2015-08-13

Function and Regulation of the Tip60-p400 Complex in Embryonic Stem Cells: A Dissertation

Poshen B. Chen
University of Massachusetts Medical School

Function and regulation of the Tip60-p400 complex in embryonic stem cells

A Dissertation Presented

By

Poshen Chen

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 13, 2015

INTERDISCIPLINARY GRADUATE PROGRAM

A Dissertation Presented By

POSHEN CHEN

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

Jennifer Benanti Ph.D., Member of Committee

Paul Kaufman Ph.D., Member of Committee

Erica Larschan Ph.D., Member of Committee

Craig Peterson Ph.D., Member of Committee

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

Oliver Rando M.D. 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

Program in Basic Biomedical Sciences

August 13, 2015

ACKNOWLEDGEMENTS

This work could not be possible without the help from the following people. First, I would like to thank my wife, Hsiuyi Chen. Because of her encouragement and support (both in the science and family), I could finally go through five-years graduate school. Second, I would like to thank my mentor, Tom Fazzio, for the insightful advice and guidance and I really appreciate the great help from all co-authors associated with my work. Especially, I would like to thank my chair of committee, Ollie Rando. I am always inspired by the way how he thinks about science. He is a truly great scientist and I am fortunate to have his guidance during my Ph.D training. Members of Fazzio Lab are important for this work. They always provide the great ideas for the experimental design and lots of help in the data interpretation. Finally, I apologize to anyone who contributed significantly to this work that I have forgotten.

ABSTRACT

The following work examines the mechanisms by which Tip60-p400 chromatin-remodeling complex regulates gene expression in embryonic stem cells (ESCs). Tip60-p400 complex has distinct functions in undifferentiated and differentiated cells. While Tip60-p400 is often associated with gene activation in differentiated cells, its most prominent function in ESCs is to repress differentiation-related genes. I show that Tip60-p400 interacts with Hdac6 and other proteins to form a unique form of the complex in ESCs. Tip60-Hdac6 interaction is stem cell specific and is necessary for Tip60-p400 mediated gene regulation, indicating that Tip60-p400 function is controlled in part through the regulation of Hdac6 during development. Furthermore, I find that Hdac6 is required for the binding of Tip60-p400 to many of its target genes, indicating Hdac6 is necessary for the unique function of Tip60-p400 in ESCs. In addition to accessory proteins like Hdac6, Tip60-p400 also interacts with thousands of coding and noncoding RNAs in ESCs. I show that R-loops, DNA-RNA hybrids formed during transcription of many genes, are important for regulation of chromatin binding by at least two chromatin regulators (Tip60-p400 and PRC2). This finding suggests that transcripts produced by many genes in ESC may serve as a signal to modulate binding of chromatin regulators. However, R-loops might also function to regulate chromatin architecture in differentiated cells as well. Future studies based on this work will be necessary to understand the full repertoire of cell types and chromatin regulators regulated by these structures.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS OR NOMENCLATURE

Acronym	Term
ESCs	Embryonic stem cells
NSCs	Neural stem cells
PEV	Position-effect variegation
ICM	The inner cell mass
LIF	Leukemia inhibitory factor
iPSCs	Induced pluripotent stem cells
DNaseI	Deoxyribonuclease I
TSSs	Transcription start sites
NDRs	Nucleosome-depleted regions
RdDM	RNA-directed DNA methylation
LncRNAs	Long noncoding RNAs
eRNAs	Enhancer RNAs
RED-seq	Deep sequencing analysis of restriction enzyme digestion
MNase-seq	Deep sequencing of micrococcal nuclease footprints
DHS	DNaseI hypersensitive site
DNase-seq	Deep sequencing of DHSs

FAIRE-seq	Deep sequencing of DNA fragments from FAIRE – Formaldehyde-Assisted Isolation of Regulatory Elements
MEF	Mouse embryonic fibroblast
RE	Restriction enzyme
RS	Restriction site
TSA	Trichostatin A
TubA	Tubastatin A
HPCs	Hematopoietic stem and progenitor cells
ChIP-seq	Chromatin immunoprecipitation followed by deep sequencing of the precipitated DNA
EBs	Embryoid bodies
RIP-seq	Deep sequencing of co-immunoprecipitate RNAs
DRIP-seq	Deep sequencing of Immunoprecipitation of RNA:DNA hybrids

Preface

Chapter I is reprinted from the following work with myself as the first author:

Unbiased chromatin accessibility profiling by RED-seq uncovers unique features of nucleosome variants in vivo

Poshen B. Chen, Lihua J. Zhu, Sarah J. Hainer, Kurtis N. McCannell and Thomas G. Fazzio

BMC Genomics. 2014 Dec 15; 15:1104. doi: 10.1186/1471-2164-15-1104.

*****Contributions are listed at the beginning of Chapter I in its own preface.

Chapter II is reprinted from the following work with myself as the first author:

Hdac6 regulates Tip60-p400 function in stem cells

Poshen B. Chen, Jui-Hung Hung, Taylor L. Hickman, Andrew H. Coles, James F. Carey, Zhiping Weng, Feixia Chu and Thomas G. Fazzio

Elife. 2013 Dec 3; 2:e01557. doi: 10.7554/eLife.01557

*Contributions are listed at the beginning of Chapter II in its own preface.

Chapter III is a manuscript under review with myself as the first author:

R-loops regulate ES cell fate by modulating chromatin binding of key regulatory complexes

Poshen B. Chen, Hsiuyi V. Chen, Diwash Acharya, Oliver J. Rando and Thomas G. Fazzio

*Contributions are listed at the beginning of Chapter III in its own preface.

INTRODUCTION

The genetic information in every cell is encoded within negatively charged double stranded DNA, which interacts with histones, non-histone proteins (including transcription factors and structural proteins), and non-coding RNA to form chromatin in eukaryotic cells. Nucleosomes, the basic unit of chromatin structure, are composed of two copies of each of the histone proteins H2A, H2B, H3 and H4, and approximately 147 bp of DNA wrapped in a left-handed orientation around the histone octamer (Finch et al., 1977; Luger et al., 1997; Richmond and Davey, 2003). Moreover, DNA within nucleosomes is often inaccessible to transcription factors and the accessibility of DNA is a major determinant controlling which genetic information is “read” by different factors. In mammals, different adult cells are derived from the same embryonic cells that can be cultured in the laboratory as embryonic stem cells (ESCs). Consequently, all adult cells share the same genetic information but must perform different functions in various tissues within the body. To meet this goal, different cells must express different subsets of a shared list of approximately 20,000 genes. Therefore, epigenetic regulation is important for regulation of cell type-specific transcriptional networks and, ultimately, cell fate decisions in all eukaryotes.

Epigenetics

Broadly speaking, epigenetics is the study of how cells display different phenotypes when they share the same genotype. In the nineteenth-century, Mendel discovered the laws of inheritance, whereby an organism inherits one allele from each parent, some alleles are dominant (and therefore expressed) while others are recessive, and alleles segregate independently during the formation of gametes. According to his finding, the physical appearance, phenotype, is determined by which alleles are found within an organism, their genotype, as well as the environment. However, the inheritance of many traits cannot be explained solely by Mendelian genetics. For instance, position-effect variegation (PEV), first discovered by Muller working in *Drosophila* in 1930, is observed in several species and can generate mosaic patterns of gene expression when a euchromatic gene is inserted near pericentric heterochromatin by a chromosome rearrangement. *Drosophila melanogaster* usually have red eyes due to the actions of the *white* gene. Muller found that when the *white* gene was placed next to pericentric heterochromatin after an inversion induced by X-rays, he observed variegated *white* expression. Because the heterochromatin-directed *cis*-inactivation only occurs in a proportion of cells during development, the eye color is variegated (red-white mosaic colored) due to expression of the *white* gene in some ommatidia and not in others. From that, many conserved epigenetic regulators, such as heterochromatin protein1α

(HP1α) and histone H3 lysine 9 methyltransferase (SU(VAR)3-9) were identified from genetic screens for dominant mutations that suppress or enhance *white* gene variegation (Elgin and Reuter, 2013). Many of these regulators function to regulate chromatin architecture, highlighting the importance of chromatin structure in epigenetic gene regulation.

Later, Dickinson and co-workers observed that seed size was dramatically changed simply by switching the direction of interploidy crosses in *Arabidopsis thaliana* (Scott et al., 1998). Seeds were much bigger when maternal diploids (2n) were crossed with with paternal tetraploids (4n) than when maternal tetraploids (4n) were crossed with paternal diploids (2n). The genotype in the offspring should not be changed by switching the direction of the cross under traditional Mendelian laws of inheritance, but the phenotype (seed size) was affected. Numerous similar phenomena have been observed in fungi, insects and animals, leading to the discovery that certain genes associated with genomic imprinting are expressed in a parent-of-origin-specific manner. Several mechanisms have been shown to function in genomic imprinting, including DNA methylation and histone methylation (Herrick and Seger, 1999; Martienssen and Colot, 2001; Wood and Oakey, 2006). Taken together, epigenetic regulation is important to control gene expression during development in multicellular organisms and defects in epigenetic inheritance often lead to defects during development, including birth defects or embryonic lethality.

Embryonic stem cells

In mammals, most cells in different adult tissues originate from cells within the inner cell mass (ICM) of blastocyst stage embryos. ESCs are derived from the ICM and have two unique abilities, self-renewal and pluripotency. Self-renewal is the ability to proliferate indefinitely in the same state, and pluripotency is the capacity to differentiate into all cell lineages of the adult organism, despite the fact that all cells contain the same genetic information.

Interestingly, nuclear structure was observed to be less dense in ESCs compared to differentiated cells (Park et al., 2004), and lamin A, one of the main nuclear structural proteins, was absent in ESCs (Constantinescu et al., 2006), suggesting that the establishment of a more open chromatin structure is important for ESCs. Eukaryotic chromatin is packed inside the nucleus and changes in nuclear structure can broadly affect chromatin structure leading to change gene expression. During differentiation, chromatin structure is dynamically changed. For instance, the staining of HP1α and tri-methylation of histone H3 lysine 9 (H3K9me3), a histone modification generally associated with heterochromatin, is highly dispersed in ESCs relative to the large “blocks” of heterochromatin found in somatic cells. Upon differentiation, the number of HP1α and H3K9me3 domains increases and those domains become small and discrete foci (Meshorer et al., 2006). Therefore, ESCs have relatively less condensed chromatin structure and they undergo global chromatin reorganization via

remodeling of nucleosomes and alterations in epigenetic marks during differentiation. Thus, the unique chromatin structure found in ESCs could provide a specialized platform for regulatory proteins that is uniquely adapted to maintain self-renewal and pluripotency (Efroni et al., 2008; Meshorer and Misteli, 2006).

In ESCs, the pluripotent state is mainly regulated by the core pluripotency transcription factors Oct4, Sox2, and Nanog, among others. Oct4 is a POU domain-containing transcription factor, which specifically interacts with the octamer motif (ATGCAAAT), and is present in the pre-implantation mouse embryo (Schöler et al., 1990). Interestingly, ICM cells from Oct4-deficient embryos are not pluripotent and are prone to differentiation into the trophoblast lineage (Nichols et al., 1998). Oct4 often interacts with Sox2, which binds to a neighboring Sox element, and functions as a heterodimer in ESCs (Ambrosetti et al., 2000; Avilion et al., 2003; Masui et al., 2007). Nanog, a homeodomain-containg protein, was identified from a functional screen to identify the factors required for the growth of ESCs in the absence of leukemia inhibitory factor (LIF) (Chambers et al., 2003; Mitsui et al., 2003). Although Nanog-deficient ESCs showed a defect in cell proliferation and were prone to differentiation, Nanog-deficient ESCs still maintained the expression of pluripotency markers and self-renewal *in vitro* (Chambers et al., 2007). However, within embryos, Nanog was required for formation of germ cells and necessary for pluripotency in ICM cells of blastocysts (Silva et al., 2009).

Interestingly, Oct4, Sox2 and Nanog not only co-occupy the promoter-proximal regions of their regulated protein-coding genes, but these factors also bind to their own promoters to form an interconnected autoregulatory and feed-forward loop (Boyer et al., 2005; Loh et al., 2006; Marson et al., 2008; Young, 2011). This interconnected autoregulatory loop generates a flexible ESC state which maintains pluripotency when those factors are expressed at appropriate levels, or enter into a differentiation program when any one of those factors is no longer functionally available. In general, these factors bind in combinations to the promoters of ESC genes and activate expression of genes necessary to maintain the ESC state, while contributing to repression of lineage-specific transcription factors which promote exit from the pluripotent state. Gene regulation by these factors is important to maintain pluripotency, as well as to allow normal differentiation when ESCs receive an appropriate differentiation signal from the environment (Young, 2011). It is not well understood how these core transcription factors promote globally open chromatin in ESCs while simultaneously repressing differentiation-related genes.

Recently, studies showed that the core pluripotency factors are highly interconnected to chromatin remodeling and modification complexes (Orkin and Hochedlinger, 2011), suggesting that the interaction of pluripotency factors and chromatin complexes contribute to different roles in gene regulation. For instance, Nanog, Oct4, and Sox2 have been shown to interact with several proteins including SMARCA2, SMARCA4 and BAF180 in the SWI/SNF (Brg/Brahma-

associated factors [BAF]) complex (Ho et al., 2009a; Liang et al., 2008). BAF complex in ESCs (called esBAF) exhibits a unique subunit composition defined by the presence of Brg1, BAF155 and BAF60A, and the absence of BAF170 and BAF60C. Furthermore, esBAF colocalizes with Oct4, Sox2 and Nanog throughout the ESC genome (Ho et al., 2009b), suggesting that Oct4, Sox2, and Nanog function with esBAF to regulate transcription and maintain ESC pluripotency. Consistent with this possibility, Brg1 null embryos die during the pre-implantation stage (Bultman et al., 2000). In addition to esBAF, Oct4, Sox2, and Nanog also interact with several repressive chromatin remodeling complexes, including the nucleosome remodeling and deacetylase (NuRD) repression complex, SetDB1, and polycomb repressive complex 2 (PRC2) (Bilodeau et al., 2009; Liang et al., 2008; Yeap et al., 2009; Yuan et al., 2009). Oct4 interacts with sumoylated SetDB1 to repress Cdx2, a key transcription factor in trophectoderm differentiation, and knockdown of some subunits in NuRD complex upregulates differentiation genes, indicating that the interactions with repressive complexes are important to poise or repress the expression of differentiation genes.

Induced pluripotent stem cells (iPSCs), are adult cells that have been reprogrammed to act like ESCs, have great potential in medical applications and are an excellent tool to study disease by making iPSCs from cells with disease-causing mutations. In 2006, Yamanaka and Takahashi introduced Oct4, Sox2, c-Myc and Klf4 to “reprogram” adult fibroblast cells into iPSCs (Takahashi and Yamanaka, 2006), indicating that the core pluripotency factors are sufficient to

induce pluripotency. Interestingly, the iPSCs generated from different adult cell types showed methylation patterns similar to their parental adult cells, suggesting that DNA methylation is important for cell memory during development (Kim et al., 2010a). Furthermore, the inhibition of either DNA methylation or histone deacetylation increased the efficiency of reprogramming, indicating that DNA methylation and histone deacetylation serve as the barriers for cellular reprogramming and help maintain specific cellular identities once they are established. Further research has shown that depletion of chromatin-modifying enzymes or chromatin-remodeling factors can also affect the efficiency of reprogramming, suggesting that epigenetic events regulate the ability of differentiated cells to be reprogramed (Onder et al., 2012; Rais et al., 2013). Together, these data support the idea that pluripotency factors and chromatin remodeling factors are necessary to regulate developmental processes, both during normal development and reprogramming.

Chromatin

Chromatin is composed of chromosomal DNA, histones, nonhistone proteins and non-coding RNAs. The basic unit of chromatin is the nucleosome, in which 147 bp of DNA is wrapped around an octamer of four core histones H3, H4, H2A, and H2B. Non-condensed nucleosomes (typically without the linker histone H1) appear as “beads on a string” when viewed using an electron microscope (Thoma et al., 1979), but chromatin structure in vivo is much more complex. In

order to fit approximately two meters of naked DNA (when stretched end to end) into the nucleus of a mammalian cell, DNA must be compacted at least 10,000-fold. To accomplish this, the simple “beads on a string” structures are further compacted into higher-order chromatin structure to facilitate the DNA storage process. The packaging of DNA is important not only for storage but also for the maintenance of cell identity during development. For example, Weintraub and co-workers found that globin genes in red blood cell nuclei are more accessible to deoxyribonuclease I (DNase I) digestion than in fibroblast nuclei (Weintraub and Groudine, 1976). Their study suggested that the level of gene expression is associated with the accessibility of DNA to DNase I and cells can “memorize” their identity by maintaining cell-type specific chromatin structure. Later, Wu showed that those DNase I hypersensitive sites contain regions important for gene regulation (Wu, 1980).

Overall, the interaction of DNA and histones restricts DNA accessibility, and nucleosome occupancy can be utilized to regulate various processes in the cell, such as directing enzymes where to read, replicate, and repair DNA. For example, RNA polymerase II (Pol II) initiates transcription from the 5’ end of genes, typically within nucleosome-depleted regions of highly expressed genes. Similarly, nucleosome occupancy is an important determinant for replication origin selection and function (Eaton et al., 2010). Finally, DNA repair enzymes are directed to sites of DNA damage after promoting the formation of relaxed chromatin structure around sites of DNA double-strand break by early chromatin-

based events (Price and Andrea, 2013). These examples highlight that nucleosomes are not only important for packaging the genome, but that their management is necessary to direct DNA dependent processes.

The work presented in this thesis relates to multiple features of chromatin regulation, including histone modifications/variants, nucleosome remodeling factors, and interaction with non-coding RNAs. Below, I review each of these features of chromatin regulation in detail, in order to set the stage for my own studies, presented in subsequent chapters.

Histone modifications and variants

Different covalent modifications, such as methylation, acetylation, phosphorylation, and ubiquitination, on the N-terminal or C-terminal tails of histones regulate the interactions of transcriptional regulators with chromatin. For instance, active genes are usually marked with tri-methylation of histone H3 at lysine 4 (H3K4me3) around their transcription start sites (TSS) and silent genes are usually marked with tri-methylation in histone H3 at lysine 27 (H3K27me3) (Jenuwein and Allis, 2001; Ringrose and Paro, 2004). Many lineage-specific genes in ESCs are marked with both H3K4me3 and H3K27me3 in the same regions (and are therefore called “bivalent”), where H3K27me3 helps to repress their gene expression for maintaining pluripotency while H3K4me3 is thought to poise genes for activation upon differentiation (Azuara et al., 2006; Bernstein et

al., 2006; Pan et al., 2007). Some histone modifications also regulate other histone modifications or incorporation of variant histones. For example, ubiquitination of histone H2B at lysine 123 (H2BK123ub) is required for the subsequent methylation of lysine 4 and lysine 79 of histone H3 (Briggs et al., 2002; Dover et al., 2002), and all three marks are important for transcriptional activity. In addition, Allis and co-workers showed that phosphorylation of histone H3 at serine 10 (H3S10p) is necessary for the dissociation of HP1 from chromatin in mitosis (Fischle et al., 2005). The mitotic kinase Aurora B phosphorylates H3S10 during mitosis and this phosphorylation causes loss of HP1 binding to heterochromatic regions without changing levels of H3K9me3. More recently, acetylation of histone H3 at lysine 56 (H3K56ac) was shown to affect nucleosome exchange (Kaplan et al., 2008) and facilitate the exchange of both H2A and H2A.Z (Watanabe et al., 2013).

Besides histone modifications, the composition of the nucleosome is often altered by exchange of canonical histones for various histone variants in specific locations of the genome, which can alter the properties of nucleosomes due to the distinct biophysical characteristics of some histone variants (Kamakaka and Biggins, 2005; Sarma and Reinberg, 2005). At active genes or at genes poised for activation, histones H2A and H3 are frequently replaced by histone variants H2A.Z and H3.3, respectively (Ahmad and Henikoff, 2002; Jin and Felsenfeld, 2007; Jin et al., 2009). H2A.Z has been linked to both transcriptional activation and repression. In *Tetrahymena thermophila*, hv1, a hybrid H2A variant with

properties of both vertebrate H2A.Z and H2AX (Allis et al., 1986), was found in the transcriptionally active macronucleus and only present in the micronucleus when that nucleus is transcriptionally active during early conjugation (Stargell et al., 1993), suggesting that H2A.Z may be involved in the activation of gene expression. In *Drosophila*, H2A.Z is present at heterochromatic loci in addition to euchromatin (Leach et al., 2000) and is required for the establishment of heterochromatin through Polycomb-mediated silencing (Swaminathan et al., 2005). Similar results in mammals showed that H2A.Z localizes to pericentric heterochromatin during early development (Rangasamy et al., 2003) and the depletion of H2A.Z disrupts HP1α and chromatin interactions (Fan et al., 2004; Rangasamy et al., 2004), suggesting that H2A.Z plays a role in transcriptional repression.

Similar to H2A.Z, H3.3 is also associated with both gene activation and repression. First, H3.3 can be incorporated into chromatin during DNA replication-coupled processes as well as in a DNA replication-independent manner (Ahmad and Henikoff, 2002). During S phase, GFP-H3.3 was found throughout the genome and became predominantly localized to rDNA arrays, which are sites of highly active transcription, in G2 phase of *Drosophila* Kc167 cells. On the other hand, artificial expression of GFP-H3.1 outside of S phase did not result in its incorporation into chromatin, indicating that incorporation of H3.1 is tightly coupled to DNA replication. Mutation of any of the H3.1-specific

residues to the corresponding residue in H3.3 contributed to partial replication-independent incorporation. In dividing cells, H3.3 is present at genes that are either poised for transcription or are actively transcribed. H3.3 has been proposed to replenish H3 at active genes as nucleosomes reform behind the transcribing polymerase. Recently, H3.3 was also found to accumulate in transcriptionally silent regions, such as telomeres and pericentric heterochromatin, in mouse ESCs and mouse embryonic fibroblasts (MEF) (Drané et al., 2010; Goldberg et al., 2010; Santenard et al., 2010; Wong et al., 2010), suggesting that H3.3 plays a role in gene repression. Therefore, the roles of histone modifications and variants are context-dependent and the factors that regulate these chromatin remodeling events (histone modification or incorporation of histone variants) are important for transcriptional regulation.

Chromatin remodeling complexes

Chromatin remodeling complexes utilize ATP hydrolysis to control chromatin structure by altering histone-DNA contacts, resulting in the repositioning of nucleosomes, nucleosome ejection, unwrapping DNA within nucleosomes, and exchanging variant histones for canonical histones within nucleosomes. Many chromatin remodeling complexes are evolutionarily conserved throughout eukaryotes, with homologous proteins for several core complexes identified in yeast, flies, plants and mammals, indicating that they play important roles in many different organisms (Clapier and Cairns, 2009; Hargreaves and Crabtree,

2011; Ho and Crabtree, 2010). All ATP-dependent nucleosome-remodeling protiens share a similar ATPase domain and can be separated into four different families based on the composition of different accessory domains. The SWI/SNF (switching defective/ sucrose nonfermenting) family contains the HSA (helicase-SANT) domain and C-terminal bromodomain, which recognizes acetylated histones, and functions to eject or slide nucleosomes (Mohrmann and Verrijzer, 2005). The ISWI (imitation switch) family contains a SANT domain adjacent to a SLIDE domain, which binds to unmodified histone tail and DNA (Boyer et al., 2004), and optimizes nucleosome spacing by sliding nucleosomes in either the 3' or 5' direction within the genome (Fazzio and Tsukiyama, 2003; Hamiche et al., 1999; Längst et al., 1999). The CHD (chromodomain, helicase, DNA binding) family contains N-terminal tandem chromodomains, which can interact with histone methylation, but the role of the CHD family in chromatin remodeling is less well understood.

Unlike other chromatin remodeling families, which usually appear in multi-subunit complexes, CHD family proteins can exist as monomers or dimers (Pray-Grant et al., 2005; Tran et al., 2000). CHD1 in *Saccharomyces cerevisiae* is targeted to sites of active transcription through the recognition of H3K4me3. Genome-wide mapping of CHD1 correlates with H3K4me3 and Pol II in ESCs (Flanagan et al., 2005; Gaspar-Maia et al., 2009; Sims et al., 2005), suggesting that CHD1 is associated with gene activation. However, CHD3 and CHD4 (or Mi-

2α and Mi-2β) are incorporated into the NuRD complex in flies and mammals. The NuRD complex exclusively containing either Mbd2 or Mbd3 could interact with methylated or hydroxymethylated DNA respectively and deacetylated histone to repress gene expression (Tong et al., 1998; Xue et al., 1998; Yildirim et al., 2011).

The INO80 (inositol requiring 80) family contains an HSA domain and a split ATPase domain with a long insertion present in the middle of the ATPase domain, which acts as the scaffold for the interaction with Rvb1, Rvb2 and Arp5 (Jónsson et al., 2004). The INO80 and SWR1 complexes in *Saccharomyces cerevisiae* belong to the INO80 family, but SWR1 has a unique function in regulation of nucleosome composition by exchanging canonical H2A-H2B dimer with H2A.Z-H2B dimer, while INO80 acts to slide or evict nucleosomes or replace H2A.Z with H2A (Shen et al., 2000)(Mizuguchi et al., 2004). INO80 and SWR1 also antagonistically regulate H2A.Z replacement at sites of DNA double-strand break (Papamichos-Chronakis et al., 2006).

In addition, many chromatin remodeling complexes contain several accessory subunits, which are important for recruitment to their target genes or maintenance of complex function/integrity, and the distinct subunit composition of many nucleosome remodeling complexes in different cell types provides cell-type specific functions (Ho and Crabtree, 2010). However, a key question is how chromatin remodeling complexes are recruited to target sites and how they

promote transcriptional activation or repression in vivo. Below, I discuss the functions of each family of chromatin remodeling complex, first from a mechanistic perspective regarding their roles in gene regulation, followed by a discussion of their broader roles during development.

The role of SWI/SNF family in gene regulation. Genetic studies in *Saccharomyces cerevisiae* suggested that the SWI/SNF family facilitates transcription by altering chromatin structure (Hirschhorn et al., 1992; Peterson and Herskowitz, 1992). Additionally, the mammalian SWI/SNF complex purified from HeLa cells was also shown to facilitate the binding of TATA-binding protein (TBP) or transcriptional activators to target DNA sites when target sequences reside within a nucleosome (Côté et al., 1994; Khavari et al., 1993; Kwon et al., 1994; Imbalzano et al., 1994). SWI/SNF can be recruited to target promoters by transcriptional activators and its recruitment is independent of promoter sequences, TBP or RNA Pol II holoenzyme (Wallberg et al., 2000; Yudkovsky et al., 1999). These data suggest that SWI/SNF is associated with transcriptional activation and can promote gene-specific activation either upon recruitment by different transcriptional activators or by establishing open chromatin structure in the promoter to facilitate the binding of transcriptional activators. In addition to gene activation, several studies have shown that the SWI/SNF complex is required for the transcriptional repression of some genes due to the activation of intergenic transcrips, suggesting that it controls transcription in diverse ways (Martens and Winston, 2002; Ng et al., 2002; Sudarsanam and Winston, 2000).

The role of ISWI family in gene regulation. The nucleosome remodeling factor (NURF) complex, containing the ISWI subunit, was originally purified from *Drosophila* embryo extracts (Tsukiyama and Wu, 1995; Tsukiyama et al., 1995). Unlike the ATPase activity of the SWI/SNF complex, which is stimulated by both nucleosomal DNA and free DNA, the ATPase activity of NURF is primary stimulated by assembled nucleosomes (Cairns et al., 1994; Côté et al., 1994; Laurent et al., 1993; Tsukiyama and Wu, 1995). Initially, the NURF complex was thought to play a role in transcriptional activation as it promotes transcription *in vitro* (Mizuguchi et al., 1997; Tsukiyama et al., 1995). However, studies in *Drosophila* showed a lack of overlap between ISWI protein and RNA Pol II on polytene chromosomes, suggesting that ISWI may instead play a role in transcriptional repression *in vivo* (Deuring et al., 2000). In *Saccharomyces cerevisiae,* the ISW2 complex, one of two ISWI family members found in yeast, was found to repress early meiotic genes during mitotic growth. Ume6p, a sequence-specific DNA binding protein, helps recruit ISW2 complex to target sites, and ISW2 establishes nuclease-inaccessible chromatin structure near the Ume6p binding sites. This compact chromatin structure is semi-redundant with Rpd3-Sin3-mediated histone deacetylation for repression of these genes (Goldmark et al., 2000). In addition, ISW2 complex was shown to slide nucleosomes closer to the promoter regions *in vivo* (Fazzio and Tsukiyama, 2003), suggesting that ISW2 organizes the positions and spacing of nucleosomes at its target sites instead of ejecting nucleosomes. Consistent with

these findings, ISW2 helps reposition nucleosomes and prevent antisense transcription from intergenic regions as well as initiation from cryptic initiation sites (Whitehouse et al., 2007).

The role of CHD family in gene regulation. Similar to the SWI/SNF complex, the yeast CHD1 complex remodels nucleosome structure *in vitro* in an ATP-dependent manner but ATP-dependent changes conferred by CHD1 were distinct from the SWI/SNF complex (Tran et al., 2000). Interestingly, CHD1 was shown to interact with elongation factors and localize to transcribed regions (Simic et al., 2003), suggesting that CHD1 plays a role in transcription elongation. In addition, CHD1 was identified to interact with SAGA (Spt-Ada-Gcn5 acetyltransferase) and SLIK (SAGA-like) HAT complexes and one of the two chromodomains of CHD1 specifically interacts with H3K4me3. These data suggest that CHD1 promotes transcription activation by recruiting HAT complexes to increase histone acetylation at sites with H3K4me3 (Pray-Grant et al., 2005). Recently, nucleosome organization around TSS was shown to be disrupted upon depletion of CHD1 (Gkikopoulos et al., 2011). Consistent with that, genome-wide mapping revealed that CHD1 is highly enriched in promoter-proximal regions and is responsible for RNA Pol II-directed nucleosome turnover, indicating that CHD1 removes the nucleosomal barrier to promote RNA Pol II exit from promoters (Skene et al., 2014; Zentner et al., 2013). However, Chd1-null mutantions in yeast are viable but have subtle phenotypes when grown under

special conditions (Jin et al., 1998; Tsukiyama et al., 1999; Woodage et al., 1997).

Unlike the role of CHD1 in yeast, CHD3/4 (Mi-2α and Mi-2β) is present in the NuRD repressive complex in both flies and mammals. The NuRD complex contains the histone deacetylases, HDAC1 and HDAC2, and methyl-CpG binding domain family proteins, MBD2 or MBD3, supporting their role in gene repression. However, the functions of CHD3 or CHD4 in the NuRD complex and how they regulate chromatin structure *in vivo* are poorly understood. Our lab showed that BRG1, an ATPase in the SWI/SNF family, and MBD3 antagonistically regulate nucleosome occupancy at promoter-proximal regions of their target genes in ESCs, suggesting that CHD3/4 could play a role in regulating nucleosomes *in vivo* (Yildirim et al., 2011).

The role of INO80 family in gene regulation. INO80 was shown to slide nucleosomes and SWR1 could evict histones from DNA vitro (Shen et al., 2003; Tsukuda et al., 2005). Rvb1 and Rvb2 are present in both INO80 and SWR1 complexes and their presence is important for chromatin remodeling activity *in vitro* and gene regulation *in vivo* (Jónsson et al., 2004). INO80 acts both positively and negatively to regulate transcription, suggesting that it utilizes distinct mechanisms for gene regulation. Human INO80 (hINO80) was shown to be recruited to specific target genes by YY1 (Yin-Yang-1), the GLI-Kruppel zinc-

finger transcription factor, and helps YY1 gain access to its binding sites in order to activate transcription (Cai et al., 2007).

SWR1 could regulate the composition of nucleosomes by catalyzing the exchange of H2A for H2A.Z (Krogan et al., 2003; Mizuguchi et al., 2004). In yeast, H2A.Z accumulates at both transcriptionally active and inactive regions. In certain telomere-proximal regions, the presence of H2A.Z may prevent the spread of heterochromatin from chromosome ends into gene-rich regions (Kobor et al., 2004; Meneghini et al., 2003). H2A.Z replacement is most prevalent at -1 and/or +1 nucleosomes around NDRs and these nucleosomes may be less stable than canonical nucleosomes, suggesting that H2A.Z plays a role to promote DNA exposure and gene activation (Raisner et al., 2005; Zhang et al., 2005). Interestingly, the Peterson laboratory showed that INO80 in yeast plays the opposite role in H2A.Z deposition. Instead of exchanging H2A/H2B with H2A.Z/H2B by SWR1, INO80 exchanges H2A.Z/H2B with H2A/H2B and regulates the dynamics of H2A.Z containing nucleosomes, suggesting that INO80 promotes the eviction of H2A.Z during transcriptional activation to maintain genome stability (Papamichos-Chronakis et al., 2011).

In mammals, Tip60-p400 complex, one of two SWR1 homologs in mammals, can regulate chromatin structure through the Tip60 lysine acetyltransferase (KAT), which acetylates lysine residues within the amino-terminal tails of histones H4 and H2A (and H2A variants), as well as many non-histone proteins and p400, an INO80 family ATPase, which mediates exchange

of H2A-H2B dimers for H2A.Z-H2B dimers within nucleosomes (Cai et al., 2005; Doyon et al., 2004; Squatrito et al., 2006). Tip60-p400 serves mainly as a transcriptional co-activator and interacts with numerous sequence-specific transcription factors to activate gene expression in somatic cells (Brady et al., 1999; Baek et al., 2002; Frank et al., 2003; Legube et al., 2004). In contrast, while Tip60-p400 promotes expression of some genes required for cellular proliferation and cell cycle regulation in ESCs, its most prominent function is to silence genes that are active during differentiation (Fazzio et al., 2008a). However, little is known about how Tip60-p400 represses gene expression and how Tip60-p400 complex is recruited in ESCs.

Roles of chromatin remodeling complexes during development

A common feature of chromatin remodeling complexes in multicellular organisms is that the same DNA-dependent ATPase appears in different functional complexes, indicating that they could play diverse roles in the same cell or in different cell types. For instance, BRG1, a SWI/SNF family protein, is found in BAF and PBAF complexes in mammalian cells. SNF2H, an ISWI family protein, is present in CHRAC and ACF complexes and associates respectively with Tip5, RSF1, and WSTF to form NoRC (nucleolar remodeling complex), RSF, and WICH (WSTF ISWI chromatin remodeling) complexes. CHD family members CHD3 and CHD4 are found in different NuRD complexes (Clapier and Cairns, 2009). In addition, some related complexes have distinct functions that are

mediated through combinatorial assembly of the complexes during development. The differentially incorporated subunits in these combinatorial complexes are often encoded by gene families, with different family members providing important for distinct functions of each specific complex assembly.

The SWI/SNF family during development. The diverse functions of SWI/SNF complexes are required for dynamic changes in gene expression and cell state during development. *Brm*, encoding the catalytic ATPase of BAP complex in fly, was initially identified as a repressor of *Polycomb* mutations (Kennison and Tamkun, 1988; Peterson and Tamkun, 1995; Tamkun et al., 1992) and is essential for oogenesis and embryogenesis (Brizuela et al., 1994; Tamkun, 1995). In mammals, alternative ATPase subunits of the SWI/SNF complex are encoded by *Brm* and *Brg1*, which are mutually exclusive within the complex. However, mutations in the genes encoding the two ATPases in mice have different phenotypes. *Brg1* null mice are embryonic lethal (pre-implantation) while *Brm* null mice develop normally (Bultman et al., 2000; Reyes et al., 1998), indicating that BAF complexes with Brg1 play the important role during early embryonic development. In ESCs, esBAF, which contains BRG1 but not BRM, and BAF155 but not BAF170, plays a crucial role in maintaining self-renewal and pluripotency. Additionally, several cell-type specific BAF complexes in which different BAF subunits are replaced with paralogs were identified and shown to play cell-type specific roles in the brain, heart and muscle (Ho and Crabtree, 2010; Lessard et al., 2007).

The ISWI family during development. In mammals, two ISWI ATPases are expressed: SNF2H, a component of multiple chromatin remodeling complexes, and SNF2L, which is found in NURF complex. In addition, SNF2L and SNF2H have non-overlapping protein expression patterns in mice (Dirscherl and Krebs, 2004), supporting the idea that they have different biological roles.

The several SNF2H containing complexes play different roles in biological processes, such as transcriptional activation/repression and DNA replication. SNF2H is a member of the NoRC complex, which is involved in transcriptional activation and repression while ACF, CHRAC, RSF and WICH complexes are required for the regulation of chromatin structure, the replication of DNA through heterchromatin, and the segregation of chromosomes (Dirscherl and Krebs, 2004). Likely due to its diverse roles, mice lacking SNF2H die during the pre-implantation stage because of the growth arrest and cell death within both the trophectoderm and ICM (Stopka and Skoultchi, 2003).

In contrast to SNF2H, SNF2L mutant mice grew normally, but had larger heads (Yip et al., 2012). Interestingly, BPTF (the largest subunit of NURF complex) null mice die between E7.5 and E8.5 due to defects in gastrulation while BPTF-null ESCs are viable but not able to differentiate into mesoderm and endoderm lineages (Landry et al., 2008). Taken together, these data suggest that SNF2H and SNF2L might play key roles at different time points during development.

The CHD family during development. Only one CHD family ATPase is present in yeast, whereas nine members are found in humans (Hall and Georgel, 2007), suggesting that the CHD family play different roles during the development in mammals. CHD proteins in mammals are broadly classified into three subfamilies based on their constituent domains: subfamily I (CHD1 and CHD2), subfamily II (CHD3 and CHD4), and subfamily III (CHD5, CHD6, CHD7, CHD8 and CHD9). The depletion of CHD1 by RNAi in ESCs results in the loss of proliferation, and pluripotency, and increases heterchromatin formation, suggesting that CHD1 maintains open chromatin structure in ESCs to promote pluripotency (Gaspar-Maia et al., 2009).

In contrast to CHD1, MBD3 (a subunit of NuRD complexes; subfamily II) null ESCs are viable but fail to fully commit to developmental lineages, as a result of impaired silencing of pluripotency genes (Kaji et al., 2006). Similar to BAF complexes, NuRD complexes also maintain distinct functions through combinatorial assembly. Each complex contains one MTA protein: MTA1, MTA2, or MTA3 and one MBD protein: MBD2 or MBD3. These combinations are mutually exclusive and provide distinct functions. In B lymphocytes, MTA3 containing NuRD complexes interact with BCL-6, a key regulator of B cell differentiation, to repress terminal differentiation into plasma cells until appropriate signals occur (such as the signal to trigger BCL-6 acetylation because the interaction of MTA3 and BCL-6 is sensitive to BCL-6 acetylation level) (Fujita et al., 2004). In genetic studies, the depletion of MBD3 in mice was

lethal during early embryogenesis while MBD2-null mice were viable and fertile (Hendrich et al., 2001). In addition, MBD2 binds to DNA marked with 5-methylcytosine modifications *in vitro* in contrast to MBD3, which binds to 5-hydroxymethylcytosine and unmethylated DNA, but not 5-methylcytosine (Hendrich and Bird, 1998; Saito and Ishikawa, 2002; Yildirim et al., 2011; Zhang et al., 1999). These data suggest that MBD2-NuRD and MBD3-NuRD play different roles during development, and the composition of the NuRD complexes may vary by cell type to give rise to a diverse set of complexes with multiple unique functions.

The INO80 family during development. The INO80 family contains the INO80 and SWR1 complexes in yeast and the INO80, SRCAP, and Tip60-p400 complexes in mammals. In mouse ESCs, Oct4 and Wdr5 were shown to be necessary for INO80 binding to many target genes. This binding of INO80 to chromatin promotes formation of open chromatin structure and recruitment of Mediator and RNA Pol II for gene activation (Wang et al., 2014).

Tip60-p400 participates in diverse cellular processes, such as transcriptional regulation, DNA damage repair and apoptosis (Ikura et al., 2000; Kusch et al., 2004; Legube et al., 2004; Squatrito et al., 2006; Xu and Price, 2011), and might regulate those processes through its enzymatic activities (acetylation, H2A.Z exchange, or both). H2A.Z is essential for the development of both mice and flies (Faast et al., 2001; van Daal and Elgin, 1992), leading to the

hypothesis that p400 plays key role during development by facilitating H2A.Z incorporation. Depletion of H2A.Z in ESCs did not change ESC proliferation or morphology but inhibits differentiation into certain lineages, indicating H2A.Z is necessary for lineage-specific gene expression during differentiation (Creyghton et al., 2008). Consistent with this possibility, knockdown of *Tip60* or *p400* by RNAi in ESCs causes changes in cell morphology consistent with initiation of ES cell differentiation and shows defects in both self-renewal and pluripotency (Fazzio et al., 2008a). In addition, *Tip60* and *p400* null mice die at the early embryonic stage, suggesting that Tip60-p400 complex is necessary during development (Hu et al., 2009; Ueda et al., 2007). SRCAP containing complexes are required for H2A.Z exchange in vivo (Wong et al., 2007), although the role of those complexes during development have not yet been described.

Chromatin structure and noncoding RNAs

In mouse cells, heterchromatin was found to contain at least two-fold more chromatin-associated RNA than euchromatin, suggesting that RNA could be a component of heterchromatin or regulate chromatin structure (Paul and Duerksen, 1975). In plants, small RNAs generated from transposable-element-derived tandem repeats can target homologous genomic DNA sequences and recruit a *de novo* DNA methyltransferase for cytosine methylation through a phenomenon termed RNA-directed DNA methylation (RdDM) (Lippman et al., 2004; Wassenegger et al., 1994). RdDM can serve as signal to promote

H3K9me3 and maintain heterchromatin structure (Bernstein and Allis, 2005; Law and Jacobsen, 2010), explaining the observation that the increased RNA abundance associated with heterchromatin in mice.

After the human genome sequencing project was completed and genomic maps of chromatin modifications were obtained, more than 1000 long noncoding RNAs (lncRNAs) were identified in human cells and found to associate with chromatin-modifying complexes to regulate gene expression (Khalil et al., 2009). LncRNAs are transcribed by RNA Pol II and defined as RNAs longer than 200 nucleotides, but which still typically maintain regular RNA modifications, such as a 5'-cap and polyadenylation at their 3' ends. LncRNAs are not thought to be translated into proteins, although this remains controversial (Guttman et al., 2013; Ingolia et al., 2011; Ingolia et al., 2014; Ruiz-Orera et al., 2014).

Dosage compensation of the sex chromosomes is a classic example in which lncRNAs are involved in gene regulation. In mammals, dosage compensation refers to the process whereby the expression levels of genes on the two X chromosomes in female cells are made equal to the single X chromosome in male cells. The lncRNA *Xist* is generated from one of two X chromosomes in female cells and alters the chromatin structure of the entire X chromosome, which becomes the inactive X (Xi), to achieve transcriptional silencing (Gendrel and Heard, 2014). *Xist* has been found to interact with PRC2 through a domain named Repeat A (RepA) *in vitro* and is reported to be recruited

to the target X chromosome by YY1, a sequence-specific transcription factor, for gene silencing (Jeon and Lee, 2011; Zhao et al., 2008). However, a recent study identified proteins specifically associated with *Xist* and showed that SHARP and HDAC3, but not RepA, are required for *Xist*-mediated recruitment of PRC2 *in vivo* (McHugh et al., 2015).

LncRNAs could repress gene expression by interacting with different chromatin modifying and remodeling complexes. In imprinting, the paternally and maternally inherited alleles are differentially expressed, and lncRNAs are often involved in distinguishing the two alleles. The histone H3 lysine 9 methylase G9a is thought to mediate H3K9me3 and transcriptional silencing of Kcnq1 or Igf2r loci by interacting with lncRNA *Air* or *Kcnq1ot1* respectively (Nagano et al., 2008; Pandey et al., 2008). Furthermore, lncRNAs can serve as modular scaffolds of chromatin modifying complexes. As one example, the Chang laboratory showed that *HOTAIR, a* lncRNA that silences the *HOXD* gene cluster (among other genes), interacts with two repressive complexes, PRC2 and LSD1/CoREST/REST, through the 5’ and 3’ domains of *HOTAIR* respectively (Tsai et al., 2010).

In addition to the roles of lncRNAs in gene repression, some lncRNAs transcribed bi-directionally by RNA Pol II from activating enhancers (eRNAs) were shown to correlate with mRNA levels of nearby genes. Those eRNAs usually lack poly A tails and are expressed at very low levels (De Santa et al.,

2010; Kim et al., 2010b). However, the observation that eRNAs are expressed from many enhancers raises the question of whether eRNAs merely correlate with active enhancers (as a “mistake” of RNA Pol II recruitment) or a functional component of enhancer activity. Several studies have used different approaches to explore this question including, RNA interference, DNA-RNA hybrid induced degradation via RNaseH, tethering eRNA with RNAs, or transcription inhibition by chemical inhibitors to test the functions of eRNA. These studies found that eRNAs are necessary in some cases to activate gene expression, but thsat eRNA production might occur after the assembly of active enhancers (Hah et al., 2013; Lam et al., 2013; Lam et al., 2014). Another possibility is that eRNAs might promote chromosomal looping between enhancers and TSSs. The Rosenfeld laboratory showed that knockdown of eRNAs reduced enhancer-promoter interactions and subsequently decreased coding gene activation (Li et al., 2013). However, the Kraus laboratory showed that looping at other enhancers is independent of eRNA expression (Hah et al., 2013). Additionally, eRNAs might collaborate with transcriptional activators for gene activation. Recently, *HOTTIP*, an enhancer-like lncRNA, was found to activate *HOXA* gene cluster by directly interacting with WDR5, a key component of the MLL complex that catalyzes the H3K4me3 mark (Wang et al., 2011).

Furthermore, lncRNAs were also found to play important roles during development (Guttman and Rinn, 2012; Pauli et al., 2011). From an RNAi screen, lncRNAs were found to interact with different chromatin regulators and

are essential for ESC self-renewal and pluripotency (Guttman et al., 2011). The functions of lncRNAs within chromatin remodeling enzymes are not well understood. However, roles in targeting chromatin remodeling complexes, maintaining complex architecture, or regulating complex activities have all been proposed (Guttman and Rinn, 2012; Rinn and Chang, 2012). Therefore, the investigation of how chromatin regulators interact with lncRNAs is a critical area of future research. These studies will provide an additional new layer of understanding of the mechanisms underlying transcriptional regulation.

Perspectives

Over the past 25 years, it has become clear that chromatin structure and chromatin remodeling complexes play crucial roles in cell fate. Prior to the beginning of my project, a large number of chromatin remodeling complexes had been found to play critical roles in ESC self-renewal and pluripotency. However, the rules governing how they are targeted to specific genes, and how they regulate cell fate are not well understood. In the beginning of my thesis work, I developed a method that allows one to probe chromatin accessibility within nucleosomes and non-nucleosomal DNA genome-wide, and can be used to study the functions of chromatin remodeling complexes. Subsequently, in order to better understand the functions and mechanisms of action of chromatin remodeling complexes during development, I focused on the roles of Tip60-p400 complex in control of ESC identity. These studies led me to identify a stem cell-

specific form of Tip60-p400 complex, as well as the roles of coding and non-coding RNAs in regulation of Tip60-p400 function. In the following chapters, I describe the findings I have made in these areas, as well as some unexpected findings regarding regulation of additional chromatin regulatory factors important for pluripotency.

