Roles for Chromatin Regulators During Differentiation of Embryonic Stem Cells

By

Warren A. Whyte
B.A., Molecular and Cell Biology
Cornell University, 2006

SUBMITTED TO THE DEPARTMENT OF BIOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 2012

Signature of Author

_____

Department of Biology
June 1, 2012

Certified by

_____

Dr. Richard A. Young
Professor of Biology
Thesis Supervisor

Accepted by

_____

Dr. Robert Sauer
Professor of Biology and
Chairperson, Biology Graduate Committee

Roles for Chromatin Regulators in Differentiation of Embryonic Stem Cells

by

Warren A. Whyte

Submitted to the Department of Biology on May 31st, 2012
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy in Biology

Abstract

Mammalian development involves the process by which a single fertilized egg develops into an adult with over 200 specialized cell types, each with a distinct gene expression pattern controlling its cellular state. As cells differentiate into specialized cell types, changes in the gene expression program occur with associated changes in chromatin. An understanding of the roles for chromatin regulators in the control of gene expression programs during differentiation is fundamental to understanding development. Although it is not yet feasible to elucidate the functions of all chromatin regulators in all vertebrate cells, recent work in embryonic stem (ES) cells has demonstrated that regulatory features of differentiation can be elucidated by focusing on the chromatin regulators involved in the changes in the pluripotent gene expression program as ES cells differentiate. New insights reveal that chromatin regulators of opposing functions share a common set of active genes in ES cells, suggesting a dynamic balance in the control of embryonic stem cell state and differentiation. I describe here the molecular mechanisms by which chromatin regulators contribute to the control of the ES cell state and differentiation, where these regulators play critical roles both in activating new gene expression programs and in silencing old programs.

Thesis Supervisor: Dr. Richard A. Young
Title: Professor of Biology

Dedication

To my loving parents Leonard R. Whyte and Cora E. Whyte,
for their immeasurable support, encouragement, and love.

Acknowledgements

I would not be able to obtain a PhD without the support of others, and I have many people to thank for supporting me over the past six years.

A thank you to my advisor, Richard Young, for encouraging my growth as an independent researcher, teaching me the skills necessary to be successful, and for providing me with constructive feedback.

A thank you to several other faculty members who have helped shape my career; Frank Solomon for making me believe in myself, Harvey Lodish for what can be achieved, Phil Sharp for his insightful advice, and Peter Reddien for perceptive comments.

A thank young to the entire Young Lab for being my colleagues, friends, and mentors. Special thanks to David Orlando for his computational help; Garrett Frampton for his hard work and true friendship; Matthew Guenther for donating his money during poker games; Steve Bilodeau for insightful comments; and a tremendous thank you to Tony Lee for being an outstanding mentor and friend.

A special thank you to my long-time friend, Terry O'Toole, for his unwavering friendship during the many times I've failed to keep in touch.

A special thank you to Jasmine De Cock, who has seen me through my struggles, and had the utmost confidence in me.

Most of all I would like to especially thank my parents, Leonard and Cora Whyte. Without their sacrifice, support, encouragement, confidence, and love I would not be obtaining this PhD. All my achievements are a testament to them.

Table of Contents

Preamble

In the first chapter of this thesis I discuss new themes that have been uncovered based on recent studies of the roles of chromatin regulators in the control of the embryonic stem (ES) cell state, and in the differentiation of embryonic stem cells. I first introduce chromatin regulators and discuss their roles in regulating the gene expression program that is necessary to maintain ES cells in a pluripotent state. I then highlight the mechanisms involved in ES cell differentiation, where chromatin regulators play critical roles both in activating new gene expression programs and in silencing old programs. I discuss new insights uncovered from my, and others' work, underscoring their potential implications. Based on my thesis work, I describe a model, wherein chromatin regulators with opposing activities co-occupy the same genes in ES cells, where a dynamic balance between these regulators governs the changes in chromatin and associated gene expression during ES cell differentiation. In Chapter 2, I describe how LSD1 and the NuRD complex function during differentiation of ES cells, which I propose works in a dynamic fashion to decommission enhancers that are active in ES cells; this work was published in 2012 in *Nature*. In the final chapter, I offer future challenges and possible approaches to test this model in other cell types. I have also appended 3 published papers, on which I participated as coauthor, that focus on other aspects of gene regulation and differentiation.

Chapter 1

Roles For Chromatin Regulators In

Transcriptional Control Of Embyronic Stem Cells

Introduction

During development, cells of the inner cell mass (ICM) of the developing blastocyst differentiate to generate an adult organism consisting of specialized cell types (Fig 1). Deciphering the molecular mechanisms controlling cellular differentiation is key to understanding development. Many diseases are the result of defects in development, and may lead to new therapies for diseases associated with developmental defects.

As cells differentiate, changes in the gene expression program occur that involve silencing of some genes and activation of others. The coordinated actions of transcription factors, chromatin regulators, signaling factors, and small RNAs that control gene expression programs have been the subject of much study in embryonic stem (ES) cells (reviewed in ref. 1). Only recently have investigators begun to understand how chromatin regulators contribute to the dynamic changes in gene expression programs as cells differentiate. I briefly describe the broad array of histone-modifying enzymes (reviewed in ref. 2), ATP-dependent nucleosome-remodeling complexes (reviewed in ref. 3), and DNA methyltransferases (reviewed in ref. 4) that are involved in the establishment and maintenance of ES cell state (Fig 2a), and discuss how they contribute to the control of this state. I then discuss new insights into how these regulators participate in the changes in the gene expression program as ES cells differentiate

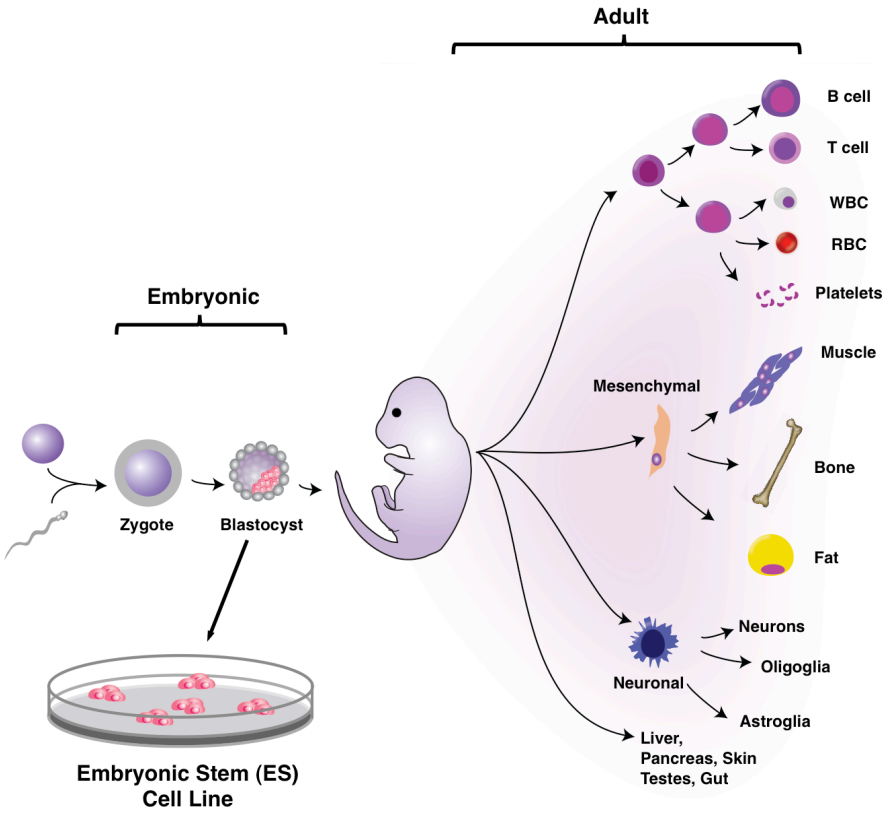

Adult
B cell
T cell
WBC
RBC
Platelets
Muscle
Mesenchymal
Bone
Fat
Neurons
Oligoglia
Neuronal
Astroglia
Liver,
Pancreas, Skin
Testes, Gut
Embryonic
Zygote
Blastocyst
Embryonic Stem (ES)
Cell Line

Figure 1

Figure 1: High level view of mammalian development

During mammalian embryonic development, a single fertilized egg develops and forms the blastocyst, which harbors the inner cell mass (ICM). Once derived from the ICM, embryonic stem (ES) cells can be derived in culture in a pluripotent state almost indefinitely (Box 1). After blastocyst formation, signaling cues trigger cells within the ICM to differentiate and generate an adult organism, consisting of over 200 specialized cell types, each with a distinct gene expression program controlling its cellular identity.

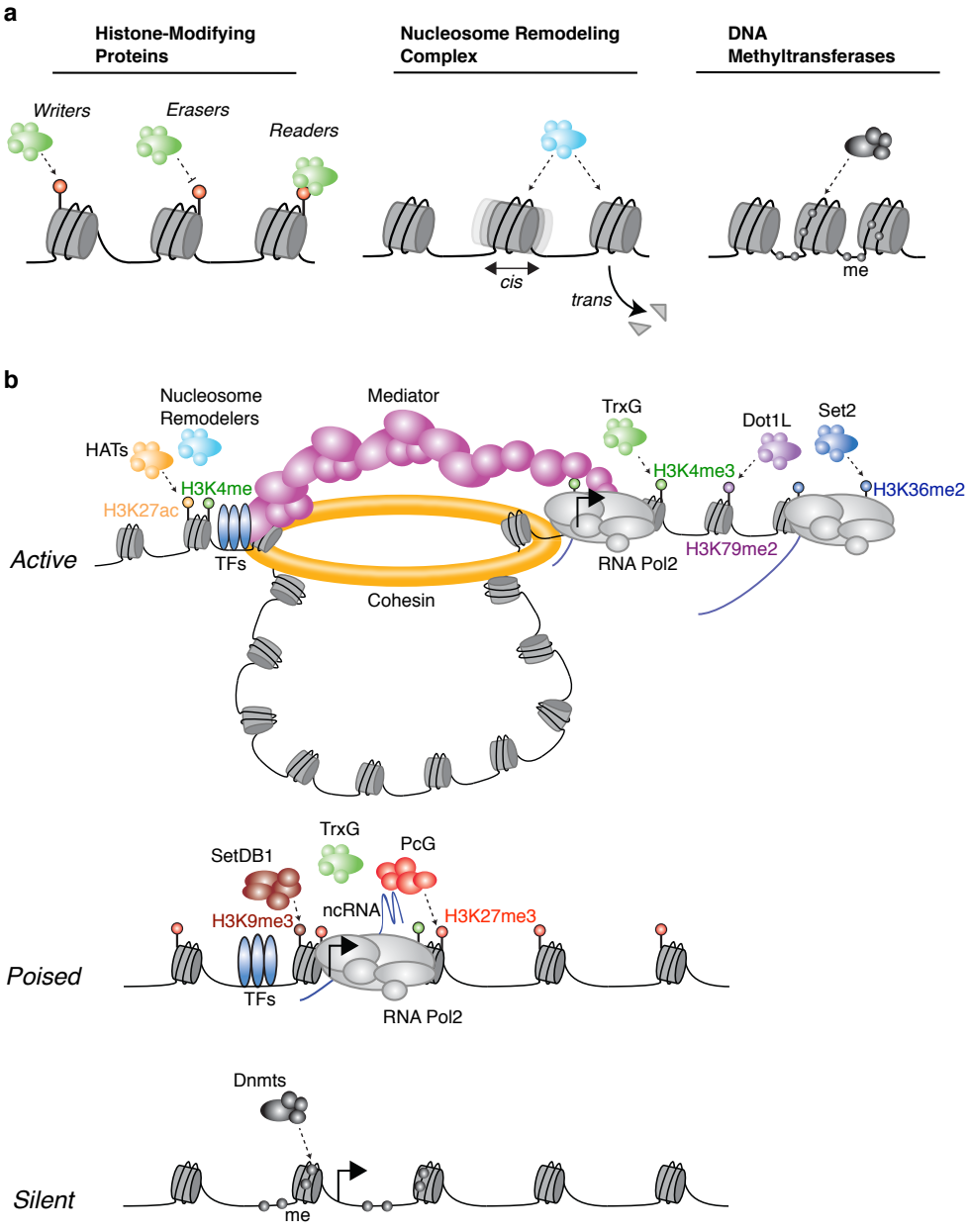

a
Histone-Modifying Proteins
Nucleosome Remodeling Complex
DNA Methyltransferases
Writers
Erasers
Readers
cis
trans
me
b
Nucleosome Remodelers
Mediator
TrxG
Dot1L
Set2
HATs
H3K4me
H3K27ac
H3K4me3
H3K36me2
Active
TFs
RNA Pol2
H3K79me2
Cohesin
TrxG
PcG
SetDB1
ncRNA
H3K9me3
H3K27me3
Poised
TFs
RNA Pol2
Dnmts
Silent
me

Figure 2

Figure 2: Models for chromatin regulators at transcriptionally active, poised and silent genes

a) Chromatin regulators generally fall into three different classes.

Left panel – Histone-modifying proteins consist of enzymes that either catalyze the addition of a modification (termed *writers*) or the removal of a modification (termed *erasers*). Histone-modifying proteins also include proteins that recognize and occupy sites containing specific histone modifications (termed *readers*).

Middle panel – ATP-dependent nucleosome remodeling proteins alter chromatin structure by sliding nucleosomes along DNA in *cis* or displacing histone proteins in *trans*. These proteins exist as multi-subunit complexes and generally belong to four different families: SWI/SNF, CHD, INO80 and ISWI.

Right panel – DNA methyltransferases catalyze the methylation of CpG dinucleotides on DNA strands.

b) Chromatin regulators contribute to the regulation of active, poised and silent genes in ES cells.

Upper panel – At active genes, enhancers are typically bound by multiple transcription factors, which recruit cofactors that can interact with RNA polymerase II (RNA Pol2) at the core promoter. Chromatin regulators, which include nucleosome-remodeling complexes such as SWI/SNF and histone-modifying proteins such as TrxG, are recruited by transcription factors or the transcription apparatus and mobilize or modify nucleosomes. The modifications catalyzed by these regulators localize in a stereotypic pattern and can serve as docking sites for factors that facilitate further transcription.

Middle panel – At poised genes, transcription initiation and recruitment of TrxG complexes can occur, but elongation and recruitment of Set2 and Dot1L do not occur. The PcG and SetDB1 chromatin regulators can contribute to this repression, and these can be recruited by some transcription factors and by noncoding RNAs. RNA Pol2 levels are lower at poised genes compared to the levels detected at active genes.

Lower panel – Silent genes show little or no evidence of transcription initiation or elongation. These genes are often occupied by DNA methyltransferases and and other chromatin regulators that methylate histone H3K9.

---

to establish new cell states. From these insights, an intriguing model is emerging involving a dynamic competition between regulators of opposing functions that control gene expression and cellular differentiation. Knowledge of the chromatin regulators involved in the control of gene expression programs as cells transition to new cell states can guide efforts to reprogram cells back to a pluripotent state and holds great promise for both disease therapeutics and regenerative medicine.

Control of the Pluripotent State

Once derived from the ICM, ES cells can be maintained in culture in a pluripotent state almost indefinitely (Fig 1). Control of the pluripotent state includes the regulation of the ES cell gene expression program by specific transcription factors and chromatin regulators, among other factors[1] (Box 1). Transcription factors recognize and bind specific DNA sequences to either activate or prevent transcription[5]. In ES cells, the gene expression program is largely governed by the core transcription factors Oct4, Sox2, and Nanog[6-8]. The core transcription factors function together to positively regulate their own promoters, forming an interconnected autoregulatory loop. This keeps ES cells in a state that is maintained by a positive-feedback-controlled gene expression program when the transcription factors are expressed at normal levels. This state is no longer maintained and differentiates when one of the core transcription factors is no longer functioning.

The core transcription factors activate a large fraction of the actively transcribed protein-coding and miRNA genes in ES cells[9,10]. Oct4, Sox2 and Nanog also occupy repressed genes encoding lineage-specific transcription factors, and the repression of these genes is essential for ES cells to maintain a stable pluripotent state[10-17]. The loss of these core transcription factors leads to activation of these developmental genes, indicating that these genes are poised for activation. In addition to Oct4, Sox2 and Nanog, other transcription factors have been shown to play important roles in control of ES cell state[18-28]. Thus, the functions of the core transcription factors are augmented by the functions of many transcription factors implicated in control of ES cell state at actively transcribed genes.

Transcription factors can bind to chromatin regulators that alter chromatin structure and influence gene expression[29-32]. Chromatin regulators generally fall into three classes: histone-modifying proteins, ATP-dependent nucleosome remodeling protein complexes, and DNA methyltransferases. These regulators play roles in ES cells, where they alter chromatin to influence gene expression. I briefly describe models for the functions of chromatin regulators in the control of the ES cell gene expression program, including gene activation, establishment of a poised state for gene activation, and gene silencing (Fig 2). These models provide the foundation for our understanding on the roles of chromatin regulators in control of the pluripotent state (Table 1).

Gene activation

The master ES cell transcription factors Oct4, Sox2 and Nanog bind to enhancer elements and recruit chromatin regulators and associated cofactors to facilitate transcription. These regulators participate in at least three aspects of gene activation (Fig 2b, *upper panel*): (1) Chromatin regulators are recruited to enhancer elements and core promoters, where they open chromatin to create accessible binding sites. (2) Chromatin regulators alter chromatin architecture to link enhancer elements to core promoters. (3) Chromatin regulators facilitate transcription initiation and (4) transcription elongation.

Gene activation – Creation of accessible binding sites

During gene activation, chromatin regulators are recruited and open chromatin to create accessible binding sites. These regulators include cofactors, which are protein complexes that contribute to activation (coactivators) and repression (corepressors) but do not have DNA-binding properties of their own. In ES cells, Oct4, Sox2 and Nanog bind active enhancers and bind coactivators such as the histone acetyltransferase p300. Histone acetyltransferases (HATs) are histone-modifying enzymes (Fig 2a) that utilize acetyl CoA as a cofactor to catalyze the transfer of an acetyl group to the ε-amino group of lysine side chains, and are generally divided into three main families (reviewed in ref. 33): Gcn5-related N-acetyltransferases (GNAT), MOZ, Ybf22/Sas3, Sas2, and Tip60 (MYST), and CREB-binding protein (CBP)/p300. The p300 cofactor occupies most active enhancers and promoters in ES cells[34], where they catalyze acetylation of histones. Histone acetylation is thought to regulate chromatin accessibility by affecting the nucleosome net charge and by reducing electrostatic interactions between histones

and DNA or between histones of neighboring nucleosomes[35-38]. In ES cells, p300 regulates Nanog levels by acetylation of the *Nanog* distal enhancer[39]. Nucleosomes containing acetylated histones also serve as docking sites for bromodomain effector proteins. For example, human TAF1, a subunit of TFIID of RNA Polymerase II (RNA Pol2), contains double bromodomains that bind multiply acetylated histone H4, and participates in the assembly of the transcription machinery[40].

Components of nucleosome-remodeling complexes have also been implicated in the regulation of chromatin accesibility[3,11,41-45]. ATP-dependent nucleosome-remodeling complexes can be recruited to gene promoters, where they utilize ATP to slide nucleosomes in *cis* along the DNA, or displace histones in *trans* (Fig 2a), resulting in positive or negative effects on gene activity (reviewed in ref. 46). These remodelers generally fall into four families: switch/sucrose nonfermentable (SWI/SNF; reviewed in ref. 3), chromodomain helicase DNA-binding (CHD; reviewed in ref. 47), inositol-requiring 80 (INO80; reviewed in ref. 48), and imitation switch (ISWI; reviewed in ref. 49). In ES cells, the CHD subunit Chd1 is associated with the enhancers and promoters of active genes, where it interacts with Oct4/Sox2/Nanog master transcription factors and RNA Pol2 and prevents heterochromatin formation[41]. It is unclear how Chd1 acts to maintain an open chromatin structure. An intriguing possibility is that Chd1 may mediate incorporation of the histone variant H3.3, which is generally associated with active genes and is less prone to nucleosome compaction and heterochromatin formation[50,51]. Thus, histone-modifying enzymes and nucleosome remodeling

complexes are present at key active ES cell genes, where they contribute to opening chromatin to create accessible binding sites.

Gene activation – Linking enhancer elements to core promoters

In addition to increasing chromatin accessibility for binding, chromatin regulators participate in the linking of enhancer elements to core promoters. The mediator coactivator physically links Oct4/Sox2/Nanog-bound enhancers to the promoters of active genes in ES cells[52]. At Oct4/Sox2/Nanog-regulated promoters, mediator associates with cohesin, and this complex forms a DNA loop between enhancers and master promoters that is necessary for normal gene activity. Mediator also plays an important role in the transcriptional response to signaling cues. CDK8, the kinase subunit of mediator, can influence the activity of signaling transcription factors[53-55]. For example, CDK8-meditated phosphorylation of the linker region within Smad1/5 or Smad2/3 complexes can activate these transcription factor complexes, but it also targets them for proteasomal degradation. A dynamic cycle of transcription factor activation and destruction ensures maintained gene activation and may facilitate rapid changes in cell state when signaling is altered. Hence, the association of mediator with cohesin and its contribution to both DNA looping and gene activity in ES cells makes it both an essential cofactor and a key chromatin regulator[52].

Gene activation – Transcripton initiation

Once enhancer elements are brought in to close proximity with core promoters,

chromatin regulators can facilitate the initiation of transcription and production of full-length messenger RNA transcripts. In ES cells, Oct4/Sox2/Nanog-regulated promoters are occupied by Trithorax group (TrxG) proteins[56,57]. These proteins were discovered in *D. melanogaster* as activators of Hox genes (reviewed in ref. 58). The SET-domain containing Mll proteins are TrxG proteins that catalyze trimethylation of histone H3 lysine 4 (H3K4me3) and facilitate maintenance of active gene states during development, in part by antagonizing the functions of Polycomb group (PcG) proteins. Mll proteins are recruited to promoters by various mechanisms[58], including the transcription apparatus[59], and short RNAs[60]. Recent evidence suggests RNA Pol2 transcribes short RNA transcripts, which then tether to TrxG proteins to either target or stabilize these proteins at Pol2-occupied promoters[60]. The TrxG protein Wdr5 has recently been implicated in control of ES cell state by regulating H3K4me3 levels and contributing to gene activity[57]. The molecular mechanisms by which this occurs is unclear, but likely involves H3K4me3 serving as a docking site for nucleosome remodeling complexes and HATs that facilitate chromatin remodeling and transcription[41,45,61,62]. For example, the nucleosome remodeling subunit Bptf has a PHD domain that recognizes methylation of histone H3K4. This interaction recruits the ISWI nucleosome remodeling complex NuRF to disrupt chromatin at the promoter and enhance RNA Pol2 clearance[61]. In ES cells, the Tip60-p400 complex, an INO80 nucleosome remodeling complex containing histone acetyltransferase activity, has been shown to be recruited directly by H3K4me3 to enhancers and promoters[45]. Histone acetylation at gene enhancers and promoters serves as a docking site for the

binding of the bromodomain-containing protein Brd4 and mediates recruitment of positive transcription elongation factor b (pTEF-b), which is necessary for the release of the promoter-proximal paused RNA Pol2 (ref. 62). Thus, the link of H3K4 methylation to histone acetylation and chromatin remodeling may regulate chromatin accessibility and increased escape of RNA Pol2 from active core promoters in ES cells.

Gene activation – Transcription elongation

After RNA Pol2 release, the mechanisms by which chromatin regulators facilitate transcription elongation at a subset of genes in ES cells are not fully understood. Based on studies in yeast, transcription elongation involves targeting of TrxG complexes via interactions with the polymerase-associated factor 1 (Paf1) elongating complex, resulting in di- and trimethylation of histone H3K4 (ref. 63). Paf1 also regulates the histone-modifying enzyme Dot1L, which is recruited to elongating RNA Pol2 through its association with various elongating factors. Dot1L catalyzes dimethylation of histone H3 lysine 79 (H3K79me2), likely to maintain active transcription and/or serve as a transcription memory[63]. The histone-modifying enzyme Set2, which catalyzes trimethylation of histone H3 lysine 36 (H3K36me3), is also linked to transcription elongation, presumably to maintain chromatin structure via the recruitment of a histone deacetylase complex to prevent cryptic initiation of transcription[63-66]. In ES cells, the Paf1 complex may play similar roles, as reduction in Paf1 proteins leads to reduced levels of histone H3K4me3 at actively transcribed genes[67].

In summary, chromatin regulators are recruited by transcription factors, the transcription apparatus or short RNAs to create accessible binding sites, alter chromatin architecture and facilitate transcription by mobilizing or modifying local nucleosomes. These modifications localize in a stereotypic pattern, and serve as docking sites for recruitment of regulators that further facilitate gene activation (Fig 2b, *upper panel*).

Poised state for gene activation

Oct4, Sox2, and Nanog also bind repressed genes encoding lineage-specific transcription factors, and the repression of these genes is essential to maintain a stable pluripotent ES cell state[10-14,17,68,69] (Fig 2b, *middle panel*). Loss of these master transcription factors leads to rapid induction of a wide spectrum of genes encoding lineage-specific regulators, indicating that these genes are poised for activation. The chromatin regulators known to have a profound impact on ES cell state are histone-modifying enzymes that repress genes encoding lineage-specific transcription factors. These include SetDB1 and the PcG protein complexes.

Multiple histone-modifying enzymes that methylate histone H3 lysine 9 (H3K9) have been implicated in control of ES cell state[11,70,71]. Oct4 can bind sumoylated SetDB1, a histone-modifying enzyme that trimethylates H3K9 (H3K9me3), at a subset of the repressed genes that encode lineage-specific developmental regulators, including those involved in generating the extraembryonic trophoblast lineage[11,70,71]. At the promoters of these genes, SetDB1 is responsible for catalyzing

H3K9 methylation[11,70,71]. Histone H3K9me3 serves as a docking site for the recruitment of the effector protein heterochromatin protein 1 (HP1), which recognizes and binds methylated histone H3K9[11,70,71]. Once HP1 occupies sites containing H3K9me3, the effector protein oligomerizes to bridge nearby nucleosomes to compact chromatin[72-74]. At SetDB1-occupied lineage-specific genes in ES cells, H3K9me3 and chromatin compaction facilitate gene repression[13,48,49]. The corepressor Trim28 can also interact with HP1 and SetDB1 to facilitate formation of repressive chromatin[75,76], but the mechanisms are not fully understood.

PcG protein complexes can associate with nucleosomes with histone H3K9me3 and further contribute to repression[77]. These proteins occupy most of the transcriptionally repressed lineage-specific genes targeted by Oct4, Sox2, and Nanog[13,14,78]. The PcG proteins form multiple polycomb repressive complexes (PRCs), the components of which are conserved from *Drosophila* to humans (reviewed in ref. 79). Prc2 catalyzes histone H3 lysine 27 (H3K27) methylation, an enzymatic activity required for Prc2-mediated gene silencing. Methylation of histone H3K27 is thought to provide a binding surface for Prc1, which facilitates oligomerization, condensation of chromatin structure, and inhibition of chromatin remodeling activity in order to maintain gene silencing. Prc1 also contains a histone ubiquitin ligase, Ring1b, whose activity appears likely to contribute to silencing in ES cells[80].

Various models have been proposed to explain how PcG proteins silence genes encoding key regulators of development yet allow them to remain in a state that is poised for activation during differentiation[13,14,78,81-84,85-90]. A particularly intriguing model involves a dynamic balance between TrxG and PcG proteins at promoters of developmental genes[91]. Recruitment of TrxG complexes and transcription initiation at promoters of developmental regulators results in catalyzation of H3K4me3 and production of short CG-rich RNAs that recruit PcG proteins. These proteins then catalyze histone H3K27 methylation and histone H2A ubiquitylation, and then spread beyond the nucleation site to establish repression. This repression prevents RNA Pol2 pause release, elongation, and recruitment of Dot1L and Set2 proteins. How this repression is achieved is not fully understood, but likely involves a variety of mechanisms[58]. Prc2-mediated methylation of histone H3K27 might directly interfere with transcriptional activation by counteracting TrxG-associated proteins and the histone modifications they mediate, including acetylation of histone H3K27 (H3K27ac) by p300 (refs. 92,93) and histone H3K4me3 (refs. 94-96). Prc1 binding and ubiquitylation of histone H2A has been suggested to induce chromatin compaction and interfere with nucleosome-remodeling activities to prevent RNA Pol2 transcriptional elongation and prevent H3K36 methylation by Set2 (ref. 97).

In conclusion, Oct4, Sox2 and Nanog may recruit SetDB1 through protein-protein interactions, and PcG complexes via interactions with both histone H3K9me3 and RNA transcripts produced as a consequence of local transcription activation. These

complexes may serve to pause RNA polymerase machinery at key regulators of development in pluripotent cells (Fig 2b, *middle panel*). At genes where the transcription apparatus is continuously recruited by activating transcription factors (Fig 2a, *upper panel*), activities associated with TrxG proteins predominate, reducing PcG complexes and their associated histone modifications.

Gene silencing

Oct4, Sox2 and Nanog are not bound at a subset of genes that are silenced and show little to no evidence of transcription initiation and elongation (Fig 2b, *lower panel*). These genes are occupied by DNA methyltransferases (Fig 2a), among other chromatin regulators. DNA methylation is essential for mammalian development[98,99], and is established and maintained by three catalytically active enzymes[100]. DNA methyltransferase (Dnmt)3a and Dnmt3b are responsible for de novo methylation, and Dnmt1 functions as a maintenance methyltransferase, ensuring newly synthesized DNA is methylated during DNA replication. Though ES cells can be established and maintained in the absence of Dnmts and DNA methylation[101,102], recent evidence suggests a subset of these silent genes are regulated by DNA methylation. For example, the silent trophectodermal gene *Elf5* is highly methylated, and cells deficient in Dnmt1 show hypomethylation of the *Elf5* promoter and gain the ability to differentiate into trophectoderm[103].

There are various mechanisms by which these genes may be silenced by DNA methylation. One possibility is that DNA methylation can directly block

transcription factor binding. A more likely scenario is that DNA methylation represses gene expression through several methyl-CpG-binding proteins (MECPs) that recognize and subsequently bind to methylated DNA. For example, in somatic cells MECP2 forms a complex with histone deacetylases and a corepressor protein, Sin3a, to repress transcription in a methylation-dependent manner[104,105]. The methyl-CpG-binding protein MBD2 forms a complex with the CHD family member CHD3/4 to form the nucleosome remodeling and deacetylase (NuRD) complex, which contains histone deacetylase (HDAC) activity. The NuRD complex can occupy methylated promoters and remodel methylated chromatin to repress genes[106,107]. The Sin3a and NuRD complexes provide a mechanistic link between DNA methylation and histone deacetylation in gene silencing.

In summary, models exist for the functions of chromatin regulators in the control of the pluripotent state (Fig 2). These regulators can be found at active, poised, and silent genes. At active genes (Fig 2b, *upper panel*), enhancers are typically bound by multiple transcription factors, which recruit cofactors that can interact with RNA Pol2 at the core promoter. RNA Pol2 generates a short transcript and pauses until pause-release and elongation factors allow further transcription. Chromatin regulators, which include nucleosome-remodeling complexes such as CHD complexes and histone-modifying complexes such as TrxG, Dot1, and Set2, are recruited by transcription factors or the transcription apparatus and mobilize or modify local nucleosomes. At poised genes (Fig 2b, *middle panel*), transcription initiation and recruitment of TrxG can occur, but pause release, elongation, and

recruitment of Dot1 and Set2 do not occur. SetDB1 and PcG proteins can contribute to this repression, and these can be recruited by transcription factors and by non-coding RNAs. These genes are rapidly activated when ES cells are stimulated to differentiate. At silent genes (Fig 2b, *lower panel*), there is little or no evidence of transcription initiation and elongation, and these genes are often occupied by chromatin regulators that methylate DNA and H3K9. Together, these regulators regulate the ES cell gene expression program to control the pluripotent state.

Exit from the Pluripotent State

Upon ES cell differentiation and establishment of new cell states, at least three things must occur (Fig 3): (1) Certain genes of the prior gene expression program must be silenced. (2) Genes of the new gene expression program must be activated. (3) Genes of alternative gene expression programs must be silenced. Recent evidence in ES cells indicates that a large fraction of genes that are either active or poised for activation upon differentiation are co-occupied by chromatin regulators with opposing activities—those that modify or mobilize nucleosomes to facilitate chromatin accessibility and gene activity, and those that modify or mobilize nucleosomes to reduce chromatin accessibility and gene activity.

