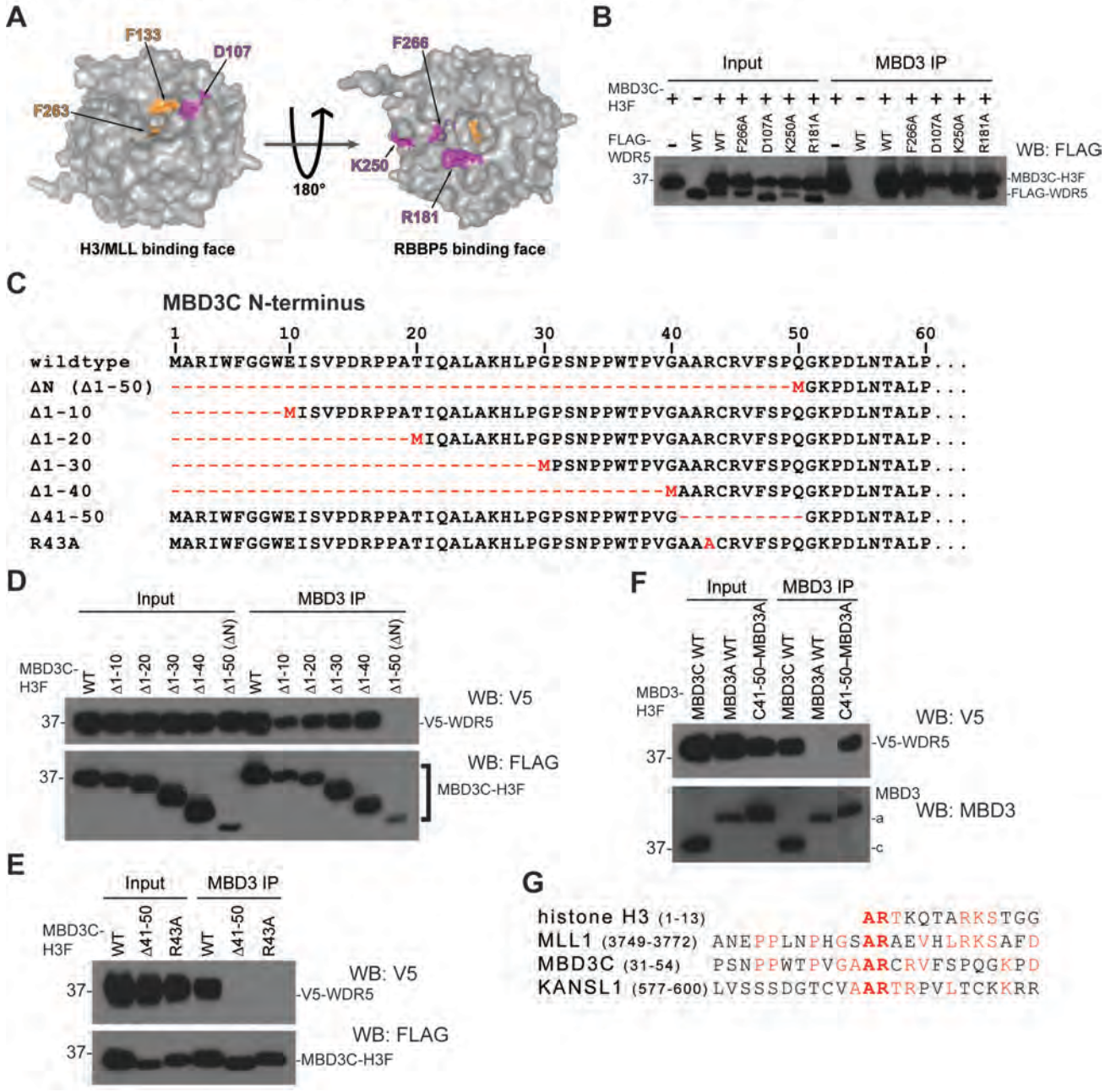
of amino acids 41-50 to the MBD3A isoform was sufficient to allow MBD3A to bind WDR5 (Figure 2.3F). These data demonstrate that amino acids 41-50 of MBD3C mediate WDR5 binding (see also Figure 2.1D).

MLL1, KANSL1, and histone H3 all bind the same domain on WDR5 via a two amino acid alanine-arginine (AR) motif present on each protein (Couture et al., 2006; Han et al., 2006; Patel et al., 2008; Ruthenburg et al., 2006; Schuetz et al., 2006; Song and Kingston, 2008); (Dias et al., 2014); Figures 2.3A and 2.3G). MBD3C contains an AR dipeptide within its N-terminal WDR5 binding domain (A42-R43; Figure 2.3G), which we hypothesized to be necessary for WDR5 binding. Confirming our hypothesis, an R43A mutant of MBD3C failed to pull down WDR5 (Figure 2.3E). Together, these data indicate that MBD3C, histone H3, and MLL1 use a common motif to bind the same surface of WDR5. We observed slightly reduced MBD3C protein levels in cells expressing mutants of MBD3C or WDR5 that disrupt MBD3C–WDR5 binding (Figures 2.3D and 2.3E), raising the possibility that interaction with WDR5 plays a role in stabilizing MBD3C.

Figure 2.3: The WDR5 histone H3 binding pocket is required to bind MBD3C

(A) PyMol depiction of WDR5 crystal structure (PDB:2GNQ) showing the H3K4/MLL1 (left) and RBBP5 (right) binding pockets. Residues individually mutated to alanine (Yang et al., 2014) are shown in magenta. Residues necessary for MLL1 R3761 or histone H3R2 binding are shown in orange. **(B)** Co-IPs with MBD3 antibody from 293T cells co-transfected with expression vectors carrying MBD3C-H3F and indicated FLAG-tagged WDR5 mutants. **(C)** Schematic of *Mbd3c* N-terminal mutant constructs used in (D) and (E). **(D-F)** Co-IPs with MBD3 antibody in 293T cells performed as in (B), using V5-tagged WDR5 constructs and H3F-tagged MBD3 constructs. For “c41-50−MBD3A”, MBD3C amino acids 41-50 were fused to the N-terminus of the MBD3A isoform. **(G)** Alignment of the MBD3C N-terminus with WDR5-binding regions of mouse MLL1, KANSL1, and histone H3.

Figure 2.3 The WDR5 histone H3 binding pocket is required to bind MBD3C



Mbd3c expression is largely restricted to pluripotent stem cells

The *Mbd3c* isoform appears to be highly expressed only in ESCs, as it is absent or weakly expressed in mouse embryonic fibroblasts (MEFs) and all adult tissues tested (Figures 2.1B and 2.4A). To determine when *Mbd3c* expression is lost during differentiation we subjected ESCs to a 10-day embryoid body (EB) differentiation time course. We observed that MBD3C protein was lost between days 4 and 6 of the time course with kinetics similar to loss of OCT4 protein during differentiation (Figure 2.4A).

Next, we tested whether *Mbd3c* expression was restored upon reprogramming of differentiated cells to induced pluripotent stem cells. Primary MEFs were infected with doxycycline (dox)-inducible lentiviruses expressing reprogramming factors OCT4, SOX2, and KLF4 marked with an mCherry reporter (OSK-mCherry), L-MYC, and a lentiviral EOS-EGFP reporter specifically activated in pluripotent cells (Hotta et al., 2009). Infected cells were cultured with dox for 20 days. After an additional 10 days in the absence of dox, cells were imaged and stained for alkaline phosphatase to verify silencing of OSK-mCherry and presence of ESC-like colonies (Figures 2.4C and 2.4D). Lysates for western blots were prepared from expanded iPSC colonies picked at 30 days. We observed that reprogrammed iPSC colonies express *Mbd3c* (Figure 2.4E). These data demonstrate *Mbd3c* expression is restored when somatic cells are reprogrammed to iPSCs.

Finally, to investigate how *Mbd3c* expression might be silenced during differentiation, we performed bisulfite pyrosequencing analysis on the *Mbd3c*

promoter during an EB differentiation time course. The *Mbd3* gene contains a ~350bp CGI which spans exon 2 and part of intron 2, and overlaps with the sequence encoding the MBD3C N-terminal domain (Figure 2.5A). We measured methylation of 11 individual CpGs within the *Mbd3c* promoter and observed a large increase in methylation at all sites over the differentiation time course (Figure 2.5B). Methylation increased most dramatically around day 4, which corresponds to the timing of MBD3C loss during differentiation (Figure 2.4A). Therefore, silencing of *Mbd3c* expression during differentiation is likely due to increased methylation of the *Mbd3c* promoter CGI.

Figure 2.4: *Mbd3c* expression is largely restricted to pluripotent cells

(A) Western blot of MBD3 from lysates of the indicated mouse tissues. Actin blot and Coomassie stain are shown as loading controls. **(B)** Western blots from ESCs differentiated over 10 days. Actin serves as a loading control. **(C)** Representative images of EOS-EGFP positive/OSK-mCherry negative iPSCs at reprogramming Day 30, 10 days after dox removal (left) and an iPSC line derived from a single colony (right). Scale bars = 400μm. **(D)** Representative AP staining of iPSCs at reprogramming Day 31. **(E)** Western blots from primary MEFs reprogrammed to iPSCs.

Figure 2.4 *Mbd3c* expression is largely restricted to pluripotent cells

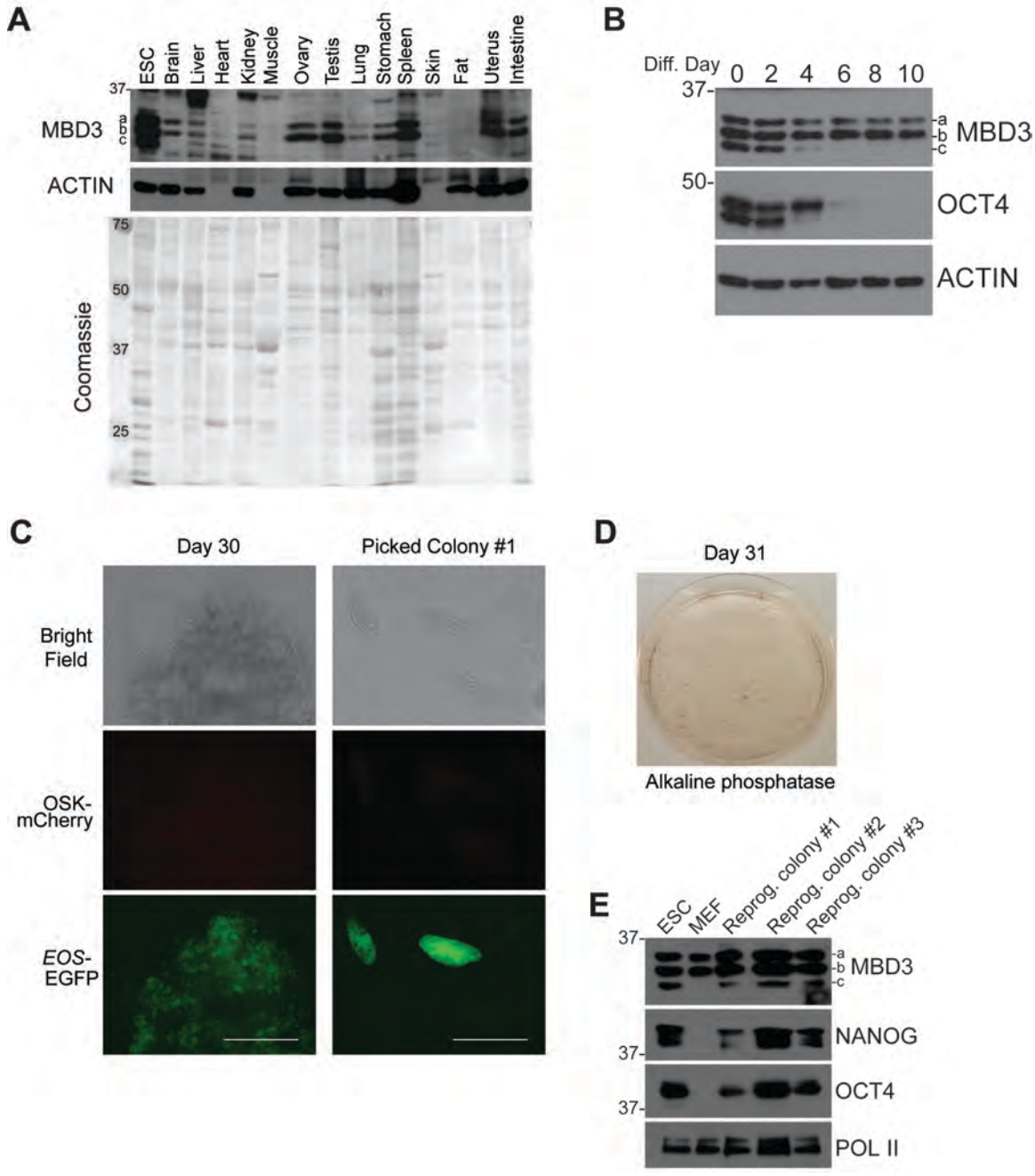


Figure 2.5 Methylation of the *Mbd3c* promoter during differentiation

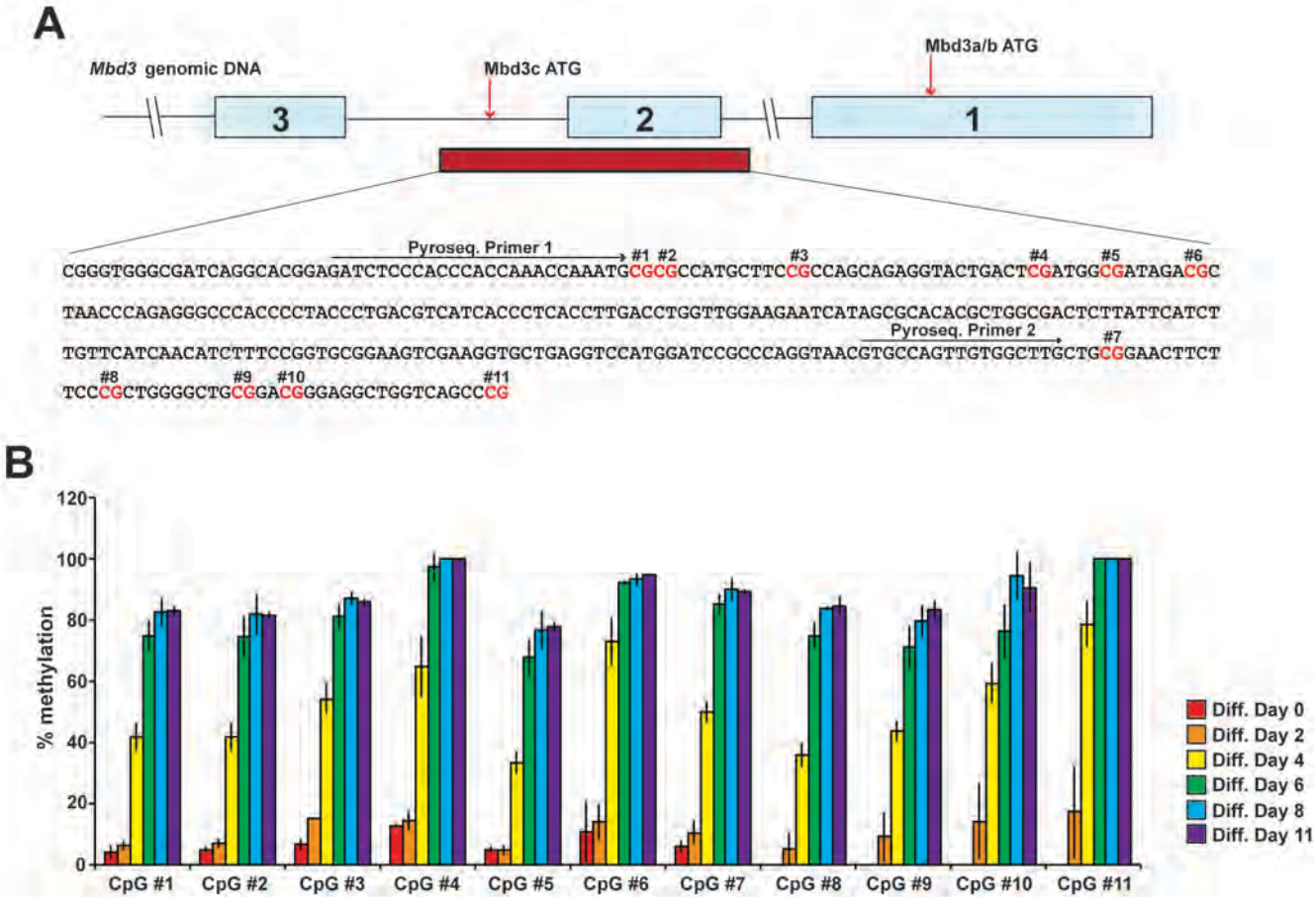


Figure 2.5: Methylation of the *Mbd3c* promoter during differentiation
(A) Schematic of the *Mbd3* gene showing the sequence and location of the *Mbd3c* promoter CpG island (red bar). Light blue boxes indicate exons. CpGs tested for methylation are highlighted in red. **(B)** Pyrosequencing of bisulfite-converted DNA from cells collected at the indicated differentiation timepoints. Error bars represent the standard deviation of three biological replicates.

Mbd3c is dispensable for ESC differentiation

Mbd3 null ESCs exhibit pluripotency defects and are capable of self-renewal in the absence of LIF (Kaji et al., 2006). We therefore wanted to test whether the MBD3C isoform was specifically required for early stages of differentiation. We generated homozygous *Mbd3c*, *Mbd3ab*, and *Mbd3abc* KO ESCs (Figure 2.6A) using CRISPR/Cas9 cleavage and error-prone DNA repair (Cong et al., 2013). We found that *Mbd3c* KO ESCs proliferate similarly to wildtype (WT) cells. In contrast, *Mbd3ab* KO and especially *Mbd3abc* KO ESCs grow more slowly than WT (Figures 2.6B and 2.6C), consistent with previous observations of *Mbd3* null ESCs (Kaji et al., 2006).

Next, we tested the differentiation capacity of each *Mbd3* mutant. Consistent with previous studies (Kaji et al., 2006), we found that ESCs lacking all MBD3 isoforms (*Mbd3abc* KO) maintained expression of both OCT4 and NANOG over 9 days in media without LIF (Figure 2.6D). Unlike *Mbd3abc* KOs, *Mbd3c* KOs did not show a noticeable differentiation defect (Figure 2.6E). ESCs expressing only *Mbd3c* (*Mbd3ab* KO) were largely defective in differentiation (although OCT4 and NANOG levels appeared slightly reduced relative to *Mbd3abc* KO lines). Since *Mbd3c* expression is lost early during differentiation (Figures 2.4B and 2.6D), *Mbd3ab* mutants are functionally equivalent to *Mbd3abc* mutants at mid-to-late differentiation time points, potentially accounting for this phenotype. We next asked whether constitutive overexpression of *Mbd3c* in the absence of MBD3A and MBD3B could allow for normal differentiation. To test this possibility, we replaced the entire *Mbd3* gene with an H3F-tagged

Mbd3c transgene overexpressed from the CAG promoter, which is not silenced during differentiation. Differentiation proceeds normally in these cells (Figure 2.7A, top panel), revealing MBD3C can compensate for MBD3A and MBD3B when it is overexpressed. Unexpectedly, ESCs overexpressing *Mbd3cΔN* were also able to differentiate (Figure 2.7A, bottom panel), in marked contrast with cells expressing *Mbd3cΔN* at endogenous levels (Figure 2.7B). We conclude that *Mbd3c* is not required for differentiation but can substitute for *Mbd3a* and *Mbd3b* when overexpressed.

Figure 2.6: MBD3C is dispensable for ESC differentiation

(A) Western blot of MBD3 expression in *Mbd3* isoform KO ESCs generated through CRISPR/Cas9. **(B-C)** Growth curves for *Mbd3c* KO **(B)** and *Mbd3ab* and *Mbd3abc* KO **(C)** ESC lines, relative to wildtype (WT) ESCs. Error bars represent +/- standard deviation of three replicate experiments performed on each clonal replicate. **(D-E)** Western blots of differentiating *Mbd3ab* and *Mbd3abc* KO cells **(D)** and two clonal *Mbd3c* KO lines **(E)**.