CHAPTER I:

PREFACE

Chapter I is reprinted from the following work with myself as the first author:

Unbiased chromatin accessibility profiling by RED-seq uncovers unique features of nucleosome variants in vivo

Poshen B. Chen, Lihua J. Zhu, Sarah J. Hainer, Kurtis N. McCannell and Thomas G. Fazzio

BMC Genomics. 2014 Dec 15; 15:1104. doi: 10.1186/1471-2164-15-1104.

Contributions:

Poshen Chen and Thomas Fazzio developed the RED-seq method and designed most experiments. Poshen Chen carried out all RED-seq and ChIP experiments, as well as most data analyses. Lihua Zhu wrote the REDseq R package and performed the initial RED-seq data analysis. Sarah Hainer performed the MNase-seq experiments and initial data analysis. Kurtis McCannell performed and analyzed the qPCR experiments. Poshen Chen, Lihua Zhu, Sarah Hainer and Thomas Fazzio wrote the paper.

CHAPTER I: Unbiased chromatin accessibility profiling by RED-seq uncovers unique features of nucleosome variants in vivo

ABSTRACT

Differential accessibility of DNA to nuclear proteins underlies the regulation of numerous cellular processes. Although DNA accessibility is primarily determined by the presence or absence of nucleosomes, differences in nucleosome composition or dynamics may also regulate accessibility. Methods for mapping nucleosome positions and occupancies genome-wide (MNase-seq) have uncovered the nucleosome landscapes of many different cell types and organisms. Conversely, methods specialized for the detection of large nucleosome-free regions of chromatin (DNase-seq, FAIRE-seq) have uncovered numerous gene regulatory elements. However, these methods are less successful in measuring the accessibility of DNA sequences within nucelosome arrays. Here we probe the genome-wide accessibility of multiple cell types in an unbiased manner using restriction endonuclease digestion of chromatin coupled to deep sequencing (RED-seq). Using this method, we identified differences in chromatin accessibility between populations of cells, not only in nucleosome-depleted regions of the genome (e.g., enhancers and promoters), but also within the majority of the genome that is packaged into nucleosome arrays. Furthermore, we identified both large differences in chromatin accessibility in distinct cell lineages and subtle but significant changes during differentiation of

mouse embryonic stem cells (ESCs). Most significantly, using RED-seq, we identified differences in accessibility among nucleosomes harboring well-studied histone variants, and show that these differences depend on factors required for their deposition. We demonstrate that different types of nucleosomes within mammalian cells exhibit different degrees of accessibility. These findings provide significant insight into the regulation of DNA accessibility.

INTRODUCTION

Eukaryotic genomes are wrapped around histone octamers to form nucleosome arrays, which are further packaged into the nucleus. Although chromatin compaction facilitates storage of large quantities of DNA within small nuclear compartments, it drastically reduces the accessibility of genomic DNA to proteins that require access. Nucleosomal DNA is relatively inaccessible to DNA binding proteins due to both the occlusion of approximately half of its surface by contacts with histones, as well as the distortion of the normal B-form structure that occurs when DNA is wrapped around a histone octamer (Luger and Richmond, 1998). Consequently, chromatin structure must be disrupted to facilitate normal cellular processes, such as DNA repair, recombination, replication, and transcription.

Although protection of DNA from nuclear factors by the formation of tight interactions with histones appears to be the major method by which DNA

accessibility is regulated, many different isoforms of the histone octamer exist within most eukaryotes, each with distinct biochemical and biophysical properties (Abbott et al., 2001; Bao et al., 2004; Doyen et al., 2006; Thambirajah et al., 2006; Jin and Felsenfeld, 2007; Luger et al., 2012; Watanabe et al., 2013). These differences are mainly derived from two sources. First, most eukaryotes express several variants each of histones H2A and H3. Within each family, differences between variants can range from a few amino acid substitutions to the presence or absence of additional, non-histone domains at their amino- or carboxyl-termini. Second, all four core histone proteins are subject to a wide array of post-translational modifications, including acetylation, methylation, phosphorylation, ubiquitylation, and others. Several of these modifications and variants change the overall charge of the histone octamer and/or create or destroy binding sites for proteins, resulting in alterations in nucleosome stability (Thambirajah et al., 2006; Li et al., 1993; Wang and Hayes, 2008; Chandrasekharan et al., 2010). Together, these differences in nucleosome structure and stability conferred by histone variants and modifications raise the possibility that accessibility of nucleosomal DNA may not be a simple binary phenomenon in which nucleosome-bound DNA is completely protected and nucleosome-free DNA is completely accessible; rather, DNA within some variants of nucleosomes may be more accessible than DNA bound by other variants. For example, nucleosomes harboring histone variants H2A.Z and/or H3.3 are extractable from bulk chromatin at lower salt and, in some cases,

protect smaller footprints of DNA from nucleases than canonical nucleosomes (Jin and Felsenfeld, 2007; Henikoff et al., 2009; Jin et al., 2009; Tolstorukov et al., 2009), raising the possibility that DNA within certain nucleosome variants is more broadly accessible, due to either biophysical properties or dynamic behavior of these nucleosomes. However, this possibility remains to be directly tested *in vivo*.

Along with differences in chromatin structure within distinct genomic regions in individual cell types, cell type-specific chromatin structural differences facilitate gene expression patterns specific to cells of different lineages (Weintraub and Groudine, 1976). In embryonic stem cells (ESCs), chromatin structure is relatively open (less heterochromatic) compared to differentiated cells, which may be necessary for their ability to self-renew (proliferate as ESCs) while maintaining the flexibility to turn on lineage-specific genes during differentiation (Meshorer and Misteli, 2006; Meshorer, 2007). As ESCs differentiate, DNA accessibility decreases, chromatin becomes less dynamic, and larger blocks of heterochromatin form, suggesting that differentiation induced chromatin alterations may stabilize cell fates by “locking down” regions of the genome in heterochromatic blocks that are relatively insensitive to transcriptional activators.

Methods have been developed to study DNA accessibility based on either the protection of nucleosomal DNA from general endonuclease digestion or the

differential solubility properties of open and closed chromatin. Deoxyribonuclease I (DNase I) (Crawford et al., 2006; Sabo et al., 2006) preferentially digests nucleosome-free DNA (Wu, 1980; Saragosti et al., 1980; Schones et al., 2008), and genomic regions that are more sensitive to DNase I digestion – called DNase I hypersensitive sites (DHSs) – can be identified by deep sequencing (DNase-seq) (Boyle et al., 2008). Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) is a second method to isolate accessible genomic regions, which uses organic extractions of formaldehyde cross-linked chromatin to enrich protein-free DNA fragments that are subsequently identified by microarrays (FAIRE-chip) (Giresi et al., 2007) or high-throughput sequencing (FAIRE-seq) (Waki et al., 2011). Consistent with the requirement of most transcription factors (TFs) for accessible binding sites on DNA, DHSs and FAIRE-seq peaks are enriched for regulatory regions of active genes (enhancers and promoters). Conversely, micrococcal nuclease digestion of chromatin followed by deep sequencing of the regions of DNA protected from digestion (MNase-seq) allows inference of the positions and occupancy levels of nucleosomes in a population (when footprints of ~150 bp are quantified) and TFs (when footprints less than ~80 bp are considered) (Schones et al., 2008; Yuan et al., 2005; Henikoff et al., 2011; Kent et al., 2011). When compared to maps of nucleosome positions, both DNase-seq and FAIRE-seq tend to identify large nucleosome-depleted regions that range from 100-300 bp in length (Mouse ENCODE Consortium et al., 2012). As a result, differences in DNA accessibility that occur within or close to

nucleosomes, or quantitative differences in accessibility of individual nucleosomes, are difficult to detect by these methods.

In addition, for more than three decades, restriction enzymes (REs) have been utilized to probe DNA accessibility at individual loci (Liberator and Lingrel, 1984; Almer and Hörz, 1986; Logie and Peterson, 1997; Narlikar et al., 2001; Ohkawa et al., 2006). Since REs digest DNA at specific nucleotide sequences known as restriction sites (RSs), REs can quantitatively probe cell type-specific differences in accessibility at individual positions, when combined with Southern blotting or PCR. The accessibility of chromatin to REs can, in principle, be quantified at any genomic location that harbors an RS, including DHSs, DNA sequences within nucleosomes, and linker regions within closely-spaced nucleosome arrays. Previously, Gargiulo *et al* developed a genome-wide method to probe chromatin structure using restriction enzymes, finding that chromatin accessibility correlated broadly with gene expression in hematopoietic cell lineages and became progressively restricted during differentiation (Gargiulo et al., 2009). Here we modified this method to reduce potential biases in library production and increase the fraction of reads within a library that directly reflect RE cleavage. We employ this modified method, termed RED-seq, to measure RE accessibility across the genome of multiple cell types.

Here we show that, as with DNase-seq and FAIRE-seq, RED-seq uncovers known regions of open chromatin, validating the method as a genome-

wide probe of chromatin accessibility. Furthermore, we find that RED-seq can quantify both large differences in chromatin accessibility between different cell types and subtle changes that occur during ESC differentiation, highlighting the sensitivity of the assay. However, unlike these methods, we find that RED-seq also identifies differences in accessibility within nucleosome arrays. Consequently, we uncover significant differences in accessibility between nucleosomes containing different histone variants, showing that DNA bound by nucleosomes containing H2A.Z or H3.3 are more accessible than the genome-wide average. Consistent with this model, RNAi-mediated depletion of factors required for H2A.Z or H3.3 deposition into chromatin results in reduction of accessibility at these sites. Therefore, these results provide *in vivo* evidence that DNA accessibility within nucleosomes is modulated by the composition of histone proteins.

RESULTS

Genome-wide measurement of chromatin accessibility by RED-seq

Due to the inherent biases of standard methods of measuring chromatin accessibility, such as DNase-seq and FAIRE-seq, toward nucleosome-free regions of DNA, these methods are not well suited to examination of chromatin accessibility in the vast majority of the genome found within nucleosome arrays. A prior RE-based method of probing chromatin accessibility genome-wide (called

NA-Seq) revealed that accessibility of regulatory regions of genes correlated with their gene expression patterns (Gargiulo et al., 2009). We therefore wished to examine the accessibility of ESC chromatin using REs, in order to probe regions of open chromatin structure that are well covered by DNase-seq and FAIRE-seq maps (to assess whether REs faithfully report known features of ESC chromatin structure), as well as examine chromatin accessibility within nucleosomes and between nucleosomes that lie within regularly-spaced nucleosome arrays.

NA-Seq was previously performed by exposing purified nuclei to REs, secondary digestion of the purified DNA with an additional RE, ligation of linkers, and 454 pyrosequencing (Gargiulo et al., 2009). We modified the NA-Seq method in several ways (Figure I.1A): First, we performed RE digestion on permeabilized cells without nuclear purification in order to reduce processing steps prior to chromatin digestion by REs. Second, we used an unbiased, sonication-based shearing approach after DNA purification to reduce potential biases in the library introduced by the genomic distribution of the restriction sites (RSs) specific for the post-DNA purification RE used in NA-Seq. Finally, we used two separate linker ligation steps to ensure that single-read Illumina sequencing would sequence the end of each DNA fragment cleaved by the RE (rather than the randomly sheared end), making nearly all mapped reads informative, rather than about half. We refer to this modified method as RED-seq to distinguish this modified protocol from the previous NA-Seq approach.

In principle, any RE or combination of REs could be used for RED-seq library preparation. We utilized Sau96I, an RE with a four base RS (GGNCC) that occurs frequently throughout the mouse genome and is abundant within gene regulatory sequences, in order to probe genome-wide accessibility at relatively high resolution. First, we compared the differences in RE accessibility between mouse ESC chromatin and naked DNA. Because chromatin and naked DNA have identical RSs, differences in RE accessibility should result directly from the influences of chromatin proteins on accessibility at each RS (e.g., nucleosome occupancy or binding of non-histone proteins). Indeed, naked DNA was more efficiently cleaved and the digestion products were more uniformly distributed compared to ESC chromatin (Figure I.1B), as expected. Next, we prepared sequencing libraries of ESC and naked DNA samples, to quantify the digestion frequency at each Sau96I RS in the genome, and sequenced the libraries. The enrichment within the sequence reads of the expected product of Sau96I digestion (GNCC) immediately following the adapter barcode confirmed the quality of the libraries (Figure I.1C).

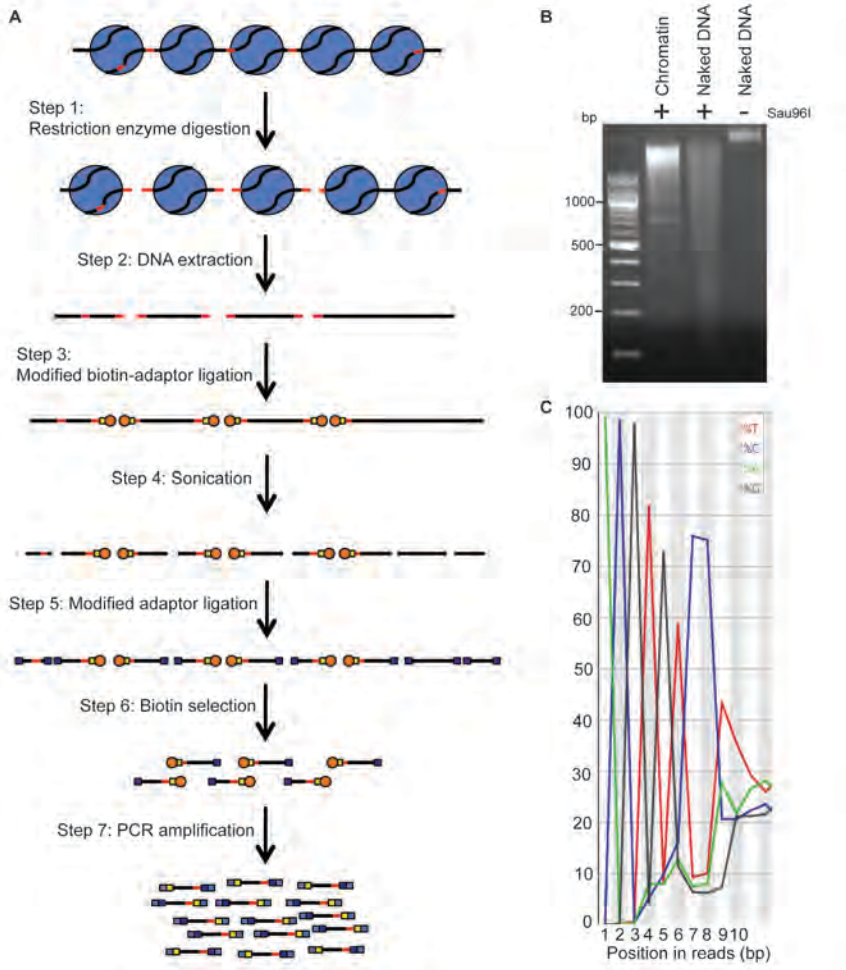


Figure I.1: The RED-seq method for genome-wide measurement of RE accessibility

(A) RED-seq workflow. RSs are shown in red, yellow boxes (Step 3) represent RS-proximal adaptors, dark blue boxes (Step 5) represent RS-distal adaptors, orange circles represent biotin, light blue boxes represent paired-end PCR primers, large blue circles (Step 1) represent nucleosomes, and DNA is shown in black.

(B) Ethidium bromide stained agarose gel indicating bulk digestion levels of chromatin and naked DNA.

(C) An example FASTQ file is shown to illustrate the near-uniform sequencing of the RS-containing end of each fragment in the library, signified by the large enrichment of G at position 5, and a CC dinucleotide at positions 7 and 8, derived from the cleaved and blunt-ended Sau96I site (GNCC).

We developed a software package (also named REDseq; available as a Bioconductor package) to assign each read to a unique RS in the mouse genome (see Methods for details), and count the relative cut frequency per site corresponding to normalized read counts assigned to each RS. As we observed by electrophoresis of digested naked DNA or chromatin (Figure I.1B), average RE accessibility, as measured by relative cut frequency per RS, was reduced in the chromatin library relative to naked DNA at most sites (Figure I.2A). As expected, due to the fact that cutting frequency at each RS was normalized to total reads in each library, we observed fragments derived from some RSs that were more abundant in the chromatin library than the naked DNA library. In addition, cleavage within the naked DNA library was not uniform at all RSs (Figure I.2A), likely due to the fact that fragments generated by two Sau96I cleavages within close proximity are selected against during library preparation, which eliminates small DNA fragments. This is less of a concern in chromatin samples, in which cleavage at most RSs is suppressed.

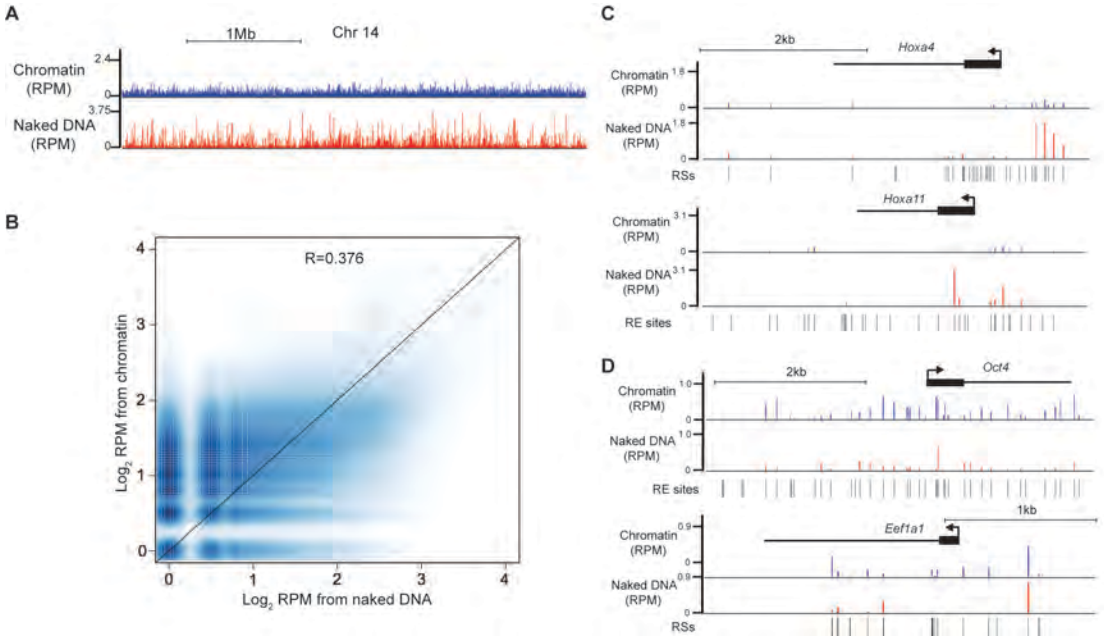


Figure I.2: Comparison of RED-seq to naked DNA digestion

(A) RE accessibility reads from mouse ESC chromatin (top) and naked DNA

(bottom) from a 3 Mb region of chromosome 14 (Chr14). Shown are normalized reads per million (RPM).

(B) Scatterplot of RE accessibility[Log2(RPM)] for Chr14 from chromatin relative to naked DNA.

(C) RE accessibility from chromatin and naked DNA of two Hox genes, Hoxa4 and Hoxa11, which are silent in ESCs. Dotted lines highlight the genomic regions with RE accessibility differences apparent between chromatin and naked DNA.

(D) RE accessibility from chromatin and naked DNA of two highly expressed genes in ESCs, Oct4 and Eef1a1.

Furthermore, we did not observe a strong correlation between the reads from chromatin DNA and naked DNA (R=0.376), confirming that the degree of RE digestion at most sites was different between chromatin and naked DNA (Figure I.2B). Thus, RED-seq accurately reflects inhibition of RE accessibility by the presence of chromatin *in vivo*.

Active genes and nucleosome-free regions are highly accessible

RE accessibility in promoter-proximal regions is usually correlated with gene expression (Pfeiffer and Zachau, 1980; Felsenfeld, 1992; Kornberg and Lorch, 1992). Homeobox (Hox) genes encode key developmental TFs that are not expressed in ESCs (Pearson et al., 2005). We observed low levels of RE accessibility around *Hox* genes relative to surrounding regions and normalized naked DNA reads (Figure I.2C). In contrast, for genes that are highly expressed in ESCs (*Oct4*, *Eef1a1*), RE accessibility was elevated within upstream regulatory regions and surrounding transcriptional start sites (TSSs) (Figure I.2D). Overall, these results showed that enhanced RE accessibility was generally associated with transcriptional activity, consistent with previous data.

DNase I is frequently used to identify open chromatin/nucleosome-free regions of the genome, and many gene regulatory elements are hypersensitive to DNase I (Saragosti et al., 1980; Schones et al., 2008; Davie and Saunders, 1981; Xi et al., 2007). Therefore, we next examined the frequency of RED-seq

reads surrounding annotated DHSs in ESCs. Since RSs are non-uniformly distributed throughout the genome, we compared RE accessibility averaged over all DHSs to average RS density to test whether DHSs were generally accessible or inaccessible. We found that RE accessibility over DHSs was strongly enhanced relative to the RS density surrounding these regions (**Figure I.3A**). DHSs are typically nucleosome-depleted and highly transcribed, relative to DNase I-insensitive regions (Saragosti et al., 1980; Schones et al., 2008; Davie and Saunders, 1981; Xi et al., 2007). Therefore, we compared our RED-seq data to nucleosome occupancy maps previously obtained by deep sequencing of nucleosome-sized DNA fragments protected from digestion by micrococcal nuclease (MNase-seq) (Carone et al., 2014), and found that nucleosomes were strongly depleted over DHSs (Figure I.3B), consistent with the higher RE accessibility we observed.

Next, we compared RE accessibility surrounding the binding sites of two key TFs in ESCs. CTCF is a sequence-specific insulator binding protein with important roles in regulation of imprinted gene expression (Fedoriw et al., 2004; Szabó et al., 2004) and higher-order chromatin structure (Kurukuti et al., 2006). RE accessibility was enriched within the regions surrounding CTCF (Figure I.3C). As previously reported (Fu et al., 2008; Cuddapah et al., 2009), CTCF binding sites are depleted of nucleosomes, with well-positioned nucleosomes flanking the nucleosome-free regions (Figure I.3D), explaining the higher accessibility we observed at these sites. Interestingly, for highly abundant nucleosome-free

regions such as CTCF binding sites and DHSs, RED-seq also revealed nucleosome phasing around nucleosome-depleted regions, with smaller phased peaks of RE accessibility found within linker regions (Figures I.3E-F).

Since the majority of inter-nucleosomal linkers are relatively small (averaging approximately 30 bp in ESCs (Cao et al., 2013), this phasing is not apparent using DNase-seq (Mouse ENCODE Consortium et al., 2012) which is specialized for identification of long stretches of nucleosome-free DNA (Figure I.3E-F). Together these results show that while the resolution of RED-seq at the level of individual loci is variable and depends on the frequency of RSs at each locus, when averaged over thousands of loci RED-seq not only identifies large nucleosome-free regions identified by DNase-seq, but can also probe DNA linker regions within nucleosome arrays.

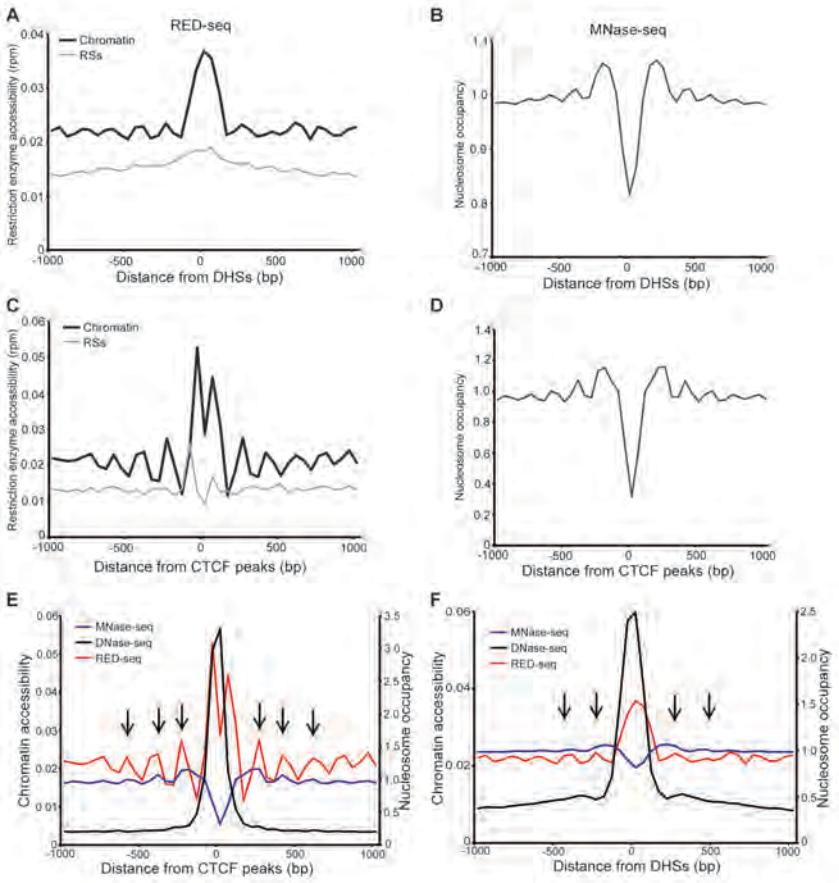


Figure I.3: RED-seq captures the enhanced accessibility of open chromatin regions

Average RE accessibility (A, C) and nucleosome occupancy (B, D) [GEO:GSM1400766] of indicated chromatin domains. RED-seq or MNase-seq data are aligned on the centers of all peaks of DHSs (A-B), or CTCF binding sites (C-D), and averaged within a 2 kb region (-1000 to +1000 bp from the peaks). Normalized RE accessibility and RS density are shown. There are 159,331 DHSs [GEO:GSM1014154] (A-B), and 15,657 CTCF binding sites [GEO:GSE11431] (C-D) plotted.

(E-F) Chromatin accessibility determined by RED-seq or DNase-seq and nucleosome occupancy are shown surrounding CTCF binding sites (E) or DHSs (F). Arrows indicate the phased peaks of RE accessibility found within linker regions.

Remodeling of chromatin accessibility during differentiation

ESC chromatin structure is relatively dynamic and is depleted of large blocks of heterochromatin, unlike many differentiated cell types, suggesting that major alterations in chromatin structure that accompany cellular differentiation may be important for lineage commitment (Meshorer and Misteli, 2006). To study chromatin accessibility during differentiation, we first tested whether RED-seq could identify distinct RE accessibility patterns in different cell types by comparing chromatin accessibility in ESCs and mouse embryonic fibroblasts (MEFs). We found that, in MEFs, nucleosome occupancy was increased and RE accessibility decreased at ESC-specific DHSs (Figure I.4A-B), consistent with the widespread differences in chromatin structure and gene expression between these two cell types. As with DHSs, RE accessibility at sites of CTCF binding in ESCs was reduced in MEFs (Figure I.4C-D), and these results were consistent in biological replicate RED- seq libraries from both cell types (Figure I.4E).

Figure I.4: Cell type-specific differences in chromatin accessibility

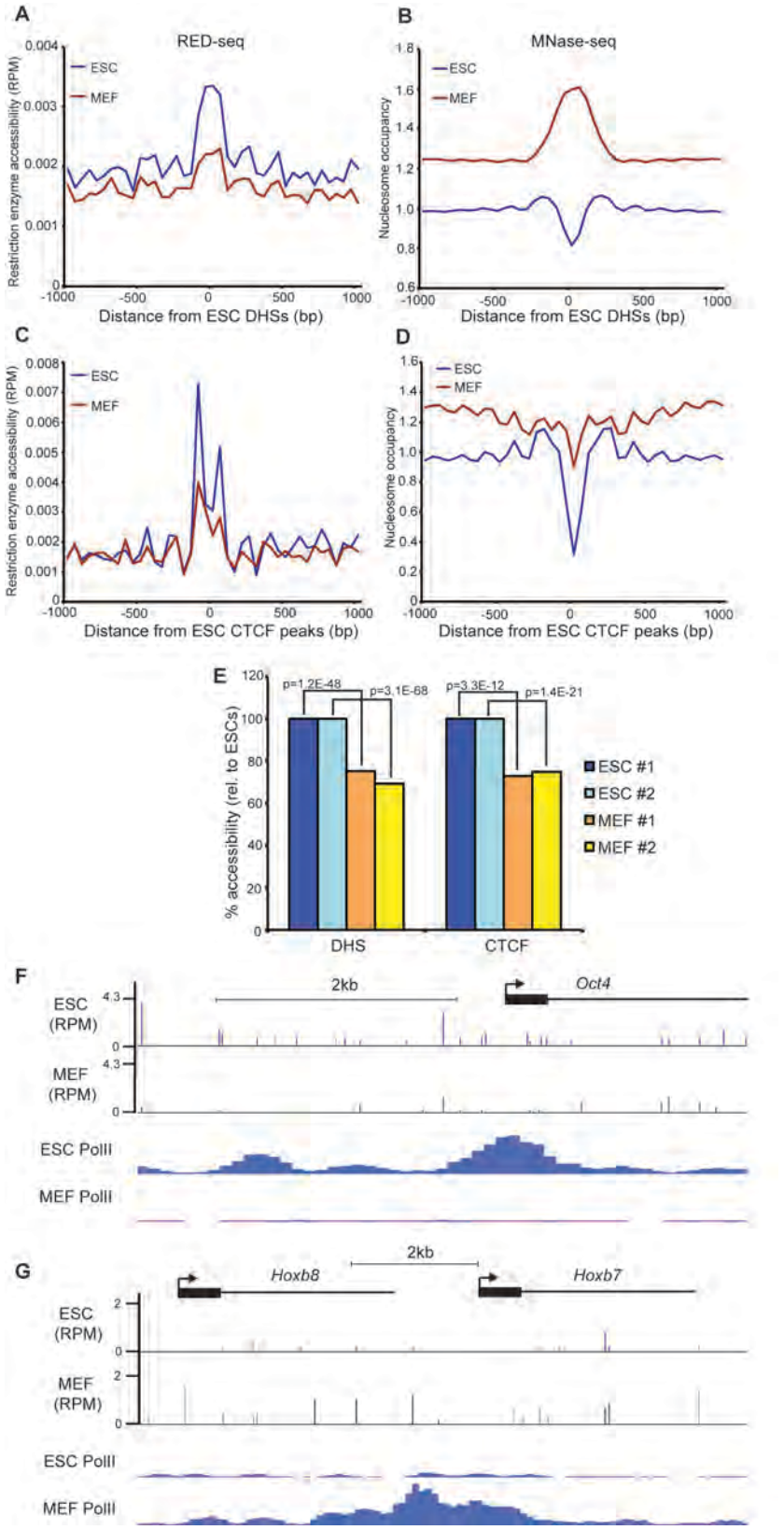
(A) Average RE accessibility of ESCs (blue) or MEFs (red) shown relative to DNase I hypersensitive sites (DHSs) identified in ESCs [GEO:GSE46588].

(B) Nucleosome occupancy of the same regions is shown for ESCs [GEO:GSM1400766] and MEFs [GEO:GSM1004654].

(C) Average RE accessibility and (D) nucleosome occupancy surrounding CTCF binding regions in ESCs [GEO:GSE11431] are shown for ESCs and MEFs.

(E) Average accessibilities over DHSs and CTCF binding sites were quantified for biological replicate experiments from –200 to +200 bp with respect to the indicated feature. P-values indicating statistical significance of accessibility between ESCs and MEFs are indicated.

(F, G) RE accessibility of ESCs and MEFs surrounding the Oct4 gene (F) and two genes within the Hoxb cluster (G). RNA Polymerase II (RNA Pol II) ChIP-seq reads [GEO:GSE29184] from ESCs and MEFs are shown for the same regions.


A
RED-seq
ESC
MEF
Restriction enzyme accessibility (RPM)
Distance from ESC DHSs (bp)
B
MNase-seq
ESC
MEF
Nucleosome occupancy
Distance from ESC DHSs (bp)
C
ESC
MEF
Restriction enzyme accessibility (RPM)
Distance from ESC CTCF peaks (bp)
D
ESC
MEF
Nucleosome occupancy
Distance from ESC CTCF peaks (bp)
E
p=1.2E-48
p=3.1E-68
p=3.3E-12
p=1.4E-21
% accessibility (rel. to ESCs)
ESC #1
ESC #2
MEF #1
MEF #2
DHS
CTCF
F
2kb
Oct4
ESC (RPM)
MEF (RPM)
ESC PolII
MEF PolII
G
2kb
Hoxb8
Hoxb7
ESC (RPM)
MEF (RPM)
ESC PolII
MEF PolII

Finally, we examined RE accessibility within regions surrounding TSSs in both cell types. TSS-proximal regions of actively transcribed genes are usually nucleosome-depleted and the degree of nucleosome-depletion correlates with transcriptional activity at many genes. As expected, RE accessibility was higher in ESCs than in MEFs surrounding the TSSs of genes that were highly expressed in ESCs (Figure I.4F), whereas genes highly expressed in MEFs were generally more accessible in MEFs (Figure I.4G). These data confirmed that RED-seq could identify differences in chromatin accessibility between two distinct cell types that reflected differences in TF binding and gene expression.

Next, to test whether we could observe more subtle changes in chromatin structure during cellular differentiation, we differentiated ESCs by RNAi-mediated knockdown (KD) of the ESC pluripotency TF Oct4. We chose this differentiation model since, unlike most other methods of differentiation that generate heterogeneous mixtures of many different cell types from all three germ layers, *Oct4* KD robustly induces differentiation to trophectoderm specifically (Niwa et al., 2000). Consistent with previous reports (Niwa et al., 2000), *Oct4* KD promoted ESC differentiation to cells with trophoblast morphology (Figure I.5A-B). Using RED-seq, we found that RE accessibility was decreased upon *Oct4* KD near ESC DHSs and CTCF binding sites (Figure I.5C-E).

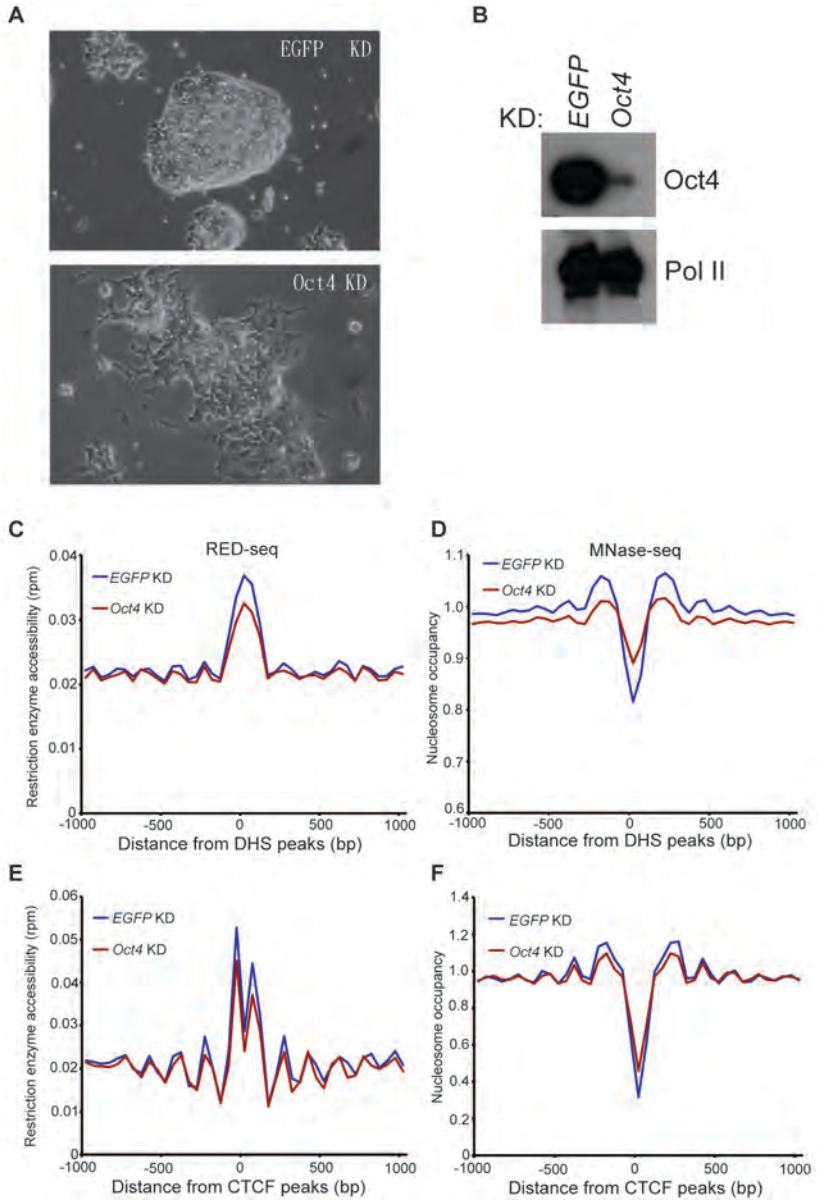


Figure I.5: Alterations in RE accessibility during ESC differentiation

(A) Brightfield images of control (EGFP) or Oct4 KD ESC colonies indicate colony flattening and elongated cellular morphology upon Oct4 depletion.

(B) Western blot of Oct4 in control (EGFP) or Oct4 KDs, indicating KD efficiency. RNA Polymerase II blot (RNA Pol II) is shown as a loading control.

(C, E) Average RE accessibility upon EGFP or Oct4 KD is shown relative to DHSs (C), or CTCF binding sites (E).

(D, F) MNase-seq data. Nucleosome occupancy over DHSs (D), or CTCF binding sites (F).

Although the reduction in DNA accessibility upon *Oct4* KD was not as severe as in MEFs, we also observed slightly increased nucleosome occupancy by MNase-seq upon *Oct4* KD at ESC DHSs and CTCF binding sites (Figure I.5D, F), consistent with the decrease in RE accessibility that we observed in these regions.

To validate these results, we used quantitative PCR (qPCR) to determine the fraction of uncut (protected) DNA after RE digestion, probing several ESC DHSs and CTCF binding sites. Consistent with the RED-seq results, higher levels of uncut DNA were observed upon *Oct4* KD at most sites tested (Figure I.6A-B). Furthermore, we tested CTCF binding at the same regions by ChIP-qPCR, and observed a reduction in binding upon *Oct4* KD wherever chromatin accessibility decreased, whereas control CTCF binding sites that showed no difference in accessibility upon *Oct4* KD showed no decrease in CTCF binding (Figure I.6C). These data indicate that CTCF binding and RE accessibility are inter-dependent. Next, we observed that RE accessibility surrounding the binding sites of the ESC TF Klf4 was also reduced upon *Oct4* KD (Figure I.6D), with concomitant increases in nucleosome occupancy over these sites (Figure I.6E). Finally, we found the alterations in accessibility we observed over DHSs, CTCF binding sites, and Klf4 binding sites were consistent in two biological RED-seq replicates from each KD (Figure I.6F), further validating these results. These results suggest that, during differentiation, many enhancers that are protected from nucleosome deposition in ESCs (presumably by TF binding) become

occupied by nucleosomes, leading to decreased RE accessibility. Taken together, RED-seq not only detects large differences in chromatin accessibility between distinct cell types (ESCs vs MEFs) but also tracks more subtle changes that occur during differentiation (control vs *Oct4* KD ESCs).

Figure I.6: Loss of chromatin accessibility at some CTCF binding sites correlates with reduced CTCF binding upon ESC differentiation

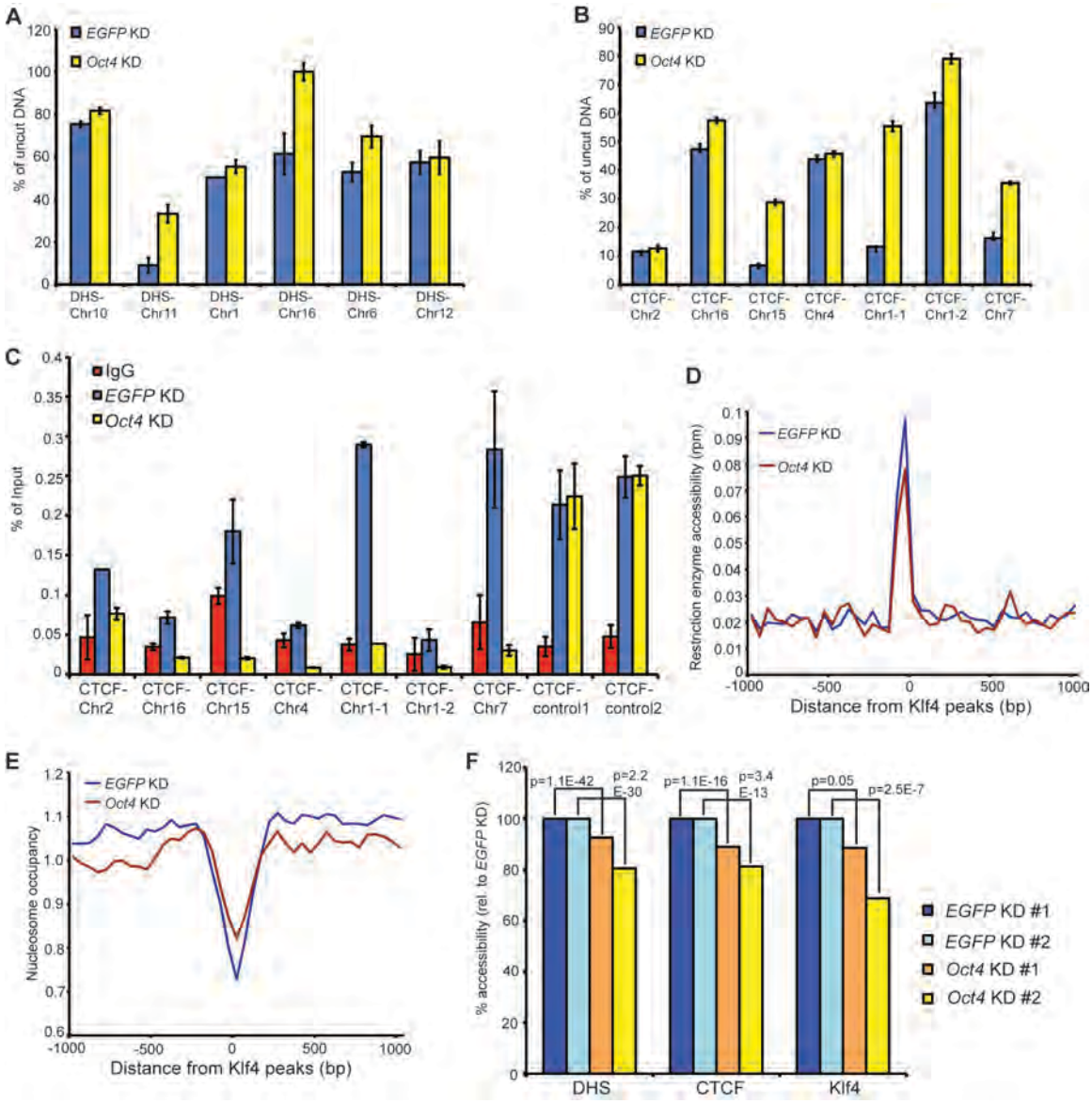
(A) Differences in RE accessibility at specific DHSs were confirmed by qPCR across an RS of interest at each locus. Remaining uncut DNA after RE digestion of each indicated KD is shown for several DHSs that exhibited accessibility differences by RED-seq. Data are normalized to uncut genomic DNA.

(B) Confirmation of restriction enzyme accessibility surrounding CTCF binding sites, as in (A).

(C) CTCF ChIP-qPCR data are shown for the indicated KDs at several CTCF binding sites. Controls are CTCF binding sites in which accessibility did not change upon Oct4 KD. Data are presented as a percentage of input DNA. Shown are the mean ± SD of three technical replicates from one representative experiment of two biological replicates performed.

(D-E) RED-seq data (D) and MNase-seq data (E) over Klf4 binding sites, plotted as in Figure I.3.

(F) Average accessibilities over DHSs, CTCF binding sites, and Klf4 binding sites were quantified for biological replicate KD experiments from –200 to +200 bp with respect to the indicated feature. P-values indicating statistical significance of accessibility between EGFP KD and Oct4 KD are indicated.


A
EGFP KD
Oct4 KD
% of uncut DNA
DHS-Chr10
DHS-Chr11
DHS-Chr1
DHS-Chr16
DHS-Chr6
DHS-Chr12
B
EGFP KD
Oct4 KD
% of uncut DNA
CTCF-Chr2
CTCF-Chr16
CTCF-Chr15
CTCF-Chr4
CTCF-Chr1-1
CTCF-Chr1-2
CTCF-Chr7
C
IgG
EGFP KD
Oct4 KD
% of input
CTCF-Chr2
CTCF-Chr16
CTCF-Chr15
CTCF-Chr4
CTCF-Chr1-1
CTCF-Chr1-2
CTCF-Chr7
CTCF-control1
CTCF-control2
D
EGFP KD
Oct4 KD
Restriction enzyme accessibility (rpm)
Distance from Klf4 peaks (bp)
E
EGFP KD
Oct4 KD
Nucleosome occupancy
Distance from Klf4 peaks (bp)
F
% accessibility (rel. to EGFP KD)
p=1.1E-42
p=2.2 E-30
p=1.1E-16
p=3.4 E-13
p=0.05
p=2.5E-7
DHS
CTCF
Klf4
EGFP KD #1
EGFP KD #2
Oct4 KD #1
Oct4 KD #2

Altered accessibility of nucleosomes harboring distinct histone variants

Genomic regions that are dynamic (i.e. experience relatively rapid exchange of chromatin proteins) are frequently marked with specific histone modifications and/or histone variants (Skene and Henikoff, 2013). However, using traditional methods such as DNase-seq or FAIRE-seq, it is difficult to identify differences in chromatin accessibility that correlate with the presence of dynamic nucleosomes, because these regions are not nucleosome-free. In principle, RED-seq does not share these limitations, due to the fact that a single RE cleavage is all that is necessary for inclusion in a RED-seq library (Figure I.1A). Therefore, we examined the accessibility of regions enriched for dynamic histone variants/modifications using RED-seq.

To establish a baseline for the examination of different types of nucleosomes, we first determined the average accessibility of a random distribution of nucleosomes across the genome. To this end, we randomly selected 1% of all nucleosomal footprints from an MNase-seq library prepared from ESCs, and plotted the average RED-seq and MNase-seq profiles within a 2 kb window surrounding their positions. Consistent with the fact that nucleosome-bound DNA is relatively inaccessible to nuclear factors, we observed a low level of RE accessibility surrounding the peak of bulk nucleosomes, relative to RS density (Figure I.7A). Therefore, as expected, nucleosome-free DNA, like that

underlying DHSs and TF binding sites, is generally more accessible than nucleosomal DNA.