At least two models that are not mutually exclusive emerge from these findings. Such opposing activities at genes may allow for precise tuning of gene expression in ES cells. A more intriguing view is that a dynamic balance exists in ES cells

between regulators of opposing functions at enhancers and core promoters of genes, and this balance poises genes to become either silenced or activated upon differentiation. In ES cells, chromatin regulators with opposing functions occupy enhancer elements and core promoters of a subset of active genes, where the actions of chromatin regulators responsible for maintained gene activation

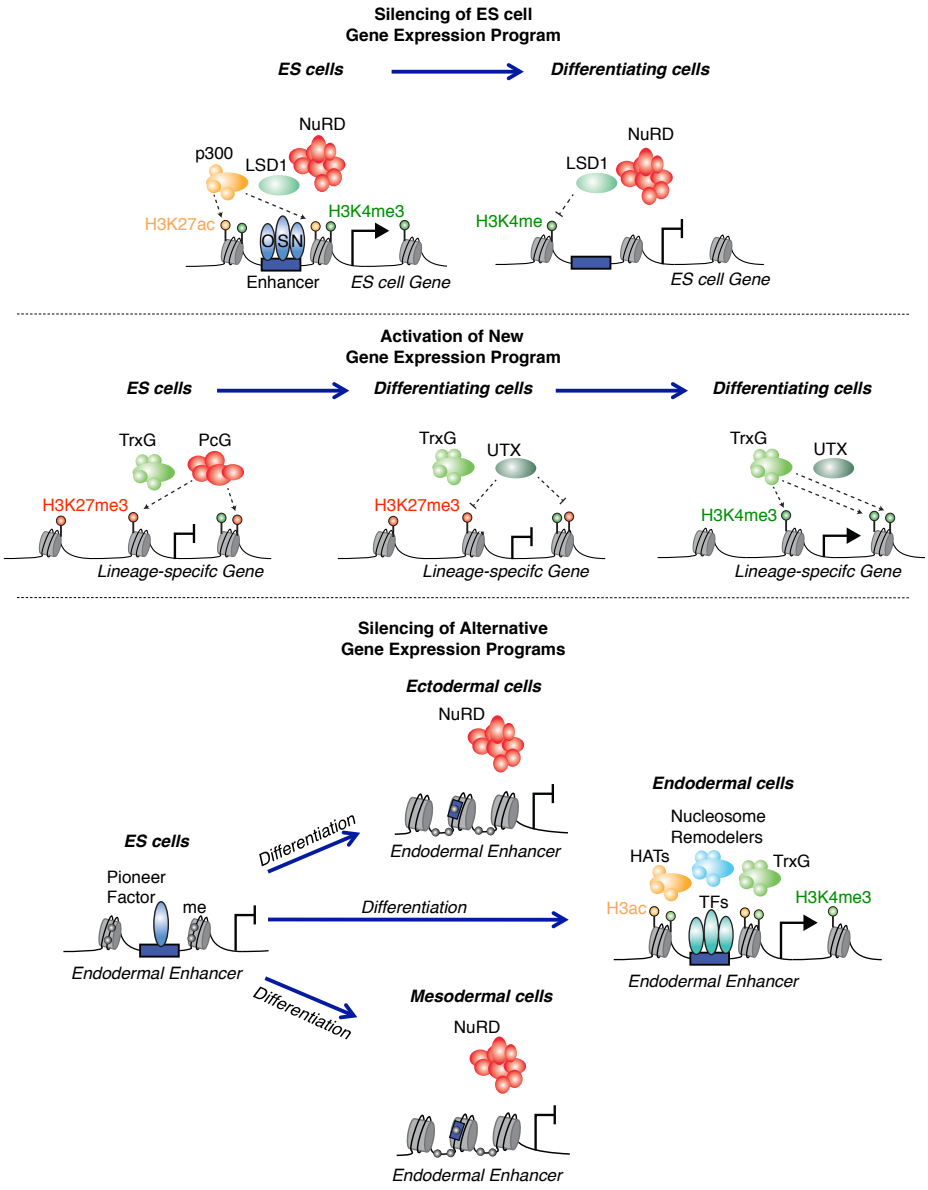


Figure 3

Figure 3: Models for chromatin regulators during ES cell differentiation

Chromatin regulators play important roles during ES cell differentiation, where they participate in the silencing of certain active genes of the ES cell gene expression program, the activation of genes of the new gene expression program, and silencing of genes of alternative gene expression programs.

Upper panel – Chromatin regulators participate in the silencing of certain active genes of the ES cell gene expression program. In ES cells, LSD1 and NuRD complexes are recruited to active enhancer elements containing H3K4 methylation. LSD1 is unable to demethylate histones, however, due to the presence of acetylated histones that are catalyzed by HATs such as p300. During differentiation, when the levels of Oct4 and associated HATs are lost, acetylated histones are reduced and LSD1 is able to demethylate H3K4, resulting in enhancer decommissioning and gene silencing.

Middle panel – Chromatin regulators participate in the selective activation of genes encoding lineage-specific transcription factors. In ES cells, these genes are occupied by TrxG and RNA Pol2, but these genes are not fully transcribed and are repressed by PcG complexes that catalyze H3K27 methylation. During differentiation, Utx removes H3K27 methylation, and the TrxG member Dpy30 is

responsible for increased H3K4 methylation at promoters of certain lineage-specific genes that become activated.

Lower panel – Chromatin regulators are also responsible for silencing genes of alternative gene expression programs. In ES cells, a pioneer factor binds the endoderm-specific enhancer of the *Alb1* gene and protects it from DNA methylation. The surrounding regions are DNA methylated, and *Alb1* is silent. During differentiation of ES cells down mesodermal and ectodermal lineages, the pioneer factor is lost at the *Alb1* enhancer and the DNA is methylated, facilitating stable silencing of *Alb1*. The NuRD complex contributes to this silencing. During differentiation of ES cells down the endodermal lineage, transcription factors bind the *Alb1* enhancer to facilitate activation. HATs and the nucleosome remodeling complex SWI/SNF are believed to facilitate gene activation.

---

predominate. Upon differentiation, when the high levels of regulators associated with gene activation are reduced, the competitive balance shifts and these genes are repressed by the chromatin regulators associated with gene repression. At lineage-specific genes in ES cells, the chromatin regulators responsible for gene repression predominate. When cells are triggered to differentiate, the levels or functions of chromatin regulators associated with repression are reduced at selected genes, and lineage-specific transcription factors are selectively activated. Hence, in ES cells a set of active genes are poised for repression upon differentiation, and a set of lineage-specific genes are poised for activation upon differentiation. We discuss new insights into how chromatin regulators might facilitate this dynamic control.

Silencing genes of the old gene expression program

Chromatin regulators control at least two major steps in the silencing of genes of the old gene expression program (Fig 3, *upper panel*)[44,108-110]. Some chromatin regulators are required in the initial silencing of genes by occupying enhancer elements and core promoters of genes, where they either modify or mobilize nucleosomes and facilitate gene repression. Other chromatin regulators are required to stably maintain the silent state by methylating DNA.

The histone-demethylating protein LSD1 is responsible for removal of mono- and dimethylation of histone H3K4 (ref. 111), two modifications typically found at active enhancer elements[112]. In ES cells, LSD1 and members of the NuRD ATP-dependent nucleosome remodeling complex, including HDACs and the ATPase CHD4, interact with Oct4 (refs. 113,114), and occupy Oct4-regulated active enhancer elements[108]. LSD1 proteins do not substantially demethylate histone H3K4 because the H3K4 demethylase activity of LSD1 is inhibited in the presence of acetylated histones[115,116]. Enhancers occupied by Oct4, Sox2 and Nanog are also occupied by the HAT p300 and nucleosomes with acetylated histones[9]. Thus, as long as the enhancer-bound transcription factors recruit HATs to enhancers, the net effect of having both HATs and NuRD-associated HDACs present is to have sufficient levels of acetylated histones to suppress LSD1 demethylase activity. During ES cell differentiation, the levels of Oct4 and p300 are decreased, thus decreasing the level of acetylated histones, which in turn permits the demethylation of H3K4 by LSD1. This model would explain why key components of LSD1 and NuRD complexes are not essential for the maintenance of ES cell state but are essential for normal differentiation, when the active enhancers and associated genes must be silenced. Several lines of evidence demonstrate that LSD1 is required for differentiation of multiple cell types[108,117-121], suggesting that LSD1 is likely to be generally involved in gene silencing during differentiation (Fig 3a).

To stably maintain gene silencing, a different set of chromatin regulators are required. Following histone deacetylation and H3K4 demethylation at the *Oct4*

gene, the histone-modifying enzyme G9a methylates histone H3K9, leading to HP1 recruitment and chromatin compaction[109]. G9a-mediated methylation is required for de novo DNA methylation by Dnmt3a and Dnmt3b, resulting in stable gene repression[110,122]. In summary, during ES cell differentiation, certain genes of the old gene expression program undergo a multi-step, tightly regulated form of repression that requires the activities of various chromatin regulators to enforce a stable form of silencing that is maintained by DNA methylation.

Activating genes of the new gene expression program

Recent studies reveal that histone-modifying and nucleosomal-remodeling enzymes play critical roles in the activation of gene expression programs important for differentiation and establishment of new cell states. From these studies, a dynamic model is emerging for the functions of PcG/TrxG complexes and associated proteins in the regulation of genes encoding lineage-specific transcription factors during differentiation. In ES cells, transcription initiation and recruitment of TrxG and associated coactivators to promoters of lineage-specific genes can occur, but pause release and elongation do not occur, due in part to PcG-mediated repression. This dynamic balance is lost when cells are stimulated to differentiate, as PcG proteins are reduced from a subset of these promoters and increased occupancy and stabilization of TrxG proteins facilitate gene activation.

The mechanisms involved in the loss of PcG proteins during differentiation is unclear, but likely involve the removal of H3K27 methylation by histone-

demethylating proteins (Fig 3, *middle panel*). The histone-demethylating enzyme Utx is responsible for demethylating histone H3K27 (refs. 96,123,124) at certain PcG-occupied promoters of developmental genes[125]. During ES cell differentiation, master lineage-specific transcription factors bind to a subset of developmental genes that contain H3K27 methylation and are occupied by PcG proteins. These factors then bind to Utx, which catalyzes the removal of H3K27 methylation at these genes to facilitate transcriptional activation. How demethylation of H3K27 contributes to gene activation is not fully understood but likely involves loss of Prc1 proteins. Brg1, the SWI/SNF ATPase and member of the ES-specific nucleosome remodeling complex esBAF (ref. 126), is also targeted to these promoters, presumably to facilitate chromatin accessibility[125]. H3K27 demethylation has also been linked to increased acetylation of H3K27 by p300, which can facilitate activation of PcG-target genes (ref. 92). Considering Utx is broadly expressed in multiple tissues (reviewed in ref. 127), the interaction between Utx and master lineage-specific transcription factors likely plays an important role in the specific activation of Utx targets.

Activation of lineage-specific transcription factors not only requires the loss of PcG proteins, but also increased occupancy of TrxG proteins and associated cofactors that facilitate gene activity. Utx associates with Mll TrxG proteins that catalyze the methylation of H3K4 (refs. 94-96). Therefore, it is likely that Utx is required for the activity or the targeting of TrxG proteins at genes encoding lineage-specific transcription factors genes, and contributes to gene activation[94,96]. The mammalian TrxG member Dpy30 is part of the Mll complex of proteins

responsible for catalyzing H3K4 methylation. In ES cells, Dpy30 occupies promoters of PcG-targeted genes, where it regulates H3K4 methylation[128]. Loss of Dpy30 has limited effects on H3K4 methylation levels and gene expression, presumably due to the presence of PcG proteins that repress gene activity. During differentiation, Dpy-30 occupancy increases at a subset of genes whose promoters lose PcG-mediated H3K27 methylation. H3K4 methylation levels are also increased at these genes, which facilitates gene activation[128]. In summary, the dynamic balance between PcG and TrxG proteins affects expression of genes that become activated during differentiation (Fig 3, *middle panel*). Loss of PcG proteins by demethylases like Utx, and increased activity of TrxG proteins such as Dpy30 contribute to the activation of the new gene expression program during differentiation.

Silencing genes of alternative programs

During differentiation and establishment of a new cell type, genes encoding lineage-specific master transcription factors of alternative gene expression programs must be silenced, because forced expression of these factors is associated with establishment and control of new cell states[129-131]. PcG proteins occupy promoters of lineage-specific genes whose promoters contain H3K27me3 and are repressed. Upon differentiation, a subset of these genes maintain H3K27 methylation and are not activated, providing a model that PcG targets are specified early in development, and continue to be repressed until activated. ES cells lacking several PcG subunits fail to silence several key regulators of development, and have differentiation defects[132].

Loss of DNA methylation impairs ES cell differentiation, also suggesting an important role for DNA methyltransferases during differentiation[101]. Most of the changes in DNA methylation associated with differentiation occur at regions distal from known promoters[133], suggesting gene enhancers of alternative gene expression programs may be methylated and silenced during differentiation[134] (Fig 3, *lower panel*). For example, the *Alb1* gene is repressed in ES cells and the endoderm-specific *Alb1* enhancer is protected from DNA methylation. During differentiation into endodermal cells, the pioneer transcription factor FoxA1 binds to the *Alb1* enhancer and facilitates chromatin remodeling and transcriptional activation[135]. This activation is likely accompanied by specific transcription factor recruitment and increased histone acetylation and H3K4 methylation, as well as by recruitment of SWI/SNF remodeling complexes[136-138]. During differentiation into mesoderm or ectoderm lineages, however, the endoderm-specific *Alb1* enhancer becomes DNA methylated and the *Alb1* gene is silenced, likely by Dnmts and the NuRD complex[134]. Hence, methylation of enhancer elements controlling developmental regulators might be a key feature of silencing alternative gene expression programs during differentiation (Fig 3, *lower panel*).

In summary, chromatin regulators with opposing functions co-occupy a large subset of active and poised genes in ES cells, providing a more dynamic view on the mechanisms controlling ES cell maintenance and differentiation. This model describes a dynamic competition that exists at genes between regulators that promote gene activity and those that promote gene silencing. At most active genes,

the activities of chromatin regulators responsible for maintained gene activation predominate, and the net effect of having regulators of opposing activities is to keep genes poised for repression during differentiation. At lineage-specific genes, the chromatin regulators responsible for gene repression predominate, and the net effect of having regulators of opposing activities at these genes is to keep genes poised for activation during differentiation. When ES cells are triggered to differentiate, the levels or functions of chromatin regulators associated with gene activation are reduced at a set of active genes, allowing these genes to become silenced. At lineage-specific genes, the levels or functions of chromatin regulators associated with repression are reduced at selected genes, and lineage-specific transcription factors are selectively activated.

References

1 Young, R.A., Control of the embryonic stem cell state. *Cell* 144 (6), 940-954 (2011).

2 Li, B., Carey, M., & Workman, J.L., The role of chromatin during transcription. *Cell* 128 (4), 707-719 (2007).

3 Ho, L. & Crabtree, G.R., Chromatin remodelling during development. *Nature* 463 (7280), 474-484 (2010).

4 Meissner, A., Epigenetic modifications in pluripotent and differentiated cells. *Nature biotechnology* 28 (10), 1079-1088 (2010).

5 Jacob, F. & Monod, J., Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3, 318-356 (1961).

6 Chambers, I. & Smith, A., Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* 23 (43), 7150-7160 (2004).

7 Niwa, H., How is pluripotency determined and maintained? *Development* 134 (4), 635-646 (2007).

8 Silva, J. & Smith, A., Capturing pluripotency. *Cell* 132 (4), 532-536 (2008).

9 Chen, X. *et al*., Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133 (6), 1106-1117 (2008).

10 Marson, A. *et al*., Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134 (3), 521-533 (2008).

11 Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., & Young, R.A., SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev* 23 (21), 2484-2489 (2009).

12 Boyer, L.A. *et al*., Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122 (6), 947-956 (2005).

13 Boyer, L.A. *et al*., Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441 (7091), 349-353 (2006).

14 Lee, T.I. *et al*., Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125 (2), 301-313 (2006).

15 Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., & Helin, K., Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J* 23 (20), 4061-4071 (2004).

16 Pasini, D. *et al*., Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev* 22 (10), 1345-1355 (2008).

17 Loh, Y.H. *et al*., The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 38 (4), 431-440 (2006).

18 Cartwright, P. *et al*., LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* 132 (5), 885-896 (2005).

19 Rahl, P.B. *et al*., c-Myc regulates transcriptional pause release. *Cell* 141 (3), 432-445 (2010).

20 Ivanova, N. *et al*., Dissecting self-renewal in stem cells with RNA interference. *Nature* 442 (7102), 533-538 (2006).

21 Zhang, X., Zhang, J., Wang, T., Esteban, M.A., & Pei, D., Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *J Biol Chem* 283 (51), 35825-35833 (2008).

22 Niwa, H., Ogawa, K., Shimosato, D., & Adachi, K., A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* 460 (7251), 118-122 (2009).

23 Galan-Caridad, J.M. *et al*., Zfx controls the self-renewal of embryonic and hematopoietic stem cells. *Cell* 129 (2), 345-357 (2007).

24 Dejosez, M. *et al*., Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. *Genes Dev* 24 (14), 1479-1484 (2010).

25 Zhang, J. *et al*., Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat Cell Biol* 8 (10), 1114-1123 (2006).

26 Wu, Q. *et al*., Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem* 281 (34), 24090-24094 (2006).

27 Jiang, J. *et al*., A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat Cell Biol* 10 (3), 353-360 (2008).

28 Chia, N.Y. *et al*., A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* 468 (7321), 316-320 (2010).

29 Kornberg, R.D. & Thomas, J.O., Chromatin structure; oligomers of the histones. *Science* 184 (4139), 865-868 (1974).

30 Olins, A.L. & Olins, D.E., Spheroid chromatin units (v bodies). *Science* 183 (4122), 330-332 (1974).

31 Brownell, J.E. *et al*., Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* 84 (6), 843-851 (1996).

32 Cote, J., Quinn, J., Workman, J.L., & Peterson, C.L., Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* 265 (5168), 53-60 (1994).

33 Lee, K.K. & Workman, J.L., Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol* 8 (4), 284-295 (2007).

34 Chen, X., Vega, V.B., & Ng, H.H., Transcriptional regulatory networks in embryonic stem cells. *Cold Spring Harbor symposia on quantitative biology* 73, 203-209 (2008).

35 Wolffe, A.P. & Hayes, J.J., Chromatin disruption and modification. *Nucleic acids research* 27 (3), 711-720 (1999).

36 Dorigo, B., Schalch, T., Bystricky, K., & Richmond, T.J., Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *Journal of molecular biology* 327 (1), 85-96 (2003).

37 Shogren-Knaak, M. *et al*., Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311 (5762), 844-847 (2006).

38 Robinson, P.J. *et al*., 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *Journal of molecular biology* 381 (4), 816-825 (2008).

39 Zhong, X. & Jin, Y., Critical roles of coactivator p300 in mouse embryonic stem cell differentiation and Nanog expression. *The Journal of biological chemistry* 284 (14), 9168-9175 (2009).

40 Jacobson, R.H., Ladurner, A.G., King, D.S., & Tjian, R., Structure and function of a human TAFII250 double bromodomain module. *Science* 288 (5470), 1422-1425 (2000).

41 Gaspar-Maia, A. *et al*., Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* 460 (7257), 863-868 (2009).

42 Klochendler-Yeivin, A. *et al*., The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. *EMBO reports* 1 (6), 500-506 (2000).

43 Schnetz, M.P. *et al*., CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS genetics* 6 (7), e1001023 (2010).

44 Schaniel, C. *et al*., Smarcc1/Baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. *Stem Cells* 27 (12), 2979-2991 (2009).

45 Fazzio, T.G., Huff, J.T., & Panning, B., An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 134 (1), 162-174 (2008).

46 Bell, O., Tiwari, V.K., Thoma, N.H., & Schubeler, D., Determinants and dynamics of genome accessibility. *Nat Rev Genet* 12 (8), 554-564 (2011).

47 Hall, J.A. & Georgel, P.T., CHD proteins: a diverse family with strong ties. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 85 (4), 463-476 (2007).

48 Bao, Y. & Shen, X., INO80 subfamily of chromatin remodeling complexes. *Mutation research* 618 (1-2), 18-29 (2007).

49 Dirscherl, S.S. & Krebs, J.E., Functional diversity of ISWI complexes. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 82 (4), 482-489 (2004).

50 McKittrick, E., Gafken, P.R., Ahmad, K., & Henikoff, S., Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proceedings of the National Academy of Sciences of the United States of America* 101 (6), 1525-1530 (2004).

51 Konev, A.Y. *et al*., CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin in vivo. *Science* 317 (5841), 1087-1090 (2007).

52 Kagey, M.H. *et al*., Mediator and cohesin connect gene expression and chromatin architecture. *Nature* (2010).

53 Alarcon, C. *et al*., Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways. *Cell* 139 (4), 757-769 (2009).

54 Fryer, C.J., White, J.B., & Jones, K.A., Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. *Molecular cell* 16 (4), 509-520 (2004).

55 Taatjes, D.J., The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends Biochem Sci* (2010).

56 Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., & Young, R.A., A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130 (1), 77-88 (2007).

57 Ang, Y.S. *et al*., Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* 145 (2), 183-197 (2011).

58 Schuettengruber, B., Martinez, A.M., Iovino, N., & Cavalli, G., Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* 12 (12), 799-814 (2011).

59 Muntean, A.G. *et al*., The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. *Cancer cell* 17 (6), 609-621 (2010).

60 Wang, K.C. *et al*., A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472 (7341), 120-124 (2011).

61 Wysocka, J. *et al*., A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442 (7098), 86-90 (2006).

62 Jang, M.K. *et al*., The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol Cell* 19 (4), 523-534 (2005).

63 Krogan, N.J. *et al*., The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* 11 (3), 721-729 (2003).

64 Carrozza, M.J. *et al*., Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123 (4), 581-592 (2005).

65 Joshi, A.A. & Struhl, K., Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol Cell* 20 (6), 971-978 (2005).

66 Keogh, M.C. *et al*., Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123 (4), 593-605 (2005).

67 Ding, L. *et al*., A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* 4 (5), 403-415 (2009).

68 Pasini, D. *et al*., Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes & development* 22 (10), 1345-1355 (2008).

69 Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., & Helin, K., Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *The EMBO journal* 23 (20), 4061-4071 (2004).

70 Yeap, L.S., Hayashi, K., & Surani, M.A., ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage. *Epigenetics & chromatin* 2 (1), 12 (2009).

71 Yuan, P. *et al*., Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes & development* 23 (21), 2507-2520 (2009).

72 Bannister, A.J. *et al*., Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410 (6824), 120-124 (2001).

73 Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., & Jenuwein, T., Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410 (6824), 116-120 (2001).

74 Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., & Grewal, S.I., Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* 292 (5514), 110-113 (2001).

75 Cammas, F., Herzog, M., Lerouge, T., Chambon, P., & Losson, R., Association of the transcriptional corepressor TIF1beta with heterochromatin protein 1 (HP1): an essential role for progression through differentiation. *Genes Dev* 18 (17), 2147-2160 (2004).

76 Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., & Rauscher, F.J., 3rd, SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* 16 (8), 919-932 (2002).

77 Margueron, R. *et al*., Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461 (7265), 762-767 (2009).

78 Bernstein, B.E. *et al*., A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125 (2), 315-326 (2006).

79 Schuettengruber, B. & Cavalli, G., Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* 136 (21), 3531-3542 (2009).

80 Stock, J.K. *et al*., Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* 9 (12), 1428-1435 (2007).

81 Azuara, V. *et al*., Chromatin signatures of pluripotent cell lines. *Nature cell biology* 8 (5), 532-538 (2006).

82 Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., & Helin, K., Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & development* 20 (9), 1123-1136 (2006).

83 Endoh, M. *et al*., Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* 135 (8), 1513-1524 (2008).

84 Shen, X. *et al*., Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* 139 (7), 1303-1314 (2009).

85 Landeira, D. *et al*., Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nature cell biology* 12 (6), 618-624 (2010).

86 Pan, G. *et al*., Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell stem cell* 1 (3), 299-312 (2007).

87 Pasini, D. *et al*., JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464 (7286), 306-310 (2010).

88 Peng, J.C. *et al*., Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139 (7), 1290-1302 (2009).

89 Li, G. *et al*., Jarid2 and PRC2, partners in regulating gene expression. *Genes & development* 24 (4), 368-380 (2010).

90 van der Stoop, P. *et al*., Ubiquitin E3 ligase Ring1b/Rnf2 of polycomb repressive complex 1 contributes to stable maintenance of mouse embryonic stem cells. *PloS one* 3 (5), e2235 (2008).

91 Guenther, M.G. & Young, R.A., Repressive transcription. *Science* 329 (5988), 150-151 (2010).

92 Pasini, D. *et al*., Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res* 38 (15), 4958-4969 (2010).

93 Tie, F. *et al*., CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* 136 (18), 3131-3141 (2009).

94 Lee, M.G. *et al*., Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* 318 (5849), 447-450 (2007).

95 Di Stefano, L., Ji, J.Y., Moon, N.S., Herr, A., & Dyson, N., Mutation of Drosophila Lsd1 disrupts H3-K4 methylation, resulting in tissue-specific defects during development. *Curr Biol* 17 (9), 808-812 (2007).

96 Agger, K. *et al*., UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* 449 (7163), 731-734 (2007).

97 Simon, J.A. & Kingston, R.E., Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* 10 (10), 697-708 (2009).

98 Jaenisch, R. & Young, R., Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* 132 (4), 567-582 (2008).

99 Jones, P.A. & Baylin, S.B., The epigenomics of cancer. *Cell* 128 (4), 683-692 (2007).

100 Goll, M.G. & Bestor, T.H., Eukaryotic cytosine methyltransferases. *Annual review of biochemistry* 74, 481-514 (2005).

101 Jackson, M. *et al*., Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Molecular and cellular biology* 24 (20), 8862-8871 (2004).

102 Fouse, S.D. *et al*., Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell stem cell* 2 (2), 160-169 (2008).

103 Ng, R.K. *et al*., Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nature cell biology* 10 (11), 1280-1290 (2008).

104 Nan, X. *et al*., Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393 (6683), 386-389 (1998).

105 Jones, P.L. *et al*., Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 19 (2), 187-191 (1998).

106 Ng, H.H. *et al*., MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* 23 (1), 58-61 (1999).

107 Feng, Q. & Zhang, Y., The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev* 15 (7), 827-832 (2001).

108 Whyte, W.A. *et al*., Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482 (7384), 221-225 (2012).

109 Feldman, N. *et al*., G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* 8 (2), 188-194 (2006).

110 Epsztejn-Litman, S. *et al*., De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature structural & molecular biology* 15 (11), 1176-1183 (2008).

111 Shi, Y. *et al*., Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119 (7), 941-953 (2004).

112 Heintzman, N.D. *et al*., Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39 (3), 311-318 (2007).

113 Pardo, M. *et al*., An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* 6 (4), 382-395 (2010).

114 van den Berg, D.L. *et al*., An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* 6 (4), 369-381 (2010).

115 Forneris, F., Binda, C., Vanoni, M.A., Battaglioli, E., & Mattevi, A., Human histone demethylase LSD1 reads the histone code. *J Biol Chem* 280 (50), 41360-41365 (2005).

116 Lee, M.G. *et al*., Functional interplay between histone demethylase and deacetylase enzymes. *Mol Cell Biol* 26 (17), 6395-6402 (2006).

117 Wang, J. *et al*., The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat Genet* 41 (1), 125-129 (2009).

118 Musri, M.M. *et al*., Histone demethylase LSD1 regulates adipogenesis. *J Biol Chem* (2010).

119 Choi, J. *et al*., Histone demethylase LSD1 is required to induce skeletal muscle differentiation by regulating myogenic factors. *Biochem Biophys Res Commun* 401 (3), 327-332 (2010).

120 Adamo, A. *et al*., LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nat Cell Biol* 13 (6), 652-660 (2011).

121 Macfarlan, T.S. *et al*., Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev* 25 (6), 594-607 (2011).

122 Feldman, N. *et al*., G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nature cell biology* 8 (2), 188-194 (2006).

123 Lan, F. *et al*., A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* 449 (7163), 689-694 (2007).

124 Hong, S. *et al*., Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proc Natl Acad Sci U S A* 104 (47), 18439-18444 (2007).

125 Lee, S., Lee, J.W., & Lee, S.K., UTX, a Histone H3-Lysine 27 Demethylase, Acts as a Critical Switch to Activate the Cardiac Developmental Program. *Dev Cell* 22 (1), 25-37 (2011).

126 Ho, L. *et al*., 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 (13), 5181-5186 (2009).

127 Kooistra, S.M. & Helin, K., Molecular mechanisms and potential functions of histone demethylases. *Nat Rev Mol Cell Biol* 13 (5), 297-311 (2012).

128 Jiang, H. *et al*., Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell* 144 (4), 513-525 (2011).

129 Graf, T. & Enver, T., Forcing cells to change lineages. *Nature* 462 (7273), 587-594 (2009).

130 Vierbuchen, T. *et al*., Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463 (7284), 1035-1041 (2010).

131 Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., & Melton, D.A., In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* 455 (7213), 627-632 (2008).

132 Leeb, M. *et al*., Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev* 24 (3), 265-276 (2010).

133 Meissner, A. *et al*., Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454 (7205), 766-770 (2008).

134 Xu, J. *et al*., Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes & development* 23 (24), 2824-2838 (2009).

135 Zhou, L. *et al*., An inducible enhancer required for Il12b promoter activity in an insulated chromatin environment. *Mol Cell Biol* 27 (7), 2698-2712 (2007).

136 Bossard, P. & Zaret, K.S., GATA transcription factors as potentiators of gut endoderm differentiation. *Development* 125 (24), 4909-4917 (1998).

137 Gualdi, R. *et al*., Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev* 10 (13), 1670-1682 (1996).

138 Cirillo, L.A. *et al*., Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* 9 (2), 279-289 (2002).

Box 1: The Embryonic Stem Cell State

Embryonic stem (ES) cells are pluripotent, self-renewing cells that are derived from the ICM of the developing blastocyst[1-3]. Pluripotency is the capacity of a single cell to generate all cell lineages of the developing and adult organism. Self-renewal is the ability of a cell to proliferate in the same state. ES cells have a gene expression program that allows them to self-renew yet remain poised to differentiate into essentially all cell types in response to developmental cues. The ES cell gene expression program is governed by transcription factors, signaling pathways, chromatin regulators, and noncoding RNAs. These four classes of regulators control the embryonic stem cell state[4].

Transcription factors recognize and bind specific DNA sequences to either activate or prevent transcription[5]. In ES cells, the gene expression program is largely governed by the core transcription factors Oct4, Sox2, and Nanog[6-8]. The core transcription factors function together to positively regulate their own promoters, forming an interconnected autoregulatory loop. This keeps ES cells in a state that is maintained by a positive-feedback-controlled gene expression program when the transcription factors are expressed at normal levels. This state is no longer maintained and differentiates when one of the core transcription factors is no longer functioning.

The core transcription factors activate a large fraction of the actively transcribed protein-coding and miRNA genes in ES cells[9,10]. Oct4, Sox2 and Nanog also

occupy repressed genes encoding lineage-specific transcription factors, and the repression of these genes is essential for ES cells to maintain a stable pluripotent state[10-17]. The loss of these core transcription factors leads to activation of these developmental genes, indicating that these genes are poised for activation. In addition to Oct4, Sox2 and Nanog, other transcription factors have been shown to play important roles in control of ES cell state[18-28]. Thus, the functions of the core transcription factors are augmented by the functions of many transcription factors implicated in control of ES cell state at actively transcribed genes.

For ES cells, maintenance of the pluripotent state is dependent on the absence of signal that stimulate differentiation[8,29]. Murine ES cells are initially cultured on a layer of fibroblasts in order to obtain factors necessary for self-renewal and pluripotency[30,31]. LIF, Wnt, and ligands of the TGF-β/BMP signaling pathway are among factors supplied by the fibroblasts and found to influence the murine ES cell state[32-36]. The transcription factors associated with the LIF, Wnt, and BMP4 signaling pathways (Stat3, Tcf3, and Smad1) tend to co-occupy enhancers bound by Oct4, Sox2, and Nanog, thereby allowing direct control of genes within the core circuitry by these signaling pathways[9,22,23,37-39]. Loss of Oct4 leads to a loss of these signaling transcription factors at Oct4-bound enhancers.