Figure 2.6 MBD3C is dispensable for ESC differentiation

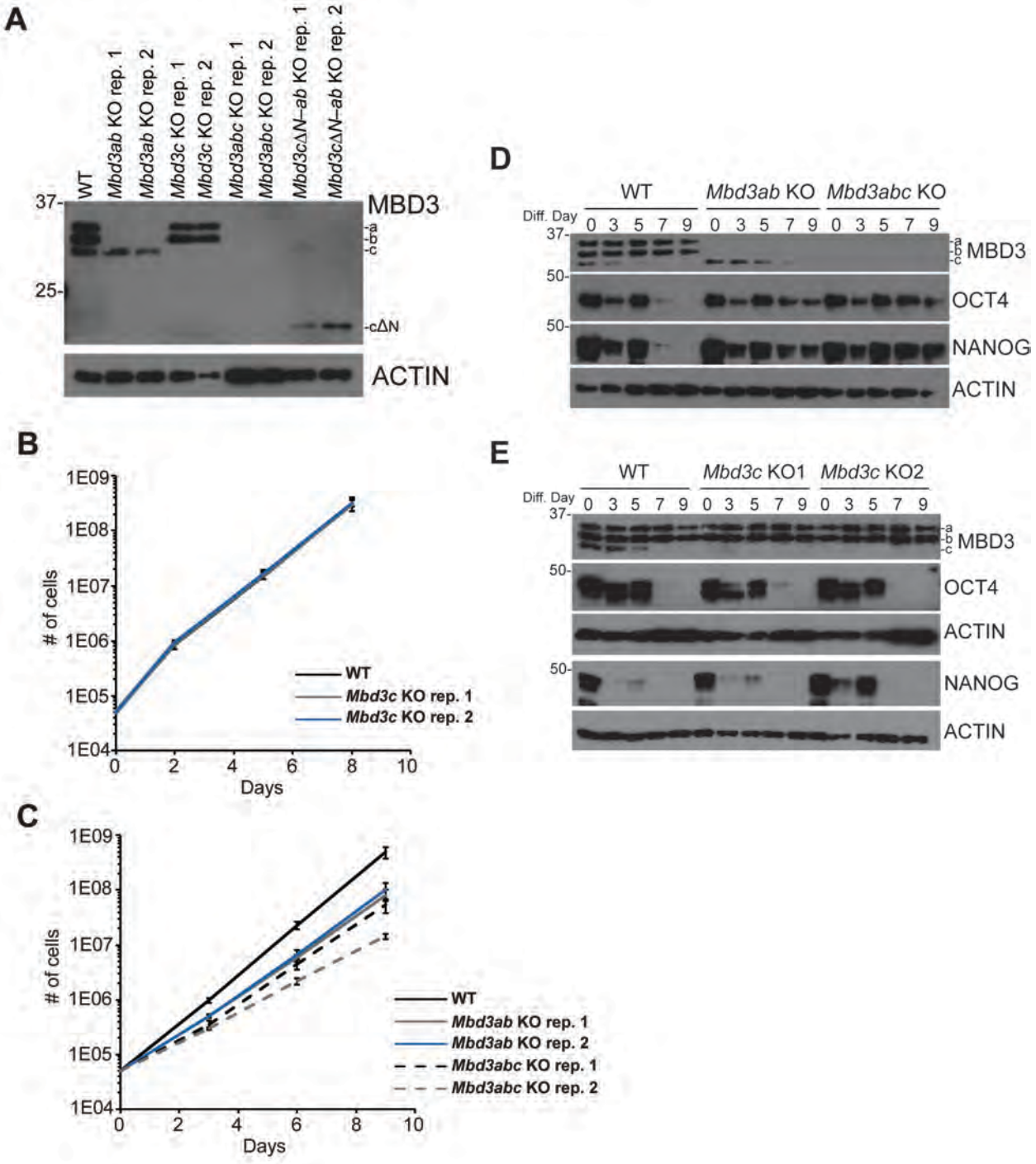


Figure 2.7 MBD3C or MBD3CΔN overexpression is sufficient for ESC differentiation

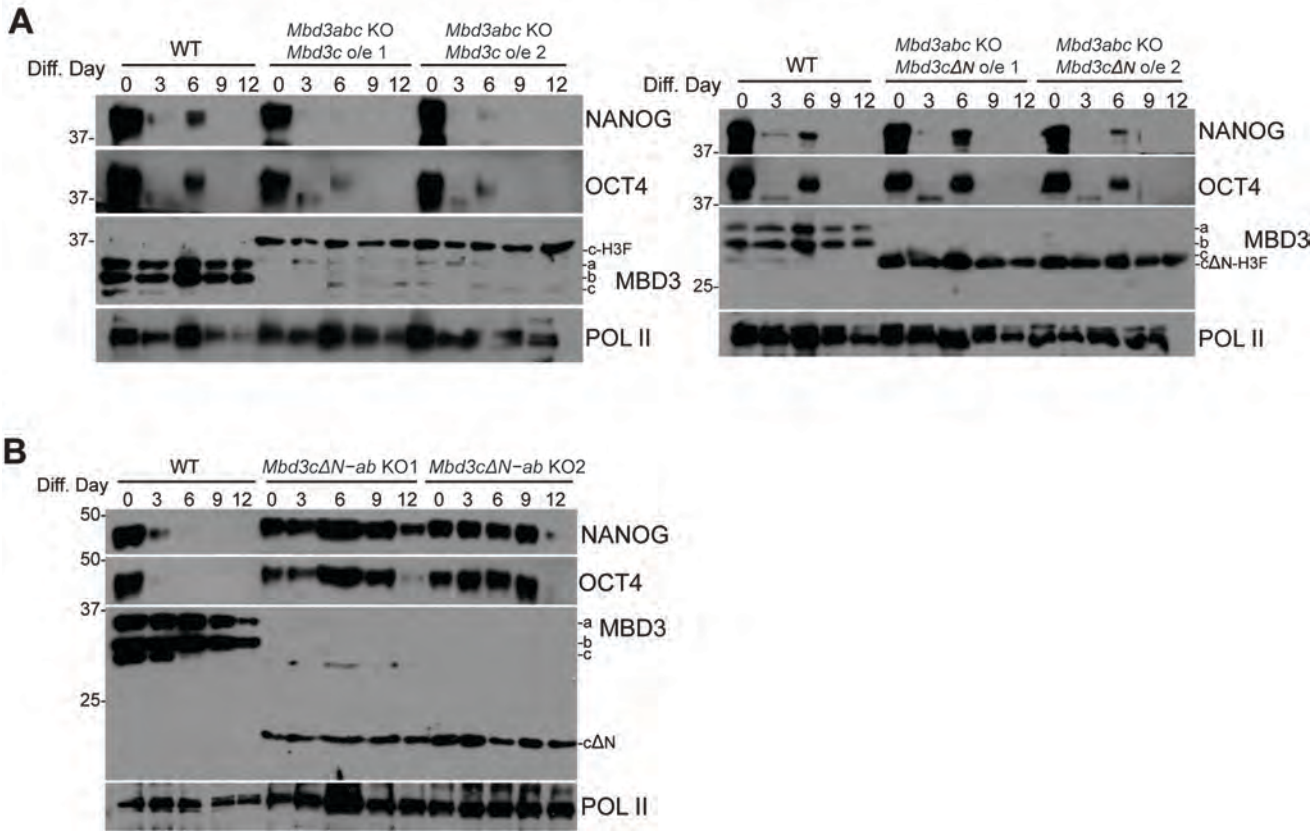


Figure 2.7: MBD3C and MBD3CΔN overexpression is suffcient for ESC differentiation

(A) Western blots of indicated proteins during differentiation of WT, *Mbd3abc* KO ESCs overexpressing *Mbd3c-H3F*, or *Mbd3abc* KO ESCs overexpressing *Mbd3cΔN-H3F.* **(B)** Western blots of indicated proteins during differentiation of WT or *Mbd3cΔN–ab* KO ESCs.

MBD3 isoforms function redundantly in gene regulation

Although it is dispensable for differentiation, MBD3C binds chromatin (Figure 2.8) and could still be important for regulation of a subset of MBD3 target genes independently of the other MBD3 isoforms. To test this possibility, we analyzed the transcriptomes of *Mbd3* isoform-specific mutant ESCs by RNA-seq. Gene expression in *Mbd3c* KO ESCs was largely normal (Figure 2.9A, left panel), with expression of only 38 genes changed more than two-fold compared to WT. Similarly, we observed relatively few genes (258) misregulated in *Mbd3ab* KO ESCs that express only *Mbd3c* (Figure 2.9A, right panel) compared with a much larger number (4,879) misregulated in ESCs where all *Mbd3* isoforms are deleted (Figure 2.9B, left panel and Figure 2.9C). These data suggest that *Mbd3c* can largely compensate for the loss of *Mbd3a* and *Mbd3b* at shared target genes.

To test whether the unique 50–amino acid MBD3C N-terminus (and thus the interaction with WDR5) is important for this compensatory effect, we also performed RNA-seq on ESCs lacking MBD3A and B and the N-terminus of MBD3C (*Mbd3cΔN−ab* KO). In contrast to the relatively few genes misregulated in *Mbd3ab* KO and *Mbd3c* KO cells, we observed 2,431 genes misregulated in *Mbd3cΔN−ab* KO cells, with nearly twice as many genes upregulated as downregulated (1,577 vs. 854 respectively; Figure 2.9B, right panel). The vast majority (~93%) of misregulated genes overlapped with genes misregulated in *Mbd3abc* KO cells (Figure 2.9D), indicating that the MBD3C N-terminus is largely required for MBD3C to compensate for loss of MBD3A and B. However, as

Mbd3abc KO has a stronger phenotype than *Mbd3cΔN−ab* KO, MBD3C/NuRD may also regulate some genes independently of its N-terminal domain. Closer examination of the 2,627 genes misregulated only in *Mbd3abc* KO ESCs (Figure 2.9D) revealed similar, but weaker misregulation in *Mbd3cΔN−ab* KO cells in most cases that fell below our two-fold cutoff. These data suggest that the *Mbd3cΔN* mutation is not a complete null and are consistent with our finding that MBD3CΔN can compensate for loss of MBD3A and MBD3B during ESC differentiation, but only when overexpressed (Figure 2.7A and B).

WDR5 is a component of multiple complexes with key regulatory functions in ESCs (Ang et al., 2011; Chelmicki et al., 2014; Li et al., 2012; Ravens et al., 2014). *Wdr5* KD results in loss of ESC self-renewal (Ang et al., 2011), precluding the use of *Wdr5* KO ESCs to compare the functions of MBD3C and WDR5 in gene regulation. However, to test whether the genes misregulated in *Mbd3c* mutant cells are targets of WDR5 and/or WDR5-associated complexes, we examined published ESC ChIP-seq data for WDR5 and the MSL/NSL subunit MOF (Ang et al., 2011; Li et al., 2012). We observed considerably higher WDR5 binding at the promoter-proximal regions of genes that were misregulated in *Mbd3abc* KO and *Mbd3cΔN−ab* KO cells compared to genes that were unaffected by *Mbd3* mutations (Figure 2.9E), consistent with a regulatory role for WDR5 at MBD3 target genes. However, it is likely that WDR5 regulates some of these genes through mechanisms independent of MBD3C/NuRD, as we also observed higher MOF binding at genes misregulated in *Mbd3abc* KO and *Mbd3cΔN−ab* KO cells (Figure 2.9F). Similarly, we found that WDR5 binding is

enriched at promoter-distal DNase I hypersensitive sites (DHSs) that are co-bound by MBD3 (Figure 2.9G), suggesting that WDR5 and MBD3 co-regulate target gene expression at both promoter-distal enhancers and promoters.

We have identified a variant ESC-specific NuRD complex that includes the histone H3 binding protein WDR5. While WDR5 contributes to H3K4 trimethlyation by the SET/MLL complex (Ang et al., 2011; Yang et al., 2014) our data suggest that a WDR5-binding MBD3C/NuRD complex functions separately from SET/MLL. Consistent with these findings, we showed that MBD3C interacts with WDR5 at the same binding surface as MLL1 and histone H3, using a conserved arginine-containing motif (Couture et al., 2006; Patel et al., 2008). As MBD3/NuRD has previously been shown to repress pluripotency genes during differentiation (Reynolds et al., 2012), it is possible that MBD3C/NuRD functions to oppose SET/MLL activity.

Our data reveal an additional layer of complexity to the composition and function of chromatin remodelers in ESCs and uncover a previously unidentified function for the WDR5 protein. The WDR5 binding domain appears to be essential for MBD3C/NuRD function in ESCs, while other MBD3/NuRD complexes are recruited to the same target genes via the MBD or other binding domains. However, since the differentiation defect of *Mbd3ab* KO cells can be overcome by constitutive overexpression of *Mbd3cΔN*, WDR5 may simply enhance the chromatin binding or remodeling activities of MBD3C/NuRD. Multiple independent mechanisms likely target different NuRD complexes to overlapping targets on chromatin, where the complexes function redundantly.

Figure 2.8 MBD3 isoforms associate with chromatin in ESCs

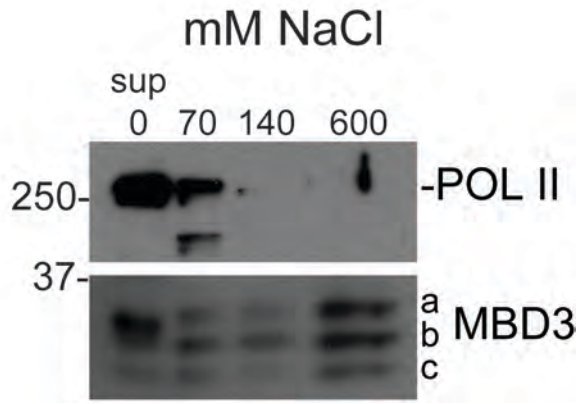
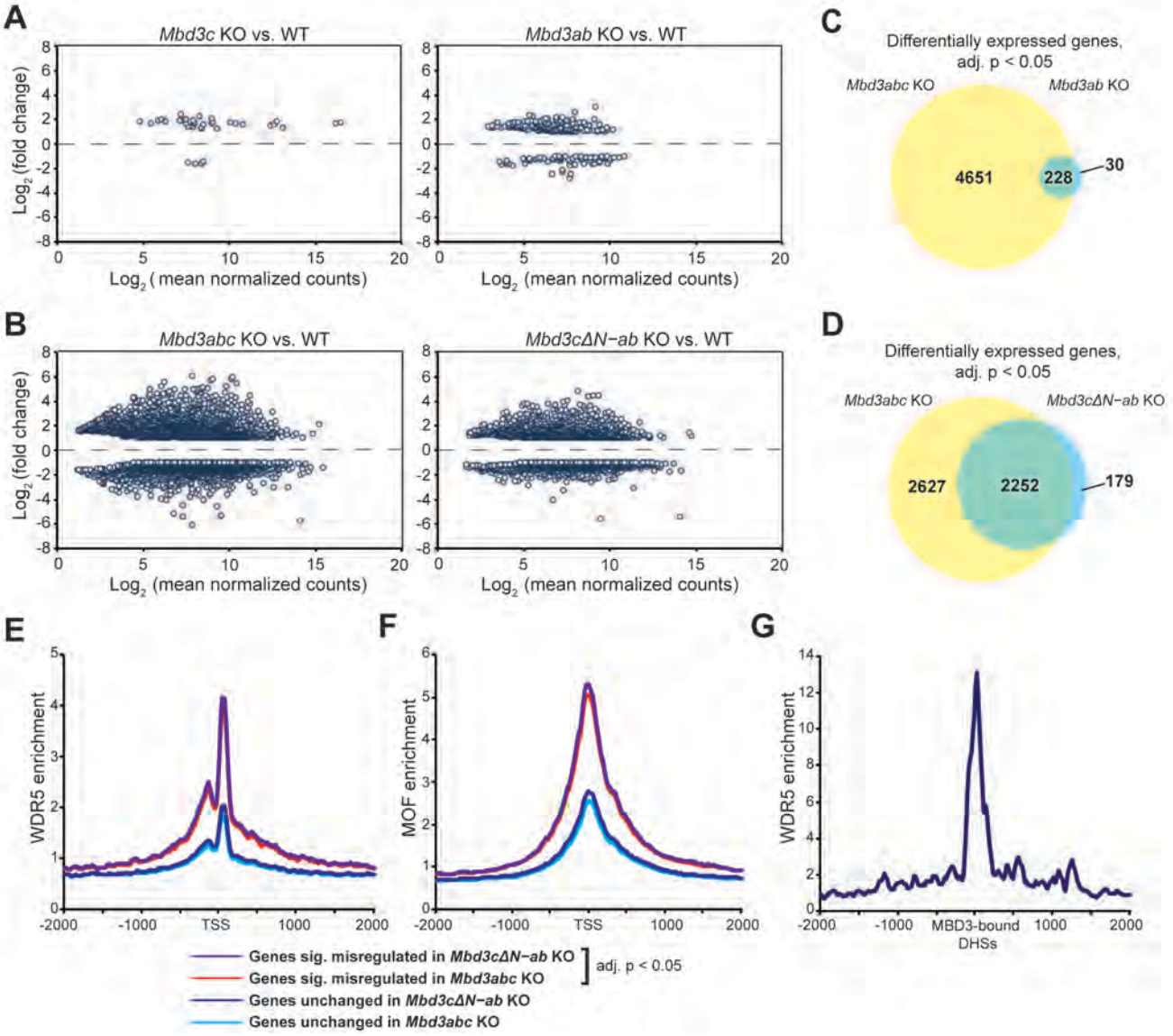


Figure 2.8: MBD3 isoforms associate with chromatin in ESCs
Western blot of MBD3 from salt fractionation of chromatin from WT ESC nuclei. Protein released at indicated concentrations of NaCl is shown. RNA Pol II is included as a control.

Figure 2.9: MBD3C is redundant with MBD3A and MBD3B in regulation of gene expression

(A-B) MA plots showing $\log_2$fold change in gene expression in *Mbd3c* KO (**(A)**, left), *Mbd3ab* KO (**(A)**, right), *Mbd3abc* KO (**(B)**, left), and *Mbd3cΔN−ab* KO (**(B)**, right) ESCs relative to WT. Genes shown are significantly misregulated ≥ 2-fold (adjusted $p < 0.05$) compared with WT. **(C-D).** Venn diagrams showing overlap between misregulated genes in ESCs of indicated genotypes. **(E-F)** WDR5 binding (Ang et al., 2011) **(E)** and MOF binding (Li et al., 2012) **(F)** averaged over transcription start sites (TSSs) of misregulated or unchanged genes in *Mbd3abc* KO (red) and *Mbd3cΔN−ab* KO ESCs (purple). **(G)** Average WDR5 binding over MBD3-bound (Yildirim et al., 2011), TSS-distal DNase I hypersensitive sites (GSM1014514).

Figure 2.9 MBD3C is redundant with MBD3A and MBD3B in regulation of gene expression



MATERIALS AND METHODS

Cell Culture and Generation of ESC lines

Murine ESCs are derived from E14 and cultured on gelatin-coated plates as previously described (Chen et al., 2013b). *Mbd3a-H3F*, *Mbd3c-H3F*, and *Mbd3cΔN-H3F* ESCs were generated by infection of E14 with pLJM1 lentiviral vectors carrying the respective constructs.

The H3F-WDR5 targeting construct was made by inserting PCR-produced homology arms (959 and 561 bp) and an H3F tag into pBluescript II SK+ (Stratagene). Oligos were inserted into pX330-puro^R to target to the N-terminus of WDR5. The plasmids were transfected with FuGENE HD (Promega) into E14, *Mbd3c* KO, and *Mbd3abc* KO cells. Clones were selected with puromycin and screened by PCR, sequencing, and western blot.

Expression of MBD3C and WDR5 domain mapping mutants

MBD3C-H3F, V5-WDR5 constructs and the indicated WDR5 mutants (Yang et al., 2014) were cloned into pCAGGS-IRES-Hygro^R using the 5’ XhoI site and the 3’ EcoRI site. pCAGGS-V5-WDR5 was created from a synthesized full-length mouse *Wdr5*. MBD3A-H3F and truncations of MBD3C-H3F with XhoI and MfeI restriction sites were derived by PCR on *Mbd3*-H3F sequences and cloned into the vector. The MBD3C Δ41-50 and R43A mutant constructs were made by PCR on pCAGGS-MBD3C-H3F with primers incorporating the mutations and flanking primers and then digesting the new sequences and ligating them into the vector. pCAGGS–cN41-50–MBD3A-H3F was made by synthesizing the N-terminus and

inserting it into the MBD3A plasmid using the XhoI site and a gene-internal BamHI site. Plasmids were transiently transfected into 293T cells using FuGENE HD (Promega), and the cells were harvested for IP after 2 days.

Generation of MBD3 isoform KO ESCs

Mbd3ab, *Mbd3c*, *Mbd3abc*, and *Mbd3cΔN-ab* KO ESC lines were generated using the CRISPR/Cas9 system (Cong et al., 2013) to introduce mutations into *Mbd3* exon 2 (*Mbd3ab* KO and *Mbd3cΔN−ab* KO), intron 2 (*Mbd3c* KO), or exon 5 (*Mbd3abc* KO). Guide RNAs targeting *Mbd3ab*, *Mbd3c*, or *Mbd3abc* (see Table 2.3 for sequences) were cloned into the pX330-puro^{R} vector and transfected into E14 ESCs as described (Hainer et al., 2015). Individual clones were screened by TOPO cloning (Invitrogen) and sequencing to verify the presence of homozygous frameshift mutations. ESC lines were further screened by western blot to verify loss of the appropriate MBD3 isoform(s). To create the *Mbd3cΔN−ab* KO line, we first transfected the pX330 plasmid targeting *Mbd3c* into E14 ESCs along with a donor plasmid containing the *Mbd3c* coding sequence with a 50-amino acid N-terminal deletion. The donor plasmid was generated by annealing oligos to make the *Mbd3cΔN* cDNA construct and cloning along with ~2kb homology arms flanking the *Mbd3c* start site into pBluescript SK II+ (Stratagene). Verified *Mbd3cΔN* ESC lines were then retargeted using the *Mbd3ab* KO guide RNA plasmid and screened as described above.

Mbd3abc KO/Mbd3c o/e ESC lines were made by replacing the endogenous *Mbd3* locus with a construct containing PCR-produced homology

arms (2070 and 1649 bp), the full CAGGS promoter, and a modified *Mbd3c*-H3F-polyA sequence without CpGs in the gene or tag. The *Mbd3abc* KO – cΔN o/e construct was made using PCR and restriction digestion to delete the 150 bp corresponding to the unique N-terminus of MBD3C. Oligos were inserted into two CRISPR plasmids (pX330-puro^{R}) and transfected into E14 cells as described above. Clones were selected with puromycin and screened by PCR, sequencing, and western blot.

5'RACE

5'RACE was performed on 4µg total RNA using the 5'RACE System Version 2.0 kit (Invitrogen). See Table 2.3 for primer sequences.

MBD3 Purification, WDR5 Purification, and LC-MS/MS

MBD3/NuRD complex was purified from MBD3-H3F, MBD3A-H3F, MBD3C-H3F, and MBD3CΔN-H3F ESCs as described (Yildirim et al., 2011). Purified samples were separated by SDS-PAGE, stained with SimplyBlue SafeStain (Invitrogen), and LC-MS/MS was performed as described in (Chen et al., 2013b). For WDR5 complex purification, nuclear fractions were isolated from H3F-WDR5 ESCs using the NE-PER kit (Thermo), diluted 1:3 in MVL buffer (50mM Tris, pH 7.5; 250mM NaCl; 0.1% Triton X-100), and purified similarly, omitting the His purification step.