Next, we tested whether the accessibility of nucleosome variants that harbor particular histone modifications or histone variants were identical to that of bulk nucleosomes. The two nucleosomes surrounding TSSs (referred to as +1 and -1 nucleosomes) are frequently marked by histone variants H2A.Z and H3.3 (Jin and Felsenfeld, 2007; Henikoff et al., 2009; Jin et al., 2009; Tolstorukov et al., 2009). Nucleosomes harboring these variants have been found to be extractible from chromatin at lower salt than is required for canonical nucleosomes (Jin and Felsenfeld, 2007; Henikoff et al., 2009), raising the possibility that they may be more highly accessible in general. H2A.Z is enriched surrounding the TSSs of many eukaryotic genes, and also found within active enhancers in mammalian cells (Hu et al., 2013). Furthermore, H2A.Z-marked nucleosomes protect smaller footprints of DNA than canonical nucleosomes, leading to the suggestion that these nucleosomes are more intrinsically accessible (Tolstorukov et al., 2009). In ESCs, H2A.Z is found near approximately 84% of all TSSs, including those of many silent genes (Creyghton et al., 2008). Interestingly, we observed increased RE accessibility over the center of the H2A.Z peaks relative to both RS density and surrounding regions ± 1 kb from the peaks of H2A.Z enrichment (Figure I.7B), suggesting that H2A.Z-containing nucleosomes are generally more accessible than canonical nucleosomes.

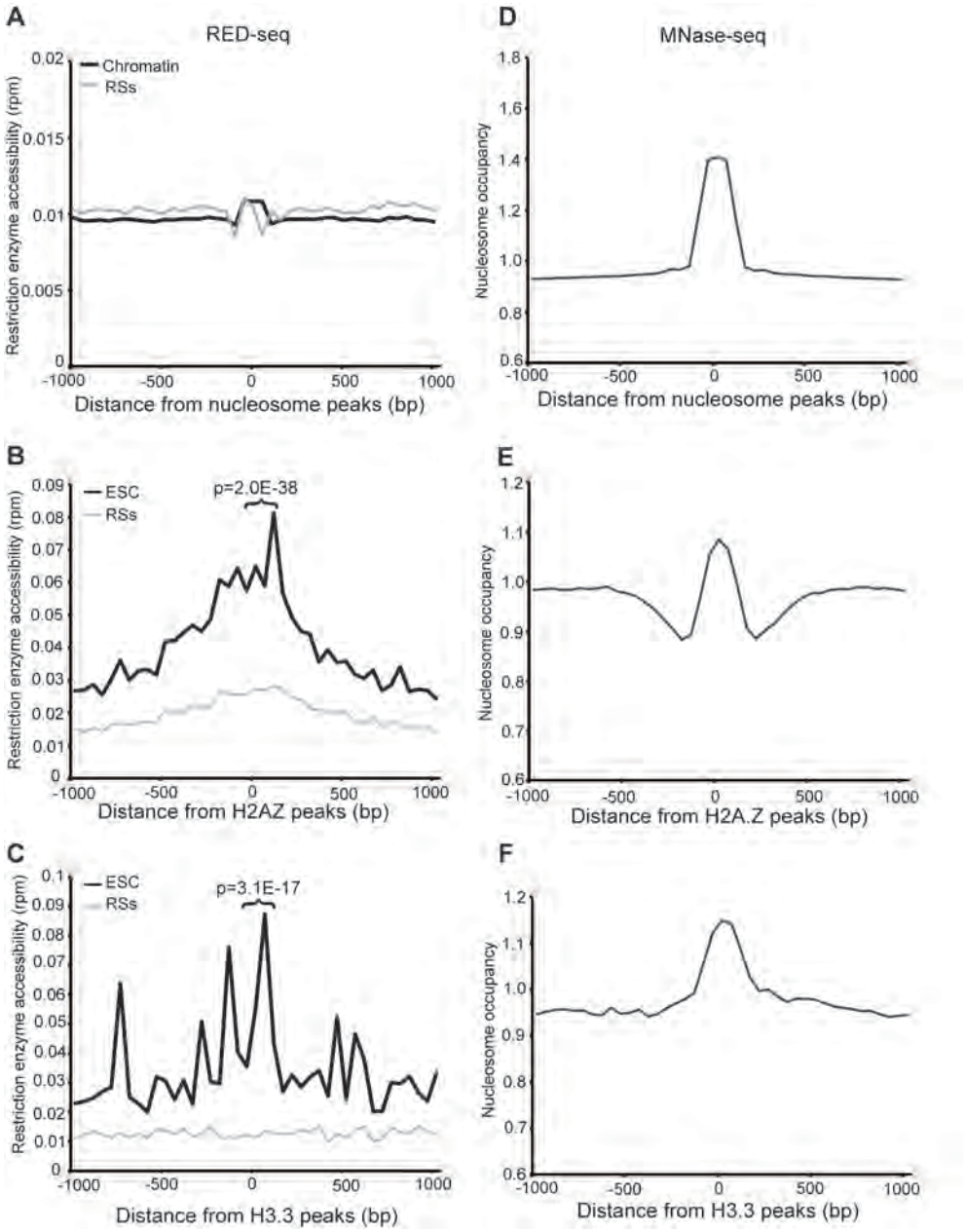


Figure I.7: Enhanced accessibility of DNA bound by H2A.Z-containing nucleosomes

Average RE accessibility (A-C) and nucleosome occupancy (D-F) shown relative to 320,135 randomly selected nucleosomes (A, D), 39,437 H2A.Z-containing nucleosomes [GEO:GSE34483] (B, E), or 8,287 H3.3-containing nucleosomes [GEO:GSE16893] (C, F). Data are plotted as in Figure I.3. P-values indicating statistical significance of accessibility between H2A.Z and average nucleosome profiles, as well as H3.3 and average nucleosomes are indicated.

Next, we examined H3.3, a variant of histone H3 that is enriched near the TSSs of both active and silent genes, as well as within gene bodies of highly expressed genes, and is incorporated into chromatin in a replication-independent manner (Ahmad and Henikoff, 2002; Goldberg et al., 2010; Szenker et al., 2011). Like H2A.Z, we found that RE accessibility over H3.3 peaks was elevated relative to RS density (Figure I.7C). These data suggest that DNA wrapped around H2A.Z- and H3.3-marked nucleosomes is more accessible than DNA found within the majority of nucleosomes genome-wide that lack these histone variants.

We considered the possibility that the elevated RE accessibilities observed over peaks of H2A.Z enrichment and broad regions surrounding H3.3 were due to reduced nucleosome occupancy at these sites. However, while the average occupancies of H2A.Z- and H3.3-containing nucleosomes were slightly lower than bulk nucleosomes (compare the peak heights in Figures I.7D-F), these modest differences are insufficient to account for the greater than 5-fold increase in accessibility observed over H2A.Z and H3.3 peaks observed by RED-seq.

To validate these data, we examined chromatin accessibility upon KD of factors necessary for incorporation of H2A.Z or H3.3 into chromatin. In mammals, H2A.Z is incorporated into chromatin in part by p400 (gene name: *Ep400*), a homolog of the yeast Swr1 ATPase, whereas H3.3 incorporation depends in part on the HIRA (*Hira*) histone chaperone (Mizuguchi et al., 2004; Tagami et al.,

2004). We tested whether the enhanced chromatin accessibility observed at sites of H2A.Z and H3.3 deposition was reduced upon depletion of their respective loading factor, and found that the elevated accessibility we observed within regions of H2A.Z and H3.3 enrichment was partially lost upon *Ep400* KD or *Hira* KD, respectively (Figure I.8A-F). When we examined alterations in chromatin accessibility upon *Ep400* or *Hira* KD over a random sampling of nucleosomes (as in Figure I.7A), we observed only a modest decrease in accessibility, suggesting that the effects of *Ep400* or *Hira* KD are specific for nucleosomes containing H2A.Z or H3.3 (Figure I.8G). Finally, we examined changes in chromatin accessibility due to *Ep400* or *Hira* KD over CTCF binding sites, due to the reported enrichment of H2A.Z- and H3.3-containing nucleosomes surrounding CTCF (Jin et al., 2009). Interestingly, while *Hira* KD resulted in significantly reduced accessibility over CTCF binding sites, *Ep400* KD did not (Figure I.8H), suggesting that either H3.3 plays a more important role than H2A.Z in regulation of chromatin structure near CTCF binding sites or that H2A.Z is incorporated into chromatin at these sites independently of p400. We observed consistent differences in accessibility over H2A.Z, H3.3, and CTCF binding sites in biological replicate KDs of *Ep400*, *Hira*, and *Hira*, respectively (Figure I.8C, F, and I), validating these data. Together, these results suggest that H2A.Z- and H3.3-containing nucleosomes are either more dynamic or more intrinsically accessible than canonical nucleosomes, consistent with their association with gene regulatory sequences.

Figure I.8: Factors required for H2A.Z or H3.3 deposition are required for enhanced accessibility of regions normally bound by these histone variants

(A) Chromatin accessibility determined by RED-seq averaged over regions of the genome bound by H2A.Z, as in Figure I.7. Shown are control (EGFP KD) and Ep400 KD ESCs.

(B) Western blot of p400 in control (EGFP) or Ep400 KDs, indicating KD efficiency. Actin is shown as a loading control.

(C) Average accessibilities over H2A.Z-marked nucleosomes were quantified for biological replicate experiments from –200 to +200 bp with respect to the H2A.Z peak. P-values indicating statistical significance of accessibility between EGFP and Ep400 KDs are indicated.

(D) Chromatin accessibility determined by RED-seq averaged over regions of the genome bound by H3.3. Shown are control (EGFP KD) and Hira KD ESCs.

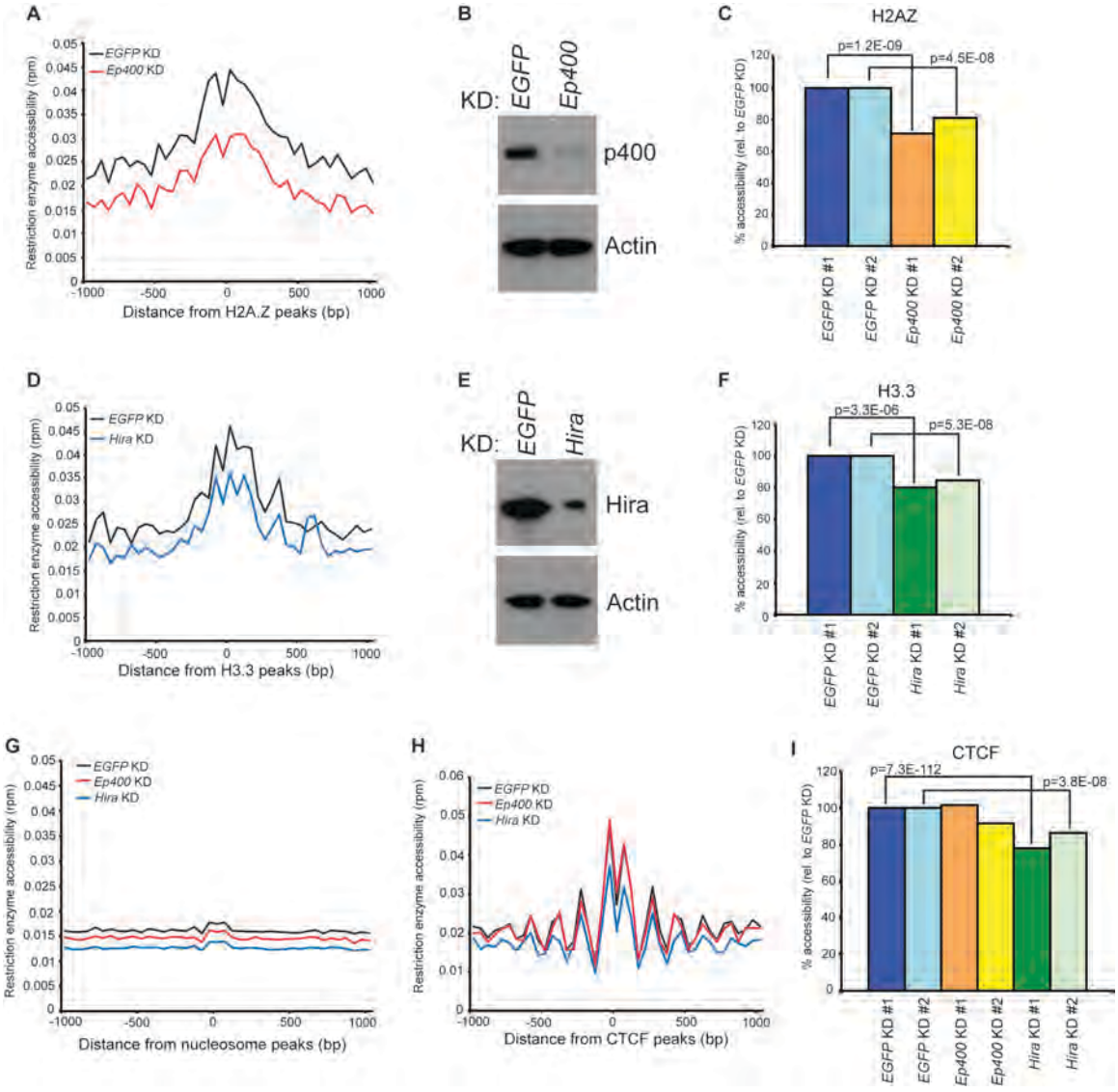
(E) Western blot of Hira in control (EGFP) or Hira KDs, indicating KD efficiency. Actin is shown as a loading control.

(F) Average accessibilities over H3.3-marked nucleosomes were quantified for biological replicate experiments from –200 to +200 bp with respect to the H3.3 peak. P-values indicating statistical significance of accessibility between EGFP and Hira KDs are indicated.

(G) Effects of Ep400 or Hira KD on average nucleosome accessibility shown by plotting RED-seq data over the same 320,135 randomly selected nucleosomes as in Figure I.7A.

(H) Effects of Ep400 or Hira KD on chromatin accessibility over CTCF binding sites, as in Figure I.3C.

(I) Average accessibilities over CTCF-binding sites were quantified for biological replicate experiments from –200 to +200 bp with respect to the peak of CTCF-binding. P-values indicating statistical significance of accessibility between EGFP and Hira KDs are indicated.


A
EGFP KD
Ep400 KD
Restriction enzyme accessibility (rpm)
Distance from H2A.Z peaks (bp)
B
KD: EGFP Ep400
p400
Actin
C
H2AZ
p=1.2E-09
p=4.5E-08
% accessibility (rel. to EGFP KD)
EGFP KD #1
EGFP KD #2
Ep400 KD #1
Ep400 KD #2
D
EGFP KD
Hira KD
Restriction enzyme accessibility (rpm)
Distance from H3.3 peaks (bp)
E
KD: EGFP Hira
Hira
Actin
F
H3.3
p=3.3E-06
p=5.3E-08
% accessibility (rel. to EGFP KD)
EGFP KD #1
EGFP KD #2
Hira KD #1
Hira KD #2
G
EGFP KD
Ep400 KD
Hira KD
Restriction enzyme accessibility (rpm)
Distance from nucleosome peaks (bp)
H
EGFP KD
Ep400 KD
Hira KD
Restriction enzyme accessibility (rpm)
Distance from CTCF peaks (bp)
I
CTCF
p=7.3E-112
p=3.8E-08
% accessibility (rel. to EGFP KD)
EGFP KD #1
EGFP KD #2
Ep400 KD #1
Ep400 KD #2
Hira KD #1
Hira KD #2

DISCUSSION

Utilizing an adaptation of a decades-old, quantitative technique for probing chromatin accessibility, we probed the chromatin structure of ESCs and differentiated cells, observing differences in chromatin accessibility in distinct regions of the genome, as well as in different cellular states. We found that both the level of nucleosome occupancy and the presence of specific histone variants present at individual loci affected the level of chromatin accessibility we observed at each site.

Over the past several years, DNase-seq and FAIRE-seq have been used to identify regions of open chromatin structure within cells. One limitation of these methods is that only nucleosome-depleted regions of DNA are typically identified. Interestingly, while RED-seq identified nucleosome-depleted regions as well, we also observed differences in chromatin accessibility within nucleosomes that harbor specific histone variants, detecting increased RE accessibility in genomic regions enriched for histones H2A.Z and H3.3. Therefore, unlike previous methods, RED-seq not only measures general chromatin “openness” but also identifies highly dynamic regions of the genome, even if they are not nucleosome-free. We believe that this feature – the ability to quantify accessibility of DNA within nucleosome-bound regions – best distinguishes RED-seq from complementary approaches such as MNase-seq and DNase-seq, which do not probe intranucleosomal accessibility.

The increased accessibility of DNA within H2A.Z- and H3.3-containing nucleosomes is due to the histone variants themselves rather than some unrelated feature of chromatin structure within these regions of the genome, since depletion of H2A.Z and H3.3 loading factors strongly reduced the accessibility of the underlying DNA. Although H2A.Z and H3.3 are also enriched near TSSs, these histone variants are also found within multiple other genomic domains. Indeed, we find that accessibility over CTCF binding sites was reduced upon KD of the H3.3 deposition factor, *Hira*, suggesting that H3.3 incorporation within nucleosomes surrounding CTCF binding sites may be important for CTCF binding and/or function.

Chromatin structure is dramatically altered during cellular differentiation. By examining regions of the genome enriched for histone modifications, TFs, or chromatin regulators, RED-seq could identify differences in chromatin structure within functionally distinct regions of the genome during ESC differentiation. We found that RE accessibility decreased at many CTCF binding sites upon *Oct4* KD and that this decrease correlated with a decrease in CTCF occupancy and an increase in nucleosome occupancy. These differences were even more apparent when comparing ESCs with MEFs. Together, these results suggest that loss of TF binding during differentiation is coincident with deposition of nucleosomes at these sites, leading to loss of chromatin accessibility.

Besides chromatin structure, restriction enzymes have been widely used in biological assays for single nucleotide polymorphisms (SNPs) (Miller et al., 2007; Baird et al., 2008) and DNA methylation (McClelland, 1981) at individual loci, by virtue of their inhibitory effect on RE cleavage. Therefore, a genome-wide method to directly quantify differences in RE cleavage would be highly desirable in these assays. Our method of directly purifying RE-digested sequences and quantifying RE cleavage at each site by high-throughput DNA sequencing could be easily adapted to perform these types of studies. Thus, we believe that RED-seq will be a valuable tool for not only the measurement of chromatin accessibility and dynamics, but also the study of any other phenomena that alter RS cleavage by REs.

METHODS

Cells

The murine ESC line used in this study was E14 (Doetschman et al., 1987). Mouse embryonic fibroblasts (MEFs) used in this study were immortalized by serial passaging, following a 3T3 protocol, to minimize day-to-day differences in these cells due to their passage number.

Preparation of RED-seq libraries

One million cells were used to construct RED-seq libraries. Cells were washed, pelleted, and resuspended in Swelling buffer (10mM Tris pH8.0, 85mM KCl, 0.5% NP-40, 10mM $MgCl_2$) with 100 units of Sau96I (NEB) and incubated in a thermomixer (Eppendroff) at 37°C for 1 hour, shaking at 900 rpm. (For testing whether two REs might increase coverage, in one experiment 100 units of Sau96I and 50 units of DdeI were used in digestion.) Digestion was terminated by adding 40µl of 10% SDS and 20µl of 0.5M EDTA and the chromatin was treated with proteinase K (Ambion) overnight at 55°C. Digested DNA was purified using Phenol/Chloroform/Isoamyl alcohol extractions and precipitated at -80°C for 1 hour. Digested DNA samples were end-repaired and A-tailed as described (Yildirim et al., 2011), and ligated with biotinylated and barcoded adaptors. DNA was purified using Zymo Research DNA clean and concentrate columns following each enzyme reaction. The biotin-adaptor ligated DNA was sonicated in a Covaris sonicator (S220) to generate DNA peak fragments of 200 bp, on average. The sonicated DNA samples were then end-repaired, A-tailed, and ligated with non-biotinylated adaptors. The ligation samples were loaded on 2% agarose gel and DNA was purified within a size range of roughly 200-350 bp in length. Gel-purified DNA was diluted to 250µl with streptavidin binding buffer (20mM HEPES pH 7.6, 500mM NaCl, 1mM EDTA, 0.02% NP-40) and incubated with 30µl of pre-washed streptavidin magnetic beads (NEB) at room temperature for 1 hour. After magnetic separation, the supernatants were removed, and the beads were washed additional three times with streptavidin binding buffer. DNA

was eluted from streptavidin magnetic beads by adding 20 µl of 0.1X TE and heating at 60°C for 3 minutes. The elution was repeated three times. The adaptor-ligated material was then PCR amplified with Phusion polymerase (NEB) using 16 cycles of PCR and its concentration was determined using a NanoDrop (Thermo). The integrity of each library was confirmed by sequencing 10-20 individual fragments per library. Libraries with different barcodes were pooled together and single-end sequencing (50 bp) was performed on an Illumina HiSeq2000 at the UMass Medical School deep sequencing core facility. For most RED-seq libraries (*GFP*, *Oct4*, *Ep400* and *Hira* KD), we added one further modification in which the sequence of the biotinylated adapters and the second, non-biotinylated, adapters were modified such that after PCR amplification of the libraries, only the end that was ligated to the biotinylated adapter would be sequenced in a single-end sequencing run (**Table I.1**). Although this alteration makes the data analysis slightly simpler, the two methods provide essentially identical results.

Name	Forward sequence	Reverse sequence
Modified-Biotin-Set2	p-GCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	Biotin-ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCT
Modified-Set2	p-GCAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGCT
Modified-Biotin-Set5	p-CGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	Biotin-ACACTCTTTCCCTACACGACGCTCTTCCGATCTACGT
Modified-Set5	p-CGTAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTACGT
Modified-Biotin-Set6	p-ATGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	Biotin-ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATT
Modified-Set6	p-ATGAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCATT
Modified-Biotin-Set8	p-GCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	Biotin-ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCT
Modified-Set8	p-GCTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTAGCT
Modified-Biotin-Set10	p-CGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	Biotin-ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGT
Modified-Set10	p-CGAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTCG

Table I.1: Sequences of barcoded and biotinylated adaptors

Preparation of MNase-seq libraries

MNase-library preparation was adapted from Henikoff et al (Henikoff et al., 2011). Formaldehyde cross-linked cells were pelleted and washed twice with PBS. Cell pellets were resuspended in MNase lysis buffer (10mM Tris pH 7.5, 10mM NaCl, 3mM $MgCl_2$, 0.5% NP-40, 1mM $CaCl_2$, and protease inhibitors) and treated with 10 units/10^6 cells of micrococcal nuclease (Roche) for 5 minutes at 37°C. The reaction was stopped with 10mM EDTA. Nuclei were then incubated with RNaseA (Ambion) for 4 hours at 4°C with rotation followed by incubation with proteinase K (Ambion) overnight at 55°C. DNA was then isolated by Phenol:Chloroform:Isoamyl Alcohol (PCI) and EtOH precipitation. Equal MNase digestion was confirmed by examining DNA size fragments through electrophoresis on a 2% agarose gel and through bioanalyzer analysis. After phosphatase (NEB) treatment, digested DNA was end-repaired and A-tailed, with PCI extraction and EtOH precipitation following each enzyme reaction. Adaptors were ligated and DNA was size selected using Agencourt Ampure XP beads (Beckman Coulter), as previously described (Henikoff et al., 2011). Equal library sizes were confirmed through electrophoresis on a 2% agarose gel and through bioanalyzer analysis. Sequencing of 10 fragments per library confirmed the integrity and libraries were sent for paired-end (100 bp) high throughput sequencing using an Illumina HiSeq at the UMass Medical School sequencing facility. Reads were mapped to the mouse genome (mm9) using Bowtie2 and uniquely mapped reads were used for further analysis.

Data analysis

Assignment of reads to individual RSs:

Sequence reads were binned according to the 4 bp barcode present at the beginning of each sequence using a Perl script written in-house. Sequences with barcodes removed were mapped to the mouse genome (mm9) using Bowtie-0.12.7 (Langmead et al., 2009) with parameters set as -n 2 -l 28 -M 1 --best --strata (i.e. uniquely mapped with at most 2 mismatches at the left 28 bp seed region). Assignment of aligned sequences to individual restriction enzyme cut sites (REs) and differential cut analysis were performed using the Bioconductor package REDseq, developed by us. The ChIPpeakAnno package (Zhu et al., 2010) was used to annotate the differentially cut sites to the nearest genes. Surprisingly, we found that the GGTCC sequence was cleaved more efficiently by Sau96I than GGACC, GGGCC, or GGCCC in digestions of chromatin or naked DNA control samples. This may be due to the different buffer conditions used for digestion of chromatin (which are optimized for permeabilization of cells) relative to the optimal conditions for Sau96I digestion recommended by the manufacturer. However, this phenomenon was observed in all samples, independent of cell type or KD, and should therefore not affect any comparisons of accessibility.

Aggregation of RED-seq data at specific genomic regions:

Data for DNase I hypersensitive sites was downloaded from mouse ENCODE Project (UCSC). ChIP-seq data for H2A.Z (GSE34483), H3.3 (GSE16893), H3K4me3 (GSE12241) were downloaded from GEO datasets (NCBI) and analyzed in HOMER software suite (Heinz et al., 2010). The MNase-seq data in ESCs was obtained from Carone et al (Carone et al., 2014). The enrichment regions were identified by using the “findPeaks” command in HOMER with default setting (1. fold enrichment over local tag count, default: 4.0. 2. Poisson p-value threshold relative to local tag count, default: 0.0001 3. False discovery rate, default = 0.001). For the binding sites of different TFs (CTCF and Klf4) in ESCs, the enriched regions were obtained from GEO datasets (GSE11431) and converted to mm9 by LiftOver (UCSC Genome Bioinformatics Group).

Calculation of restriction enzyme accessibility:

RED-seq data was processed in HOMER by using “annotatePeaks” command to bin the regions of interest in 50 bp windows and sum the reads within each window. Average RE accessibility was calculated by normalizing the reads in each window to total reads, dividing by the number of regions of interest, and presented in reads per million. To calculate the genome-wide distribution of restriction enzyme sites, we manually assigned one read to each site and calculated average RE accessibility as mentioned above.

Measurement of Restriction enzyme accessibility at individual loci

DNA from RE-digested chromatin was prepared as above, up to the first DNA purification step (prior to library preparation). DNA was resuspended in 50μl of 0.1X TE and 10ng of DNA subjected was to quantitative PCR (qPCR) using SYBR FAST universal reagents (KAPA Biosystems) with specific primers (**Table I.2, 3**) flanking RSs of interest.

Name	Forward sequence	Reverse sequence
DHS-Chr1	CAGGTGGAGGGAGACAGAGA	TGGCAAGAAGGACTGTTGGT
DHS-Chr6	TGCTCTCTTCCACCCTCTGT	GGAGAGCCCTAGACTGGAAC
DHS-Chr10	CAGTGCCCAGACTCATCCTG	TGATGTCACATGCCTAGCCA
DHS-Chr11	TACCTGGAACCGGTTTACCC	CCTGAGCATCGACTCTCCAA
DHS-Chr12	GGTTGGAGTTGGAATCCGCA	AGCAAAGAAGGCACAGCTATG
DHS-Chr16	GCATGCCTCGTACCTGTGTA	GCAGGTAGGCGCTCTTACAG

Table I.2: Sequences of qPCR primers for DHSs

Name	Forward sequence	Reverse sequence
CTCF-Chr2	GGTTACCAAACCAGGTAAGATTTG	CTGCTCAGAAGAAGAAACTGGA
CTCF-Chr16	CATGTATGCCTACTTGCCAGA	TGCTCTTACAGAGGTCCTGA
CTCF-Chr15	CACCCACATGGCAGCTAATA	CTGCTTGTGTGTGCACTTTATG
CTCF-Chr4	ACCACTGACTGCTGAAAGTT	CTGTTAGAAGGACTGACTGGTG
CTCF-Chr1-1	ACTATGCATGCAGTACCTGTG	TGGCTCACGACTGTCTCTAA
CTCF-Chr1-2	GTACATGTAGTGCTCACAGAGG	CAATGGCTGCTCTTCCAGTA
CTCF-Chr7	TGCATCATGTGTGGACCTAAT	CAATTCCCAGAGCCCATGTA
CTCF-control1	GGTCCTGAACCCTTGAGAGA	GCCCTCTAGTGGCAAAGAAA
CTCF-control2	GGCGATGAACTTTACCCATC	TGTTTGCTATAAAACAGGACCAGA

Table I.3: Sequences of qPCR primers for CTCF binding sites

RNAi

RNAi-mediated KD of *Oct4*, *p400*, *Hira* or *GFP* (control) was performed using esiRNAs as described (Yang et al., 2002; Fazzio et al., 2008a). For differentiation experiments, *GFP* (control) or *Oct4* esiRNAs were transfected into ESCs using Lipofectamine 2000 (Invitrogen). Chromatin was isolated and used for RED-seq or MNase-seq library construction 5 days after transfection.

Chromatin Immunoprecipitation

ChIP samples were prepared as described (Chen et al., 2013). Briefly, chromatin from *GFP* or *Oct4* KD ESCs was crosslinked, lysed and sonicated to generate 300-1000 base-pair fragments. 50 µl of Protein A Magnetic beads (NEB) were washed twice with PBS containing 5mg/ml BSA and 10 µl of anti-CTCF antibody (Millipore) was coupled in 500µl PBS with 5mg/ml BSA overnight at 4°C. Immunoprecipitation was performed with antibody-coupled beads and sonicated supernatants in ChIP buffer (20mM Tris-HCl pH8.0, 150mM NaCl, 2mM EDTA, 1% Triton X-100) overnight at 4°C. Magnetic beads were washed twice with ChIP buffer, once with ChIP buffer including 500mM NaCl, 4 times with RIPA buffer (10mM Tris-HCl pH 8.0, 0.25M LiCl, 1mM EDTA, 0.5% NP-40, 0.5% Na·Deoxycholate), and once with TE buffer (pH 8.0). Chromatin was eluted twice from washed beads by adding elution buffer (20mM Tris-HCl pH 8.0, 100mM NaCl, 20mM EDTA, 1% SDS) and incubating for 15 minutes at 65°C.

Crosslinking was reversed at 65°C for 6hr and RNase A/ T1 (Ambion) was added for 1hr at 37°C followed by proteinase K (Ambion) treatment overnight at 50°C. ChIP-enriched DNA was purified using Phenol/Chloroform/Isoamyl alcohol extractions in phase-lock tubes. Then, chromatin was analyzed by qPCR as described above, using primers specific for CTCF sites of interest (**Table I.3**).

Data Access

The genome-wide data sets generated in this study can be obtained from GEO [GEO:GSE51821].

CHAPTER II:

PREFACE

Chapter II is reprinted from the following work with myself as the first author:

Hdac6 regulates Tip60-p400 function in stem cells

Poshen B. Chen, Jui-Hung Hung, Taylor L. Hickman, Andrew H. Coles, James F. Carey, Zhiping Weng, Feixia Chu and Thomas G. Fazzio

Elife. 2013 Dec 3; 2:e01557. doi: 10.7554/eLife.01557

Contributions:

Poshen Chen, Feixia Chu and Thomas Fazzio designed all experiments. Poshen carried out most experiments as well as most data analyses. Jui-Hung Hung performed the initial ChIP-seq data analysis with the assistance of Zhiping Weng. Isolation of NSCs and HPCs were performed by Poshen Chen and Andrew Coles. LC-MS/MS was performed by Taylor Hickman. James Carey developed and optimized the FLAG ChIP protocol. Thomas Fazzio performed the ESC gene targeting and initial purification. Poshen Chen and Thomas Fazzio wrote the paper.

CHAPTER II: Hdac6 regulates Tip60-p400 function in stem cells

ABSTRACT

In embryonic stem cells (ESCs), the Tip60 histone acetyltransferase activates genes required for proliferation and silences genes that promote differentiation. Here we show that the class II histone deacetylase Hdac6 co-purifies with Tip60-p400 complex from ESCs. Hdac6 is necessary for regulation of most Tip60-p400 target genes, particularly those repressed by the complex. Unlike differentiated cells, where Hdac6 is mainly cytoplasmic, Hdac6 is largely nuclear in ESCs, neural stem cells (NSCs), and some cancer cell lines, and interacts with Tip60-p400 in each. Hdac6 localizes to promoters bound by Tip60-p400 in ESCs, binding downstream of transcription start sites. Surprisingly, Hdac6 does not appear to deacetylate histones, but rather is required for Tip60-p400 binding to many of its target genes. Finally, we find that, like canonical subunits of Tip60-p400, Hdac6 is necessary for robust ESC differentiation. These data suggest that Hdac6 plays a major role in the modulation of Tip60-p400 function in stem cells.

INTROUDCTION

ESC self-renewal and differentiation are controlled by multiple pathways: exogenous factors that act through well-defined signaling pathways that are also employed in adult cells, and a network of nuclear factors that regulate the ESC transcriptome (Hanna et al., 2010). Regulators of gene expression can be further sub-divided into (i) sequence-specific transcription factors, including ESC-specific master regulators, (ii) non-coding RNAs that act both in cis and in trans to regulate specific subsets of genes, and (iii) chromatin regulatory complexes, most of which are expressed in multiple cell and tissue types, and often act very broadly in the genome to covalently modify histones, remodel nucleosomes, or modify higher-order chromatin folding (Young, 2011). A number of chromatin regulators have been identified from RNA-interference screens or traditional knockout studies that are important for various features of ESC identity. However, for most chromatin regulatory complexes, several key questions remain, including how they find their genomic targets, how their catalytic activities lead to alteration of gene expression, and how the activities of these factors are altered to facilitate differentiation.

In mammals, several chromatin remodeling complexes are modular, with distinct forms expressed in different cell or tissue types, or sometimes within the same cells (Fazzio and Panning, 2010; Ho et al., 2009a; Ramírez and Hagman, 2009; Wang et al., 1996). For example, the mammalian SWI/SNF complex BAF

(Brg1/Brahma Associated Factor) is actually a family of related complexes with many shared subunits, along with a few subunits that are specific to each particular cell type. In particular, when neural progenitors differentiate into neurons in mouse, two BAF subunits are replaced with two paralogous subunits that shift BAF from a factor promoting self-renewal to one that promotes differentiation (Lessard et al., 2007; Yoo et al., 2009). Another unique combination of subunits, different from those observed in differentiated cells, comprises BAF complex from ESCs (esBAF) (Ho et al., 2009a). Similarly, multiple forms of PRC1 (Polycomb Repressive Complex 1) have been purified from human and mouse cells that each contain the Ring1a/b ubiquitin ligase, but have different arrays of accessory proteins that confer distinct target specificity and activities (Gao et al., 2012; Tavares et al., 2012).

Tip60-p400 has been purified from cancer cell lines as a 17 subunit chromatin remodeling complex with two chromatin remodeling activities: the Tip60 (also known as Kat5) subunit acetylates the N-terminal tails of histones H2A, H4, and a number of transcription factors, while the p400 subunit mediates exchange of H2A-H2B dimers for H2A.Z-H2B dimers within nucleosomes (Cai et al., 2005; Doyon et al., 2004; Squatrito et al., 2006). In somatic cells, Tip60-p400 serves mainly as a transcriptional co-activator that functions with numerous sequence-specific transcription factors to activate gene expression (Baek et al., 2002; Brady et al., 1999; Frank et al., 2003; Legube et al., 2004). In contrast, while Tip60-p400 promotes expression of some genes required for cellular

proliferation and cell cycle regulation in ESCs, its most prominent function is to silence genes that are active during differentiation (Fazzio et al., 2008a; Fazzio et al., 2008b). RNAi-mediated knockdown (KD) of several Tip60-p400 subunits individually in ESCs induces a phenotype in which differentiation and ESC markers are expressed simultaneously, proliferation is reduced, the cell cycle is altered, and cells exhibit diminished self-renewal and pluripotency (Fazzio et al., 2008a). Consistent with these phenotypes, mice homozygous for a *Tip60* deletion allele die at the pre-implantation stage (Hu et al., 2009). It remains unknown why Tip60-p400 functions mainly as a repressor of differentiation gene expression in ESCs rather than an activator of expressed genes, as it does in most cell types examined.

Similarly, treatment of ESCs with Trichostatin A (TSA), a drug that broadly inhibits class I and II HDACs and results in elevated acetylation of most lysines targeted by HATs, promotes morphological changes similar to those observed upon KD of Tip60-p400 subunits (Karantzali et al., 2008; McCool et al., 2007). Therefore, maintenance of proper levels of histone acetylation appears to be essential to perpetuate the pluripotent state, as neither significant increases nor decreases in histone acetylation appear to be compatible with ESC self-renewal. However, TSA also inhibits several HDAC family members known to target acetylated lysines on non-histone proteins, leaving open the possibility that these targets play an equal or greater role in maintenance of ESC self-renewal. Furthermore, deletion or KD of several individual HDACs in ESCs produces

phenotypes that differ substantially from those of Tip60-p400 subunits (Dovey et al., 2010; McBurney et al., 2003). A more careful analysis of the individual contributions of the 18 HDACs expressed in mammals will be required to identify the key acetylation targets of each that are necessary for the maintenance of pluripotency and the ESC state.

Here we interrogate the composition of Tip60-p400 complex in mouse ESCs in order to identify unique interacting proteins that might account for its altered functional repertoire in this cell type. We find that the class II histone deacetylase (HDAC), Hdac6, is a stable interaction partner with Tip60-p400 in ESCs, but not mouse embryo fibroblasts (MEFs). Subsequent analyses revealed that Hdac6 also interacts with Tip60-p400 in adult neural stem cells from the brain, but is sequestered away from Tip60-p400 in the cytoplasm of most differentiated cell types, as previously reported (Hubbert et al., 2002; Kawaguchi et al., 2003; Valenzuela-Fernández et al., 2008; Verdel et al., 2000). We show that Hdac6 is necessary for regulation of most Tip60-target genes in ESCs, particularly differentiation genes repressed by Tip60-p400 in ESCs. Surprisingly, while its deacetylase domains are required for silencing of differentiation genes, Hdac6 does not regulate gene expression by deacetylating histones near the promoters of Tip60-p400 targets. Instead, the catalytic domains of Hdac6 are required for its interaction with Tip60-p400. Furthermore, we find that Hdac6 is necessary for normal Tip60-p400 enrichment at its gene targets, just downstream of their transcription start sites (TSSs). These data support a model in which

Hdac6 recruits Tip60-p400 complex to many target gene promoters. Finally, we show that Hdac6 is necessary for several major functions of Tip60-p400 in ESCs, as *Hdac6*-deficient ESCs are defective in silencing of differentiation genes, have a reduced proliferation rate and defects in formation of single colonies, and differentiate with altered kinetics. However, unlike KD of canonical Tip60-p400 subunits, Hdac6 KD does not prevent ESC self-renewal. Thus, Hdac6 is a component of a novel, stem cell-specific, form of Tip60-p400 complex that is necessary for gene regulation and normal differentiation in ESCs.

RESULTS

A class II HDAC, Hdac6, co-purifies with Tip60-p400 complex in ESCs and NSCs

Tip60-p400 complex has roles in both activation and repression of transcription in most cell types where it has been examined. However, in ESCs, Tip60-p400 is required for repression of many more genes than it activates, raising the possibility that a unique form of Tip60-p400 complex that might be expressed in ESCs that shifts the balance of its activity toward a more repressive role. To test this possibility, we targeted a 36 amino acid 6-histidine-3-FLAG (H3F) tag to the C-terminus of one copy of the endogenous *Tip60* gene in murine ESCs (Figure II.1A), performed double affinity purifications from tagged or untagged ESCs (Figure II.1B), and identified proteins that co-purified with Tip60

Figure II.1: Identification of Tip60-p400-interacting proteins in ESCs

(A) Shown are the 3’ end of the Tip60 gene plus downstream sequence (below) and the targeting construct for introducing the C-terminal tandem 6-His-3-FLAG (H3F) tag (above). The counter-selection cassette within the targeting construct is omitted for brevity. See methods for details of ESC targeting.

(B) Silver stained gel of purified Tip60 complex from *Tip60-H3F* ESCs and MEFs, along with untagged control cells.

(C) Validation of Hdac6 interaction. Western blots for Hdac6, p400, and Dmap1 following immunoprecipitation with anti-FLAG antibody from nuclear extracts derived from the indicated ESC lines.

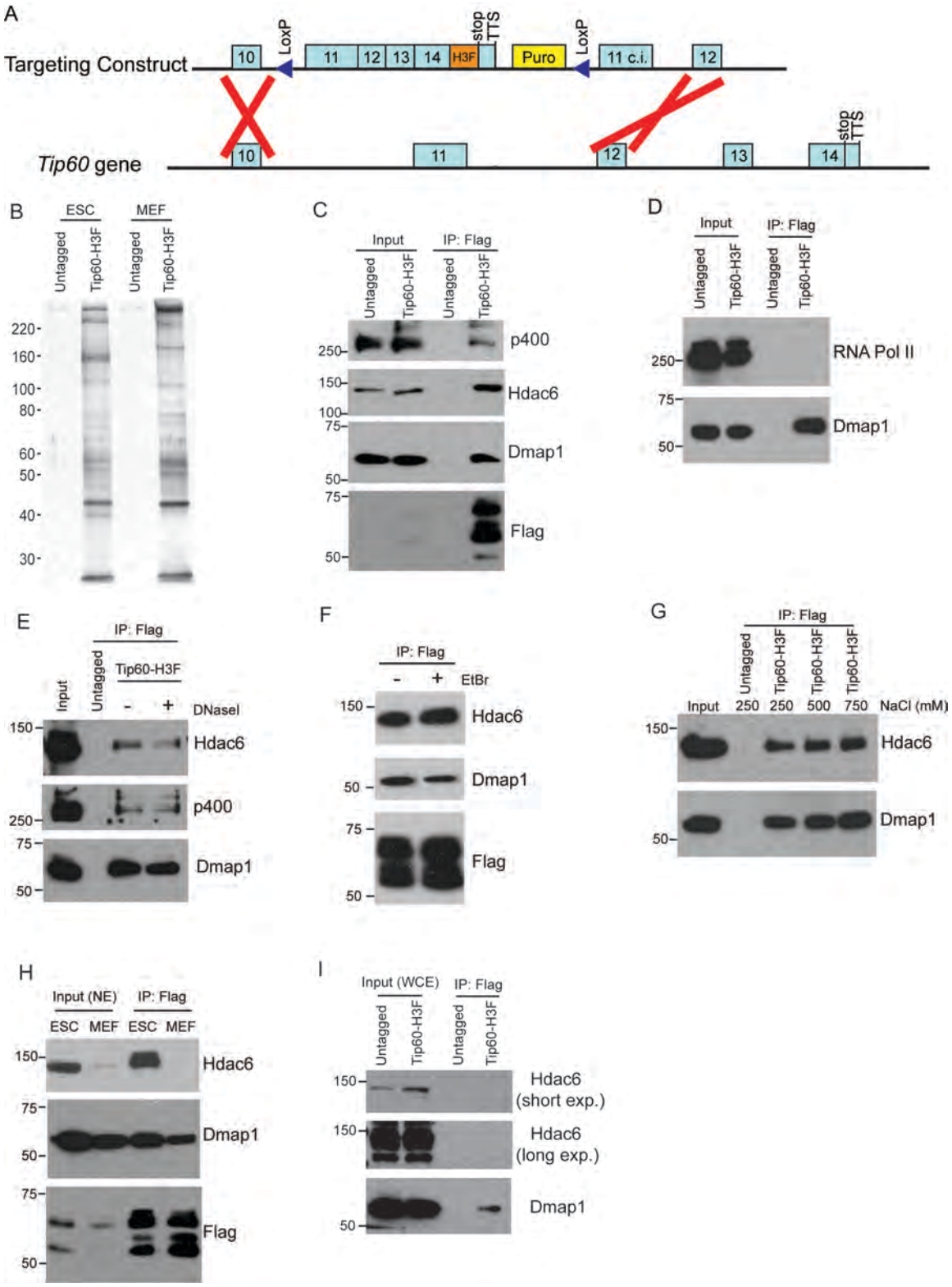
(D) Western blot of purified Tip60 complex from ESCs for RNA polymerase II (RNA Pol II) and Dmap1. (Tip60-p400 subunit Dmap1 is shown to confirm the presence of subunits of Tip60-p400 complex after purification.)

(E-F) Treatment with DNaseI (E) or ethidium bromide (F) does not affect Tip60-Hdac6 interaction. DnaseI or ethidium bromide was added into nuclear extracts during immunoprecipitation and the Tip60-interacting proteins were analyzed by Western blotting with the indicated antibodies.

(G) The Tip60-Hdac6 interaction is resistant to high salt. Tip60-H3F was purified from ESCs as above and subjected to washing with buffer containing the salt concentrations indicated.

(H) Tip60 complexes purified from *Tip60-H3F* ESC and MEF nuclear extracts were subjected to Western blotting for Hdac6, Dmap1, and FLAG.

(I) Tip60-H3F was immunoprecipitated from MEF whole cell extracts as above, and blotted for the indicated proteins.


A
Targeting Construct
10
LoxP
11
12
13
14
H3F
stop
TTS
Puro
LoxP
11 c.i.
12
Tip60 gene
10
11
12
13
14
stop
TTS
B
ESC
MEF
Untagged
Tip60-H3F
Untagged
Tip60-H3F
220
160
100
80
60
50
40
30
C
Input
IP: Flag
Untagged
Tip60-H3F
Untagged
Tip60-H3F
250
150
100
75
50
75
50
p400
Hdac6
Dmap1
Flag
D
Input
IP: Flag
Untagged
Tip60-H3F
Untagged
Tip60-H3F
250
75
50
RNA Pol II
Dmap1
E
IP: Flag
Input
Untagged
Tip60-H3F
-
+
DNaseI
150
250
75
50
Hdac6
p400
Dmap1
F
IP: Flag
-
+
EtBr
150
50
75
50
Hdac6
Dmap1
Flag
G
IP: Flag
Untagged
Tip60-H3F
Tip60-H3F
Tip60-H3F
Input
250
250
500
750
NaCl (mM)
150
50
Hdac6
Dmap1
H
Input (NE)
IP: Flag
ESC
MEF
ESC
MEF
150
75
50
75
50
Hdac6
Dmap1
Flag
I
Input (WCE)
IP: Flag
Untagged
Tip60-H3F
Untagged
Tip60-H3F
150
150
50
Hdac6
(short exp.)
Hdac6
(long exp.)
Dmap1

using LC-MS/MS (Table II.1). By this approach, we identified 16 of 17 known subunits (Altaf et al., 2009; Cai et al., 2003; Cai et al., 2005; Doyon et al., 2004; Ikura et al., 2000) of Tip60-p400, suggesting that expression of the tagged form of Tip60 from its endogenous locus allowed for normal complex formation. Furthermore, we observed a number of novel Tip60-p400-interacting proteins, including chromatin regulatory factors and transcription factors, among others. To test for cell type specificity of Tip60-p400 complexes we generated a knock-in mouse harboring *Tip60-H3F*, isolated embryonic fibroblasts and repeated the purification. We observed several bands within Tip60 purifications from ESCs that were not observed in purifications from *Tip60-H3F* MEFs or untagged cells (Figure II.1B), consistent with the possibility that ESCs express a distinct form of Tip60-p400 complex.

We were intrigued by the finding that Hdac6 co-purified with Tip60-p400 in ESCs (Table II.1). Hdac6 is a class II histone deacetylase (HDAC) that is expressed in many different cell types but is usually localized to the cytoplasm (Hubbert et al., 2002; Kawaguchi et al., 2003; Valenzuela-Fernández et al., 2008; Verdel et al., 2000), as are its well-established substrates: a-tubulin (Hubbert et al., 2002), Hsp90 (Kovacs et al., 2005), and cortactin (Zhang et al., 2007). Moreover, despite its homology to proteins that deacetylate histone tails, Hdac6 has not been found to target histones in vivo (Haggarty et al., 2003).