Transcription factors can bind to chromatin regulators that alter chromatin structure and influence gene expression[40-45]. Chromatin regulators generally fall into three classes: histone-modifying proteins, ATP-dependent nucleosome

remodeling protein complexes, and DNA methyltransferases. These regulators play important roles in ES cells, and is the focus of my thesis.

Non-coding RNAs (ncRNAs) have long been implicated in the regulation of gene expression[5,46], and are critical in the control of ES cell state[10,47,48]. These include miRNAs, which can regulate messenger RNAs and play essential roles in self-renewal and differentiation[10]. They also include longer ncRNAs of various types, which have been implicated in recruitment of chromatin regulators such as the PcG complexes[47-53].

ES cells provide a powerful system for discovering the regulators and mechnasms that control cell states, and serve as a resource for understanding the changes that occur as cells differentiate.

References

1 Thomson, J.A. et al., Embryonic stem cell lines derived from human blastocysts. *Science* 282 (5391), 1145-1147 (1998).

2 Evans, M.J. & Kaufman, M.H., Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292 (5819), 154-156 (1981).

3 Martin, G.R., Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci* U S A 78 (12), 7634-7638 (1981).

4 Young, R.A., Control of the embryonic stem cell state. *Cell* 144 (6), 940-954 (2011).

5 Jacob, F. & Monod, J., Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3, 318-356 (1961).

6 Chambers, I. & Smith, A., Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* 23 (43), 7150-7160 (2004).

7 Niwa, H., How is pluripotency determined and maintained? Development 134 (4), 635-646 (2007).

8 Silva, J. & Smith, A., Capturing pluripotency. *Cell* 132 (4), 532-536 (2008).

9 Chen, X. et al., Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133 (6), 1106-1117 (2008).

10 Marson, A. et al., Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134 (3), 521-533 (2008).

11 Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., & Young, R.A., SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev* 23 (21), 2484-2489 (2009).

12 Boyer, L.A. et al., Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122 (6), 947-956 (2005).

13 Boyer, L.A. et al., Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441 (7091), 349-353 (2006).

14 Lee, T.I. et al., Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125 (2), 301-313 (2006).

15 Pasini, D. et al., Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev* 22 (10), 1345-1355 (2008).

16 Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., & Helin, K., Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO* J 23 (20), 4061-4071 (2004).

17 Loh, Y.H. et al., The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 38 (4), 431-440 (2006).

18 Cartwright, P. et al., LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* 132 (5), 885-896 (2005).

19 Rahl, P.B. et al., c-Myc regulates transcriptional pause release. *Cell* 141 (3), 432-445 (2010).

20 Ivanova, N. et al., Dissecting self-renewal in stem cells with RNA interference. *Nature* 442 (7102), 533-538 (2006).

21 Zhang, X., Zhang, J., Wang, T., Esteban, M.A., & Pei, D., Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *J Biol Chem* 283 (51), 35825-35833 (2008).

22 Wu, Q. et al., Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem* 281 (34), 24090-24094 (2006).

23 Zhang, J. et al., Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat Cell Biol* 8 (10), 1114-1123 (2006).

24 Niwa, H., Ogawa, K., Shimosato, D., & Adachi, K., A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* 460 (7251), 118-122 (2009).

25 Galan-Caridad, J.M. et al., Zfx controls the self-renewal of embryonic and hematopoietic stem cells. *Cell* 129 (2), 345-357 (2007).

26 Dejosez, M. et al., Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. *Genes Dev* 24 (14), 1479-1484 (2010).

27 Jiang, J. et al., A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat Cell Biol* 10 (3), 353-360 (2008).

28 Chia, N.Y. et al., A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* 468 (7321), 316-320 (2010).

29 Pera, M.F. & Tam, P.P., Extrinsic regulation of pluripotent stem cells. *Nature* 465 (7299), 713-720 (2010).

30 Smith, A.G., Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* 17, 435-462 (2001).

31 Smith, T.A. & Hooper, M.L., Medium conditioned by feeder cells inhibits the differentiation of embryonal carcinoma cultures. *Exp Cell Res* 145 (2), 458-462 (1983).

32 Okita, K. & Yamanaka, S., Intracellular signaling pathways regulating pluripotency of embryonic stem cells. *Curr Stem Cell Res Ther* 1 (1), 103-111 (2006).

33 Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., & Brivanlou, A.H., Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med* 10 (1), 55-63 (2004).

34 Smith, S.K., Charnock-Jones, D.S., & Sharkey, A.M., The role of leukemia inhibitory factor and interleukin-6 in human reproduction. *Hum Reprod* 13 Suppl 3, 237-243; discussion 244-236 (1998).

35 Williams, R.L. et al., Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* 336 (6200), 684-687 (1988).

36 Ying, Q.L., Nichols, J., Chambers, I., & Smith, A., BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* 115 (3), 281-292 (2003).

37 Chen, X., Vega, V.B., & Ng, H.H., Transcriptional regulatory networks in embryonic stem cells. *Cold Spring Harb Symp Quant Biol* 73, 203-209 (2008).

38 Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., & Young, R.A., Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev* 22 (6), 746-755 (2008).

39 Tam, W.L. et al., T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. *Stem Cells* 26 (8), 2019-2031 (2008).

40 Kornberg, R.D. & Thomas, J.O., Chromatin structure; oligomers of the histones. *Science* 184 (4139), 865-868 (1974).

41 Olins, A.L. & Olins, D.E., Spheroid chromatin units (v bodies). *Science* 183 (4122), 330-332 (1974).

42 Brownell, J.E. et al., Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* 84 (6), 843-851 (1996).

43 Cote, J., Quinn, J., Workman, J.L., & Peterson, C.L., Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* 265 (5168), 53-60 (1994).

44 Imbalzano, A.N., Kwon, H., Green, M.R., & Kingston, R.E., Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* 370 (6489), 481-485 (1994).

45 Kwon, H., Imbalzano, A.N., Khavari, P.A., Kingston, R.E., & Green, M.R., Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature* 370 (6489), 477-481 (1994).

46 Britten, R.J. & Davidson, E.H., Gene regulation for higher cells: a theory. *Science* 165 (3891), 349-357 (1969).

47 Guenther, M.G. & Young, R.A., Repressive transcription. *Science* 329 (5988), 150-151 (2010).

48 Surface, L.E., Thornton, S.R., & Boyer, L.A., Polycomb group proteins set the stage for early lineage commitment. *Cell Stem Cell* 7 (3), 288-298 (2010).

49 Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., & Helin, K., Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* 20 (9), 1123-1136 (2006).

50 Wilusz, J.E., Sunwoo, H., & Spector, D.L., Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 23 (13), 1494-1504 (2009).

51 Zhao, X.Y. et al., iPS cells produce viable mice through tetraploid complementation. *Nature* 461 (7260), 86-90 (2009).

52 Zhao, J. et al., Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* 40 (6), 939-953 (2010).

53 Kanhere, A. et al., Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol Cell* 38 (5), 675-688 (2010).

Table 1: **Key chromatin regulators implicated in control of ES cell state**

Family	Complex or protein	Function*	Key References
Histone-modifying complexes and proteins			
Polycomb (PcG)	PRC2	Differentiation	1-3
	PRC1	Maintenance	4-7
	Jarid2	Differentiation	8-12
	Pcl2	Differentiation	13
Trithorax (TrxG)	MLL	Differentiation	14,15
	Wdr5	Maintenance	16
	Ash2L	Maintenance	16
	Dpy30	Differentiation	17
Eset	SetDB1	Maintenance	18-20
G9a/Glp	Ehmt2	Differentiation	21,22
	Ehmt1	Maintenance	22
Dot1	Dot1L	Differentiation	23
Set2	Setd2	Differentiation	24
Histone Deacetylase	Hdac1	Differentiation	25
Histone Demethylase	LSD1	Differentiation	26-30
	Jmjd1a	Maintenance	31
	Jmjd2c	Maintenance	31
	Jarid1a	Differentiation	32
	UTX	Differentiation	33
Histone Arginine Methylation	Carm1/Prmt4	Maintenance	34,35
	Prmt5	Maintenance	36
	Prmt6	Differentiation	37
Histone Clipping	Cathepsin L	Differentiation	38,39
ATP-dependent Nucleosome Remodeling			
SWI/SNF	esBAF	Maintenance	40-44
CHD	NuRD	Differentiation	26,45,46
	Chd1	Maintenance	47
ISWI	NURF	Differentiation	48

INO80	Tip60-p400	Maintenance	49
DNA Methyltransferases			
DNMT	Dnmt3a/b	Differentiation	50,51
	Dnmt1	Differentiation	50,51
Cofactors			
Mediator	Mediator	Maintenance	21,52
SMC	Cohesin	Maintenance	21,49,52
KAP	Trim28	Maintenance	21,49
PAF	Paf1	Maintenance	53
CBP/p300	p300	Differentiation	54,55

* - Reduced levels or activity of these regulators result in either loss of the ES cell gene expressi
therefore these regulators function in the maintenance of cell state) or defects in ES cell differen

References

1 Shen, X. *et al*., EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell* 32 (4), 491-502 (2008).

2 Chamberlain, S.J., Yee, D., & Magnuson, T., Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* 26 (6), 1496-1505 (2008).

3 Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., & Helin, K., The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* 27 (10), 3769-3779 (2007).

4 Endoh, M. *et al*., Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* 135 (8), 1513-1524 (2008).

5 Roman-Trufero, M. *et al*., Maintenance of undifferentiated state and self-renewal of embryonic neural stem cells by Polycomb protein Ring1B. *Stem Cells* 27 (7), 1559-1570 (2009).

6 van der Stoop, P. *et al*., Ubiquitin E3 ligase Ring1b/Rnf2 of polycomb repressive complex 1 contributes to stable maintenance of mouse embryonic stem cells. *PLoS One* 3 (5), e2235 (2008).

7 Voncken, J.W. *et al*., Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci U S A* 100 (5), 2468-2473 (2003).

8 Peng, J.C. *et al*., Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139 (7), 1290-1302 (2009).

9 Shen, X. *et al*., Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* 139 (7), 1303-1314 (2009).

10 Landeira, D. *et al*., Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat Cell Biol* 12 (6), 618-624 (2010).

11 Li, G. *et al*., Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev* 24 (4), 368-380 (2010).

12 Pasini, D. *et al*., JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464 (7286), 306-310 (2010).

13 Walker, E. *et al*., Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* 6 (2), 153-166 (2010).

14 Glaser, S. *et al*., Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* 133 (8), 1423-1432 (2006).

15 Lubitz, S., Glaser, S., Schaft, J., Stewart, A.F., & Anastassiadis, K., Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. *Mol Biol Cell* 18 (6), 2356-2366 (2007).

16 Ang, Y.S. *et al*., Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* 145 (2), 183-197 (2011).

17 Jiang, H. *et al*., Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell* 144 (4), 513-525 (2011).

18 Bianchi, T.I., Aviles, G., & Sabattini, M.S., Biological characteristics of an enzootic subtype of western equine encephalomyelitis virus from Argentina. *Acta Virol* 41 (1), 13-20 (1997).

19 Yeap, L.S., Hayashi, K., & Surani, M.A., ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage. *Epigenetics Chromatin* 2 (1), 12 (2009).

20 Yuan, P. *et al*., Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes Dev* 23 (21), 2507-2520 (2009).

21 Hu, G. *et al*., A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev* 23 (7), 837-848 (2009).

22 Feldman, N. *et al*., G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* 8 (2), 188-194 (2006).

23 Barry, E.R. *et al*., ES cell cycle progression and differentiation require the action of the histone methyltransferase Dot1L. *Stem Cells* 27 (7), 1538-1547 (2009).

24 Hu, M. *et al*., Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *Proc Natl Acad Sci U S A* 107 (7), 2956-2961 (2010).

25 Dovey, O.M., Foster, C.T., & Cowley, S.M., Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proc Natl Acad Sci U S A* 107 (18), 8242-8247 (2010).

26 Whyte, W.A. *et al*., Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482 (7384), 221-225 (2012).

27 Adamo, A. *et al*., LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nat Cell Biol* 13 (6), 652-660 (2011).

28 Macfarlan, T.S. *et al*., Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev* 25 (6), 594-607 (2011).

29 Foster, C.T. *et al*., Lysine-specific demethylase 1 regulates the embryonic transcriptome and CoREST stability. *Mol Cell Biol* 30 (20), 4851-4863 (2010).

30 Wang, J. *et al*., The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat Genet* 41 (1), 125-129 (2009).

31 Loh, Y.H., Zhang, W., Chen, X., George, J., & Ng, H.H., Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev* 21 (20), 2545-2557 (2007).

32 Pasini, D. *et al*., Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev* 22 (10), 1345-1355 (2008).

33 Lee, S., Lee, J.W., & Lee, S.K., UTX, a Histone H3-Lysine 27 Demethylase, Acts as a Critical Switch to Activate the Cardiac Developmental Program. *Dev Cell* 22 (1), 25-37 (2011).

34 Wu, Q. *et al*., CARM1 is required in embryonic stem cells to maintain pluripotency and resist differentiation. *Stem Cells* 27 (11), 2637-2645 (2009).

35 Torres-Padilla, M.E., Parfitt, D.E., Kouzarides, T., & Zernicka-Goetz, M., Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* 445 (7124), 214-218 (2007).

36 Tee, W.W. *et al*., Prmt5 is essential for early mouse development and acts in the cytoplasm to maintain ES cell pluripotency. *Genes Dev* 24 (24), 2772-2777 (2010).

37 Lee, Y.H. *et al*., Protein arginine methyltransferase 6 regulates Embryonic Stem cell identity. *Stem Cells Dev* (2012).

38 Duncan, J.A. *et al*., Neisseria gonorrhoeae activates the proteinase cathepsin B to mediate the signaling activities of the NLRP3 and ASC-containing inflammasome. *J Immunol* 182 (10), 6460-6469 (2009).

39 Duncan, E.M. *et al*., Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation. *Cell* 135 (2), 284-294 (2008).

40 Ho, L. *et al*., 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 (13), 5187-5191 (2009).

41 Ho, L. *et al*., 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 (13), 5181-5186 (2009).

42 Kidder, B.L., Palmer, S., & Knott, J.G., SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells* 27 (2), 317-328 (2009).

43 Yan, Z. *et al*., BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells. *Stem Cells* 26 (5), 1155-1165 (2008).

44 Gao, X. *et al*., ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc Natl Acad Sci U S A* 105 (18), 6656-6661 (2008).

45 Kaji, K. *et al*., The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol* 8 (3), 285-292 (2006).

46 Kaji, K., Nichols, J., & Hendrich, B., Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. *Development* 134 (6), 1123-1132 (2007).

47 Gaspar-Maia, A. *et al*., Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* 460 (7257), 863-868 (2009).

48 Landry, J. *et al*., Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. *PLoS Genet* 4 (10), e1000241 (2008).

49 Fazzio, T.G., Huff, J.T., & Panning, B., An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 134 (1), 162-174 (2008).

50 Jackson, M. *et al*., Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol Cell Biol* 24 (20), 8862-8871 (2004).

51 Tsumura, A. *et al*., Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* 11 (7), 805-814 (2006).

52 Kagey, M.H. *et al*., Mediator and cohesin connect gene expression and chromatin architecture. *Nature* (2010).

53 Ding, L. *et al*., A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* 4 (5), 403-415 (2009).

54 Zhong, X. & Jin, Y., Critical roles of coactivator p300 in mouse embryonic stem cell differentiation and Nanog expression. *J Biol Chem* 284 (14), 9168-9175 (2009).

55 Yao, T.P. *et al*., Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* 93 (3), 361-372 (1998).

Chapter 2

Enhancer Decommissioning By LSD1 During Embryonic Stem Cell Differentiation

Warren A. Whyte[1,2*], Steve Bilodeau[1*], David A. Orlando[1], Heather A. Hoke[1,2], Garrett M. Frampton[1,2], Charles T. Foster[3,4], Shaun M. Cowley[4], and Richard A. Young[1,2]

[1] Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2] Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3] Department of Molecular Biology, Adolf-Butenandt Institut, Ludwig-Maximilians-Universität München, Munich, Germany 80336. [4] Department of Biochemistry, University of Leicester, Leicester LE1 9HN, United Kingdom.

*These authors contributed equally to this work

Personal Contribution to the Project

This chapter is the result of a close collaboration with Steve Bilodeau on work that was published in 2012 in Nature. This work was inspired by prior work (Bilodeau et al., 2009). I suggested that we examine the role of LSD1 in ES cell differentiation. I conceived the initial experiments to test if LSD1 could promote the silencing of active enhancers belonging to the ES cell gene expression program, and conducted the experiments described here. Steve and I prepared the manuscript.

Summary

Transcription factors and chromatin modifiers play important roles in programming and reprogramming of cellular states during development[1,2]. Transcription factors bind to enhancer elements and recruit coactivators and chromatin modifying enzymes to facilitate transcription initiation[3,4]. During differentiation, a subset of these enhancers must be silenced, but the mechanisms underlying enhancer silencing are poorly understood. Here we show that the H3K4/K9 histone demethylase LSD1 (ref. 5) plays an essential role in decommissioning enhancers during differentiation of embryonic stem cells (ESCs). LSD1 occupies enhancers of active genes critical for control of ESC state. However, LSD1 is not essential for maintenance of ESC identity. Instead, ESCs lacking LSD1 activity fail to fully differentiate and ESC-specific enhancers fail to undergo the histone demethylation events associated with differentiation. At active enhancers, LSD1 is a component of the NuRD complex, which contains additional subunits that are necessary for ESC differentiation. We propose that the LSD1-NuRD complex decommissions enhancers of the pluripotency program upon differentiation, which is essential for complete shutdown of the ESC gene expression program and the transition to new cell states.

The histone H3K4/K9 demethylase LSD1 (Lysine-specific-demethylase-1, KDM1A) is among the chromatin regulators that have been implicated in control of early embryogenesis[6-8]. Loss of LSD1 leads to embryonic lethality and ESCs lacking LSD1 function fail to differentiate into embryoid bodies[6-8]. These results suggest that LSD1 contributes to changes in chromatin that are critical to differentiation of ESCs, but LSD1's role in this process is not yet understood. To investigate the function of LSD1 in ESCs, we first identified the sites it occupies in the genome using chromatin immunoprecipitation coupled with massively parallel DNA sequencing (ChIP-Seq; Fig. 1, and Supplementary Fig. 1). The results revealed that LSD1 occupies the enhancers and core promoters of a substantial population of actively transcribed and bivalent genes (Fig. 1a, b, and Supplementary Table 1). Inspection of individual gene tracks showed that LSD1 co-occupies well-characterized enhancer regions with the ESC master transcription factors Oct4, Sox2 and Nanog and the Mediator coactivator (Fig. 1b, and Supplementary Fig. 1). Loci bound by Oct4, Sox2 and Nanog are generally associated with Mediator and p300 coactivators and have enhancer activity[9,10]. A global view of Oct4, Sox2, Nanog and Mediator -occupied enhancer regions confirmed that 97% of the 3,838 high confidence enhancers were co-occupied by LSD1 ($p < 10^{-9}$) (Fig. 1c, and Supplementary Table 2). This is consistent with evidence that LSD1 can interact with Oct4 (refs. 11,12). LSD1 signals were also observed at core promoter regions with RNA Polymerase II (Pol II) and TBP (Fig. 1d). The density of LSD1 signals at enhancers was higher than at core promoters ($p < 10^{-16}$; Supplementary Fig. 1), indicating that LSD1 is associated predominantly with the enhancers of actively transcribed genes in ESCs.

It was striking to find that LSD1 is associated with active genes in ESCs because previous studies have shown that LSD1 is not essential for maintenance of ESC state, but is required for normal differentiation[6-8]. We used an ESC differentiation assay to further investigate the involvement of LSD1 in cell state transitions (Fig. 2a, b). Prolonged depletion of Oct4 in ZHBTc4 ESCs using doxycycline causes loss of pluripotency and differentiation into trophectoderm[13]. As expected, loss of Oct4 expression led to a rapid loss of ESC morphology and a marked reduction in SSEA-1 and alkaline phosphatase, two markers of ESCs (Fig. 2c, and Supplementary Fig. 2). When these ESCs were treated with the LSD1 inhibitor tranylcypromine (TCP) during Oct4 depletion, they failed to undergo the morphological changes associated with differentiation of ESCs (Fig. 2c). Instead, the TCP-treated cells formed small colonies resembling those of untreated ESCs and maintained expression of SSEA-1 and alkaline phosphatase (Fig. 2c, and Supplementary Fig. 3). Very similar results were obtained in LSD1 knockout ESCs (Supplementary Fig. 4, 5), and in cells treated with another LSD1 inhibitor, pargyline (Prg), or an shRNA against LSD1 (Supplementary Fig. 2, 3). LSD1 inhibition also caused an increase in cell death during differentiation, as has been observed with cells lacking LSD1 in other assays[7,8]. These results suggest that LSD1 may be required for ESCs to completely silence the ESC gene expression program.

Further analysis of ESCs forced to differentiate in the absence of LSD1 activity confirmed that these cells failed to fully transition from the ESC gene expression

program; while key genes of the trophectoderm gene expression program were activated, including *Cdx2* and *Esx1* (ref. 14), there was incomplete repression of many ESC genes, including *Sox2* and *Fbox15* (Fig. 2d). A global analysis confirmed that a set of genes neighboring LSD1-occupied enhancers in ESCs are repressed upon differentiation, and that repression of this set of genes is partially relieved in the presence of TCP (Fig. 2e, and Supplementary Table 3). Similar results were obtained with LSD1 knockout cells (Supplementary Fig. 4, 5), and cells treated with either pargyline or an shRNA against LSD1 (Supplementary Fig. 3). These results indicate that the trophectoderm differentiation program can be induced in cells lacking LSD1 function, but the ESC program is not fully silenced in these cells.

To gain further insight into the role of LSD1 in ESC differentiation, we investigated whether LSD1 is associated with previously described complexes, including NuRD (Nucleosome remodeling and histone deacetylase), CoREST (Cofactor of REST) and the AR/ER (Androgen receptor/ Estrogen receptor) complexes[8,15-17]. We first studied whether the LSD1 found at Oct4-occupied genes is a component of NuRD because Oct4 and Nanog have been reported to interact with several components of NuRD[11,12,18]. ChIP-Seq experiments confirmed that NuRD subunits Mi-2β, HDAC1 and HDAC2 co-occupy sites with LSD1 at enhancers ($p < 10^{-9}$; Fig. 3 and Supplementary Table 1). Immunoprecipitation of LSD1 confirmed its association with Mi-2β, HDAC1 and HDAC2 (Fig. 3b, c). We then investigated whether LSD1 is associated with CoREST; ChIP-Seq data revealed that a minor fraction of LSD1 co-occupies sites with CoREST and Rest (2% and 6%,

respectively)(Supplementary Fig. 6 and Supplementary Table 1). As expected, LSD1-REST sites were frequently found associated with neuronal genes (Supplementary Fig. 7 and Supplementary Table 4). Immunoprecipitation experiments confirmed that LSD1 is associated with CoREST (Fig. 3b, c). AR and ER are not expressed in ESCs based on the lack of histone H3K79me2 and H3K36me3 (modifications associated with transcriptional elongation) at the genes encoding these proteins (Supplementary Table 1). Further examination of the ChIP-Seq data revealed that enhancers were significantly more likely to be occupied by the LSD1 and NuRD proteins as compared to REST and CoREST ($p < 10^{-9}$) (Fig. 3d and Supplementary Fig. 8). Multiple components of NuRD are dispensable for ESC state but required for normal differentiation[6,19-21]. ESCs with reduced levels of the core NuRD ATPase Mi-2β failed to differentiate properly and partially maintained expression of SSEA-1, alkaline phosphatase and ESC genes (Supplementary Fig. 9), which are the same phenotypes we observed with reduced levels of LSD1. These results indicate that LSD1 at enhancers is associated with a NuRD complex that is essential for normal cell state transitions.

Nucleosomes with histone H3K4me1 are commonly found at enhancers of active genes, and are a substrate for LSD1 (refs. 5,22). If LSD1-dependent H3K4me1 demethylase activity is involved in enhancer silencing during ESC differentiation, LSD1 inhibition should cause retention of H3K4me1 levels at active ESC enhancers when differentiation is induced. During trophectoderm differentiation with control ESCs, we found p300 and H3K27ac levels reduced at a set of active ESC enhancers,

suggesting these enhancers were being silenced (Supplementary Fig. 10). The levels of H3K4me1 at enhancers were also reduced, as seen for example at *Lefty1* (Fig. 4a, and Supplementary Table 5), while the levels of H3K4me1 increased at newly active trophectoderm genes such as *Gata2* (Fig. 4b). In contrast, H3K4me1 signals were higher at LSD1- occupied enhancers in differentiating ESCs treated with TCP than in control cells, including *Lefty1 and Sox2* (Fig. 4a, c). The majority of enhancers (1,722 of 2,755) that were occupied by LSD1 and that experienced reduced levels of H3K4me1 during differentiation retained H3K4me1 in TCP-treated ESCs compared to untreated control differentiating ESCs (Fig. 4d, e). These results are consistent with the model that LSD1 demethylates H3K4me1 at the enhancers of ESC-specific genes during differentiation, and that this activity is essential to fully repress the genes associated with these enhancers.

Our results indicate that an LSD1-NuRD complex is required for silencing of ESC enhancers during differentiation, which is essential for complete shutdown of the ESC gene expression program and the transition to new cell states. These results, together with those of previous studies on NuRD function[18,21,23,24], suggest the following model for LSD1-NuRD in enhancer decommissioning. LSD1-NuRD complexes occupy Oct4-regulated active enhancers in ESCs, but do not substantially demethylate histone H3K4 because LSD1's H3K4 demethylase activity is inhibited in the presence of acetylated histones[23,24]. Enhancers occupied by Oct4, Sox2 and Nanog are co-occupied by the HAT p300 and nucleosomes with acetylated histones (Supplementary Fig. 10 and ref. 10). Thus, as long as the enhancer-bound

transcription factors recruit HATs to enhancers, the net effect of having both HATs and NuRD-associated HDACs present is to have sufficient levels of acetylated histones to suppress LSD1 demethylase activity. During ESC differentiation, the levels of Oct4 and p300 are reduced, thus reducing the level of acetylated histones, which in turn permits demethylation of H3K4 by LSD1. Consistent with this model, we find that the shutdown of Oct4 leads to reduced levels of p300 and histone H3K27ac at enhancers that are occupied by Oct4 and LSD1 (Supplementary Fig. 10, 11), and this is coincident with reduced levels of methylated H3K4 (Fig. 4, and Supplementary Figs. 12, 13). This model would explain why key components of LSD1-NuRD complexes are not essential for maintenance of ESC state, but are essential for normal differentiation, when the active enhancers must be silenced. Additional HATs expressed in ESCs may also contribute to the dynamic balance of nucleosome acetylation. Future biochemical analysis of HAT, HDAC and demethylase complexes at enhancers will be valuable for testing this model and for further understanding how enhancers are regulated during differentiation.

We conclude that LSD1-NuRD complexes present at active promoters in ESCs are essential for normal differentiation, when the active enhancers must be silenced. Given evidence that LSD1 is required for differentiation of multiple cell types[6,25,26], LSD1 is likely to be generally involved in enhancer silencing during differentiation. The ESC gene expression program can be maintained in the absence of many other chromatin regulators[2], and it is possible that some of these also play key roles in the transition from one transcriptional program to another during differentiation.

Methods Summary

ESC Cell Culture Conditions

ESCs were grown on irradiated murine embryonic fibroblasts (MEFs) and passaged as previously described[9]. In drug treatment experiments, ESCs were split off MEFs and treated with tranylcypromine (TCP, 1mM) or pargyline (Prg, 3mM) to inhibit LSD1 activity. Lentiviral constructs were purchased from Open Biosystems and produced according to the *Trans*-lentiviral shRNA Packaging System (TLP4614).

Differentiation assay, Immunofluorescence, and Alkaline Phosphatase Staining

ZHBTc4 ESCs were split off MEFs in ESC media containing 2μg/ml doxycycline to reduce Oct4 expression levels. For Immunofluorescence, ESCs were crosslinked, blocked and permeabilized before incubation with Oct4 (Santa Cruz, sc-9081x; 1:200 dilution) or SSEA1 (mc-480, DHSB, 1:20 dilution) antibodies. Alexa-conjugated secondary antibodies were used for detection. Staining of ESCs for alkaline phosphatase was achieved using the Alkaline Phosphatase Detection Kit (Millipore, SCR004). Cells were harvested at indicated time points for ChIP-Seq, qPCR or expression array analyses.

ChIP-Seq

Chromatin immunoprecipitations (ChIPs) were performed and analyzed as previously described[9]. The following antibodies were used: LSD1 (Abcam, ab17721), Mi-2b (Abcam, ab72418), HDAC1 (Abcam, ab7028), HDAC2 (Abcam, ab7029), REST

(Millipore, 07-579), CoREST (Abcam, ab32631), H3K4me1 (Abcam, ab8895), p300 (Santa-Cruz, sc-584) and H3K27Ac (Abcam, ab4729).

For ChIP-Seq analyses, reads were aligned with Bowtie and analyzed as described in Supplemental Information.

Full Methods and any associated references are available in Supplementary Information of the paper at www.nature.com/nature.

ACCESSION NUMBERS

ChIP-Seq and GeneChip expression data have been deposited in Gene Expression Omnibus with accession number GSE27844.

Acknowledgements

We thank Jakob Lovén, Michael H. Kagey, Jill Dowen, Alan C. Mullen, Alla Sigova, Peter B. Rahl, Tony Lee, and members of Yang Shi's lab for experimental assistance, reagents and helpful discussions; and Jeong-Ah Kwon, Vidya Dhanapal, Jennifer Love, Sumeet Gupta, Thomas Volkert, Wendy Salmon and Nicki Watson for assistance with ChIP-Seq, expression arrays, and immunofluorescence imaging acquisition. This work was supported by a Canadian Institutes of Health Research Fellowship (SB), a Career Development Award from the Medical Research Council (SMC), and by NIH grants HG002668 and NS055923 (RY).

Author Contributions

The ChIP-Seq, immunofluorescence and expression experiments were designed, conducted and interpreted by W.A.W., S.B., H.A.H., C.T.F., S.M.C and R.A.Y. W.A.W., D.A.O. and G.M.F. performed data analysis. The manuscript was written by S.B., W.A.W. D.A.O., H.A.H., G.M.F. and R.A.Y.

References

1 Graf, T. & Enver, T., Forcing cells to change lineages. *Nature* 462 (7273), 587-594 (2009).

2 Young, R.A., Control of the embryonic stem cell state. *Cell* 144 (6), 940-954 (2011).

3 Fuda, N.J., Ardehali, M.B., & Lis, J.T., Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* 461 (7261), 186-192 (2009).

4 Li, B., Carey, M., & Workman, J.L., The role of chromatin during transcription. *Cell* 128 (4), 707-719 (2007).

5 Shi, Y. *et al*., Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119 (7), 941-953 (2004).

6 Wang, J. *et al*., Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* 446 (7138), 882-887 (2007).

7 Wang, J. *et al*., The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat Genet* 41 (1), 125-129 (2009).

8 Foster, C.T. *et al*., Lysine-specific demethylase 1 regulates the embryonic transcriptome and CoREST stability. *Mol Cell Biol* 30 (20), 4851-4863 (2010).

9 Kagey, M.H. *et al*., Mediator and cohesin connect gene expression and chromatin architecture. *Nature* (2010).

10 Chen, X. *et al*., Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133 (6), 1106-1117 (2008).

11 Pardo, M. *et al*., An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* 6 (4), 382-395 (2010).

12 van den Berg, D.L. *et al*., An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* 6 (4), 369-381 (2010).

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

14 Rossant, J. & Cross, J.C., Placental development: lessons from mouse mutants. *Nat Rev Genet* 2 (7), 538-548 (2001).

15 Metzger, E. *et al*., LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* 437 (7057), 436-439 (2005).

16 Shi, Y.J. *et al*., Regulation of LSD1 histone demethylase activity by its associated factors. *Mol Cell* 19 (6), 857-864 (2005).

17 Wang, Y. *et al*., LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. *Cell* 138 (4), 660-672 (2009).

18 Liang, J. *et al*., Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol* 10 (6), 731-739 (2008).

19 Dovey, O.M., Foster, C.T., & Cowley, S.M., Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proc Natl Acad Sci U S A* 107 (18), 8242-8247 (2010).