Western Blotting and Immunoprecipitation

Western blots were performed with the following antibodies: anti-MBD3 (Bethyl A302-528A and Abgent AM2203B), anti-FLAG (M2, Sigma F1804), anti-WDR5 (Bethyl A302-429A and A302-430A), anti-OCT4 (Santa Cruz sc-8628), anti-NANOG (Bethyl A300-398A), anti-MTA1 (Bethyl A300-911A), anti-MTA2 (Santa Cruz sc-28731), anti-CHD4 (Bethyl A301-082A) anti–β-ACTIN (Sigma A1978), anti–RNA POLYMERASE II (Santa Cruz sc-899), anti-RBBP4 (Bethyl A301-206A), anti-RBBP7 (Bethyl A300-958A), anti-LSD1 (Bethyl A300-215A), anti-p66α (Bethyl A302-358A), anti-p66β (Bethyl A301-281A), anti-HDAC1 (Bethyl A300-713A), anti-HDAC2 (Bethyl A300-705A), anti-ASH2L (Bethyl A300-112A), anti-V5 (Invitrogen 46-0705). Mouse tissue lysates were prepared by homogenizing indicated tissues in lysis buffer (50mM Tris-HCl pH 7.5; 150mM NaCl; 0.5% Triton X-100; 5% glycerol; 1mM PMSF). For IP, nuclear lysates were prepared using the NE-PER kit (Thermo). FLAG IP was performed as described in (Chen et al., 2013b). MBD3 IPs in ESCs and 293T cells were performed similarly, except washes were performed in MVL buffer + 1mM EDTA.

Glycerol Gradient Analysis

Nuclear extracts were prepared from *Mbd3-H3F* ESCs using the NE-PER kit (Thermo). Nuclear extract (1470μg) was diluted in MVL buffer and spun in a 10-40% glycerol gradient at 37krpm for 17 hours in a Beckmann L-90K ultracentrifuge. 29 fractions were collected and odd fractions western blotted with the indicated antibodies.

Embryoid Body Differentiation

ESCs were differentiated to embryoid bodies (EBs) in suspension culture. 2.5 x 10^6 cells were plated in ESC medium without LIF in bacteriological plates. Cells were replated to non-gelatinized cell culture plates after 3 days and harvested for western blots at the indicated timepoints.

Chromatin Extraction Assay

Chromatin was extracted from WT ESCs by salt fractionation as described (Henikoff et al., 2009). Briefly, 4 x 10^7 cells were pelleted, washed in PBS, and resuspended in TM2 buffer (10mM Tris-HCl pH 7.5, 2mM $MgCl_2$, 1.5% NP-40, 1x HALT protease inhibitor cocktail) on ice for 5min. Cells were pelleted and the nuclei incubated in TM2 + 70mM NaCl for 2h at 4°C. Nuclei were pelleted and incubated as before in TM2 + 140mM NaCl, and further pelleted and incubated in TM2 + 600mM NaCl overnight at 4°C. Supernatants from each incubation were saved as 0, 70,140 and 600mM NaCl fractions, clarified by centrifugation at full speed, and western blotted with the indicated antibodies.

Reprogramming to pluripotent stem cells

Lentiviral plasmids for dox-inducible Oct4, Sox2, and Klf4 expression (pLenti-Tet-OSK-mCherry) and for L-Myc expression (pLenti-Tet-L-Myc) were generated by digestion of CMV-OSK and CMV-L-Myc cDNA from FUW-OSKM and pMXs-Ms-L-Myc (Addgene 20328 and 26023 respectively) and cloning into pcDNA3.1 with HIV1-based 5’ and 3’ LTRs from pGIPZ. To package lentivirus, 293T cells were

transfected with 5µg pLenti-Tet-OSK-mCherry, pLenti-Tet-L-Myc, FUW-rTtA (Addgene 20342), or EOS-EGFP (Addgene 21313) lentiviral reporter plasmids along with packaging plasmids (5µg psPAX2 and 2.5µg pCMV-VSV-G (Addgene)). Primary MEFs were infected with day 2 viral supernatant using 8µg/mL hexadimethrine bromide (Sigma), and re-infected after 24 hours. MEFs were replated in ESC media after 48 hours and induced with 2µg/ml dox 4 days after the first infection. Dox was removed at day 20 and cells were cultured for an additional 10 days. On Day 30 cells were imaged to assess loss of OSK transgenes and EOS-EGFP reporter activation, and on day 31 were stained for alkaline phosphatase according to kit instructions (Millipore). Single colonies were picked on Day 30, expanded, and western blotted with the indicated antibodies. Media were changed every other day.

Bisulfite pyrosequencing

WT ESCs were subjected to embryoid body differentiation and harvested at the indicated timepoints. Briefly, genomic DNA was phenol-chloroform extracted from cells incubated in ES cell lysis buffer (10 mM Tris, pH 7.5; 10 mM EDTA; 10 mM NaCl; 0.5% sarkosyl) with 1 µg/µL proteinase K at 55°C overnight. The DNA was bisulfite converted using the EpiTect Bisulfite Kit (QIAGEN). Primers were designed for the *Mbd3c* CpG island using PyroMark Q24 software (QIAGEN, see Table 2.3 for sequences), with one PCR primer in a pair biotinylated, and PCR was performed on the converted DNA with KAPA HiFi HotStart Uracil+ ReadyMix (Kapa Biosystems). PCR products were bound to streptavidin sepharose beads

(GE Healthcare) and sequenced using a PyroMark Q24 (QIAGEN). Data were analyzed using PyroMark Q24 software.

RNA-seq

Total RNA was isolated from two biological replicate WT, *Mbd3ab* KO, *Mbd3c* KO, *Mbd3abc* KO, and *Mbd3cΔN−ab* KO ESC lines using TRIzol (Life Technologies), and purified with the Zymo RNA Clean and Concentrator kit. 2μg RNA was used for library preparation. RNA was rRNA-depleted (NEB and Clontech) and strand-specific libraries were prepared using the TruSeq Stranded mRNA LT kit (Illumina) by Applied Biological Materials, Inc.

RNA-seq data analysis

Reads were mapped to the mouse mm10 genome with TopHat2 (Kim et al., 2013), using parameters --library-type fr-firststrand --segment-length 38. Mapped reads were processed in HOMER (Heinz et al., 2010) using the "analyzeRepeats" command to calculate raw counts and normalized reads per kilobase per million mapped reads (rpkm) for each gene. Differential gene expression was calculated with DESeq2 (Love et al., 2014) using the "getDiffExpression" command in HOMER. Genes with an adjusted p value < 0.05 and log2 (fold change) were considered significantly changed.

To map WDR5 and Mof binding at TSSs of significantly changed genes, WDR5 (Ang et al., 2011) and Mof (Li et al., 2012) ChIP-seq data were downloaded from GEO (GSE22934 and GSE37268) and aligned to the mouse

mm10 genome using Bowtie (Langmead et al., 2009). Mapped reads were processed in HOMER using the “annotatePeaks” command. For Mof ChIP-seq, 2 replicate libraries were averaged in the aggregation plot.

To map WDR5 at MBD3-bound DNase I hypersensitive sites (DHSs), peaks were called from MBD3 ChIP-seq data (Yildirim et al., 2011); GSE31690) using the HOMER “findPeaks” command, and MBD3-bound DHSs were identified using the “mergePeaks” command with peaks called from mouse ENCODE DHSs (GSM1014154) without TSSs. The WDR5 ChIP-seq library was aligned to the MBD3-bound DHS peak data using the “annotatePeaks” command.

ACCESSION NUMBERS

RNA-seq data was deposited at GEO with accession # GSE80708.

Table 2.3: Oligonucleotides used in Chapter II

Name	Sequence	Purpose
Mbd3c-10trunc-XhoI-f	CTCGAGACCATGATCTCCGTGCCTGATCGC	clone Mbd3cΔ1-10-FLAG into pCAGGS
Mbd3c-20trunc-XhoI-f	CTCGAGACCATGATTCAGGCTCTGGCTAAGCAC	clone Mbd3cΔ1-20-FLAG into pCAGGS
Mbd3c-30trunc-XhoI-f	CTCGAGACCATGCCCTCCAACCCTCCATGGAC	clone Mbd3cΔ1-30-FLAG into pCAGGS
Mbd3c-40trunc-XhoI-f	CTCGAGACCATGGCGGCCCGCTGCAGA	clone Mbd3cΔ1-40-FLAG into pCAGGS
XhoI-Mbd3ab-f	CTCGAGACCATGGAGCGGAAGAGGTGG	clone Mbd3a-FLAG into pCAGGS
3xFLAG-MfeI-r	CAATTGCTACTTGTCATCGTCATCCTTG	clone Mbd3a/c-FLAG sequences into pCAGGS
pCAGGS-insertseq-f	GCAACGTGCTGGTTATTGTGC	mutagenesis PCR on pCAGGS-Mbd3c-FLAG
Mbd3 seq9 R	CCACACACCAGGGTTCTTCT	mutagenesis PCR on pCAGGS-Mbd3c-FLAG
Mbd3c-R43A-r	GAAGACTCTGCAGGCGGCCGCTCCGACC	R43A mutagenesis on pCAGGS-Mbd3c-FLAG
Mbd3c-41del50-r	CAGGCTTGCCTCCGACCGGGGTCC	Δ41-50 mutagenesis on pCAGGS-Mbd3c-FLAG
XhoI-V5-BamHI+	AGCTTCTCGAGACCATGGGTAAGCCTATCCCTAACCCTCTCCTCGG TCTCGATTCTACGGGATCCGCCAC	clone V5 tag into pCAGGS-Wdr5 (replaces GST tag)
XhoI-V5-BamHI-	GTGGCGGATCCCGTAGAATCGAGACCGAGGAGAGGGTTAGGGATA GGCTTACCCATGGTCTCGAGAAGCT	clone V5 tag into pCAGGS-Wdr5 (replaces GST tag)
Mbd3a_pLJM1 AfeI	AGCGCTACCATGGAGCGGAAGAGGTG	clone Mbd3a-H3F into pLJM1
Mbd3c_pLJM1 AfeI	AGCGCTACCATGGCGCGCATTTGGTTTG	clone Mbd3c-H3F into pLJM1
Mbd3ctrunc_pLJM1AfeI	AGCGCTACCATGGGCAAGCCTGACCTGAA	clone Mbd3cΔN-H3F into pLJM1
Mbd3FLAG_pLJM1BstBI	TTCGAACTACTTGTCATCGTCATCCTTG	clone Mbd3-H3F into pLJM1
Mbd3a_NotIEcoRI	AAGCGGCCGCACCATGGAGCGGAAGAGGTGGGAGTGC	Amplify Mbd3a from cDNA
Mbd3a/b/c R	CGGAATTCCACTCGCTCTGGCTCCGGCTTTCCTCCTC	Amplify Mbd3 (all isoforms) from cDNA
Mbd3c v2 F	AAGCGGCCGCACCATGGCGCGCATTTGGTTTGGTGGG	Amplify Mbd3c from cDNA
Mbd3c v2 dN	AAGCGGCCGCACCATGGGCAAGCCTGACCTGAACACC	Amplify Mbd3cΔN from cDNA
GSP1-3	TCCTGACCAGTTCTTCT	5'RACE sequencing primer
GSP1-2	GTGTAGAGCACTCGCAATG	5'RACE sequencing primer
Mbd3c Xho1	CTCGAGACCATGGCGCGCATTTGGTTTG	clone Mbd3c-H3F into pCAGGS
Mbd3ctrunc Xho1	CTCGAGACCATGGGCAAGCCTGACCTGAA	clone Mbd3cΔN-H3F into pCAGGS
Mbd3FLAG EcoR1	GAATTCCTACTTGTCATCGTCATCCTTG	clone Mbd3c- and Mbd3cΔN-H3F into pCAGGS
FLAG Xho1	CTCGAGACCATGGACTACAAAGACGATGA	clone FLAG-Wdr5 domain muts into pCAGGS
hWdr5 EcoR1	GAATTCTTATTAGCAGTCACTCTTCCACA	clone FLAG-Wdr5 domain muts into pCAGGS
pX330-Mbd3abc-1	CACCGGTGTGTAGAGCACTCGCAA	*Mbd3abc* KO guide RNA oligo
pX330-Mbd3abc-2	AAACTTGCGAGTGCTCTACACACC	*Mbd3abc* KO guide RNA oligo
pX330-Mbd3ab-1	CACCGCTTTCCGGTGCGGAAGTCGA	*Mbd3ab* KO guide RNA oligo

pX330-Mbd3ab-2	AAACTCGACTTCCGCACCGGAAAGC	*Mbd3ab* KO guide RNA oligo
pX330-Mbd3c-1	CACCGTCCATGGACCCCGGTCGGAG	*Mbd3c* KO guide RNA oligo
pX330-Mbd3c-2	AAACCTCCGACCGGGGTCCATGGAC	*Mbd3c* KO guide RNA oligo
Mbd3abcCRISPR seq F	GACTTACAGGGAGTTGTGAGCC	PCR *Mbd3abc* KO from gDNA to verify sequence
Mbd3abcCRISPR seq R	TGCTTTCTCACTGCTATTTCCCCA	PCR *Mbd3abc* KO from gDNA to verify sequence
Mbd3abCRISPR seq F	CCTTAGGCTTCCCAGATGAACT	PCR *Mbd3ab* KO from gDNA to verify sequence
Mbd3abCRISPR seq R	GGTGCTTAGCCAGAGCCTGAA	PCR *Mbd3ab* KO from gDNA to verify sequence
Mbd3cCRISPR seq F	GAAGAAGTTCCGCAGCAAGC	PCR *Mbd3c* KO from gDNA to verify sequence
Mbd3cCRISPR seq R	AGGCAATGGTTTCTTCCACCC	PCR *Mbd3c* KO from gDNA to verify sequence
Mbd3ctrunc Sal1	TCGACATGGGCAAGCCTGACCTGAACACCGCGCTGCCTGTACGGCAGACTGCATCCATCTTCAAGCAA	Annealing oligo to clone Mbd3cΔN into pBluScript for CRISPR
Mbd3ctrunc Age1	CCGGTTGCTTGAAGATGGATGCAGTCTGCCGTACAGGCAGCGCGGTGTTCAGGTCAGGCTTGCCCATG	Annealing oligo to clone Mbd3cΔN into pBluScript for CRISPR
Mbdc BS+	AGGGAGGTGTTTAGTTAGAGT	PCR bisulfite-converted gDNA from *Mbd3c* CpG island
Mbd3c BS-	[Btn]AAAAAAAATTCCCCAACAAACCACAACT	PCR bisulfite-converted gDNA from *Mbd3c* CpG island. Adds biotin tag
Mbd3c-1_BS+	ATTTTTATTTTGATTTGGTTGGAAGAATTA	PCR bisulfite-converted gDNA from *Mbd3c* CpG island
Mbd3c-1_BS-	[Btn]CTTCCCACCTACCTTCCTAATAACCTA	PCR bisulfite-converted gDNA from *Mbd3c* CpG island. Adds biotin tag
Mbd3c_BSseq	GATTTTTTATTTATTAAATTAAATG	Pyrosequencing primer #1
Mbd3c-1b_BSseq	GTGTTAGTTGTGGTTTG	Pyrosequencing primer #2
EcoRI-5'-Wdr5-f	TTGAATTCCTTGGGACAGTTGATTTGTTGGAGG	clone H3F-Wdr5 5' homology arm
BamHI-5'-Wdr5-r-PAMmut	AAGGATCCCATGGCTCTGAAGACCACAGGGC	clone H3F-Wdr5 5' homology arm
XbaI-3'-Wdr5-f	AATCTAGAGCCACAGAGGAGAAGAAGCCA	clone H3F-Wdr5 3' homology arm
SacI-3'-Wdr5-r	TAGAGCTCATCAAGCACTAGCTGGTTCCAAAC	clone H3F-Wdr5 3' homology arm
BamHI-6xHis-f	TAGGATCCCATCACCACCATCATCACGC	clone 6xHis,3xFLAG tag for H3F-Wdr5
XbaI-3xFLAG-r	TTTCTAGACTTGTCATCGTCATCCTTGTAGTC	clone 6xHis,3xFLAG tag for H3F-Wdr5
pX330-Wdr5-Nterm+	CACCGTTCACGGTGTCCTGCCCTGT	CRISPR oligo targeting the N terminus of Wdr5
pX330-Wdr5-Nterm-	AAACACAGGGCAGGACACCGTGAAC	CRISPR oligo targeting the N terminus of Wdr5
XhoI-5'Mbd3-homol-f	CTCGAGTCCTGCATACGTCTCTTCCCAG	clone *Mbd3abc* KO/Mbd3c o/e 5' homology arm
SalI-5'Mbd3-homol-r	GTCGACACGTGGTAGACTTTCTGCC	clone *Mbd3abc* KO/Mbd3c c o/e 5' homology arm
EcoRI-3'Mbd3-homol-f	GAATTCTGCTGCAGGCCAGGGTGG	clone *Mbd3abc* KO/Mbd3c o/e 3' homology arm
BamHI-3'Mbd3-homol-r	GGATCCTACCTCTGGTGCCACCATCCT	clone *Mbd3abc* KO/Mbd3c o/e 3' homology arm

pCAGGS-5'seq	GAGCCTCTGCTAACCATGTTCA	mutagenesis PCR on *Mbd3abc* KO/Mbd3c o/e construct
Mbd3c-noCpG-afterBsrGI-r	AGAGCACTGGCAATGGCAGA	mutagenesis PCR on *Mbd3abc* KO/Mbd3c o/e construct
Mbd3c-noCpG-50trunc-r	CAGGTCAGGCTTGCCCATGGTGGCAAGCT	Δ1-50 mutagenesis on *Mbd3abc* KO/Mbd3c o/e construct
pX330-5'Mbd3+	CACCGAGAAAGCAGAACCTTACACG	CRISPR oligo targeting 5' of the *Mbd3* locus
pX330-5'Mbd3−	AAACCGTGTAAGGTTCTGCTTTCTC	CRISPR oligo targeting 5' of the *Mbd3* locus
pX330-3'Mbd3+	CACCGAGCCAGAGCGAGTGTAGCAC	CRISPR oligo targeting 5' of the *Mbd3* locus
pX330-3'Mbd3−	AAACGTGCTACACTCGCTCTGGCTC	CRISPR oligo targeting 5' of the *Mbd3* locus
Mbd3-exon6-f	GGAAGCAGGAGGAGCTGGT	PCR check primer for *Mbd3abc* KO/Mbd3c o/e
Mbd3c-noCpG-outcheck-r	TGGCCTAGCAACATTCTGGC	PCR check primer for *Mbd3abc* KO/Mbd3c o/e
Mbd3c-noCpG-outcheck-f	GATTTCCAGACAGTGTGCACTAAGTG	PCR check primer for *Mbd3abc* KO/Mbd3c o/e
CMVenhancer-r	GGCGGGCCATTTACCGTAAG	PCR check primer for *Mbd3abc* KO/Mbd3c o/e

CHAPTER III: CHARACTERIZATION OF H3.3K56AC IN PLURIPOTENCY AND DIFFERENTIATION

Contributions

Tom Fazzio designed and cloned H3.3K56R guide RNAs and homology arm templates and constructed H3.3K56R mutant ESC lines. Unpublished data from Feixia Chu is referenced in the Introduction. Ly-sha Ee performed all other experiments. Figure 3.1B was made by Kurtis McCannell and used with permission.

ABSTRACT

Post-translational modifications of histones play numerous roles in the modulation of chromatin structure and gene expression. Acetylation of lysine 56 of histone H3 (H3K56ac) is a modification in the histone core that is well-characterized and abundant in yeast but less common in mammalian cells, where its functions are largely unknown. In ESCs, H3K56ac interacts with OCT4, SOX2, and NANOG at gene promoters to maintain pluripotency but the precise roles of H3K56ac in ESC pluripotency and during differentiation are still undescribed. It is also unclear whether K56ac has separate functions when it occurs on different histone H3 variants. In this chapter we show that ESC lines depleted for H3.3K56ac by K56R substitution in both *H3.3* genes are largely unchanged in morphology and self-renewal. However, although loss of H3.3K56ac does not appear to be critical to exit the pluripotent state, H3.3K56ac is important for differentiation in certain contexts, particularly the formation of neurons. Our work complements previous data indicating that H3K56

hyperacetylation leads to upregulation of the ectodermal lineage during differentiation.

INTRODUCTION

Histone modifications regulate chromatin structure by binding and recruiting other chromatin regulators that, at genomic regions such as enhancers and promoters can in turn activate or repress transcription. In embryonic stem cells (ESCs), many histone modifications are linked to the transcription factor network that includes the pluripotency factors OCT4, SOX2, and NANOG and which is essential to maintain ESCs in an undifferentiated state. Pluripotent chromatin is characterized by a dynamic, more open structure as compared with somatic cell chromatin (Meshorer and Misteli, 2006). In addition to “bivalent” domains – regions co-marked by modifications associated with activation and repression (H3K4me3 and H3K27me3 respectively) – ESCs contain elevated levels of activation-associated histone marks accompanied by higher global transcription (Efroni et al., 2008).

While most well-studied modifications are located on the N-terminal histone tails, modifications of amino acids within the globular histone fold domain have also been identified and characterized. Lysine 56 is located in the α-N helix of histone H3 (Figure 3.1A and B) just N-terminal to the histone fold at the entry and exit points of DNA (Masumoto et al., 2005; Ozdemir et al., 2005; Xu et al., 2005). Acetylation of lysine 56 (by the RTT109 and p300 and GCN5 histone acetyltransferases in yeast and mammals respectively) disrupts a water-

mediated contact between the histone and the DNA (Das et al., 2009; Driscoll et al., 2007; Han et al., 2007; Luger et al., 1997; Neumann et al., 2009; Schneider et al., 2006; Tjeertes et al., 2009) and possibly destabilizes nucleosomes, although a recent study reported no change in folding of H3K56Q nucleosomal arrays (Watanabe et al., 2010).