Name	Description	# of Peptides	MW (kD)	Peptides/MW	Gel Slices
Stk38	Serine/threonine-protein kinase 38	78	54	1.44	7,14,15,16,17
Ruvbl1	RuvB-like 1	42	50	0.84	7,8,13,15,16,17
Ruvbl2	RuvB-like 2	38	51	0.75	15,16,17
Acta1 or other iso	Actin	21	42	0.50	18
Sun2	Protein unc-84 homolog B	36	82	0.44	11,12,13
Hdac6	Histone deacetylase 6	38	126	0.30	1,2,3,4,5,6,7,8
Kat5	Histone acetyltransferase KAT5	17	59	0.29	13,14,15
Actb	Actin, cytoplasmic 1	12	42	0.29	5,8,9
Trrap	Transformation/transcription domain-associated protein	72	292	0.25	1,2
Epc1	Enhancer of polycomb homolog 1	22	90	0.24	9,10
Brd8	Bromodomain-containing protein 8	24	103	0.23	6,7
Yeats4	YEATS domain-containing protein 4	6	27	0.22	20
Epc2	Enhancer of polycomb homolog 2	20	91	0.22	10,11
H2B	Histone H2B	3	14	0.21	21
Ing3	Inhibitor of growth protein 3	10	47	0.21	16,17
Ep400	E1A-binding protein p400	70	337	0.21	1,2
Dmap1	DNA methyltransferase 1-associated protein 1	10	53	0.19	14,15
Hspa8	Heat shock cognate 71 kDa protein	13	69	0.19	13
Lima1	LIM domain and actin-binding protein 1	15	84	0.18	9,10
Vps72	Vacuolar protein sorting-associated protein 72 homolog	7	41	0.17	16,17
Actl6a	Actin-like protein 6A	8	47	0.17	17,18
Actg1	Actin, cytoplasmic 1	6	42	0.14	19
H2afv	Histone H2A.V	2	14	0.14	22
Meaf6	Chromatin modification-related protein MEAF6	3	22	0.14	20
Mbtd1	MBT domain-containing protein 1	8	71	0.11	13,14
Rps18	40S ribosomal protein S18	2	18	0.11	21
Tubb5	Tubulin beta-5 chain	5	50	0.10	16
Tuba1a or other iso	Tubulin alpha chain	5	50	0.10	16
Trim28	Transcription intermediary factor 1-beta	8	89	0.09	9,10
Morf4l2	Mortality factor 4-like protein 2	3	34	0.09	19
Mrgbp	MRG-binding protein	2	24	0.08	20
Rangap1	Ran GTPase-activating protein 1	5	64	0.08	12
Spna2	Spectrin alpha chain, brain	18	285	0.06	3
Setx	Probable helicase senataxin	18	298	0.06	2,3
Sf3b1	Splicing factor 3B subunit 1	3	54	0.06	6
Sfpq	Splicing factor, proline- and glutamine-rich	4	75	0.05	10
Rab5c	Ras-related protein Rab-5C	1	23	0.04	20
Lrrfip2	Leucine-rich repeat flightless-interacting protein 2	2	47	0.04	16
Stat2	Signal transducer and activator of transcription 2	2	50	0.04	8
Nono	Non-POU domain-containing octamer-binding protein	2	55	0.04	15
Tpr	Nucleoprotein TPR	6	274	0.02	3
Hnrnpf	Heterogeneous nuclear ribonucleoprotein F	1	46	0.02	17
Spnb2	Spectrin beta chain, brain 1	5	274	0.02	3
Flna	Filamin-A	4	281	0.01	3
Flii	Protein flightless-1 homolog	2	145	0.01	6
Hdx	Highly divergent homeobox	1	77	0.01	11

Table II.1: Proteins co-purifying with Tip60-H3F in ESCs

*Proteins in blue represent known Tip60-p400 subunits

To confirm that Hdac6 is a *bona fide* Tip60-p400 interacting protein, we performed reciprocal co-immunoprecipitation experiments in ESCs, observing the Tip60-Hdac6 interaction no matter which protein was immunoprecipitated (Figures II.1C). Previously, both Hdac6 and Tip60 were separately found to interact with RNA Polymerase II (RNA Pol II) in human $CD4^+$ T-cells (Wang et al., 2009), raising the possibility that the Hdac6-Tip60 interaction we observed in ESCs might be mediated by RNA Pol II. However, there were no peptides corresponding to Pol II in our LC-MS/MS data, and we could not detect Pol II in Tip60-H3F immunoprecipitates (Figure II.1D), arguing against this explanation. Furthermore, the Tip60-Hdac6 interaction was independent of DNA and resistant to high salt (Figure II.1E-G), verifying that Hdac6 is a stable interaction partner within Tip60-p400 complex. Finally, we tested whether Hdac6 interacts with Tip60-p400 complex in a differentiated cell type, MEFs. Unlike ESCs, Tip60-H3F immunoprecipitated from MEF nuclear extracts or whole cell lysates did not pull down Hdac6 (Figures II.1H-I), suggesting that Hdac6 interacts with Tip60-p400 complex in only a subset of cell types.

The reported cytoplasmic localization of Hdac6 in multiple types of cells (Hubbert et al., 2002; Kawaguchi et al., 2003; Valenzuela-Fernández et al., 2008; Verdel et al., 2000) raised the question of whether its interaction with Tip60-p400 complex was physiologically relevant. To address this issue, we first confirmed that Hdac6 exhibited significant nuclear localization in ESCs (Figure II.2A).

Figure II.2: Hdac6 is partially nuclear in multiple types of undifferentiated cells and interacts with Tip60-p400 in NSCs

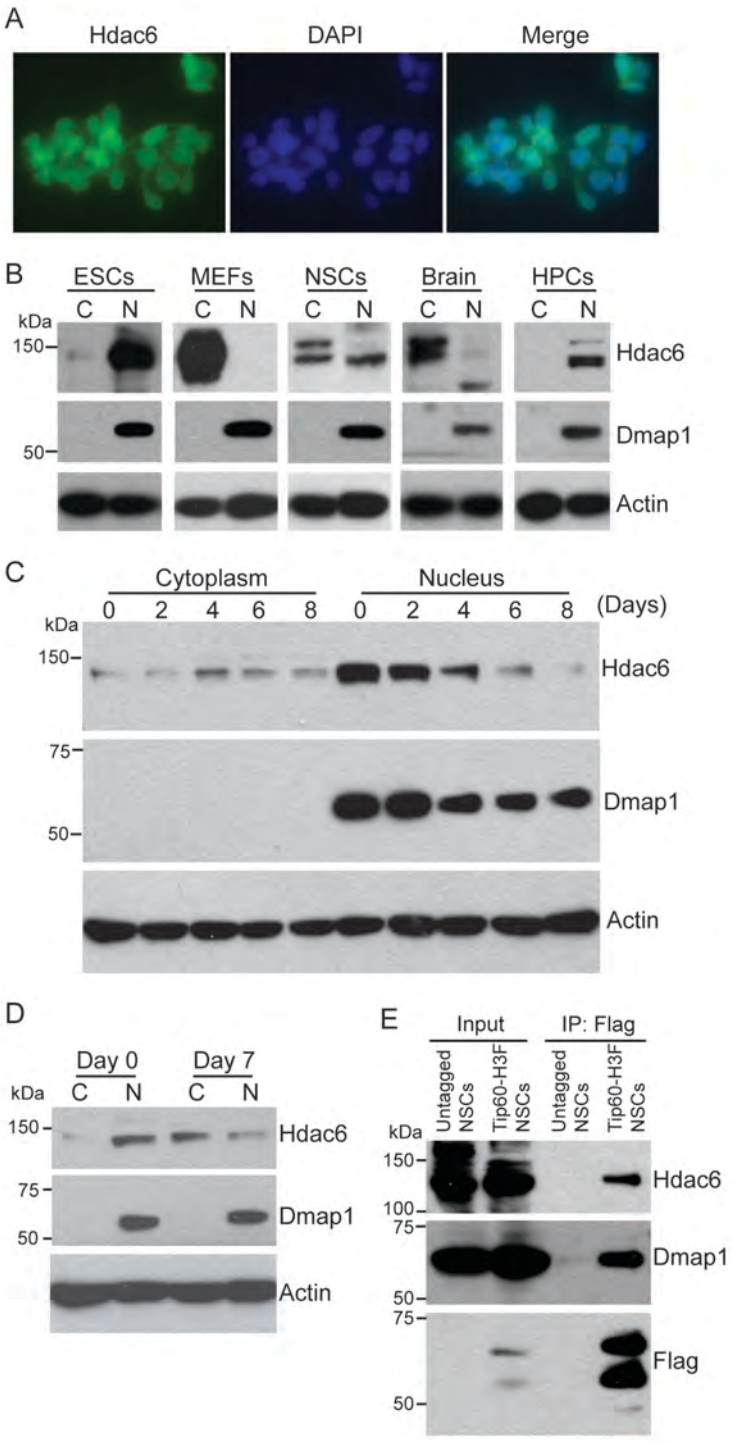
(A) ESCs were subjected to immunofluorescence using an antibody recognizing Hdac6. DAPI staining is shown to identify nuclei.

(B) High levels of nuclear Hdac6 in stem cells, but not differentiated cells. Western blots of Hdac6 and known Tip60-p400 subunit Dmap1 in cytoplasmic (C) and nuclear (N) fractions of indicated cells are shown, with β-actin serving as a loading control.

(C) Hdac6 relocalizes to the cytoplasm during ESC differentiation. Cytoplasmic (C) and nuclear (N) fractions from ESCs differentiated for the indicated numbers of days were Western blotted for the indicated proteins.

(D) Hdac6 relocalizes to the cytoplasm during NSC differentiation. Cytoplasmic (C) and nuclear (N) fractions from undifferentiated NSCs (day 0) or NSCs differentiated for 7 days were Western blotted for the indicated proteins.

(E) Hdac6 interacts with Tip60-p400 in NSCs. Shown are Western blots for the proteins indicated of input or Tip60-p400 complex immunoprecipitated from NSCs.


A
Hdac6
DAPI
Merge
B
ESCs
MEFs
NSCs
Brain
HPCs
C N
kDa
150
50
Hdac6
Dmap1
Actin
C
Cytoplasm
Nucleus
0 2 4 6 8 0 2 4 6 8 (Days)
kDa
150
75
50
Hdac6
Dmap1
Actin
D
Day 0
Day 7
C N C N
kDa
150
75
50
Hdac6
Dmap1
Actin
E
Input
IP: Flag
Untagged NSCs
Tip60-H3F NSCs
kDa
150
100
75
50
75
50
Hdac6
Dmap1
Flag

Next, we prepared cytoplasmic and nuclear protein fractions to examine the cellular localization of Hdac6 in ESCs and several adult cell types by Western blotting. Interestingly, while differentiated cells (MEFs, whole brain) exhibited the reported cytoplasmic sequestration, high levels of Hdac6 in undifferentiated cells, including ESCs, NSCs, and hematopoietic stem and progenitor cells (HPCs), were found in the nucleus (Figure II.2B). Consistent with these data, differentiation of either ESCs or NSCs caused a dramatic decrease in nuclear Hdac6, accompanied by increased Hdac6 within the cytoplasm (Figure II.2C-D). To test whether nuclear localization of Hdac6 promoted its interaction with Tip60-p400 complex, we immunoprecipitated Tip60 from NSCs isolated from *Tip60-H3F* knock-in mice. Indeed, Hdac6 was present in Tip60-H3F immunoprecipitates from NSC nuclear extracts (Figure II.2E). These data suggest that the interaction of Hdac6 with Tip60-p400 in undifferentiated cells is lost during the course of certain types of differentiation, due to nuclear exclusion of Hdac6. However, cytoplasmic sequestration in differentiated cells cannot be the sole factor preventing Hdac6 from associating with Tip60-p400, since we did not observe Hdac6 within Tip60-H3F immunoprecipitates from whole cell lysates (in which nuclear and cytoplasmic proteins are mixed; Figure II.1I). As an orthogonal approach, we used glycerol gradient sedimentation to examine the extent to which Tip60, p400, and Hdac6 co-fractionate in ESCs and MEFs. Consistent with our co-immunoprecipitation data, we found that a small portion of Hdac6 co-migrated with Tip60 and p400 within higher molecular weight fractions

in ESCs (Figure II.S1A). In contrast, we observed very little overlap between fractions containing Hdac6 and Tip60-p400 from MEFs (Figure II.S1B), consistent with our finding that Hdac6 does not interact with Tip60-p400 complex in this cell type. Together, these data show that Hdac6 exhibits significant nuclear localization in some types of embryonic and adult stem and progenitor cells, where it associates with Tip60-p400 complex. Furthermore, stem cell differentiation promotes re-localization of Hdac6 to the cytoplasm, where it is sequestered away from Tip60-p400.

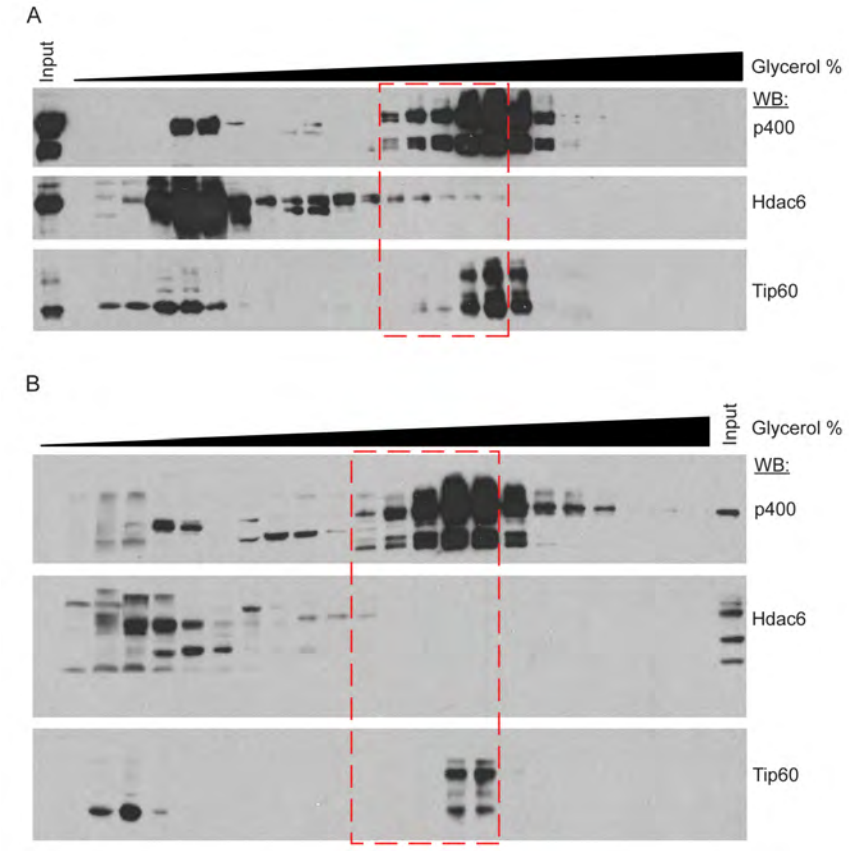


Figure II.S1: A portion of Hdac6 co-fractionates with Tip60-p400 in ESCs

Glycerol gradient sedimentation analyses from ESC (A) or MEF (B) protein extracts.

Hdac6 is necessary for normal regulation of most Tip60-p400 target genes

To test whether Hdac6 is necessary for gene regulation by Tip60-p400 complex in ESCs, we used DNA microarrays to examine the changes in mRNA levels upon *Tip60* or *Hdac6* KD. We observed highly correlated gene expression profiles in ESCs knocked down individually for *Tip60* and *Hdac6* (R=0.63), suggesting a significant overlap in their sets of target genes, although *Hdac6* KD generally had weaker effects on expression of common targets (Figure II.3A-B). Double KD of *Tip60* and *Hdac6* was nearly identical to the *Tip60* single KD, consistent with the model that Hdac6 functions within Tip60-p400 complex (Figure II.3C). Next we performed unsupervised hierarchical clustering of mRNA expression data comparing ESCs depleted of Tip60, Hdac6, or both to control KD ESCs. We observed four main clusters of genes differentially expressed in *Tip60* KD cells: genes upregulated upon *Tip60* KD but unaffected by *Hdac6* KD (Figure II.3C, cluster 1), genes upregulated in both single KDs (cluster 2), genes downregulated in *Tip60* KD cells but unaffected by *Hdac6* KD (cluster 3), and genes downregulated in both single KDs (cluster 4). Genes downregulated upon Tip60 KD were significantly overrepresented (relative to the number expected by chance) among Hdac6-independent Tip60 targets, while genes upregulated upon Tip60 KD were significantly overrepresented among Hdac6-dependent target genes, suggesting that Tip60-dependent repression in ESCs usually requires Hdac6 (Figure II.3D). We tested individual genes from each cluster by RT-qPCR, which generally confirmed these results (Figure II.3E).

Figure II.3: Overlapping effects of *Hdac6* KD and *Tip60* KD on gene expression in ESCs

(A) Scatter plot of gene expression (Log2 fold change relative to control KD ESCs) upon *Tip60* KD relative to *Hdac6* KD. Genes misregulated upon *Tip60* KD (adjusted p value < 0.1) are shown in red.

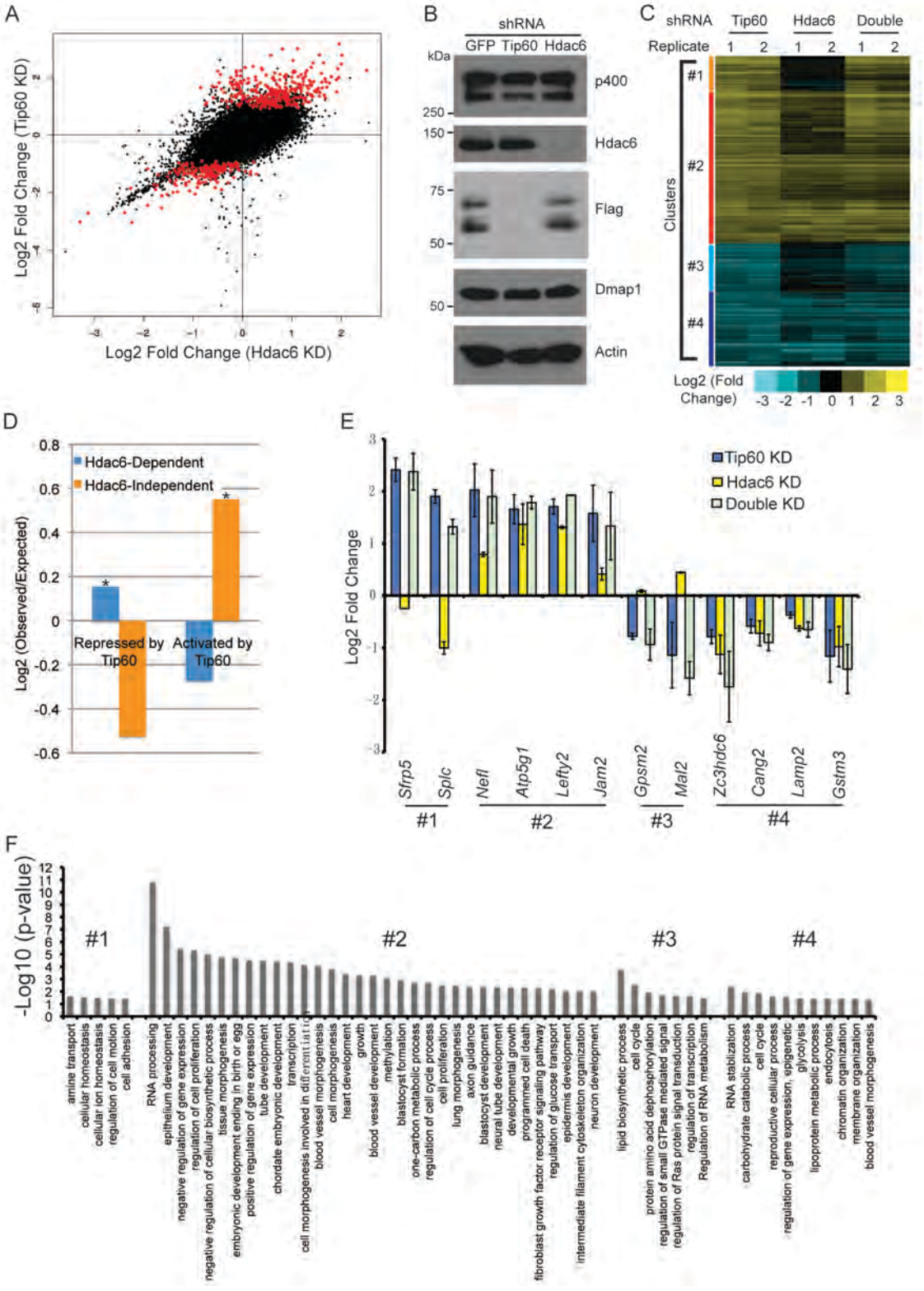
(B) Western blot showing the levels of p400, Dmap1, Hdac6 and Tip60 (FLAG) upon *Tip60* or *Hdac6* KD. β-actin is shown as a loading control.

(C) Unsupervised hierarchical clustering of genes misregulated (adjusted p-value < 0.1) upon *Tip60* KD. Up-regulated genes are indicated in yellow and downregulated genes are indicated in blue. The first cluster (#1) includes 200 genes that were upregulated upon *Tip60* KD but not *Hdac6* KD, the second cluster (#2) includes 867 genes upregulated in both *Tip60* KD and *Hdac6* KD ESCs, the third cluster (#3) includes 277 genes that were downregulated only in *Tip60* KD cells and the forth cluster (#4) includes 424 genes that were downregulated in both *Tip60* KD and *Hdac6* KD ESCs.

(D) Hdac6-dependent target genes are biased toward genes repressed by Tip60. Genes repressed or activated by Tip60 were split based on their Hdac6-dependence, and each group was plotted as the Log2 (ratio) of genes observed in each group relative to the expected number of genes if Hdac6-dependence was randomly distributed. Asterisk indicates statistically significant enrichment ($p < 10^{-20}$).

(E) Validation of microarray datasets. The expression levels of genes from each cluster were measured by RT-qPCR in the indicated KDs and expressed as Log2 (fold change) values relative to control (*GFP*) KD ESCs after normalization. Data shown are mean ± SD of three technical replicates from one representative experiment of two biological replicates performed.

(F) The significance of enrichment [- Log10 (p-value)] of Gene Ontology (GO) terms overrepresented in each cluster of genes in Figure II.3C. GO terms that were partially redundant with those listed were eliminated for brevity.


A
Log2 Fold Change (Tip60 KD)
Log2 Fold Change (Hdac6 KD)
B
shRNA
GFP Tip60 Hdac6
kDa
p400
Hdac6
Flag
Dmap1
Actin
C
shRNA
Tip60
Hdac6
Double
Replicate
Clusters
Log2 (Fold Change)
D
Hdac6-Dependent
Hdac6-Independent
Log2 (Observed/Expected)
Repressed by Tip60
Activated by Tip60
E
Tip60 KD
Hdac6 KD
Double KD
Log2 Fold Change
Sfrp5
Splc
Nefl
Atp5g1
Lefty2
Jam2
Gpsm2
Mal2
Zc3hdc6
Cang2
Lamp2
Gstm3
F
-Log10 (p-value)
amine transport
cellular homeostasis
cellular ion homeostasis
regulation of cell motion
cell adhesion
RNA processing
epithelium development
negative regulation of gene expression
regulation of cell proliferation
negative regulation of cellular biosynthetic process
tissue morphogenesis
embryonic development ending in birth or egg
positive regulation of gene expression
tube development
chordate embryonic development
transcription
cell morphogenesis involved in differentiation
blood vessel morphogenesis
cell morphogenesis
heart development
growth
blood vessel development
methylation
blastocyst formation
one-carbon metabolic process
regulation of cell cycle process
cell proliferation
lung morphogenesis
axon guidance
blastocyst development
neural tube development
developmental growth
programmed cell death
fibroblast growth factor receptor signaling pathway
regulation of glucose transport
epidermis development
intermediate filament cytoskeleton organization
neuron development
lipid biosynthetic process
cell cycle
protein amino acid dephosphorylation
regulation of small GTPase mediated signal
regulation of Ras protein signal transduction
regulation of transcription
Regulation of RNA metabolism
RNA stabilization
carbohydrate catabolic process
cell cycle
reproductive cellular process
regulation of gene expression, epigenetic
glycolysis
lipoprotein metabolic process
endocytosis
chromatin organization
membrane organization
blood vessel morphogenesis
#1
#2
#3
#4

Furthermore, we validated these results using two *Hdac6* mutant ESC lines: one that harbors a hypomorphic allele that expresses reduced Hdac6 levels (*Hdac6*^{reduced}*;* Figure II.S2A), and one that contains an Hdac6 deleteion (*Hdac6*^{null}*;* Figure II.S2B). Both lines exhibited misregulation of Tip60/Hdac6 target genes similar to that observed upon Hdac6 KD, with greater effects usually observed in the more severe *Hdac6*^{null} line (Figure II.S2C-D). Finally, we examined which classes of genes were enriched in each regulatory cluster. While Hdac6-independent targets of Tip60-p400 were enriched for genes involved in cellular growth, homeostasis, and the cell cycle (Figure II.3F, clusters 1 and 3), the majority of Hdac6-dependent Tip60-p400 target genes were differentiation-induced genes (Figure II.3F, cluster 2), consistent with the idea that Hdac6 is broadly important for Tip60-p400-dependent repression of developmental genes, although it also plays a smaller role in activation of some Tip60-dependent proliferation genes (Figure II.3F, cluster 4).

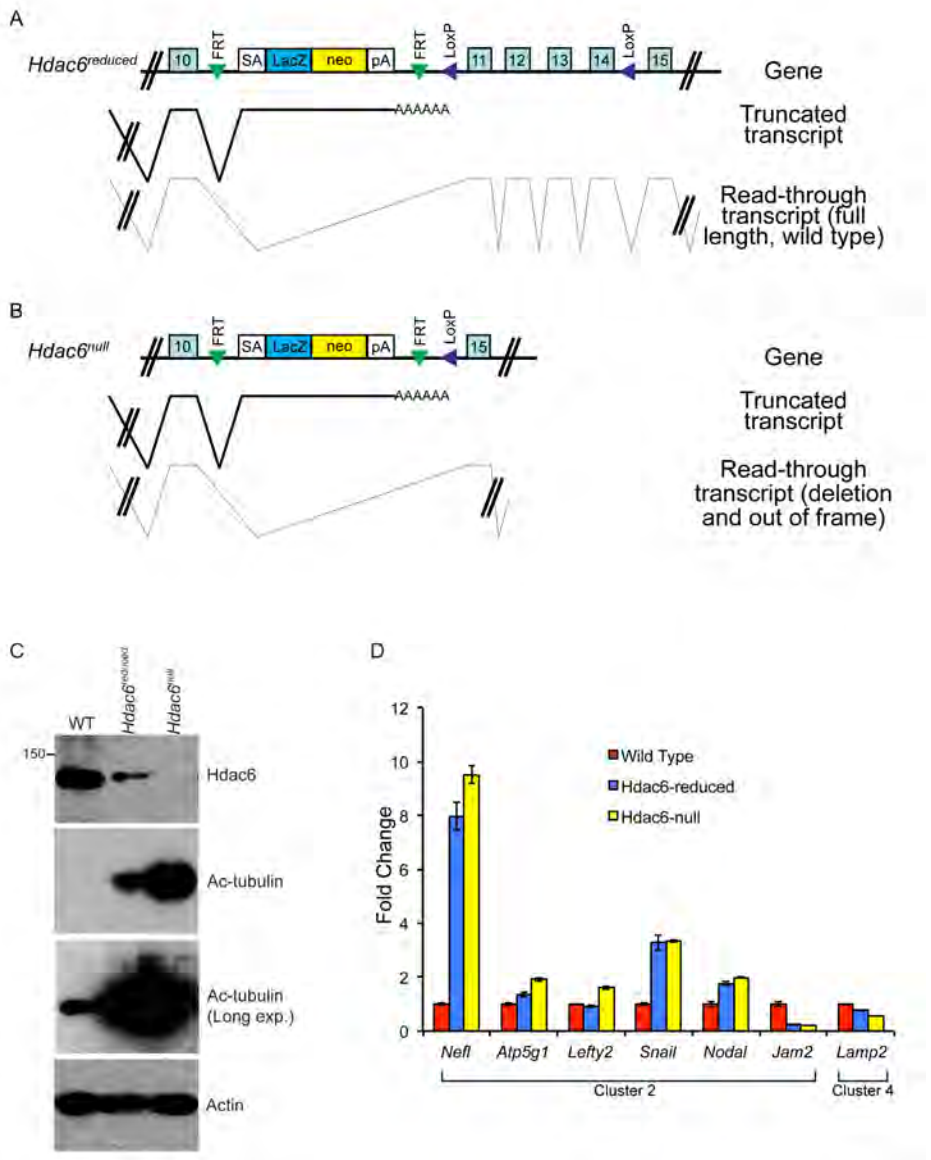


Figure II.S2: *Hdac6* mutant ESCs exhibit alterations in Tip60 target gene expression consistent with the KD phenotypes

(A-B) Diagram of Hdac6 mutant ESC line *Hdac6*$^{tm1a(EUCOMM)Wtsi}$ before (A; hereafter, *Hdac6*^{reduced}) and after (B; hereafter, *Hdac6*^{null}) CRE expression. Since, prior to CRE expression, this line produces a low level of Hdac6 protein, we infer that some transcripts fail to include the gene trap allele, and therefore produce full-length, wild type transcript. SA: splice acceptor; pA: cleavage and polyadenylation sequence.

(C) *Hdac6*$^{tm1a(EUCOMM)Wtsi}$ ESCs are hypomorphic prior to CRE expression and null alleles after. Whole cell extracts from normal (WT) ESCs or *Hdac6* mutant ESCs diagrammed above were prepared, followed by Western blotting for the proteins indicated.

(D) Hdac6 mutant ESCs exhibit defects in gene regulation similar to Hdac6 KD cells. Cells described above were harvested for RNA and subjected to RT-qPCR for the genes indicated.

Hdac6 binding overlaps one of two peaks of Tip60 enrichment at common target promoters

Despite overlapping changes in gene expression upon KD of *Tip60* or *Hdac6* in ESCs, the functional role of Hdac6 within Tip60-p400 complex remained unclear. Hdac6 is not thought to bind chromatin or regulate gene expression in most cell types (Haggarty et al., 2003; Hubbert et al., 2002; Verdel et al., 2000), therefore we considered the possibility that Hdac6 modulates Tip60-p400 function prior to chromatin binding by the complex. To determine whether Hdac6 associates with chromatin-bound Tip60-p400, we tested whether Hdac6 co-localizes with Tip60-p400 on chromatin. To this end, we examined the genome-wide distributions of Tip60 and Hdac6 in ESCs using chromatin immunoprecipitation followed by deep sequencing of the precipitated DNA (ChIP-seq). To facilitate these analyses, we generated an ESC line in which the H3F tag utilized above was fused to the C-terminus of Hdac6 at the endogenous *Hdac6* locus. This line allowed us to directly compare the genomic binding profiles of Tip60 and Hdac6 with untagged control cells using the same antibody, thereby eliminating differences in background.

We found that Tip60 binding was enriched at the 5' ends of many genes in ESCs, with two peaks of binding flanking the promoter regions of most targets, one at approximately 400 base pairs upstream and another at approximately 100 base pairs downstream of the TSS (Figures II.4A-B).

Figure II.4: Tip60 and Hdac6 co-localize on chromatin

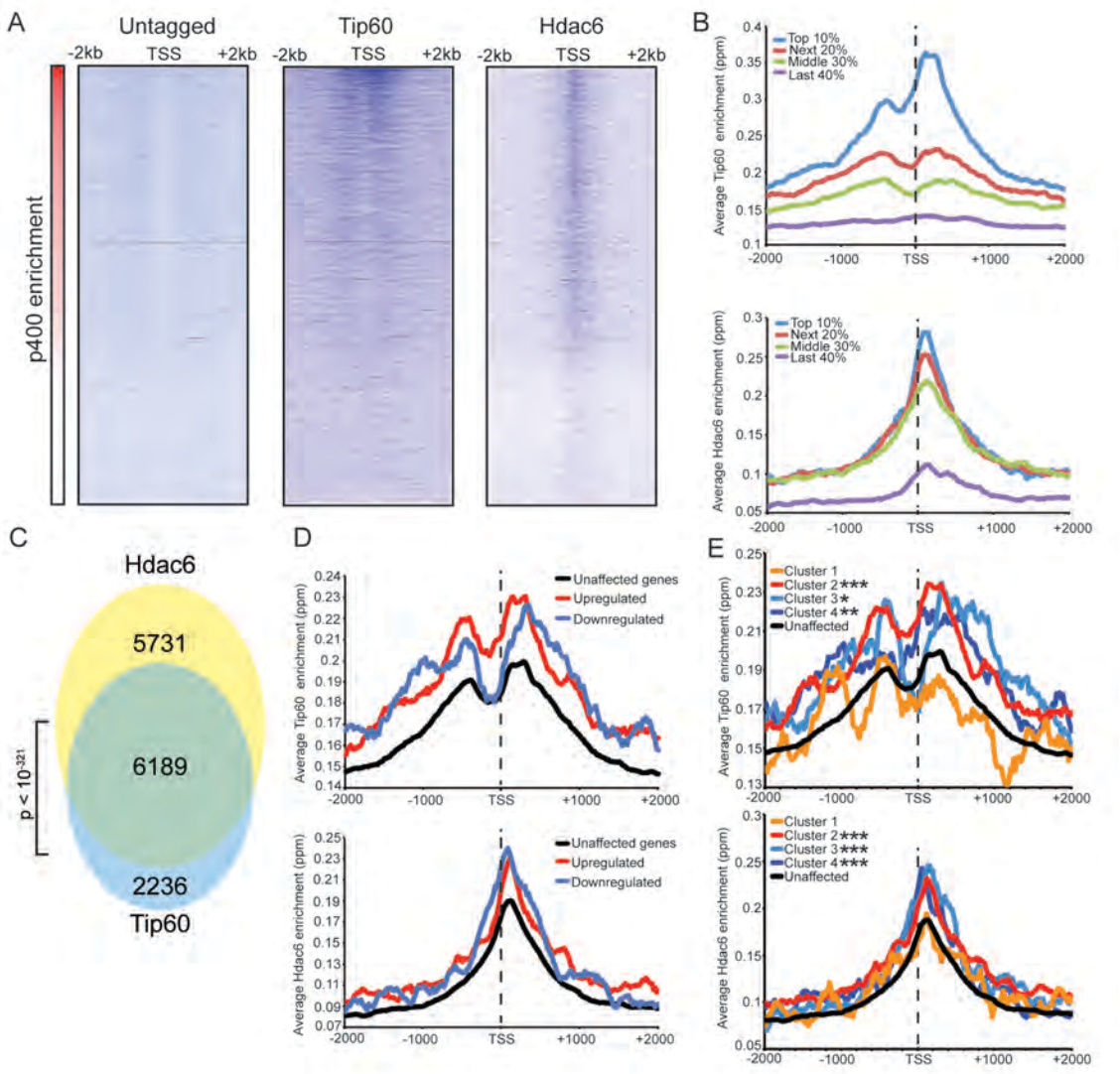
(A) Heat map representation of ChIP-seq data for H3F-tagged Tip60 and Hdac6 comprising the 2kb surrounding the transcriptional start sites (TSS) of 10,507 genes for which published p400 ChIP-chip data (Fazzio et al., 2008a) (p400 enrichment) were available. ChIP-seq data from E14 (Untagged) cells is shown as a control. All panels are sorted by decreasing p400 binding for the 1kb surrounding the TSS, ranging from high levels of p400 binding (red) to genes unbound by p400 (white).

(B) Tip60 and Hdac6 binding correlate with p400 binding. Genes in the p400 ChIP-chip dataset were grouped by the intensity of p400 enrichment: The groups of genes exhibiting the top 10% of p400 enrichment (Top 10%), the 11^{th}-30^{th} percentile (Next 20%), the 31^{st}-60^{th} percentile (Middle 30%) and the rest of genes in dataset (Last 40%). Upper panel: Averaged Tip60 enrichment for groups of genes at each level of p400 binding are shown relative to the TSS. Lower panel: Averaged Hdac6 binding data for genes in the same groups.

(C) Correlation of Tip60 and Hdac6 binding. Shown is a Venn diagram delineating the overlap between the gene sets bound by Tip60 and Hdac6. The p-value was calculated by summing the hypergeometric probabilities of Tip60/Hdac6 overlap below the number observed and subtracting from one.

(D) Hdac6 and Tip60 are enriched at genes regulated by these factors. Upper panel: Tip60 binding segregated by genes that are upregulated, downregulated, or unchanged by Tip60 KD. Bottom panel: Hdac6 binding at genes segregated as in upper panel.

(E) Tip60 and Hdac6 are enriched at genes within clusters 2, 3, and 4. Tip60 (upper panel) and Hdac6 (bottom panel) binding data are shown for genes segregated by cluster, as in Figure 2C. Asterisks mark clusters exhibiting statistically significantly higher promoter-proximal (-500 to +500) binding of indicated factor than does the set of genes not regulated by *Tip60* and *Hdac6*: *($p < 0.05$); **($p < 0.01$); ***($p < 10^{-5}$).


A
Untagged
Tip60
Hdac6
-2kb
TSS
+2kb
p400 enrichment
B
Top 10%
Next 20%
Middle 30%
Last 40%
Average Tip60 enrichment (ppm)
Average Hdac6 enrichment (ppm)
C
Hdac6
5731
6189
2236
Tip60
p < 10-321
D
Unaffected genes
Upregulated
Downregulated
E
Cluster 1
Cluster 2***
Cluster 3*
Cluster 4**
Cluster 3***
Cluster 4***
Unaffected

This two peak pattern of Tip60 binding and the gene set bound by Tip60 were very similar to previous mapping data examining the distribution of the p400 subunit of Tip60-p400 complex in ESCs (Fazzio et al., 2008a). Interestingly, we found that while Hdac6 was also enriched at p400- (Figures II.4A-B) and Tip60-target genes (Figure II.4C) near their TSSs, its pattern of binding was very different from Tip60 and p400, forming one peak of enrichment at approximately 100 base pairs downstream of the TSS that overlapped with the downstream Tip60 peak (Figure II.4A-B). Like p400 (Fazzio et al., 2008a), we found that Hdac6 was enriched at genes marked by H3K4me3, including bivalent genes also marked by H3K27me3 (Figure II.S3). Furthermore, we found that, on average, both Hdac6 and Tip60 were both enriched to significantly higher levels at the promoter regions of genes that are misregulated upon *Hdac6* or *Tip60* KD compared to genes unaffected by KD of these factors (Figure II.4D), suggesting that many of these genes are direct targets. Interestingly, while levels of Hdac6 binding were elevated at genes from clusters 2 and 4 (Figure II.3C), whose expression levels are regulated by Hdac6, they were also elevated at cluster 3 genes, whose expression levels are Tip60-dependent but Hdac6-independent (Figure II.4E), suggesting that Tip60-p400 can act independently of Hdac6 to activate these genes. Together, these data show that Hdac6 binds in an asymmetrical pattern with respect to the transcription start site at its genomic targets.

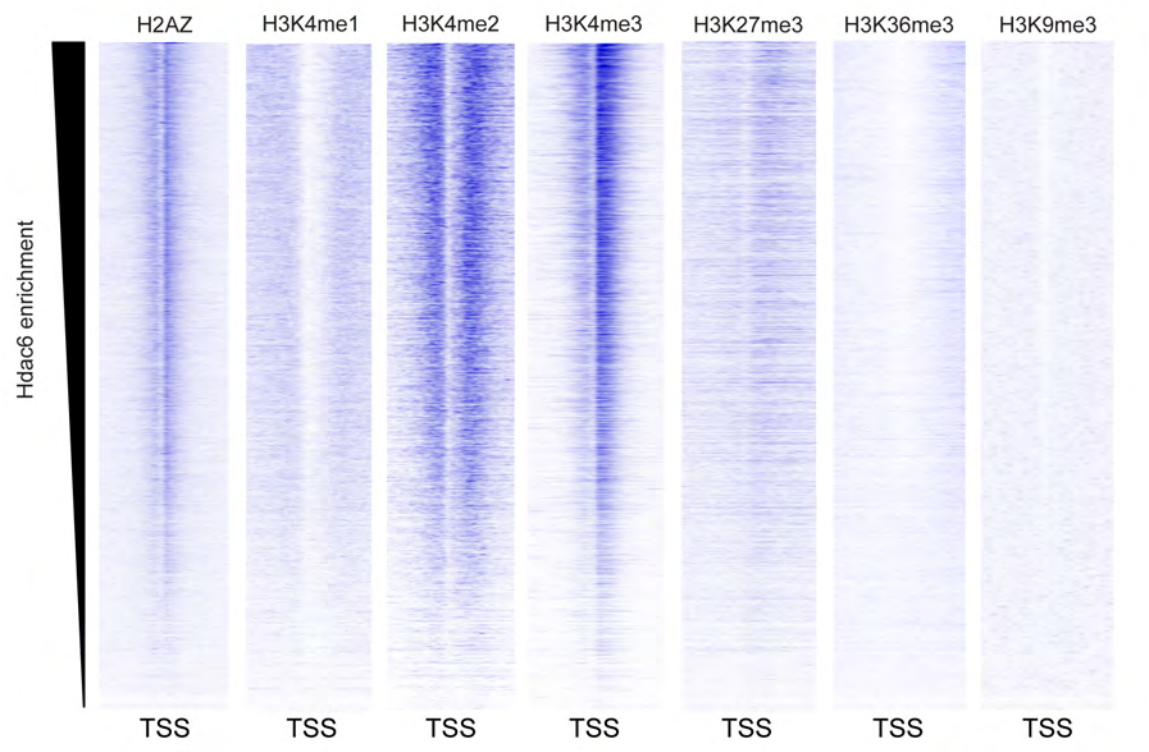


Figure II.S3: Hdac6 is enriched at genes marked by H3K4me3

Heatmaps showing the enrichment of various histone modifications or variants within promoter-proximal regions (TSS +/- 2kb) of all genes, sorted by Hdac6 enrichment. H3K4me3 is strongly enriched at genes with high levels of Hdac6 binding, including bivalent genes marked by both H3K4me3 and H3K27me3, similar to previous findings for the p400 subunit of Tip60-p400.

In addition, the overlap between Hdac6 and one of the two peaks of Tip60-p400 binding is consistent with a model in which Hdac6 functions within chromatin-bound Tip60-p400 complex to regulate a set of common target genes.

The catalytic domains of Hdac6 are necessary for interaction with Tip60-p400 complex, but not for deacetylation of histones at target promoters

We next tested several possible models by which Hdac6 might function within Tip60-p400 complex to regulate gene expression. First, we considered the possibility that Hdac6 was necessary for complex formation or stability. To test this model, we immunoprecipitated Tip60-H3F in control and *Hdac6* KD ESCs, but found that Tip60-p400 subunits p400 and Dmap1 were maintained in the absence of Hdac6 (Figure II.S4A), arguing against this explanation. Alternatively, Hdac6 might be necessary for localization of Tip60-p400 to its target genes in ESCs. Finally, Tip60-p400 may recruit Hdac6 to its target genes, where it regulates gene expression by deacetylating histones. To distinguish between these latter two possibilities, we performed a series of experiments. First, we tested whether treatment of ESCs with Tubastatin A, a chemical inhibitor that prevents Hdac6-dependent deacetlyation of acetylated lysines by binding within the catalytic channel of one or both deacetylase domains (Butler et al., 2010), had similar effects on gene expression as *Hdac6* KD.

Figure II.S4: *Hdac6* KD reduces histone acetylation at Tip60 target genes

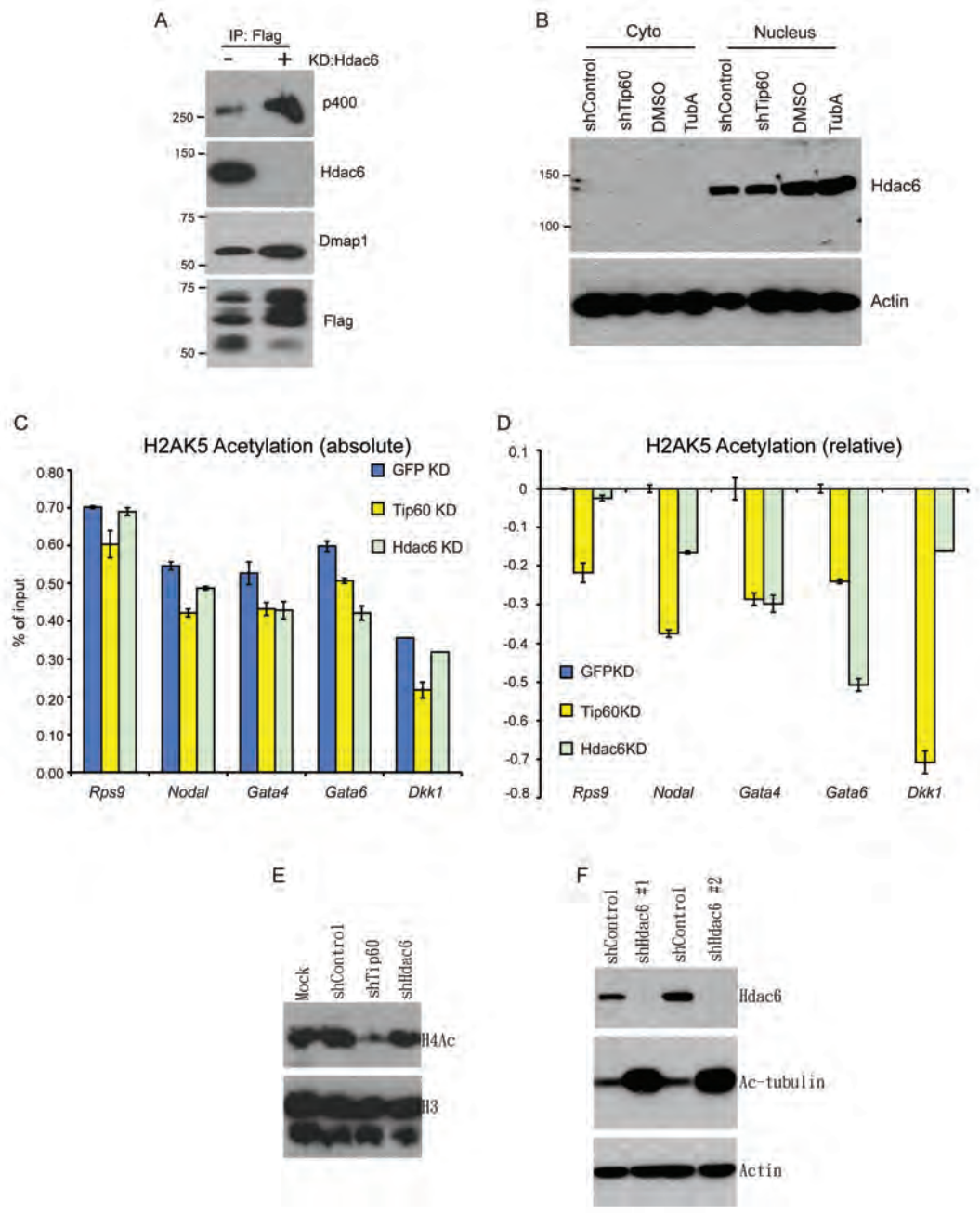
(A) *Hdac6* KD does not disrupt formation of Tip60 complex. Western blot of purified Tip60 complex from *Hdac6* KD ESCs for indicated proteins.