20 Kaji, K. *et al*., The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol* 8 (3), 285-292 (2006).

21 Scimone, M.L., Meisel, J., & Reddien, P.W., The Mi-2-like Smed-CHD4 gene is required for stem cell differentiation in the planarian Schmidtea mediterranea. *Development* 137 (8), 1231-1241 (2010).

22 Heintzman, N.D. *et al*., Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39 (3), 311-318 (2007).

23 Forneris, F., Binda, C., Vanoni, M.A., Battaglioli, E., & Mattevi, A., Human histone demethylase LSD1 reads the histone code. *J Biol Chem* 280 (50), 41360-41365 (2005).

24 Lee, M.G. *et al*., Functional interplay between histone demethylase and deacetylase enzymes. *Mol Cell Biol* 26 (17), 6395-6402 (2006).

25 Musri, M.M. *et al*., Histone demethylase LSD1 regulates adipogenesis. *J Biol Chem* (2010).

26 Choi, J. *et al*., Histone demethylase LSD1 is required to induce skeletal muscle differentiation by regulating myogenic factors. *Biochem Biophys Res Commun* 401 (3), 327-332 (2010).

27 Okumura-Nakanishi, S., Saito, M., Niwa, H., & Ishikawa, F., Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *J Biol Chem* 280 (7), 5307-5317 (2005).

28 Yeom, Y.I. *et al*., Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* 122 (3), 881-894 (1996).

29 Nakatake, Y. *et al*., Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. *Mol Cell Biol* 26 (20), 7772-7782 (2006).

30 Ray, S. *et al*., Context-dependent function of regulatory elements and a switch in chromatin occupancy between GATA3 and GATA2 regulate Gata2 transcription during trophoblast differentiation. *The Journal of biological chemistry* 284 (8), 4978-4988 (2009).

FIGURE LEGENDS

Figure 1: LSD1 is associated with enhancer and core promoter regions of active genes in ESCs

a, LSD1 occupies a substantial population of actively transcribed genes in murine ESCs. Pie charts depict active (green), bivalent (yellow) and silent (red) genes, and the proportion (black lines) occupied by either LSD1, Pol II, or the Polycomb protein Suz12 (Supplementary Table 1, and Supplementary Information). The numbers depict the number of genes bound over the total number of genes in each of the active, bivalent, and silent classes. LSD1 ChIP-Seq data is from combined biological replicates using an antibody specific for LSD1 as determined by knockdown experiments (Supplementary Fig. 1). The p-value for each category was determined by a hypergeometric test.

b, LSD1 occupies enhancers and core promoter regions of actively transcribed genes. ChIP-Seq binding profiles (reads/million) for ESC transcription factors (Oct4, Sox2, Nanog), coactivator (Med1), chromatin regulator (LSD1), the transcriptional apparatus (Pol II, TBP) and histone modifications (H3K4me1, H3K4me3, H3K79me2, H3K36me3) at the *Oct4 (Pou5f1)* and *Lefty1* loci in ESCs, with the y-axis floor set to 1. Gene models, and previously described enhancer regions[27-29] are depicted below the binding profiles.

c, LSD1 occupies enhancer sites. Density map of ChIP-Seq data at Oct4, Sox2, Nanog, and Med1 co-occupied enhancer regions. Data is shown for ESC regulators (Oct4), coactivators (Med1 and p300) and a chromatin regulator (LSD1) in ESCs. Enhancers

were defined as Oct4, Sox2, Nanog and Mediator co-occupied regions. Over 96% of the 3,838 high confidence enhancers were co-occupied by LSD1 ($p < 10^{-9}$). Color scale indicates ChIP-seq signal in reads per million.

d, LSD1 occupies core promoter sites. Density map of ChIP-Seq data at transcriptional start sites (TSS) of genes neighboring the 3,838 previously defined enhancers (Fig. 1c). Data is shown for components of the transcription apparatus (Pol II and TBP) and the chromatin regulator LSD1 in ESCs. Core promoters were defined as the closest TSS from each enhancer. Color scale indicates ChIP-Seq signal in reads per million.

Figure 2: LSD1 inhibition results in incomplete silencing of ESC genes during differentiation

a, Schematic representation of trophectoderm differentiation assay using doxycycline-inducible Oct4 shutdown murine ESC line ZHBTc4. Treatment with doxycycline for 48 hours leads to depletion of Oct4 and early trophectoderm specification. Cells were treated with DMSO (control) or the LSD1 inhibitor tranylcypromine (TCP) for 6 hours before 2μg/ml doxycycline was added for an additional 24 or 48 hours.

b, Treatment of ZHBTc4 ESCs with doxycycline leads to loss of Oct4 proteins. Oct4 and LSD1 protein levels in nuclear extracts (NE) determined by Western blot (WB) before and after treatment of ZHBTc4 ESCs with 2μg/ml doxycycline. Tubulin served as loading control.

c, Doxcycline-treated cells treated with TCP maintained SSEA-1 cell surface marker expression. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 100μM

d, Expression of selected ESC and trophectodermal genes 48 hours after Oct4 depletion in DMSO- versus TCP-treated cells (black versus grey bars, respectively). Treatment of TCP partially relieved repression of ESC genes, but did not affect upregulation of trophectodermal genes. Error bars reflect standard deviation from biological replicates.

e, Genes neighboring LSD1-occupied enhancers are less downregulated during ESC differentiation following TCP treatment. Mean fold change in expression of the 630 downregulated (at least 1.25 fold; $p < 0.01$) genes nearest LSD1- occupied enhancers (Fig. 1c) during differentiation of TCP-treated and untreated control cells. Alleviation of repression is significantly higher ($p < 0.005$) for LSD1 enhancer-bound repressed genes compared to all repressed genes.

Figure 3: LSD1 is associated with a NuRD complex at active enhancers in ESCs

a, NuRD components occupy enhancers and core promoter regions of actively transcribed genes. ChIP-Seq binding profiles (reads/million) for transcription factors (Oct4, Sox2, Nanog), coactivator (Med1), and chromatin regulators (LSD1, Mi-2β, HDAC1, HDAC2), at the *Oct4 (Pou5f1)* and *Lefty1* loci in ESCs, with the y-axis floor set to 1. Gene models, and previously described enhancer regions[27-29] are depicted below the binding profiles.

b, LSD1 is associated with NuRD components Mi-2β, HDAC1, HDAC2, as well as CoREST. LSD1 and HDAC1 are detected by Western blot (WB) after immunoprecipitation of crosslinked chromatin using LSD1, HDAC1, HDAC2, Mi-2β, or CoREST antibodies. IgG is shown as a control.

c, LSD1 and HDAC1 are detected by Western blot (WB) after immunoprecipitation of uncrosslinked nuclear extracts (NE) using LSD1, HDAC1, HDAC2, Mi-2β, or CoREST antibodies. IgG is shown as a control.

d, The occupancy of enhancers by NuRD proteins (Mi-2β, HDAC1 and HDAC2) is significantly greater than the occupancy by CoREST or REST ($p < 10^{-9}$). The height of the bars represents the percentage of the 3,838 enhancers co-occupied by LSD1, NuRD proteins (Mi-2β, and either HDAC1 or HDAC2), CoREST and REST.

Figure 4: LSD1 is required for H3K4me1 removal at ESC enhancers

a, H3K4me1 levels are reduced at LSD1-occupied enhancers upon ESC differentiation, and this effect is partially blocked upon TCP treatment. **b,** TCP treatment does not affect the increase in H3K4me1 levels at trophectodermal genes during differentiation. ChIP-Seq binding profiles (reads/million) for Oct4 and LSD1 at the *Lefty1* and *Gata2* loci in ESCs. Below these profiles, histone H3K4me1 levels are shown for ZHBTc4 control ESCs, cells treated with doxycycline for 48 hours to repress Oct4 and induce differentiation (ESCs +Dox), and ESCs treated with doxycycline and TCP (ESCs +Dox,TCP). For appropriate normalization, ChIP-Seq data for histone H3K4me1 is shown as rank normalized reads/million with the y-axis floor set to 1 (Supplementary Information). Gene models, and previously described enhancer regions[29,30] are depicted below the binding profiles.

c, Sum of the normalized H3K4me1 density +/- 250 nucleotides surrounding LSD1-occupied enhancer regions before and during trophectoderm differentiation in the presence or absence of TCP. The associated genes were identified based on their

proximity to the LSD1-occupied enhancers.

d, Sum of the normalized H3K4me1 density +/- 250 nucleotides surrounding 1,722 LSD1-occupied enhancers before and during differentiation in the presence or absence of TCP. Of the 2,755 LSD1-occupied enhancers having reduced levels of H3K4me1 upon differentiation, 63% (1,722) display higher H3K4me1 levels after TCP treatment ($p < 10^{-16}$).

e, Heatmap displaying the sum of the normalized H3K4me1 density +/- 250 nucleotides surrounding the 1,722 LSD1-occupied enhancers that retained H3K4me1 in TCP-treated ESCs compared to untreated control differentiating ESCs. Color scale indicates ChIP-Seq signal in normalized reads per million.

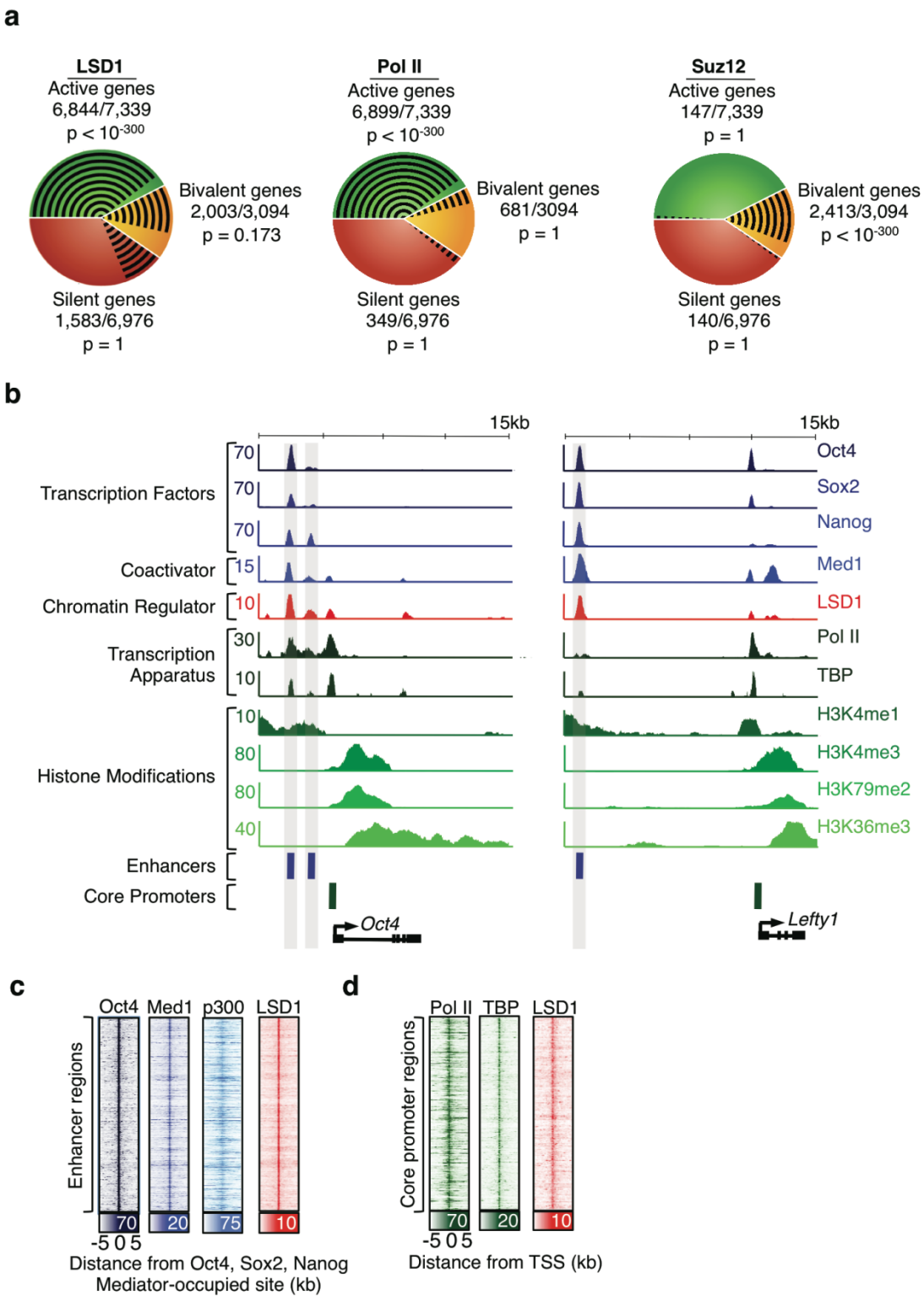

a
LSD1
Active genes
6,844/7,339
p < 10-300
Bivalent genes
2,003/3,094
p = 0.173
Silent genes
1,583/6,976
p = 1
Pol II
Active genes
6,899/7,339
p < 10-300
Bivalent genes
681/3094
p = 1
Silent genes
349/6,976
p = 1
Suz12
Active genes
147/7,339
p = 1
Bivalent genes
2,413/3,094
p < 10-300
Silent genes
140/6,976
p = 1
b
15kb
15kb
Transcription Factors
Coactivator
Chromatin Regulator
Transcription Apparatus
Histone Modifications
Enhancers
Core Promoters
Oct4
Sox2
Nanog
Med1
LSD1
Pol II
TBP
H3K4me1
H3K4me3
H3K79me2
H3K36me3
Oct4
Lefty1
c
Oct4 Med1 p300 LSD1
Enhancer regions
Distance from Oct4, Sox2, Nanog
Mediator-occupied site (kb)
d
Pol II TBP LSD1
Core promoter regions
Distance from TSS (kb)

Figure 1

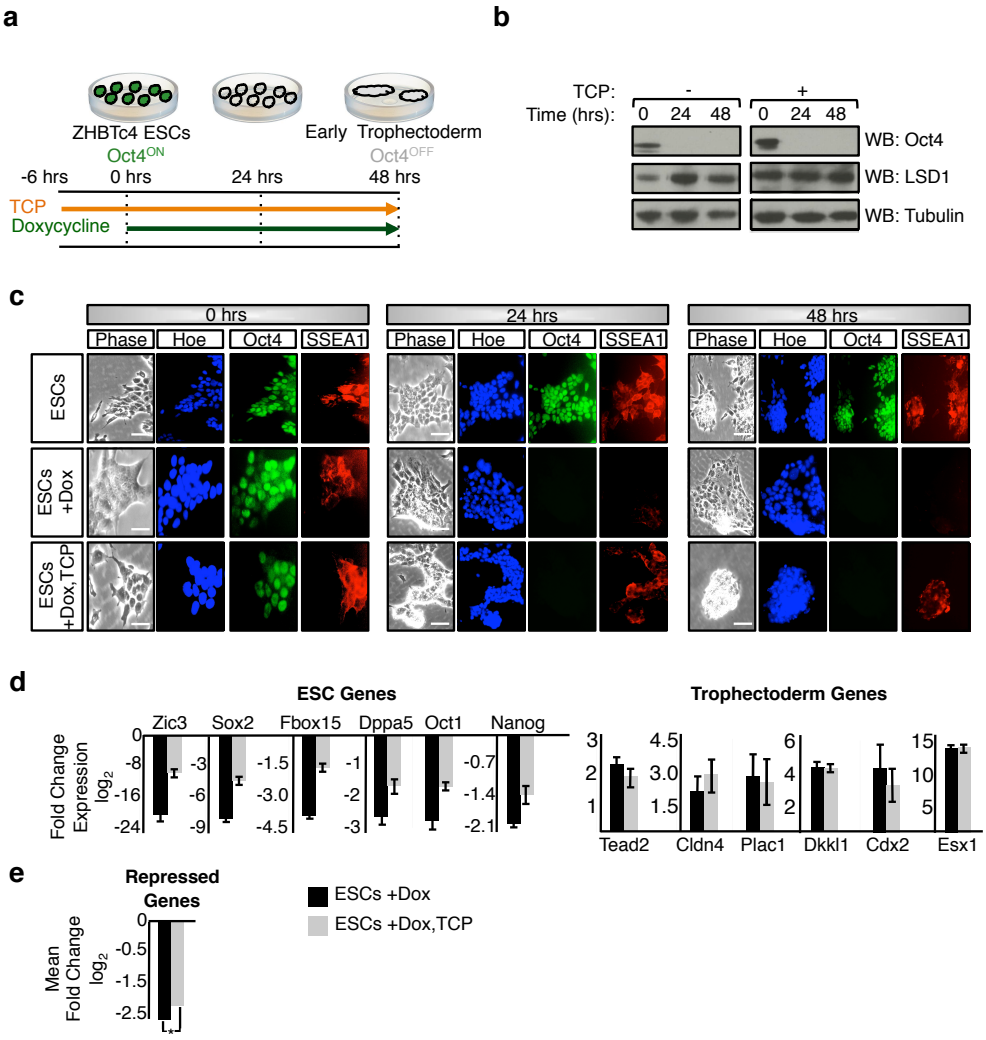

a
ZHBTc4 ESCs
Oct4ON
Early Trophectoderm
Oct4OFF
-6 hrs
0 hrs
24 hrs
48 hrs
TCP
Doxycycline
b
TCP:
Time (hrs):
0 24 48
0 24 48
WB: Oct4
WB: LSD1
WB: Tubulin
c
0 hrs
24 hrs
48 hrs
Phase
Hoe
Oct4
SSEA1
ESCs
ESCs +Dox
ESCs +Dox,TCP
d
ESC Genes
Trophectoderm Genes
Fold Change Expression log2
Zic3
Sox2
Fbox15
Dppa5
Oct1
Nanog
Tead2
Cldn4
Plac1
Dkkl1
Cdx2
Esx1
e
Repressed Genes
Mean Fold Change log2
ESCs +Dox
ESCs +Dox,TCP

Figure 2

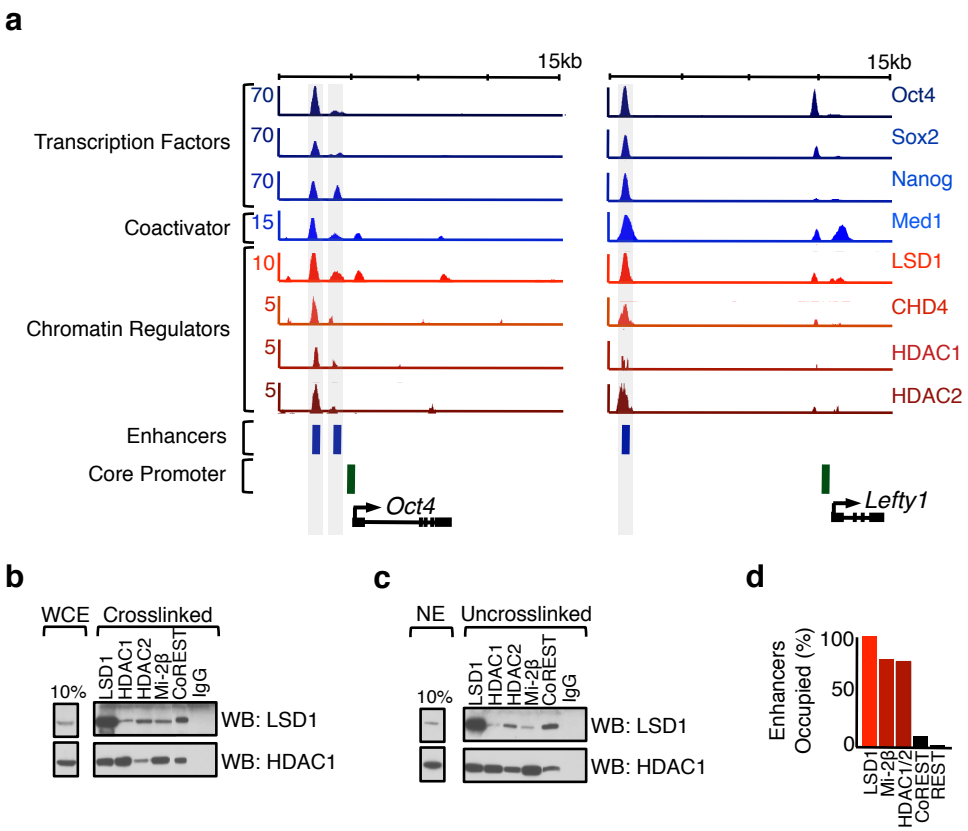

a
15kb
15kb
Transcription Factors
Coactivator
Chromatin Regulators
Enhancers
Core Promoter
70
70
70
15
10
5
5
5
Oct4
Sox2
Nanog
Med1
LSD1
CHD4
HDAC1
HDAC2
Oct4
Lefty1
b
WCE
Crosslinked
LSD1
HDAC1
HDAC2
Mi-2β
CoREST
IgG
10%
WB: LSD1
WB: HDAC1
c
NE
Uncrosslinked
LSD1
HDAC1
HDAC2
Mi-2β
CoREST
IgG
10%
WB: LSD1
WB: HDAC1
d
Enhancers
Occupied (%)
100
50
0
LSD1
Mi-2β
HDAC1/2
CoREST
REST

Figure 3

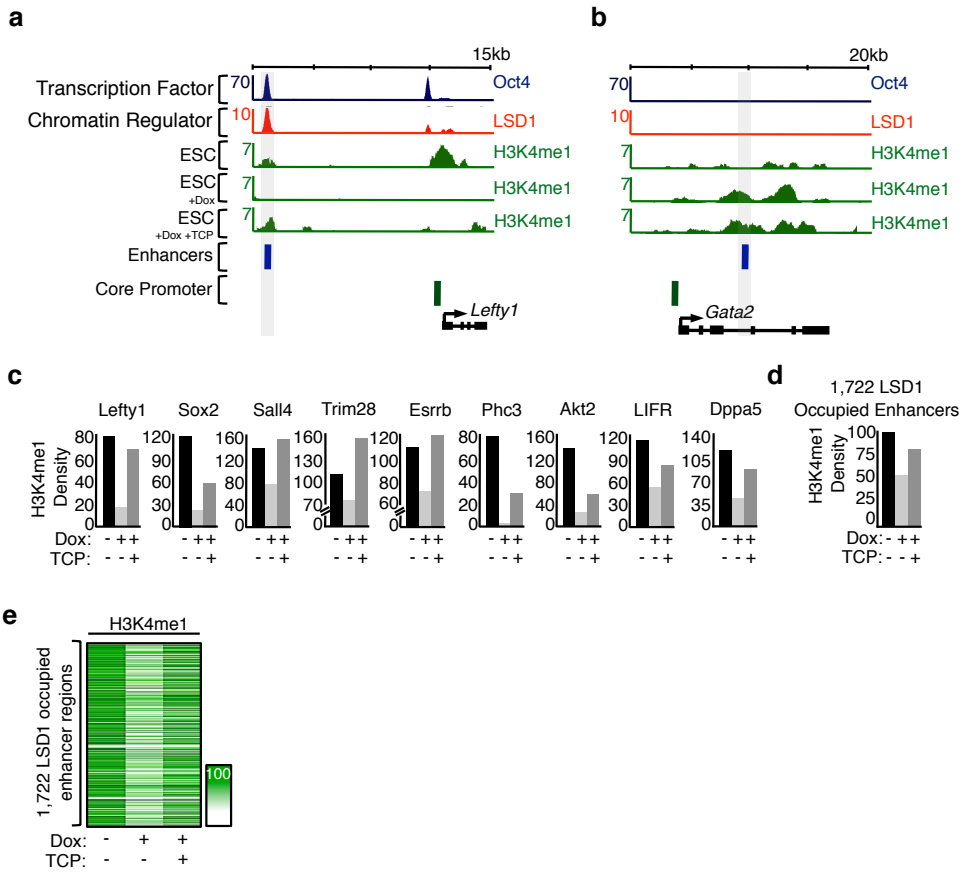

a
15kb
Transcription Factor
Chromatin Regulator
ESC
ESC +Dox
ESC +Dox +TCP
Enhancers
Core Promoter
70
Oct4
10
LSD1
7
H3K4me1
Lefty1
b
20kb
Gata2
c
Lefty1
Sox2
Sall4
Trim28
Esrrb
Phc3
Akt2
LIFR
Dppa5
H3K4me1 Density
Dox:
TCP:
d
1,722 LSD1 Occupied Enhancers
e
H3K4me1
1,722 LSD1 occupied enhancer regions
100

Figure 4

Chapter 3

Summary And Future Perspectives

Cellular differentiation is the developmental process by which cells become more specialized. Differentiation depends on changes in gene expression programs that are associated with changes in chromatin. In previous chapters, I described how deciphering the molecular mechanisms controlling differentiation is paramount to understanding development. Specifically, I discussed new insights revealing that chromatin regulators of opposing functions share a common set of active genes in embryonic stem (ES) cells, suggesting a model of a dynamic balance in the transcriptional control of cell state and differentiation. I described the mechanisms by which chromatin regulators contribute to the control of the ES cell state and differentiation, where these regulators play critical roles both in activating new gene expression programs and in silencing old programs.

In this chapter, I discuss how reprogramming is another valuable method for studying the molecular mechanisms involved in cell transitioning. I then discuss how these mechanisms can go awry and lead to diseased states. I then conclude by offering future challenges and possible approaches to test this dynamic model in other cell types.

Insights into Reprogramming to a Pluripotent State

The study of chromatin regulators in ES cell control has provided new insights into mechanisms that are involved in generation of induced pluripotent stem (iPS) cells (Box 2). These cells can been generated from a broad range of somatic cell types by using forced expression of Oct4, Sox2 and other transcription factors[1-4], and are apparently equivalent to ES cells in their ability to generate all cell types of the developing organism[5-8], and in gene expression[7,9].

Chromatin regulators play critical roles in either the reactivation of the pluripotent gene expression program or the silencing of the somatic gene expression program during iPS cell generation (reviewed in ref. 4)(Fig 1). Reactivation of the pluripotent gene expression program requires the removal of the DNA methylation and H3K9 methylation that normally occurs during differentiation when active ES cell genes are silenced. The mechanisms by which DNA methylation is lost at the promoters of ES cell transcription factors in iPS cells have not been elucidated, but the removal of H3K9 methylation likely involves H3K9 histone-demethylating proteins such as Jmjd1a and Jmjd2c[10]. Reprogramming of somatic cells to iPS cells is accompanied by increased expression of Jmjd1a and Jmjd2c. These two proteins occupy active genes in ES cells, where they remove H3K9 methylation and positively regulate gene expression[10].

Silencing of the somatic cell gene expression program during the generation of iPS cells requires a different set of regulators (Fig 1). Repressive chromatin regulators, which include histone-modifying proteins such as SetDB1 and PcG, are important for heterochromatin formation and transcriptional silencing of somatic genes. Interestingly, the H3K79 methyltransferase Dot1L inhibits this process, as prolonged H3K79 methylation at

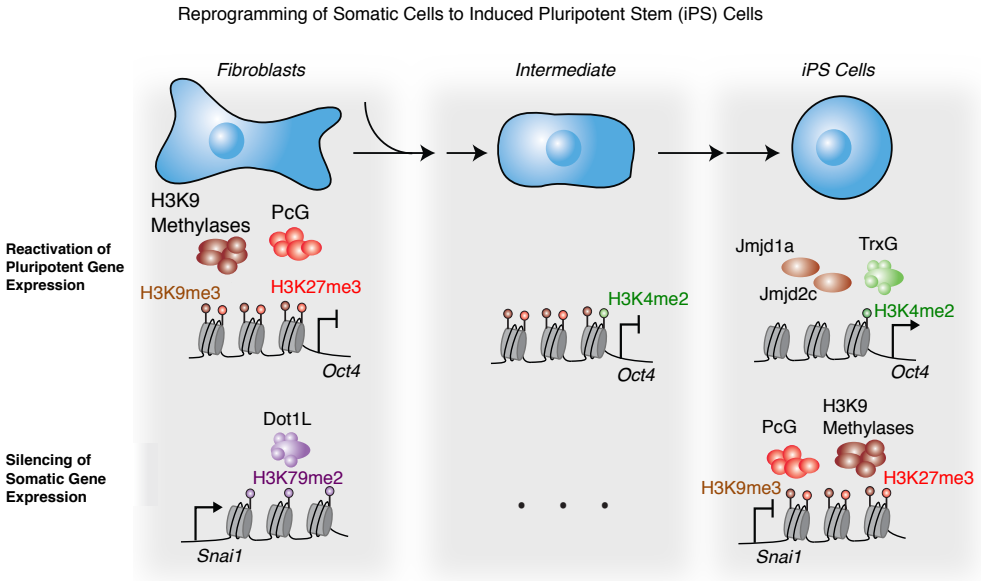

Reprogramming of Somatic Cells to Induced Pluripotent Stem (iPS) Cells
Fibroblasts
Intermediate
iPS Cells
Reactivation of Pluripotent Gene Expression
Silencing of Somatic Gene Expression
H3K9 Methylases
PcG
H3K9me3
H3K27me3
Oct4
H3K4me2
Jmjd1a
TrxG
Jmjd2c
Dot1L
H3K79me2
Snai1

Figure 1

Figure 1: Chromatin regulators participate in reprogramming of fibroblasts to induced pluripotent stem (iPS) cells

Reprogramming of murine somatic cells is a multistep process by which transcription factors or small molecules induce a pluripotent stem cell state. During generation of iPS cells, chromatin regulators participate in the reactivation of certain genes of the pluripotent gene expression program. In somatic fibroblast cells, many of the enhancers and promoters of genes associated with pluripotency such as *Oct4* and *Nanog* contain nucleosomes with H3K9 and H3K27 trimethylation. These genes are DNA methylated and are silent. During reprogramming, enhancer and promoter regions lacking DNA methylation gain H3K4 dimethylation, but pluripotency-associated genes remain silent until later steps when the histone modifications associated with repression are lost. How DNA methylation is lost from these genes is not fully understood, but loss of H3K9 and H3K27 methylation likely involves the activities of histone-demethylating enzymes such as Jmjd1a and Jmjd2c. Chromatin regulators also participate in the silencing of the somatic gene expression program. Recently, Dot1L was shown to prevent silencing of fibroblast genes such as *Snai1*. In fibroblasts, a subset of somatic genes are active and contain nucleosomes with H3K79 dimethylation (H3K79me2). Removal of Dot1L leads to loss of H3K79me2, and facilitates the repression of these genes by PcG proteins and H3K9-methyltransferases.

somatic genes during the initial phases of reprogramming prevents silencing and delays activation of the pluripotent gene expression program. Reduction in the function or levels of Dot1L enhances reprogramming, at least in part, by facilitating the removal of H3K79 methylation at somatic genes that need to be silenced. Hence, Dot1L and H3K79 methylation antagonizes the efficient repression of the somatic gene expression program during reprogramming[11]. Given this antagonistic function is observed in other systems[12], Dot1L inhibition may enhance reprogramming in a broad range of somatic cell types by facilitating the silencing of lineage-specific gene expression programs. In summary, the levels of specific chromatin regulators can be modulated to increase the efficiency by which iPS cells are generated.

Insights into Disease States

The study of chromatin regulators in ES cell control has also provided new insights into mechanisms that are involved in several human diseases (Fig 2). For example, improved understanding of the functions of nucleosome remodeling complexes such as the NuRD complex in cell transitioning has provided new insights into the molecular pathways affected by deregulation of this complex.