In yeast H3K56ac is important for histone gene expression through recruitment of SWI/SNF, and has numerous roles in CAF1-mediated nucleosome assembly, response to DNA damage, DNA repair, and H2A.Z dimer exchange (Chen et al., 2008a; Li et al., 2008; Masumoto et al., 2005; Vempati et al., 2010; Watanabe et al., 2013; Xu et al., 2005). H3K56ac is far less abundant in mammalian cells (about 1% of total H3 in HeLa cells) (Xie et al., 2009) and its functions in mammalian cells are largely unknown. In both human and mouse ESCs H3K56ac is linked to the pluripotency network, co-localizing with OCT4, SOX2, and NANOG at target gene promoters (Tan et al., 2013; Xie et al., 2009). Depletion of H3K56ac by knockdown of the histone chaperone *Asf1a* led to decreased expression of pluripotency factors and increased expression of markers of all three germ layers (Tan et al., 2013), and *Oct4* KD ESCs exhibit a marked decrease in H3K56ac levels by mass spec analysis (T. Fazzio and F. Chu, unpublished observation) suggesting that H3K56ac is important for ESC self-renewal and pluripotency and/or for preventing spontaneous differentiation. Additional work from the Grunstein lab in human ESCs showed that H3K56ac becomes enriched at developmental gene promoters during RA-induced differentiation, although transcription of a subset of enriched genes was

unchanged (Xie et al., 2009). A recent study showed that H3K56ac levels in ESCs could be regulated by a long noncoding RNA, *lncPRESS1*, through sequestration of the NAD-dependent deacetylase SIRT6 (Jain et al., 2016).

Here, we have constructed ESC lines with homozygous K56R mutations in each *H3.3* gene to deplete K56ac at histone H3.3, which marks active gene promoters and enhancers as well as heterochromatic and repeat regions in ESCs (Elsässer et al., 2015; Goldberg et al., 2010). We found that ESC self-renewal is largely unchanged in H3.3K56R mutants. Loss of OCT4 and NANOG during differentiation, as well as timely upregulation of many lineage markers were similarly unaffected. However, H3.3K56R mutants appear to be defective in formation of neurons during retinoic acid- (RA)-induced differentiation, although we did not observe decreased or delayed expression of two common neuroectoderm markers. Together, our data suggest that H3.3K56ac is important to maintain or enhance ESC pluripotency and is also required during later stages of differentiation. Our work further indicates that differently-modified histone variants could play varied roles in pluripotency and differentiation.

Figure 3.1 The mammalian core and replication-independent histone H3 variants, and position of H3K56 on the nucleosome

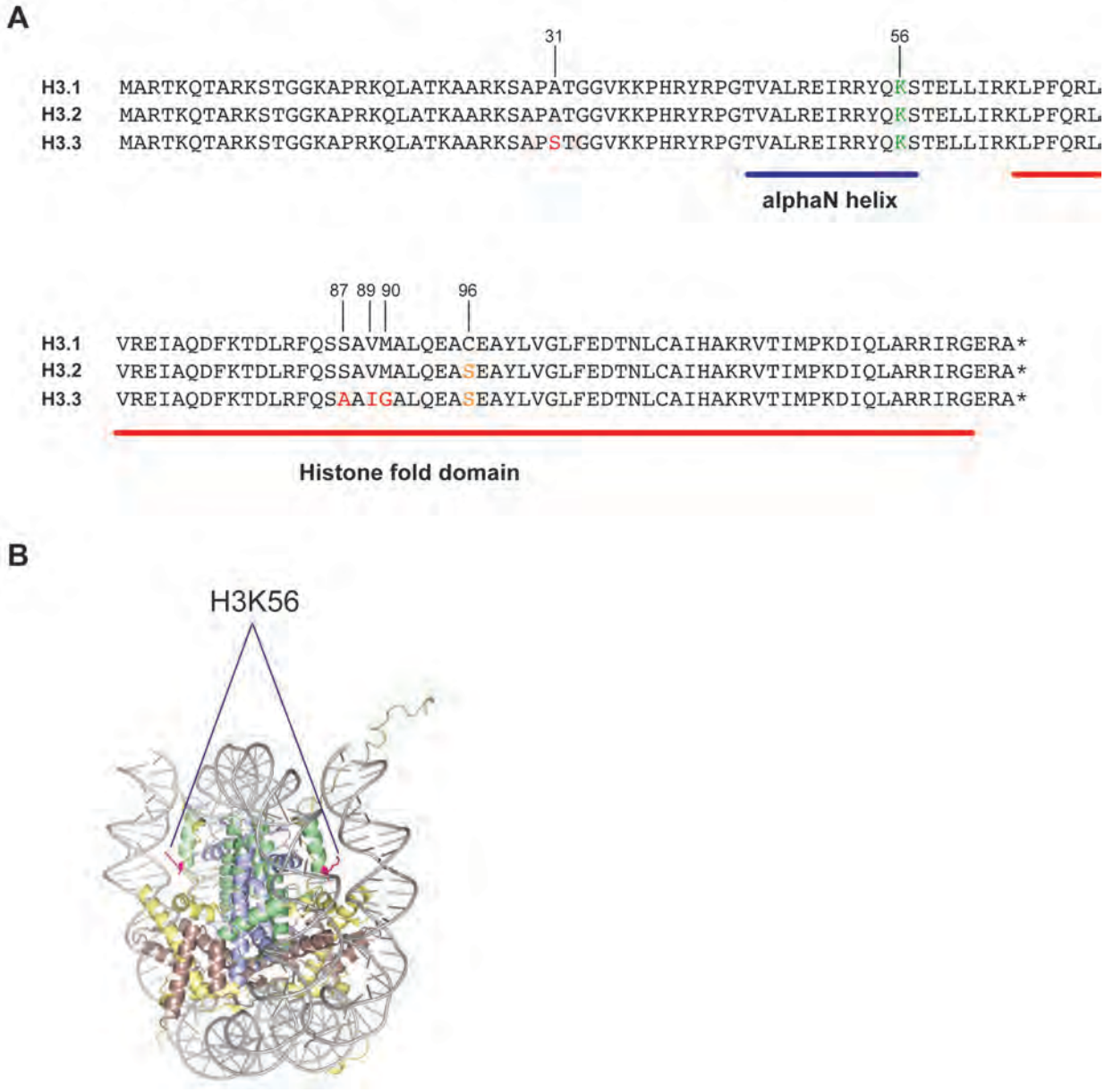


Figure 3.1: The mammalian core and replication-independent histone H3 variants, and position of H3K56 on the nucleosome
(A) Sequence alignment of the mammalian core (H3.1 and H3.2) and replication-independent (H3.3) H3 variants. Amino acids that are not conserved in H3.3 are shown in red, and differing between H3.1 and H3.2/3 in orange. Lysine 56 is indicated in green. **(B)** Pymol crystal structure of the nucleosome (Luger et al., 1997; PDB:1AOI) showing the position of both H3K56 residues (magenta). H2A, H2B, H3, and H4 are shown in yellow, brown, green, and blue respectively. Used with permission from Kurtis McCannell.

RESULTS AND DISCUSSION

To assess the effects of H3K56ac loss on ESC self-renewal and pluripotency we created an ESC line where lysine 56 of both *H3.3* alleles was homozygously mutated to arginine (henceforth referred to as the H3.3K56R mutant) using the CRISPR/Cas9 system (Cong et al., 2013). As both H3.1 and H3.2 are expressed from tandem, multicopy clusters at multiple loci (Marzluff et al., 2002) we were unable to create H3.1 or H3.2 mutant lines, as we could not target every *H3.1* or *H3.2* gene copy (T. Fazzio, personal communication). In contrast, H3.3 is expressed from two discrete genes (*H3f3a* and *H3f3b*) with different coding sequences, allowing for sequential targeting with two homology constructs (see Methods for details). Although H3.3 comprises a minority of total H3 in ESCs and the fraction of H3.3 acetylated at K56 is unknown, H3.3 has highly distinct genomic localization and functions in ESCs and development (Banaszynski et al., 2013; Elsässer et al., 2015; Goldberg et al., 2010; Jang et al., 2015). We therefore hypothesized that loss of H3.3K56ac could result in similar or overlapping phenotypes as those observed in H3.3-depleted ESCs.

H3.3K56ac is not required for ESC self-renewal

To determine whether H3K56ac levels were reduced in H3.3K56R mutants we prepared histones by acid extraction and western blotted for H3K56ac. We observed a very slight decrease in H3K56ac levels in H3.3K56R mutants (Figure 3.2A). This result was not unexpected given that K56ac has been detected in less than 1% of total H3 (Das et al., 2009; Xie et al., 2009; Yuan

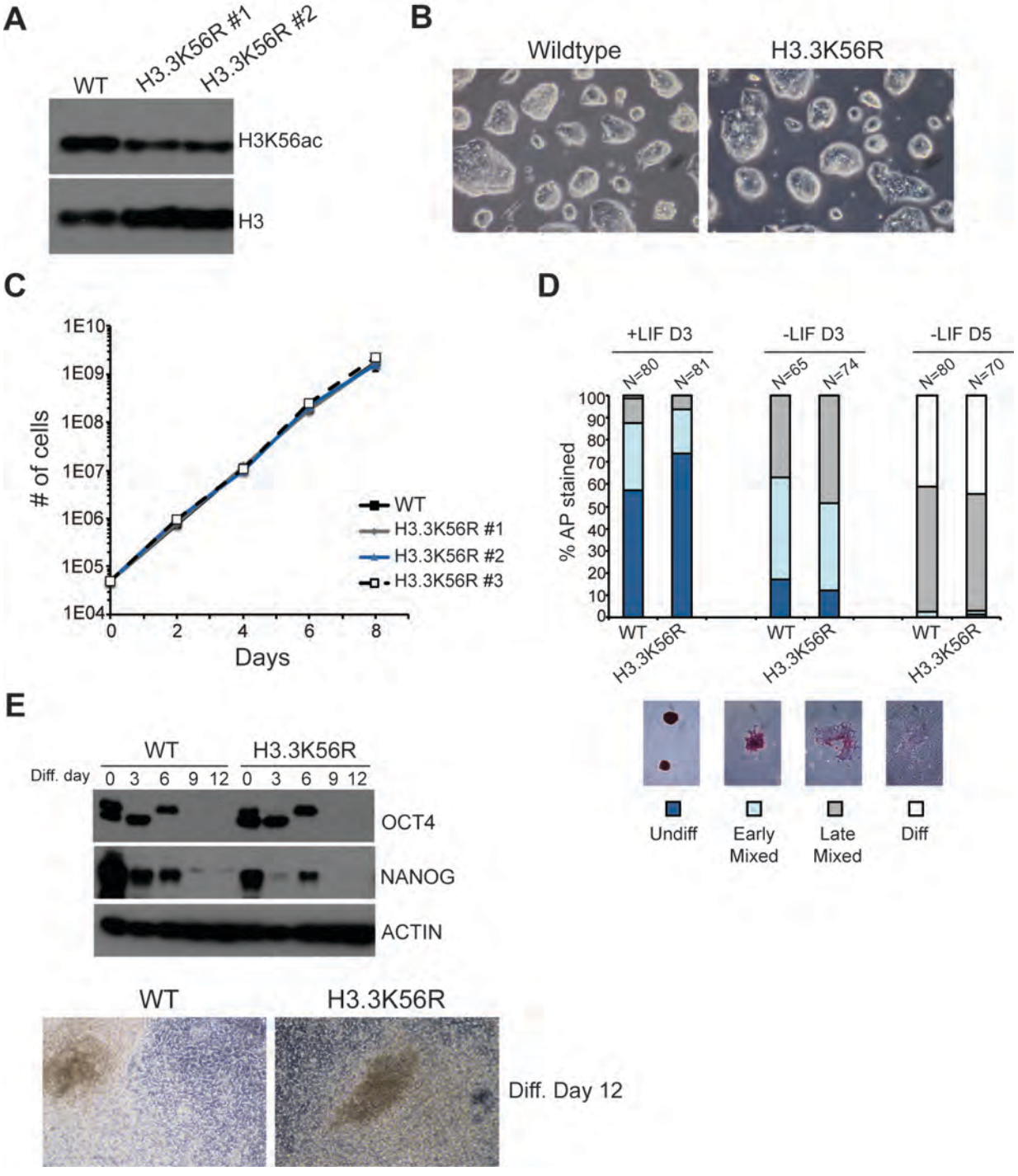
et al., 2009). It is however important to note that several antibodies against H3K56ac including the Active Motif antibody used in this study were shown to be non-specific (Drogaris et al., 2012; Pal et al., 2016). It is therefore likely that the blot in Figure 3.2A is skewed by cross-reactivity with H3K9ac and other acetylated H3 residues. H3.3K56R ESCs exhibit colony morphology similar to wildtype (Figure 3.2B), and unchanged proliferation rates (Figure 3.2C). Moreover, H3.3K56R cells do not appear to be defective in self-renewal as determined by similar levels of alkaline phosphatase (AP) staining to WT when maintained in leukemia inhibitory factor (LIF) (Figure 3.2D). We next tested whether H3.3K56R ESCs were defective for differentiation in an embryoid body (EB) timecourse. We observed normal rates of OCT4 and NANOG pluripotency factor loss during the timecourse (Figure 3.2E, top panel) and similar morphology in WT and H3.3K56R cells differentiated for 12 days (Figure 3.2E, bottom panel), consistent with a flattening out and loss of AP staining in H3.3K56R ESCs cultured for 3-5 days in the absence of LIF (Figure 3.2D). Together, these data indicate that H3.3K56ac is not required for ESC self-renewal or for mediating loss of pluripotency factors during differentiation. Interestingly, a previous study showed that OCT4, SOX2, and NANOG expression is upregulated in *Sirt6* KO ESCs and stabilized during *Sirt6* KO EB differentiation (Etchegaray et al., 2015). SIRT6 is an NAD-dependent Class III histone deacetylase that deacetylates H3K9 and H3K56, among other H3 lysine residues (Michishita et al., 2008; 2009; Yang et al., 2009). Thus, H3K56 hyperacetylation appears to block differentiation by preventing loss of pluripotency factors. Additionally, Etchegaray et al. showed

that expression of endoderm and mesoderm lineage markers was downregulated while expression of ectoderm markers was upregulated with respect to wildtype in *Sirt6* KO EBs. To further investigate the role of H3K56ac in differentiation we are currently constructing an H3.3K56Q mutant ESC line. As the K56Q mutation mimics constitutive H3K56 acetylation (Masumoto et al., 2005) we hypothesize that this mutant will behave similarly to the *Sirt6* KO in differentiation assays. The H3.3K56Q cell line will also allow us to distinguish H3K56ac-specific phenotypes from those resulting from H3K9 hyperacetylation that are observed in the *Sirt6* KO.

Figure 3.2: H3.3K56R ESCs exhibit normal morphology, growth, and EB differentiation

(A) Western blot of acid-extracted histones from the indicated ESC lines. H3 serves as a loading control. **(B)** Representative images of WT and H3.3K56R mutant ESCs. **(C)** Growth curve for H3.3K56R ESC lines, relative to WT ESCs. **(D)** Histogram of AP staining of WT and H3.3K56R ESCs at the indicated timepoints in the presence or absence of LIF. The number of scored colonies for each sample is indicated above the corresponding bar. Representative images for scored classes are shown. **(E)** Western blots of differentiating WT and H3.3K56R ESCs (top) and representative images of WT and H3.3K56R ESCs at differentiation day 12 (bottom).

Figure 3.2 H3.3K56R ESCs exhibit normal morphology, growth, and EB differentiation



Differentiating H3.3K56R cells are defective for some mesoderm and endoderm lineage markers

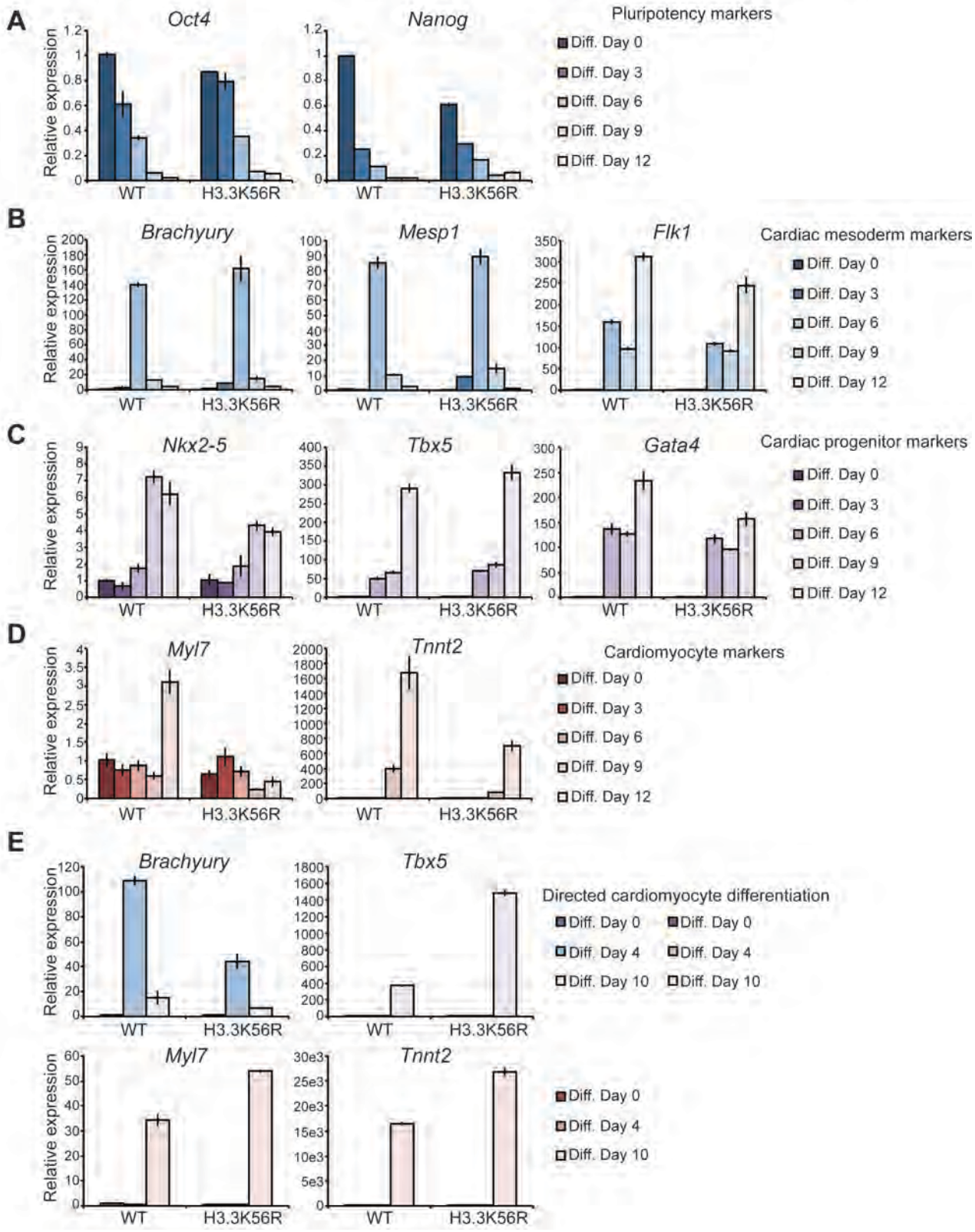
To further investigate pluripotency of the H3.3K56R ESCs, we measured expression levels of markers from the three germ layers using RT-qPCR. Consistent with the western blots in Figure 3.2E, *Oct4* and *Nanog* RNA levels decreased in WT and H3.3K56R ESCs during EB differentiation (Figure 3.3A). We first examined expression of markers for mesoderm, the central embryonic germ layer that gives rise to most blood cells, heart and muscle tissue. Cardiac specification can be divided into several stages of mesoderm differentiation, each of which is marked by the presence of specific genes (Rajala et al., 2011; Wamstad et al., 2012). As we had observed beating cardiomyocytes in late-stage differentiating H3.3K56R cultures, we predicted that cardiac genes would be largely unchanged. Expression and kinetics of early cardiac mesoderm genes were largely unchanged in the H3.3K56R mutant, save for a very slight decrease in *Flk1* (Figure 3.3B). Expression of two cardiac progenitor markers, *Nkx2-5* and *Gata4* was also slightly reduced, although there was no delay in upregulation (Figure 3.3C). Interestingly, the expression of the mature cardiomyocyte markers *Myl7* and *Tnnt2* was highly downregulated in H3.3K56R (Figure 3.3D), raising the possibility that the H3.3K56R mutants are less efficient in generating beating cardiomyocytes than WT. Since EBs produce cells from multiple differentiated lineages, we next performed directed differentiation (Kattman et al., 2011; Wamstad et al., 2012) to better enrich for cardiomyocytes and possibly quantify the differences in efficiency. Surprisingly, despite downregulation of *Brachyury*

the expression of the cardiac progenitor *Tbx5* and the cardiomyocyte markers *Myl7* and *Tnnt2* were all upregulated in the H3.3K56 mutant (Figure 3.3E), suggesting no defect in cardiac differentiation. Although further experimental replicates will be necessary to confirm this result, these data suggest that while loss of H3.3K56ac leads to mis-regulation of some mesodermal markers, acetylation is not required to form functional cardiac progenitors and cardiomyocytes, particularly if formation of cardiac mesoderm is enriched using a directed differentiation protocol.