(B) Nuclear localization of Hdac6 in ESCs is unaffected by Tip60 depletion and HDAC inhibition. ESCs were treated as shown and nuclear and cytoplasmic fractions were Western blotted for the proteins indicated.

(C-D) H2AK5 acetylation levels for several common Tip60/Hdac6 target genes in *Tip60* and *Hdac6* KD ESCs was measured by ChIP-qPCR. H2AK5Ac levels in *Tip60* or *Hdac6* KD cells are expressed as a fraction of the input (C) or as Log2 ratios relative to control (GFP) KD (D).

(E) Hdac6 depletion has minimal effect on histone acetylation. ESCs knocked down as indicated were Western blotted for acetylated histone H4 or total H3.

(F) Hdac6 KD results in accumulation of high levels of acetylated tubulin. Indicated KD ESCs were Western blotted with the indicated antibodies.


A
IP: Flag
- +
KD:Hdac6
250
p400
150
Hdac6
75
Dmap1
50
75
Flag
50
B
Cyto
Nucleus
shControl
shTip60
DMSO
TubA
shControl
shTip60
DMSO
TubA
150
100
Hdac6
Actin
C
H2AK5 Acetylation (absolute)
GFP KD
Tip60 KD
Hdac6 KD
% of input
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Rps9
Nodal
Gata4
Gata6
Dkk1
D
H2AK5 Acetylation (relative)
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
GFPKD
Tip60KD
Hdac6KD
Rps9
Nodal
Gata4
Gata6
Dkk1
E
Mock
shControl
shTip60
shHdac6
H4Ac
H3
F
shControl
shHdac6 #1
shControl
shHdac6 #2
Hdac6
Ac-tubulin
Actin

Like *Hdac6* KD or *Tip60* KD ESCs, most target genes were de-repressed upon Tubastatin A treatment, although generally to lower levels than observed upon KD of *Hdac6* or *Tip60* (Figure II.5A), consistent with the possibility that histone deacetylation by Hdac6 was necessary for repression of its gene targets. Tubastatin A treatment did not affect expression or nuclear localization of Hdac6 (Figure II.S4B), suggesting that it acts directly on the Hdac6 deacetylase domains.

Next we tested whether *Hdac6* KD resulted in increased acetylation of histone tails at Hdac6 target promoters by examining acetylation of the N-terminal tails of histones H4 and H2A, two major targets of the Tip60 acetyltransferase activity (Altaf et al., 2010; Doyon et al., 2004; Kimura and Horikoshi, 1998; Yamamoto and Horikoshi, 1997; Yan et al., 2000). ChIP-qPCRs using antibodies recognizing histone H4 acetylated at all four N-terminal lysines or acetylated lysine-5 of histone H2A were performed in control, *Tip60* KD, and *Hdac6* KD ESCs. Surprisingly, like *Tip60* KD, *Hdac6* KD resulted in a decrease in both H4 and H2A acetylation at most shared targets of Tip60 and Hdac6 that were examined (Figures II.5B, II.S4C-D). These findings show that Hdac6 does not silence differentiation genes by counteracting Tip60-dependent histone acetylation, despite the fact that treatment with the Hdac6 inhibitor Tubastatin A resulted in gene expression changes similar to Hdac6 KD. Unlike *Tip60* KD, *Hdac6* KD had no significant effect on bulk histone H4 acetylation (Figure II.S4E), although tubulin acetylation was strongly enhanced by *Hdac6* KD (Figure

II.S4F), consistent with a model in which Hdac6 is required for normal Tip60 function only at common target promoters.

A few subunits of Tip60-p400 complex are known to be acetylated (Choudhary et al., 2009), raising the possibility that Hdac6 regulates Tip60-p400 function by deacetylating the complex, which then leads to altered biochemical activity or chromatin binding by the complex. We tested this possibility by performing mass spectrometry on p400, the most highly acetylated subunit in Tip60-p400 complex (Choudhary et al., 2009), purified from control or *Hdac6* KD ESCs. However, we did not observe any notable differences in p400 acetylation levels (data not shown). Thus, neither *Hdac6* KD nor inhibition of Hdac6's deacetylase activity leads to any observable increase in p400 acetylation or acetylation of histone tails, but both perturbations lead to de-repression of Tip60-p400 target genes. Next we considered the possibility that Hdac6 binds Tip60-p400 through one or both of its deacetylase domains. In this scenario, despite the fact that Hdac6 does not appear to deacetylate Tip60-p400 subunits, Tubastatin A could cause de-repression of Tip60- and p400-target genes simply by preventing Hdac6 from binding Tip60-p400. We tested this possibility by immunoprecipitating Tip60-p400 complex from *Tip60-H3F* ESCs treated with either vehicle alone or Tubastatin A and determining whether Hdac6 co-precipitates with Tip60. Interestingly, Tubastatin A treatment completely prevented Hdac6 association with Tip60-p400 (Figure II.5C).

We reasoned that if Tubastatin A directly inhibited binding of Hdac6 to Tip60-p400, then addition of the drug to purified Hdac6-containing Tip60-p400 complex should disrupt this interaction. In contrast, if Hdac6 deacetylates some unknown protein, which then activates it to bridge the interaction of Hdac6 with Tip60-p400, treatment of cells with Tubastatin A should disrupt the Hdac6 interaction with Tip60-p400, while in vitro Tubastatin A treatment of Tip60-p400 complex should not. To distinguish between these two possibilities, we immunoprecipitated Tip60-p400 from untreated *Tip60-H3F* ESCs and, after washing unbound proteins away from beads, subjected them to five additional washes with or without Tubastatin A. After harvesting the protein eluted in the Tubastatin A washes or remaining bound to beads, we examined the distribution of Hdac6 by Western blotting. Interestingly, we found that the Tubastatin A washes removed a large fraction of Hdac6 from bead-bound Tip60-p400 complex (Figure II.5D), suggesting that Tubastatin A disrupts the interaction of Hdac6 with Tip60-p400, rather than preventing deacetylation of histones or other proteins. Consistent with these findings, we found that mutation of the Hdac6 deacetylase domains had the same effect as Tubastatin A treatment: mutation of deacetylase domain 1 partially disrupted Hdac6's interaction with Tip60-p400 in ESCs, while mutation of both deacetylase domains abolished this interaction (Figure II.5E). Together, these data indicate that Hdac6 interacts with Tip60-p400 via its deacetylase domains and that Tubastatin A directly disrupts this interaction.

Figure II.5: The Hdac6 deacetylase domains are necessary for Tip60-p400 binding but do not reverse histone acetylation catalyzed by Tip60

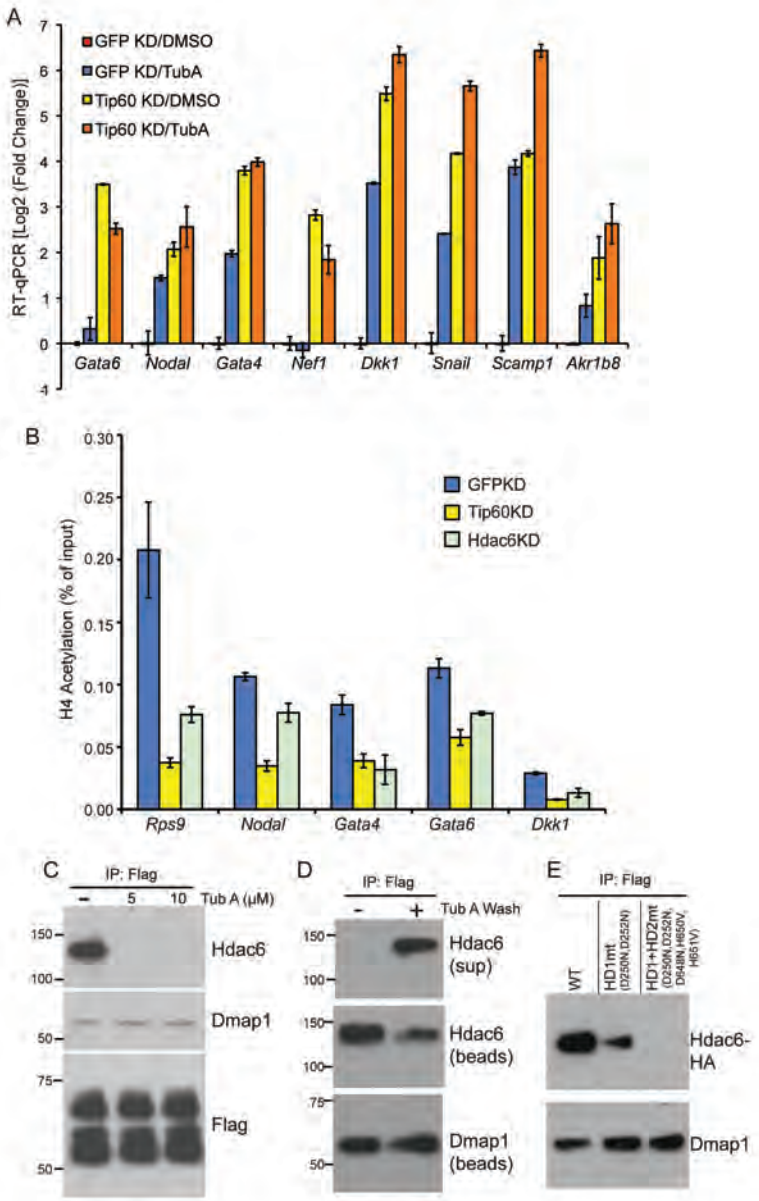
(A) Treatment of ESCs with Hdac6 catalytic domain inhibitor Tubastatin A causes de-repression of Tip60-p400 target genes. RT-qPCRs for indicated genes are shown for *GFP* KD or *Tip60* KD ESCs treated with either Tubastatin A (Tub A) or DMSO vehicle. Data are expressed as Log2 (fold change) values relative to DMSO treated *GFP* KD ESCs after normalization. Shown are the mean ± SD values of three technical replicates from one representative experiment of two biological replicates performed.

(B) H4 acetylation levels for several common Tip60/Hdac6 target genes in *Tip60* KD and *Hdac6* KD ESCs were measured by ChIP-qPCR, using an antibody specific for tetra-acetylated histone H4. H4 acetylation levels in cells knocked down as indicated are expressed as a fraction of the input. Shown are the mean ± SD values of three technical replicates from one representative experiment of two biological replicates performed.

(C) ESCs were treated overnight with the indicated amounts of Hdac6 inhibitor Tubastatin A in their growth medium, Tip60-H3F was immunoprecipitated as above, and co-immunoprecipitating proteins were examined by Western blotting.

(D) Tip60-H3F was immunoprecipitated from ESCs grown under normal conditions and the beads were washed in buffer with or without Tubastatin A. The Hdac6 eluted in the Tubastatin A wash or remaining bound to beads is shown by Western blotting, along with canonical Tip60 subunit Dmap1.

(E) HA-tagged wild type, deacetylase domain 1 (HD1) mutant, or double deacetylase domain mutant (HD1 + HD2) *Hdac6* were stably expressed in *Tip60-H3F* ESCs, Tip60-H3F was immunoprecipitated from nuclear extracts, and co-precipitating proteins were examined by Western blotting. Co-IP of canonical Tip60 complex subunit Dmap1 is shown as a control.


A
GFP KD/DMSO
GFP KD/TubA
Tip60 KD/DMSO
Tip60 KD/TubA
RT-qPCR [Log2 (Fold Change)]
Gata6
Nodal
Gata4
Nef1
Dkk1
Snail
Scamp1
Akr1b8
B
GFPKD
Tip60KD
Hdac6KD
H4 Acetylation (% of input)
Rps9
Nodal
Gata4
Gata6
Dkk1
C
IP: Flag
Tub A (μM)
Hdac6
Dmap1
Flag
D
IP: Flag
Tub A Wash
Hdac6 (sup)
Hdac6 (beads)
Dmap1 (beads)
E
IP: Flag
WT
HD1mt (D250N,D252N)
HD1+HD2mt (D250N,D252N, D944N,H650V, H651V)
Hdac6-HA
Dmap1

Hdac6 is necessary for normal Tip60 and p400 recruitment

The finding that Hdac6 does not appear to deacetylate histones or Tip60-p400 subunits, but that its deacetylase domains are necessary for interaction with Tip60-p400 complex, was consistent with a model in which Hdac6 regulates gene expression by helping recruit Tip60-p400 complex to its targets in vivo. We tested this possibility by comparing Tip60 enrichment in control KD and *Hdac6* KD ESCs, predicting that if Hdac6 is necessary for Tip60-p400 recruitment, *Hdac6* KD should result in diminished Tip60 enrichment at many of its target genes. We observed a striking difference in Tip60 binding upon *Hdac6* KD compared to control cells: Tip60 enrichment downstream of the TSS was strongly reduced at many target genes in *Hdac6* KD ESCs, along with a modest decrease in enrichment upstream of the TSS (Figure II.6A-B). Importantly, the reduction of Tip60 binding upon *Hdac6* KD was most severe at genes whose expression is regulated by Tip60-p400, compared to genes not regulated by the complex (Figures II.6C, II.S5A-B). Similarly, *Hdac6* KD strongly reduced binding of the p400 subunit (Figures II.6D-E, II.S5C); these results were also recapitulated in $Hdac6^{reduced}$ and $Hdac6^{null}$ cells (Figure II.S5D). Although we found that Hdac6 was required for Tip60 binding downstream of the TSS at most of its target genes, it remained possible that Tip60 was also required for Hdac6 to bind the same genes. To test this possibility, we mapped Hdac6 binding in control and *Tip60* KD ESCs. In contrast to Tip60's requirement for Hdac6, *Tip60* KD had no effect on Hdac6 localization (Figures II.6F, II.S5E).

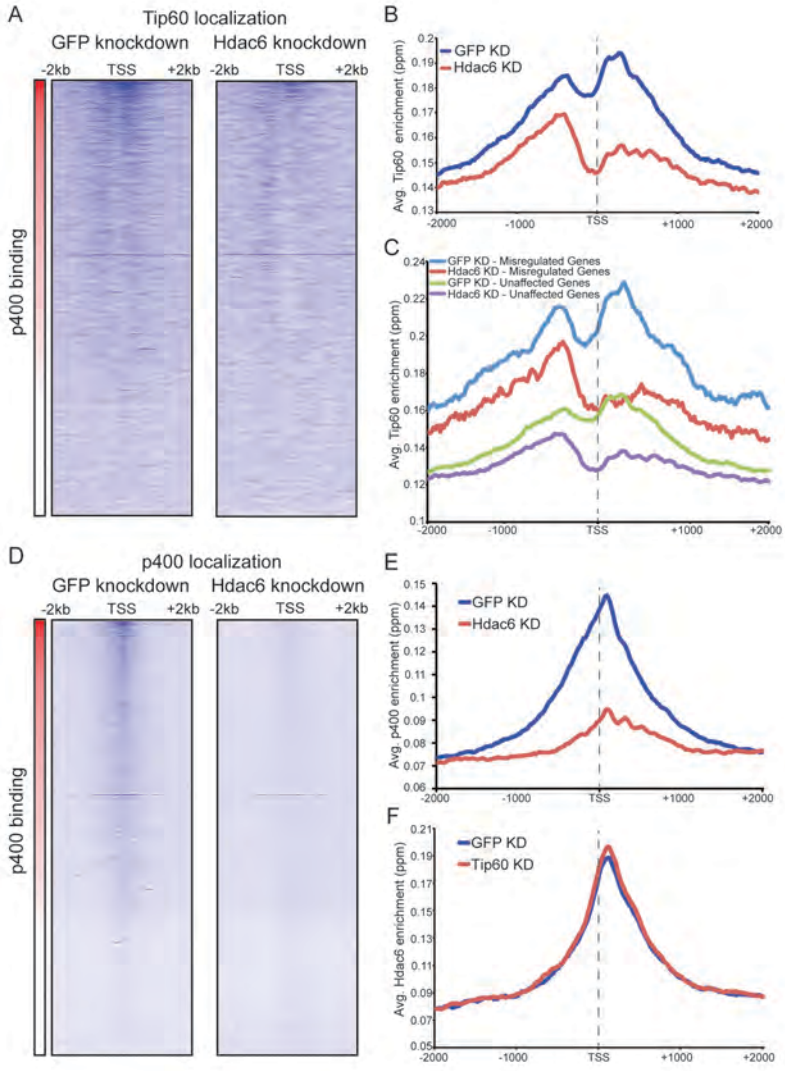


Figure II.6: Hdac6 is necessary for normal Tip60 binding to its targets on chromatin

(A) Heat map representations of Tip60 binding as measured by ChIP-seq in *Tip60-H3F* ESCs upon *GFP* KD or *Hdac6* KD. Data are sorted by p400 binding as in Figure 3A.

(B) Averaged Tip60 binding upon *GFP* KD or *Hdac6* KD are shown relative to the TSS.

(C) *Hdac6* KD mainly reduces Tip60 enrichment at genes that are misregulated upon *Tip60* KD. Average Tip60 binding profiles upon *GFP* KD or *Hdac6* KD are shown for genes misregulated upon *Tip60* KD (adjusted $p < 0.1$) and genes that are unaffected.

(D) Heat map representations of p400 binding as measured by ChIP-seq using an antibody against endogenous p400 in control and Hdac6 KD ESCs. Genes are sorted exactly as in (A).

(E) Averaged p400 binding upon *GFP* KD or *Hdac6* KD are shown relative to the TSS.

(F) *Tip60* KD does not affect Hdac6 binding. Average Hdac6 binding upon *GFP* KD or *Hdac6* KD are plotted as in (B and E).

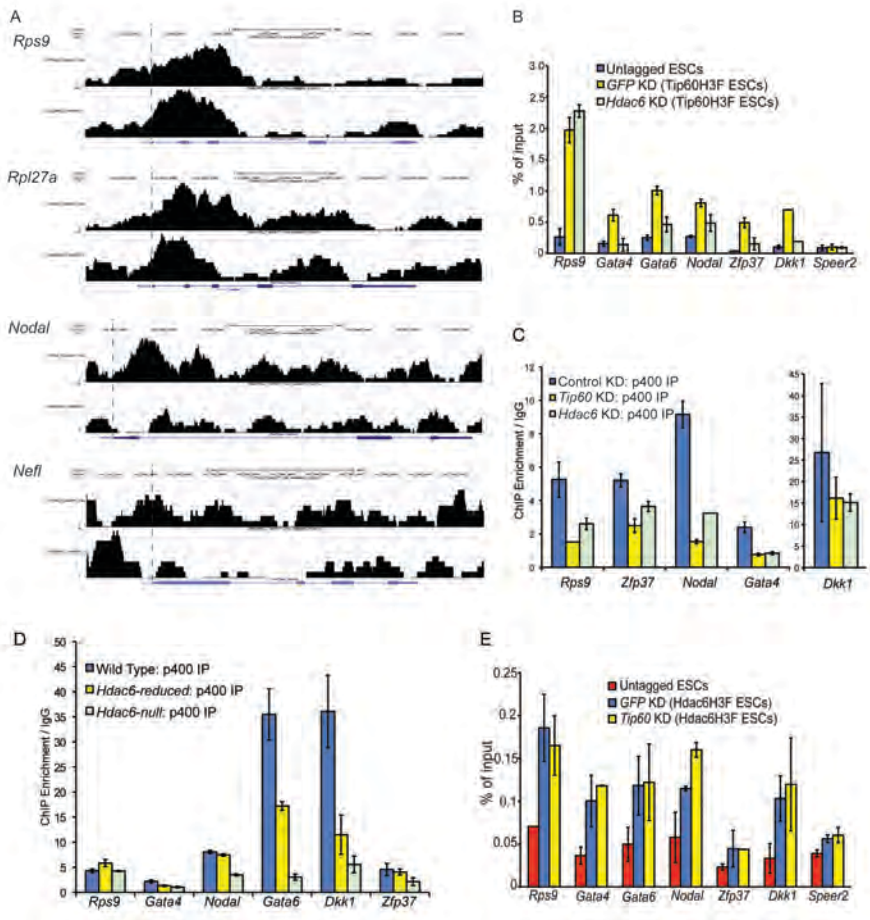


Figure II.S5: Reduced levels of Tip60 binding to differentiation genes upon *Hdac6* KD

(A) Example genomic loci showing Tip60 binding as measured by ChIP-seq upon GFP KD or *Hdac6* KD. The first two panels show that Tip60 binding to the promoter regions of highly expressed genes (*Rps9*, *Rpl27a*) is relatively unaffected by *Hdac6* KD. The second two panels show that Tip60 binding to the promoter regions of differentiation genes (*Nodal*, *Nefl*) is strongly reduced downstream of the TSS upon *Hdac6* KD.

(B) Effects of *Hdac6* KD on Tip60 binding to target genes. Anti-FLAG ChIP-qPCR data are shown for the indicated ESCs at the promoter-proximal regions of the target genes shown. Note that *Speer2* is a negative control locus to which Tip60-p400 does not bind.

(C) Effects of *Hdac6* KD on p400 binding to target genes. Anti-p400 ChIP-qPCR data are shown for the indicated KDs at the promoter-proximal regions of target genes. Data are normalized to IgG background ChIPs. Note that *Dkk1* is shown on a different scale.

(D) Effects of *Hdac6* mutations on p400 binding to target promoters. Anti-p400 ChIP-qPCR data are shown for the indicated ESCs at the promoter-proximal regions of target genes.

(E) Effects of *Tip60* KD on Hdac6 binding to target genes. Anti-FLAG ChIP-qPCR data are shown for the indicated ESCs at the promoter-proximal regions of target genes.

Together, these data support a model in which Hdac6 binding functions (through its deacetylase domains) to recruit or maintain binding of Tip60-p400 complex. In addition, these data rule out the possibility that Tip60-p400 represses differentiation genes by recruiting Hdac6 to these sites, since common targets of Tip60-p400 and Hdac6 were misregulated upon *Tip60* KD even though Hdac6 localization was unaffected.

Hdac6 is required for normal ESC proliferation, colony formation, and differentiation but not self-renewal

Previously, we showed that ESCs depleted of Tip60-p400 subunits exhibited two prominent phenotypes: a failure to self-renew under conditions favoring ESC growth and a defect in the formation of embryoid bodies (EBs) under growth conditions favoring differentiation (Fazzio et al., 2008a). In contrast, while deletion of *Hdac6* in ESCs was reported to cause a defect in colony formation, as well as a slight proliferation defect, *Hdac6* KO ESCs could nonetheless continue to self-renew (Zhang et al., 2003), and mice derived from *Hdac6* KO ESCs are viable. However, it was unclear from previous reports whether *Hdac6* KO or KD might result in impaired ESC differentiation in vitro, similar to that of *Tip60* KD. To test this possibility, we first confirmed that *Hdac6* KD ESCs exhibited colony formation and proliferation defects similar to those described for *Hdac6* KO ESCs. Indeed, we observed a significant decrease in colony formation and size in both *Hdac6* KD and *Tip60* KD ESCs (Figure II.7A).

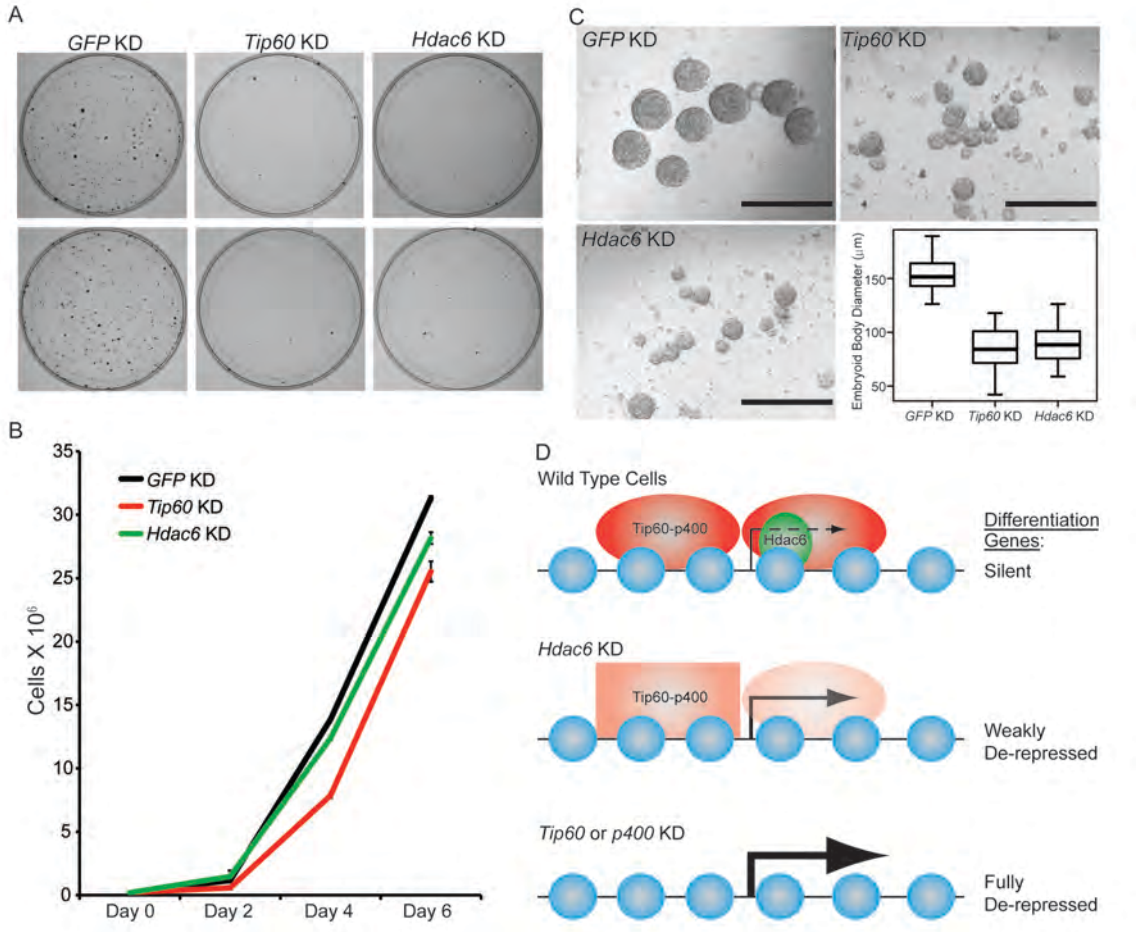


Figure II.7: Hdac6 is necessary for ESC colony formation and normal differentiation

(A) Colony formation assays. Indicated KD ESCs were plated at clonal density and grown for seven days, at which time they were stained with crystal violent for visualization.

(B) ESCs were infected with shRNA expressing viruses as shown and cultured in normal ESC medium. Cells were counted at the times indicated.

(C) Left and upper right: Brightfield images of EBs formed by hanging drop cultures of ESCs knocked down as indicated, then cultured in differentiation medium, as described in the Experimental Procedures. Scale bars equal 400 μm. Lower right: box plot quantification of the range of EB sizes by diameter. The upper and lower limits of the box correspond to the 75th and 25th percentiles of each KD, respectively, and the dark line corresponds to the median of each box. At least 88 EBs were measured for each KD.

(D) Model for Hdac6- and Tip60-p400-dependent repression of differentiation genes in ESCs. In the presence of Hdac6, Tip60-p400 binds both upstream and downstream of TSSs and differentiation genes are silenced. In the absence of Hdac6, Tip60-p400 binding is reduced causing de-repression of differentiation genes. Tip60-p400 binding downstream of target TSSs may appears to be more strongly affected by *Hdac6* KD. Note that Hdac6 also appears to affect the activation of some genes (not depicted), also by recruitment of Tip60-p400 complex.

Furthermore, *Hdac6* KD ESCs had a small but reproducible decrease in proliferation rate, while *Tip60* KD caused a somewhat more severe proliferation defect (Figure II.7B), consistent with previous studies (Fazzio et al., 2008a; Zhang et al., 2003). Next we tested whether EB formation or cellular differentiation might be impaired upon *Hdac6* KD, as we previously observed upon KD of genes encoding known Tip60-p400 subunits *Tip60*, *p400*, or *Dmap1* (Fazzio et al., 2008a). EBs are thought to roughly mimic the embryonic state, as cells proliferate within spherical aggregates that subsequently develop cystic structures and differentiate into cells from all three germ layers (Martin and Evans, 1975). Indeed we found that, similar to *Tip60* KD ESCs, *Hdac6* KD ESCs suspended in differentiation medium formed EBs that were significantly smaller and more heterogeneous than control KD ESCs (Figure II.7C). Furthermore, both *Hdac6* and *Tip60* KD EBs exhibited delayed induction of several differentiation markers during a time course of ESC differentiation (Figure II.S6A). Thus, while *Hdac6* loss results in only modest phenotypes in self-renewing ESCs, it is necessary for normal EB formation during cellular differentiation in vitro, consistent with its role in Tip60-p400 recruitment to a subset of Tip60-p400 target genes.

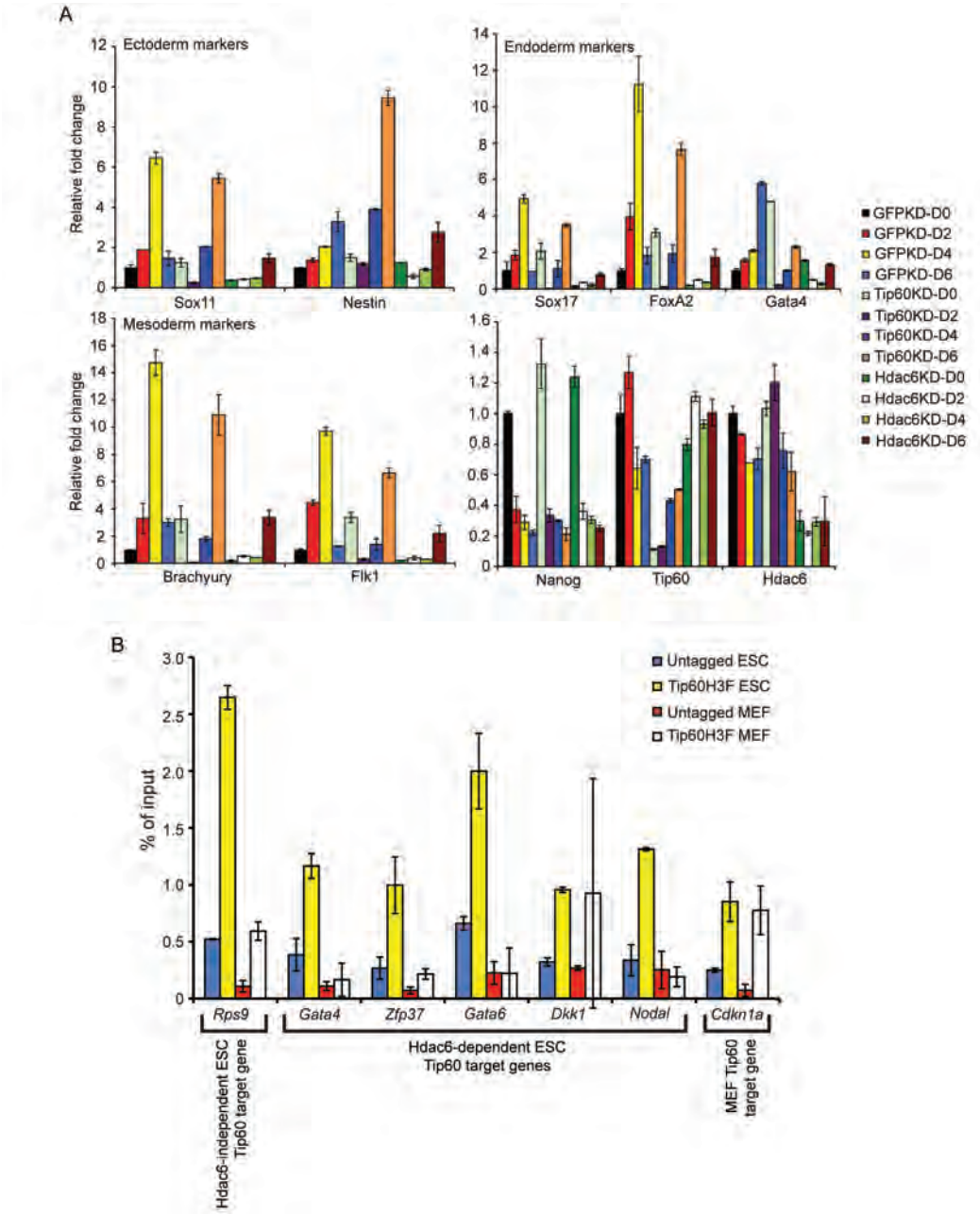


Figure II.S6: *Hdac6* KD ESCs have impaired induction of differentiation markers

(A) Indicated KDs were differentiated for 0, 2, 4, or 6 days.. At the indicated time points, RNA was isolated and RT-qPCR quantification of several differentiation markers of each primary germ layer was performed.

(B) Tip60 enrichment near the promoters of the genes indicated was determined by ChIP-qPCR in ESCs and MEFs. We examined binding to *Cdkn1a,* a known Tip60 target in MEFs, as a positive control for ChIP in MEFs. Data shown are the mean ± SD of three technical replicates from one representative experiment of two biological replicates performed.

DISCUSSION

Here we showed that Hdac6 interacts with Tip60-p400 complex in ESCs and NSCs and is necessary for the proper regulation of most genes regulated by Tip60-p400. For a majority of Tip60-p400 target genes, we found that Hdac6 facilitated Tip60-p400 binding to its gene targets. Interestingly, while Hdac6 does not appear to deacetylate Tip60-p400 subunits or histones in vivo, we found that its catalytic domains were necessary for interaction with Tip60-p400 complex. Hdac6 has well established deacetylase activity directed against several cytoplasmic proteins, most notably tubulin (Hubbert et al., 2002), leaving open the possibility that, in stem cells, it may deacetylate some nuclear protein(s) that remain to be discovered.

Tip60-p400 binds chromatin in vivo near the 5' ends of genes, in two peaks surrounding the TSS: an upstream peak that does not overlap with Hdac6 and is moderately sensitive to its loss, and a downstream peak that overlaps with and more strongly requires Hdac6 (Figure II.7D). We found that p400 binding also required Hdac6 for normal levels of enrichment, consistent with these data. Recruitment of Tip60-p400 is the only apparent function of Hdac6 in gene regulation, since Hdac6 binding is maintained upon *Tip60* KD, while gene silencing is not. These data contrast with previously observed genetic interactions of mammalian Tip60 with other HDACs, in which Tip60 recruitment counteracts the repressive effects of HDACs on common targets (Baek et al.,

2002), although similarities in gene expression profiles between Tip60 loss of function and chemical inhibition of HDACs have been reported in *Drosophila* (Schirling et al., 2010).

We found that *Hdac6* KD or mutation recapitulates most phenotypes of ESCs lacking Tip60 or p400: de-repression of many differentiation genes, impaired colony formation, a slower proliferation rate, impaired EB formation, and delayed kinetics of differentiation. However, unlike *Tip60* (Hu et al., 2009), *Hdac6* KO ESCs can self-renew and are competent for mouse development (Zhang et al., 2008), strongly suggesting that the defect observed for *Hdac6* KD ESCs in vitro is overcome in the embryo by compensatory mechanisms. Since the inner cell mass of blastocyst-stage embryos is attached to a layer of trophectoderm cells, it is possible that the colony-formation and differentiation defects observed for *Hdac6* KD cells in vitro do not pose a barrier to development. Alternatively, since Hdac6 is only partially required for Tip60-p400 function in ESCs, the levels to which Tip60 and Hdac6 target genes are misregulated within the ICM of $Hdac6^{-/-}$ embryos may not be severe enough to induce a developmental arrest. In addition, since Hdac6 appears to regulate Tip60-p400 complex in only a subset of cell types (stem and progenitor cells) the effect of *Hdac6* loss on gene regulation is likely much more limited than that of *Tip60* loss throughout the embryo.

The finding that Hdac6 is partially to completely nuclear in two types of stem cells and a mixed population of stem and progenitor cells (HPCs), and the fact that it interacts with Tip60-p400 in both ESCs and NSCs, but not MEFs, suggests that nuclear exclusion of Hdac6 during differentiation may play a major role in gene regulation by reducing Tip60-p400 binding to specific sets of genes. Indeed, upon differentiation of ESCs or NSCs, we observed a strong decrease in nuclear Hdac6 and a coincident increase in Hdac6 within the cytoplasm (Figure II.2A). Also consistent with this model, we found that Hdac6-dependent Tip60 target genes in ESCs were not bound by Tip60 in MEFs (Figure II.S6B). It remains to be determined whether alternative proteins substitute for Hdac6 in differentiated cells to recruit Tip60-p400 to different promoters, or if alternative mechanisms of recruitment (e.g., binding to histone modifications or interaction with sequence-specific DNA-binding proteins) are more important upon differentiation.

Together, these studies identify a new, stem cell-specific mechanism by which Tip60-p400 is regulated, and uncover a role for Hdac6 in gene regulation. These findings differ substantially from established models of Hdac6 function as mainly a regulator of cellular motility and clearance of misfolded proteins via deacetylation of a small set of cytoplasmic targets, and suggest that cell type must be carefully considered when examining the phenotypes observed upon *Hdac6* loss of function. In addition, these studies lend support to the idea that different types of stem cells share some common features of chromatin structure

that allow them to maintain their developmental potency, and provide one possible mechanistic link underlying this commonality. Re-localization of Hdac6 to the cytoplasm during ESC and NSC differentiation could help establish a distinct set of Tip60-p400 targets in differentiated cells compared to stem cells, and lead to an altered pattern of Tip60-p400 binding to these sites. Future work focusing on developmental regulation of Hdac6 re-localization and alterations in Tip60 binding will shed further light on these possibilities.

METHODS

Cells

ESCs were grown under feeder-free conditions and used for all ESC experiments. The *Tip60-H3F* and *Hdac6-H3F* lines were made by targeting into E14 (Hooper et al., 1987) and *Hdac6tm1a(EUCOMM)Wtsi* ESCs were obtained from EUCOMM (clone EPD0519_4_C03). This clone expresses low levels of Hdac6 protein before introduction of CRE, while CRE addition converts the mutation to a deletion. Therefore, we refer to the Hdac6 alleles as *Hdac6reduced* and *Hdac6null* before and after CRE introduction, respectively. CRE was introduced by Lenti-LucS (Addgene plasmid 22778) lentiviral infection, and the cells were harvested 3 days later for RT-qPCR or ChIP. Tip60-EGFP and Hdac6-HD mutant-HA expression: ES cells were infected with pLJM1puro lentiviral vectors containing mouse Tip60 cDNA fused to EGFP. The infected cells were

selected by puromycin 3 days post-infection and Tip60 expression was checked by Western-blot. Hdac6 was cloned into pBabe HAII-hygro retroviral vector with or without previously described HD1 and HD2 mutations (Zhang et al., 2006). MEFs were infected with pBabe retroviral vectors containing wild-type or mutant mouse Hdac6 cDNAs. The infected cells were selected by hygromycin 3 days post-infection and Hdac6 expression was checked by Western-blot.

For imaging, embryoid bodies (EBs) were generated by placing 300 cells, after infection with shRNA-expressing lentiviruses as indicated, in 30 µl of medium lacking LIF and incubating in hanging-drop cultures. For examination of differentiation markers, 106 cells were suspended in non-cell culture treated petri dishes for 2 days, and transferred to gelatin-coated cell culture dishes for another 4 days. RNA was isolated at the indicated time points. Mouse embryonic fibroblasts (MEFs) were isolated from *Tip60-H3F*/+ or wild-type littermates using standard protocols (Coles et al., 2007) and mouse NSCs were isolated as previously described (Li et al., 2008).

Colony formation assay: control, *Tip60*, or *Hdac6* KD ESCs were plated 3 days after lentiviral shRNA infection. 2000 cells were seeded into each 10 cm dish and cultured for 7 days. The cells were fixed and stained with fixation-staining solution (6% glutaraldehyde, 0.5% crystal violet) for 30 min at room temperature followed by three washes with water.

For examination of Hdac6 redistribution during stem cell differentiation, ESCs were plated in N2B27 medium (Ying et al., 2003) and NSCs were plated in DMEM with 10% fetal bovine serum (FBS) for the number of days indicated. The cancer cell lines were grown in DMEM with 10% fetal bovine serum.

RNAi

Lentiviral shRNA expression vectors from the TRC library (Thermo Fisher, Waltham, MA, USA) were obtained from the UMMS RNAi Core Facility. After screening through multiple shRNAs for each gene to be knocked down, we identified the most effective hairpin for each gene for subsequent experiments, listed as follows:

pLKO.1/shGFP: GCAAGCTGACCCTGAAGTTCAT

pLKO.1/shTip60: CGGAGTATGACTGCAAAGGTT

pLKO.1/shHdac6: CGCTGACTACATTGCTGCTTT

Lentiviral vectors and Lenti-X packaging plasmids were transfected into 293.T cells using Fugene6 (Roche, Branford, CT, USA). At 48, 60 and 72 hr after transfection, lentivirus-containing media were collected and concentrated over a 20% sucrose cushion by centrifugation at 14,000 rpm for 4 hr in an SW-28 rotor.

Concentrated virus was re-suspended in 200 µl PBS, aliquoted, and stored at −80°C.

Western blotting and immunoprecipitation

Cells were lysed using an NE-PER Extraction kit (Thermo Fisher) to isolate cytoplasmic and nuclear fractions. Western blotting was performed with antibodies against Hdac6 (Cat. 07-732; Millipore, Billerica, MA, USA), acetyl-Histone H4 (Cat. 06-866; Millipore), β-actin (Cat. A5316; Sigma, St. Louis, MO, USA), Flag-M2 (Cat. F1804; Sigma), p400 (Cat. A300-541A; Bethyl Labs, Montgomery, TX, USA), Dmap1 (Cat. 10411-1-AP; Proteintech Group, Chicago, IL, USA), GFP (Cat. ab290; Abcam, Cambridge, MA, USA), Pol II (Cat. sc-899; Santa Cruz Biotechnology, Santa Cruz, CA, USA), Acetylated Lysine (Cat. 9441S; Cell Signaling Technologies, Danvers, MA, USA), ubiquitin (P4D1), and HA (12CA5). For immunoprecipitation, the aliquots of nuclear extract were incubated with specific antibodies conjugated with protein G magnetic beads (New England Biolabs, Ipswich, MA, USA) or FLAG-M2 Agarose beads (Sigma) in IP buffer (50 mM Tris-HCl pH7.4, 250 mM NaCl, 0.1% Triton X-100, plus 1X HALT protease inhibitors (Thermo Fisher) overnight at 4°C. To examine the effect of inhibitors of Hdac6 on its interaction with Tip60 complex, beads were subjected to five additional washes with or without 10 µM tubastatin A (ChemieTek, Indianapolis, IN, USA).

Tip60-p400 purification

Tip60 complex was purified from nuclear extracts of Tip60-H3F ESCs as described previously for Mbd3 (Yildirim et al., 2011). Briefly, nuclear extracts were subjected to sequential affinity purification steps using FLAG-M2 Agarose (Sigma) and TALON Agarose beads (Clontech Laboratories, Mountain View, CA, USA). The proteins purified from untagged control and Tip60-H3F cells were separated by SDS-PAGE and were either stained with SimplyBlue SafeStain (Invitrogen, Grand Island, NY, USA) after TCA precipitation and re-suspension in sample buffer or SilverXpress (Invitrogen) for visualization in *Figure 1*.

LC-MS/MS

Affinity-purified samples were separated by SDS-PAGE gel, in-gel digested and analyzed by LC-MS and LC-MS/MS as described previously (Chu et al., 2006). Briefly, 1 ml aliquot of the digestion mixture was injected into a Dionex Ultimate 3000 RSLCnano UHPLC system with an autosampler (Dionex Corporation, Sunnyvale, CA, USA), and the eluant was connected directly to a nanoelectrospray ionization source of an LTQ Orbitrap XL mass spectrometer (Thermo Fisher). LC-MS data were acquired in an information-dependent acquisition mode, cycling between a MS scan (m/z 310-2000) acquired in the Orbitrap, followed by low-energy CID analysis in the linear ion trap. The centroided peak lists of the CID spectra were generated by PAVA (Guan and

Burlingame, 2010) and searched against a database that consisted of the Swiss-Prot protein database, to which a randomized version had been concatenated, using Batch-Tag, a program in the in-house version of the University of California San Francisco Protein Prospector version 5.9.2. A precursor mass tolerance of 15 ppm and a fragment mass tolerance of 0.5 Da were used for protein database search. Protein hits are reported with a Protein Prospector protein score ≥22, protein discriminant score ≥0.0 and a peptide expectation value ≤0.01 (Chalkley et al., 2005). This set of protein identification parameters threshold did not return any substantial false positive protein hit from the randomized half of the concatenated database.

Chromatin immunoprecipitation

The cells from 80% confluent 10 cm dishes were crosslinked by adding fixation solution (1% formaldehyde, 0.1M NaCl, 1 mM EDTA, 50 mM HEPES·KOH pH 7.6) for 10 min at room temperature. Crosslinking was quenched with 125 mM Glycine for 5 min. The cells were washed twice with cold PBS containing protease inhibitors (Roche), and pelleted at 1000×*g* for 5 min at 4°C. The cell pellets were either flash frozen in liquid nitrogen and stored at −80°C or immediately sonicated. The pellets were resuspended in Lysis buffer 1 (50 mM HEPES−KOH pH 7.6, 140 mM NaCl, 1 mM EDTA, 10% (vol/vol) Glycerol, 0.5% NP-40, 0.25% Triton X-100) including protease inhibitors and incubated for 10 min at 4°C. After centrifugation at 1350×*g* for 5 min, the pellets

were resuspended in Lysis buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) containing protease inhibitors and incubated for another 10 min at 4°C. The pellets were collected after centrifugation at 1350×*g* for 5 min and resuspended in Lysis buffer 2 for sonication. The samples were sonicated in a Bioruptor (UCD-200; Diagenode, Delville, NJ, USA) set to high for three cycles (10 min per cycle with 30 s on/ 30 s off) to generate 300–1000 base-pair fragments. The supernatants were collected after a 13,000 rpm spin for 10 min at 4°C. 50-µl Protein G Magnetic beads (New England Biolabs) were washed twice with PBS with 5 mg/ml BSA and 10 µg of anti-Flag M2 antibody (Sigma) coupled in 500 µl PBS with 5 mg/ml BSA overnight at 4°C. Immunoprecipitation was performed with antibody-coupled beads and sonicated supernatants in ChIP buffer (20 mM Tris-HCl pH8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100) overnight at 4°C. The magnetic beads were washed twice with ChIP buffer, once with ChIP buffer including 500 mM NaCl, four times with RIPA buffer (10 mM Tris-HCl pH8.0, 0.25M LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% Na−Deoxycholate), and once with TE buffer (pH 8.0). Chromatin was eluted twice from washed beads by adding elution buffer (20 mM Tris-HCl pH8.0, 100 mM NaCl, 20 mM EDTA, 1% SDS) and incubating for 15 min at 65°C. The crosslinking was reversed at 65°C for 6 hr and RNase A (Sigma) was added for 1 hr at 37°C followed by proteinase K (Ambion, Carlsbad, CA, USA) treatment overnight at 50°C. ChIP-enriched DNA was purified using Phenol/Chloroform/Isoamyl alcohol extractions in phase-lock tubes. Then, chromatin was analyzed by qPCR using a

SYBR FAST universal kit (KAPA Biosystems, Woburn, MA, USA) with specific primers.