Of the various NuRD subunits, the metastasis-associated (MTA) subunits have been extensively studied in the context of human disease. Deregulation of MTA-containing NuRD complexes have been implicated in a variety of cancers, including breast, colorectal, gastric, oesophageal, endometrial, pancreatic, ovarian, non-small-cell lung, prostate, hepatocellular carcinoma, and diffuse large

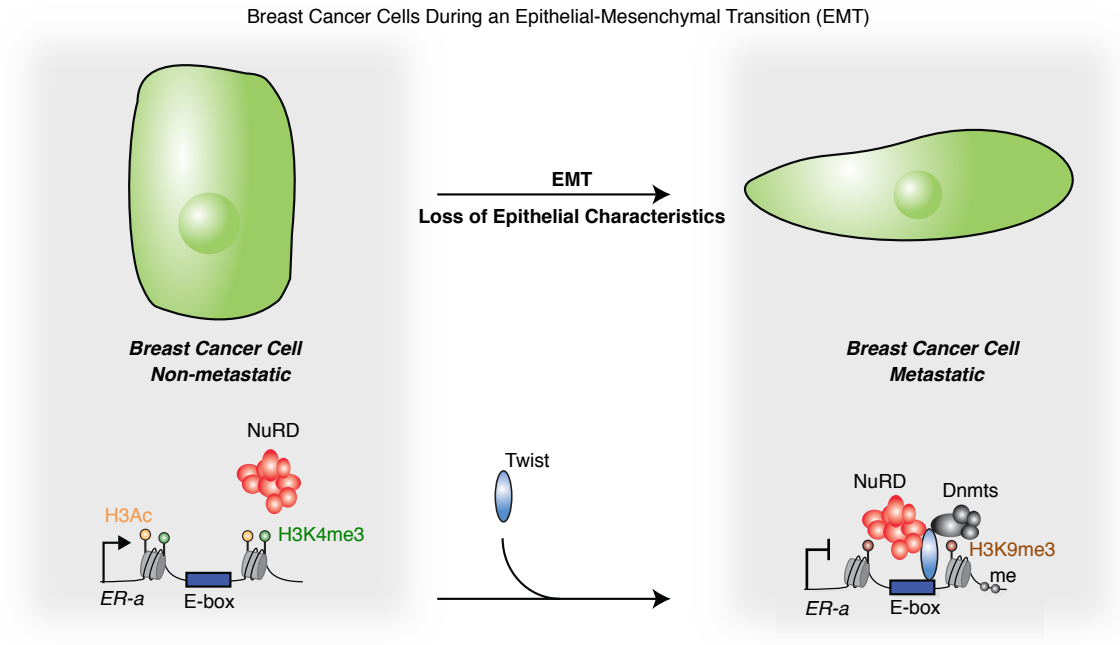

Breast Cancer Cells During an Epithelial-Mesenchymal Transition (EMT)
EMT
Loss of Epithelial Characteristics
Breast Cancer Cell
Non-metastatic
Breast Cancer Cell
Metastatic
NuRD
H3Ac
H3K4me3
ER-a
E-box
Twist
NuRD
Dnmts
H3K9me3
me
ER-a
E-box

Figure 2

Figure 2: NuRD participates in an epithelial-mesenchymal transition (EMT) in breast cancer

Non-metastatic breast cancer cells express epithelial markers such as E-cadherin and estrogen-receptor alpha (ER-a). During EMT, expression of certain epithelial genes are lost as cells acquire mesenchymal characteristics and metastatic ability. The oncogenic transcription factor Twist binds to the E-Box motif at the *ER-a* gene, where it binds to NuRD to facilitate gene silencing. Reduced levels of histone H3 acetylation by NuRD is accompanied by increased H3K9 methylation and chromatin compaction. DNA methyltransferases catalyze DNA methylation to enforce stable silencing.

B cell lymphoma (reviewed in ref. 13). Recently, NuRD was found to interact with the oncogene Twist[14-16], a master transcription factor that, when overexpressed, promotes an epithelial-mesenchymal transition (EMT) in various cancers[17]. During EMT, cells lose expression of certain epithelial genes and gain expression of certain mesenchymal genes, which is crucial for initiation of cancer metastasis[18]. In breast cancer cells expressing high levels of Twist, NuRD is associated with the oncogene in silencing of the epithelial markers E-cadherin and Estrogen Receptor-alpha (ER-a)[14-16]. During EMT, the NuRD complex deacetylates histones, which results in increased H3K9 methylation and chromatin condensation. Twist also interacts with Dnmt3b to methylate DNA and facilitate gene repression[16]. Reduced expression of these genes contributes to an EMT in these cells and cancer metastasis (Fig 2).

Knowledge that LSD1 and NuRD associate with master transcription factors in ES cells and promote the multi-step silencing of the ES cell gene expression program during differentiation suggests therapies that might compensate for the loss of the epithelial gene expression program during EMT and cancer metastasis. Histone deacetyltransferases reside within the NuRD complex[19-21], so it is conceivable that small molecule antagonists of HDACs would lead to an increase in transcriptional activity of NuRD target genes. Alternatively, as the NuRD complex frequently associates with oncogenic transcription factors to repress transcription (reviewed in ref. 13), drugs interfering with the interactions of these proteins may represent a more selective approach to inhibiting NuRD functions in cancer cells. Chromatin regulators with enzymatic activities are a new class of targets for small-molecule

drug discovery, and we can expect new developments in this field in the near future.

Future Perspective

Chromatin regulators play important roles in controlling the ES cell gene expression program to either maintain self-renewing cells in a pluripotent state, or to allow them to differentiate into all cell types of the developing and adult organism. Master transcription factors recruit chromatin regulators to contribute to the activation of a large fraction of active genes, and to the poised activation of lineage-specific genes. At actively transcribed genes, chromatin regulators implicated in maintenance of cell state occupy enhancer elements and core promoters and modulate RNA expression levels through mechanisms that include creating accessible binding sites, looping in of enhancer elements to core promoters, and releasing paused RNA polymerase. At lineage-specific genes, chromatin regulators occupy promoter regions and help create a poised state by pausing RNA polymerase and creating repressive chromatin architecture. During differentiation, chromatin regulators facilitate the silencing of genes encoding ES cell transcription factors and the selective activation of specific lineage-specific genes.

An interesting regulatory feature of ES cells is that chromatin regulators of opposing functions share common target genes in ES cells, suggesting a more dynamic view on the mechanisms governing ES cell state. An intriguing model is that in ES cells a dynamic balance exists at genes between regulators that promote

gene activity and those that promote gene silencing. At active genes, the activities of chromatin regulators responsible for maintained gene activation predominate, and the net effect of having regulators of opposing activities is to keep genes poised for repression. At lineage-specific genes, the chromatin regulators responsible for gene repression predominate, and the net effect of having regulators of opposing activities at these genes is to keep genes poised for activation. During ES cell differentiation, the levels or functions of chromatin regulators associated with gene activation are reduced at a set of active genes, allowing these genes to become silenced. At lineage-specific genes, the levels or functions of chromatin regulators associated with repression are reduced at selected genes, and lineage-specific transcription factors are selectively activated.

ES cells are a powerful model system for discovering regulatory features of the control of cell state, and it would be important to test if this model is true in other cell types, such as adult stem and progenitor cells that can also differentiate, to improve our understanding on the regulation of cell identity, the process of mammalian development, and how regulation goes awry in disease. Insights from ES cells, reprogramming and disease support the idea that chromatin regulators are important in the control of cell state in various cell types. Therefore, if this model holds true for most cell types, then identifying the functions and localization of all chromatin regulators in all cell types would significantly improve our understanding of the mechanisms involved in cell state control and differentiation. For example, the concept that some regulators activate transcription and others repress transcription is almost certainly operative in all cell types, suggesting that

improved understanding of global control on gene expression will depend on ascertaining which of these functions applies to each chromatin regulator in each cell type. With emerging evidence that signaling cues can transmit information about the surrounding environment to genes occupied by master transcription factors and associated regulators in a cell type-specific-manner[22], better perspective on the cell-type-specific recruitment of chromatin regulators to genes in each cell type will be crucial to elucidating the potential roles for these regulators in control of cell state and differentiation.

References

1 Hanna, J.H., Saha, K., & Jaenisch, R., Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* 143 (4), 508-525 (2010).

2 Stadtfeld, M. & Hochedlinger, K., Induced pluripotency: history, mechanisms, and applications. *Genes Dev* 24 (20), 2239-2263 (2010).

3 Yamanaka, S. & Blau, H.M., Nuclear reprogramming to a pluripotent state by three approaches. *Nature* 465 (7299), 704-712 (2010).

4 Plath, K. & Lowry, W.E., Progress in understanding reprogramming to the induced pluripotent state. *Nat Rev Genet* 12 (4), 253-265 (2011).

5 Boland, M.J. *et al*., Adult mice generated from induced pluripotent stem cells. *Nature* 461 (7260), 91-94 (2009).

6 Kang, L., Wang, J., Zhang, Y., Kou, Z., & Gao, S., iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. *Cell Stem Cell* 5 (2), 135-138 (2009).

7 Stadtfeld, M. & Hochedlinger, K., Induced pluripotency: history, mechanisms, and applications. *Genes Dev* 24 (20), 2239-2263 (2010).

8 Zhao, X.Y. *et al*., iPS cells produce viable mice through tetraploid complementation. *Nature* 461 (7260), 86-90 (2009).

9 Guenther, M.G. *et al*., Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* 7 (2), 249-257 (2010).

10 Loh, Y.H., Zhang, W., Chen, X., George, J., & Ng, H.H., Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev* 21 (20), 2545-2557 (2007).

11 Onder, T.T. *et al*., Chromatin-modifying enzymes as modulators of reprogramming. *Nature* 483 (7391), 598-602 (2012).

12 Stulemeijer, I.J. *et al*., Dot1 binding induces chromatin rearrangements by histone methylation-dependent and -independent mechanisms. *Epigenetics Chromatin* 4 (1), 2 (2011).

13 Lai, A.Y. & Wade, P.A., Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer* 11 (8), 588-596 (2011).

14 Fu, J. *et al*., The TWIST/Mi2/NuRD protein complex and its essential role in cancer metastasis. *Cell Res* 21 (2), 275-289 (2011).

15 Fu, J. *et al*., TWIST represses estrogen receptor-alpha expression by recruiting the NuRD protein complex in breast cancer cells. *Int J Biol Sci* 8 (4), 522-532 (2012).

16 Vesuna, F. *et al*., Twist contributes to hormone resistance in breast cancer by downregulating estrogen receptor-alpha. *Oncogene* (2011).

17 Peinado, H., Olmeda, D., & Cano, A., Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer* 7 (6), 415-428 (2007).

18 Kalluri, R. & Weinberg, R.A., The basics of epithelial-mesenchymal transition. *J Clin Invest* 119 (6), 1420-1428 (2009).

19 Tong, J.K., Hassig, C.A., Schnitzler, G.R., Kingston, R.E., & Schreiber, S.L., Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* 395 (6705), 917-921 (1998).

20 Wade, P.A., Jones, P.L., Vermaak, D., & Wolffe, A.P., A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. *Curr Biol* 8 (14), 843-846 (1998).

21 Xue, Y. *et al*., NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell* 2 (6), 851-861 (1998).

22 Mullen, A.C. *et al*., Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* 147 (3), 565-576 (2011).

Box 2: Changes in Chromatin Structure During Reprogramming to a Pluripotent State

Since 2006 (ref. 1), the stepwise mechanism by which transcription factors or small molecules induce a pluripotent state has been well studied (reviewed in ref. 2)(Fig 1). For example, reprogramming of mouse fibroblast cells requires the stepwise transition through key intermediate steps, with fewer and fewer cells advancing. Increase in proliferation rate and decrease in cell size are the first noticeable changes in the reprogramming of fibroblasts[3]. These changes are accompanied by the induction of proliferation genes and silencing of the somatic gene expression program[4-6]. Mesenchymal cells such as fibroblasts need to acquire epithelial characteristics similar to ES cells, and thus undergo a mesenchymal-epithelial transition (MET) by upregulating epithelial genes such as *CDH1* (encodes E-cadherin), and downregulating mesenchymal genes such as *Snai1* (refs. 5,7). After these characteristics have been established, other ES cell markers such as SSEA1 are induced[7-9], as well as many embryonic genes involved in housekeeping functions[5,6,10]. Expression of the pluripotency-related genes *Oct4*, *Sox2* and Nanog, and the jump-start of the pluripotency gene expression program, are considered to be the final step of reprogramming[8].

Alterations in the levels or functions of chromatin regulators affect reprogramming efficiency, suggesting that changes in chromatin are crucial for inducing a pluripotent state[11]. In fibroblasts and other somatic cells, promoters of key pluripotency-related genes such as *Oct4* and *Nanog* lack histone H3K4me3, the surrounding DNA is

methylated, and these genes are silenced[1,6]. Histones H3K27 and H3K9 methylation can also often be found at many of these promoters[6,10,12,13]. It is unclear when DNA demethylation and the loss of these histone methylation marks occur during reprogramming, however recent evidence suggests that specific chromatin changes precede the activation of pluripotency-related genes[4,10]. During the initial steps of reprogramming, enhancer elements and promoter regions of certain pluripotency-related genes that lack DNA methylation in fibroblasts gain histone H3K4 dimethylation (H3K4me2). Increased H3K4 methylation does not affect the levels of H3K27 methylation at these genes, which remain silent until later steps of reprogramming when H3K27 methylation levels are reduced[4]. Changes in chromatin structure also occur at enhancer elements and core promoters of fibroblast genes that must be silenced. H3K4 methylation is lost early from core promoters of these genes, while DNA methylation of enhancer elements occurs much later in reprogramming[4]. Given DNA methylation is required to stably silence genes during ES cell differentiation[13], this observation would explain why some cells in the intermediate steps of reprogramming can return to a fibroblast-like cell state upon removal of reprogramming factors[8,9].

In summary, knowledge of the mechanisms underlying reprogramming, including the changes in chromatin structure that occur during this process, is critical for advancing therapeutic application of induced pluripotent stem (iPS) cells.

References

1 Takahashi, K. & Yamanaka, S., Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126 (4), 663-676 (2006).

2 Feng, B., Ng, J.H., Heng, J.C., & Ng, H.H., Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. *Cell Stem Cell* 4 (4), 301-312 (2009).

3 Smith, Z.D., Nachman, I., Regev, A., & Meissner, A., Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nat Biotechnol* 28 (5), 521-526 (2010).

4 Koche, R.P. *et al*., Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* 8 (1), 96-105 (2011).

5 Samavarchi-Tehrani, P. *et al*., Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* 7 (1), 64-77 (2010).

6 Mikkelsen, T.S. *et al*., Dissecting direct reprogramming through integrative genomic analysis. *Nature* 454 (7200), 49-55 (2008).

7 Li, R. *et al*., A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* 7 (1), 51-63 (2010).

8 Stadtfeld, M. *et al*., Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* 465 (7295), 175-181 (2008).

9 Brambrink, T. *et al*., Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* 2 (2), 151-159 (2008).

10 Sridharan, R. *et al*., Role of the murine reprogramming factors in the induction of pluripotency. *Cell* 136 (2), 364-377 (2009).

11 Plath, K. & Lowry, W.E., Progress in understanding reprogramming to the induced pluripotent state. *Nat Rev Genet* 12 (4), 253-265 (2011).

12 Hawkins, R.D. *et al*., Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 6 (5), 479-491 (2010).

13 Feldman, N. *et al*., G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* 8 (2), 188-194 (2006).

Appendix A:

Supplemental Information to Chapter 2: Enhancer Decommissioning by LSD1 During Embryonic Stem Cell Differentiation

Warren A. Whyte[1,2*], Steve Bilodeau[1*], David A. Orlando[1], Heather A. Hoke[1,2], Garrett M. Frampton[1,2], Charles T. Foster[3,4], Shaun M. Cowley[4], and Richard A. Young[1,2]

[1] Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2] Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3] Department of Molecular Biology, Adolf-Butenandt Institut, Ludwig-Maximilians-Universität München, Munich, Germany 80336. [4] Department of Biochemistry, University of Leicester, Leicester LE1 9HN, United Kingdom.

SUPPLEMENTARY INFORMATION

Enhancer Decommissioning by LSD1 During Embryonic Stem Cell Differentiation

Warren A. Whyte, Steve Bilodeau, David A. Orlando, Heather A. Hoke, Garrett M. Frampton, Charles T. Foster, Shaun M. Cowley, and Richard A. Young

CONTENTS

Generation and Expression of LSD1 Expression Constructs in ESCs (Supplementary Fig. 5)

Differentiation of ESCs (Fig. 2, 4, and Supplementary Fig. 2, 3, 9, and 10)

Inhibition of p300 using HAT inhibitor C-646 (Supplementary Fig. 12)

Lentiviral Production and Infection (Supplementary Fig. 2, 3, 4, 5, and 9)

Immunofluorescence (Fig. 2b and Supplementary Fig. 3, 4, 5, and 9)

Alkaline Phosphatase Staining (Supplementary Fig. 2, 4, 5, and 9)

RNA Extraction, cDNA, and TaqMan Expression Analysis (Supplementary Fig. 1, 3, 5, and 9)

Chromatin Immunoprecipitation (ChIP)

Antibody Specificity (Supplementary Fig. 1, 6, and 9)

ChIP Protocol

Gene Specific ChIPs (Supplementary Fig. 9, 10, 11, 12)

ChIP-Seq Sample Preparation and Analysis

Sample Preparation

Polony Generation and Sequencing

ChIP-Seq Data Analysis

Assigning ChIP-Seq Enriched Regions to Genes (Supplementary Table 1)

Enrichment of LSD1, Pol II, and Suz12 at Genes (Fig. 1)

Definition of Enhancer and Core Promoter (Fig. 1, 4, Supplementary Fig. 8, 13, and Supplementary Table 1)

Determining LSD1 density at Enhancer and Core Promoter (Supplementary Fig. 1)

ChIP-Seq Density Heatmaps (Fig. 1, 4, Supplementary Fig. 8, and 13)

Calculation of the Statistical Significance of the Overlap Between Sets of Genomic Regions (Fig. 1, 3, 4, Supplementary Fig. 8, and 12)

Rank Normalization of H3K4me1 ChIP-Seq Data (Fig. 4, Supplementary Fig. 13, and Supplementary Table 5)

Gene Ontology (GO) Analysis (Supplementary Fig. 7)

Public availability of ChIP-Seq Datasets

Microarray Analysis (Fig. 2)

Cell Culture and RNA isolation

Microarray hybridization and Analysis

Public availability of Microarray Gene Expression Datasets

Protein Extraction and Western Blot Analysis (Fig. 2, Supplementary Fig. 1, 4, 5, 6, and 9)

ChIP-Western and Co-Immunoprecipitation (Fig. 3)

Supplementary References

Supplementary Tables

Supplementary Table 1 – Summary of occupied genes and regions

Supplementary Table 2 – Summary of LSD1 occupancy at active and poised enhancers

Supplementary Table 3 – Gene expression changes 48 hours after Oct4 repression in ZHBTc4 DMSO (control)- and TCP-treated cells

Supplementary Table 4 – Genes co-occupied by LSD1 and REST

Supplementary Table 5 – Enhancer decommissioning 48 hours after Oct4 depletion in ZHBTc4 DMSO (control)- and TCP-treated cells

Supplementary Data File 1

Supplementary Data File 1 contains ChIP-Seq data in compressed WIG format (WIG.GZ) for upload into the UCSC genome browser[1]. This file contains data for mES_H3K4me1, DMSO_H3K4me1, DMSO_48HR_H3K4me1, TCP_48HR_H3K4me1, mES_H3K4me3, mES_H3K79me2, mES_H3K36me3, mES_Oct4, mES_Sox2, mES_Nanog, mES_Pol II, mES_TBP, mES_LSD1, mES_p300, mES_REST, mES_CoREST, mES_Suz12, mES_H3K27ac, mES_HDAC1, mES_HDAC2, mES_Mi-2b, mES_DMSO_H3K27ac, mES_DMSO_H3K4me1, mES_C646_H3K27ac, mES_C646_H3K4me1, WCE_DMSO, WCE_48HR_DMSO, WCE_48HR_TCP, and WCE_mES.

The first track for each data set contains the ChIP-Seq density across the genome in 25bp bins. The minimum ChIP-Seq density shown in these files is 1.0 reads per million total reads. For DMSO_H3K4me1, DMSO_48HR_H3K4me1, and TCP_48HR_H3K4me1 datasets, the minimum ChIP-Seq density is 1 normalized read per million total reads. Subsequent tracks identify genomic regions identified as enriched at a p-value threshold of 10^{-9}.

Supplementary Discussion

To address LSD1 function in ESCs and during ESC differentiation, cells were treated with the monoamine oxidase inhibitors (MAOIs) tranylcypromine (TCP) or pargyline (Prg). Mechanistically, LSD1 is unique relative to other demethylases, as it demethylates lysine residues via a flavin–adenine dinucleotide–dependent reaction[2,3]. This reaction is inhibited by MAOIs, which are used in the treatment of certain psychiatric and neurological disorders[4-8]. Furthermore, TCP inhibits LSD1 at levels comparable to MAO inhibition of clinical mitochondrial MAOI targets[6,7]. Finally, the effects of inhibition by TCP and Prg were highly similar in our assays. Therefore, we found TCP and Prg suitable to study LSD1 activity during differentiation of ES cells.

Previous studies demonstrated that LSD1 is required for mouse development beyond e6.5 (ref. 9). The differentiation defect is recapitulated *in vitro* in embryoid body (EB) formation assays, where ESCs are plated in suspension in media lacking leukemia

inhibitory factor (LIF), causing ESC differentiation. In these assays, deletion of LSD1 leads to considerable cell death and formation of small EBs, with the remaining cells expressing markers of all three germ layers[9,10]. The differentiation defect is also recapitulated *in vitro* in the Oct4 depletion assay used in the present study, where ESCs preferentially differentiate into the trophectodermal lineage. In this assay, inhibition of LSD1 with either TCP or Prg leads to considerable cell death, and surviving cells express trophectodermal differentiation markers. Thus, very similar effects are observed in these two assays.

In the Oct4 depletion assay used in the present study, full repression of key ESC regulators such as Nanog is not achieved in the presence of LSD1 inhibitors 48 hours after Oct4 depletion is initiated. Similar results were obtained for Nanog in the assay used by Foster et al. (2010) with LSD1 mutant ESCs 48 hours after LIF removal[9]. In this latter assay, Nanog levels were ultimately fully reduced in ESCs lacking LSD1. Our interpretation of all these data is that inhibition of LSD1 delays differentiation because there is a delay in reducing levels of key ESC regulators whose enhancers are occupied by LSD1.

Previous studies also report that deletion of LSD1 protein destabilizes Co-REST, which leads to a less active LSD1-CoREST-HDAC complex[9]. Accordingly, developmental regulators targeted by the LSD1-CoREST-HDAC complex are de-repressed, and this may explain the defect observed in ESC differentiation. In experiments using small molecule inhibitors of LSD1 activity, we did not observe

significant changes in genes associated with LSD1-REST co-occupied sites. One explanation for this discrepancy may be that loss of LSD1 proteins destabilizes the CoREST-HDAC complex, while small molecules inhibiting LSD1 enzymatic activity may not alter the stability of the complex, thereby giving a different transcriptional phenotype. Thus, the different methods (gene deletion versus enzymatic inhibition) of LSD1 inhibition may explain the discrepancy in gene expression.

Differences were observed for LSD1 function in mouse and human ESCs. Although mouse ESCs fail to differentiate into embryoid bodies in absence of LSD1 (refs. 9-11), human ESCs differentiate when LSD1 levels are reduced[12]. In both mouse and human ESCs, LSD1 is found at Oct4 and Nanog -occupied regions. Similar to our observations in the Oct4 depletion assay, a subpopulation of differentiating human ESCs with reduced LSD1 levels retained the expression of pluripotency markers[12].

Supplementary Experimental Procedures

Cell Culture Conditions and Differentiation Assays

Embryonic Stem Cells

V6.5, ZHBTc4, and E14 murine ESCs were grown on irradiated murine embryonic fibroblasts (MEFs), unless otherwise stated. Cells were grown under standard ESC conditions as described previously[13]. Briefly, cells were grown on 0.2% gelatinized (Sigma, G1890) tissue culture plates in ESC media; DMEM-KO (Invitrogen, 10829-018) supplemented with 15% fetal bovine serum (Hyclone, characterized SH3007103), 1000 U/mL LIF (ESGRO, ESG1106), 100 μM nonessential amino acids (Invitrogen, 11140-050), 2 mM L-glutamine (Invitrogen, 25030-081), 100 U/mL penicillin, 100 μg/mL streptomycin (Invitrogen, 15140-122), and 8 nL/mL of 2-mercaptoethanol (Sigma, M7522).

Generation of LSD1 knockout ESC lines (Supplementary Fig. 4, 5)

An E14 ESC line expressing a Cre-estrogen receptor fusion protein from the *ROSA26* locus was used to generate cells in which LSD1 can be inactivated conditionally. Briefly, LSD1 Lox/Δ3 ESCs were generated as previously described[12], in which one LSD1 allele has exon 3 flanked by LoxP sites (floxed) and the second has exon 3 deleted. Induction of Cre activity by addition of 4-hydroxytamoxifen (4-OHT, Sigma, H7904) to the growth medium resulted in complete recombination of the remaining floxed allele (to generate LSD1 Δ3/Δ3) within 6 hours. Cells were grown in the absence of irradiated MEFs under standard ESC conditions as described previously[13].

Generation and Expression of LSD1 Expression Constructs in ESCs (Supplementary Fig. 5)

Two forms of human LSD1 were generated by PCR with primers containing tails of family D vector homology for cloning into pCAGGs expression vectors. An enhanced green fluorescent protein (EGFP) tag was added to the N-terminus of either the full-length LSD1 protein, or the catalytically inactive form of LSD1 with a lysine to alanine mutation introduced at residue 661 by site-directed mutagenesis[14].

For expression of LSD1 constructs in ESCs, LSD1 Lox/Δ3 and Δ3/Δ3 E14 ESCs were plated in 6-well culture plates. The following day, cells were transfected with 2μg pCAGGs construct and 2μg of the pMONO-hygro plasmid containing the hygromycin resistance gene (Invivogen, pmonoh-mcs) using 10μl Lipofectamine 2000 (Invitrogen, 11668-019). After 24 hours, the media was removed and replaced with ESC media containing 400μg/ml hygromycin. 24 hours after hygromycin addition, hygromycin concentration was reduced to 200μg/ml for the duration of the experiment.

Differentiation of ESCs (Fig. 2, 4, and Supplementary Fig. 2, 3, 9, and 10)

For immunofluorescence, monitoring H3K4me1 levels and ESC expression analysis during differentiation following either DMSO, TCP or Prg treatment, ZHBTc4 ESCs were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs, and plated on either 6-well or 15cm cell culture plates. The following day, cells were treated with either DMSO, 1mM TCP or 3mM Prg for 6 hours in ESC media. After 6 hours, the media was removed and replaced with ESC media containing 2μg/ml doxycyline and either DMSO, TCP or Prg. 24 hours after doxycycline addition, the media was replaced with ESC media containing doxycycline and either DMSO, TCP or Prg for another 24 hours (48 hours total).

Inhibition of p300 using HAT inhibitor C-646 (Supplementary Fig. 12)

During differentiation, it is presently unclear which acetylated histone residue(s) need to be deacetylated to trigger activation of LSD1 demethylation activity. While p300 and H3K27Ac are very good candidates at enhancers, many other HATs (CBP, GCN5, Tip60, Elp3, Myst3, Myst4) are active in ESCs [15-17]. Therefore, regulation of acetylation levels at enhancers is most likely to be the result of a combination of HATs and HDACs complexes.

To test if acetylation of H3K27 prevents demethylation of H3K4 by LSD1, we treated ESCs with the p300-specific acteyltransferase inhibitor C-646 (ref. 18). V6.5 ESCs were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs, and plated in 15cm cell culture plates. The following days, cells were treated with either DMSO (control) or C-646 (Tocris, 4200) for 24 hours. Cells

were then crosslinked and collected to generate ChIP-seq datasets for H3K27ac and H3K4me1.

Lentiviral Production and Infection (Supplementary Fig. 2, 3, 4, 5, and 9)

Lentivirus was produced according to Open Biosystems *Trans*-lentiviral shRNA Packaging System (TLP4614). The shRNA constructs targeting LSD1, Oct4, CoREST and Mi-2b are listed below. The shRNAs targeting either GFP (RHS4459) or Luciferase (SHC007) were used as controls.

GFP	RHS4459
Luciferase	SHC007
LSD1	TRCN0000071373
Oct4	TRCN0000009611
Mi-2b #1	TRCN0000086143
Mi-2b #2	TRCN0000086145
CoREST	TRCN0000071371

For GFP, LSD1 and Mi-2b, ZHBTc4 ESCs were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs, and plated on either 6-well or 12-well cell culture plates. The following day, cells were infected in ESC media containing 8 μg/ml polybrene (Sigma, H9268-10G). After 24 hours the media was removed and replaced with ESC media containing 3.5 μg/mL puromycin (Sigma, P8833). ESC media with puromycin was changed daily. Three days post infection,

the media was removed and replaced with ESC media containing 2μg/ml doxycycline to shutdown Oct4 and induce differentiation. Cells were harvested 48 hours later.

For Luciferase and Oct4, LSD1 Lox/Δ3 or Δ3/Δ3 E14 ESCs were plated on 6-well culture plates. The following day, cells were infected in ESC media containing 8 μg/ml polybrene. Cells transfected with LSD1 expression constructs were infected in ESC media lacking antibiotics and polybrene. After 16 hours the media was removed and replaced with ESC media containing 3.5 μg/mL puromycin (Sigma, P8833). The media of ESCs transfected with the LSD1 expression constructs also contained hygromycin, as described earlier. ESC media with either puromycin, or puromycin and hygromycin, was changed daily. Cells were harvested 72 hours later.

Immunofluorescence (Fig. 2b and Supplementary Fig. 3, 4, 5, and 9)

Following crosslinking, the cells were washed once with PBS, twice with blocking buffer (PBS with 0.25% BSA, Sigma, A3059-10G) and then permeabilized for 15 minutes with 0.2% Triton X-100 (Sigma, T8797-100ml). After two washes with blocking buffer cells were stained overnight at 4 degrees C for either Oct4 (Santa Cruz Biotechnology, sc-9081x; 1:200 dilution) or SSEA1 (mc-480, DHSB, 1:20 dilution) and washed twice with blocking buffer. Cells were incubated for 1 hour at room temperature with either goat anti-rabbit-conjugated Alexa Fluor 488, goat anti-rabbit-conjugated Alexa Fluor 647 or goat anti-mouse conjugated 568 (Invitrogen; 1:1000 dilution) and Hoechst 33342 (Invitrogen; 1:2000 dilution). Finally, cells were washed twice with blocking buffer and twice with PBS before imaging. Images were acquired

on either Nikon Inverted TE300 or Zeiss Axiovert 200m Inverted microscopes with a Hamamatsu Orca camera. Openlab (http://www.improvision.com/products/openlab/) was used for image acquisition.

Alkaline Phosphatase Staining (Supplementary Fig. 2, 4, 5, and 9)

Staining of ZHBTc4 cells for alkaline phosphatase was achieved using the Alkaline Phosphatase Detection Kit (Millipore, SCR004). Briefly, cells were crosslinked at various timepoints before addition of Fast Red Violet solution and Napthol AS-BI phosphate solution. Cells were visualized on a Nikon Inverted TE300 with a Hamamatsu Orca camera. Openlab (http://www.improvision.com/products/openlab/) was used for image acquisition.

RNA Extraction, cDNA, and TaqMan Expression Analysis (Supplementary Fig. 1, 3, 5, and 9)

RNA utilized for real-time qPCR was extracted with TRIzol according to the manufacturer protocol (Invitrogen, 15596-026). Purified RNA was reverse transcribed using Superscript III (Invitrogen) with oligo dT primed first-strand synthesis following the manufacturer protocol.

Real-time qPCR were carried out on the 7000 ABI Detection System using the following Taqman probes according to the manufacturer protocol (Applied Biosystems).

Gapdh	Mm99999915_g1
LSD1	Mm01181030_g1
Mi-2b	Mm01190896_m1
CoREST	Mm03053471_s1
Sox2	Mm00488369_s1
Nanog	Mm02019550_s1
Lefty1	Mm00438615_m1
Lefty2	Mm00774547_m1
Esrrb	Mm00442411_m1
Trim28	Mm00495594_m1
Klf4	Mm00516104_g1
Oct4	Mm03053917_g1
Apobec1	Mm00482895_m1
Dppa3	Mm01184198_g1
Tcl1	Mm00493477_m1

Expression levels were normalized to Gapdh levels. All comparisons were made relative to either DMSO (control)-treated, Luciferase, or GFP-infected cells.

Chromatin Immunoprecipitation (ChIP)

Antibody Specificity (Supplementary Fig. 1, 6, and 9)

For LSD1 (AOF2/KDM1A)-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab17721 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide conjugated to KLH derived from within residues 800 to the C-terminus of human LSD1. Antibody specificity was determined by shRNA-mediated knockdown (Open Biosystems) of LSD1 proteins, followed by Western blot analysis. Knockdowns were carried out using the following shRNAs according to the manufacturer protocol (Open Biosystems). Cells were infected with indicated short hairpins for 5 days, followed by protein extraction and Western blotting.

GFP	RHS4459
LSD1	TRCN0000071374

For Mi-2b (CHD4)-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab72418 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide conjugated to KLH derived from within residues 25 to 75 of human CHD4. Antibody specificity was determined by shRNA-mediated knockdown (Open Biosystems) of Mi-2b proteins, followed by Western blot analysis. Knockdowns were carried out using the following shRNAs according to the manufacturer protocol (Open Biosystems). Cells were infected with indicated short hairpins for 5 days, followed by protein extraction and Western blotting.