Figure 3.3: Expression of mesoderm and cardiac lineage genes in H3.3K56R mutant ESCs

(A-D) RT-qPCR of known pluripotency **(A)**, cardiac mesoderm **(B)**, cardiac progenitor **(C)**, and cardiomyocyte **(D)** marker genes in WT and H3.3K56R mutant ESCs differentiating from EBs. Data are normalized to WT Diff. Day 0. Error bars represent +/- standard deviation of three technical replicates. Representative data from at least two biological replicate experiments are shown. **(E)** RT-qPCR of cardiac mesoderm (blue), cardiac progenitor (purple), and cardiomyocyte (red) lineage genes from the indicated timepoints during directed cardiomyocyte differentiation. One replicate is shown. Normalization and error bars are as above.

Figure 3.3 Expression of mesoderm and cardiac lineage genes in H3.3K56R mutant ESCs



Markers for endodermal differentiation (Figure 3.4A) and hematopoietic mesoderm (Figure 3.4B) were similarly affected: subsets of genes were slightly downregulated but not delayed in H3.3K56R mutants. Thus, it is possible that H3.3K56ac is only necessary to enhance the expression of certain differentiation markers. To further test this possibility it will be necessary to perform RNA-Seq to compare global gene expression during directed differentiation of WT, H3.3K56R, and H3.3K56Q cells. We hypothesize that H3.3K56Q mutants will be defective in cardiomyocyte and hematopoietic differentiation, with stabilized pluripotency markers similar to a *Sirt6* KO (Etchegaray et al., 2015). However, it is also highly possible that hyperacetylation of H3.1 and H3.2 in addition to H3.3 is necessary to observe said defects.

Figure 3.4 Expression of endoderm and hematopoietic lineage genes in H3.3K56R mutant ESCs

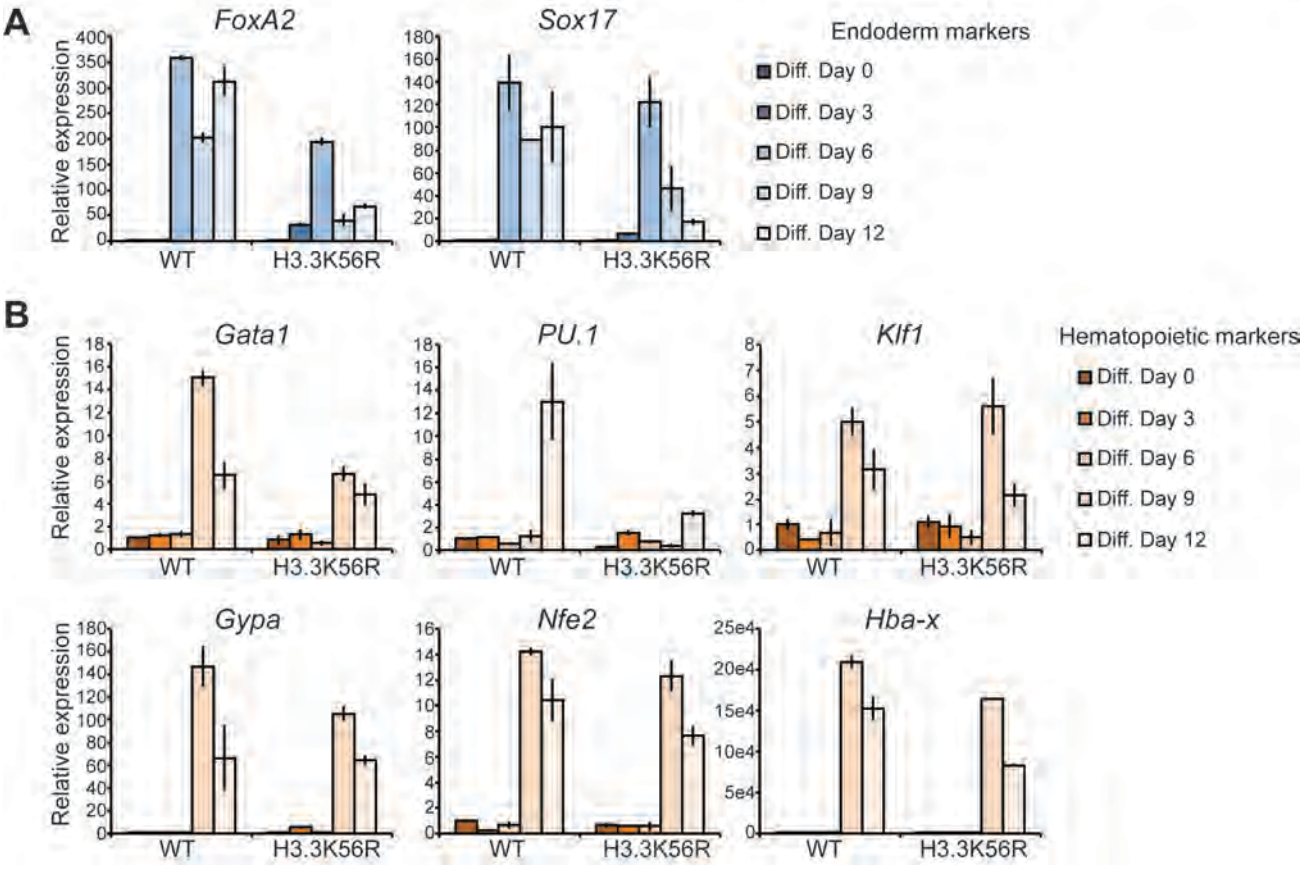


Figure 3.4: Expression of endoderm and hematopoietic lineage genes in H3.3K56R mutant ESCs

(A-B) RT-qPCR of known endoderm **(A)**, and hematopoietic **(B)** marker genes in WT and H3.3K56R mutant ESCs differentiating from EBs. Data are normalized to WT Diff. Day 0. Error bars represent +/- standard deviation of three technical replicates. Representative data from at least two biological replicate experiments are shown.

H3.3K56R mutants are defective in neuronal differentiation

We next tested whether H3.3K56R mutants were defective for ectodermal differentiation. To induce neuronal differentiation we added retinoic acid (RA) to EB cultures at Day 3. Addition of RA enhanced levels of ectoderm markers *Nestin* and *Sox11* in both WT and H3.3K56R mutants (Figure 3.5A). As few neurons could be visually detected in EB cultures we attempted to enrich for neurons through an alternative RA induction where EBs were cultured from hanging drops and individually incubated with RA for 14 days (Jiang et al., 2011). Cultures were stained for the postmitotic neuronal marker TUJ1 (Figure 3.5B, top). We observed decreased numbers of TUJ1-positive cells in H3.3K56R cultures (Figure 3.5B, bottom), with positive cells frequently exhibiting abnormal axons. These data complement data from Etchegaray et al., which suggest that H3K56 hyperacetylation in *Sirt6* KO cells pushes differentiation towards the ectoderm lineage (Etchegaray et al., 2015). If H3K56ac is important for ectoderm formation we would hypothesize that H3.3K56Q mutants would behave similarly to the *Sirt6* KO, although it is also possible that stabilization of the pluripotency factors could lead to differentiation defects, particularly if the upregulated ectodermal differentiation observed in the *Sirt6* KO is due to constitutive H3K9ac or acetylation of other SIRT6 targets. A third possibility, as stated previously for other germ layers, is that the H3K56Q phenotype can be detected only if the H3.1 and H3.2 variants are mutated along with H3.3. For future studies it will also be necessary to optimize the neuronal differentiation protocol, as fewer than 20% of WT cells were TUJ1-positive. It is likely that improved induction can be

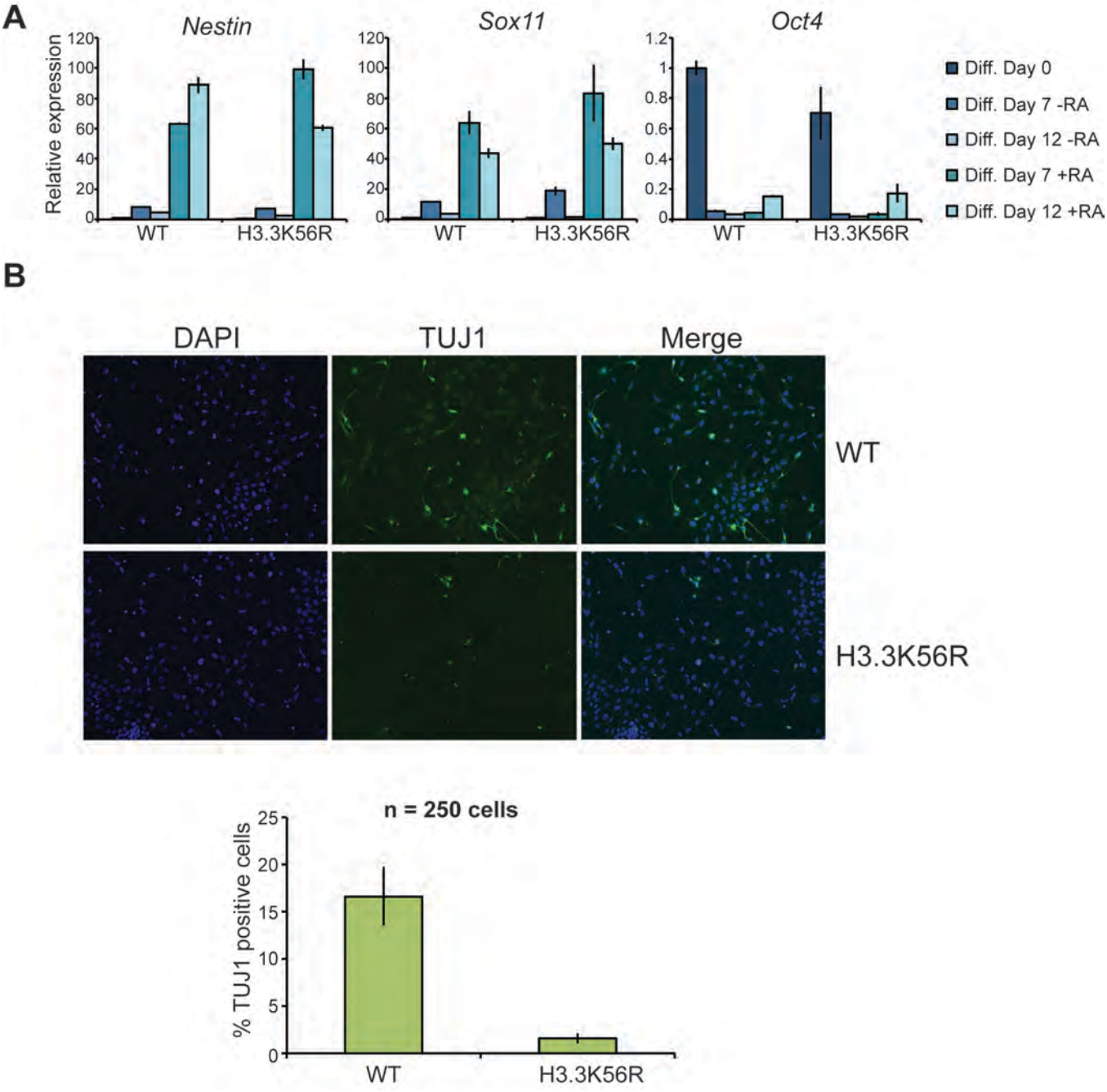
achieved by addition of RA earlier in the timecourse (Rohwedel et al., 1999) or by titrating RA concentration.

While it will be much more informative to repeat the experiments here with an H3.3K56Q mutant, our data are consistent with previous work indicating that H3K56ac interacts with OCT4, SOX2, and NANOG to maintain ESC pluripotency. We also show that H3.3K56ac is important during neural differentiation although further ChIP-Seq or RNA-Seq studies will be necessary at later stages of differentiation to ascertain whether H3.3K56ac binds and regulates neural progenitor genes, or is indirectly required. Although ChIP-Seq experiments are currently limited by the lack of availability of non-crossreacting H3K56ac antibodies (Pal et al., 2016), it is possible that H3.3K56ac could regulate gene expression through H3.3 or H2A.Z. As H3K56Q yeast mutants have decreased H2A.Z at promoters (Watanabe et al., 2013), and H2A.Z mutant ESCs harbor differentiation-defective phenotypes (Creyghton et al., 2008; Hu et al., 2013) it is likely that H2A.Z function is compromised in H3.3K56Q mutants. Thus, future studies will aim to elucidate the mechanisms by which K56ac, pluripotency factors, and histone variants interact to regulate ESC function.

Figure 3.5: H3.3K56R mutant ESCs are defective in neuronal differentiation

(A) RT-qPCR of ectoderm markers (left and center), and OCT4 (right) in EB differentiation timecourses of WT and H3.3K56R mutant ESCs. Retinoic acid (RA) was added at Diff. Day 3 to induce neuronal differentiation in +RA samples. Data are normalized to WT Diff. Day 0 and error bars represent +/- standard deviation of three technical replicates. Representative data from at least two biological replicates are shown. **(B)** TUJ1 (green) and DAPI (blue) staining of H3.3K56R and WT ESCs neuronally differentiated using the hanging drop method and RA (Top panel). Quantification of TUJ1-positive cells (bottom panel). 250 cells were counted for each of two biological replicates of WT and H3.3K56R. Error bars represent +/- standard deviation from two biological replicates.

Figure 3.5 H3.3K56R mutant ESCs are defective in neuronal differentiation



MATERIALS AND METHODS

Generation of H3.3K56R mutant ESC lines

H3.3K56R mutant ESCs were made using CRISPR/Cas9 genome editing (Cong et al., 2013) as described previously (Hainer et al., 2015). Guide RNAs targeting K56 in *H3f3a* and *H3f3b* were cloned into pX330-puro^{R}. K56R repair templates for *H3f3a* and *H3f3b* were synthesized as gblocks (Integrated DNA Technologies) containing ~1kb of homology and mutated PAM sites were cloned into pCR2.1. The cloned guide RNA and repair template vectors were transfected into E14 ESCs as described above. Clones were screened by TOPO cloning and sequencing. *H3f3a* and *H3f3b* were targeted sequentially. See Table 3.1 for guide RNA sequences.

Cell Culture and Embryoid Body Differentiation

E14 ESCs were maintained and EB differentiation was performed as described (Chapter II; (Ee et al., 2017)).

RT-qPCR

Total RNA from ESCs was prepared as described (Chapter II; (Ee et al., 2017)). 1µg of total RNA was reverse-transcribed to cDNA using random hexamers (Promega). 1µl of cDNA was amplified for each RT-qPCR technical replicate using FAST SYBR mix (KAPA Biosystems) as described (Hainer et al., 2015). Three technical replicates were performed per sample. GAPDH was used as a loading control. See Table 3.1 for primer sequences.

Histone Extraction and Western Blotting

Histones were prepared using the Gozani lab protocol: http://web.stanford.edu/group/gozani/cgi-bin/gozanilab/wp-content/uploads/2014/01/Histone-extraction-protocol.pdf. Briefly, cells were washed twice in cold PBS, resuspended in TEB buffer (1X PBS; 0.5% Triton X-100; HALT protease inhibitors), lysed on ice for 10 minutes, pelleted at 2,000rpm, washed once in TEB, and resuspended in 0.2N HCl. Pellets were incubated overnight at 4°C, and supernatants collected by centrifugation.

Whole cell lysates were prepared from EB timecourses using We16th buffer (25mM Tris pH 7.5; 125mM NaCl; 2.5mM EDTA; 0.05% SDS; 0.5% NP-40; 10% w/v glycerol). The following antibodies were used for western blotting: anti-OCT4 (Santa Cruz sc-8628), anti-NANOG (Bethyl A300-398A), anti-β-ACTIN (Sigma A1978), anti-H3 (Abcam ab1791), anti-H3K56ac (Active Motif 39281).

Retinoic acid and Hanging drop Neuronal Differentiation

For neuronal differentiation of ESCs in suspension culture, EBs were formed by plating in ESC media without LIF as described in Chapter II and 0.1μM all-trans RA (Sigma R2625) was added at Diff. Day 3. Cells were maintained in RA for 9 days and harvested for total RNA collection at the indicated timepoints.

Hanging drop neuronal differentiation was performed as described (Jiang et al., 2011). Briefly, EBs were formed in ESC media without LIF by hanging drop for 2 days and transferred to 96 well Ultralow attachment plates (Corning) for 3

days. EBs were replated in ungelatinized 48 well plates for 11 days in ESC media without LIF and 0.1μM all-trans RA. 50,000 cells were replated onto chamber slides (Lab-Tek) overnight for TUJ1 staining.

Immunofluorescence

Cultures from neuronal differentiation in chamber slides (see above) were fixed with 4% paraformaldehyde and immunofluorescence performed as described (Chen et al., 2015) using 1:400 anti-TUJ1 (SigmaT8660) and 1:1000 Alexa Fluor 488 rabbit anti-mouse (Life Technologies) and DAPI. Cells were imaged on a Nikon Eclipse E400 microscope. For quantification, 250 stained cells were counted each for WT and H3.3K56R.

Directed Cardiomyocyte Differentiation

Cardiomyocyte differentiation was performed as described (Kattman et al., 2011; Wamstad et al., 2012) with modifications. WT E14 and H3.3K56R ESCs were cultured for two days in serum-free media (3 parts IMDM (Cellgro 15-016-CV); 1 part Ham's F12 (Cellgro 10-080-CV); 0.05% BSA; 2mM L-glutamine; 50μg/ml ascorbic acid; N2B27 supplement without vitamin A (GIBCO); 4.5×10^{-4}M monothioglycerol) to induce EB formation. EBs were dissociated with TrypLE (Invitrogen) and reaggregated in 5ng/ml human VEGF (PeproTech 100-20); 8ng/ml human Activin A (PeproTech 120-14E), and 0.5ng/ml human BMP4 (PeproTech 120-05ET) for 40 hours. EBs were dissociated again and replated in gelatinized 48-well plates in StemPro-34 (GIBCO 10639011); 2mM L-glutamine;

5ng/ml VEGF, 10ng/ml human FGF-basic (PeproTech 100-18C) and 25ng/ml human FGF10 (PeproTech 100-26). Beating cardiomyocytes in were first observed in WT cells at differentiation day 8. Cells were harvested for RT-qPCR at differentiation days 0, 4, and 10.

Table 3.1: Oligonucleotides used in Chapter III

Name	Sequence	Purpose
H3.3a_gRNA-f	[Phos]CACCGACTTTTTAGGCCTGGTACTG	H3.3a (H3f3a) guideRNA for knocking in K56 mutations
H3.3a_gRNA-r	[Phos]AAACCAGTACCAGGCCTAAAAAGTC	H3.3a (H3f3a) guideRNA for knocking in K56 mutations
H3.3b_gRNA-f	[Phos]CACCGTGACTGCAGGCCAGGGACCG	H3.3b (H3f3b) guideRNA for knocking in K56 mutations
H3.3b_gRNA-r	[Phos]AAACCGGTCCCTGGCCTGCAGTCAC	H3.3b (H3f3b) guideRNA for knocking in K56 mutations
H3.3aK56chk+	AGCTTTAGGCATTGCTTTCAAC	PCR for checking K56 mutation knock-in in H3f3a
H3.3aK56chk-	ACTAGGCAGCCTCTACCAACAG	PCR for checking K56 mutation knock-in in H3f3a
H3.3bK56chk+	TCCCTTCTGCGTATTAGCAACT	PCR for checking K56 mutation knock-in in H3f3b
H3.3bK56chk-	CCTTGAACGTCGCTTGTCTC	PCR for checking K56 mutation knock-in in H3f3b
Oct4-2 F	AGAGGGAACCTCCTCTGAGC	RT-qPCR primer
Oct4-2 R	TTCTAGCTCCTTCTGCAGGG	RT-qPCR primer
Nanog_RTqPCRf	ATTCTTGCTTACAAGGGTCTGC	RT-qPCR primer
Nanog_RTqPCRr	TTGAGAGCTTTTGTTTGGGACT	RT-qPCR primer
Brachyury-F	CCAAGGACAGAGAGACGGCT	RT-qPCR primer
Brachyury-R	AGTAGGCATGTTCCAAGGGC	RT-qPCR primer
Mesp1 F	TTGTCCCCTCCACTCTTCAG	RT-qPCR primer
Mesp1 R	AGAAACAGCATCCCAGGAAA	RT-qPCR primer
Flk1-F	GCTTGCTCCTTCCTCATCTC	RT-qPCR primer
Flk1-R	CCATCAGGAAGCCACAAAGC	RT-qPCR primer
Nkx2-5 F	GGCTTTGTCCAGCTCCACT	RT-qPCR primer
Nkx2-5 R	CATTTTACCCGGGAGCCTAC	RT-qPCR primer
Tbx5 F	GGCAGTGATGACCTGGAGTT	RT-qPCR primer
Tbx5 R	TGGTTGGAGGTGACTTTGTG	RT-qPCR primer
Gata4 F	TGATAGAGGCCACAGGCATT	RT-qPCR primer
Gata4 R	CTGGAAGACACCCCAATCTC	RT-qPCR primer
Myl7 F	CTCTTCCTTGTTCACCACCC	RT-qPCR primer
Myl7 R	CTCACACTCTTCGGGGAGAA	RT-qPCR primer
Tnnt2 F	GTGTGCAGTCCCTGTTCAGA	RT-qPCR primer
Tnnt2 R	ACCCTCAGGCTCAGGTTCA	RT-qPCR primer
FoxA2-F	CCCTACGCCAACATGAACTCG	RT-qPCR primer

FoxA2-R	GTTCTGCCGGTAGAAAGGGA	RT-qPCR primer
Sox17 f	CTCGGGGATGTAAAGGTGAA	RT-qPCR primer
Sox17 r	GCTTCTCTGCCAAGGTCAAC	RT-qPCR primer
Gata1 F	AGGGCAGAATCCACAAACTG	RT-qPCR primer
Gata1 R	CCACTAAGGTGGCTGAATCC	RT-qPCR primer
PU.1 F	TGCAGCTCTGTGAAGTGGTT	RT-qPCR primer
PU.1 R	AGCGATGGAGAAAGCCATAG	RT-qPCR primer
Klf1 F	GAGCGAACCTCCAGTCACAG	RT-qPCR primer
Klf1 R	TACTCCAAGAGCTCGCACCT	RT-qPCR primer
Gypa F	CCAATGTGTGGTGAGACAGG	RT-qPCR primer
Gypa R	CCAAGAAGAGCATTCACCATC	RT-qPCR primer
Nfe2 F	CAGGTCTCCACAAGCACAAA	RT-qPCR primer
Nfe2 R	CCAGCCTCTCAGGGACACTA	RT-qPCR primer
Hba-x F	GTAGGTCTTCGTCTGGGGGT	RT-qPCR primer
Hba-x R	TCATCATGTCCATGTGGGAG	RT-qPCR primer
Nestin F	TGGCACACCTCAAGATGTCCCTTA	RT-qPCR primer
Nestin R	AAGGAAATGCAGCTTCAGCTTGGG	RT-qPCR primer
Sox11-F	ACGACCTCATGTTCGACCTGAGCT	RT-qPCR primer
Sox11-R	CACCAGCGACAGGGACAGGTTC	RT-qPCR primer

CHAPTER IV: CONCLUSIONS AND DISCUSSION

Nearly two decades ago, the canonical NuRD complex was isolated and characterized in four separate studies as a transcriptional repressor with a unique dual chromatin remodeling/histone deacetylase activity (Wade et al., 1998; Tong et al., 1998; Zhang et al., 1998; Xue et al., 1998). Since this initial characterization, subsequent studies have largely focused on addressing three primary questions: 1) what is the stoichiometry of each NuRD subunit, and which subunits directly interact with which other subunits within the complex; 2) how is NuRD recruited to its genome-wide targets; and 3) how does the complex function at its target sites to regulate chromatin structure and gene expression? Additionally, further studies describe NuRD's function with peripheral interacting proteins in myriad tissue, organism, developmental, and disease contexts. The work in this thesis provides novel insight into questions 1) and 2), and opens additional routes of inquiry for all three questions through the characterization of a largely unstudied isoform of the NuRD subunit MBD3, MBD3C. Our work suggests that it is necessary to consider the existence of multiple subunit isoforms to precisely determine the composition of NuRD and other complexes. Although the expression of MBD3C appears to be limited to ESCs, multiple isoforms or splice variants of MBD family and numerous other chromatin regulators are known to exist (and continue to be identified with the advent of advanced RNA-Seq methods). Furthermore, we have demonstrated that a single subunit isoform can uniquely interact with WDR5, a chromatin regulator that

binds histone H3, suggesting an alternative recruitment mechanism for MBD3 beside 5hmc recognition, and for NuRD in alongside to histone binding by RBBP4/7 and recognition of methylated histones by the CHD3/4 chromodomains. Further studies will be necessary to elucidate the precise role(s) of WDR5 within NuRD (see below).