ChIP-seq

Library construction

Chromatin immunoprecipitation and deep sequencing library construction were performed using minor modifications of our chromatin immunoprecipitation protocol. The samples were crosslinked and prepared as described previously (Yildirim et al., 2011). The samples were transferred into 15 ml Falcon tubes and sonicated in a Bioruptor set to high for two cycles (10 min for one cycle with 30 s on/ 30 s off). ChIP samples were end-repaired, A-tailed, and adaptor-ligated using barcoded adaptors according to the manufacturer's instructions (Illumina, San Diego, CA, USA). DNA purification on Zymo Research PCR purification columns was performed following each enzyme reaction (Zymo Research, Irvine, CA, USA). The adaptor-ligated material was then PCR amplified with Phusion polymerase using 16 cycles of PCR before size selection of 200–300 bp fragments on a 2% agarose gel. The library was purified using a Zymo Gel Extraction kit, its concentration was determined using a NanoDrop (Thermo), and the integrity of each library was confirmed by sequencing 10–20 individual fragments per library. Libraries with different barcodes were pooled together and

single-end sequencing (50 bp) was performed on an Illumina HiSeq2000 at the UMass Medical School deep sequencing core facility.

Data analysis

Raw FastQ reads were first collapsed by sequence and the read occurrences were kept. The reads were then mapped to the mm9 genome using bowtie allowing at most one mismatch in every alignment. For multimappers, only one alignment was chosen randomly by the M 1 parameter setting. Each aligned location was extended downstream to a length of 150 bp. Any extension that exceeded the end of the chromosome was clipped. The extended mapped locations overlapping with simple repeats annotated by RepBase were removed. For each remaining read along with its occurrence, we calculated the relative distance to the nearest TSS and for each TSS tallied the sum of read occurrence from its upstream 2000 bp to downstream 2000 bp. The occurrences were normalized and binned in 20 bp intervals. Deep sequencing data can be obtained from GEO (http://www.ncbi.nlm.nih.gov/geo/), accession: GSE42329.

Microarray analysis

5 µg of total RNA from control (GFP), *Tip60*, *Hdac6* KD or double KD ESCs was subjected to RNA amplification and labeling using the Low Input Quick Amp Labeling Kit protocol (Agilent, Santa Clara, CA, USA) with minor modifications. Briefly, cRNA was amplified by in vitro transcription with amino-

allyl UTP (3:2 ratio for amino-ally UTP: UTP) overnight at 37°C. Then, cRNA was purified using Zymo RNA purification columns and labeled with Cy3 (GE Healthcare, Uppsala, Sweden) at room temperature for 60 min in the dark. The fluorescence intensity of Cy3 was determined by NanoDrop and 50 picomoles of cRNA was used for fragmentation and hybridization on Agilent 4X44K mouse whole-genome microarrays. Slides were scanned on Agilent DNA microarray scanner G2565CA and fluorescence data were obtained using Agilent Feature Extraction software at the UMass Medical School genomics core facility. The expression profiles from two biological replicates were analyzed as previously described (Yildirim et al., 2011). Enrichment of Gene Ontology terms and categories was performed with DAVID 6.7 (Huang et al., 2007a; Huang et al., 2007b). Microarray data can be obtained from GEO (http://www.ncbi.nlm.nih.gov/geo/), accession: GSE42329.

CHAPTER III:

PREFACE

Chapter III is a manuscript under review with myself as the first author:

R-loops regulate ES cell fate by modulating chromatin binding of key regulatory complexes

Poshen B. Chen, Hsiuyi V. Chen, Diwash Acharya, Oliver J. Rando and Thomas G. Fazzio

Contributions:

Poshen Chen performed all experiments and bioinformatics analyses, except Hsiuyi Chen prepared libraries for RIP-seq. Diwash Acharya generated several of the ESC lines. Thomas Fazzio performed Suz12 ChIP-seq and analyzed the data. Poshen Chen, Oliver Rando and Thomas Fazzio designed the experiments. Poshen Chen and Thomas Fazzio wrote the manuscript with input from all authors.

CHAPTER III: R-loops regulate ES cell fate by modulating chromatin binding of key regulatory complexes

ABSTRACT

Numerous chromatin-remodeling factors are regulated by interactions with RNA, although the contexts in which they encounter RNA and the functions of RNA binding are poorly understood. Here we show that R-loops, RNA:DNA hybrids consisting of nascent transcripts hybridized to template DNA, facilitate embryonic stem cell (ESC) differentiation by modulating the binding of two key chromatin regulatory complexes, Tip60-p400 and polycomb repressive complex 2 (PRC2). Like PRC2, the Tip60-p400 histone acetyltransferase and nucleosome-remodeling complex binds *in cis* to nascent transcripts, but unlike PRC2, transcription promotes chromatin binding by Tip60-p400. Interestingly, we observed higher Tip60-p400 and lower PRC2 levels at genes marked by R-loops near their transcription start sites. Disruption of R-loops broadly reduced Tip60-p400 and increased PRC2 occupancy, caused widespread changes in gene expression, and impaired ESC differentiation. These results define a novel mechanism by which Rloops act both positively and negatively in recruitment of key regulators of pluripotency.

INTRODUCTION

With the recent discovery of thousands of long non-coding RNAs (lncRNAs) that are expressed in mammalian cells, a considerable effort is underway to uncover the roles of specific lncRNAs in the nucleus, as well as to elucidate broadly generalizable mechanisms of action that govern their biological functions. LncRNAs function both *in cis* and *in trans* to regulate gene expression (Bonasio and Shiekhattar, 2014; Rinn and Chang, 2012), raising the possibility that these transcripts act specifically to modulate the functions of individual transcription factors or other regulatory proteins. Indeed, numerous lncRNAs have been shown to interact with transcriptional regulatory proteins, consistent with this hypothesis (Bonasio and Shiekhattar, 2014; Flynn and Chang, 2014; Rinn and Chang, 2012).

Interestingly, in a survey of 74 lncRNAs expressed in ESCs, several chromatin regulatory complexes with key roles in ESC self-renewal and/or pluripotency were shown to bind lncRNAs (Guttman et al., 2011). Many of these complexes bound to more than 30% of lncRNAs tested, and many lncRNAs were bound by multiple complexes, suggesting that either these factors are differentially programmed by dozens of individual lncRNAs to target multiple sets of genomic loci, or that these complexes bind lncRNAs relatively non-specifically. In the latter scenario, the distinct sequence of each lncRNA bound by a complex would not be predicted to impart a unique function (such as targeting the

complex to specific genomic loci), but lncRNA binding in general may serve some structural or regulatory role on the complex.

Among the first chromatin regulatory complexes shown to bind lncRNAs was polycomb repressive complex 2 (PRC2) (Pandey et al., 2008; Riising et al., 2014; Zhao et al., 2008), a highly conserved histone H3 lysine-27 methyltransferase complex important for gene silencing during development (Aloia et al., 2013). PRC2 binding to the A-repeat of the *Xist* lncRNA is thought to play a role in recruitment of the complex to the inactive X chromosome (Cifuentes-Rojas et al., 2014; Zhao et al., 2008). In addition to interacting with lncRNAs, PRC2 has been shown to bind promiscuously to nascent RNA transcripts expressed from thousands of genes, and the level of RNA binding by PRC2 catalytic subunit Ezh2 was found to correlate with RNA abundance (Davidovich et al., 2013; Kaneko et al., 2013). At first glance, PRC2 binding of nascent transcripts from active genes appears to conflict with models in which lncRNA-dependent PRC2 recruitment promotes gene silencing. However, RNA binding by PRC2 has been shown to inhibit its histone H3 lysine-27 methyltransferase activity (Cifuentes-Rojas et al., 2014; Kaneko et al., 2014). Consistent with these findings, PRC2 components bind to both silent and active genes, and active genes bound by PRC2 are not marked by H3K27me3 (Davidovich et al., 2013; Kaneko et al., 2013). These findings support a revised model in which binding of nascent transcripts at active genes helps recruit PRC2 to these loci, but maintains the complex in an inactive state (Cifuentes-Rojas et

al., 2014; Kaneko et al., 2014). In this model, PRC2 is poised to generate repressive chromatin structure and enforce silencing at these genes at a later time, should their expression be silenced by an independent mechanism. On the other hand, chemical inhibition of transcription promotes binding of PRC2 to CpG islands (including many promoter-proximal regions) throughout the genome, arguing against a model in which nascent transcripts are necessary for recruitment of PRC2 (Riising et al., 2014). Therefore, the roles of nascent transcripts in regulation of PRC2 binding and chromatin structure appear to be complex and contextspecific.

Tip60-p400 is another chromatin-remodeling complex with essential functions in ESC self-renewal and pluripotency that has also been shown to bind lncRNAs (Guttman et al., 2011). Tip60-p400 comprises a 17 subunit chromatin-remodeling complex with two catalytic subunits: the Tip60 (also known as Kat5) protein lysine acetyltransferase (KAT), which acetylates multiple lysines on histones H4 and H2A, among other proteins, and the p400 ATPase, which incorporates the H2A.Z histone variant into chromatin (Squatrito et al., 2006). We previously found that Tip60-p400 was essential for normal ESC self-renewal and pluripotency, acting simultaneously to repress some differentiation genes and activate proliferation genes (Chen et al., 2013; Fazzio et al., 2008a). Although it is not clear how Tip60-p400 simultaneously activates one group of genes and silences another, interaction with lncRNAs could potentially target the complex to

specific regions of the genome and/or tune its catalytic activities at specific targets to favor activation or silencing.

Here, we begin by addressing two fundamental questions regarding the role of RNA binding by Tip60-p400 in mouse ESCs. First, we examine whether Tip60-p400 binds specifically to lncRNAs or promiscuously to multiple categories of RNAs, similar to PRC2. Second, we elucidate the function of RNA binding by Tip60-p400. We find that, like PRC2, Tip60-p400 binds promiscuously to nascent RNAs from both coding and non-coding genes. However, unlike PRC2, whose binding to chromatin is inhibited by transcription (Riising et al., 2014), Tip60-p400 requires transcription for normal binding to most of its target promoters. Given that either nascent transcripts or the act of transcription itself promotes Tip60-p400 and inhibits PRC2 association with chromatin, we considered the context in which nascent transcripts might be encountered by these complexes. We find that Tip60-p400 is enriched, and PRC2 depleted, at genes marked by promoter-proximal R-loops, RNA:DNA hybrid structures formed when G-rich sequences on RNA hybridize with their DNA template (Ginno et al., 2012; Ginno et al., 2013), suggesting that R-loops might stimulate Tip60-p400 binding while inhibiting binding of PRC2. We tested this directly by disrupting R-loop formation in ESCs via overexpression of *RNaseH1*, and observed reduced Tip60-p400 binding and enhanced PRC2 binding at most genes targeted by each complex. Consistent with these findings, *RNaseH1* overexpressing ESCs exhibited widespread alterations in gene expression and defects in differentiation. These results

demonstrate that R-loops play a major role in regulating chromatin structure near the 5’ regulatory regions of thousands of genes in ESCs, acting both positively and negatively to control binding of chromatin-remodeling factors. More broadly, these findings suggest that RNA binding can have different effects on chromatin regulators, depending on the molecular context in which the RNA is presented.

RESULTS

Tip60-p400 interacts with nascent transcripts

Previously, in a survey of chromatin-remodeling complexes with key roles in ESCs, Tip60-p400 was shown to interact with 9 of 74 long non-coding RNAs (lncRNAs) tested (Guttman et al., 2011), raising the possibility that lncRNAs might be important for interaction of the complex with chromatin or remodeling of chromatin structure by the complex. Alternatively, Tip60-p400 might bind promiscuously to RNA, as shown for the well-studied lncRNA-interacting complex, PRC2 (Davidovich et al., 2013; Davidovich et al., 2015; Kaneko et al., 2013). To distinguish between these possibilities, we first performed unbiased identification of Tip60-p400-interacting transcripts by deep sequencing of RNA that co-immunoprecipitate with Tip60-p400 (RIP-seq). We performed biological replicate IPs of two different Tip60-p400 subunits, p400 and Ruvbl1, and observed strong correlations between replicates (Figure III.S1A-B). To elucidate the set of high-confidence Tip60-p400-binding RNAs, we focused on those enriched greater than two-fold in both replicates of p400 and Ruvbl1 RIPs compared to control RIPs, identifying approximately 2,500 transcripts in this category (Figure III.1A-D). Among these, we identified 608 enriched lncRNAs (Figure III.1C), confirming that Tip60-p400 binds to many non-coding transcripts in ESCs. More interestingly, we also observed that Tip60-p400 interacts with 1,909 coding gene transcripts (Figure III.1D), suggesting Tip60-p400 does not

bind specifically to lncRNAs, but rather interacts with a broad array of both coding and noncoding transcripts in vivo. We also tested whether Tip60-p400 only interact with highly expressed RNAs and found that these transcripts occupied a broad range of expression levels, and functional categories, consistent with the diverse set of target genes bound by Tip60-p400 (Chen et al., 2013; Fazzio et al., 2008a) (Figure III.S1C-D).

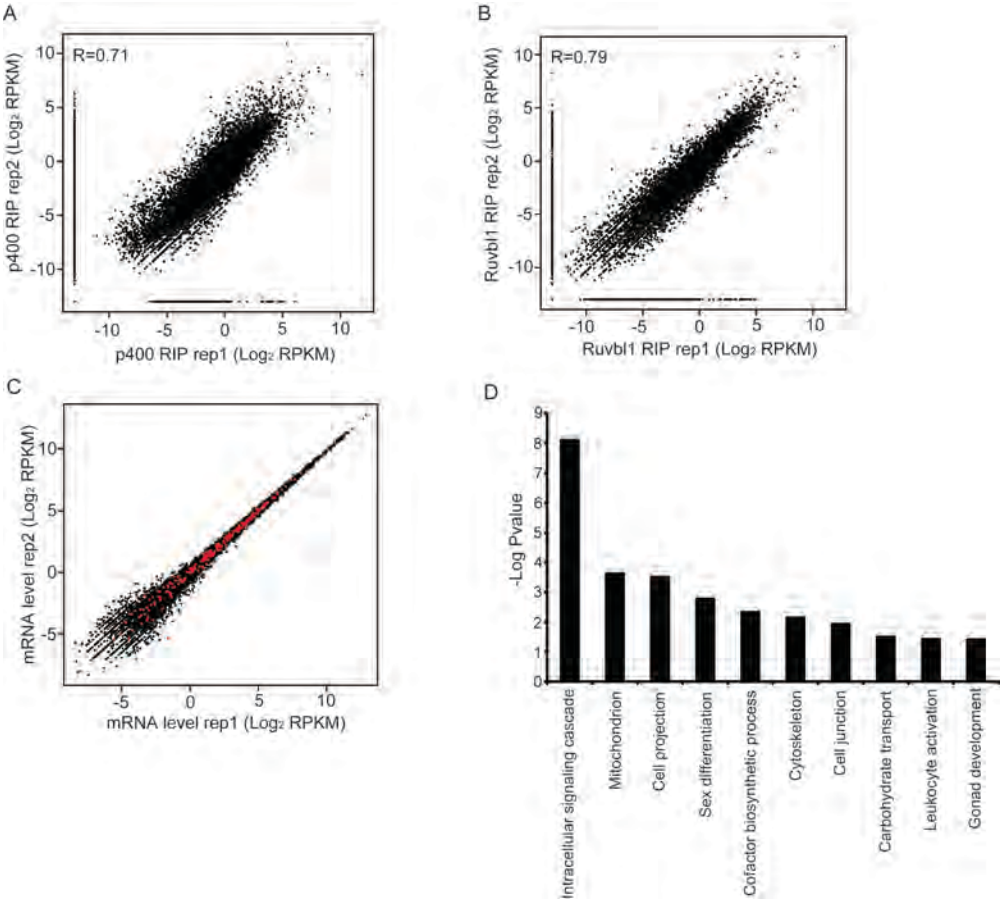


Figure III.S1: Characteristics of Tip60-p400-bound transcripts

(A-B) Scatter plots of biological replicate RIP-seq data for p400 RIPs (A) and Ruvbl1 RIPs (B). Correlation coefficients (R) are noted.

(C) Scatter plot of biological replicate RNA-seq data overlayed with Tip60-p400-bound transcripts (red). (D) Gene Ontology (GO) terms significantly overrepresented among Tip60-p400-bound transcripts, plotted as $-\text{Log}_{10}$ (p-value).

Given the abundance of coding RNAs interacting with Tip60-p400, we considered whether this complex might also interact with nascent transcripts and therefore examined the genomic locations of reads within our RIP-seq libraries. Aggregation of TSS-proximal reads from p400 and Ruvbl1 RIPs revealed peaks of interacting transcripts just downstream of TSSs (Figure III. 1E). Consistent with this observation, we observed a large fraction of reads within the first exon and first intron of many Tip60-p400-interacting RNAs (example shown in Figure III. 1F), suggesting the complex interacts with non-spliced (pre-mRNA) transcripts. Finally, when we counted all reads within each gene rather than only those within spliced mRNAs, we observed stronger enrichment in p400 and Ruvbl1 RIPs relative to controls (Figure III. 1G-H). We therefore conclude that Tip60-p400, like PRC2, binds primarily to nascent transcripts near their initiation sites.

Figure III.1: Tip60-p400 binds nascent transcripts

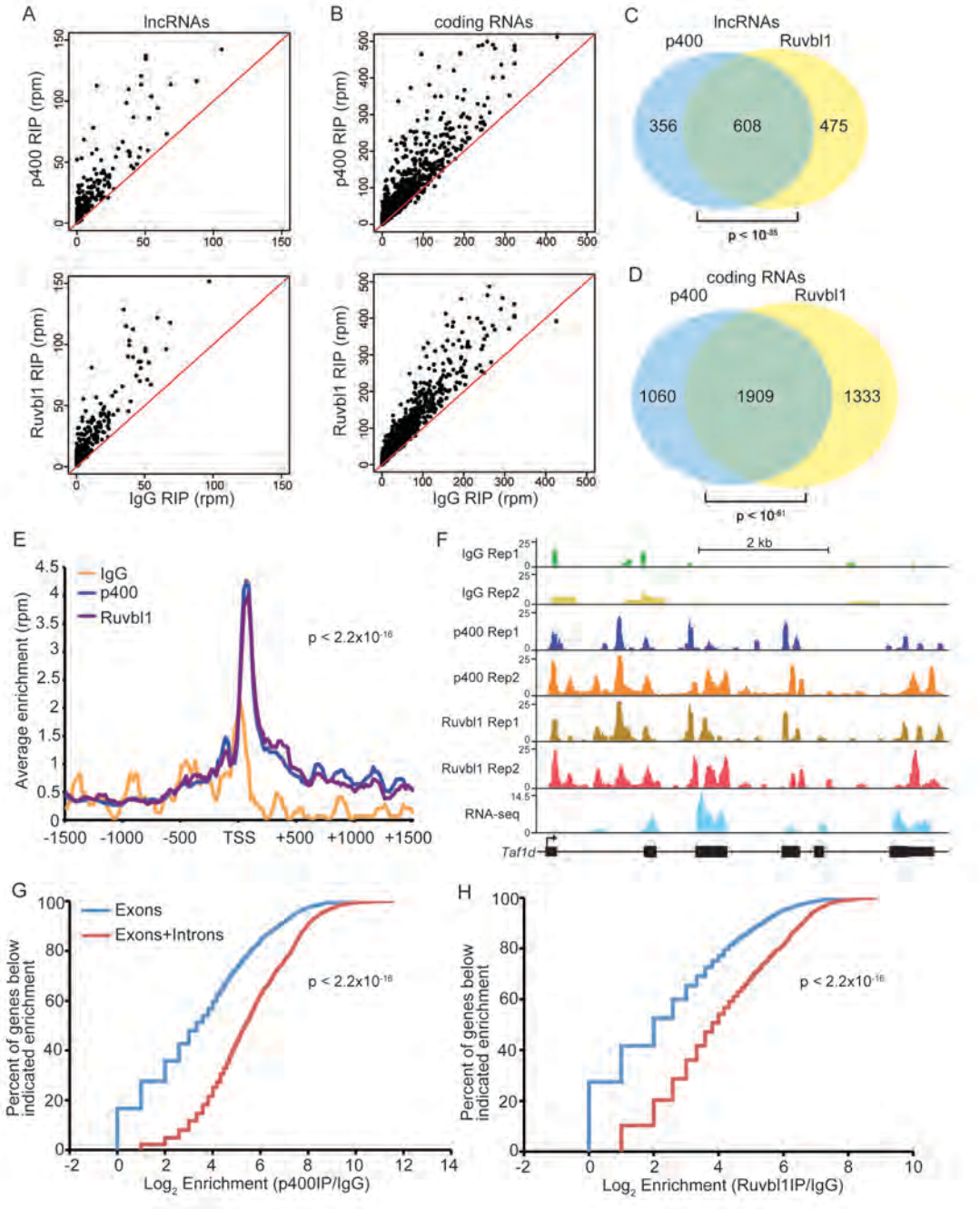
(A-B) Enrichment of transcripts in p400 (upper) or Ruvbl1 (lower) RIP-seq libraries relative to control (IgG) RIP-seq. Normalized reads from biological replicate RIP-seq experiments were averaged and plotted for lncRNAs RNAs (A) or coding RNAs (B).

(C-D) Overlap of significantly enriched lncRNAs (C) or coding RNAs (D) are shown as Venn diagrams with significance of overlap (hypergeometric test) indicated.

(E) Aggregation plot of RIP-seq data over annotated transcription start sites (TSSs).

(F) Example browser track showing locations of RIP-seq reads for the *Taf1d* gene relative to introns (thin line) and exons (black boxes).

(G-H) Cumulative distribution plots showing enrichment of reads over the entire gene (red) or only within exons (blue) in p400 and Ruvbl1 RIP-seq compared to IgG, expressed as a log2 ratio. P-values indicating significantly different distributions were calculated using a Kolmogorov-Smirnov Test.


A
lncRNAs
p400 RIP (rpm)
Ruvbl1 RIP (rpm)
IgG RIP (rpm)
B
coding RNAs
p400 RIP (rpm)
Ruvbl1 RIP (rpm)
IgG RIP (rpm)
C
lncRNAs
p400
Ruvbl1
356
608
475
p < 10-85
D
coding RNAs
p400
Ruvbl1
1060
1909
1333
p < 10-61
E
IgG
p400
Ruvbl1
Average enrichment (rpm)
p < 2.2x10-16
-1500 -1000 -500 TSS +500 +1000 +1500
F
2 kb
IgG Rep1
IgG Rep2
p400 Rep1
p400 Rep2
Ruvbl1 Rep1
Ruvbl1 Rep2
RNA-seq
Taf1d
G
Exons
Exons+Introns
Percent of genes below indicated enrichment
p < 2.2x10-16
Log2 Enrichment (p400IP/IgG)
H
Percent of genes below indicated enrichment
p < 2.2x10-16
Log2 Enrichment (Ruvbl1IP/IgG)

Transcription is necessary for chromatin binding by Tip60-p400

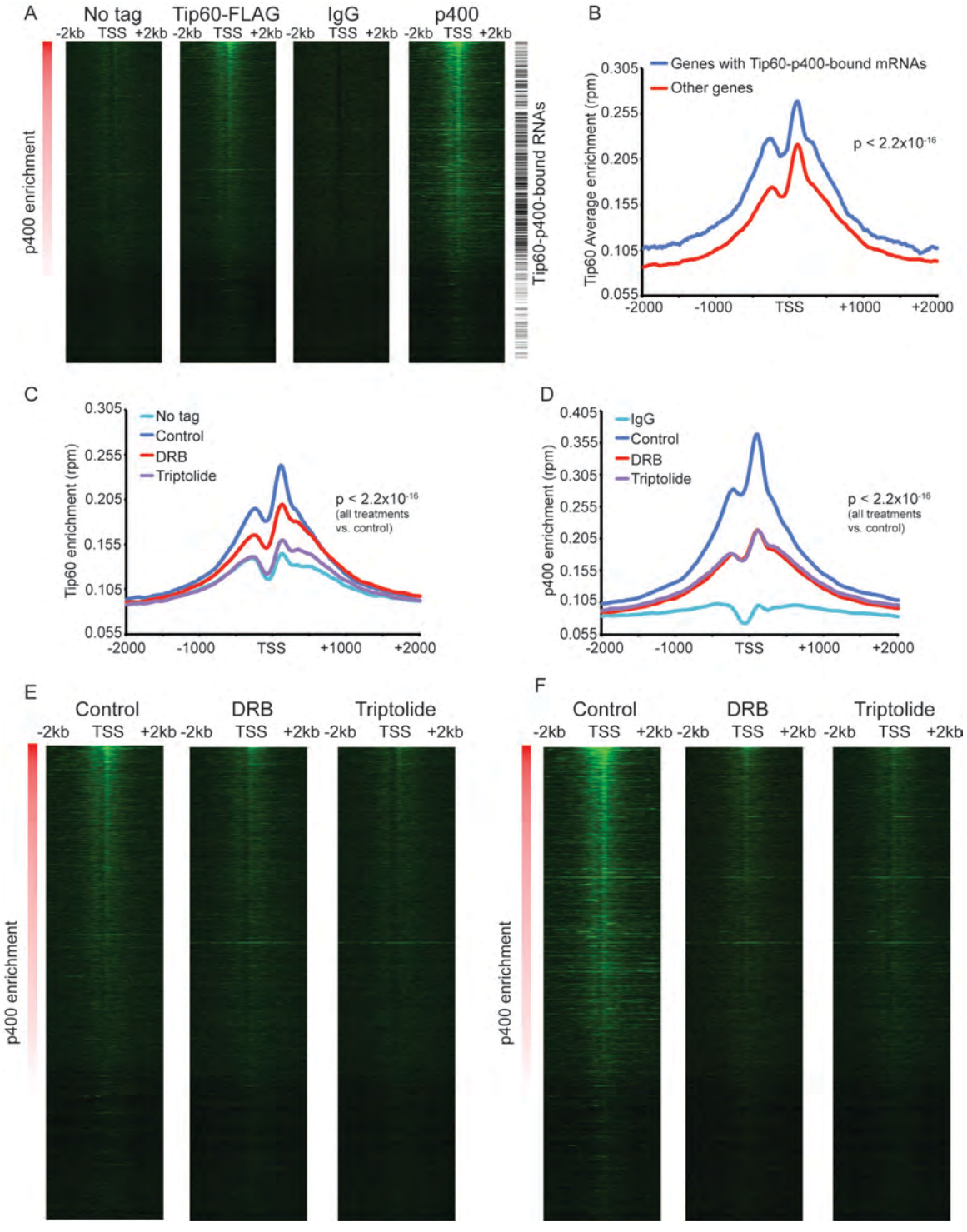
To dissect the role of RNA binding by Tip60-p400, we first tested whether the complex binds to the same regions of chromatin from which Tip60-p400-interacting RNAs are transcribed. To this end, we compared ChIP-seq maps of Tip60 and p400 localization near annotated TSSs to the set of RNAs bound by the complex. We observed significantly higher levels of Tip60 and p400 enrichment near the promoters of genes from which Tip60-p400-interacting RNAs are transcribed than all other genes, suggesting Tip60-p400 binds transcripts *in cis* (Figure III. 2A-B). The enrichment of Tip60-p400 at genes whose transcripts are bound by the complex suggested that interaction with RNA may promote chromatin binding by Tip60-p400. To address this possibility, we tested whether transcription was required for interaction of Tip60-p400 with its target genes on chromatin by addition of transcription inhibitors 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole riboside (DRB) or Triptolide to culture media for 9 or 4 hours, respectively. We determined the optimal time of inhibitor treatment based on the shortest time we observed efficient inhibition of transcription while having no effect on protein levels of Tip60 subunits (Figure III.S2A). DRB blocks transcriptional elongation by inhibiting Serine 2 phosphorylation (Ser2-P) of RNA Polymerase II (RNA Pol II) and Triptolide inhibits the ATPase activity of the XPB helicase subunit of TFIIH to induce proteasomal degradation of RNA Pol II (Bensaude, 2011). Inhibition of transcription caused a substantial reduction in the abundance of short-lived

transcripts, but did not affect levels of long-lived transcripts or any of several Tip60-p400 subunits tested (Figure III.S2B). Interestingly, both inhibitors strongly reduced Tip60 and p400 binding to a large fraction of their genomic targets (Figure III.2C-F). We validated these data at multiple genomic targets, obtaining results consistent with the genome-level data (Figure III.S2C). Together, these data demonstrate that binding of nascent transcripts by Tip60-p400, the act of transcription itself, or both are necessary for binding of the complex to a large fraction of its target genes in ESCs.

Figure III.2: Transcription promotes promoter-proximal association of Tip60-p400 with chromatin

(A) Comparison of ChIP-seq maps of Tip60 (C-terminally FLAG tagged at the endogenous *Tip60* locus), p400, or control IPs (anti-FLAG ChIP of ESCs lacking FLAG-tagged Tip60 and IgG ChIP) with Tip60-p400-interacting RNAs (overlapping in p400 and Ruvbl1 RIP-seq libraries). ChIP data is shown as heatmaps extending from -2 kb to + 2 kb from each transcription start site (TSS), with each row representing a gene, and enrichment denoted in green. Heatmaps are sorted by previously published p400 ChIP-chip data[10].

(B) Average Tip60 enrichment over TSSs of genes whose transcripts are bound by Tip60-p400 and those that are not.
(C-F) Tip60 (C,E) or p400 (D,F) enrichment by ChIP-seq (as in (A)) in control treated ESCs or ESCs treated with indicated transcription inhibitors.


A
No tag
Tip60-FLAG
IgG
p400
-2kb TSS +2kb
p400 enrichment
Tip60-p400-bound RNAs
B
Genes with Tip60-p400-bound mRNAs
Other genes
p < 2.2x10-16
Tip60 Average enrichment (rpm)
0.305
0.255
0.205
0.155
0.105
0.055
-2000
-1000
TSS
+1000
+2000
C
No tag
Control
DRB
Triptolide
p < 2.2x10-16
(all treatments
vs. control)
Tip60 enrichment (rpm)
D
IgG
Control
DRB
Triptolide
p < 2.2x10-16
(all treatments
vs. control)
p400 enrichment (rpm)
0.405
0.355
E
Control
DRB
Triptolide
p400 enrichment
F
Control
DRB
Triptolide
p400 enrichment

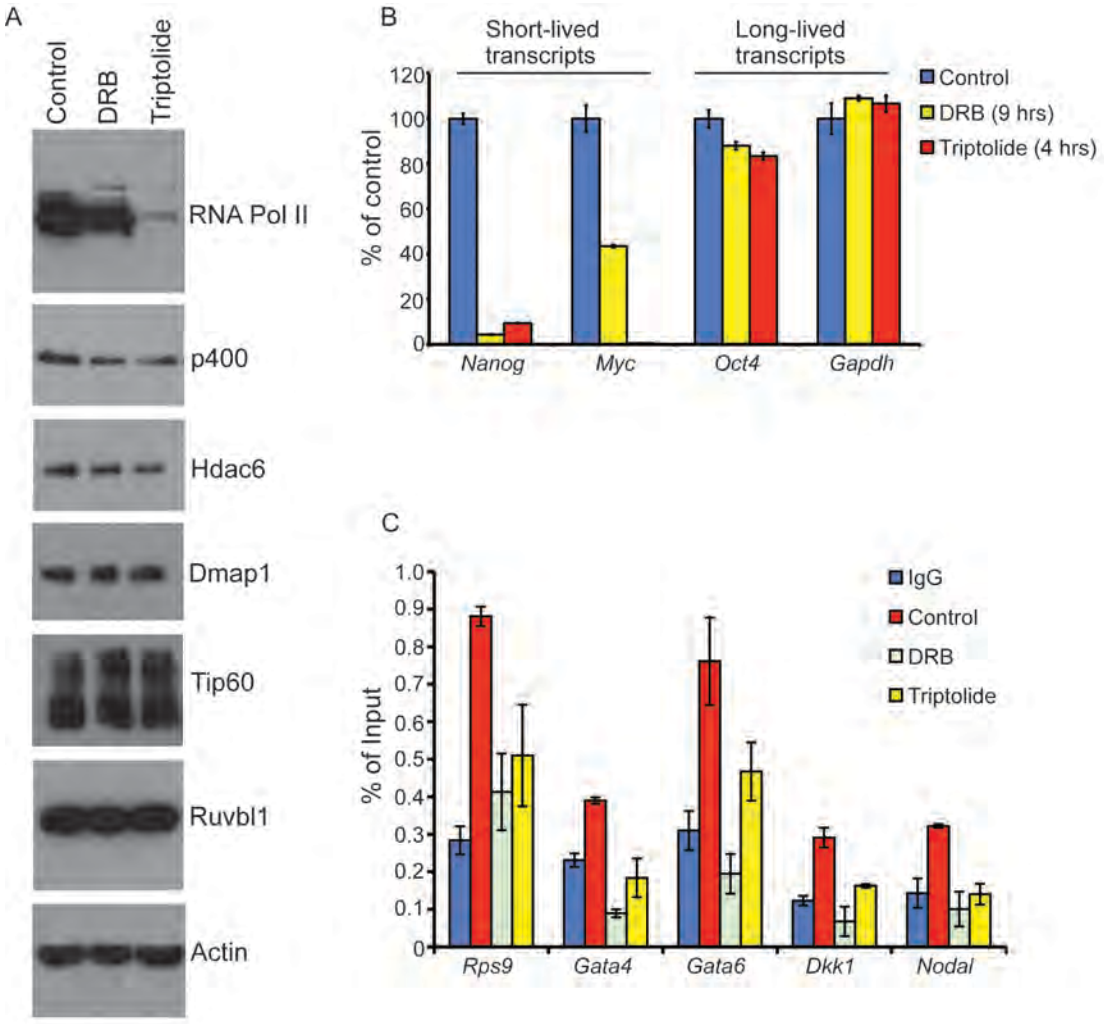


Figure III.S2: Acute transcription inhibition inhibits Tip60-p400 binding to chromatin

(A) Western blots of indicated proteins showing unchanged levels of Tip60-p400 subunits p400, Hdac6, Dmap1, Tip60, and Ruvbl1. Actin serves as a loading control.

(B) Transcript levels of rapidly-degraded transcripts (*Nanog*, *Myc*) and slowly-degraded transcripts (*Oct4*, *Gapdh*), measured by RT-qPCR. Levels in cells treated with indicated transcription inhibitors are expressed relative to control cells.

(C) ChIP-qPCR of the p400 subunit of Tip60-p400 complex with or without transcription inhibitor treatment. IgG serves as a negative control for ChIP. Enrichment levels are expressed for each locus as a fraction of input.

R-loops regulate the binding of chromatin remodelers

Nascent transcripts with G-rich domains have been shown to form R-loops near the 5' and 3' ends of transcribed genes in multiple cell types (Castel et al., 2014; Ginno et al., 2012; Ginno et al., 2013; Sun et al., 2013; Skourti-Stathaki et al., 2014). Although unresolved R-loops induce DNA damage and genomic instability (Britton et al., 2014; Groh et al., 2014; Helmrich et al., 2011; Li and Manley, 2005), 3' R-loops also regulate transcription termination (Ginno et al., 2013; Sun et al., 2013; Skourti-Stathaki et al., 2014) and R-loop formation over CpG islands functions to keep these regions relatively free of DNA methylation (Ginno et al., 2012; Ginno et al., 2013). Since Tip60-p400 binds primarily near the 5' ends of transcripts, we considered the possibility that nascent transcripts are encountered by Tip60-p400 in the form of R-loops, and that 5' R-loops may play a role in recruitment or stabilization of Tip60-p400 binding at these loci. To test this possibility, we first mapped the locations of R-loops across the genome of mouse ESCs. Immunoprecipitation of DNA-RNA hybrids using an antibody (S9.6) specific for these structures coupled to deep sequencing (DRIP-seq) has been used to map R-loops in multiple cell types (Castel et al., 2014; Ginno et al., 2012; Skourti-Stathaki et al., 2015). Due to the moderately high background and imprecise boundaries of R-loops mapped using this technique (Figure III.S3A), we modified the DRIP-seq protocol to sequence only RNAs enriched within immunoprecipitates of RNA-DNA hybrids (DRIP-RNA-seq) (Figure III.S3B). Using this DRIP-RNA-seq approach, we observed R-loops near the 5' ends of

10,595 genes (Figure III. 3A-C). We confirmed the specificity of DRIP signals in two ways: First, signals were strongly reduced when samples were treated with RNaseH prior to immunoprecipitation (Figure III.3A-C). Second, in our strand-specific DRIP-RNA-seq libraries, we observed a strong strand bias in reads in favour of annotated transcripts (Figure III.3A-C; Figure III.S3C), demonstrating the specificity of the signals we observed.

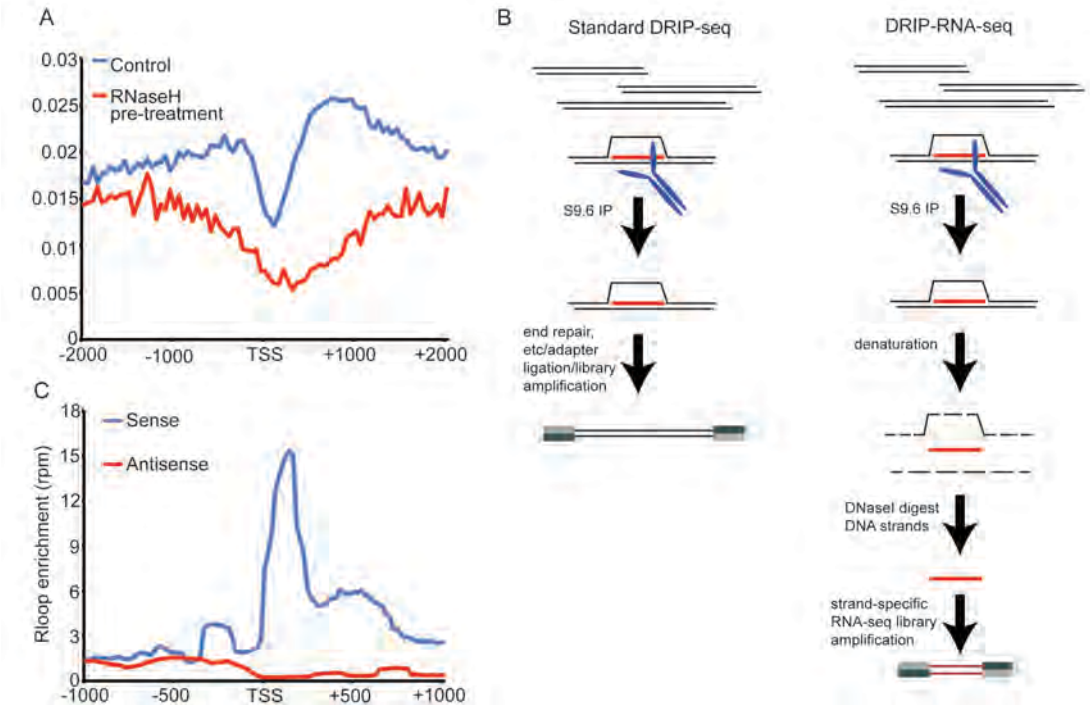


Figure III.S3: Comparison of DRIP-RNA-seq method to standard DRIP-seq

(A) Published DRIP data from human embryonal carcinoma cells[8] were aggregated over TSSs. Control: DRIP from unperturbed sample. RNaseH pre-treatment: samples were digested with RNaseH prior to DRIP, as a negative control for immunoprecipitation with the S9.6 antibody.

(B) Comparison of DRIP-seq[8] and DRIP-RNA-seq modification.

(C) Comparison of sense and antisense reads of DRIP data aggregated near promoters as in (A).

Interestingly, we observed higher average enrichment of Tip60 and p400 at genes with associated R-loops than those without (Figure III.3C-D), consistent with the possibility that R-loops may promote Tip60-p400 binding. To test this hypothesis, we overexpressed RNaseH1 in ESCs and examined its effect on Tip60-p400 binding to target genes. RNaseH1 overexpression disrupts R-loop formation by degradation of RNA species within RNA:DNA heteroduplexes, and has been shown to suppress transcription-associated recombination (Huertas and Aguilera, 2003) and DNA double strand breaks (Helmrich et al., 2011; Li and Manley, 2005) induced by R-loop formation. We observed a striking reduction in both Tip60 and p400 localization to many Tip60-p400 target genes in RNaseH1 overexpressing cells (Figure III.3E-F; Figure III.S4), demonstrating R-loops are necessary for normal association of the complex with chromatin. We conclude from these data that promoter-proximal R-loops are necessary for normal binding by Tip60-p400 to a large fraction of its target genes.

To test whether promoter-proximal R-loops specifically function in Tip60-p400 recruitment, or are required for chromatin binding by additional regulatory complexes, we focused on PRC2, due to its established RNA-binding activity in multiple cell types (Davidovich et al., 2013; Kanhere et al., 2010; Kaneko et al., 2013; Pandey et al., 2008; Rinn et al., 2007; Zhao et al., 2008). Like Tip60-p400, PRC2 binds nascent transcripts, the substrates for R-loop formation, consistent with the possibility that R-loops might promote PRC2 binding. However, since inhibition of transcription broadly stimulates PRC2 association with chromatin

(Riising et al., 2014), it was also possible that R-loops might inhibit PRC2 binding to a portion of its target genes, or have no effect at all.

Figure III.3: R-loops exert opposing effects on chromatin binding by Tip60-p400 and PRC2

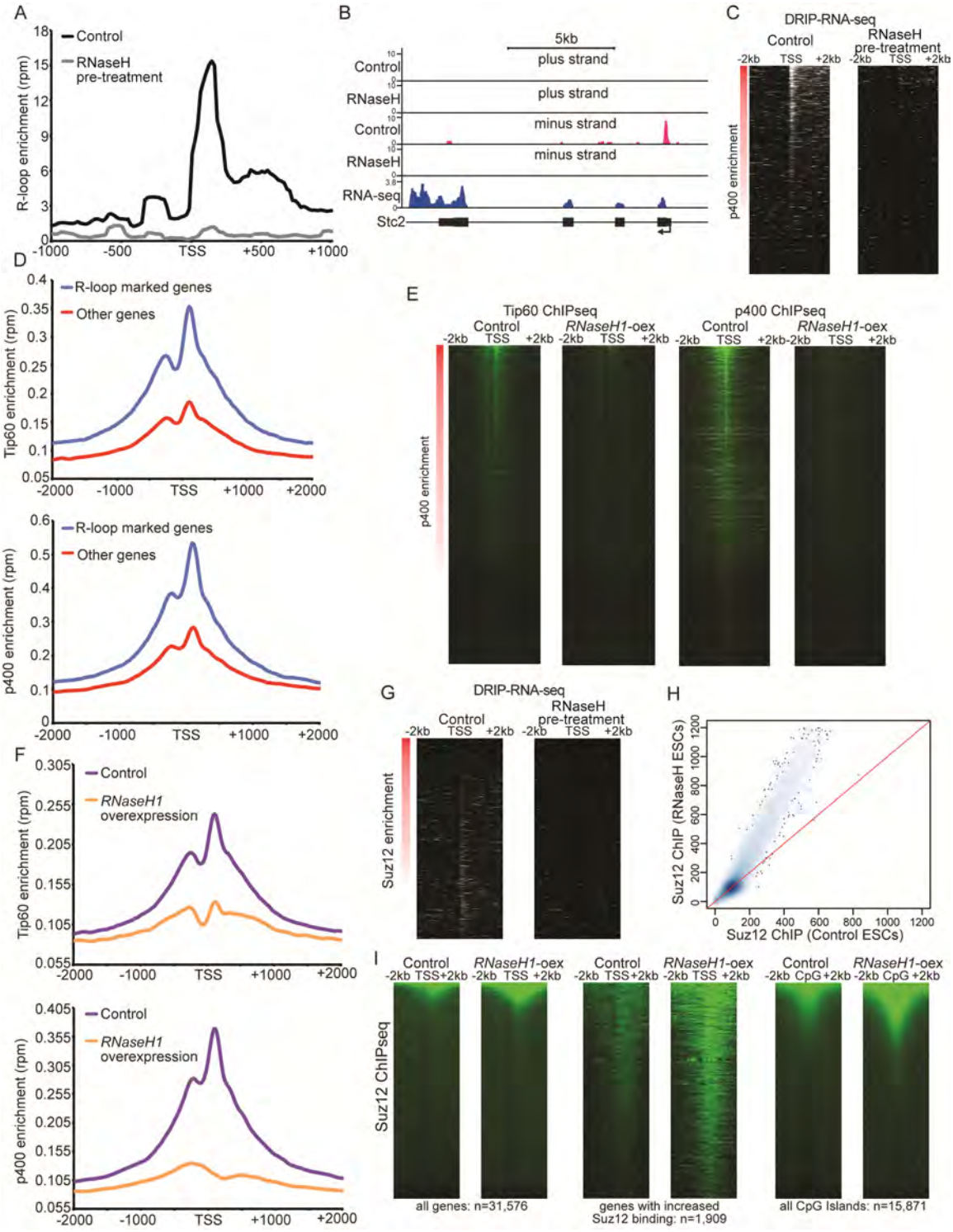
(A-C) Mapping R-loops in mouse ESCs using DRIP-RNA-seq. (A) Average R-loop enrichment over TSSs in control samples or samples pre-treated with RNaseH in vitro prior to DRIP (see Methods). (B) R-loop localization at an example gene. DRIP-RNA-seq reads were split into plus and minus strands (minus strand corresponds to *Stc2* mRNA in this example). (C) DRIP-RNA-seq data represented as a heatmap sorted by p400 ChIP-seq enrichment.

(D) Average Tip60 (upper) and p400 (lower) binding over promoters marked by R-loops (blue) and those that are not (red).

(E-F) Changes in Tip60 and p400 chromatin binding in ESCs overexpressing RNaseH1, expressed as heatmaps (E) or averaged over promoters (F).

(G) Heatmap showing DRIP-RNA-seq data as in (C), but sorted by Suz12 enrichment measured by ChIP-seq.

(H-I) Changes in Suz12 chromatin binding upon RNaseH1 overexpression, shown as a density scatter plot (H; red line corresponds to equal enrichment in both cell types) or heatmaps (I) over all genes (left) genes with increased Suz12 association in RNaseH1 overexpressing cells (middle) or annotated CpG islands (right).