GFP	RHS4459
shMi-2b #1	TRCN0000086143

shMi-2b #2 TRCN0000086147

For HDAC1-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab7028 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide conjugated to KLH derived from within residues 466 to 482 of human HDAC1. Antibody specificity was previously determined[19].

For HDAC2-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab7029 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide conjugated to KLH derived from within residues 471 to 488 of human HDAC2. Antibody specificity was previously determined[19].

For REST (NRSF)-occupied genomic regions, we performed ChIP-Seq experiments using Millipore 07-579 rabbit polyclonal antibody. The antibody was raised with a GST fusion protein corresponding to residues 801-1097 of human REST. Antibody specificity was previously determined[20].

For CoREST-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab32631 rabbit polyclonal **antibody. The antibody was raised** with a synthetic peptide conjugated to KLH derived from within residues 450 to the C-terminus of human CoREST. Antibody specificity was determined by shRNA-mediated knockdown (Open Biosystems) of CoREST proteins, followed by Western blot analysis. Knockdowns were carried out using the following shRNAs according to

the manufacturer protocol (Open Biosystems). Cells were infected with indicated short hairpins for 5 days, followed by protein extraction and Western blotting.

GFP	RHS4459
CoREST	TRCN0000071371

For H3K4me1-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab8895 rabbit polyclonal **antibody. The antibody was raised with a** synthetic peptide conjugated to KLH derived from within residues 1 to 100 of human H3K4me1. Antibody specificity was previously determined[21].

For p300-occupied genomic regions, we performed ChIP-PCR experiments using Santa Cruz sc-584 rabbit polyclonal **antibody. The antibody was raised with a synthetic** peptide derived from the N-terminus of human p300. Antibody specificity was previously determined[22].

For H3K27Ac-occupied genomic regions, we performed ChIP-PCR experiments using Abcam ab4729 rabbit polyclonal **antibody. The antibody was raised with a s**ynthetic peptide conjugated to KLH derived from within residues 1 - 100 of human histone H3, acetylated at K27. Antibody specificity was previously determined[23].

ChIP Protocol

Protocols describing chromatin immunoprecipitation materials and methods have been previously described[24]. v6.5 or ZHBTc4 ESCs were grown to a final count of 5-10 x 10^7 cells for each ChIP experiment. Cells were chemically crosslinked by the addition of one-tenth volume of fresh 11% formaldehyde solution for 15 minutes at room temperature. Cells were rinsed twice with 1X PBS and harvested using a silicon scraper and flash frozen in liquid nitrogen. Cells were stored at –80ºC prior to use. Cells were resuspended, lysed in lysis buffers and sonicated to solubilize and shear crosslinked DNA. Sonication conditions vary depending on cells, culture conditions, crosslinking and equipment.

For LSD1, Mi-2b, HDAC1, HDAC2, REST, CoREST, p300, H3K27Ac, and H3K4me1 the sonication buffer was 20mM Tris-HCl pH8, 150mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100. We used a Misonix Sonicator 3000 and sonicated at approximately 24 watts for 10 x 30 second pulses (60 second pause between pulses). Samples were kept on ice at all times. The resulting whole cell extract was incubated overnight at 4 degrees C with 100ul of Dynal Protein G magnetic beads that had been pre-incubated with approximately 10 ug of the appropriate antibody. Beads were washed 1X with the sonication buffer, 1X with 20mM Tris-HCl pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1%Triton X-100, 1X with 10mM Tris-HCl pH8, 250mM LiCl, 2mM EDTA, 1% NP40 and 1X with TE containing 50 mM NaCl.

Bound complexes were eluted from the beads (50 mM Tris-Hcl, pH 8.0, 10 mM EDTA and 1% SDS) by heating at 65 degrees C for 1 hour with occasional vortexing and crosslinking was reversed by overnight incubation at 65°C. Whole cell extract DNA reserved from the sonication step was also treated for crosslink reversal.

Gene Specific ChIPs (Supplementary Fig. 9, 10, and 11)

For gene specific ChIPs carried out in ZHBTc4 ESCs, approximately $5x10^7$ ES cells were grown on two 15cm plates. Plates were treated or not treated with 2ng/ml doxycyline for 12 hours, and crosslinked. SYBR Green real-time qPCR was carried out on the 7000 ABI Detection System according to the manufacturer protocol (Applied Biosystems). Data was normalized to the whole cell extract and control regions. Primer pairs are listed below.

Lefty1

5'-GTAGCCAGCAGACAGGACAA-3'

5'-ATCCCCAATCCACATTCACT-3'

Lefty2

5'-AGGCCTAGCTTTTGCATCAC-3'

5'-TCTCCCAGAGTCGATCTTCC-3'

Sall4

5'- GAAATAAACATCTGGGAGAAGGA-3'

5’- GGAAACCCCAGATTGAGAGA-3’

Sox2

5’- TGGCGAGTGGTTAAACAGAG-3’

5’- TAGCGAGAACTAGCCAAGCA-3’

Trim28

5’- GGTCTGCAATTGAAGGAAGG-3’

5’- TTAAACAGCAGGGGGTAAGG-3’

Esrrb

5’- CGAGCTTCAGCTGGCTATTT-3’

5’- GAGCTCCAGATCCCCTACAC-3’

Nanog

5’-GGAATTTCCTTCCCAGGTTT-3’

5’- GGTTGGAACTAGCTGTGTGG-3’

Ctrl (Olfr460)

5’-AACTGTTATTGTGCCCGTGA -3’

5’-CATTGCTCCAAGCAAAGAAA -3’

ChIP-Seq Sample Preparation and Analysis

All protocols for Illumina/Solexa sequence preparation, sequencing and quality control are provided by Illumina (http://www.illumina.com/pages.ilmn?ID=203). A brief summary of the technique and minor protocol modifications are described below.

Sample Preparation

Purified chromatin immunoprecipitated (ChIP) DNA was prepared for sequencing according to a modified version of the Illumina/Solexa Genomic DNA protocol. Approximately 200 ng of ChIP DNA was prepared for ligation of Solexa linkers by repairing the ends and adding a single adenine nucleotide overhang to allow for directional ligation. A 1:200 dilution of the Adaptor Oligo Mix (Illumina) was used in the ligation step. A subsequent PCR step with 18 amplification cycles added additional linker sequence to the fragments to prepare them for annealing to the Genome Analyzer flow-cell. Amplified material was purified by Qiaquick MinElute (Qiagen, Valencia, CA) and a narrow range of fragment sizes was selected by separation on a 2% agarose gel and excision of a band between 150-300bp, representing ChIP fragments between 50 and 200 nt in length and ~100bp of primer sequence. The DNA was purified from the agarose and diluted to 10nM for loading on the flow cell.

For multiplexed samples, libraries were prepared using the Illumina TruSeq adapters (to enable multiplexing) and prepared using Beckman-Coulter's SPRIworks system. For library preparations, ChIP samples were used in their entirety and whole-cell extracts control samples were prepared using 100ng. Adapters were diluted to 1:200.

Size selection was 200–400bp before PCR, and the samples were amplified using KAPA Hi-Fi polymerase and 18 cycles of PCR according to manufacturer's cycling recommendations. Amplified material was purified using Agencourt Ampure XP beads using a 0.93 ratio of beads to sample.

Polony Generation and Sequencing

The DNA library (2–5pM) was applied to the flow-cell (8 samples per flow-cell) using the Cluster Station device from Illumina. The concentration of library applied to the flow-cell was calibrated such that polonies generated in the bridge amplification step originate from single strands of DNA. Multiple rounds of amplification reagents were flowed across the cell in the bridge amplification step to generate polonies of approximately 1,000 strands in 1mm diameter spots. Double-stranded polonies were visually checked for density and morphology by staining with a 1:5000 dilution of SYBR Green I (Invitrogen) and visualizing with a microscope under fluorescent illumination. Validated flow-cells were stored at 4°C until sequencing. Flow-cells were removed from storage and subjected to linearization and annealing of sequencing primer on the Cluster Station. Primed flow-cells were loaded into the Illumina Genome Analyzer II or Hi-seq. After the first base was incorporated in the Sequencing by-Synthesis reaction the process was paused for a quality control checkpoint. A small section of each lane was imaged and the average intensity value for all four bases was compared to minimum thresholds. Flow-cells with low first base intensities were reprimed and if signal was not recovered the flow-cell was aborted. Flow-cells with signal intensities meeting the minimum thresholds were resumed and sequenced for 26

cycles. For multiplexed samples Truseq V2.5 kits were used to cluster them on the cBot and Truseq V2 were used to do a multiplex 40+7 cycle run on the Hi-seq

ChIP-Seq Data Analysis

Images acquired from the Illumina/Solexa sequencer were processed through the bundled Solexa image extraction pipeline which identified polony positions, performed base-calling and generated QC statistics.

ChIP-Seq reads were aligned using the software Bowtie[25] to NCBI build 36 (mm8) of the mouse genome with default settings. Sequences uniquely mapping to the genome with zero or one mismatch were used in further analysis

When multiple reads mapped to the same genomic position, a maximum of two reads mapping to the same position were used. ChIP-Seq datasets profiling the genomic occupancy of H3K36me2 (ref. 26), H3K79me2 (ref. 26), H3K4me3 (ref. 26), H3K27me3 (ref. 27), Oct4 (ref. 26), Sox2 (ref. 26), Nanog26, RNA Polymerase 2 (ref. 28), TBP29, Med1 (ref. 29), p300 (ref. 22), H3K4me1 (ref. 21), H3K27ac23 and Suz12 (ref. 26) in mouse ESCs were obtained from previous publications. Below is the list of ChIP-Seq datasets used and corresponding GEO Accession numbers.

Dataset	**GEO Accession Numbers**
H3K27ac	GSM594578, GSM594579
H3K27me3	GSM307619

H3K36me3	GSM307152, GSM307153
H3K4me1	GSM281695
H3K4me3	GSM307146
H3K79me2	GSM307150, GSM307151
Med1	GSM560347, GSM560348
WCE mES	GSM307154, GSM307155, GSM560357
WCE mES (matching H3K4me1 mES)	GSM307625
Nanog	GSM307140, GSM307141
Oct4	GSM307137
p300	GSM288359
RNA Pol II	GSM318444
Sox2	GSM307138, GSM307139
Suz12	GSM307144, GSM307145
TBP	GSM555160, GSM555162
DMSO H3K4me1	*This Paper*
48HR DMSO H3K4me1	*This Paper*
48HR TCP H3K4me1	*This Paper*
HDAC1	*This Paper*
HDAC2	*This Paper*
LSD1	*This Paper*
Mi-2b	*This Paper*
REST	GSM656525, *This Paper*

CoREST	*This Paper*
WCE – DMSO	*This Paper*
WCE – 48HR DMSO	*This Paper*
WCE – 48HR TCP	*This Paper*

All ChIP-Seq datastsets, including those obtained elsewhere, were analyzed using the methods described below.

Analysis methods were derived from previously published methods[26,27,30,31]. Sequence reads from multiple flow cells for each IP target and/or biological replicates were combined. For all datasets, excluding H3K4me1, H3K4me3, H3K79me2, H3K36me3, H3K27ac, and H3K27me3 each read was extended 200bp, towards the interior of the sequenced fragment, based on the strand of the alignment. For H3K4me1, H3K4me3, H3K79me2, H3K36me3, H3K27ac, and H3K27me3 datasets, each read was extended 600bp towards the interior and 400bp towards the exterior of the sequenced fragment, based on the strand of the alignment. Across the genome, in 25 bp bins, the number of extended ChIP-Seq reads was tabulated. The 25bp genomic bins that contained statistically significant ChIP-Seq enrichment were identified by comparison to a Poissonian background model. Assuming background reads are spread randomly throughout the genome, the probability of observing a given number of reads in a genomic bin can be modeled as a Poisson process in which the expectation can be estimated as the number of mapped reads multiplied by the number of bins (8 for all sequences datasets except H3K4me1, H3K4me3, H3K79me2, H3K36me3, H3K27ac,

and H3K27me3, which was 40) into which each read maps, divided by the total number of bins available (we estimated 70% of the genome). Enriched bins within 200bp of one another were combined into regions.

The Poissonian background model assumes a random distribution of background reads, however we have observed significant deviations from this expectation. Some of these non-random events can be detected as sites of apparent enrichment in negative control DNA samples and can create many false positives in ChIP-Seq experiments. To remove these regions, we compared genomic bins and regions that meet the statistical threshold for enrichment to a set of reads obtained from Solexa sequencing of DNA from whole cell extract (WCE) in matched cell samples. We required that enriched bins and enriched regions have five-fold greater ChIP-Seq density in the specific IP sample, compared with the control sample, normalized to the total number of reads in each dataset. This served to filter out genomic regions that are biased to having a greater than expected background density of ChIP-Seq reads. A summary of the bound regions and genes for each antibody is provided (Supplementary Table 1).

Assigning ChIP-Seq Enriched Regions to Genes (Supplementary Table 1)

The complete set of RefSeq genes was downloaded from the UCSC table browser (http://genome.ucsc.edu/cgi-bin/hgTables) on June 1, 2010. For all datasets except H3K4me1, H3K4me3, H3K79me2, H3K36me3, H3K27ac, and H3K27me3, genes with enriched regions within 10kb of their transcription start site were called bound.

For H3K4me1, H3K4me3, H3K79me2, H3K36me3, H3K27ac, and H3K27me3 datasets, genes with enriched regions within the gene body were called bound.

Enrichment of LSD1, Pol II, and Suz12 at genes (Fig. 1)

Each gene in the mouse genome was classified into active, bivalent, or silent groups based on the presence of co-occupancy of H3K4me3 (GSE11724), H3K79me2 (GSE11724), and H3K27me3 (GSM307619). See the table below for a description of the number of genes in each category as well as the classification rules. Using a hypergeometric test, the p-value for enrichment of LSD1, Pol II, and Suz12 binding to the genes in each of these classes was determined.

	Classification rules			
	H3K4me3 within +/- 2kb of the TSS	H3K79me2 within first 5kb of gene body	H3K27me3 within +/- 5kb of the TSS	Number of genes
Active	Required	Required	-	7339
Bivalent	Required	-	Required	3094
Silent	Absent	Absent	-	6976

Definition of Enhancer and Core Promoter (Fig. 1, 4, Supplementary Fig. 8, 13, and Supplementary Table 1)

An enhancer was defined as a Med1 high-confidence enriched region that is overlapped at least 1bp by enriched regions of Oct4, Sox2, and Nanog in mES cells. Using this definition we identified 3,838 enhancers, and the genomic locations of those enhancers are listed in Supplementary Table 1. These enhancers were assigned to the nearest genes to determine the neighboring core promoter regions.

Determining LSD1 density at Enhancer and Core Promoter (Supplementary Fig. 1)

The mean LSD1 density at enhancers (15.91) and core promoters (13.46) was determined by calculating the sum of LSD1 density within 5kb of each enhancer and core promoter, and then taking the average across the 3,838 enhancers and core promoters. A t-test was then performed to obtain a p-value ($p < 10^{-16}$) indicating that LSD1 density was statistically larger at enhancers compared to core promoters.

ChIP-Seq Density Heatmaps (Fig. 1, 4, Supplementary Fig. 8, and 13)

Selected enriched regions were aligned with each other according to the position of either Med1 (Fig. 1c, 4e, and Supplementary Fig. 5b) or the TSS (Fig. 1d). For each experiment, the ChIP-Seq density profiles were normalized to the density per million total reads. Heatmaps were generated using Java Treeview (http://jtreeview.sourceforge.net/) with color saturation as indicated.

Calculation of the Statistical Significance of the Overlap Between Sets of Genomic Regions (Fig. 1, 3, 4, and Supplementary Fig. 8)

In the manuscript, when assessing the overlap between two sets of genomic regions,

we report p-values calculated using a Monte Carlo method. In order to make this type of comparison, a number of different statistical methods could be applied. One could use a chi-square test, and compare the total size of the overlap of two sets of genomic regions to the size of this overlap that would be expected at random based on the total size of the genome and the total size of the genomic regions in each group. This method accurately models the expected overlap between two datasets if they were truly randomly associated. Unfortunately, since the genome is so large, this method will call the overlap between two datasets statistically significant even if it is not much larger than would be expected at random.

Another way to assess the statistical significance of the overlap between two sets of genomic regions is to use a Monte Carlo method. In Monte Carlo methods, a tremendous number of simulated datasets are created and the overlaps between the simulated datasets are tabulated. The statistical significant of the actual overlap is then estimated based on the frequency of an overlap at least that large occurring in the simulated datasets. We used a Monte Carlo method to assess the statistical significance of the overlap between two sets of genomic regions in this manuscript. To create the simulated datasets, we shifted each of the genomic regions in one of the datasets a random distance between -2,000bp and +2,000bp from its original position. For this manuscript, Monte Carlo simulations were run one billion times.

We noted that, if two sets of genomic regions that tend to occur in the same place, then this overlap is statistically significant. This will be true whether the regions occur

together 10% of the time, or 100% of the time. Thus, the test for statistical significance is not a very good at describing the degree of overlap between to datasets. Consequently, in each instance that we assessed the statistical significance of the overlap between two set of genomic regions, we also calculated the fold enrichment of the overlap between the two datasets. This was calculated as the total size of the actual overlap between the two sets of regions and the overlap between the two datasets that would be expected at random.

Rank Normalization of H3K4me1 ChIP-Seq Data (Fig. 4, Supplementary Fig. 13, and Supplementary Table 5)

In order to compare the levels of H3K4me1 between three conditions a rank normalization method was used. The three cells types and conditions in which H3K4me1 occupancy was profiled were ZHBTc4 ESCs, ZHBTc4 ESCs treated with doxycycline for 48 hours, and ZHBTc4 ESCs treated with doxycycline for 48 hours and TCP for 54 hours. In each of the three datasets, the genomic bin with the greatest ChIP-Seq signal was identified. The average of these three values was calculated, and the greatest bin in each dataset was assigned this average value. This was repeated for all genomic bins from the greatest signal to the least, assigning each the average ChIP-Seq signal for all bins of that rank across all datasets. Subsequently, the total number of ChIP-Seq reads in the three H3K4me1 datasets was calculated, and this value was used to scale each dataset to the units, rank normalized reads/million.

Using this method, we computed the rank normalized reads/million density in each of the three datasets in 10 bp bins across the genome. The H3K4me1 signal at an enhancer was then calculated as the sum of the observed densities in the 50 bins (+/- 250bp) surrounding each of the 3,838 identified enhancers. This number captures the normalized H3K4me1 signal in dataset, and the value for each enhancer in each of the three datasets is reported in Supplementary Table 5.

Gene Ontology (GO) Analysis (Supplementary Fig. 7)

Gene ontology analysis was performed using the web tool David Bioinformatics Database (http://david.abcc.ncifcrf.gov/home.jsp)[32,33]. The complete set of all RefSeq genes was used as a background.

Public availability of ChIP-Seq datasets

ChIP-Seq data have been submitted to the Gene Expression Omnibus Database (http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE27844 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=jfcvlkoimyymyzm&acc=GSE27844).

Microarray Analysis (Fig. 2)

Cell Culture and RNA isolation

For ESC expression analysis during differentiation following either DMSO or TCP treatment, ZHBTc4 ESCs were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs and plated in 6-well plates. The following day

cells were treated with either DMSO or 1mM TCP for 6 hours in ESC media. After 6 hours a subset of cells from DMSO treated wells was removed for TRIzol RNA isolation, the media was removed and replaced with ESC media containing 2μg/ml doxycyline and either DMSO or 1mM TCP. 24 hours after doxycycline addition, a subset of cells was removed for RNA isolation and the media was replaced with ESC media containing doxycycline and either DMSO or TCP. After another 48 hours, RNA was isolated using TRIzol (Invitrogen, 15596-026) from all time points (at indicated times), further purified with RNeasy columns (Qiagen, 74104) and DNase treated on column (Qiagen, 79254) following the manufacturer's protocols.

Micorarray hybridization and Analysis

For microarray analysis, Cy3 and Cy5 labeled cRNA samples were prepared using Agilent's QuickAmp sample labeling kit starting with 1ug total RNA. Briefly, double-stranded cDNA was generated using MMLV-RT enzyme and an oligo-dT based primer. *In vitro* transcription was performed using T7 RNA polymerase and either Cy3-CTP or Cy5-CTP, directly incorporating dye into the cRNA.

Agilent mouse 4x44k expression arrays were hybridized according to our laboratory's standard method, which differs slightly from the standard protocol provided by Agilent. The hybridization cocktail consisted of 825 ng cy-dye labeled cRNA for each sample, Agilent hybridization blocking components, and fragmentation buffer. The hybridization cocktails were fragmented at 60 degrees C for 30 minutes, and then Agilent 2X hybridization buffer was added to the cocktail prior to application to the

array. The arrays were hybridized for 16 hours at 60 degrees C in an Agilent rotor oven set to maximum speed. The arrays were treated with Wash Buffer #1 (6X SSPE / 0.005% n-laurylsarcosine) on a shaking platform at room temperature for 2 minutes, and then Wash Buffer #2 (0.06X SSPE) for 2 minutes at room temperature. The arrays were then dipped briefly in acetonitrile before a final 30 second wash in Agilent Wash 3 Stabilization and Drying Solution, using a stir plate and stir bar at room temperature.

Arrays were scanned using an Agilent DNA microarray scanner. Array images were quantified and statistical significance of differential expression for each hybridization was calculated using Agilent's Feature Extraction Image Analysis software with the default two-color gene expression protocol. For each gene in the RefSeq gene list (see ChIP-Seq analysis section), the log10 ratio values (ESC+Dox / ESC, or ESC+DOX+TCP / ESC) and p-value for that gene were determined by averaging log10 ratios and p-values of all Agilent Features annotated to that gene. Genes with no annotated features were reported as NA (Supplementary Table 3). Genes with a log10 ratio of at least 0.0969 (1.25x) and a p-value less than or equal to 10^{-2} were determined to have a significant change in expression.

Public availability of microarray gene expression datasets

Microarray gene expression data have been submitted to the Gene Expression Omnibus Database (http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE27844 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=jfcvlkoimyymyzm&acc=GSE

27844).

Protein Extraction and Western Blot Analysis (Fig. 2, Supplementary Fig. 1, 4, 5, 6, and 9)

ESCs were lysed with CelLytic Reagent (Sigma, C2978-50ml) containing protease inhibitors (Roche). After SDS-PAGE, Western blots were revealed with antibodies against Oct4 (Santa Cruz Biotechnology, sc-5279), LSD1 (Abcam, ab17721), Mi-2b (Abcam, ab72418), CoREST (Abcam, ab32631), p300 (Santa Cruz, sc-584), GAPDH (Abcam, ab9484) or Tubulin (Millipore, 05-661).

ChIP-Western and Co-Immunoprecipitation (Fig. 3)

For ChIP-Western, same conditions as for ChIP-Seq were used. For co-immunoprecipitation, murine ES cells were harvested in cold PBS and extracted for 30 min at 4 degrees C in TNEN250 (50 mM Tris pH 7.5, 5 mM EDTA, 250 mM NaCl, 0.1% NP-40) with protease inhibitors. After centrifugation, supernatant was mixed to 2 volumes of TNENG (50 mM Tris pH 7.5, 5 mM EDTA, 100 mM NaCl, 0.1% NP-40, 10% glycerol). Protein complexes were immunoprecipitated overnight at 4 degrees C using 5 micrograms of either LSD1 (Abcam, ab17721), HDAC1 (Abcam, ab7028), HDAC2 (Abcam, ab7029), Mi-2b (Abcam, ab72418), CoREST (Abcam, ab32631) or Rabbit IgG (Upstate, 12-370), bound to 50ul of Dynabeads®. Immunoprecipitates were washed three times with TNEN125 (50 mM Tris pH 7.5, 5 mM EDTA, 125 mM NaCl, 0.1% NP-40). For both ChIP-Western and co-immunoprecipitation, beads were boiled for 10 minutes in XT buffer (Biorad) containing 100mM DTT to elute proteins.

After SDS-PAGE, Western blots were revealed with antibodies against either LSD1 (Abcam, ab17721), or HDAC1 (Abcam, ab7028).

Supplementary References

1 Kent, W.J. *et al*., The human genome browser at UCSC. *Genome Res* 12 (6), 996-1006 (2002).

2 Forneris, F., Binda, C., Vanoni, M.A., Mattevi, A., & Battaglioli, E., Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. *FEBS Lett* 579 (10), 2203-2207 (2005).

3 Shi, Y. *et al*., Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119 (7), 941-953 (2004).

4 Lee, M.G., Wynder, C., Schmidt, D.M., McCafferty, D.G., & Shiekhattar, R., Histone H3 lysine 4 demethylation is a target of nonselective antidepressive medications. *Chem Biol* 13 (6), 563-567 (2006).

5 Schmidt, D.M. & McCafferty, D.G., trans-2-Phenylcyclopropylamine is a mechanism-based inactivator of the histone demethylase LSD1. *Biochemistry* 46 (14), 4408-4416 (2007).

6 Yang, M. *et al*., Structural basis for the inhibition of the LSD1 histone demethylase by the antidepressant trans-2-phenylcyclopropylamine. *Biochemistry* 46 (27), 8058-8065 (2007).

7 Mimasu, S. *et al*., Structurally designed trans-2-phenylcyclopropylamine derivatives potently inhibit histone demethylase LSD1/KDM1. *Biochemistry* 49 (30), 6494-6503.

8 Liang, Y., Vogel, J.L., Narayanan, A., Peng, H., & Kristie, T.M., Inhibition of the histone demethylase LSD1 blocks alpha-herpesvirus lytic replication and reactivation from latency. *Nat Med* 15 (11), 1312-1317 (2009).

9 Foster, C.T. *et al*., Lysine-specific demethylase 1 regulates the embryonic transcriptome and CoREST stability. *Mol Cell Biol* 30 (20), 4851-4863 (2010).

10 Wang, J. *et al*., The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat Genet* 41 (1), 125-129 (2009).

11 Wang, J. *et al*., Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* 446 (7138), 882-887 (2007).

12 Adamo, A. *et al*., LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nat Cell Biol* 13 (6), 652-660 (2011).

13 Boyer, L.A. *et al*., Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122 (6), 947-956 (2005).

14 Binda, C. *et al*., A 30-angstrom-long U-shaped catalytic tunnel in the crystal structure of polyamine oxidase. *Structure* 7 (3), 265-276 (1999).

15 Lin, W. *et al*., Developmental potential of Gcn5(-/-) embryonic stem cells in vivo and in vitro. *Developmental dynamics : an official publication of the American Association of Anatomists* 236 (6), 1547-1557 (2007).

16 Ura, H. *et al*., Eed/Sox2 regulatory loop controls ES cell self-renewal through histone methylation and acetylation. *EMBO J* 30 (11), 2190-2204 (2011).

17 Miyabayashi, T. *et al*., Wnt/beta-catenin/CBP signaling maintains long-term murine embryonic stem cell pluripotency. *Proceedings of the National*

Academy of Sciences of the United States of America 104 (13), 5668-5673 (2007).

18 Bowers, E.M. *et al*., Virtual ligand screening of the p300/CBP histone acetyltransferase: identification of a selective small molecule inhibitor. *Chemistry & biology* 17 (5), 471-482 (2010).

19 Wang, Z. *et al*., Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* 138 (5), 1019-1031 (2009).

20 Singh, S.K., Kagalwala, M.N., Parker-Thornburg, J., Adams, H., & Majumder, S., REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature* 453 (7192), 223-227 (2008).

21 Meissner, A. *et al*., Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454 (7205), 766-770 (2008).

22 Chen, X. *et al*., Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133 (6), 1106-1117 (2008).

23 Creyghton, M.P. *et al*., Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* (2010).

24 Boyer, L.A. *et al*., Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441 (7091), 349-353 (2006).

25 Langmead, B., Trapnell, C., Pop, M., & Salzberg, S.L., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* 10 (3), R25 (2009).

26 Marson, A. *et al*., Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134 (3), 521-533 (2008).

27 Mikkelsen, T.S. *et al*., Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448 (7153), 553-560 (2007).

28 Seila, A.C. *et al*., Divergent transcription from active promoters. *Science* 322 (5909), 1849-1851 (2008).

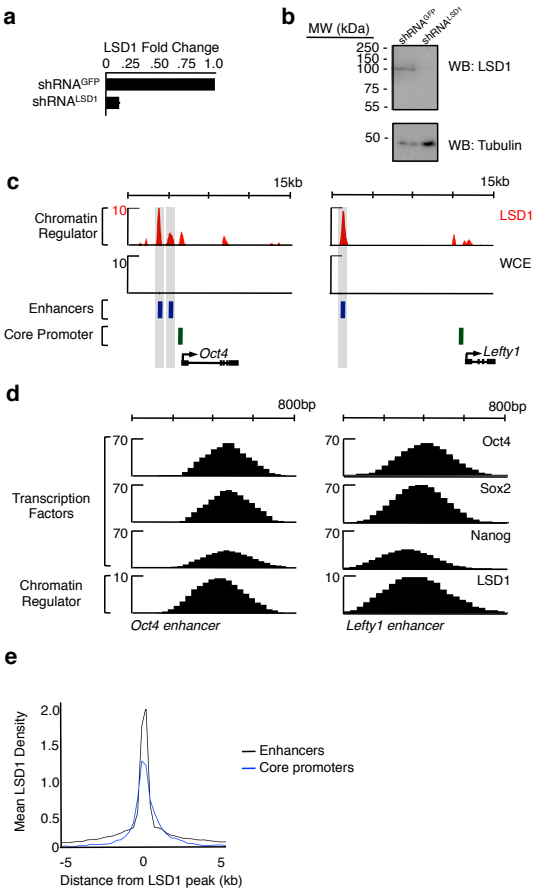
29 Kagey, M.H. *et al*., Mediator and cohesin connect gene expression and chromatin architecture. *Nature* (2010).

30 Guenther, M.G. *et al*., Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes Dev* 22 (24), 3403-3408 (2008).

31 Johnson, D.S., Mortazavi, A., Myers, R.M., & Wold, B., Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316 (5830), 1497-1502 (2007).

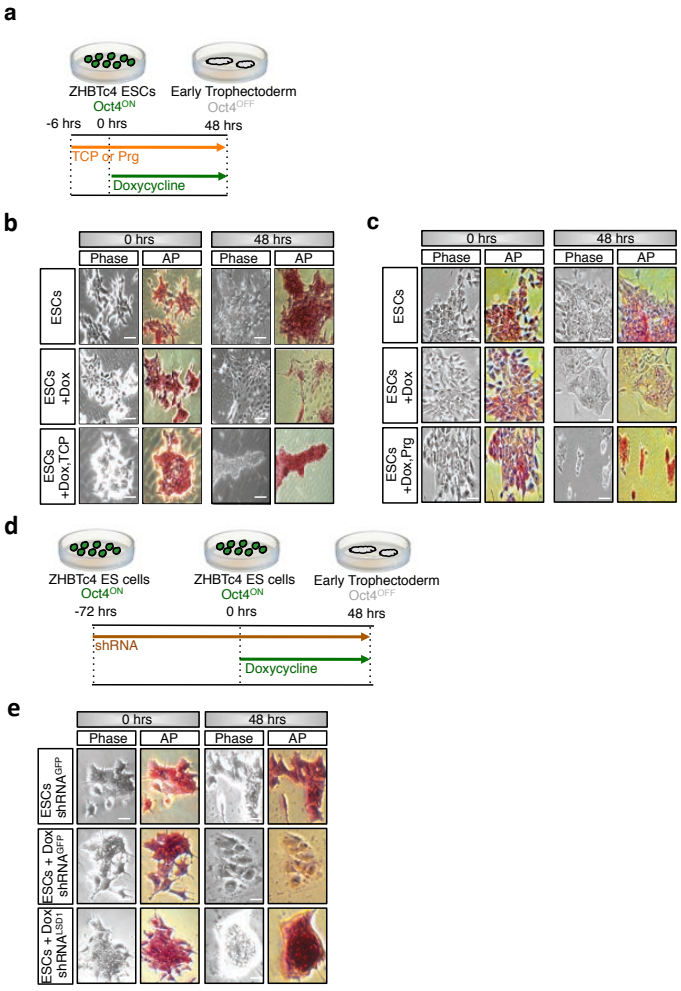
32 Huang da, W., Sherman, B.T., & Lempicki, R.A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4 (1), 44-57 (2009).