An ESC-specific MBD3 isoform and NuRD complex

We have identified a novel NuRD complex formed through an interaction between the smallest MBD3 isoform MBD3C and the histone H3 binding protein WDR5. This form of NuRD appears to be specific to ESCs, as MBD3C expression is lost as ESCs differentiate, and WDR5 expression is similarly reduced during EB formation assays (Ang et al., 2011). WDR5 and MBD3 have each been shown to be required for pluripotency (Kaji et al., 2006; 2007; Yang et al., 2014), possibly through association with OCT4 and regulation of OCT4 target genes (Ang et al., 2011; Liang et al., 2008). WDR5 is also required for ESC self-renewal (Ang et al., 2011). While WDR5 is hypothesized to be important for maintaining expression of pluripotent genes through H3K4 trimethlyation by the SET/MLL complex (Ang et al., 2011; Yang et al., 2014) our data suggest that a WDR5/MBD3C/NuRD complex functions separately from other WDR5 complexes, as we did not observe other SET/MLL, MOF, or NSL subunits interacting with MBD3. Interestingly however, we found that MBD3C interacts with the WDR5 at the same binding surface as MLL1 and histone H3, using the same conserved arginine-containing motif (Couture et al., 2006; Patel et al.,

2008). It is therefore also of interest to determine whether MBD3C can bind WDR5 only when the latter is not bound to H3/MLL1. Due to the simplicity of the WDR5 interaction motif, it is not unexpected that WDR5 is a member of numerous protein complexes. Outside of the SET/MLL protein family no amino acid is completely conserved within the motif apart from the alanine-arginine dipeptide (Figure 2.3; Patel et al., 2008). A cursory examination of the motif alignment in Figure 2.3 suggests a requirement for the two amino acids 5' to the arginine to be relatively small (or absent, as in H3), but it is still unclear which if any motif amino acids are required for MBD3C to bind WDR5 aside from Arg43.

A role for WDR5 within the NuRD complex?

Although *Mbd3c* is expressed at lower levels than the other larger *Mbd3* isoforms and does not appear to be required on its own for ESC differentiation, we show that MBD3C can function redundantly with MBD3A and MBD3B to regulate gene expression (Figure 4.1A) with less than 40 and only ~250 genes misregulated in *Mbd3c* and *Mbd3ab* KO respectively, compared with nearly 5,000 genes changed in *Mbd3abc* KO. Consistent with the idea that MBD3C can functionally compensate for the other MBD3 isoforms, a previous study found that while *Mbd3* null ESCs exhibit loss of DNA methylation (and subsequent de-repression of trophectoderm lineage-specific genes), the methylation phenotype was rescued in ESCs expressing *Mbd3c* alone (Latos et al., 2012). The MBD3C N-terminus (and therefore also the interaction with WDR5) appears to be important for MBD3C's function (Figure 4.1B), as the genes changed in *Mbd3abc*

KO overlap with almost all of the genes changed in *Mbd3cΔN−ab* KO cells. Furthermore, of the 2,627 genes classified as misregulated only in *Mbd3abc* KO cells (Figure 2.9D), 837 (roughly 32%) had $\log_2$fold change values > 0.8 in the *Mbd3cΔN−ab* KO and thus were reported as unchanged due to falling just below the $\log_2$fold change cutoff of 1. Since *Mbd3c* is identical to *Mbd3a* and *b* outside of its N-terminus, it is not entirely unexpected that MBD3C can still regulate many genes when its N-terminus is deleted. While our data show that WDR5 appears to bind promoters of most genes changed in *Mbd3abc* and *Mbd3cΔN−ab* KO cells, it is likely that WDR5 regulates a large subset of these genes through mechanisms independent of Mbd3c/NuRD, as it has been shown that WDR5 also functions as a member of the MOF acetyltransferase complex independently of MLL H3K4 methlyase activity (Dias et al., 2014; Li et al., 2012). Further studies will be necessary to determine the genomic regions specifically targeted by MBD3C and WDR5, as well as pinpoint unique functions of the individual MBD3 and WDR5 complexes. For all such studies it would be important to be able to distinguish the effects of *Wdr5* KD on MBD3/NuRD from the phenotypes caused by the loss of WDR5 function in SET/MLL or MOF complexes. It is moreover important to note that there are alternative models to the sequestration model proposed in Figure 4.1. Rather than acting as a molecular sponge that prevents WDR5 from associating with activating complexes, MBD3C/NuRD could be recruiting WDR5 to fine-tune the expression of NuRD target genes (by counteracting HDAC- or CHD-mediated repression) or using WDR5 as an additional histone-binding module. As WDR5 is known to interact with a large

subset of lncRNAs in ESCs using a binding surface opposite to that of MBD3C and H3K4 (Yang et al., 2014) it would be interesting to test whether any of these lncRNAs also bind and/or recruit MBD3C/NuRD.

Much has been made of MBD3's function as a transcriptional co-repressor; however, in agreement with transcriptome analyses from previous studies (Günther et al., 2013; Hainer et al., 2016; Luo et al., 2015) our RNA-Seq analysis indicates that MBD3 can also function as an activator, with several thousand genes downregulated in *Mbd3abc* KO ESCs. Likewise, this work is the first known example of WDR5 functioning as part of a repressive complex.

Figure 4.1 MBD3C is a redundant regulator of gene expression in ESCs

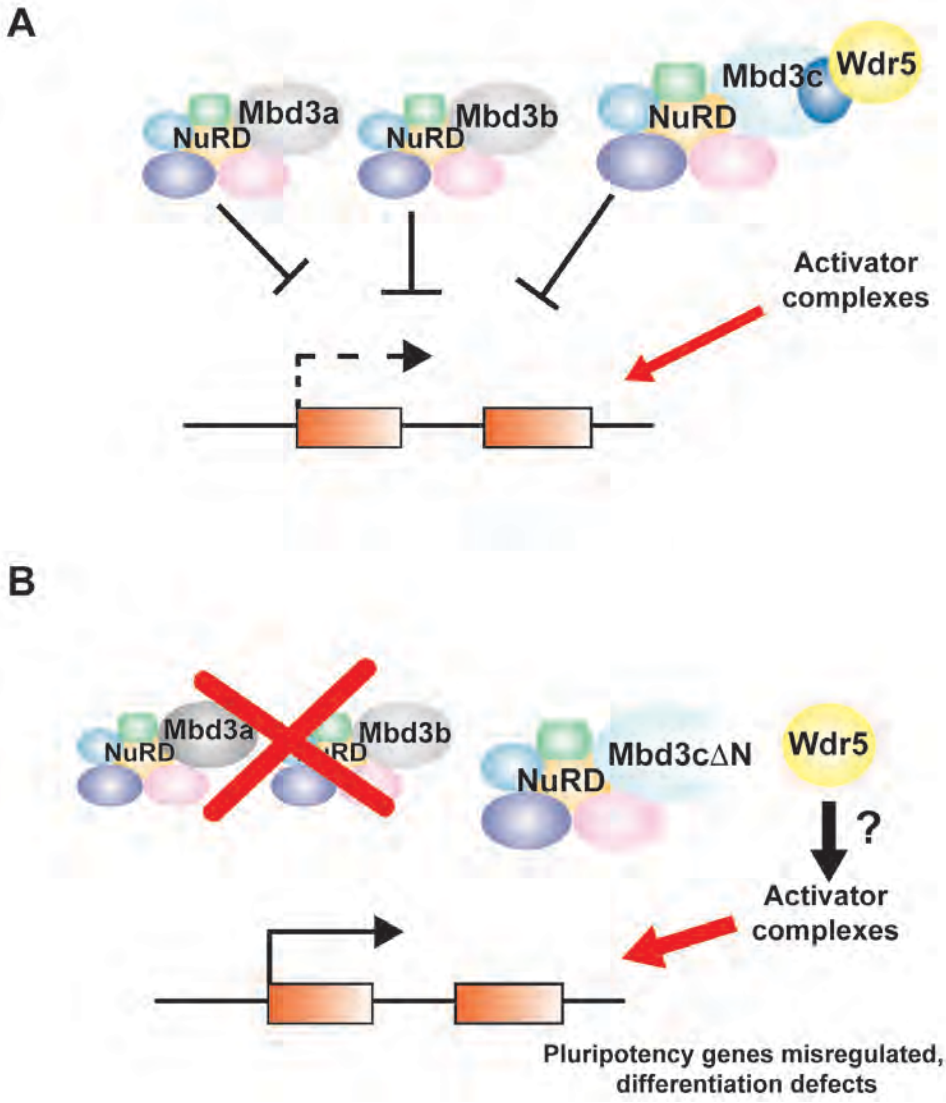


Figure 4.1: MBD3C is a redundant regulator of gene expression in ESCs
(A) MBD3C binds WDR5 via a conserved N-terminal motif (dark blue). MBD3C/NuRD is redundant in gene regulation with MBD3A/NuRD and MBD3B/NuRD. **(B)** Interaction with WDR5 is required for gene regulation by MBD3C. Pluripotency genes are derepressed if all MBD3/NuRD activity is lost. A possible scenario is that further upregulation of genes may occur through increased association of WDR5 with transcriptional activators. See text for more details.

Significance of the methyl-CpG binding domain

DNA methylation is an epigenetic modification common to vertebrates and conserved in numerous invertebrate, plant, and fungal species. Methylated genes are commonly silenced, and the question of how this repression occurs continues to be a focus of the methylation field today. While the most straightforward mechanism is that DNA methylation physically blocks or renders chromatin at gene promoters inaccessible to the transcriptional machinery, the presence of the MBD family proteins in organisms exhibiting DNA methylation implies that methylation could provide binding sites or signals for repressive chromatin remodelers, with the MBD proteins as “readers” of the methylation mark. Mammalian MBD3 contains substitutions of His30 and Phe34 for lysine and tyrosine in the MBD, resulting in the loss of the ability to bind 5mc. Although the relative affinity of MBD3 for 5hmc is still disputed in the field (Cramer et al., 2014; Hashimoto et al., 2012; Mellén et al., 2012; Spruijt et al., 2013; Yildirim et al., 2011)) our lab’s finding that TET1 catalytic activity is required for MBD3 localization in ESCs (Hainer et al., 2016; Yildirim et al., 2011) suggests that the MBD3 MBD has further evolved to recognize and recruit the NuRD complex to genomic regions undergoing demethylation.

The question of the MBD’s biological significance was first raised soon after the initial characterization of the MBD proteins. DNA methylation is essential during embryonic development (Li et al., 1992; Okano et al., 1999), but of the MBD proteins only MBD3 appears to also be necessary. As the MBD3 MBD does not bind 5mc, it was postulated that the MBD could be unimportant for either

development or for transcriptional repression. Our observation that MBD3C (which completely lacks the MBD), can functionally compensate for MBD3A and B in the regulation of gene expression in ESCs seems to cast further doubt on whether the MBD is actually essential for NuRD complex function and/or methylation-induced gene repression. It is noteworthy that the MBD3B isoform lacks almost the entire MBD but can also rescue wildtype levels of gene expression, proliferation, and methylation when expressed in *Mbd3* null ESCs (Kaji et al., 2006; Latos et al., 2012). We have further found that constitutive *Mbd3cΔN* overexpression in *Mbd3abc* KO ESCs allows for normal kinetics of OCT4 and NANOG loss during differentiation (Figure 2.7A). Taken together, these data point to a greater role for the MBD3 C-terminal region in ESC pluripotency and differentiation, and suggests a diminished importance for the MBD. Further studies would be necessary to elucidate significant functional domains within the C-terminal region. The three MBD3 isoforms are identical in sequence from exon 3 to the stop codon, which creates some functional redundancy and underscores the importance of the C-terminus. As MBD3/NuRD is required for ESC differentiation and embryonic viability it therefore seems likely that ESCs express all of the isoforms as a means to ensure proper development as much as possible, with WDR5 allowing for MBD3C to function in gene regulation as an additional backup mechanism in the absence of MBD3A and B.

Interestingly, although *Mbd3* and *Mbd2* null mice have very different phenotypes (Hendrich et al., 2001) the genome-wide localization of MBD3 and MBD2 is highly interdependent and also dependent on DNMT1 and TET1 (and

thus also on 5mc and 5hmc) in ESCs (Hainer et al., 2016). Therefore, the importance of the MBD as a NuRD recruitment module cannot be entirely dismissed, as MBD3 may require recognition of methylated DNA by MBD2's MBD to bind to a subset of targets. Furthermore, 5mc and 5hmc levels are decreased in both *Mbd3* and *Mbd2* KD cells, resulting in a regulatory loop of gene expression (Hainer et al., 2016) and consistent with previous studies showing that 5mc and 5hmc are dependent on MBD3 (Latos et al., 2012; Yildirim et al., 2011). Although our lab has shown that MBD3 and TET1 physically interact (Yildirim et al., 2011), the MBD3 isoforms that bind TET1 or 5hmc and the mechanism of 5mc and 5hmc regulation by MBD proteins are still unknown. It is therefore possible that the MBD is required for DNMT and/or TET1 function.

Why MBD3C?

Although MBD family proteins are typically found in organisms where DNA methylation is also present, MBD3C is not highly conserved. A protein BLAST search of the 50-amino acid N-terminal domain revealed significant similar sequences only among rodent species. Strikingly, amino acids 21-50 appear to be the most highly conserved, hinting at an important role for the WDR5-interacting motif in rodent development (Figure 4.2). In contrast, homology to the first 20 amino acids in MBD3C was detected only in two species of rat, *Rattus norvegicus* and *Neotoma lepida*. Thus, the evolutionary question remains unanswered: what necessitated the MBD3C isoform in rodents, or why has the isoform been lost in other species? As MBD3 is required for embryonic

development and MBD3C is largely functionally redundant with MBD3A and B in ESCs, MBD3C could have arisen simply as a backup for the other isoforms in case expression from the MBD3A/B start site was somehow lost. Since MBD3C is subsequently lost during differentiation, it is likely that MBD3 function becomes absolutely critical in the blastocyst before and in the early stages of lineage commitment. Mouse development during implantation differs slightly from that of humans, particularly in the formation of extraembryonic tissues from the trophectoderm. In the mouse the trophectoderm requires FGF4 to proliferate into extraembryonic tissues; in humans this proliferation occurs later in development and does not rely on FGF signaling (reviewed in (Rossant, 2015)). Although the expression patterns of individual MBD3 isoforms during embryogenesis have not yet been described, it is possible that MBD3C is also critical at this stage for, or oppositely to counterbalance, FGF signaling as *Mbd3* null ESCs have been shown to upregulate expression of trophectoderm markers (Kaji et al., 2006; Zhu et al., 2009).

A second difference between mouse and human ESCs is utilization of the threonine dehydrogenase (TDH) catabolism pathway. While threonine is required for growth and is catabolized by all mammalian cells as an essential amino acid, the *Tdh* gene has been shown to be inactive in humans (Edgar, 2002), while TDH enzyme is very highly expressed and active in undifferentiated mouse ESCs (Wang et al., 2009a), entailing the hypothesis that the TDH pathway facilitates the fast proliferation rate of mouse ESCs that is not observed in humans. Although we do not observe a proliferation defect in *Mbd3c* KO ESCs (Figure

2.6B), it is possible that MBD3C could also function redundantly with MBD3A and B, and evolved as a backup for the other isoforms in the regulation of the TDH pathway. ESCs expressing only MBD3C (*Mbd3ab* KO) do grow more slowly than WT, but the proliferation defect is not as severe (Figure 2.6C), and our mass spec analysis detected small numbers of TDH peptides in all MBD3 purifications (Tables 2.1 and 2.2). While it is unlikely that MBD3 directly regulates TDH as most TDH enzyme is situated in the mitochondria, it is possible that MBD3 is important for expression of the *Tdh* gene. Consistent with this idea, *Tdh* expression is significantly decreased in *Mbd3abc*, *Mbd3ab*, and *Mbd3cΔN−ab* KO cell lines (log_2fold change of -1.84, -0.57, and -1.17 respectively from our RNA-Seq data). Although MBD3C is not present in many species with a functional *Tdh* gene, it would be of interest to determine whether ESCs from these species utilize the TDH pathway to enhance proliferation, and whether MBD3 is also involved.

Figure 4.2 Protein BLAST alignment of the MBD3C N-terminus with rodent MBD3 isoforms

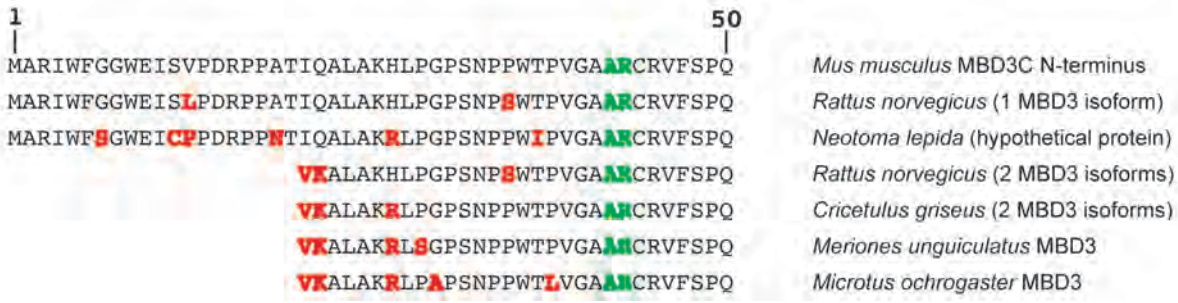


Figure 4.2: Protein BLAST alignment of the MBD3C N-terminus with rodent MBD3 isoforms

Protein BLAST alignment of the mouse 50-amino acid MBD3C N-terminus with MBD3 isoforms from other rodent species is shown. Differing amino acids are depicted in red; the alanine-arginine motif required for WDR5 interaction (Figure 2.3; Patel et al., 2008) is shown in green.

Further insight into the functions of H3K56ac in pluripotency and differentiation

While our data in Chapter III are largely preliminary and experiments are ongoing, several conclusions can still be drawn. We have shown using acetylation-depleted H3.3K56R mutants that H3.3K56ac is dispensable for ESC self-renewal, but not for differentiation, as we observe that expression of some germ layer markers is decreased and H3.3K56R mutants are defective in neuronal development. The effects of H3.3K56ac on pluripotency appear to be somewhat context-dependent, as some cardiac lineage genes downregulated in EB differentiation are expressed normally during directed differentiation with defined growth factors and serum-free culture conditions. Our data showing that ESCs depleted for H3.3K56ac are defective in differentiating into the ectoderm lineage and in forming neurons are consistent with previous data in *Sirt6* KO cells. H3K56 is hyperacetylated in the *Sirt6* KO and these cells are skewed during differentiation towards the ectoderm lineage (Etchegaray et al., 2015). We anticipate that we will observe a similar phenotype upon repeating differentiation assays using the acetylation mimic (H3.3K56Q) ESC lines, in addition to failure to downregulate pluripotency factors and decreased expression of mesoderm and endoderm markers. However, as briefly discussed in Chapter III SIRT6 targets other histone residues including H3K9 and H3K18 (Michishita et al., 2008; Tasselli et al., 2016)), and it is possible that the phenotypes observed by Etchegaray et al. result from pan-H3 hyperacetylation. Taken together, the current evidence points toward a role for H3K56ac in maintaining ESC

pluripotency in conjunction with the “master” regulators such as OCT4, with deacetylation of H3K56 necessary to exit the pluripotent state. Consistent with a shift in H3K56ac localization to developmental genes during differentiation of human ESCs (Xie et al., 2009) our data suggest that H3.3K56ac likely enhances but is not absolutely required for expression of some mesodermal and endodermal lineage markers during development. Interestingly, we did not observe mis-expression of two common ectoderm markers, *Nestin* and *Sox11* during RA-induced EB differentiation despite a clear defect in the formation of TUJ1-positive neurons using a directed differentiation protocol (Jiang et al., 2011) although it is likely *Nestin* and *Sox11* are downregulated in the directed differentiation). Optimization of our neuronal differentiation timecourse protocols and ChIP- and RNA-Seq analysis will be necessary to determine if a similar shift in H3K56ac localization occurs in mouse ESCs and which neuronal genes are affected by depletion of H3.3K56ac.