A
Control
RNaseH pre-treatment
R-loop enrichment (rpm)
-1000
-500
TSS
+500
+1000
B
5kb
plus strand
Control
RNaseH
minus strand
RNA-seq
Stc2
C
DRIP-RNA-seq
Control
RNaseH pre-treatment
-2kb
TSS
+2kb
p400 enrichment
D
R-loop marked genes
Other genes
Tip60 enrichment (rpm)
p400 enrichment (rpm)
-2000
-1000
+1000
+2000
E
Tip60 ChIPseq
p400 ChIPseq
Control
RNaseH1-oex
p400 enrichment
F
Control
RNaseH1 overexpression
Tip60 enrichment (rpm)
p400 enrichment (rpm)
G
DRIP-RNA-seq
Suz12 enrichment
H
Suz12 ChIP (RNaseH ESCs)
Suz12 ChIP (Control ESCs)
I
Suz12 ChIPseq
CpG
all genes: n=31,576
genes with increased Suz12 binding: n=1,909
all CpG Islands: n=15,871

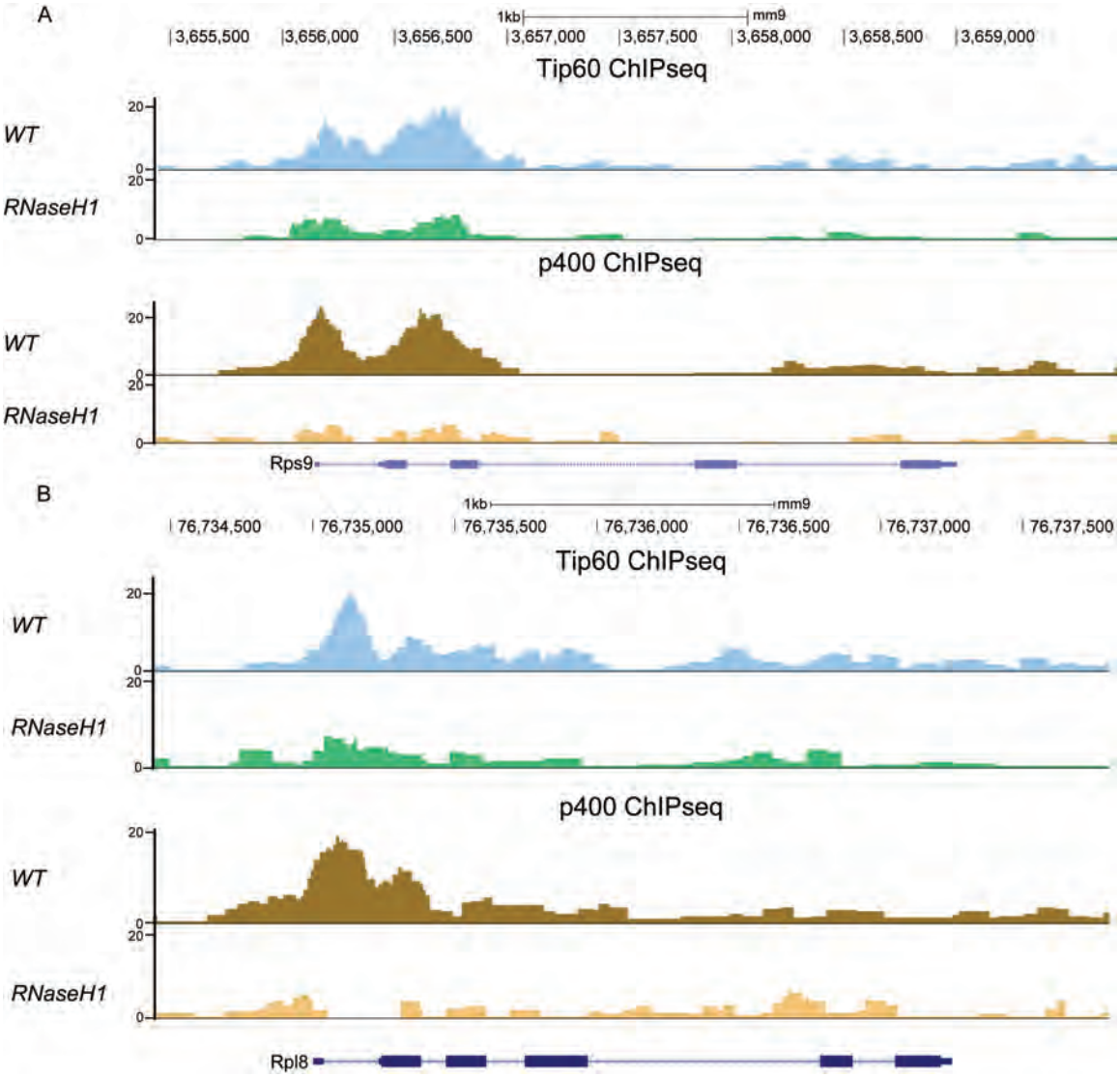


Figure III.S4: Reduction in Tip60-p400 binding to target genes upon RNaseH1 overexpression

(A-B) Browser tracks of Tip60 or p400 ChIP-seq data at two Tip60-p400 target genes, *Rps9* (A) and *Rpl8* (B) in control cells or RNaseH1 overexpressing cells.

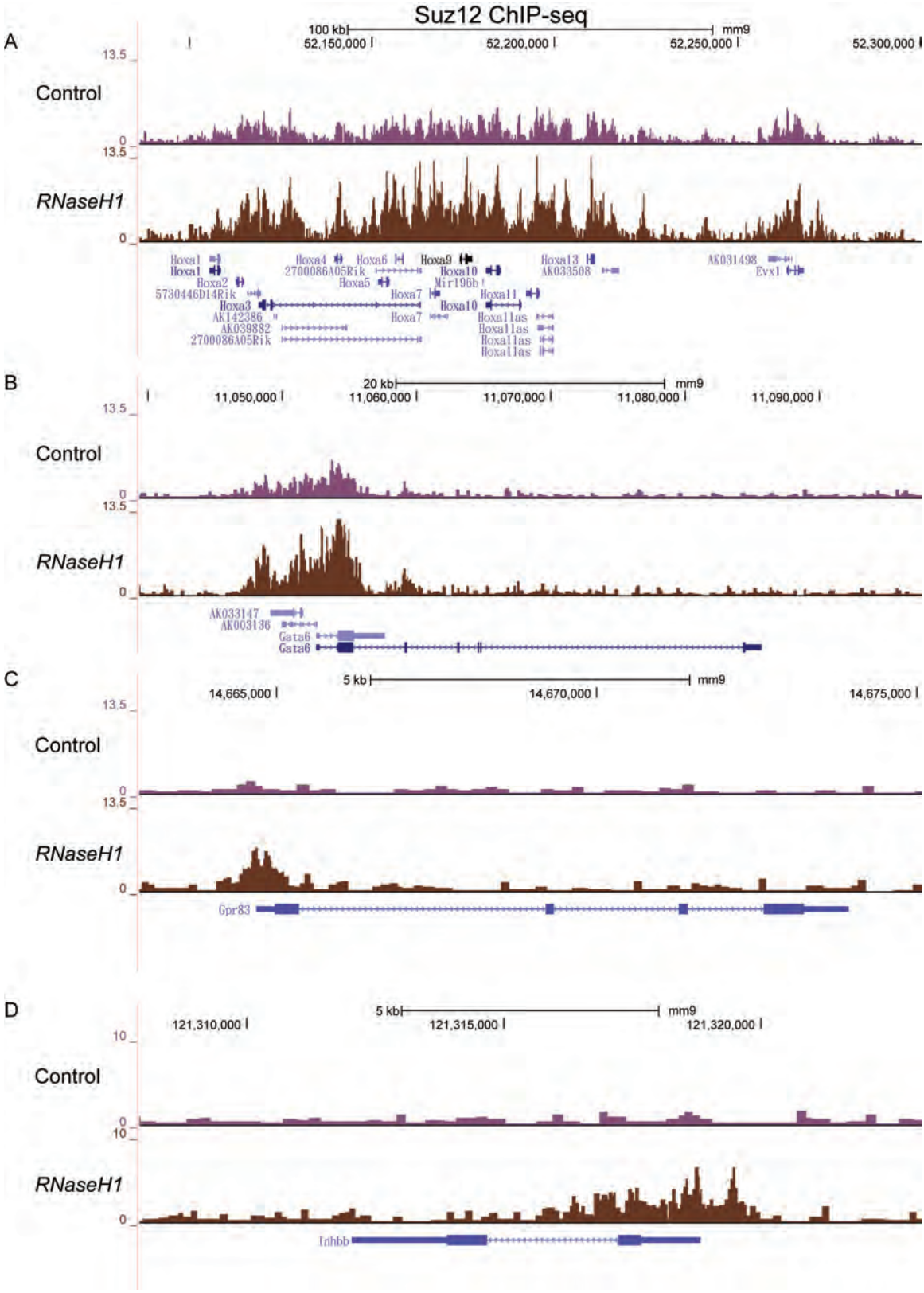
To distinguish among these possibilities, we first compared our maps of promoter-proximal R-loops to ChIP-seq maps of Suz12. Interestingly, DRIP-RNA-seq reads were poorly enriched near the promoter-proximal regions of Suz12 target genes, compared to non-Suz12 targets (Figure III.3G), suggesting that moderate to high levels of promoter-proximal R-loops inhibit PRC2 association. We tested this possibility directly by mapping Suz12 binding in the RNaseH1 overexpressing ESCs, and observed substantially increased Suz12 binding (Figure III.3G-I; Figure III.S5A-B). Some genes not significantly bound by Suz12 in control cells gained peaks of Suz12 binding (Figure III.S5C-D), and Suz12 enrichment at promoter-proximal regions normally bound by the complex increased approximately two-fold on average upon RNaseH1 overexpression (Figure III.3H). A modest increase in the width of promoter-proximal Suz12-bound domains was also evident at many targets (Figure III.3I). PRC2 has been shown to bind unmethylated CpG islands (Brinkman et al., 2012; Lynch et al., 2012; Mendenhall et al., 2010), which make up a large fraction of mammalian promoters and regulatory elements. CpG islands are kept unmethylated, in part, by the presence of R-loops (Ginno et al., 2012; Ginno et al., 2013), suggesting R-loops may help recruit PRC2 complex to these regions. However, we observed a significant enhancement of Suz12 association with CpG islands in RNaseH1 overexpressing cells (Figure III.3I), suggesting that R-loops produced from nascent transcripts inhibit PRC2 binding to these sites. Together, these data reveal that R-loop formation underlies differential recruitment of chromatin

regulatory complex at thousands of genes in ESCs, promoting Tip60-p400 association and inhibiting PRC2 association with many R-loop-associated genes.

Figure III.S5: Enhanced Suz12 binding to target genes, and emergence of novel targets upon RNaseH1 overexpression

(A-B) Genes normally bound by Suz12 show increased Suz12 enrichment in RNaseH1 overexpressing cells. Browser tracks are depicted as in Figure III.S4.

(C-D) Some genes not normally bound by Suz12 gain peaks of Suz12 binding in RNaseH1-overexpressing cells.


Suz12 ChIP-seq
A
100 kb
mm9
52,150,000
52,200,000
52,250,000
52,300,000
13.5
Control
0
13.5
RNaseH1
0
Hoxa1
Hoxa1
Hoxa2
5730446D14Rik
Hoxa3
AK142386
AK039882
2700086A05Rik
Hoxa4
Hoxa6
2700086A05Rik
Hoxa5
Hoxa7
Hoxa7
Hoxa9
Hoxa10
Mir196b
Hoxa10
Hoxa11
Hoxa11as
Hoxa11as
Hoxa11as
Hoxa11as
Hoxa13
AK033508
AK031498
Evx1
B
20 kb
mm9
11,050,000
11,060,000
11,070,000
11,080,000
11,090,000
13.5
Control
0
13.5
RNaseH1
0
AK033147
AK003136
Gata6
Gata6
C
5 kb
mm9
14,665,000
14,670,000
14,675,000
13.5
Control
0
13.5
RNaseH1
0
Gpr83
D
5 kb
mm9
121,310,000
121,315,000
121,320,000
10
Control
0
10
RNaseH1
0
Inhbb

R-loops are necessary for proper ESC differentiation

Tip60 and *p400* KD ESCs exhibit partial defects in ESC self-renewal and differentiation (Chen et al., 2013; Fazzio et al., 2008a), raising the possibility that R-loop-deficient ESCs might also be defective in one or both of these processes. When grown in medium that maintains ESC self-renewal and pluripotency, RNaseH1-overexpressing cells exhibited slightly altered colony morphology and a modest reduction in proliferation rate relative to control ESCs, but appeared to self-renew relatively normally (Figure III.4A, Figure III.S6A). Next, we examined the effect of RNaseH1-overexpression during differentiation by comparing RNaseH1 overexpressing ESCs to control ESCs or homozygous p400-mutant ESCs made by CRISPR/Cas9 cleavage and non-homologous end-joining (Mali et al., 2013). In RNaseH1 overexpressing ESCs and, to a lesser extent, *p400* mutant ESCs, we observed numerous clusters of cells with ESC-like morphology after 14 days of differentiation that stained positive for alkaline phosphatase (Figure III.4B) and the ESC-specific transcription factor Nanog (Figure III.4C), in contrast to control cells. These data suggest one major role of R-loops in ESCs is to enable their efficient response to differentiation cues, in part by promoting high levels of Tip60-p400 association and limiting levels of PRC2 association with their target genes.

Figure III.4: Disruption of R-loops impairs ESC differentiation

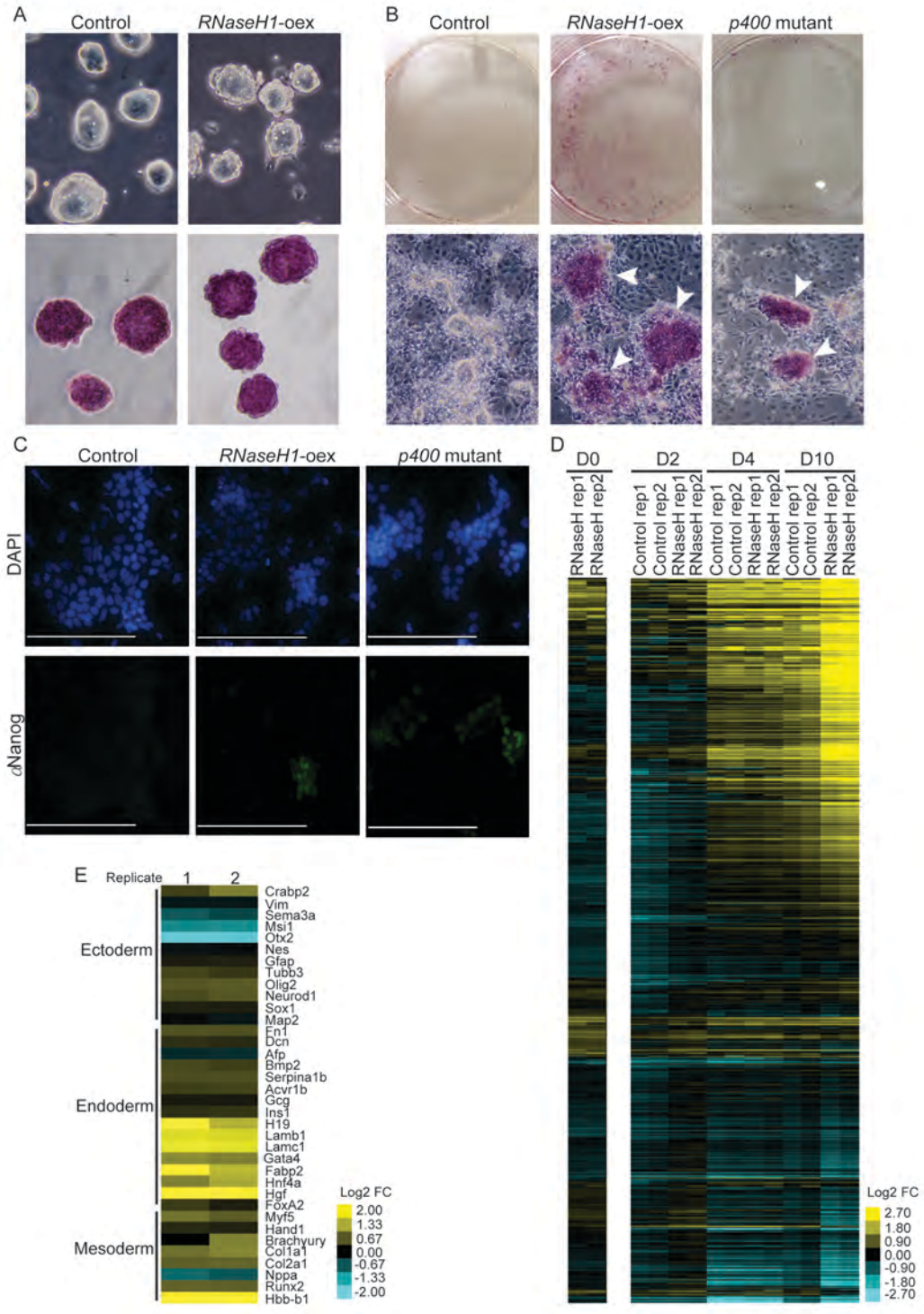
(A) Brightfield images (above) and alkaline phosphatase (AP) staining (below) of control and RNaseH1 overexpressing ESCs.

(B) AP staining of control, RNaseH1-overexpressing ESCs and p400-mutant ESCs after culture in differentiation-promoting conditions for 14 days. White arrows indicate clusters of strongly AP staining cells in RNaseH1 overexpressing cells and p400-mutant ESCs.

(C) Nanog immunofluorescence staining of control, RNaseH1-overexpressing ESCs and p400-mutant ESCs after culture in differentiation-promoting conditions for 14 days. DAPI staining is shown to identify nuclei. Scale bars in measure 100 μm.

(D) Changes in gene expression upon RNaseH1 overexpression in ESCs (d0) or during 2, 4, or 10 days of differentiation (d2, d4, and d10). Genes with significant changes (adjusted $p < 0.05$) in expression in RNaseH1 overexpressing cells relative to control cells at any time point are shown. Levels of up-regulation (yellow) or down-regulation (blue) relative to control d0 samples are illustrated for each gene.

(E) Effect of RNaseH1 overexpression on markers of primary germ layers. Shown are log2 (RNaseH1 cells/control cells) of indicated marker genes at day 10 of differentiation.


A
Control
RNaseH1-oex
B
Control
RNaseH1-oex
p400 mutant
C
Control
RNaseH1-oex
p400 mutant
DAPI
αNanog
D
D0
D2
D4
D10
RNaseH rep1
RNaseH rep2
Control rep1
Control rep2
RNaseH rep1
RNaseH rep2
Log2 FC
2.70
1.80
0.90
0.00
-0.90
-1.80
-2.70
E
Replicate 1 2
Ectoderm
Endoderm
Mesoderm
Crabp2
Vim
Sema3a
Msi1
Otx2
Nes
Gfap
Tubb3
Olig2
Neurod1
Sox1
Map2
Fn1
Dcn
Afp
Bmp2
Serpina1b
Acvr1b
Gcg
Ins1
H19
Lamb1
Lamc1
Gata4
Fabp2
Hnf4a
Hgf
FoxA2
Myf5
Hand1
Brachyury
Col1a1
Col2a1
Nppa
Runx2
Hbb-b1
Log2 FC
2.00
1.33
0.67
0.00
-0.67
-1.33
-2.00

To further characterize the role of RNaseH1 overexpressing cells in ESC differentiation, we performed RNA-seq on control or RNaseH1 overexpressing ESCs during a time course of differentiation. In ESC medium (d0), 5690 genes were significantly altered in RNaseH1 overexpressing cells (adjusted $p < 0.05$), including 2691 upregulated and 2999 downregulated genes (Figure III.4D). After two days of differentiation, several thousand genes were significantly changed in control cells relative to d0 ESCs, while RNaseH1 overexpressing cells exhibited an attenuated response (Figure III.4D), consistent with delayed or impaired differentiation of these cells. By contrast, at day 10 (d10), RNaseH1-overexpressing cells exhibited more extreme changes in expression than control cells, reflecting a broadly altered transcriptional profile within the population. Genes differentially expressed between control d10 and RNaseH1 overexpressing d10 cells were enriched within several functional categories related to differentiation and development (Figure III.S6B), consistent with the observed differentiation defect in RNaseH1 overexpressing cells. Examination of several markers of each primary germ layer revealed an increase in expression of some endoderm and a decrease in some ectoderm genes, suggesting a skewing in lineage specification in RNaseH1 overexpressing cells (Figure III.4E). We conclude that massive alterations in gene expression due to loss of R-loops in RNaseH1 overexpressing ESCs lead to a defect in differentiation in which some cells fail to differentiate and the overall population of cells is skewed toward endodermal lineages.

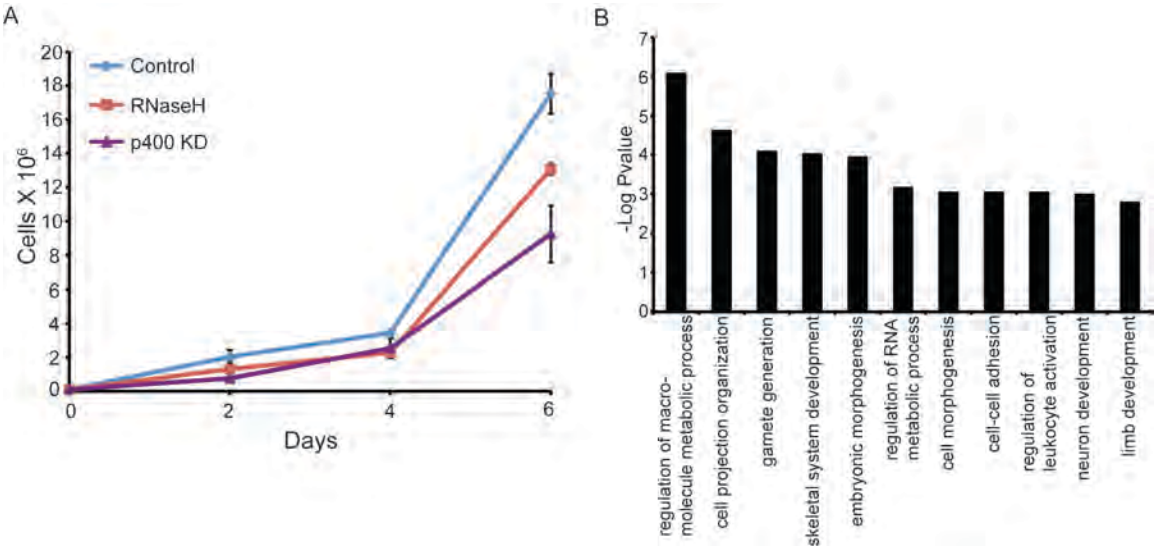


Figure III.S6: Characteristics of RNaseH1 overexpressing ESCs

(A) Growth curve of control or RNaseH1 overexpressing ESCs compared to *p400* knockdown (KD) ESCs.

(B) Gene Ontology (GO) terms significantly overrepresented among genes differentially expressed in RNaseH1 overexpressing cells compared to control cells after 10 days of differentiation, plotted as $-\text{Log}_{10}$ (p-value).

Sfpq promotes RNA binding by Tip60-p400

Next, we sought to determine how Tip60-p400 interacts with RNA. Previously, we purified Tip60-p400 complex from ESCs and identified several new interacting proteins (Chen et al., 2013). Among the new Tip60-p400 interacting proteins, some are known RNA-binding proteins (RBPs) (Kwon et al., 2013; Matias et al., 2006), raising the possibility that RBPs in Tip60 complex are important for Tip60-RNA interaction. Splicing factor, proline- and glutamine- rich (Sfpq), also commonly known as polypyrimidine tract-binding protein-associated-splicing factor (PSF), is one such interacting factor that contains two RNA recognition motifs (RRMs) (Patton et al., 1991). Sfpq has been shown to bind DNA as well as RNA and plays a role in transcriptional regulation (Akhmedov and Lopez, 2000; Hirose et al., 2014; Imamura et al., 2014; Song et al., 2005). To confirm that Sfpq is a bona fide Tip60-p400 interacting protein, we performed co-immunoprecipitation experiments and observed Tip60-Sfpq interaction (Figure III.5A). Heat-shock 70KD protein 8 (Hspa8), another potential Tip60-interaction protein from our previous mass spectrometry data, was also shown to interact with RNA (Kwon et al., 2013). However, we didn’t observe Tip60-Hspa8 interaction from co-immunoprecipitation experiments, suggesting this interaction is less stable or transient (data not shown). To identify which subunit of Tip60-p400 complex directly interacts with RNA, we adapted a method that was systematically used to identify RBPs by combining UV-crosslinking of protein and RNA in living cells, oligo (dT) capture of polyadenylated RNAs, and western-

blotting (Castello et al., 2013; Kwon et al., 2013). This procedure incorporates stringent washes (500 mM salt, 0.5% sodium deoxycholate), which removes most proteins not directly cross-linked to RNA. Sfpq and Ruvbl1 both could directly interact with RNA after UV-crosslinking, whereas but not other subunits of Tip60-p400 that we tested did not, indicating that Sfpq and Ruvbl1 are important for Tip60-RNA interaction (Figure III.5B). Next, we sought to determine whether Sfpq is important for Tip60-RNA interaction. Strikingly, we observed a reduction in p400 binding to many non-coding and coding target RNAs upon *Sfpq* KD (Figure III.5D), suggesting that Sfpq promotes Tip60 complex binding to target RNAs.

To test whether Sfpq is necessary for gene regulation by Tip60-p400 complex in ESCs, we used DNA microarrays to examine changes in gene expression upon *Tip60*, *p400*, *Ruvbl1*, *Hdac6* or *Sfpq* KD. Consistent with previous results (Chen et al., 2013; Fazzio et al., 2008a), 2021 genes were up-regulated and 1139 genes were down-regulated upon *Tip60* KD (adjusted $p < 0.05$), indicating the major role of Tip60 to repress gene expression in ESCs (Figure III.5C). Among those misregulated genes, many of them were also misregulated upon *p400*, *Ruvbl1*, *Hdac6* or *Sfpq* KD, suggesting Sfpq, like other subunits of Tip60 complex, is required for Tip60-mediated gene regulation (Figure III.5C). Furthermore, similar to the defect upon *p400* KD during differentiation, embryoid body formation was significantly impaired in *Sfpq* KD ESCs (Figure III.5E-F). Overall, Sfpq interacts with Tip60-p400 and regulates

Tip60-targeted genes. It will be important to determine whether Sfpq interacts with R-loops and affects recruitment of Tip60-p400 to R-loop marked genes in the future study.

Figure III.5: Sfpq is necessary for Tip60-RNA interaction

(A) Validation of Sfpq interaction. Western blots for Sfpq, p400 and Dmap1 following Immunoprecipitation with anti-FLAG antibody from nuclear extracts derived from the indicated ESC lines.

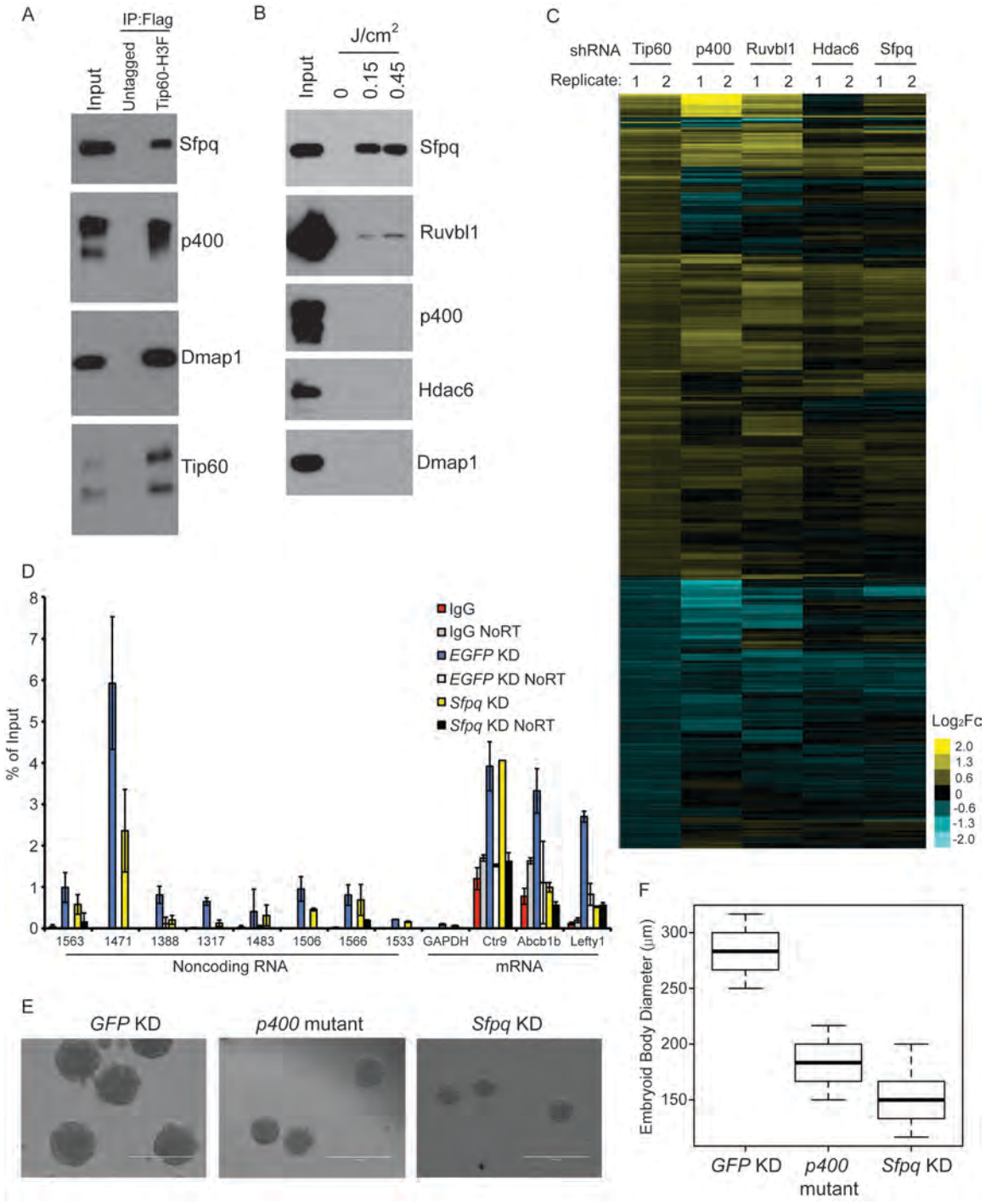
(B) Western blots for Sfpq, Ruvbl1, p400, Hdac6 and Dmap1 following Immunoprecipitation with oligo-dT beads from UV-crosslinked nuclear extracts.

(C) Unsupervised hierarchical clustering of genes misregulated (adjusted $p<0.05$) upon *Tip60* KD. Up-regulated genes are indicated in yellow, and down-regulated genes are indicated in blue.

(D) RIP-qPCR from indicated KD ESCs. Co-immunoprecipitating RNAs with anti-p400 antibody from cells knocked down as indicated are expressed as a fraction of the input. Shown are the mean ± SD values of three technical replicates from one representative experiment of two biological replicates performed.

(E) EB formation assay. Brightfield images of EBs formed by hanging-drop cultures of ESCs knocked down as indicated and p400-mutant ESCs, then cultured in differentiation medium. Scale bars equal 400 μm.

(F) Box plot quantification of the range of EB size by diameter. The upper and lower limits of the box correspond to the 75th and 25^{th} percentiles of indicated samples, respectively, and the dark line corresponds to the median of each box. At least 40 EBs were measured for each sample.


A
IP:Flag
Input
Untagged
Tip60-H3F
Sfpq
p400
Dmap1
Tip60
B
J/cm2
Input
0
0.15
0.45
Sfpq
Ruvbl1
p400
Hdac6
Dmap1
C
shRNA Tip60 p400 Ruvbl1 Hdac6 Sfpq
Replicate: 1 2 1 2 1 2 1 2 1 2
Log2Fc
2.0
1.3
0.6
0
-0.6
-1.3
-2.0
D
% of Input
IgG
IgG NoRT
EGFP KD
EGFP KD NoRT
Sfpq KD
Sfpq KD NoRT
1563 1471 1388 1317 1483 1506 1566 1533 GAPDH Ctr9 Abcb1b Lefty1
Noncoding RNA
mRNA
E
GFP KD
p400 mutant
Sfpq KD
F
Embryoid Body Diameter (μm)
GFP KD
p400 mutant
Sfpq KD

DISCUSSION

We have uncovered a role for promoter-proximal R-loops in shaping the chromatin landscape and controlling the differentiation program in ESCs. R-loops are necessary for normal chromatin binding at most Tip60-p400 target genes, but inhibit binding of PRC2 to its targets. Therefore, with regards to these key regulators of ESC pluripotency, R-loops help segregate genes into classes that are highly bound by Tip60-p400 but not PRC2, those strongly bound by PRC2 but not Tip60-p400, and some genes that are bound minimally by both complexes. Whether additional chromatin regulators are affected positively or negatively by the presence of R-loops to further compartmentalize the chromatin landscape of genes in ESCs remains to be tested. However, given the large number of chromatin regulatory complexes found to bind lncRNAs (Bonasio and Shiekhattar, 2014; Flynn and Chang, 2014; Guttman et al., 2011; Rinn and Chang, 2012), it seems likely additional factors will bind nascent transcripts in the form of R-loops.

Context-dependent effects of RNA binding on PRC2 function

While the effects of RNA on Tip60-p400 function have not been studied in detail, transcription appears to exert both positive and negative effects on the function of polycomb complexes in multiple systems (Cavalli and Paro, 1998; Cifuentes-Rojas et al., 2014; Davidovich et al., 2015; Davidovich et al., 2013; Kaneko et al.,

2010; Kanhere et al., 2013; Klymenko and Müller, 2004; Kaneko et al., 2014; Riising et al., 2014; Schmitt et al., 2005; Zhao et al., 2008). PRC2 binds to the A-repeat of the *Xist* lncRNA, and this is thought to help recruit the complex to the inactive X-chromosome (Cifuentes-Rojas et al., 2014; Zhao et al., 2008). Ezh2 binds nascent transcripts from numerous active genes, and has been shown to bind near the promoters of most active genes at low levels (Davidovich et al., 2013; Kaneko et al., 2013). In addition, RNA binding inhibits the histone methyltransferase activity of PRC2, suggesting that binding of nascent transcripts holds PRC2 activity in check at active promoters, and PRC2 remains poised for histone methylation at these genes once transcription is silenced by another mechanism (Cifuentes-Rojas et al., 2014; Kaneko et al., 2014). However, as with inhibition of transcription (Riising et al., 2014), we find that disruption of R-loops broadly stimulates PRC2 binding, suggesting that the effects of nascent transcription on PRC2 recruitment may be context-dependent. For example, nascent transcripts with G-rich sequences prone to R-loop formation may prevent PRC2 binding while different nascent transcripts that do not form R-loops may allow PRC2 binding, while inhibiting its methyltransferase activity.

Multifaceted recruitment of Tip60-p400 to target genes

Although inhibition of transcription enhances PRC2 binding at CpG islands, including many promoters (Riising et al., 2014), transcriptional inhibitors strongly reduced Tip60-p400 association with target gene promoters. Importantly,

RNaseH1 overexpression mimicked the effect of transcription inhibition on both complexes – enhancing Suz12 association and inhibiting Tip60-p400 association. Since RNaseH1 degrades RNA species only within RNA:DNA hybrids, this finding demonstrates that nascent transcripts, rather than the act of transcription itself, are necessary for Tip60-p400 recruitment and reduction of chromatin binding at many genes by PRC2. In addition, these data show that chromatin regulatory complexes encounter nascent transcripts at many genes in the form of R-loops, rather than free RNA. Although we observed a strong correlation between the presence of promoter-proximal R-loops and Tip60-p400 binding, several lines of data indicate R-loops are not sufficient for Tip60-p400 recruitment. First, R-loops are also prevalent at transcriptional termini (Castel et al., 2014; Ginno et al., 2013; Skourti-Stathaki et al., 2014), which are not strongly bound by Tip60-p400. Secondly, the PHD domain of the Ing3 subunit of Tip60-p400 was previously shown to bind histone H3 methylated on lysine-4 (H3K4me3) (Shi et al., 2006), and we previously showed that knockdown of enzymes required for H3K4me3 deposition led to a moderate reduction of Tip60-p400 binding to several target genes (Fazzio et al., 2008a). While it remains possible that H3K4me3 is necessary for formation of R-loops, this possibility is unlikely, given the modest effect of knockdown of core Set1/MLL subunits Dpy-30 or Rbbp5 on gene expression (Jiang et al., 2011) compared to what we observed upon disruption of R-loops in *RNaseH1*-overexpressing ESCs. These

data suggest that recruitment of the large 17 subunit Tip60-p400 complex to target sites on chromatin is a function of multiple different mechanisms.

Disruption of R-loops alters gene expression in ESCs and impairs differentiation

RNaseH1 overexpression causes misregulation of thousands of genes in ESCs, resulting in impaired differentiation, consistent with altered functions of both complexes. These effects of *RNaseH1* overexpression result in only a modest reduction in growth rate in ESCs, but inhibit normal ESC differentiation. In addition to the persistence of undifferentiated cells persisting for at least 14 days after induction of differentiation, we observed skewing in the pattern of lineage markers after 10 days, with an overrepresentation of endodermal and underrepresentation of ectodermal markers.

Although loss of Tip60-p400 binding and gain of PRC2 binding in *RNaseH1* overexpressing ESCs likely contributes to these phenotypes, it is also possible R-loops modulate the binding of additional factors that regulate ESC differentiation. In support of this possibility, a large number of genes misregulated in *RNaseH1* overexpressing ESCs were not targets of either Tip60-p400 or PRC2. Nonetheless, the opposing effects of R-loops on Tip60-p400 and PRC2, and their importance for normal ESC differentiation, suggest an additional layer of complexity controlling gene regulation in ESCs and changes in expression

during differentiation. These findings also suggest that factors regulating R-loop formation or clearance may have additional roles in gene regulation in multiple cell types.

METHODS

Antibodies

Anti-p400 (A300-541A; Bethyl), anti-RUVBL1 (10210-2-AP; Proteintech), anti-FLAG-M2 (F1804; Sigma), anti-DNA-RNA hybrid S9.6 (ENH001; KeraFAST), anti-SUZ12 (A302-407A; Bethyl), anti-NANOG (A300-398A; Bethyl), anti-rabbit-IgG (ab37415; Abcam), anti-HDAC6 (07-732; Millipore), anti-DMAP1 (10411-1-AP; Proteintech), anti-RNA Polymerase II (sc-899; Santa Cruz Biotechnology), anti-ACTIN (A5316; Sigma) antibodies were used.

Cell culture and treatment

ESCs were grown under feeder-free conditions as described (Chen et al., 2013). The *Tip60-H3F* (+/+) line was generated by CRISPR/Cas-mediated genome editing into E14 using homology arms of approximately 900 bp surrounding a 6-Histidine-3XFLAG tag described previously (Chen et al., 2013), immediately 5’ of the endogenous stop codon (guide RNA sequence: 5`-AAGCCAGTTATCCTCGGAGT-3`). The *p400* homozygous mutant line was made similarly, using a guide RNA (5`-TGGCTGATGAAGCAGGGCTT-3`)

specific for the ATPase domain of the p400 gene and no homology template. Both alleles of the *Ep400* gene were sequenced, revealing deletions of 135 bp in exon 15 in the two alleles. Full-length mouse RNaseH1 (NM_001286865.1) including an N-terminal 3XHA tag was synthesized (gBlocks, Integrated DNA Technologies) and cloned into the EcoRI-XhoI fragment of the pCAGGS-ires-Hygro vector. The RNaseH1-overexpressing cells were generated by transfection of pCAGGS-RNaseH1-ires-Hygro plasmid into the *Tip60-H3F* line and selection with Hygromycin B (Roche). For inhibition of transcription, cells were treated with 100 μM or 10 μM of DRB or Triptolide (Sigma), respectively, as described (Riising et al., 2014). We tested several time points of treatment for inhibition of transcription by RT-qPCR and the protein levels of several subunits in Tip60 complex by Western blotting. We determined the optimal time of inhibitor treatment based on the shortest time we observed efficient inhibition of transcription while having no effect on protein levels of Tip60 subunits. For ESC differentiation, 10^6 ESC cells were suspended in medium lacking LIF and cultured in non-cell culture treated petri dishes for 2 days. Subsequently, cells were transferred to gelatin-coated cell culture dishes in media lacking LIF for the number of days indicated. Cells were fixed or RNA was isolated at the indicated time points.

Alkaline phophatase staining

After 14-days of differentiation, cells were stained for AP activity using a kit (EMD Millipore) according to the manufacturer's instructions.

Immunofluorescence staining

Cells were fixed with 4% paraformaldehyde, blocked with blocking buffer (10% normal goat serum, 0.3% Triton X-100 in PBS) for 1 hour and stained with anti-Nanog antibody (1:100 dilution) overnight at 4°C. The next day, cells were washed and stained with Alexa Fluor 488-conjugated secondary antibodies (1:1,000) (Life Technologies). The nuclei were stained with DAPI, and the slides were imaged on a EVOS FL microscope (Life Technologies).

Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation and library construction (ChIP-seq) were performed as described previously (Chen et al., 2013). Libraries with different barcodes were pooled and single-end sequencing (50bp) was performed on an Illumina HiSeq2000 at the UMass Medical School deep sequencing core facility.

RIP-seq

Cells were lysed using an NE-PER Extraction kit (Thermo Fisher) to isolate cytoplasmic and nuclear fractions. For immunoprecipitation, 1.5 mg of

nuclear extracts were treated with DNase I (New England Biolabs) and pre-cleared with Protein A magnetic beads (New England Biolabs) for 3 hours. Cleared nuclear extract was incubated with specific antibodies in IP buffer (50 mM Tris-HCl pH7.4, 250 mM NaCl, 0.1% Triton X-100), plus 1X HALT protease inhibitors (Thermo Fisher) and SUPERaseIn (Life Technologies) overnight at 4°C. The next day, pre-washed Protein A magnetic beads were added to IP samples and incubated for another 4 hours at 4°C. The magnetic beads were sequentially washed with IP buffer twice, high-salt IP buffer (50 mM Tris-HCl pH7.4, 500 mM NaCl, 0.1% Triton X-100, 0.5 % sodium deocycholate) four times, and IP buffer two more times. RNA was eluted from beads and purified by TRIzol (Life Technologies) extraction and precipitated at -80°C for at least 2 hours. For RIP-seq, 10-50 ng of RIP enriched RNA and Adaptor 1 (5'-CTGAACCGCTCTTCCGATCTNNNNNN-3') were used for first-strand cDNA synthesis with Superscript III Reverse Transcription Kit (Life Technologies). After first-strand cDNA synthesis, RNA was degraded by sodium hydroxide and cDNA was purified by SILAN beads (Life Technologies). To preserve strand information, Adaptor 2 with the modification of 5' phosphorylated and 3' dideoxy-C (5'-p-NNACGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'ddC) was ligated to the 3' end of first-strand cDNA using T4 RNA ligase 1 (New England Biolabs). The ligated material was purified by SILAN beads and PCR amplified with Illumina primers using 18 cycles of PCR. To remove PCR primers, libraries were purified by AMPure XP beads (Beckman Coulter). Libraries with

different barcodes were pooled together and sequenced as described for the ChIP-seq libraries.

DRIP-RNA-seq

Nucleic acid extraction, immunoprecipitation and library preparation were performed as described previously (Ginno et al., 2012) with the following modifications (cartooned in Figure III.S3b). The immunoprecipitated material (with and without RNase H treatment) were denatured at 94°C for 1 min and cooled on ice. To reduce DNA background, the samples were treated with DNaseI at 37°C for 30 minutes and RNA was purified using Phenol/Chloroform/Isoamylalcohol extraction. 38 pmol of adaptor 1 (CTGAACCGCTCTTCCGATCTNNNNNN) was combined with 50 ng of S9.6 enriched RNA for first-strand cDNA synthesis with a Superscript III Reverse Transcription Kit (Life Technologies). After first-strand cDNA synthesis, RNA was degraded by sodium hydroxide and cDNA was purified by SILAN beads (Life Technologies). To preserve strand information, Adaptor 2 (5'-p-NNACGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'ddC) was ligated to the 3' end of first-strand synthesized cDNA. The ligation material was purified by SILAN beads and then PCR amplified with Illumina primers using 18 cycles of PCR. To remove PCR primers, libraries were purified by AMPure XP beads (Beckman Coulter). Libraries with different barcodes were pooled together and sequenced as described above.

RNA-seq

Strand-specific RNA-seq libraries for ESCs and differentiated ESCs were performed as described previously (Hainer et al., 2015).

Sequencing data analysis

Barcodes were removed and reads were mapped to the mouse genome (mm9) using Bowtie-1.0.0 (Langmead et al., 2009) for ChIP-seq and TopHat-2 (Kim et al., 2013) for RIP-seq and DRIP-RNA-seq. For ChIP-seq and DRIP-RNA-seq, aligned sequences were processed in HOMER (Heinz et al., 2010) by using “annotatePeaks” command to bin the regions of interest in 20 bp windows and sum the reads within each window. Average enrichment was calculated by normalizing the reads in each window to total reads, dividing by the number of regions of interest and presented in reads per million (rpm). For RIP-seq data, aligned sequences were processed in HOMER by using “analyzeRNA” command to calculate, normalize and present in reads per kilobase per million mapped reads (rpkm) for each reference gene. Fold-change was caluculated by dividing rpkm from IP with IgG control. Noncoding RNA data was obtained from GENCODE release M1 dataset (Pervouchine et al., 2015) and previously published lncRNAs (Guttman et al., 2011). For RNA-seq, rRNA sequences were removed before transcript quantification using RSEM (Li and Dewey, 2011). Differentially expressed genes were identified by DESeq2 (Love et al., 2014) and

significantly changed genes were selected using a cutoff of adjusted p-value < 0.05, comparing RNaseH1-overexpressing cells to control cells at each time point during differentiation.

CHAPTER IV: CONCLUSIONS AND PERSPECTIVES

RED-seq

We developed RED-seq to measure chromatin accessibility not only within open regions of the genome (as is typical of genome-scale techniques routinely used), but in all regions of the genome, as well as to compare chromatin accessibility in different cellular states. In addition to open regions that can also be identified by DNase-seq and FAIRE-seq, we found that RED-seq can also be used to probe intranucleosomal accessibility. We observed increased accessibility of DNA within H2A.Z- and H3.3-containing nucleosomes and strongly reduced DNA accessibility in those regions after depletion of H2A.Z and H3.3 loading factors. Therefore, RED-seq is a powerful tool to study the role of chromatin regulators *in vivo* since the product of the enzymatic function of many chromatin regulators is some form of nucleosome array.

Chromatin regulators are frequently mutated in different types of cancer (Ma et al., 2012), suggesting that they play key roles in limiting cellular proliferation. As a specific example, subunits of BAF complex are mutated in more than 20 percent of all human cancers (Kadoch et al., 2013). So far, it is still not clear how different mutations affect the functions of chromatin regulators or cancer progression *in vivo* within tumors. MNase-seq provides a high-resolution snapshot of nucleosome occupancy and is often used to study the roles of nucleosome remodeling factors. However, MNase-seq requires at least a few

hundred million sequencing reads to get good depth of coverage for each experimental condition in mammals, which is quite expensive for high-throughput screening. Alternatively, RED-seq requires fewer reads than MNase-seq and usually 10-20X coverage is sufficient for RED-seq analysis. Specifically, if we use a restriction enzyme with 1-3 million restriction sites in the genome, 10-60 million sequencing reads will provide sufficient coverage, making RED-seq an ideal and cost-effective tool to probe chromatin accessibility in dozens of tumors with mutations in nucleosome remodeling factors, or within other high-throughput settings. Additionally, future studies to understand the effect of different mutations in chromatin regulators could be achieved by combining high-throughput targeted mutagenesis using CRISPR/Cas9 (Findlay et al., 2014) and RED-seq as the readout for chromatin accessibility in mutated cells. In addition to targeted mutagenesis approaches, RED-seq could also be used for the screening of small molecular inhibitors targeting to chromatin regulators. Overall, these experiments will significantly contribute to our understanding of how different mutations affect the biochemical function of chromatin regulators.