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



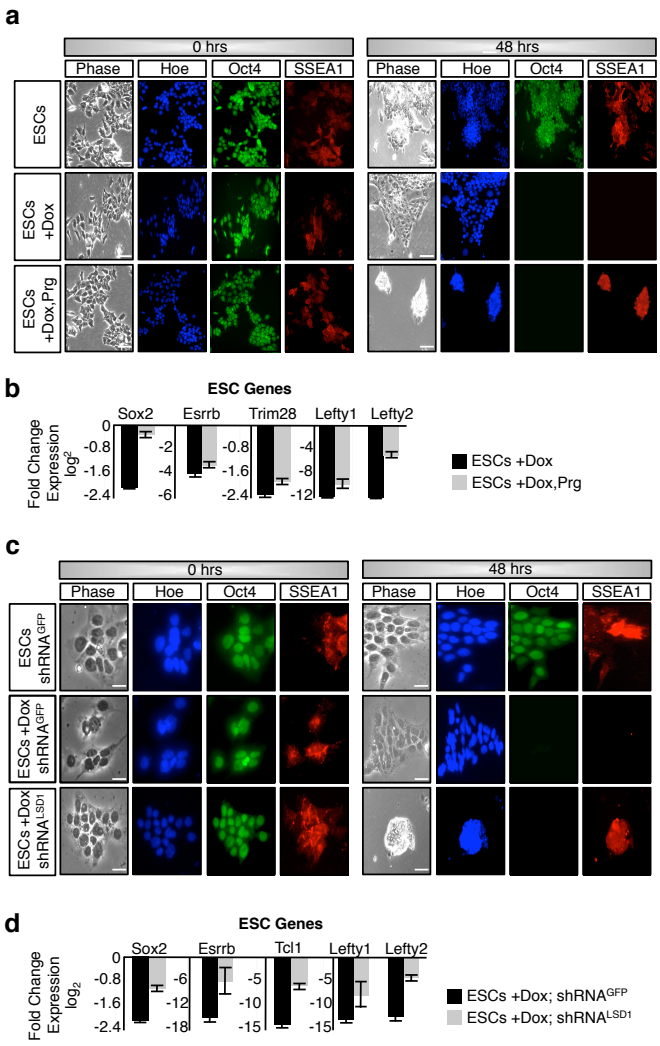
Supplementary Fig. 1: Validation and characterization of LSD1 ChIP-Seq dataset

a, Validation of shRNA targeting LSD1. qPCR of LSD1 transcripts in v6.5 ESCs infected with short hairpins (Open Biosystems) against either LSD1 or GFP (control) for 5 days. The data is normalized to Gapdh. The error bars represent the standard deviation of triplicate PCR reactions. **b,** LSD1 antibody is specific. Western blot of LSD1 protein levels in GFP or LSD1 knockdown cells, showing a decrease in signal of one band corresponding to LSD1 molecular weight (MW = 92.9 kDa). Tubulin served as loading control. **c,** LSD1 ChIP-Seq signals are above experimental noise. ChIP-Seq binding profiles (reads/million) for LSD1 at the *Oct4 (Pou5f1)* and *Lefty1* loci in ESCs, with the y-axis floor set to 1. Gene models, and previously described enhancer regions[1-3] are depicted below the binding profiles. **d,** ChIP-Seq data, at high resolution, showing coincidental binding of Oct4, Sox2, and Nanog with LSD1. High-resolution ChIP-Seq binding profiles (reads/million) at the *Oct4* (*Pou5f1*) and *Lefty1* enhancer loci in ESCs, with y-axis floor at 1.0 reads/million. **e,** Metagene showing mean LSD1 density is significantly higher at the 3,838 ESC enhancers compared to 3,838 neighboring core promoters ($p<10^{-16}$).

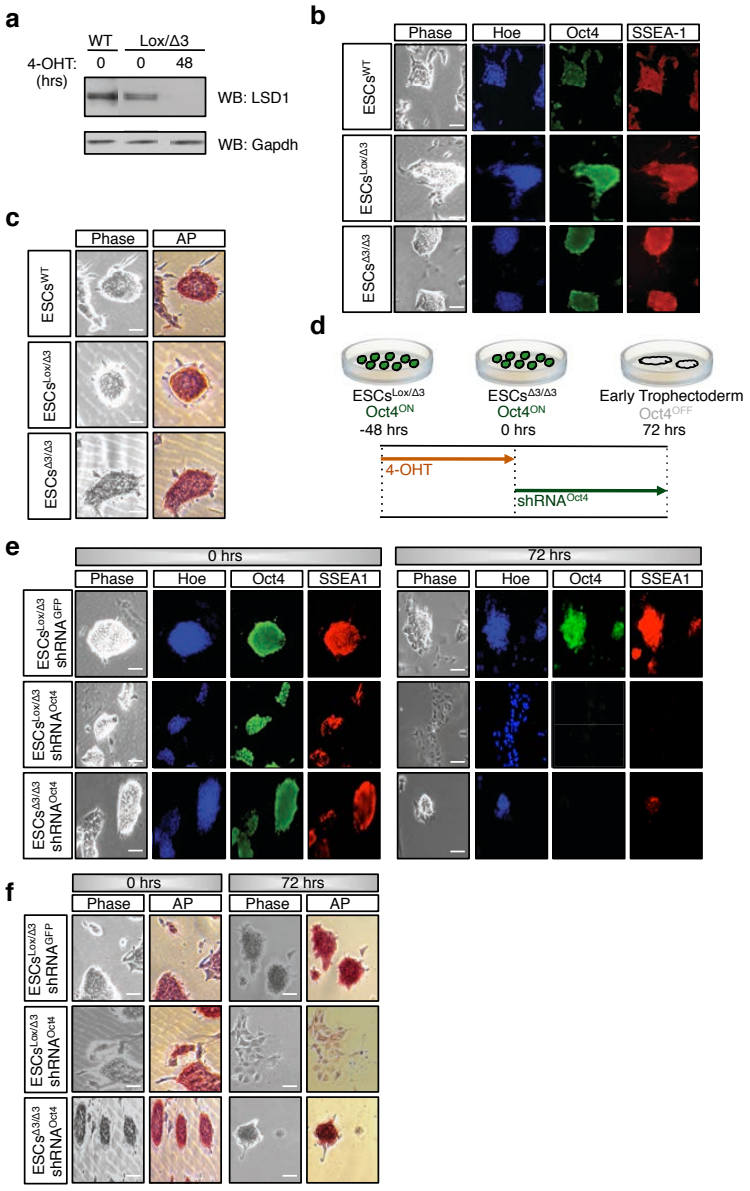


Supplementary Fig. 2: LSD1 is required for doxcycline-induced differentiation of ZHBTc4 ESCs

a, Schematic representation of trophectoderm differentiation assay using doxycycline-inducible Oct4 depletion murine ESC line ZHBTc4. Treatment with doxycyline for 48 hours leads to depletion of Oct4 and early trophectoderm specification. ZHBTC4 cells were treated with DMSO (control) or the LSD1 inhibitors Tranylcypromine (TCP) and Pargyline (Prg) for 6 hours before 2µg/ml doxycycline was added for an additional 48 hours. **b, c,** ESCs treated with either (b) 1mM tranylcypromine (TCP) or (c) 3mM pargyline (Prg) maintained ESC colony morphology and alkaline phosphatase (AP) staining in doxycycline-treated cells despite depletion of Oct4 protein levels**.** Scale bar = 20µM**. d,** Schematic representation of trophectoderm differentiation assay using doxycycline-inducible Oct4 depletion murine ESC line ZHBTc4. ESCs were infected with a GFP or LSD1 shRNA for 72 hours before 2µg/ml doxycycline was added for an additional 48 hours. **e,** ESCs infected with an shRNA targeting LSD1 maintained ESC colony morphology and alkaline phosphatase (AP) staining in doxycycline-treated cells despite depletion of Oct4 protein levels. Scale bar = 20µM.

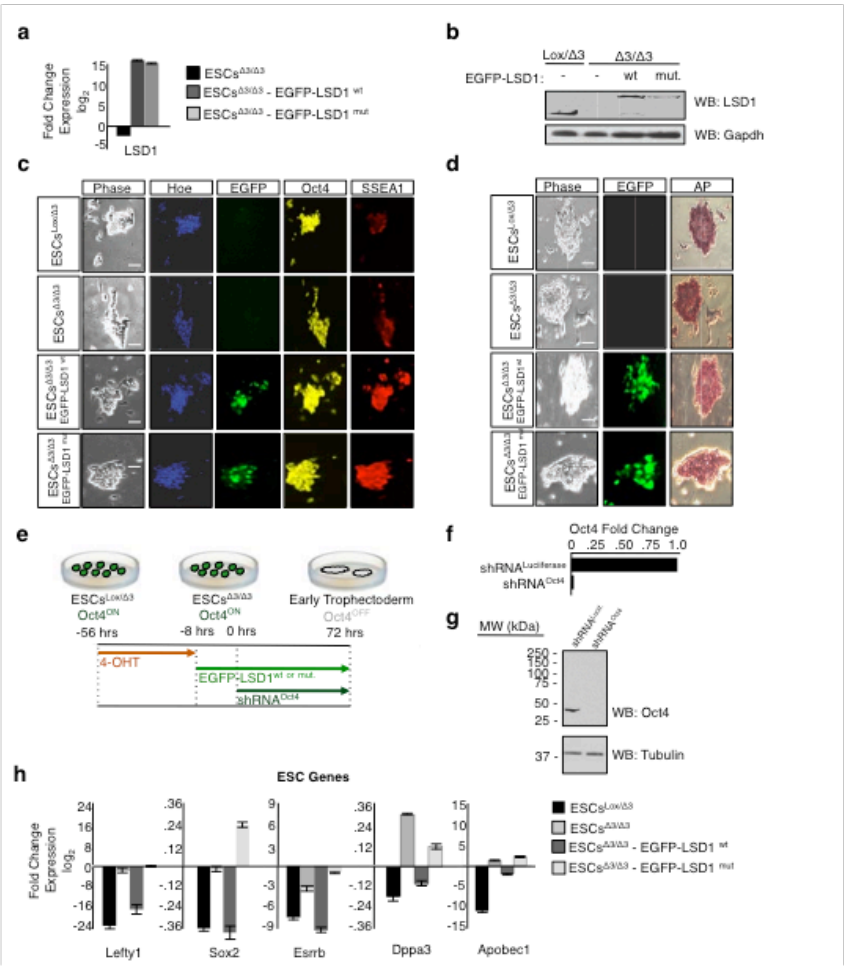


Supplementary Fig. 3: LSD1 is required for repression of ESC genes
a, Treatment of Prg partially maintained SSEA-1 cell surface marker expression in doxycycline-treated cells. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 100μM. **b,** Treatment of Prg partially relieved repression of ESC genes 48 hours after Oct4 depletion in DMSO- versus Prg-treated cells. Error bars reflect standard deviation of either duplicate or triplicate PCR reactions. **c,** ESCs with decreased LSD1 levels partially maintained SSEA-1 cell surface marker expression in doxycycline-treated cells. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 20μM. **d,** Reduced LSD1 levels partially relieved repression of ESC genes 48 hours after Oct4 depletion in GFP- versus LSD1 shRNA. Error bars reflect standard deviation of either duplicate or triplicate PCR reactions.



Supplementary Fig. 4: Characterization of LSD1 knockout ESCs
a, Beginning with an E14 ESC line (WT), cells were engineered to express an inducible Cre-ER fusion protein from the endogenous *ROSA26* locus[12]. Sequential gene targeting produced cells in which one LSD1 allele has exon 3 flanked by LoxP sites (floxed) and the second has exon 3 deleted (Lox/Δ3). Induction of Cre activity by addition of 4-hydroxytamoxifen (4-OHT) to the growth medium resulted in complete recombination of the remaining floxed LSD1 allele, generating LSD1 homozygous null (Δ3/Δ3) cells. Loss of exon 3 disrupts the open reading frame of LSD1 such that a premature stop codon is introduced into

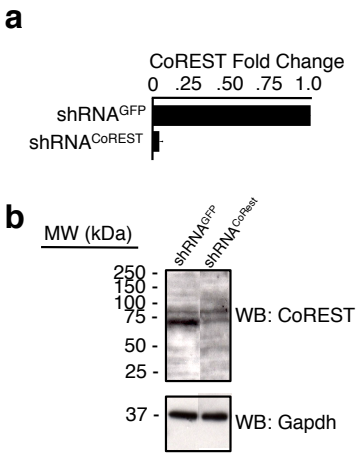
exon 4, resulting in the progressive loss of LSD1 protein by 48 hours, as detected by Western blot (WB). Gapdh served as a loading control. **b,** LSD1 heterozygous (Lox/Δ3) and homozygous null (Δ3/Δ3) ESCs maintain Oct4 and SSEA-1 cell surface marker expression compared to E14 wild type (WT) cells. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 100μM. **c,** LSD1 heterozygous (Lox/Δ3) or homozygous null (Δ3/Δ3) ESCs maintain ESC morphology and alkaline phosphatase (AP) staining compared to E14 wild type (WT) ESCs. Scale bar = 100μM. LSD1 Lox/Δ3 ESCs are used as a control throughout the remainder of the study. **d,** Schematic representation of trophectoderm differentiation assay. LSD1 Lox/Δ3 cells were treated for 48 hours with either ethanol (control) or 4-OHT to generate LSD1 Δ3/Δ3 ESCs before being infected with a GFP ($shRNA^{GFP}$, control) or Oct4 ($shRNA^{Oct4}$) shRNA to knockdown Oct4 and induce differentiation. **e,** LSD1 Δ3/Δ3 ESCs partially maintained SSEA-1 cell surface marker expression when Oct4 levels are decreased. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 100μM. **f,** LSD1 Δ3/Δ3 ESCs partially maintained ESC colony morphology and alkaline phosphatase (AP) staining when Oct4 levels are decreased.



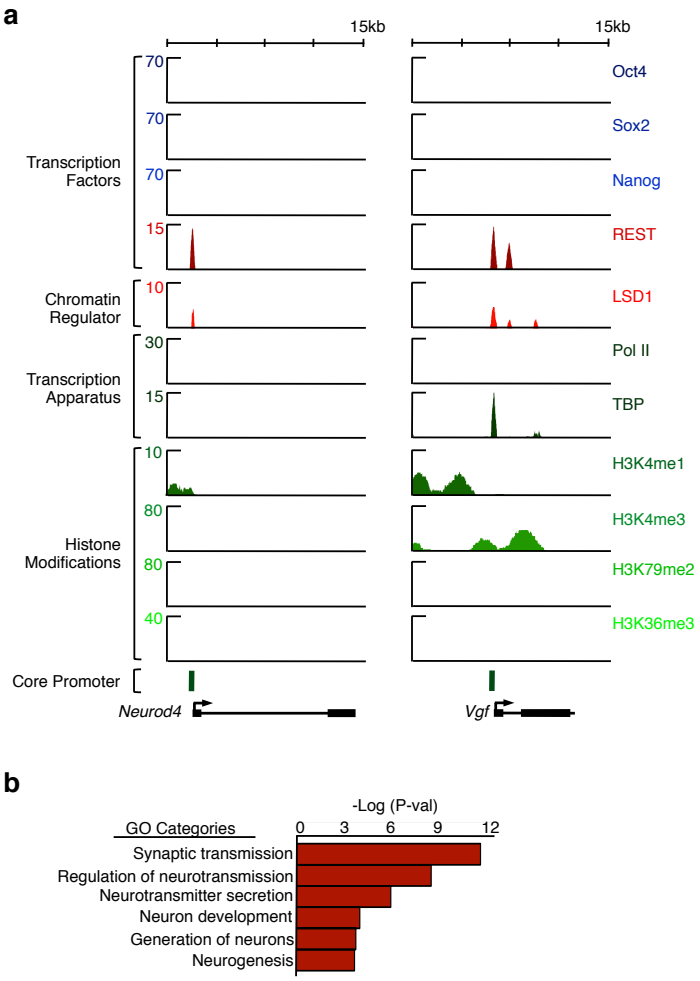
Supplementary Fig. 5: Catalytic activity of LSD1 is required for repression of ESC genes

a, Overexpression of LSD1 wildtype and catalytic mutant in LSD1 homozygous null (Δ3/Δ3) ESCs. Homozygous null (Δ3/Δ3) cells were transfected with LSD1 wildtype (EGFP-LSD1^{wt}) and a catalytic mutant (EGFP-LSD1^{mut}) fused to EGFP. LSD1 expression in Δ3/Δ3 ESCs are relative to signal in LSD1 heterozygous (Lox/Δ3) cells. The data is normalized to Gapdh. The error bars represent the standard deviation of triplicate PCR reactions. **b,** Western blot (WB) showing homozygous null (Δ3/Δ3) ESCs transfected with LSD1 transgenes express high levels of LSD1 protein. **c,** Homozygous null (Δ3/Δ3) and homozygous null (Δ3/Δ3) ESCs expressing the LSD1 transgenes maintain Oct4 and SSEA-1 cell surface marker expression compared to LSD1 heterozygous (Lox/Δ3) ESCs. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 100μM. **d,** Homozygous null (Δ3/Δ3) ESCs expressing LSD1 transgenes maintain ESC morphology and alkaline phosphatase (AP) staining compared to LSD1 heterozygous (Lox/Δ3) ESCs. Scale bar = 100μM. **e,** Schematic representation

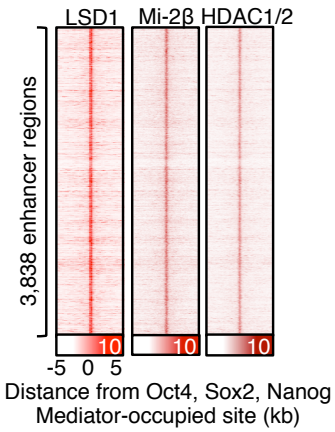
of trophectoderm differentiation assay. LSD1 Lox/Δ3 cells were treated for 48 hours with either ethanol (control) or 4-OHT to generate LSD1 Δ3/Δ3 ESCs before transfected with either an LSD1 wildtype or catalytic mutant (K661A) transgene fused to EGFP (EGFP-LSD1^{wt} or EGFP-LSD1^{mut}). Cells were transduced with an shRNA targeting either Luciferase (control) or Oct4 (shRNA^{Oct4}) to knockdown Oct4 and induce differentiation. **f,** Validation of shRNA targeting Oct4. qPCR of Oct4 transcripts in LSD1 heterozygous (Lox/Δ3) ESCs infected with a short hairpin (Open Biosystems) against either Oct4 or Luciferase (control) for 72 hours. The data is normalized to Gapdh. The error bars represent the standard deviation of triplicate PCR reactions. **g,** Western blot of Oct4 protein levels in Luciferase (control) or Oct4 knockdown cells, showing a decrease in signal of one band corresponding to Oct4 molecular weight (MW = 38 kDa). Tubulin served as loading control. **h,** Expression of ESC genes was downregulated during differentiation when wildtype, but not the catalytic mutant, LSD1 was introducted into homozygous null (Δ3/Δ3) cells. Expression of ESC genes was maintained in differentiating LSD1 null cells compared to control cells. Overexpression of wildtype LSD1, but not the catalytic mutant, rescued this phenotype. Error bars reflect standard deviation of triplicate PCR reactions.



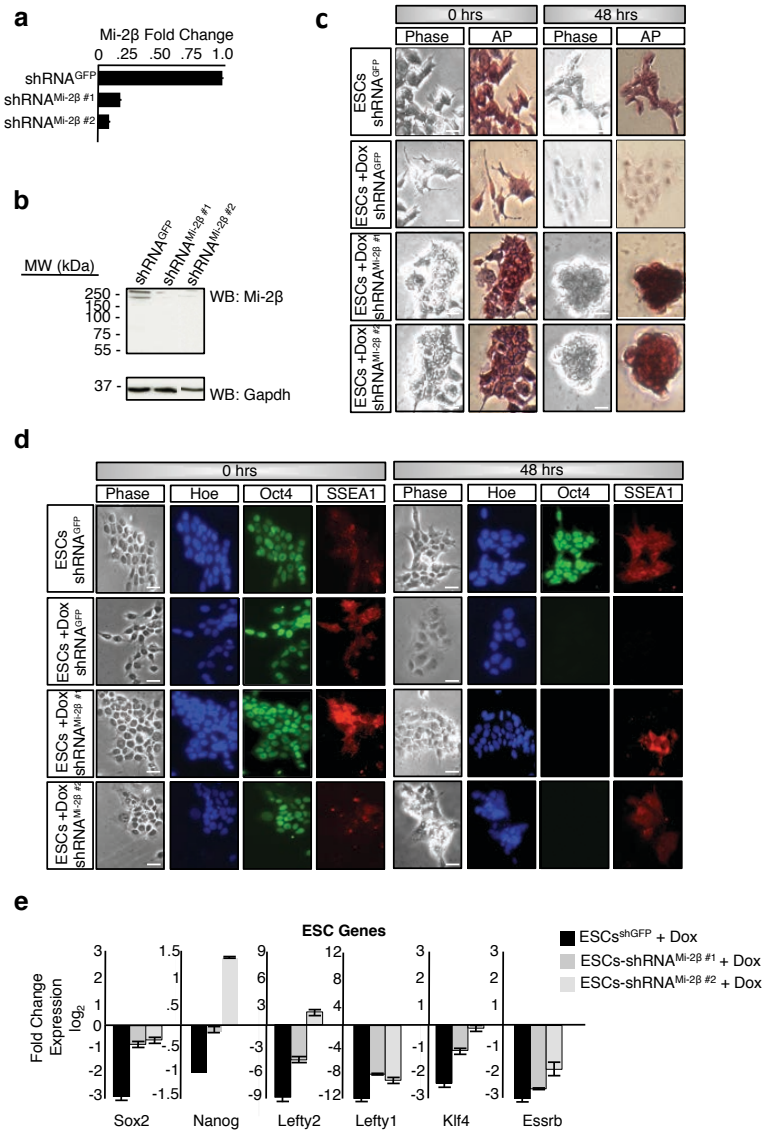
Supplementary Fig. 6: Validation of CoREST antibody
a, Validation of shRNA targeting CoREST. qPCR of CoREST transcripts in v6.5 ESCs infected with short hairpins (Open Biosystems) against either CoREST or GFP (control) for 5 days. The data is normalized to Gapdh. The error bars represent the standard deviation of triplicate PCR reactions. **b,** CoREST antibody is specific. Western blot of CoREST protein levels in GFP or CoREST knockdown cells showing a decrease in signal of one band corresponding to CoREST molecular weight (MW = 70 kDa). Gapdh served as loading control.



Supplementary Fig. 7: LSD1 occupies a subset of genomic sites with REST
a, LSD1 occupies a subset of genomic sites with REST. ChIP-Seq binding profiles (reads/million) for transcription factors (Oct4, Sox2, Nanog, REST), chromatin regulator (LSD1), the transcriptional apparatus (Pol II, TBP) and histone modifications (H3K4me1, H3K4me3, H3K79me2, H3K36me3) at the *NeuroD4* and *VGF* loci in ESCs. Gene models are depicted below the binding profiles. **b,** GO analysis on the top 500 (based on peak height) REST bound genes co-occupied by LSD1 reveals significant enrichment for genes involved in neurogenesis.



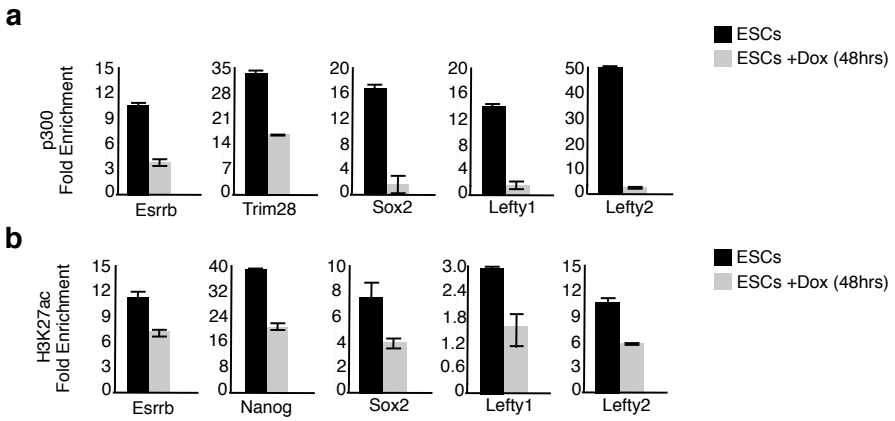
Supplementary Fig. 8: LSD1 occupies ESC enhancers with NuRD
LSD1 occupies ESC enhancer sites with Mi-2b and either HDAC1 or HDAC2. Density map of ChIP-Seq data at Oct4, Sox2, Nanog, and Med1 co-occupied enhancer regions. Data is shown for chromatin regulators (LSD1, Mi-2b, and either HDAC1 or HDAC2) in ESCs. Over 70% of the 3,838 high confidence enhancers were co-occupied by LSD1 and NuRD ($p < 10^{-9}$). Color scale indicates ChIP-seq signal in reads per million.



Supplementary Fig. 9: NuRD is required for doxycycline-induced differentiation of ZHBTc4 ESCs

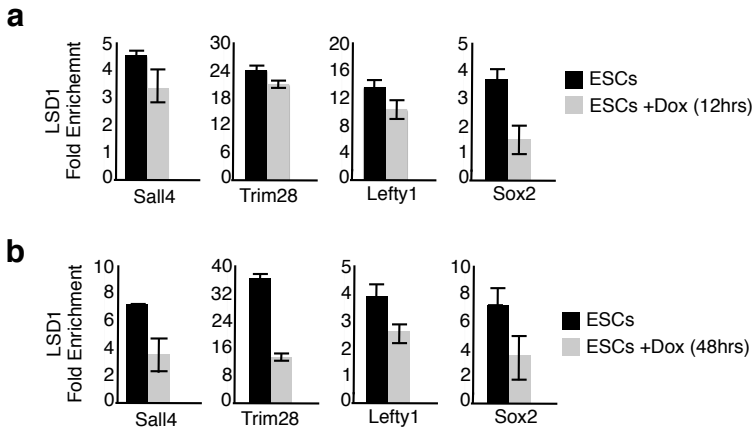
a, Validation of shRNAs targeting Mi-2b. qPCR of Mi-2b transcripts in v6.5 ESCs infected with short hairpins (Open Biosystems) against either Mi-2b or GFP (control) for 5 days. The data is normalized to Gapdh. The error bars represent the standard deviation of triplicate PCR reactions. **b,** Mi-2b antibody is specific. Western blot of Mi-2b protein levels in GFP or Mi-2b knockdown cells, showing a decrease in signal of two bands corresponding to Mi-2b molecular weight (MW = 280 and 218 kDa). Gapdh served as loading control. **c,** ESCs infected with shRNAs targeting Mi-2b partially maintained ES cell colony morphology and alkaline phosphatase (AP) staining in doxycycline-treated cells despite depletion

of Oct4 protein levels. Scale bar = 20μM**. d,** ESCs with decreased Mi-2b levels partially maintained SSEA-1 cell surface marker expression in doxycycline-treated cells. ESCs were infected with Mi-2b shRNA as represented in Supplementary Fig. 2d and were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 20μM. **e,** Mi-2b is required for downregulation of ESC genes. Expression of ESC genes was maintained 48 hours after Oct4 depletion in GFP- versus Mi-2b shRNA treated cells. Error bars reflect standard deviation of either duplicate or triplicate PCR reactions.



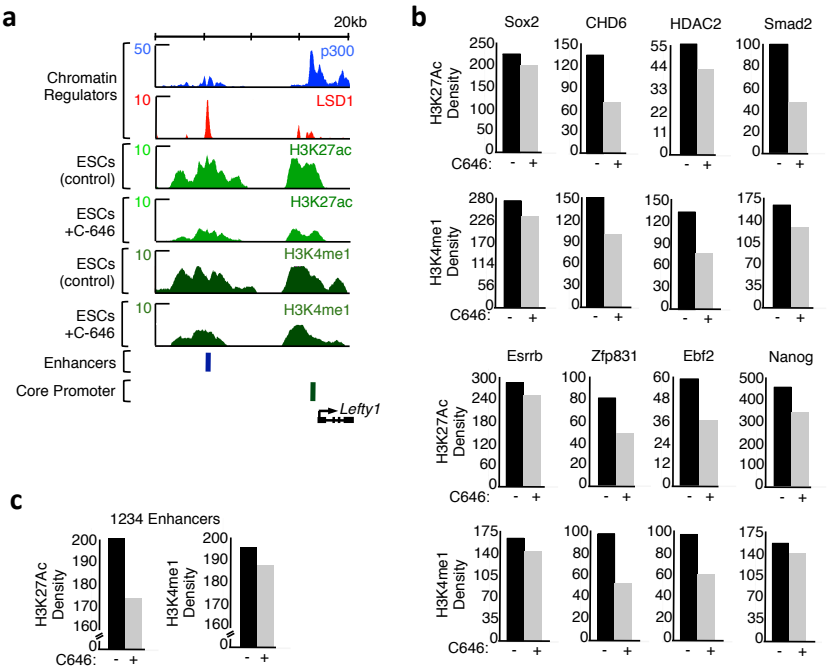
Supplementary Fig. 10: Reduced p300 and H3K27Ac levels at ESC enhancers during differentiation

a, b, Reduced levels of transcriptional coactivator p300 and H3K27Ac at enhancers during ESC differentiation. ChIP–PCR analysis of p300 and H3K27Ac enrichment at enhancer loci in doxycycline-treated (+dox) and untreated (-dox) ZHBTc4 cells. Error bars reflect standard deviation of either duplicate or triplicate PCR reactions.



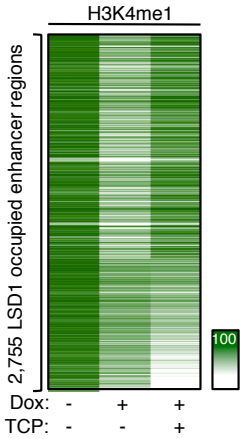
Supplementary Fig. 11: LSD1 occupies ESC enhancers during differentiation

a, b, LSD1 is approximately 0-5% reduced at 12 hours and 30-50% at 48 hours following Oct4 depletion and differentiation of ESCs. ChIP–PCR analysis of LSD1 enrichment at enhancer loci in doxycycline-treated (+dox) and untreated (-dox) ZHBTc4 cells. Error bars reflect standard deviation from biological replicates.



Supplementary Figure 12: p300 inhibition results in reduced H3K27ac and H3K4me1 levels at ESC enhancers

a, Reduced levels of H3K27ac and H3K4me1 at enhancers of ESCs treated with p300 inhibitor C-646. ChIP-Seq binding profiles (reads/million) for chromatin regulators (p300, LSD1) at the *Lefty1* locus in ESCs. Below these profiles, histone H3K27ac and H3K4me1 levels are shown for DMSO treated (control) ESCs, and cells treated with p300 inhibitor C-646 for 24 hours. For appropriate normalization, ChIP-Seq data for histone H3K27ac and H3K4me1 is shown as rank normalized reads/million with the y-axis floor set to 1 (Supplementary Information). Gene models, and previously described enhancer regions[3] are depicted below the binding profiles. **b,** Mean of the normalized H3K27ac and H3K4me1 density +/- 250 nucleotides surrounding LSD1-occupied enhancer regions in the presence or absence of C-646. The associated genes were identified based on their proximity to the LSD1-occupied enhancers (Supplementary Information). **c,** Mean of the normalized H3K27ac and H3K4me1 density +/- 250 nucleotides surrounding 1,234 LSD1-occupied enhancers showing reduced levels of H3K27ac in the presence of C-646. Of the 1,234 LSD1-occupied enhancers having reduced levels of H3K27ac upon C-646 treatment, 63% (773) display reduced H3K4me1 levels after C-646 treatment ($p < 10^{-3}$).



Supplementary Fig. 13: LSD1 is required for H3K4me1 removal at ESC enhancers

Heatmap of normalized H3K4me1 density +/- 250 nucleotides surrounding the 2,755 LSD1 occupied enhancers having reduced levels of H3K4me1 upon differentiation. Of the 2,755 LSD1 occupied enhancers having reduced levels of H3K4me1 upon differentiation, 63% (1,722) display higher H3K4me1 levels after TCP treatment ($p < 10^{-16}$). Color scale indicates ChIP-Seq signal in normalized reads per million.

Appendix B:

Embryonic Stem Cell-Based System For Discovery Of Developmental Transcriptional Programs

Esteban O. Mazzoni[1], Shaun Mahony[2], Jun-An Chen[1], Michelina Iacovino[3], Carolyn A. Morrison[1], George Mountoufaris[1], Yuan-Ping Huang[1], Warren Whyte[4], Rick A. Young[4], Michael Kyba[3], David K. Gifford[2] and Hynek Wichterle[1].