It is of particular importance to determine which phenotypes result specifically from deacetylation or hyperacetylation of H3.3 versus the core histone variants. Although targeting every endogenous copy of H3.1/2 using CRISPR is technically difficult, and knockdown of the H3K56ac acetyltransferase p300 would impact multiple acetylated sites, hyperacetylation of H3.1/2 could be studied using overexpression vectors. It is also currently unknown whether H3.3 or H3.3/H2A.Z-containing nucleosomes are enriched for K56ac. Assuming that suitable antibodies against H3K56ac can be synthesized, it would be interesting to perform ChIP-Seq of H3.3, H2A.Z and H3K56ac in parallel to test whether they

are enriched at the same genomic regions, and in H3.3K56R and H3.3K56Q mutants to determine whether localization patterns are interdependent and/or correlate with gene expression. Because the H3K56Q mutant allows the yeast SWR-C complex to promiscuously exchange H2A.Z and H2A dimers (Watanabe et al., 2013), it is possible that examining INO80 complex function in the K56R and K56Q mutants will provide more insight into H3K56ac- and H2A.Z-mediated gene regulation in ESCs and other mammalian cells. In their characterization of the *Sirt6* KO, Etchegaray et al. describe another possible mechanism for H3K56ac function. They were able to rescue the *Sirt6* KO differentiation phenotypes by knocking down the *Tet1* or *Tet2* genes, which are target genes of OCT4 and SOX2 (Koh et al., 2011; Wu et al., 2013) and function in oxidation of 5mc to 5hmc (Ito et al., 2010; Tahiliani et al., 2009). In the *Sirt6* KO, increased 5hmc as well as the activation-associated histone mark H3K4me2 were observed at ectodermal genes (Etchegaray et al., 2015), leading to a model where H3K56 deacetylation by SIRT6 during differentiation represses OCT4 and SOX2, which then prevents excessive TET-mediated demethylation and activation of ectodermal genes. The *Tet1/2* KD rescue phenotype is observed both in cell lines and in teratomas from *Sirt6* KO ESCs. While this model does not entirely explain why other lineage genes are downregulated in H3K56 hyperacetylation mutants, further examination of DNA methylation and TET enzyme function in H3.3 and H3.3K56 mutants would be informative, particularly as *Tet1* KD, like *H3.3* depletion allows ESCs to upregulate trophectodermal markers (Banaszynski et al., 2013; Koh et al., 2011). In conclusion, our work corroborates

previous studies showing that H3K56ac is important for maintaining pluripotency. Further studies are necessary to elucidate how H3K56ac affects lineage genes during differentiation, and how H3K56ac functions alongside the other H3 variants, INO80 and H2A.Z, and regulators of DNA methylation at different developmental stages.

BIBLIOGRAPHY

Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol. Cell *9*, 1191–1200.

Alexander, J.M., Hota, S.K., He, D., Thomas, S., Ho, L., Pennacchio, L.A., and Bruneau, B.G. (2015). Brg1 modulates enhancer activation in mesoderm lineage commitment. Development *142*, 1418–1430.

ALLFREY, V.G., FAULKNER, R., and MIRSKY, A.E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. Proc. Natl. Acad. Sci. U.S.a. *51*, 786–794.

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

Ang, Y.-S., Tsai, S.-Y., Lee, D.-F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., et al. (2011). Wdr5 Mediates Self-Renewaland Reprogramming via the Embryonic Stem Cell Core Transcriptional Network. Cell *145*, 183–197.

Apostolou, E., and Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. Nature *502*, 462–471.

Avdic, V., Zhang, P., Lanouette, S., Groulx, A., Tremblay, V., Brunzelle, J., and Couture, J.-F. (2011). Structural and biochemical insights into MLL1 core complex assembly. Structure *19*, 101–108.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M.,

Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532–538.

Banaszynski, L.A., Wen, D., Dewell, S., Whitcomb, S.J., Lin, M., Diaz, N., Elsässer, S.J., Chapgier, A., Goldberg, A.D., Canaani, E., et al. (2013). Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. Cell *155*, 107–120.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Bar-Nur, O., Brumbaugh, J., Verheul, C., Apostolou, E., Pruteanu-Malinici, I., Walsh, R.M., Ramaswamy, S., and Hochedlinger, K. (2014). Small molecules facilitate rapid and synchronous iPSC generation. Nat. Methods *11*, 1170–1176.

Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. Proc. Natl. Acad. Sci. U.S.a. *99*, 8695–8700.

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

Bertone, P., Hendrich, B., and Silva, J.C.R. (2015). Mbd3 and deterministic reprogramming. bioRxiv doi: 10.1101/013904.

Bhutani, N., Burns, D.M., and Blau, H.M. (2011). DNA demethylation dynamics. Cell *146*, 866–872.

Bode, D., Yu, L., Tate, P., Pardo, M., and Choudhary, J. (2016). Characterization of Two Distinct Nucleosome Remodeling and Deacetylase (NuRD) Complex Assemblies in Embryonic Stem Cells. Mol. Cell Proteomics *15*, 878–891.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536–2539.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Brackertz, M., Boeke, J., Zhang, R., and Renkawitz, R. (2002). Two highly related p66 proteins comprise a new family of potent transcriptional repressors interacting with MBD2 and MBD3. J. Biol. Chem. *277*, 40958–40966.

Brackertz, M., Gong, Z., Leers, J., and Renkawitz, R. (2006). p66alpha and

p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction. Nucleic Acids Research *34*, 397–406.

Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G., et al. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. Mol. Cell *6*, 1287–1295.

Cai, Y., Jin, J., Swanson, S.K., Cole, M.D., Choi, S.H., Florens, L., Washburn, M.P., Conaway, J.W., and Conaway, R.C. (2010). Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. Journal of Biological Chemistry *285*, 4268–4272.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development *132*, 885–896.

Chambers, I., and Smith, A. (2004). Self-renewal of teratocarcinoma and embryonic stem cells. Oncogene *23*, 7150–7160.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643–655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chelmicki, T., Dündar, F., Turley, M.J., Khanam, T., Aktas, T., Ramírez, F., Gendrel, A.-V., Wright, P.R., Videm, P., Backofen, R., et al. (2014). MOF-associated complexes ensure stem cell identity and Xist repression. eLife *3*, e02024.

Chen, C.-C., Carson, J.J., Feser, J., Tamburini, B., Zabaronick, S., Linger, J., and Tyler, J.K. (2008a). Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair. Cell *134*, 231–243.

Chen, P., Zhao, J., Wang, Y., Wang, M., Long, H., Liang, D., Huang, L., Wen, Z., Li, W., Li, X., et al. (2013a). H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. Genes Dev. *27*, 2109–2124.

Chen, P.B., Chen, H.V., Acharya, D., Rando, O.J., and Fazzio, T.G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nature Publishing Group *22*, 999–1007.

Chen, P.B., Hung, J.-H., Hickman, T.L., Coles, A.H., Carey, J.F., Weng, Z., Chu, F., and Fazzio, T.G. (2013b). Hdac6 regulates Tip60-p400 function in stem cells.

eLife *2*, e01557.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008b). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Cimmino, L., Abdel-Wahab, O., Levine, R.L., and Aifantis, I. (2011). TET family proteins and their role in stem cell differentiation and transformation. Cell Stem Cell *9*, 193–204.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu. Rev. Biochem. *78*, 273–304.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67–79.

Couture, J.-F., Collazo, E., and Trievel, R.C. (2006). Molecular recognition of histone H3 by the WD40 protein WDR5. Nat. Struct. Mol. Biol. *13*, 698–703.

Cramer, J.M., Scarsdale, J.N., Walavalkar, N.M., Buchwald, W.A., Ginder, G.D., and Williams, D.C. (2014). Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. Journal of Biological Chemistry *289*, 1294–1302.

Creyghton, M.P., Markoulaki, S., Levine, S.S., Hanna, J., Lodato, M.A., Sha, K., Young, R.A., Jaenisch, R., and Boyer, L.A. (2008). H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. Cell *135*, 649–661.

Cross, S.H., Meehan, R.R., Nan, X., and Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat. Genet. *16*, 256–259.

Das, C., Lucia, M.S., Hansen, K.C., and Tyler, J.K. (2009). CBP/p300-mediated acetylation of histone H3 on lysine 56. Nature *459*, 113–117.

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

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. Oncogene *26*, 5433–5438.

Dias, J., Van Nguyen, N., Georgiev, P., Gaub, A., Brettschneider, J., Cusack, S., Kadlec, J., and Akhtar, A. (2014). Structural analysis of the KANSL1/WDR5/KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex. Genes Dev. *28*, 929–942.

Dominski, Z., and Marzluff, W.F. (1999). Formation of the 3' end of histone mRNA. Gene *239*, 1–14.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. *16*, 1299–1309.

Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D., and Roeder, R.G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat. Struct. Mol. Biol. *13*, 713–719.

Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., Chait, B.T., Hess, J.L., and Roeder, R.G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell *121*, 873–885.

Drané, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev. *24*, 1253–1265.

Driscoll, R., Hudson, A., and Jackson, S.P. (2007). Yeast Rtt109 promotes genome stability by acetylating histone H3 on lysine 56. Science *315*, 649–652.

Drogaris, P., Villeneuve, V., Pomiès, C., Lee, E.-H., Bourdeau, V., Bonneil, E., Ferbeyre, G., Verreault, A., and Thibault, P. (2012). Histone deacetylase inhibitors globally enhance h3/h4 tail acetylation without affecting h3 lysine 56 acetylation. Sci Rep *2*, 220.

Edgar, A.J. (2002). The human L-threonine 3-dehydrogenase gene is an expressed pseudogene. BMC Genet. *3*, 18.

Ee, L.-S., McCannell, K.N., Tang, Y., Fernandes, N., Hardy, W.R., Green, M.R., Chu, F., and Fazzio, T.G. (2017). An Embryonic Stem Cell-Specific NuRD

Complex Functions through Interaction with WDR5. Stem Cell Reports *8*, 1488–1496.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D.G., Buetow, K.H., et al. (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437–447.

Egan, C.M., Nyman, U., Skotte, J., Streubel, G., Turner, S., O'Connell, D.J., Rraklli, V., Dolan, M.J., Chadderton, N., Hansen, K., et al. (2013). CHD5 is required for neurogenesis and has a dual role in facilitating gene expression and polycomb gene repression. Dev. Cell *26*, 223–236.

Elsässer, S.J., Noh, K.-M., Diaz, N., Allis, C.D., and Banaszynski, L.A. (2015). Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature *522*, 240–244.

Etchegaray, J.-P., Chavez, L., Huang, Y., Ross, K.N., Choi, J., Martinez-Pastor, B., Walsh, R.M., Sommer, C.A., Lienhard, M., Gladden, A., et al. (2015). The histone deacetylase SIRT6 controls embryonic stem cell fate via TET-mediated production of 5-hydroxymethylcytosine. Nat Cell Biol *17*, 545–557.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Fazzio, T.G., and Panning, B. (2010). Control of embryonic stem cell identity by nucleosome remodeling enzymes. Curr. Opin. Genet. Dev. *20*, 500–504.

Fazzio, T.G., Huff, J.T., and Panning, B. (2008). An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell *134*, 162–174.

Fujita, N., Jaye, D.L., Geigerman, C., Akyildiz, A., Mooney, M.R., Boss, J.M., and Wade, P.A. (2004). MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. Cell *119*, 75–86.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58–64.

Gao, Z., Huang, Z., Olivey, H.E., Gurbuxani, S., Crispino, J.D., and Svensson, E.C. (2010). FOG-1-mediated recruitment of NuRD is required for cell lineage re-enforcement during haematopoiesis. The EMBO Journal *29*, 457–468.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J., Heidersbach, A., Ramalho-Santos, J., McManus, M.T., Plath, K., Meshorer, E., et al. (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature *460*, 863–868.

Goldberg, A.D., Banaszynski, L.A., Noh, K.-M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell *140*, 678–691.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. *74*, 481–514.

Gomez, J.A., Wapinski, O.L., Yang, Y.W., Bureau, J.-F., Gopinath, S., Monack, D.M., Chang, H.Y., Brahic, M., and Kirkegaard, K. (2013). The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-γ locus. Cell *152*, 743–754.

Gong, Z., Brackertz, M., and Renkawitz, R. (2006). SUMO modification enhances p66-mediated transcriptional repression of the Mi-2/NuRD complex. Mol. Cell. Biol. *26*, 4519–4528.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295–300.

Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell *154*, 240–251.

Günther, K., Rust, M., Leers, J., Boettger, T., Scharfe, M., Jarek, M., Bartkuhn, M., and Renkawitz, R. (2013). Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. Nucleic Acids Research *41*, 3010–3021.

Hainer, S.J., and Fazzio, T.G. (2015). Regulation of Nucleosome Architecture and Factor Binding Revealed by Nuclease Footprinting of the ESC Genome. Cell Rep *13*, 61–69.

Hainer, S.J., Gu, W., Carone, B.R., Landry, B.D., Rando, O.J., Mello, C.C., and Fazzio, T.G. (2015). Suppression of pervasive noncoding transcription in embryonic stem cells by esBAF. Genes Dev. *29*, 362–378.

Hainer, S.J., McCannell, K.N., Yu, J., Ee, L.-S., Zhu, L.J., Rando, O.J., and Fazzio, T.G. (2016). DNA methylation directs genomic localization of Mbd2 and Mbd3 in embryonic stem cells. eLife *5*, 2996.

Hake, S.B., Garcia, B.A., Duncan, E.M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D.P., Allis, C.D., and Hunt, D.F. (2006). Expression patterns and post-translational modifications associated with mammalian histone H3 variants. J. Biol. Chem. *281*, 559–568.

Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S.,

Humphreys, P., Mansfield, W., Walker, R., et al. (2009). Oct4 and LIF/Stat3 additively induce Krüppel factors to sustain embryonic stem cell self-renewal. Cell Stem Cell *5*, 597–609.

Han, J., Zhou, H., Horazdovsky, B., Zhang, K., Xu, R.-M., and Zhang, Z. (2007). Rtt109 acetylates histone H3 lysine 56 and functions in DNA replication. Science *315*, 653–655.

Han, Z., Guo, L., Wang, H., Shen, Y., Deng, X.W., and Chai, J. (2006). Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. Mol. Cell *22*, 137–144.

Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Res. *21*, 396–420.

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

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

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

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol. Cell. Biol. *18*, 6538–6547.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev. *15*, 710–723.

Hendrich, B., and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet. *19*, 269–277.

Henikoff, S., Henikoff, J.G., Sakai, A., Loeb, G.B., and Ahmad, K. (2009). Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome Res. *19*, 460–469.

Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. Nature *463*, 474–484.

Ho, L., Miller, E.L., Ronan, J.L., Ho, W.Q., Jothi, R., and Crabtree, G.R. (2011). esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3

signalling and by regulating polycomb function. Nat Cell Biol *13*, 903–913.

Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proc. Natl. Acad. Sci. U.S.a. *106*, 5181–5186.

Hong, W., Nakazawa, M., Chen, Y.-Y., Kori, R., Vakoc, C.R., Rakowski, C., and Blobel, G.A. (2005). FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. The EMBO Journal *24*, 2367–2378.

Hota, S.K., and Bruneau, B.G. (2016). ATP-dependent chromatin remodeling during mammalian development. Development *143*, 2882–2897.

Hotta, A., Cheung, A.Y.L., Farra, N., Vijayaragavan, K., Séguin, C.A., Draper, J.S., Pasceri, P., Maksakova, I.A., Mager, D.L., Rossant, J., et al. (2009). Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. Nat. Methods *6*, 370–376.

Hsieh, T.-H.S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O.J. (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell *162*, 108–119.

Hu, G., Cui, K., Northrup, D., Liu, C., Wang, C., Tang, Q., Ge, K., Levens, D., Crane-Robinson, C., and Zhao, K. (2013). H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. Cell Stem Cell *12*, 180–192.

Hu, G., and Wade, P.A. (2012). NuRD and pluripotency: a complex balancing act. Cell Stem Cell *10*, 497–503.

Hyland, E.M., Cosgrove, M.S., Molina, H., Wang, D., Pandey, A., Cottee, R.J., and Boeke, J.D. (2005). Insights into the role of histone H3 and histone H4 core modifiable residues in Saccharomyces cerevisiae. Mol. Cell. Biol. *25*, 10060–10070.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129–1133.

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

Jack, A.P.M., Bussemer, S., Hahn, M., Pünzeler, S., Snyder, M., Wells, M., Csankovszki, G., Solovei, I., Schotta, G., and Hake, S.B. (2013). H3K56me3 is a novel, conserved heterochromatic mark that largely but not completely overlaps with H3K9me3 in both regulation and localization. PLoS ONE *8*, e51765.

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

Jain, A.K., Xi, Y., McCarthy, R., Allton, K., Akdemir, K.C., Patel, L.R., Aronow, B., Lin, C., Li, W., Yang, L., et al. (2016). LncPRESS1 Is a p53-Regulated LncRNA that Safeguards Pluripotency by Disrupting SIRT6-Mediated De-acetylation of Histone H3K56. Mol. Cell *64*, 967–981.

Jang, C.-W., Shibata, Y., Starmer, J., Yee, D., and Magnuson, T. (2015). Histone H3.3 maintains genome integrity during mammalian development. Genes Dev. *29*, 1377–1392.

Jiang, H., Shukla, A., Wang, X., Chen, W.-Y., Bernstein, B.E., and Roeder, R.G. (2011). Role for Dpy-30 in ES Cell-Fate Specification by Regulation of H3K4 Methylation within Bivalent Domains. Cell *144*, 513–525.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev. *21*, 1519–1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark “nucleosome-free regions” of active promoters and other regulatory regions. Nat. Genet. *41*, 941–945.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat. Genet. *19*, 187–191.

Kaji, K., Caballero, I.M., MacLeod, R., Nichols, J., Wilson, V.A., and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat Cell Biol *8*, 285–292.

Kaji, K., Nichols, J., and Hendrich, B. (2007). Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. Development *134*, 1123–1132.

Kaplan, T., Liu, C.L., Erkmann, J.A., Holik, J., Grunstein, M., Kaufman, P.D., Friedman, N., and Rando, O.J. (2008). Cell cycle- and chaperone-mediated regulation of H3K56ac incorporation in yeast. PLoS Genet. *4*, e1000270.

Kattman, S.J., Witty, A.D., Gagliardi, M., Dubois, N.C., Niapour, M., Hotta, A., Ellis, J., and Keller, G. (2011). Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. Cell Stem Cell *8*, 228–240.

Kidder, B.L., and Palmer, S. (2012). HDAC1 regulates pluripotency and lineage

specific transcriptional networks in embryonic and trophoblast stem cells. Nucleic Acids Research *40*, 2925–2939.

Kidder, B.L., Palmer, S., and Knott, J.G. (2009). SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. Stem Cells *27*, 317–328.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049–1061.

King, H.W., and Klose, R.J. (2017). The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. eLife *6*, 380.

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

Koh, F.M., Lizama, C.O., Wong, P., Hawkins, J.S., Zovein, A.C., and Ramalho-Santos, M. (2015). Emergence of hematopoietic stem and progenitor cells involves a Chd1-dependent increase in total nascent transcription. Proc. Natl. Acad. Sci. U.S.a. *112*, E1734–E1743.

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

Kolla, V., Naraparaju, K., Zhuang, T., Higashi, M., Kolla, S., Blobel, G.A., and Brodeur, G.M. (2015). The tumour suppressor CHD5 forms a NuRD-type chromatin remodelling complex. Biochem. J. *468*, 345–352.

Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. Mol. Cell. Biol. *25*, 4388–4396.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868–871.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Krogan, N.J., Keogh, M.-C., Datta, N., Sawa, C., Ryan, O.W., Ding, H., Haw, R.A., Pootoolal, J., Tong, A., Canadien, V., et al. (2003). A Snf2 family ATPase

complex required for recruitment of the histone H2A variant Htz1. Mol. Cell *12*, 1565–1576.

la Serna, de, I.L., Carlson, K.A., and Imbalzano, A.N. (2001). Mammalian SWI/SNF complexes promote MyoD-mediated muscle differentiation. Nat. Genet. *27*, 187–190.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Lai, A.Y., and Wade, P.A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. Nat. Rev. Cancer *11*, 588–596.

Landry, J.W., Banerjee, S., Taylor, B., Aplan, P.D., Singer, A., and Wu, C. (2011). Chromatin remodeling complex NURF regulates thymocyte maturation. Genes Dev. *25*, 275–286.

Landry, J., Sharov, A.A., Piao, Y., Sharova, L.V., Xiao, H., Southon, E., Matta, J., Tessarollo, L., Zhang, Y.E., Ko, M.S.H., et al. (2008). Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. PLoS Genet. *4*, e1000241.

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

Latos, P.A., Helliwell, C., Mosaku, O., Dudzinska, D.A., Stubbs, B., Berdasco, M., Esteller, M., and Hendrich, B. (2012). NuRD-dependent DNA methylation prevents ES cells from accessing a trophectoderm fate. Biology Open *1*, 341–352.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet. *32*, 42–56.

Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A.M., Cohen, A., Lasonder, E., and Stunnenberg, H.G. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol. Cell. Biol. *26*, 843–851.

Lessard, J., Wu, J.I., Ranish, J.A., Wan, M., Winslow, M.M., Staahl, B.T., Wu, H., Aebersold, R., Graef, I.A., and Crabtree, G.R. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. Neuron *55*, 201–215.

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.

Lewis, P.W., Elsaesser, S.J., Noh, K.-M., Stadler, S.C., and Allis, C.D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proc. Natl. Acad. Sci. U.S.a. *107*, 14075–14080.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

Li, Q., Zhou, H., Wurtele, H., Davies, B., Horazdovsky, B., Verreault, A., and Zhang, Z. (2008). Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. Cell *134*, 244–255.

Li, W., Wu, J., Kim, S.-Y., Zhao, M., Hearn, S.A., Zhang, M.Q., Meistrich, M.L., and Mills, A.A. (2014). Chd5 orchestrates chromatin remodelling during sperm development. Nat Commun *5*, 3812.

Li, X., Li, L., Pandey, R., Byun, J.S., Gardner, K., Qin, Z., and Dou, Y. (2012). The histone acetyltransferase MOF is a key regulator of the embryonic stem cell core transcriptional network. Cell Stem Cell *11*, 163–178.

Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S.Y., Qin, J., Wong, J., Cooney, A.J., Liu, D., et al. (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat Cell Biol *10*, 731–739.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T., and Downing, J.R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia *17*, 637–641.

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

Low, J.K.K., Webb, S.R., Silva, A.P.G., Saathoff, H., Ryan, D.P., Torrado, M., Brofelth, M., Parker, B.L., Shepherd, N.E., and Mackay, J.P. (2016). CHD4 Is a Peripheral Component of the Nucleosome Remodeling and Deacetylase Complex. Journal of Biological Chemistry *291*, 15853–15866.

Luger, K., Mäder, 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.

Luo, M., Ling, T., Xie, W., Sun, H., Zhou, Y., Zhu, Q., Shen, M., Zong, L., Lyu, G.,

Zhao, Y., et al. (2013). NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. Stem Cells *31*, 1278–1286.

Luo, Z., Gao, X., Lin, C., Smith, E.R., Marshall, S.A., Swanson, S.K., Florens, L., Washburn, M.P., and Shilatifard, A. (2015). Zic2 is an enhancer-binding factor required for embryonic stem cell specification. Mol. Cell *57*, 685–694.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Martello, G., and Smith, A. (2014). The Nature of Embryonic Stem Cells. Annual Review of Cell and Developmental Biology *30*, 647–675.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U.S.a. *78*, 7634–7638.

Martin, G.R., and Evans, M.J. (1975). Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. Proc. Natl. Acad. Sci. U.S.a. *72*, 1441–1445.

Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J., and Maltais, L.J. (2002). The human and mouse replication-dependent histone genes. Genomics *80*, 487–498.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol *9*, 625–635.

Masumoto, H., Hawke, D., Kobayashi, R., and Verreault, A. (2005). A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. Nature *436*, 294–298.

Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., and Yokota, T. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. The EMBO Journal *18*, 4261–4269.

McDonel, P., Costello, I., and Hendrich, B. (2009). Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. Int. J. Biochem. Cell Biol. *41*, 108–116.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc. Natl. Acad. Sci. U.S.a. *101*, 1525–1530.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989).

Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*, 499–507.

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

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat. Rev. Mol. Cell Biol. *7*, 540–546.

Michishita, E., McCord, R.A., Berber, E., Kioi, M., Padilla-Nash, H., Damian, M., Cheung, P., Kusumoto, R., Kawahara, T.L.A., Barrett, J.C., et al. (2008). SIRT6 is a histone H3 lysine 9 deacetylase that modulates telomeric chromatin. Nature *452*, 492–496.

Michishita, E., McCord, R.A., Boxer, L.D., Barber, M.F., Hong, T., Gozani, O., and Chua, K.F. (2009). Cell cycle-dependent deacetylation of telomeric histone H3 lysine K56 by human SIRT6. Cell Cycle *8*, 2664–2666.

Millard, C.J., Varma, N., Saleh, A., Morris, K., Watson, P.J., Bottrill, A.R., Fairall, L., Smith, C.J., and Schwabe, J.W.R. (2016). The structure of the core NuRD repression complex provides insights into its interaction with chromatin. eLife *5*, e13941.

Millard, C.J., Watson, P.J., Celardo, I., Gordiyenko, Y., Cowley, S.M., Robinson, C.V., Fairall, L., and Schwabe, J.W.R. (2013). Class I HDACs share a common mechanism of regulation by inositol phosphates. Mol. Cell *51*, 57–67.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc. Natl. Acad. Sci. U.S.a. *98*, 12902–12907.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science *303*, 343–348.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schübeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol. Cell *30*, 755–766.

Musselman, C.A., Ramírez, J., Sims, J.K., Mansfield, R.E., Oliver, S.S., Denu,

J.M., Mackay, J.P., Wade, P.A., Hagman, J., and Kutateladze, T.G. (2012). Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. Proc. Natl. Acad. Sci. U.S.a. *109*, 787–792.

Nair, S.S., Li, D.-Q., and Kumar, R. (2013). A core chromatin remodeling factor instructs global chromatin signaling through multivalent reading of nucleosome codes. Mol. Cell *49*, 704–718.

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

Neigeborn, L., and Carlson, M. (1984). Genes affecting the regulation of SUC2 gene expression by glucose repression in Saccharomyces cerevisiae. Genetics *108*, 845–858.

Neri, F., Krepelova, A., Incarnato, D., Maldotti, M., Parlato, C., Galvagni, F., Matarese, F., Stunnenberg, H.G., and Oliviero, S. (2013). Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell *155*, 121–134.

Neumann, H., Hancock, S.M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D., and Chin, J.W. (2009). A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Mol. Cell *36*, 153–163.

Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat. Genet. *23*, 58–61.

Ng, R.K., Dean, W., Dawson, C., Lucifero, D., Madeja, Z., Reik, W., and Hemberger, M. (2008). Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. Nat Cell Biol *10*, 1280–1290.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379–391.

Nishimoto, M., Fukushima, A., Okuda, A., and Muramatsu, M. (1999). The Gene for the Embryonic Stem Cell Coactivator UTF1 Carries a Regulatory Element Which Selectively Interacts with a Complex Composed of Oct-3/4 and Sox-2. Mol. Cell. Biol. *19*, 5453–5465.

Nitarska, J., Smith, J.G., Sherlock, W.T., Hillege, M.M.G., Nott, A., Barshop, W.D., Vashisht, A.A., Wohlschlegel, J.A., Mitter, R., and Riccio, A. (2016). A

Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. Cell Rep *17*, 1683–1698.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev. *12*, 2048–2060.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. *24*, 372–376.

Odho, Z., Southall, S.M., and Wilson, J.R. (2010). Characterization of a novel WDR5-binding site that recruits RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage leukemia protein-1. Journal of Biological Chemistry *285*, 32967–32976.

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

Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. (2005). Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J. Biol. Chem. *280*, 5307–5317.

Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., and Hayashi, Y. (2002). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). Cancer Res. *62*, 4075–4080.

Orkin, S.H., and Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. Cell *145*, 835–850.

Ozdemir, A., Spicuglia, S., Lasonder, E., Vermeulen, M., Campsteijn, C., Stunnenberg, H.G., and Logie, C. (2005). Characterization of lysine 56 of histone H3 as an acetylation site in Saccharomyces cerevisiae. J. Biol. Chem. *280*, 25949–25952.

Pal, S., Graves, H., Ohsawa, R., Huang, T.-H., Wang, P., Harmacek, L., and Tyler, J. (2016). The Commercial Antibodies Widely Used to Measure H3 K56 Acetylation Are Non-Specific in Human and Drosophila Cells. PLoS ONE *11*, e0155409.

Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011). Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell *144*, 200–213.

Patel, A., Vought, V.E., Dharmarajan, V., and Cosgrove, M.S. (2008). A conserved arginine-containing motif crucial for the assembly and enzymatic

activity of the mixed lineage leukemia protein-1 core complex. J. Biol. Chem. *283*, 32162–32175.

Pegoraro, G., Kubben, N., Wickert, U., Göhler, H., Hoffmann, K., and Misteli, T. (2009). Ageing-related chromatin defects through loss of the NURD complex. Nat Cell Biol *11*, 1261–1267.

Pradhan, S.K., Su, T., Yen, L., Jacquet, K., Huang, C., Côté, J., Kurdistani, S.K., and Carey, M.F. (2016). EP400 Deposits H3.3 into Promoters and Enhancers during Gene Activation. Mol. Cell *61*, 27–38.

Quan, J., Adelmant, G., Marto, J.A., Look, A.T., and Yusufzai, T. (2014). The chromatin remodeling factor CHD5 is a transcriptional repressor of WEE1. PLoS ONE *9*, e108066.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., et al. (2014). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65–70.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell *123*, 233–248.

Rajala, K., Pekkanen-Mattila, M., and Aalto-Setälä, K. (2011). Cardiac differentiation of pluripotent stem cells. Stem Cells Int *2011*, 383709–383712.

Rando, O.J. (2012). Combinatorial complexity in chromatin structure and function: revisiting the histone code. Curr. Opin. Genet. Dev. *22*, 148–155.

Ravens, S., Fournier, M., Ye, T., Stierle, M., Dembele, D., Chavant, V., and Tora, L. (2014). Mof-associated complexes have overlapping and unique roles in regulating pluripotency in embryonic stem cells and during differentiation. eLife *3*.

Rege, M., Subramanian, V., Zhu, C., Hsieh, T.-H.S., Weiner, A., Friedman, N., Clauder-Münster, S., Steinmetz, L.M., Rando, O.J., Boyer, L.A., et al. (2015). Chromatin Dynamics and the RNA Exosome Function in Concert to Regulate Transcriptional Homeostasis. Cell Rep *13*, 1610–1622.

Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., et al. (2012). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. Cell Stem Cell *10*, 583–594.

Reynolds, N., Salmon-Divon, M., Dvinge, H., Hynes-Allen, A., Balasooriya, G., Leaford, D., Behrens, A., Bertone, P., and Hendrich, B. (2011). NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. The EMBO Journal *31*, 593–605.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., and Stewart, A.F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. The EMBO Journal *20*, 7137–7148.

Rohwedel, J., Guan, K., and Wobus, A.M. (1999). Induction of cellular differentiation by retinoic acid in vitro. Cells Tissues Organs (Print) *165*, 190–202.

Rossant, J. (2015). Mouse and human blastocyst-derived stem cells: vive les differences. Development *142*, 9–12.

Rufiange, A., Jacques, P.-E., Bhat, W., Robert, F., and Nourani, A. (2007). Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. Mol. Cell *27*, 393–405.

Ruhl, D.D., Jin, J., Cai, Y., Swanson, S., Florens, L., Washburn, M.P., Conaway, R.C., Conaway, J.W., and Chrivia, J.C. (2006). Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes. Biochemistry *45*, 5671–5677.

Ruthenburg, A.J., Wang, W., Graybosch, D.M., Li, H., Allis, C.D., Patel, D.J., and Verdine, G.L. (2006). Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. Nat. Struct. Mol. Biol. *13*, 704–712.

Saito, M., and Ishikawa, F. (2002). The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. J. Biol. Chem. *277*, 35434–35439.

Santos, dos, R.L., Tosti, L., Radzisheuskaya, A., Caballero, I.M., Kaji, K., Hendrich, B., and Silva, J.C.R. (2014). MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. Cell Stem Cell *15*, 102–110.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407–411.

Sarraf, S.A., and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol. Cell *15*, 595–605.

Schneider, J., Bajwa, P., Johnson, F.C., Bhaumik, S.R., and Shilatifard, A. (2006). Rtt109 is required for proper H3K56 acetylation: a chromatin mark associated with the elongating RNA polymerase II. J. Biol. Chem. *281*, 37270–37274.

Schuettengruber, B., Martinez, A.-M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nat. Rev. Mol. Cell

Biol. *12*, 799–814.

Schuetz, A., Allali-Hassani, A., Martín, F., Loppnau, P., Vedadi, M., Bochkarev, A., Plotnikov, A.N., Arrowsmith, C.H., and Min, J. (2006). Structural basis for molecular recognition and presentation of histone H3 by WDR5. The EMBO Journal *25*, 4245–4252.

Seelig, H.P., Moosbrugger, I., Ehrfeld, H., Fink, T., Renz, M., and Genth, E. (1995). The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation. Arthritis Rheum. *38*, 1389–1399.

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844–847.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688–690.

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

Song, J.J., and Kingston, R.E. (2008). WDR5 Interacts with Mixed Lineage Leukemia (MLL) Protein via the Histone H3-binding Pocket. Journal of Biological Chemistry *283*, 35258–35264.

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

Stern, M., Jensen, R., and Herskowitz, I. (1984). Five SWI genes are required for expression of the HO gene in yeast. J. Mol. Biol. *178*, 853–868.

Stopka, T., and Skoultchi, A.I. (2003). The ISWI ATPase Snf2h is required for early mouse development. Proc. Natl. Acad. Sci. U.S.a. *100*, 14097–14102.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

Szenker, E., Ray-Gallet, D., and Almouzni, G. (2011). The double face of the histone variant H3.3. Cell Res. *21*, 421–434.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell *116*, 51–61.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-

methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Tan, Y., Xue, Y., Song, C., and Grunstein, M. (2013). Acetylated histone H3K56 interacts with Oct4 to promote mouse embryonic stem cell pluripotency. Proc. Natl. Acad. Sci. U.S.a. *110*, 11493–11498.

Tasselli, L., Xi, Y., Zheng, W., Tennen, R.I., Odrowaz, Z., Simeoni, F., Li, W., and Chua, K.F. (2016). SIRT6 deacetylates H3K18ac at pericentric chromatin to prevent mitotic errors and cellular senescence. Nature Publishing Group *23*, 434–440.

Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. Nat. Rev. Mol. Cell Biol. *15*, 703–708.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082–1086.

Tjeertes, J.V., Miller, K.M., and Jackson, S.P. (2009). Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. The EMBO Journal *28*, 1878–1889.

Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M., and Okuda, A. (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. Nucleic Acids Research *30*, 3202–3213.

Tong, J.K., Hassig, C.A., Schnitzler, G.R., Kingston, R.E., and Schreiber, S.L. (1998). Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. Nature *395*, 917–921.

Tsubota, T., Berndsen, C.E., Erkmann, J.A., Smith, C.L., Yang, L., Freitas, M.A., Denu, J.M., and Kaufman, P.D. (2007). Histone H3-K56 acetylation is catalyzed by histone chaperone-dependent complexes. Mol. Cell *25*, 703–712.

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

Vempati, R.K., Jayani, R.S., Notani, D., Sengupta, A., Galande, S., and Haldar,

D. (2010). p300-mediated acetylation of histone H3 lysine 56 functions in DNA damage response in mammals. Journal of Biological Chemistry *285*, 28553–28564.

Vidal, S.E., Amlani, B., Chen, T., Tsirigos, A., and Stadtfeld, M. (2014). Combinatorial modulation of signaling pathways reveals cell-type-specific requirements for highly efficient and synchronous iPSC reprogramming. Stem Cell Reports *3*, 574–584.

Wade, P.A., Jones, P.L., Vermaak, D., and Wolffe, A.P. (1998). A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. Curr. Biol. *8*, 843–846.

Wamstad, J.A., Alexander, J.M., Truty, R.M., Shrikumar, A., Li, F., Eilertson, K.E., Ding, H., Wylie, J.N., Pico, A.R., Capra, J.A., et al. (2012). Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage. Cell *151*, 206–220.

Wang, J., Alexander, P., Wu, L., Hammer, R., Cleaver, O., and McKnight, S.L. (2009a). Dependence of mouse embryonic stem cells on threonine catabolism. Science *325*, 435–439.

Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. Mol. Cell *43*, 904–914.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Wang, L., Du, Y., Ward, J.M., Shimbo, T., Lackford, B., Zheng, X., Miao, Y.-L., Zhou, B., Han, L., Fargo, D.C., et al. (2014). INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. Cell Stem Cell *14*, 575–591.

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

Watanabe, S., Radman-Livaja, M., Rando, O.J., and Peterson, C.L. (2013). A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. Science *340*, 195–199.

Watanabe, S., Resch, M., Lilyestrom, W., Clark, N., Hansen, J.C., Peterson, C., and Luger, K. (2010). Structural characterization of H3K56Q nucleosomes and nucleosomal arrays. Biochim. Biophys. Acta *1799*, 480–486.

Weider, M., Küspert, M., Bischof, M., Vogl, M.R., Hornig, J., Loy, K., Kosian, T.,

Müller, J., Hillgärtner, S., Tamm, E.R., et al. (2012). Chromatin-remodeling factor Brg1 is required for Schwann cell differentiation and myelination. Dev. Cell *23*, 193–201.

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.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature *336*, 684–687.

Wirbelauer, C., Bell, O., and Schübeler, D. (2005). Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. Genes Dev. *19*, 1761–1766.

Wu, H., and Zhang, Y. (2011). Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes Dev. *25*, 2436–2452.

Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. *25*, 679–684.

Wu, Y., Guo, Z., Liu, Y., Tang, B., Wang, Y., Yang, L., Du, J., and Zhang, Y. (2013). Oct4 and the small molecule inhibitor, SC1, regulates Tet2 expression in mouse embryonic stem cells. Mol. Biol. Rep. *40*, 2897–2906.

Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H., and Allis, C.D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell *121*, 859–872.

Xie, W., Song, C., Young, N.L., Sperling, A.S., Xu, F., Sridharan, R., Conway, A.E., Garcia, B.A., Plath, K., Clark, A.T., et al. (2009). Histone h3 lysine 56 acetylation is linked to the core transcriptional network in human embryonic stem cells. Mol. Cell *33*, 417–427.

Xu, F., Zhang, K., and Grunstein, M. (2005). Acetylation in histone H3 globular domain regulates gene expression in yeast. Cell *121*, 375–385.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol. Cell *2*, 851–861.

Yang, B., Zwaans, B.M.M., Eckersdorff, M., and Lombard, D.B. (2009). The

sirtuin SIRT6 deacetylates H3 K56Ac in vivo to promote genomic stability. Cell Cycle *8*, 2662–2663.

Yang, Y.W., Flynn, R.A., Chen, Y., Qu, K., Wan, B., Wang, K.C., Lei, M., and Chang, H.Y. (2014). Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. eLife *3*, e02046.

Yildirim, O., Li, R., Hung, J.-H., Chen, P.B., Dong, X., Ee, L.-S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. Cell *147*, 1498–1510.

Yip, D.J., Corcoran, C.P., Alvarez-Saavedra, M., DeMaria, A., Rennick, S., Mears, A.J., Rudnicki, M.A., Messier, C., and Picketts, D.J. (2012). Snf2l regulates Foxg1-dependent progenitor cell expansion in the developing brain. Dev. Cell *22*, 871–878.

Yoshida, T., Hazan, I., Zhang, J., Ng, S.Y., Naito, T., Snippert, H.J., Heller, E.J., Qi, X., Lawton, L.N., Williams, C.J., et al. (2008). The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation. Genes Dev. *22*, 1174–1189.

You, A., Tong, J.K., Grozinger, C.M., and Schreiber, S.L. (2001). CoREST is an integral component of the CoREST- human histone deacetylase complex. Proc. Natl. Acad. Sci. U.S.a. *98*, 1454–1458.

Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940–954.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev. *9*, 2635–2645.

Yuan, J., Pu, M., Zhang, Z., and Lou, Z. (2009). Histone H3-K56 acetylation is important for genomic stability in mammals. Cell Cycle *8*, 1747–1753.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zegerman, P., Canas, B., Pappin, D., and Kouzarides, T. (2002). Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. J. Biol. Chem. *277*, 11621–11624.

Zhang, Y., LeRoy, G., Seelig, H.P., Lane, W.S., and Reinberg, D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. Cell *95*, 279–289.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg,

D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. *13*, 1924–1935.

Zhao, H., Han, Z., Liu, X., Gu, J., Tang, F., Wei, G., and Jin, Y. (2017). The chromatin remodeler Chd4 maintains embryonic stem cell identity by controlling pluripotency- and differentiation-associated genes. Journal of Biological Chemistry *292*, 8507–8519.

Zhao, X., Su, J., Wang, F., Liu, D., Ding, J., Yang, Y., Conaway, J.W., Conaway, R.C., Cao, L., Wu, D., et al. (2013). Crosstalk between NSL histone acetyltransferase and MLL/SET complexes: NSL complex functions in promoting histone H3K4 di-methylation activity by MLL/SET complexes. PLoS Genet. *9*, e1003940.

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat. Genet. *38*, 1341–1347.

Zhu, D., Fang, J., Li, Y., and Zhang, J. (2009). Mbd3, a component of NuRD/Mi-2 complex, helps maintain pluripotency of mouse embryonic stem cells by repressing trophectoderm differentiation. PLoS ONE *4*, e7684.

Zhuang, T., Hess, R.A., Kolla, V., Higashi, M., Raabe, T.D., and Brodeur, G.M. (2014). CHD5 is required for spermiogenesis and chromatin condensation. Mech. Dev. *131*, 35–46.

Zupkovitz, G., Tischler, J., Posch, M., Sadzak, I., Ramsauer, K., Egger, G., Grausenburger, R., Schweifer, N., Chiocca, S., Decker, T., et al. (2006). Negative and positive regulation of gene expression by mouse histone deacetylase 1. Mol. Cell. Biol. *26*, 7913–7928.

Zviran, A., Rais, Y., Mor, N., Novershtern, N., Hanna, J.H. (2015). Mbd3/NuRD is a key inhibitory module during the induction and maintenance of naïve pluripotency. bioRxiv doi: 10.1101/013961.