Regulation of Tip60-p400 by Hdac6

Hdac6 interacts with Tip60-p400 in both ESCs and NSCs, but not MEFs, suggesting that Tip60-p400 forms a distinct complex in stem cells. In addition, we observed loss of Tip60-p400 binding downstream of TSSs and misregulated Tip60-target genes after depletion of Hdac6, suggesting that Hdac6 is important for Tip60-mediated gene regulation by facilitating the binding of Tip60-p400 to its

target genes in ESCs. It remains to be determined whether other proteins substitute for Hdac6 in differentiated cells to recruit Tip60-p400 to different promoters, or if alternative mechanisms of recruitment (e.g., binding to histone modifications or interaction with sequence-specific DNA-binding proteins) are more important upon differentiation.

During embryonic development, cells have to respond to various environmental signals in order to properly differentiate. However, the mechanisms by which cells sense environmental signals and trigger downstream signaling pathways to regulate gene expression are not completely clear. The finding that nuclear Hdac6 is strongly decreased, along with a coincident increase in cytoplasmic Hdac6 upon differentiation of ESCs or NSCs, suggests the localization of Hdac6 is regulated by environmental signals during differentiation. Re-localization of Hdac6 to the cytoplasm during ESC and NSC differentiation could help establish a distinct set of Tip60-p400 targets in differentiated cells compared to stem cells, and regulate gene expression by altering the pattern of Tip60-p400 binding to these sites. Future work focusing on developmental regulation of Hdac6 re-localization and alterations in Tip60 binding will provide a better understanding how cells response to environmental cues during development.

Interestingly, we found that Hdac6 re-localizes to the nucleus of several cancer cell lines, whereas it is mainly in the cytoplasm of well-differentiated cells. These findings support the hypothesis that nuclear Hdac6 is important for the

function of Tip60-p400 during cancer development or progression. Therefore, if nuclear Hdac6 is necessary for cancer cell proliferation or viability, a new strategy for cancer therapy could be to block the signaling pathway or shuttling mechanism whereby Hdac6 is re-localized to the nucleus in cancer cells.

R-loops and chromatin regulators

LncRNAs have been shown to play multiple roles during development in many different organisms. Tip60-p400 was found to interact with several lncRNAs in ESCs and knockdown of those lncRNAs caused misregulation of Tip60-target genes (Guttman et al., 2011), suggesting that lncRNAs regulate the function of Tip60-p400. From our unbiased genome-wide approach using RIP-seq, we identified 608 lncRNAs associated with Tip60-p400, supporting the hypothesis that lncRNAs play diverse roles in regulating Tip60-p400. Recently, we found that knockdown of one Tip60-p400-interacting lncRNA, *lnc1563*, results in loss of Hdac6 from Tip60-p400 complex (data not shown), raising the possibility that lncRNAs regulate the composition, genomic targets, and/or function of Tip60-p400. Future work focusing on the roles of Tip60-p400 associated lncRNAs will shed further light on these possibilities.

Although we performed RIP-seq primarily to identify Tip60-p400 associated lncRNAs, these data led to the unexpected discovery of a new role for promoter-proximal R-loops in shaping chromatin architecture and controlling ESC differentiation. R-loops are necessary for Tip60-p400 binding to most target

genes, but inhibit binding of PRC2 to its targets in ESCs. However, it is still unknown how R-loops regulate the binding of chromatin regulators. R-loops could affect the binding of chromatin regulators either by changing chromatin structure around TSSs, since R-loops likely disfavor nucleosome formation, or by altering the binding affinity of chromatin regulators, since RNA:DNA hybrids are conformationally and chemically different from double-stranded DNA. Future work to map nucleosome occupancy before and after the removal of R-loops or to test the binding affinity of each complex to different substrates (i.e. DNA-DNA, DNA-RNA, ssRNA, ssDNA) will help dissect the role of R-loops in modulating chromatin regulators.

In order to identify R-loops throughout the genome of ESC, we modified a published DRIP-seq protocol to map only the RNA component of R-loops (called DRIP-RNA-seq). From our DRIP-RNA-seq data, we observed that R-loops are formed at both the TSS and TTS of many genes and observed very low background from RNaseH pre-treated DNA, indicating the specificity of this method. DRIP-RNA-seq not only maps the RNA fraction in R-loops regions but also provides strand information, which further confirms the specificity of the observed DRIP peaks. The secondary structure of lncRNAs has been shown to act as a scaffold for the interaction of chromatin regulators and chromatin. However, how lncRNAs interact with chromatin is still largely unknown, although lncRNAs have also been proposed to form R-loops (Chu et al., 2011). Therefore, DRIP-RNA-seq could potentially be used to identify which part of lncRNAs

interacts with chromatin. Since we only sequence the RNA component of the R-loop, additional modifications will be necessary to distinguish *trans* interactions by lncRNAs from interactions *in cis*. However, if one were to sequence both the DNA and RNA components of the DRIP IP, and identify potential *trans* sites of hybridization by comparing the signal enrichment from the DNA and RNA components of the DRIP IP in lncRNA loci, this method could potentially be very powerful. If the interaction between lncRNAs and chromatin is transient or unstable, a fixation step could be added before isolating genomic DNA from cells. These results would greatly contribute to our understanding of the mechanisms by which lncRNAs regulate gene expression.

BIBLIOGRAPHY

Abbott, D.W., Ivanova, V.S., Wang, X., Bonner, W.M., and Ausió, J. (2001). Characterization of the stability and folding of H2A.Z chromatin particles: implications for transcriptional activation. J. Biol. Chem. *276*, 41945–41949.

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

Akhmedov, A.T., and Lopez, B.S. (2000). Human 100-kDa homologous DNA-pairing protein is the splicing factor PSF and promotes DNA strand invasion. Nucleic Acids Res. *28*, 3022–3030.

Allis, C.D., Richman, R., Gorovsky, M.A., Ziegler, Y.S., Touchstone, B., Bradley, W.A., and Cook, R.G. (1986). hv1 is an evolutionarily conserved H2A variant that is preferentially associated with active genes. J. Biol. Chem. *261*, 1941–1948.

Almer, A., and Hörz, W. (1986). Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast. EMBO J. *5*, 2681–2687.

Aloia, L., Di Stefano, B., and Di Croce, L. (2013). Polycomb complexes in stem cells and embryonic development. Dev. Camb. Engl. *140*, 2525–2534.

Altaf, M., Auger, A., Covic, M., and Côté, J. (2009). Connection between histone H2A variants and chromatin remodeling complexes. Biochem. Cell Biol. Biochim. Biol. Cell. *87*, 35–50.

Altaf, M., Auger, A., Monnet-Saksouk, J., Brodeur, J., Piquet, S., Cramet, M., Bouchard, N., Lacoste, N., Utley, R.T., Gaudreau, L., et al. (2010). NuA4-dependent acetylation of nucleosomal histones H4 and H2A directly stimulates incorporation of H2A.Z by the SWR1 complex. J. Biol. Chem. *285*, 15966–15977.

Ambrosetti, D.C., Schöler, H.R., Dailey, L., and Basilico, C. (2000). Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. J. Biol. Chem. *275*, 23387–23397.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. *17*, 126–140.

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.

Baek, S.H., Ohgi, K.A., Rose, D.W., Koo, E.H., Glass, C.K., and Rosenfeld, M.G. (2002). Exchange of N-CoR corepressor and Tip60 coactivator complexes links gene expression by NF-kappaB and beta-amyloid precursor protein. Cell *110*, 55–67.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS One *3*, e3376.

Bao, Y., Konesky, K., Park, Y.-J., Rosu, S., Dyer, P.N., Rangasamy, D., Tremethick, D.J., Laybourn, P.J., and Luger, K. (2004). Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. EMBO J. *23*, 3314–3324.

Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. Genes Dev. *19*, 1635–1655.

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.

Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., and Young, R.A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. Genes Dev. *23*, 2484–2489.

Bonasio, R., and Shiekhattar, R. (2014). Regulation of transcription by long noncoding RNAs. Annu. Rev. Genet. *48*, 433–455.

Boyer, L.A., Latek, R.R., and Peterson, C.L. (2004). The SANT domain: a unique histone-tail-binding module? Nat. Rev. Mol. Cell Biol. *5*, 158–163.

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.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311–322.

Brady, M.E., Ozanne, D.M., Gaughan, L., Waite, I., Cook, S., Neal, D.E., and Robson, C.N. (1999). Tip60 is a nuclear hormone receptor coactivator. J. Biol. Chem. *274*, 17599–17604.

Briggs, S.D., Xiao, T., Sun, Z.-W., Caldwell, J.A., Shabanowitz, J., Hunt, D.F., Allis, C.D., and Strahl, B.D. (2002). Gene silencing: trans-histone regulatory pathway in chromatin. Nature *418*, 498.

Brinkman, A.B., Gu, H., Bartels, S.J.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., et al. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. *22*, 1128–1138.

Britton, S., Dernoncourt, E., Delteil, C., Froment, C., Schiltz, O., Salles, B., Frit, P., and Calsou, P. (2014). DNA damage triggers SAF-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal. Nucleic Acids Res. *42*, 9047–9062.

Brizuela, B.J., Elfring, L., Ballard, J., Tamkun, J.W., and Kennison, J.A. (1994). Genetic analysis of the brahma gene of Drosophila melanogaster and polytene chromosome subdivisions 72AB. Genetics *137*, 803–813.

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.

Butler, K.V., Kalin, J., Brochier, C., Vistoli, G., Langley, B., and Kozikowski, A.P. (2010). Rational design and simple chemistry yield a superior, neuroprotective HDAC6 inhibitor, tubastatin A. J. Am. Chem. Soc. *132*, 10842–10846.

Cai, Y., Jin, J., Tomomori-Sato, C., Sato, S., Sorokina, I., Parmely, T.J., Conaway, R.C., and Conaway, J.W. (2003). Identification of new subunits of the multiprotein mammalian TRRAP/TIP60-containing histone acetyltransferase complex. J. Biol. Chem. *278*, 42733–42736.

Cai, Y., Jin, J., Florens, L., Swanson, S.K., Kusch, T., Li, B., Workman, J.L., Washburn, M.P., Conaway, R.C., and Conaway, J.W. (2005). The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone acetyltransferase and SRCAP complexes. J. Biol. Chem. *280*, 13665–13670.

Cai, Y., Jin, J., Yao, T., Gottschalk, A.J., Swanson, S.K., Wu, S., Shi, Y., Washburn, M.P., Florens, L., Conaway, R.C., et al. (2007). YY1 functions with INO80 to activate transcription. Nat. Struct. Mol. Biol. *14*, 872–874.

Cairns, B.R., Kim, Y.J., Sayre, M.H., Laurent, B.C., and Kornberg, R.D. (1994). A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. Proc. Natl. Acad. Sci. U. S. A. *91*, 1950–1954.

Cao, K., Lailler, N., Zhang, Y., Kumar, A., Uppal, K., Liu, Z., Lee, E.K., Wu, H., Medrzycki, M., Pan, C., et al. (2013). High-resolution mapping of h1 linker histone variants in embryonic stem cells. PLoS Genet. *9*, e1003417.

Carone, B.R., Hung, J.-H., Hainer, S.J., Chou, M.-T., Carone, D.M., Weng, Z., Fazzio, T.G., and Rando, O.J. (2014). High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. Dev. Cell *30*, 11–22.

Castel, S.E., Ren, J., Bhattacharjee, S., Chang, A.-Y., Sánchez, M., Valbuena, A., Antequera, F., and Martienssen, R.A. (2014). Dicer promotes transcription termination at sites of replication stress to maintain genome stability. Cell *159*, 572–583.

Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L.M., Krijgsveld, J., and Hentze, M.W. (2013). System-wide identification of RNA-binding proteins by interactome capture. Nat. Protoc. *8*, 491–500.

Cavalli, G., and Paro, R. (1998). The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell *93*, 505–518.

Chalkley, R.J., Baker, P.R., Huang, L., Hansen, K.C., Allen, N.P., Rexach, M., and Burlingame, A.L. (2005). Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: II. New developments in Protein Prospector allow for reliable and comprehensive automatic analysis of large datasets. Mol. Cell. Proteomics MCP *4*, 1194–1204.

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.

Chandrasekharan, M.B., Huang, F., and Sun, Z.-W. (2010). Histone H2B ubiquitination and beyond: Regulation of nucleosome stability, chromatin dynamics and the trans-histone H3 methylation. Epigenetics Off. J. DNA Methylation Soc. *5*, 460–468.

Chen, P.B., Hung, J.-H., Hickman, T.L., Coles, A.H., Carey, J.F., Weng, Z., Chu, F., and Fazzio, T.G. (2013). Hdac6 regulates Tip60-p400 function in stem cells. eLife *2*, e01557.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science *325*, 834–840.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol. Cell *44*, 667–678.

Chu, F., Nusinow, D.A., Chalkley, R.J., Plath, K., Panning, B., and Burlingame, A.L. (2006). Mapping post-translational modifications of the histone variant MacroH2A1 using tandem mass spectrometry. Mol. Cell. Proteomics MCP *5*, 194–203.

Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory interactions between RNA and polycomb repressive complex 2. Mol. Cell *55*, 171–185.

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

Coles, A.H., Liang, H., Zhu, Z., Marfella, C.G.A., Kang, J., Imbalzano, A.N., and Jones, S.N. (2007). Deletion of p37Ing1 in mice reveals a p53-independent role for Ing1 in the suppression of cell proliferation, apoptosis, and tumorigenesis. Cancer Res. *67*, 2054–2061.

Constantinescu, D., Gray, H.L., Sammak, P.J., Schatten, G.P., and Csoka, A.B. (2006). Lamin A/C expression is a marker of mouse and human embryonic stem cell differentiation. Stem Cells Dayt. Ohio *24*, 177–185.

Côté, J., Quinn, J., Workman, J.L., and Peterson, C.L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science *265*, 53–60.

Crawford, G.E., Davis, S., Scacheri, P.C., Renaud, G., Halawi, M.J., Erdos, M.R., Green, R., Meltzer, P.S., Wolfsberg, T.G., and Collins, F.S. (2006). DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nat. Methods *3*, 503–509.

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.

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.-Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res. *19*, 24–32.

van Daal, A., and Elgin, S.C. (1992). A histone variant, H2AvD, is essential in Drosophila melanogaster. Mol. Biol. Cell *3*, 593–602.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. Nat. Struct. Mol. Biol. *20*, 1250–1257.

Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K.J., Gooding, A.R., Lee, J.T., and Cech, T.R. (2015). Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. Mol. Cell *57*, 552–558.

Davie, J.R., and Saunders, C.A. (1981). Chemical composition of nucleosomes among domains of calf thymus chromatin differing in micrococcal nuclease accessibility and solubility properties. J. Biol. Chem. *256*, 12574–12580.

Deuring, R., Fanti, L., Armstrong, J.A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S.L., Berloco, M., et al. (2000). The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. Mol. Cell *5*, 355–365.

Dirscherl, S.S., and Krebs, J.E. (2004). Functional diversity of ISWI complexes. Biochem. Cell Biol. Biochim. Biol. Cell. *82*, 482–489.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature *330*, 576–578.

Dover, J., Schneider, J., Tawiah-Boateng, M.A., Wood, A., Dean, K., Johnston, M., and Shilatifard, A. (2002). Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J. Biol. Chem. *277*, 28368–28371.

Dovey, O.M., Foster, C.T., and Cowley, S.M. (2010). Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. Proc. Natl. Acad. Sci. U. S. A. *107*, 8242–8247.

Doyen, C.-M., Montel, F., Gautier, T., Menoni, H., Claudet, C., Delacour-Larose, M., Angelov, D., Hamiche, A., Bednar, J., Faivre-Moskalenko, C., et al. (2006). Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. EMBO J. *25*, 4234–4244.

Doyon, Y., Selleck, W., Lane, W.S., Tan, S., and Côté, J. (2004). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Mol. Cell. Biol. *24*, 1884–1896.

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.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. (2010). Conserved nucleosome positioning defines replication origins. Genes Dev. *24*, 748–753.

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.

Elgin, S.C.R., and Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. Cold Spring Harb. Perspect. Biol. *5*, a017780.

Faast, R., Thonglairoam, V., Schulz, T.C., Beall, J., Wells, J.R., Taylor, H., Matthaei, K., Rathjen, P.D., Tremethick, D.J., and Lyons, I. (2001). Histone variant H2A.Z is required for early mammalian development. Curr. Biol. CB *11*, 1183–1187.

Fan, J.Y., Rangasamy, D., Luger, K., and Tremethick, D.J. (2004). H2A.Z alters the nucleosome surface to promote HP1alpha-mediated chromatin fiber folding. Mol. Cell *16*, 655–661.

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., and Tsukiyama, T. (2003). Chromatin remodeling in vivo: evidence for a nucleosome sliding mechanism. Mol. Cell *12*, 1333–1340.

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

Fazzio, T.G., Huff, J.T., and Panning, B. (2008b). Chromatin regulation Tip(60)s the balance in embryonic stem cell self-renewal. Cell Cycle Georget. Tex *7*, 3302–3306.

Fedoriw, A.M., Stein, P., Svoboda, P., Schultz, R.M., and Bartolomei, M.S. (2004). Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science *303*, 238–240.

Felsenfeld, G. (1992). Chromatin as an essential part of the transcriptional mechanism. Nature *355*, 219–224.

Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., and Klug, A. (1977). Structure of nucleosome core particles of chromatin. Nature *269*, 29–36.

Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C., and Shendure, J. (2014). Saturation editing of genomic regions by multiplex homology-directed repair. Nature *513*, 120–123.

Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116–1122.

Flanagan, J.F., Mi, L.-Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature *438*, 1181–1185.

Flynn, R.A., and Chang, H.Y. (2014). Long noncoding RNAs in cell-fate programming and reprogramming. Cell Stem Cell *14*, 752–761.

Frank, S.R., Parisi, T., Taubert, S., Fernandez, P., Fuchs, M., Chan, H.-M., Livingston, D.M., and Amati, B. (2003). MYC recruits the TIP60 histone acetyltransferase complex to chromatin. EMBO Rep. *4*, 575–580.

Fu, Y., Sinha, M., Peterson, C.L., and Weng, Z. (2008). The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. PLoS Genet. *4*, e1000138.

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.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol. Cell *45*, 344–356.

Gargiulo, G., Levy, S., Bucci, G., Romanenghi, M., Fornasari, L., Beeson, K.Y., Goldberg, S.M., Cesaroni, M., Ballarini, M., Santoro, F., et al. (2009). NA-Seq: a discovery tool for the analysis of chromatin structure and dynamics during differentiation. Dev. Cell *16*, 466–481.

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.

Gendrel, A.-V., and Heard, E. (2014). Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. Annu. Rev. Cell Dev. Biol. *30*, 561–580.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol. Cell *45*, 814–825.

Ginno, P.A., Lim, Y.W., Lott, P.L., Korf, I., and Chédin, F. (2013). GC skew at the 5’ and 3’ ends of human genes links R-loop formation to epigenetic regulation and transcription termination. Genome Res. *23*, 1590–1600.

Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., and Lieb, J.D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. *17*, 877–885.

Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., Workman, J.L., Barton, G.J., and Owen-Hughes, T. (2011). A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. Science *333*, 1758–1760.

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.

Goldmark, J.P., Fazzio, T.G., Estep, P.W., Church, G.M., and Tsukiyama, T. (2000). The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. Cell *103*, 423–433.

Groh, M., Lufino, M.M.P., Wade-Martins, R., and Gromak, N. (2014). R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet. *10*, e1004318.

Guan, S., and Burlingame, A.L. (2010). Data processing algorithms for analysis of high resolution MSMS spectra of peptides with complex patterns of posttranslational modifications. Mol. Cell. Proteomics MCP *9*, 804–810.

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339–346.

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.

Haggarty, S.J., Koeller, K.M., Wong, J.C., Grozinger, C.M., and Schreiber, S.L. (2003). Domain-selective small-molecule inhibitor of histone deacetylase 6 (HDAC6)-mediated tubulin deacetylation. Proc. Natl. Acad. Sci. U. S. A. *100*, 4389–4394.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Res. *23*, 1210–1223.

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.

Hall, J.A., and Georgel, P.T. (2007). CHD proteins: a diverse family with strong ties. Biochem. Cell Biol. Biochim. Biol. Cell. *85*, 463–476.

Hamiche, A., Sandaltzopoulos, R., Gdula, D.A., and Wu, C. (1999). ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. Cell *97*, 833–842.

Hanna, J.H., Saha, K., and Jaenisch, R. (2010). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. Cell *143*, 508–525.

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

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.

Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Mol. Cell *44*, 966–977.

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.

Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. Proc. Natl. Acad. Sci. U. S. A. *108*, 18318–18323.

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.

Herrick, G., and Seger, J. (1999). Imprinting and paternal genome elimination in insects. Results Probl. Cell Differ. *25*, 41–71.

Hirose, T., Virnicchi, G., Tanigawa, A., Naganuma, T., Li, R., Kimura, H., Yokoi, T., Nakagawa, S., Bénard, M., Fox, A.H., et al. (2014). NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. Mol. Biol. Cell *25*, 169–183.

Hirschhorn, J.N., Brown, S.A., Clark, C.D., and Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. Genes Dev. *6*, 2288–2298.

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

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. (2009a). 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.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crabtree, G.R. (2009b). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc. Natl. Acad. Sci. U. S. A. *106*, 5187–5191.

Hooper, M., Hardy, K., Handyside, A., Hunter, S., and Monk, M. (1987). HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature *326*, 292–295.

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, Y., Fisher, J.B., Koprowski, S., McAllister, D., Kim, M.-S., and Lough, J. (2009). Homozygous disruption of the Tip60 gene causes early embryonic lethality. Dev. Dyn. Off. Publ. Am. Assoc. Anat. *238*, 2912–2921.

Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2007a). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. *8*, R183.

Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C., et al. (2007b). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. *35*, W169–W175.

Hubbert, C., Guardiola, A., Shao, R., Kawaguchi, Y., Ito, A., Nixon, A., Yoshida, M., Wang, X.-F., and Yao, T.-P. (2002). HDAC6 is a microtubule-associated deacetylase. Nature *417*, 455–458.

Huertas, P., and Aguilera, A. (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell 12, 711–721.

Ikura, T., Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., and Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell *102*, 463–473.

Imamura, K., Imamachi, N., Akizuki, G., Kumakura, M., Kawaguchi, A., Nagata, K., Kato, A., Kawaguchi, Y., Sato, H., Yoneda, M., et al. (2014). Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. Mol. Cell *53*, 393–406.

Imbalzano, A.N., Kwon, H., Green, M.R., and Kingston, R.E. (1994). Facilitated binding of TATA-binding protein to nucleosomal DNA. Nature *370*, 481–485.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R., and Weissman, J.S. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. Cell Rep. *8*, 1365–1379.

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

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133.

Jiang, H., Shukla, A., Wang, X., Chen, W., 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 Y.H., Yoo E.J., Jang Y.K., Kim S.H., Kim M.J., Shim Y.S., Lee J.S., Choi I.S., Seong R.H., Hong S.H., Park S.D. (1998). Isolation and characterization of hrp1+, a new member of the SNF2/SWI2 gene family from the fission yeast Schizosaccharomyces pombe. Mol Gen Genet. 257(3):319-29.

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.

Jónsson, Z.O., Jha, S., Wohlschlegel, J.A., and Dutta, A. (2004). Rvb1p/Rvb2p recruit Arp5p and assemble a functional Ino80 chromatin remodeling complex. Mol. Cell *16*, 465–477.

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat. Genet. *45*, 592–601.

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.

Kamakaka, R.T., and Biggins, S. (2005). Histone variants: deviants? Genes Dev. *19*, 295–310.

Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1258–1264.

Kaneko, S., Son, J., Bonasio, R., Shen, S.S., and Reinberg, D. (2014). Nascent RNA interaction keeps PRC2 activity poised and in check. Genes Dev. *28*, 1983–1988.

Kanhere, A., Viiri, K., Araújo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol. Cell *38*, 675–688.

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.

Karantzali, E., Schulz, H., Hummel, O., Hubner, N., Hatzopoulos, A., and Kretsovali, A. (2008). Histone deacetylase inhibition accelerates the early events of stem cell differentiation: transcriptomic and epigenetic analysis. Genome Biol. *9*, R65.

Kawaguchi, Y., Kovacs, J.J., McLaurin, A., Vance, J.M., Ito, A., and Yao, T.P. (2003). The deacetylase HDAC6 regulates aggresome formation and cell viability in response to misfolded protein stress. Cell *115*, 727–738.

Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. Proc. Natl. Acad. Sci. U. S. A. *85*, 8136–8140.

Kent, N.A., Adams, S., Moorhouse, A., and Paszkiewicz, K. (2011). Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing. Nucleic Acids Res. *39*, e26.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. U. S. A. *106*, 11667–11672.

Khavari, P.A., Peterson, C.L., Tamkun, J.W., Mendel, D.B., and Crabtree, G.R. (1993). BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. Nature *366*, 170–174.

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, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.R., et al. (2010a). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285–290.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010b). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kimura, A., and Horikoshi, M. (1998). Tip60 acetylates six lysines of a specific class in core histones in vitro. Genes Cells Devoted Mol. Cell. Mech. *3*, 789–800.

Klymenko, T., and Müller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. EMBO Rep. *5*, 373–377.

Kobor, M.S., Venkatasubrahmanyam, S., Meneghini, M.D., Gin, J.W., Jennings, J.L., Link, A.J., Madhani, H.D., and Rine, J. (2004). A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. PLoS Biol. *2*, E131.

Kornberg, R.D., and Lorch, Y. (1992). Chromatin structure and transcription. Annu. Rev. Cell Biol. *8*, 563–587.

Kovacs, J.J., Murphy, P.J.M., Gaillard, S., Zhao, X., Wu, J.-T., Nicchitta, C.V., Yoshida, M., Toft, D.O., Pratt, W.B., and Yao, T.-P. (2005). HDAC6 regulates Hsp90 acetylation and chaperone-dependent activation of glucocorticoid receptor. Mol. Cell *18*, 601–607.

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.

Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc. Natl. Acad. Sci. U. S. A. *103*, 10684–10689.

Kusch, T., Florens, L., Macdonald, W.H., Swanson, S.K., Glaser, R.L., Yates, J.R., Abmayr, S.M., Washburn, M.P., and Workman, J.L. (2004). Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. Science *306*, 2084–2087.

Kwon, H., Imbalzano, A.N., Khavari, P.A., Kingston, R.E., and Green, M.R. (1994). Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. Nature *370*, 477–481.

Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1122–1130.

Lam, M.T.Y., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., et al. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature *498*, 511–515.

Lam, M.T.Y., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends Biochem. Sci. *39*, 170–182.

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.

Längst, G., Bonte, E.J., Corona, D.F., and Becker, P.B. (1999). Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. Cell *97*, 843–852.

Laurent, B.C., Treich, I., and Carlson, M. (1993). Role of yeast SNF and SWI proteins in transcriptional activation. Cold Spring Harb. Symp. Quant. Biol. *58*, 257–263.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220.

Leach, T.J., Mazzeo, M., Chotkowski, H.L., Madigan, J.P., Wotring, M.G., and Glaser, R.L. (2000). Histone H2A.Z is widely but nonrandomly distributed in chromosomes of Drosophila melanogaster. J. Biol. Chem. *275*, 23267–23272.

Legube, G., Linares, L.K., Tyteca, S., Caron, C., Scheffner, M., Chevillard-Briet, M., and Trouche, D. (2004). Role of the histone acetyl transferase Tip60 in the p53 pathway. J. Biol. Chem. *279*, 44825–44833.

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.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell *122*, 365–378.

Li, W., Nagaraja, S., Delcuve, G.P., Hendzel, M.J., and Davie, J.R. (1993). Effects of histone acetylation, ubiquitination and variants on nucleosome stability. Biochem. J. *296 (Pt 3)*, 737–744.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516–520.

Li, X., Barkho, B.Z., Luo, Y., Smrt, R.D., Santistevan, N.J., Liu, C., Kuwabara, T., Gage, F.H., and Zhao, X. (2008). Epigenetic regulation of the stem cell mitogen Fgf-2 by Mbd1 in adult neural stem/progenitor cells. J. Biol. Chem. *283*, 27644–27652.

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.

Liberator, P.A., and Lingrel, J.B. (1984). Restriction endonuclease accessibility of the developmentally regulated goat gamma-, beta C-, and beta A-globin genes in chromatin. Differences in 5’ regions which show unusually high sequence homology. J. Biol. Chem. *259*, 15497–15501.

Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature *430*, 471–476.

Lo, W.S., Trievel, R.C., Rojas, J.R., Duggan, L., Hsu, J.Y., Allis, C.D., Marmorstein, R., and Berger, S.L. (2000). Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. Mol. Cell *5*, 917–926.

Logie, C., and Peterson, C.L. (1997). Catalytic activity of the yeast SWI/SNF complex on reconstituted nucleosome arrays. EMBO J. *16*, 6772–6782.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. *38*, 431–440.

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.

Luger, K., and Richmond, T.J. (1998). DNA binding within the nucleosome core. Curr. Opin. Struct. Biol. *8*, 33–40.

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.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat. Rev. Mol. Cell Biol. *13*, 436–447.

Lynch, M.D., Smith, A.J.H., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L., et al. (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J. *31*, 317–329.

Ma, Q.C., Ennis, C.A., and Aparicio, S. (2012). Opening Pandora's Box--the new biology of driver mutations and clonal evolution in cancer as revealed by next generation sequencing. Curr. Opin. Genet. Dev. *22*, 3–9.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

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.

Martens, J.A., and Winston, F. (2002). Evidence that Swi/Snf directly represses transcription in S. cerevisiae. Genes Dev. *16*, 2231–2236.

Martienssen, R.A., and Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science *293*, 1070–1074.

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.

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.

Matias, P.M., Gorynia, S., Donner, P., and Carrondo, M.A. (2006). Crystal structure of the human AAA+ protein RuvBL1. J. Biol. Chem. *281*, 38918–38929.

McBurney, M.W., Yang, X., Jardine, K., Bieman, M., Th'ng, J., and Lemieux, M. (2003). The absence of SIR2alpha protein has no effect on global gene silencing in mouse embryonic stem cells. Mol. Cancer Res. MCR *1*, 402–409.

McClelland, M. (1981). The effect of sequence specific DNA methylation on restriction endonuclease cleavage. Nucleic Acids Res. *9*, 5859–5866.

McCool, K.W., Xu, X., Singer, D.B., Murdoch, F.E., and Fritsch, M.K. (2007). The role of histone acetylation in regulating early gene expression patterns during early embryonic stem cell differentiation. J. Biol. Chem. *282*, 6696–6706.

McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. *6*, e1001244.

Meneghini, M.D., Wu, M., and Madhani, H.D. (2003). Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. Cell *112*, 725–736.

Meshorer, E. (2007). Chromatin in embryonic stem cell neuronal differentiation. Histol. Histopathol. *22*, 311–319.

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

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev. Cell *10*, 105–116.

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., and Johnson, E.A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. *17*, 240–248.

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., Tsukiyama, T., Wisniewski, J., and Wu, C. (1997). Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. Mol. Cell *1*, 141–150.

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.

Mohrmann, L., and Verrijzer, C.P. (2005). Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. Biochim. Biophys. Acta *1681*, 59–73.

Mouse ENCODE Consortium, Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaul, R., et al. (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol. *13*, 418.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science *322*, 1717–1720.

Narlikar, G.J., Phelan, M.L., and Kingston, R.E. (2001). Generation and interconversion of multiple distinct nucleosomal states as a mechanism for catalyzing chromatin fluidity. Mol. Cell *8*, 1219–1230.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2002). Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex. Genes Dev. *16*, 806–819.

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.

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.

Ohkawa, Y., Marfella, C.G.A., and Imbalzano, A.N. (2006). Skeletal muscle specification by myogenin and Mef2D via the SWI/SNF ATPase Brg1. EMBO J. *25*, 490–501.

Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Marcarci, B.O., Unternaehrer, J., Gupta, P.B., et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598–602.

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

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

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell *32*, 232–246.

Papamichos-Chronakis, M., Krebs, J.E., and Peterson, C.L. (2006). Interplay between Ino80 and Swr1 chromatin remodeling enzymes regulates cell cycle checkpoint adaptation in response to DNA damage. Genes Dev. *20*, 2437–2449.

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.

Park, S.-H., Park, S.H., Kook, M.-C., Kim, E.-Y., Park, S., and Lim, J.H. (2004). Ultrastructure of human embryonic stem cells and spontaneous and retinoic acid-induced differentiating cells. Ultrastruct. Pathol. *28*, 229–238.

Patton, J.G., Mayer, S.A., Tempst, P., and Nadal-Ginard, B. (1991). Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. Genes Dev. *5*, 1237–1251.

Paul, J., and Duerksen, J.D. (1975). Chromatin-associated RNA content of heterochromatin and euchromatin. Mol. Cell. Biochem. *9*, 9–16.

Pauli, A., Rinn, J.L., and Schier, A.F. (2011). Non-coding RNAs as regulators of embryogenesis. Nat. Rev. Genet. *12*, 136–149.

Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nat. Rev. Genet. *6*, 893–904.

Pervouchine, D.D., Djebali, S., Breschi, A., Davis, C.A., Barja, P.P., Dobin, A., Tanzer, A., Lagarde, J., Zaleski, C., See, L.-H., et al. (2015). Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. Nat. Commun. *6*, 5903.

Peterson, C.L., and Herskowitz, I. (1992). Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. Cell *68*, 573–583.

Peterson, C.L., and Tamkun, J.W. (1995). The SWI-SNF complex: a chromatin remodeling machine? Trends Biochem. Sci. *20*, 143–146.

Pfeiffer, W., and Zachau, H.G. (1980). Accessibility of expressed and non-expressed genes to a restriction nuclease. Nucleic Acids Res. *8*, 4621–4638.

Pray-Grant, M.G., Daniel, J.A., Schieltz, D., Yates, J.R., and Grant, P.A. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature *433*, 434–438.

Price, B.D., and Andrea, A.D. D' (2013). Chromatin remodeling at DNA double-strand breaks. Cell *152*, 1344–1354.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., et al. (2013). 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.

Ramírez, J., and Hagman, J. (2009). The Mi-2/NuRD complex: a critical epigenetic regulator of hematopoietic development, differentiation and cancer. Epigenetics Off. J. DNA Methylation Soc. *4*, 532–536.

Rangasamy, D., Berven, L., Ridgway, P., and Tremethick, D.J. (2003). Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. EMBO J. *22*, 1599–1607.

Rangasamy, D., Greaves, I., and Tremethick, D.J. (2004). RNA interference demonstrates a novel role for H2A.Z in chromosome segregation. Nat. Struct. Mol. Biol. *11*, 650–655.

Reyes, J.C., Barra, J., Muchardt, C., Camus, A., Babinet, C., and Yaniv, M. (1998). Altered control of cellular proliferation in the absence of mammalian brahma (SNF2alpha). EMBO J. *17*, 6979–6991.

Richmond, T.J., and Davey, C.A. (2003). The structure of DNA in the nucleosome core. Nature *423*, 145–150.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol. Cell *55*, 347–360.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu. Rev. Genet. *38*, 413–443.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Ruiz-Orera, J., Messeguer, X., Subirana, J.A., and Alba, M.M. (2014). Long non-coding RNAs as a source of new peptides. eLife *3*, e03523.

Sabo, P.J., Kuehn, M.S., Thurman, R., Johnson, B.E., Johnson, E.M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., et al. (2006). Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. Nat. Methods *3*, 511–518.

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.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.-L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. *8*, e1000384.

Santenard, A., Ziegler-Birling, C., Koch, M., Tora, L., Bannister, A.J., and Torres-Padilla, M.-E. (2010). Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. Nat. Cell Biol. *12*, 853–862.

Saragosti, S., Moyne, G., and Yaniv, M. (1980). Absence of nucleosomes in a fraction of SV40 chromatin between the origin of replication and the region coding for the late leader RNA. Cell *20*, 65–73.

Sarma, K., and Reinberg, D. (2005). Histone variants meet their match. Nat. Rev. Mol. Cell Biol. *6*, 139–149.

Schirling, C., Heseding, C., Heise, F., Kesper, D., Klebes, A., Klein-Hitpass, L., Vortkamp, A., Hoffmann, D., Saumweber, H., and Ehrenhofer-Murray, A.E. (2010). Widespread regulation of gene expression in the Drosophila genome by the histone acetyltransferase dTip60. Chromosoma *119*, 99–113.

Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev. *19*, 697–708.

Schöler, H.R., Ruppert, S., Suzuki, N., Chowdhury, K., and Gruss, P. (1990). New type of POU domain in germ line-specific protein Oct-4. Nature *344*, 435–439.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887–898.

Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G. (1998). Parent-of-origin effects on seed development in Arabidopsis thaliana. Dev. Camb. Engl. *125*, 3329–3341.

Shen, X., Mizuguchi, G., Hamiche, A., and Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. Nature *406*, 541–544.

Shen, X., Xiao, H., Ranallo, R., Wu, W.-H., and Wu, C. (2003). Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates. Science *299*, 112–114.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M.R., et al. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96–99.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722–737.

Simic, R., Lindstrom, D.L., Tran, H.G., Roinick, K.L., Costa, P.J., Johnson, A.D., Hartzog, G.A., and Arndt, K.M. (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. EMBO J. *22*, 1846–1856.

Sims, R.J., Chen, C.-F., Santos-Rosa, H., Kouzarides, T., Patel, S.S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. J. Biol. Chem. *280*, 41789–41792.

Skene, P.J., and Henikoff, S. (2013). Histone variants in pluripotency and disease. Dev. Camb. Engl. *140*, 2513–2524.

Skene, P.J., Hernandez, A.E., Groudine, M., and Henikoff, S. (2014). The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. eLife *3*, e02042.

Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature *516*, 436–439.

Song, X., Sun, Y., and Garen, A. (2005). Roles of PSF protein and VL30 RNA in reversible gene regulation. Proc. Natl. Acad. Sci. U. S. A. *102*, 12189–12193.

Squatrito, M., Gorrini, C., and Amati, B. (2006). Tip60 in DNA damage response and growth control: many tricks in one HAT. Trends Cell Biol. *16*, 433–442.

Stargell, L.A., Bowen, J., Dadd, C.A., Dedon, P.C., Davis, M., Cook, R.G., Allis, C.D., and Gorovsky, M.A. (1993). Temporal and spatial association of histone H2A variant hv1 with transcriptionally competent chromatin during nuclear development in Tetrahymena thermophila. Genes Dev. *7*, 2641–2651.

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.

Sudarsanam, P., and Winston, F. (2000). The Swi/Snf family nucleosome-remodeling complexes and transcriptional control. Trends Genet. TIG *16*, 345–351.

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science *340*, 619–621.

Swaminathan, J., Baxter, E.M., and Corces, V.G. (2005). The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of Drosophila heterochromatin. Genes Dev. *19*, 65–76.

Szabó, P.E., Tang, S.-H.E., Silva, F.J., Tsark, W.M.K., and Mann, J.R. (2004). Role of CTCF binding sites in the Igf2/H19 imprinting control region. Mol. Cell. Biol. *24*, 4791–4800.

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.

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.

Tamkun, J.W. (1995). The role of brahma and related proteins in transcription and development. Curr. Opin. Genet. Dev. *5*, 473–477.

Tamkun, J.W., Deuring, R., Scott, M.P., Kissinger, M., Pattatucci, A.M., Kaufman, T.C., and Kennison, J.A. (1992). brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. Cell *68*, 561–572.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., et al. (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell *148*, 664–678.

Thambirajah, A.A., Dryhurst, D., Ishibashi, T., Li, A., Maffey, A.H., and Ausió, J. (2006). H2A.Z stabilizes chromatin in a way that is dependent on core histone acetylation. J. Biol. Chem. *281*, 20036–20044.

Thoma, F., Koller, T., and Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J. Cell Biol. *83*, 403–427.

Tolstorukov, M.Y., Kharchenko, P.V., Goldman, J.A., Kingston, R.E., and Park, P.J. (2009). Comparative analysis of H2A.Z nucleosome organization in the human and yeast genomes. Genome Res. *19*, 967–977.

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.

Tran, H.G., Steger, D.J., Iyer, V.R., and Johnson, A.D. (2000). The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor. EMBO J. *19*, 2323–2331.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689–693.

Tsukiyama, T., and Wu, C. (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell *83*, 1011–1020.

Tsukiyama, T., Daniel, C., Tamkun, J., and Wu, C. (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. Cell *83*, 1021–1026.

Tsukiyama T., Palmer J., Landel C.C., Shiloach J., Wu C. (1999). Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in Saccharomyces cerevisiae. Genes Dev. 13(6):686-97.

Tsukuda, T., Fleming, A.B., Nickoloff, J.A., and Osley, M.A. (2005). Chromatin remodelling at a DNA double-strand break site in Saccharomyces cerevisiae. Nature *438*, 379–383.

Ueda, T., Watanabe-Fukunaga, R., Ogawa, H., Fukuyama, H., Higashi, Y., Nagata, S., and Fukunaga, R. (2007). Critical role of the p400/mDomino chromatin-remodeling ATPase in embryonic hematopoiesis. Genes Cells Devoted Mol. Cell. Mech. *12*, 581–592.

Valenzuela-Fernández, A., Cabrero, J.R., Serrador, J.M., and Sánchez-Madrid, F. (2008). HDAC6: a key regulator of cytoskeleton, cell migration and cell-cell interactions. Trends Cell Biol. *18*, 291–297.

Verdel, A., Curtet, S., Brocard, M.P., Rousseaux, S., Lemercier, C., Yoshida, M., and Khochbin, S. (2000). Active maintenance of mHDA2/mHDAC6 histone-deacetylase in the cytoplasm. Curr. Biol. CB *10*, 747–749.

Waki, H., Nakamura, M., Yamauchi, T., Wakabayashi, K., Yu, J., Hirose-Yotsuya, L., Take, K., Sun, W., Iwabu, M., Okada-Iwabu, M., et al. (2011). Global mapping of cell type-specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation. PLoS Genet. *7*, e1002311.

Wallberg, A.E., Neely, K.E., Hassan, A.H., Gustafsson, J.A., Workman, J.L., and Wright, A.P. (2000). Recruitment of the SWI-SNF chromatin remodeling complex as a mechanism of gene activation by the glucocorticoid receptor tau1 activation domain. Mol. Cell. Biol. *20*, 2004–2013.

Wang, X., and Hayes, J.J. (2008). Acetylation mimics within individual core histone tail domains indicate distinct roles in regulating the stability of higher-order chromatin structure. Mol. Cell. Biol. *28*, 227–236.

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., 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, W., Côté, J., Xue, Y., Zhou, S., Khavari, P.A., Biggar, S.R., Muchardt, C., Kalpana, G.V., Goff, S.P., Yaniv, M., et al. (1996). Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. EMBO J. *15*, 5370–5382.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell *138*, 1019–1031.

Wassenegger, M., Heimes, S., Riedel, L., and Sänger, H.L. (1994). RNA-directed de novo methylation of genomic sequences in plants. Cell *76*, 567–576.

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.

Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science *193*, 848–856.

Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. Nature *450*, 1031–1035.

Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S. (1997). Characterization of the CHD family of proteins. Proc Natl Acad Sci U S A. 94(21):11472-7.

Wong, L.H., McGhie, J.D., Sim, M., Anderson, M.A., Ahn, S., Hannan, R.D., George, A.J., Morgan, K.A., Mann, J.R., and Choo, K.H.A. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res. *20*, 351–360.

Wong, M.M., Cox, L.K., and Chrivia, J.C. (2007). The chromatin remodeling protein, SRCAP, is critical for deposition of the histone variant H2A.Z at promoters. J. Biol. Chem. *282*, 26132–26139.

Wood, A.J., and Oakey, R.J. (2006). Genomic imprinting in mammals: emerging themes and established theories. PLoS Genet. *2*, e147.

Wu, C. (1980). The 5’ ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature *286*, 854–860.

Xi, H., Shulha, H.P., Lin, J.M., Vales, T.R., Fu, Y., Bodine, D.M., McKay, R.D.G., Chenoweth, J.G., Tesar, P.J., Furey, T.S., et al. (2007). Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. PLoS Genet. *3*, e136.

Xu, Y., and Price, B.D. (2011). Chromatin dynamics and the repair of DNA double strand breaks. Cell Cycle Georget. Tex *10*, 261–267.

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.

Yamamoto, T., and Horikoshi, M. (1997). Novel substrate specificity of the histone acetyltransferase activity of HIV-1-Tat interactive protein Tip60. J. Biol. Chem. *272*, 30595–30598.

Yan, Y., Barlev, N.A., Haley, R.H., Berger, S.L., and Marmorstein, R. (2000). Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases. Mol. Cell *6*, 1195–1205.

Yang, D., Buchholz, F., Huang, Z., Goga, A., Chen, C.-Y., Brodsky, F.M., and Bishop, J.M. (2002). Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. Proc. Natl. Acad. Sci. U. S. A. *99*, 9942–9947.

Yeap, L.-S., Hayashi, K., and Surani, M.A. (2009). ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage. Epigenetics Chromatin *2*, 12.

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.

Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat. Biotechnol. *21*, 183–186.

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.

Yoo, A.S., Staahl, B.T., Chen, L., and Crabtree, G.R. (2009). MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. Nature *460*, 642–646.

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

Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626–630.

Yuan, P., Han, J., Guo, G., Orlov, Y.L., Huss, M., Loh, Y.-H., Yaw, L.-P., Robson, P., Lim, B., and Ng, H.-H. (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes Dev. *23*, 2507–2520.

Yudkovsky, N., Logie, C., Hahn, S., and Peterson, C.L. (1999). Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators. Genes Dev. *13*, 2369–2374.

Zentner, G.E., Tsukiyama, T., and Henikoff, S. (2013). ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. PLoS Genet. *9*, e1003317.

Zhang, H., Roberts, D.N., and Cairns, B.R. (2005). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. Cell *123*, 219–231.

Zhang, X., Yuan, Z., Zhang, Y., Yong, S., Salas-Burgos, A., Koomen, J., Olashaw, N., Parsons, J.T., Yang, X.-J., Dent, S.R., et al. (2007). HDAC6 modulates cell motility by altering the acetylation level of cortactin. Mol. Cell *27*, 197–213.

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.

Zhang, Y., Li, N., Caron, C., Matthias, G., Hess, D., Khochbin, S., and Matthias, P. (2003). HDAC-6 interacts with and deacetylates tubulin and microtubules in vivo. EMBO J. *22*, 1168–1179.

Zhang, Y., Gilquin, B., Khochbin, S., and Matthias, P. (2006). Two catalytic domains are required for protein deacetylation. J. Biol. Chem. *281*, 2401–2404.

Zhang, Y., Kwon, S., Yamaguchi, T., Cubizolles, F., Rousseaux, S., Kneissel, M., Cao, C., Li, N., Cheng, H.-L., Chua, K., et al. (2008). Mice lacking histone deacetylase 6 have hyperacetylated tubulin but are viable and develop normally. Mol. Cell. Biol. *28*, 1688–1701.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750–756.

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237.