1.Departments of Pathology, Neurology, and Neuroscience, Center for Motor Neuron Biology and Disease, Columbia University Medical Center, 630 168 street, New York, NY 10032, USA. 2.Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, MA 02139, USA. 3. Lillehei Heart Institute and Department of Pediatrics, University of Minnesota, 312 Church St. SE, Minneapolis, MN 55455, USA. 4. Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA.

Personal Contribution to the Project

This chapter is the result of a collaboration with Esteban Mazzoni on work that was published in 2011 in *Nature Methods*. Esteban requested assistance in the tagging of lineage-specific transcription factors and determining their genome-wide localization in ES-cell-derived motor neurons and their progenitor cells. I conducted ChIP-seq experiments, and performed some data analysis, which is described in this appendix.

Study of developmentally regulated transcription factors by chromatin immunoprecipitation and sequencing (ChIP-seq) faces two major obstacles: availability of ChIP grade antibodies and access to sufficient number of cells. We describe a versatile method for genome-wide analysis of transcription factor binding sites by combining directed differentiation of embryonic stem cells and inducible expression of tagged proteins. We demonstrate its utility by mapping transcription factors involved in motor neuron specification.

The study of transcriptional networks provides an opportunity to gain fundamental insight into complex molecular processes that govern cell fate specification and embryonic development. While numerous transcription factors controlling cell differentiation have been functionally characterized, their cell type specific patterns of DNA binding remain largely unknown.

The method of choice for genome-wide mapping of transcription factor binding sites is chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) [1]. Although powerful, current ChIP-seq technology is limited by two critical factors when applied to developmental studies. First, ChIP-seq profiling demands a large number of cells (20-50 million) separated from other cell types expressing the transcription factor of interest, and second, it requires antibodies with high affinity and specificity that recognize transcription factors in their native form bound to DNA. To overcome these two hurdles, we combined a versatile system for generating mouse embryonic stem cell (ESC) lines harboring inducible and epitope-tagged transcription factors with directed differentiation of ESCs along defined cellular lineages.

To overcome the inconsistency and inefficiency of classical transgenic ESC line production, we relied on a recently developed inducible cassette exchange (ICE) system 2. The resulting transgenic lines harbor a single copy of the transgene recombined into a defined expression-competent locus. To further streamline the generation of inducible cell lines, we introduced Gateway (Invitrogen) landing sites into the shuttle vector and a short epitope tag either at the amino- (Flag-Bio) or carboxy-terminus (His-V5) of the protein (Fig. 1a). Because of the high efficiency of all steps, parallel production of multiple inducible tagged lines can be accomplished in as little as three weeks.

Differentiation of mouse ESCs to spinal motor neurons yields scalable and largely homogeneous populations of cells mirroring developmentally relevant motor neuron differentiation states in mouse 3. To test our approach, we first investigated genome-wide binding of the bHLH transcription factor Olig2 in motor neuron progenitors (pMNs) 4, a rare population of cells (<1% of spinal cells on e9.5) found in the embryonic ventral spinal cord 5.We generated an inducible Olig2 ESC line in which Olig2 protein is carboxy-terminal tagged with the V5 epitope (iOlig2-V5). To mimic the normal Olig2 pattern of expression, doxycycline (Dox) was administered late on Day 3 and the expression of the transgene was analyzed on Day 4 (Fig. 1b) when cells reach pMN stage. The transgenic Olig2-V5 protein was expressed uniformly in pMNs, exhibited correct nuclear localization is ~4 fold higher than native Olig2 (Suppl. Fig. 1a-b). The V5 sequence did not perturb the function of the tagged Olig2-V5 protein. As expected, ectopic expression of Olig2-V5 resulted in the repression of Nkx2.2 in ventral interneuron progenitors (Fig. 1c) [4] and in the repression of Pax6 and Irx3 in

dorsal interneuron progenitors (Fig. 1d) 6. Therefore, a tagged version of Olig2 recapitulates in differentiating ESCs the normal function of native Olig2 during spinal cord development [7].

To profile Olig2 binding, we induced Olig2-V5 in pMNs and performed a ChIP-seq experiment with an anti-V5 antibody. We observed that Olig2-V5 binds in the proximity of the downregulated genes Irx3, Nkx2.2 and Pax6 (Fig 2a and Suppl. Fig. 1c), indicating that Olig2 specifies pMN identity by direct repression of interneuron transcriptional programs.

The overexpression of the Olig2 transgene or the addition of the short tag sequence might affect the genomic binding pattern of the Olig2-V5 protein. For comparison, we therefore performed a ChIP-seq experiment in ESC-derived pMNs with antibodies against the native Olig2 protein in the absence of Dox. The endogenous Olig2 and inducible Olig2-V5 ChIP-seq experiments revealed a remarkable level of agreement. The proteins bind to the same regulatory sequences of Irx3, Nkx2.2 and Pax6 (Fig 2a and Suppl. Fig. 1c). As expected for a bHLH transcription factor, motif discovery within the ChIP-enriched sites revealed an E-box motif consensus (Fig 2b) that is present at 58.8% of Olig2-V5 and 60.4% of native Olig2 binding sites (10% false discovery rate (FDR) motif scoring threshold [8]). To determine whether enriched sequences lacking E-box motif represent real binding events we employed an in vitro ELISA based DNA-protein interaction assay. We demonstrate that Olig2 transcription factor can be recruited to all tested ChIP-seq identified sequences regardless whether they contain E-box motif or not (Suppl. Fig.1d), supporting the notion that ChIP-seq data reflect Olig2 binding events.

The distribution of binding sites found in both experiments is also highly coincident (Fig. 2c). Comparing the read counts at enriched peaks shows that only 0.2 % and 1.1% are differentially enriched in the native Olig2 and Olig2-V5 ChIP experiments, respectively (Fig. 2c). Globally, the levels of ChIP-seq enrichment are highly correlated between experiments with a Pearson's correlation coefficient of 0.83, indicating that neither the overexpression of Olig2-V5 in Olig2^{+} pMNs, nor the addition of an epitope tag, affects Olig2 activity or DNA binding preference.

Next we compared the binding site preference of tagged transcription factors in a postmitotic motor neuron stage. We have previously demonstrated that Hoxc9 represses cervical programs and promotes specification of thoracic motor neurons [9]; the study of Hoxc9-V5 (iHoxc9-V5) (Fig. 1e) revealed a direct repression of cervical Hox genes [9]. We compared binding sites of C- and N-terminally epitope tagged Hoxc9, reasoning that overlapping sites are most likely to reflect native Hoxc9 binding events. We modified the inducible system to accommodate a Flag-Bio (FlagB) amino-terminal tag (Fig. 1a) that can be used for ChIP pull-downs either with anti-Flag antibodies or streptavidin-based purification in combination with the biotinylation enzyme BirA [10]. We determined that FlagB tagged Hoxc9 retained its ability to repress cervical Hoxc4 and Hoxa5 genes (Fig. 3e and data not shown). Importantly, the genome-wide binding of the iFlagB-Hoxc9 with anti-Flag antibodies shows a high degree of agreement with the Hoxc9-V5 binding profile. Both Hoxc9 proteins associate with rostral Hox genes regulatory elements, indicating their direct repression (Fig. 2d). At the genomic level, both proteins share an identical sequence preference, depicted by a typical Hox binding primary motif (Fig. 2e). Moreover, 47.1% of the

peaks in the V5-tagged and 39.1% in the Flag-tagged experiments contain the primary motif at a 10% FDR scoring threshold (Suppl. Fig. 2). Although we estimate that the proportion of ChIP-seq reads located in enriched regions is approximately three times higher in the FlagB-Hoxc9 experiment than in the Hoxc9-V5 experiment, the detected peaks are highly coincident across experiments (Fig. 2f). Out of 22,458, only 156 peaks (0.7%) are differentially enriched in V5 ChIP and 799 peaks (3.6%) are differentially enriched in the Flag experiment (Fig. 2f). We conclude that genomic regions shared between C- and N- terminally tagged ChIP-seq experiments are likely to represent native Hoxc9 binding events.

The high degree of overlap between ChIP-seq experiments for one transcription factor contrasts with the binding profiles of two unrelated transcription factors. The comparison of the Olig2-V5 and Hoxc9-V5 ChIP experiments revealed a large fraction of non-overlapping peaks, which is in striking contrast to biological replicates of Hoxc9-V5 ChIP-seq experiments that are virtually indistinguishable (Suppl. Fig 3a-c). Detailed analysis of predicted binding positions by the GPS algorithm [8] in co-bound peaks reveals that Olig2 and Hoxc9 occupy proximal but distinct sites within the peaks (Suppl. Fig 2e and f). Because a typical ChIP-seq peak covers ~200 bp, these experiments might be revealing enhancers that are active in both motor neuron progenitors and postmitotic motor neurons.

The system we present here is robust and allows the generation of multiple inducible cell lines in parallel. Of twenty-four generated lines, only three exhibited problems with inducible protein expression, likely due to the inherent toxicity of introduced transgenes (data not shown). While the system is versatile and can be employed to

study both progenitors and differentiated cells, we observed that the efficiency and homogeneity of transgene induction declines in postmitotic neurons. Inducing the transgene at late progenitor stage results in maintained and homogenous expression in postmitotic neurons, offering a reasonable workaround for this problem. Some transcription factors control their targets in a concentration dependent manner. In those instances, it will be important to first establish the Dox concentration and timing of the treatment that result in desired phenotypes, to ensure that the transcription factor binding studies produce biologically relevant information.

In summary, we provide a set of tools for rapid generation of ESC lines and production of unlimited quantities of isogenic differentiated cells that enable identification of developmentally relevant transcription factor binding sites in a genome-wide manner. The cell lines can also be utilized for other biochemical studies, including the isolation and identification of transcription factor binding partners by coimmunoprecipitation followed by mass spectrometry [11]. We believe that the combination of these powerful techniques will pave the way to a detailed mechanistic understanding of transcriptional networks that govern mammalian development.

References

1. P. J. Park, Nat Rev Genet 10 (10), 669 (2009).

2. M. Iacovino, D. Bosnakovski, H. Fey et al., Stem Cells 29 (10), 1580 (2011).

3. H. Wichterle, I. Lieberam, J. A. Porter et al., Cell 110 (3), 385 (2002).

4. B. G. Novitch, A. I. Chen, and T. M. Jessell, Neuron 31 (5), 773 (2001).

5. Y. S. Mukouyama, B. Deneen, A. Lukaszewicz et al., Proc Natl Acad Sci U S A 103 (5), 1551 (2006).

6. J. A. Chen, Y. P. Huang, E. O. Mazzoni et al., Neuron 69 (4), 721.

7. T. M. Jessell, Nat Rev Genet 1 (1), 20 (2000).

8. Y. Guo, G. Papachristoudis, R. C. Altshuler et al., Bioinformatics 26 (24), 3028 (2010).

9. H. Jung, J. Lacombe, E. O. Mazzoni et al., Neuron 67 (5), 781.

10. J. Kim, A. B. Cantor, S. H. Orkin et al., Nat Protoc 4 (4), 506 (2009).

11. J. Wang, A. B. Cantor, and S. H. Orkin, Curr Protoc Stem Cell Biol Chapter 1, Unit1B 5 (2009).

Competing financial interest

The authors declare no competing financial interests.

Acknowledgments

EOM is the David and Sylvia Lieb Fellow of the Damon Runyon Cancer Research Foundation (DRG-1937-07). The authors would like to thank Richard Sherwood for sharing unpublished observations. Personnel and work were supported by NIH grant P01 NS055923 (DKG, RAY, HW), R01 NS058502 (HW) and Helmsley Stem Cell Starter Grant (HW).

Author contribution

EOM and GM generated the transcription factor inducible lines and EOM performed phenotypic analysis of the derived lines. EOM, WAW and CAM performed ChIP experiments. MI and MK developed the ICE cell lines and vectors. EOM performed expression analysis. EOM and MC performed the WB and protein binding to immobilized DNA. SM analyzed the ChIP-seq data. EOM, RAY, DKG and HW designed the experiments. EOM, SM and HW wrote the manuscript, DKG revised the manuscript.

Methods

Cell culture

ES cells were cultured over a layer of Mitomycin-C treated fibroblast resistant to Neomycin (Fisher) in EmbryoMax D-MEM (Fisher) supplemented with 10% ES-FBS (Invitrogen), L-Glutamine (Gibco), 0.1 mM b-mercaptoethanol and 100 U/ml LIF. Motor neuron differentiation of ES cells was performed as previously described[1]. Briefly, ES cells were trypsinized (Invitrogen) and seeded at $5x10^5$ cells/ml in ANDFK medium (Advanced DMEM/F12:Neurobasal (1:1) Medium, 10% Knockout-SR, Pen/Strep, 2 mM L-Glutamine, and 0.1 mM 2-mercaptoethanol) to initiate formation of embryoid bodies (Day 0). Medium was exchanged on Day 1, Day 2 and Day 5 of differentiation. Patterning of embryoid bodies was induced by supplementing media on Day 2 with 1 μM all-*trans*-Retinoic acid (RA, Sigma) and 0.5 μM agonist of hedgehog signaling (SAG, Calbiochem). For ChIP experiments, the same conditions were used but scaled to seed $1x10^7$ cells on Day 0. Doxycycline (Sigma) was added to the culture medium at 1ug/ml when required.

Generation of inducible lines

The p2Lox-V5 plasmid was generated by replacing GFP with the L1-L2 Gateway cassette from pDEST-40 (Invitrogen) in the p2Lox plasmid. The cassette contains a V5-His double epitope tag in frame downstream of the L2 recombination site. p2Lox-FlagB was generated by replacing GFP in the p2Lox plasmid with the L1-L2 Gateway cassette from pDEST-40 without the V5-His sequence but with the addition of a Flag-Biotin sequence[2] in frame and upstream the L1 recombination site.

Open reading frames of genes are cloned by polymerase chain reaction (PCR). To minimize the introduction of mutations during PCR amplification, Phusion polymerase was used (New England Biolabs). Open reading frames were directionally inserted into pENTR/D-TOPO vector (Invitrogen) following manufacturer instructions. The 5' primer always contains the addition of the CACC sequence to ensure directional integration. For each coding sequence, two alternative 3' primers were used: with and without STOP codon, generating two pENTR plasmids for each gene.

The LR recombination scheme is as follow: 1) When constructing a V5-His C-terminal fusion protein the pENTR plasmid with NO STOP codon is recombined with the p2Lox-V5. 2) Non-tagged proteins are generated by recombining the pENTR plasmid with STOP codon with the p2Lox-V5 plasmid. 3) To generate N-terminal tagged proteins, the pENTR plasmid with STOP codon is recombined with p2Lox-FlagB.

Inducible lines were generated by treating the recipient ESCs for 16 hours with doxycycline to induce Cre followed by electroporation of either p2Lox-V5 and p2Lox-FlagB plasmids harboring the desired construct. After G418 selection, on average three resistant clones were picked, characterized and expanded.

Immunocytochemistry

Embryoid bodies were fixed with 4% paraformaldehyde in PBS, embedded in OCT (Tissue-Tek) and sectioned for staining: 24 hours at 4C for primary antibodies and 4 hours at RT for secondary antibodies. After staining, samples were mounted with Aqua Poly Mount (Polyscience). Images were acquired with a LSM 510 Carl Zeiss confocal microscope. Antibodies used in this study include: Rabbit anti-Olig2

(Millipore); mouse anti-V5 (Invitrogen); mouse anti-Flag M2 (Sigma); rabbit anti-Hoxc4 are gifts from T Jessell. Alexa488-, FITC-, Cy3- and Cy5-conjugated secondary antibodies were obtained from either Invitrogen or Jackson Immunoresearch.

ChIP-seq

Differentiating embryoid bodies were washed with PBS and then dissociated by mild Trypsinization (Invitrogen) followed by mechanical dissociation until single-cell suspension was obtained. Cells were fixed with 1% formaldehyde for 15 minutes at room temperature. Pellets containing ~40 $x10^6$ cells were flash frozen and stored at -80 ºC. Cells were thawed on ice, resuspended in 5ml of Lysis Buffer A and incubated for 10 minutes at 4 ºC in a rotating platform. Samples were spun down for 5 minutes at 1,350*g*, resuspended in 5ml Lysis Buffer B and incubated for 10 minutes at 4 ºC in a rotating platform. Samples were spun down for 5 minutes at 1,350*g*, resuspended in 3ml of Sonication Buffer (SB).

Nuclear extracts were sonicated using a Misonix 3000 model sonicator to sheer cross-linked DNA to an average fragment size of approximately 500bp. Sonicated chromatin was incubated for 16 hours at 4C with Protein-G beads (Invitrogen) conjugated with either rabbit anti-V5 (Abcam, ab15828), mouse anti-Flag M2 (Sigma, ab15828) or rabbit anti-Olig2 (Millipore, AB15328). After incubation and with the aid of a magnetic device, beads were washed once with SB+500nM NaCl, once with LiCl Wash Buffer (LiClB) and 1ml of TE. Then, beads were centrifugated at 950*g* for 3 min and residual TE removed with a pipette. 210 *ul* of Elution Buffer was added to the beads followed by incubation at 65 ºC for 45 minutes with a brief pulse of vortex

every 10 minutes. 200 *ul* of supernatant was removed after a 1 minute centrifugation at 16,000*g*. The crosslink was reversed by 16 hours incubation at 65 ºC.

RNA was digested by the addition of 200 *ul* of TE and RNAseA (Sigma) at a final concentration of 0.2mg/ml and incubated for 2 hours at 37C. Protein was digested by the addition of Proteinase K (0.2 mg/ml final, Invitrogen) supplemented with $CaCl^2$ followed by a 30 minutes incubation at 55 ºC. DNA was extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and then recovered with an ethanol precipitation with glycogens as carrier. The pellets were suspended in 70 *ul* of water. Purified DNA fragments were processed according to the Illumina/Solexa sequencing protocol using a Genome Analyzer II (Illumina, http://www.illumina.com/pages.ilmn?ID=252).

Lysis Buffer A (50 mM Hepes-KOH, pH 7.5; 140 mM NaCl; 1 mM EDTA; 10% Glycerol; 0.5% Igepal; 0.25% Triton X-100). **Lysis Buffer B** (10 mM Tris-HCl, pH 8.0; 200 mM NaCl; 1 mM EDTA, pH 8.0; 0.5 mM EGTA, pH 8.0)**. Sonication Buffer** (50mM Hepes pH 7.5; 40mM NaCl; 1mM EDTA; 1mM EGTA; 1% Trition X-100; 0.1% Na-deoxycholate; 0.1% SDS). **Sonication Buffer High Salt** (50mM Hepes pH 7.5; 500mM NaCl; 1mM EDTA; 1mM EGTA; 1% Trition X-100; 0.1% Na-deoxycholate; 0.1% SDS). **IgG LiCl Wash Buffer** (20mM Tris-HCL pH8.0; 1mM EDTA; 250mM LiCl; 0.5% NP-40; 0.5% Na-deoxycholate). **Elution Buffer** (50 mM Tris-HCl, pH 8.0; 10 mM EDTA, pH 8.0; 1% SDS).

ChIP-seq analysis

Sequence reads were aligned to the mouse genome (version mm9) using Bowtie[3] version 0.12.5 with options "-q --best --strata -m 1 -p 4 --chunkmbs 1024". Only uniquely mapping reads were analyzed further. Binding events were detected using GPS[4]. In GPS, the scaling ratio between IP and control channels was estimated using the median ratio of all 10Kbp windows along the genome. The GPS binding model was initialized to the default and iteratively updated over up to 3 training rounds. In this study, we require that reported peaks contain a ChIP-seq enrichment level that is significantly greater than 1.5 times the control level with p-value <0.01 as tested using the Binomial distribution. Signal-to-noise ratios are estimated by comparing the ChIP-seq read count occurring at any peak found for a given transcription factor in any condition to the count of remaining reads in that experiment.

When comparing enrichment levels between two ChIP-seq experiments, we first scale the read counts assigned to each peak using the median ratio of observed read counts

across all peaks. The read counts of one experiment are always scaled down to match the scale of the other experiment. We define differentially enriched sites as those that have a scaled read count in one experiment that is significantly greater than 1.5 times the scaled read count from the other experiment (p<0.01, Binomial test, adjusted for multiple testing using Benjamini & Hochberg's method). All microarray and ChIP-seq data are available from the

GEO database under accession number GSE31456

DNA motif analysis

De-novo motif-finding was performed in 200bp windows centered on the 2,000 top-ranked peaks for each examined ChIP-seq experiment. GimmeMotifs[5] was used to discover motifs by running and combining results from the motif-finders MDmodule, MEME, GADEM, MotifSampler, trawler, Improbizer, MoAn, and BioProspector. The settings "-w 200 -a large -g mm9 -f 0.5 -l 500" were used with GimmeMotifs. STAMP[6] was used to determine the similarity of discovered motifs to known DNA-binding preferences. Log-likelihood scoring thresholds for the discovered motifs were calculated by simulating 1,000,000 200bp sequences using a 3rd-order Markov model of the mouse genome. The motif scoring thresholds that yield false discovery rates of 10% in this set of sequences were recorded and used to scan 200bp sequences centered on the Olig2 and Hoxc9 GPS-predicted peak positions.

ELISA DNA Binding

PCR amplified and biotin-labeled genomic fragments are gel purified. The fragments are between 500 and 600 bp with the 5' primers containing a single biotin molecule at the 5'end. Streptavidin coated 96 well plates (Fisher, PI-15500) are washed three times

with 200 µl of Wash Buffer. The biotin-labeled PCR reaction is loaded into each well in Blocking Buffer up to 125 µl final volume to saturate the binding capacity of each well (Plates can bind ~125pmol/well). The plates were incubated overnight at 4ºC then washed each well three times with 200 µl of Wash Buffer. Serial dilutions of the cell extract in PBS were added to each well and incubated for 3 hrs with shaking at room temperature. After washing each well three times with 200 µl with Wash Buffer the primary antibody in 0.5% BSA in PBS was added to final volume of 100 µl to each well and incubate plate for 2 hours with shaking at room temperature. After washing each well three times with 200µl of Wash Buffer, HRP-conjugated secondary antibody diluted in 0.5% BSA in PBS to final volume of 100 µl to was added to each well for 30-45 minutes with shaking at room temperature. The wells were washed three times with 200 µl of Wash Buffer. 100 μl of equilibrate the TMB (Fisher EN-N301) to room temperature was added to each well. After 10 minutes incubation, the reaction was stopped by adding 100 µl of 2 M sulfuric acid to each well. The absorbance was measured of each well at 450 nm using a plate reader.

Wash Buffer: 0.05% Tween-20 + .1% BSA in PBS.

Blocking Buffer: 0.5% BSA in PBS.

Primers to amplify the ChIp-seq identified genomic regions in Suppl. Table 1.

REFERENCES

1. H. Wichterle, I. Lieberam, J. A. Porter et al., Cell 110 (3), 385 (2002).

2. J. Wang, A. B. Cantor, and S. H. Orkin, Curr Protoc Stem Cell Biol Chapter 1, Unit1B 5 (2009).

3. B. Langmead, C. Trapnell, M. Pop et al., Genome Biol 10 (3), R25 (2009).

4. Y. Guo, G. Papachristoudis, R. C. Altshuler et al., Bioinformatics 26 (24), 3028 (2010).

5. S. J. van Heeringen and G. J. Veenstra, Bioinformatics 27 (2), 270.

6. S. Mahony and P. V. Benos, Nucleic Acids Res 35 (Web Server issue), W253 (2007).

Supplementary File	Title
Supplementary Figure 1	Characterization of Olig2-V5 expression and *in vitro* binding.
Supplementary Figure 2	Primary motif distribution of ChIp-seq binding data.
Supplementary Figure 3	Comparison of Olig2-V5 vs. Hoxc9-V5 and Hoxc9-V5 vs. Flag-Hoxc9 ChIP-seq experiemtns.
Supplementary Table 1	Primer sequence for *in vitro* binding experiemtns.

Figure legend

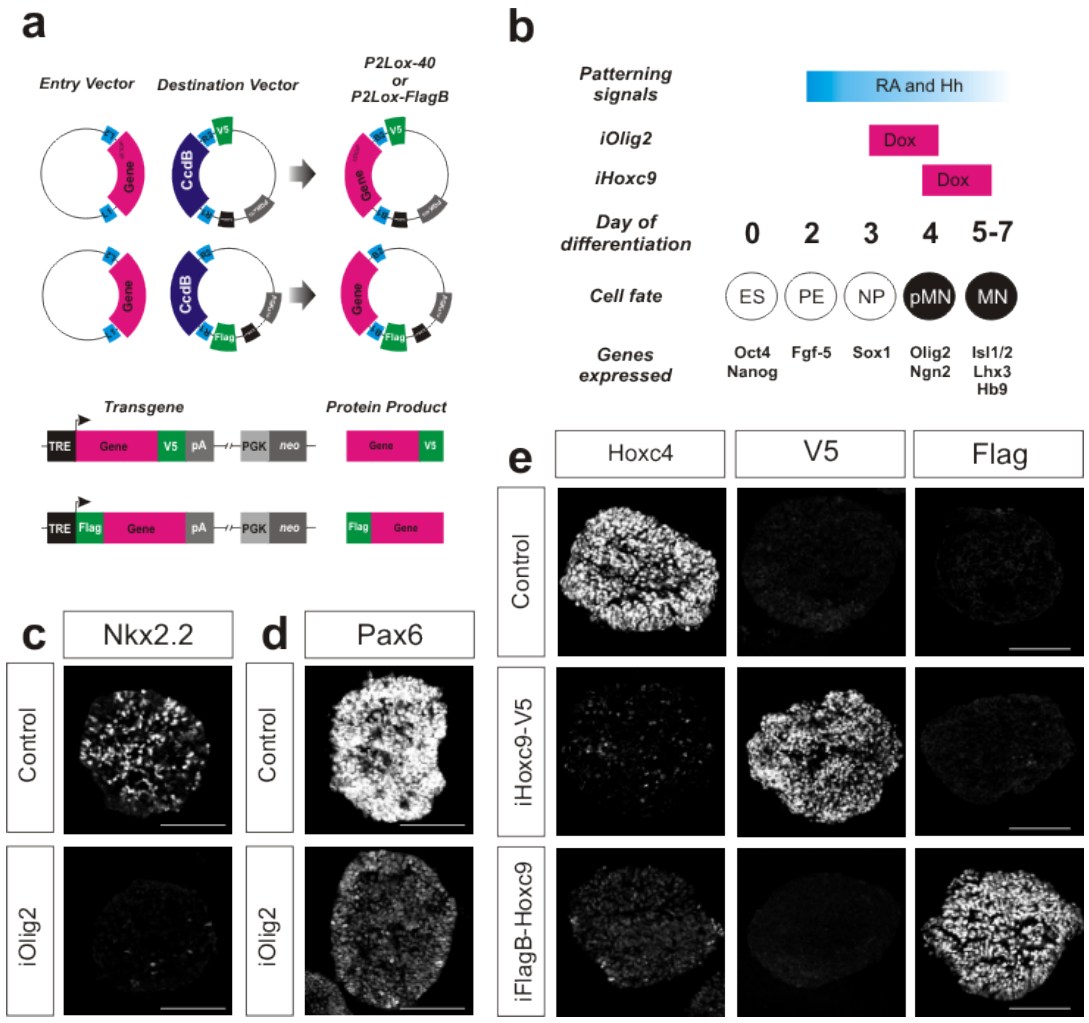
Figure 1: Generation of inducible cell lines

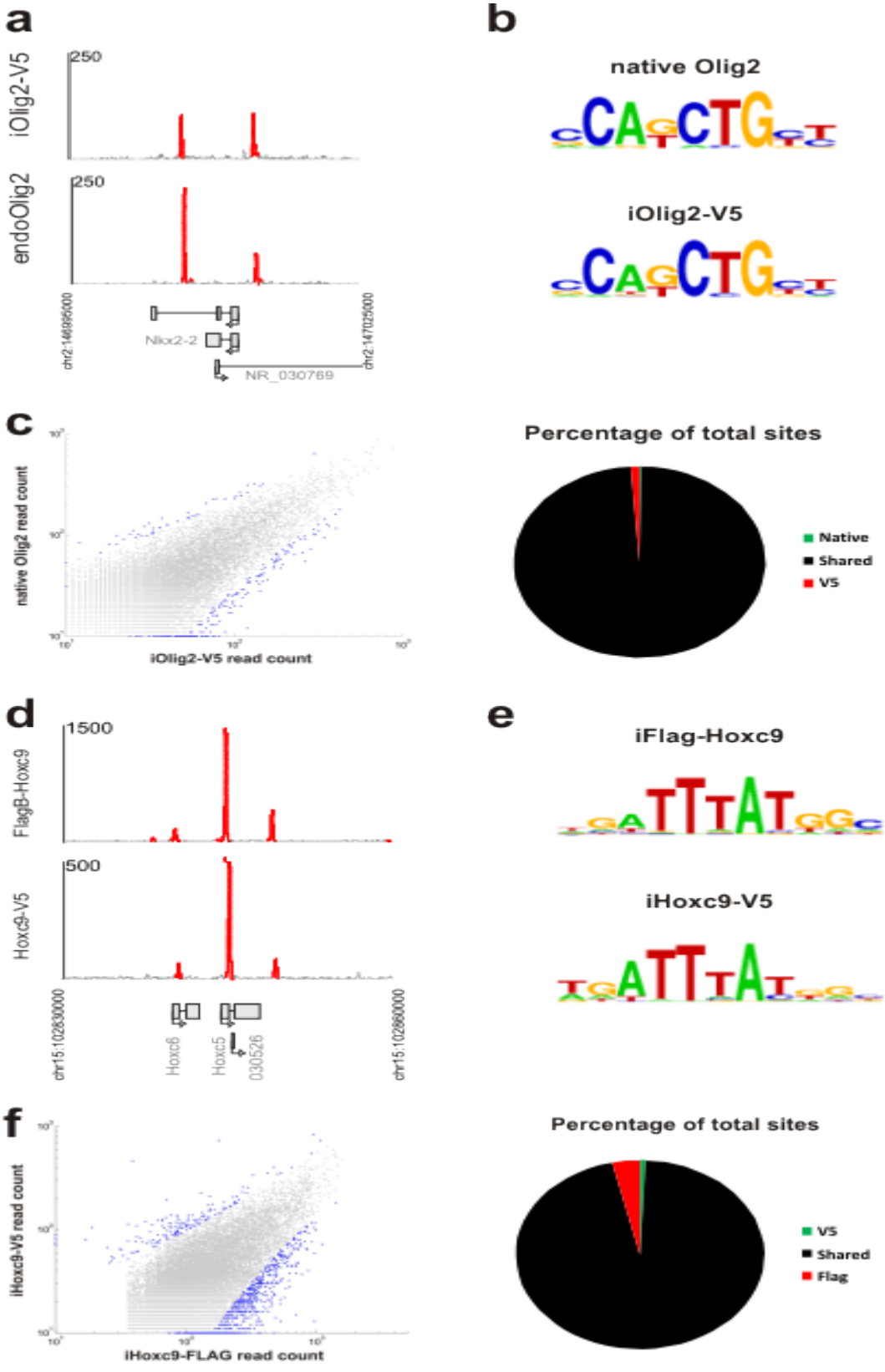
(**a**) The schematic depicts the cloning strategy to generate inducible lines with epitope tagged genes. Coding sequences lacking the STOP codon are V5-tagged at the C-terminus. (**b**) Overview of embryonic stem cell (ESC) directed differentiation. Differentiating cells become motor neuron progenitors (pMNs) at Day 4 and produce motor neurons (MNs) from Day 5-7. Doxycyline (Dox) is added late on Day 3 or Day 4 to mimic the expression pattern of the endogenous Olig2 and Hoxc9 respectively. Olig2-V5 is analyzed at Day 4 and Hoxc9-V5 or FlagB-Hoxc9 at Day 5. (**c**) Nkx2.2 staining in control or Olig2-V5 expressing cells at Day 4 of differentiation under high Hh concentration (500 nM). (**d**) Pax6 staining in control or Olig2-V5 expressing cells at Day 4 of differentiation under low Hh (5 nM). (**e**) iHoxc9-V5 and iFlagB-Hoxc9 day 5 embryoid bodies stained with anti-Hoxc4, V5 and Flag antibodies.

Figure 2: Native and tagged ChIP comparisons

(**a**) ChIP signal tracks over Nkx2.2 genomic loci for endogenous and V5-tagged Olig2. Red peaks represent significant ($p<0.01$) enrichment over control. (**b**) The most over-represented motifs discovered under ChIP-seq peaks for native Olig2 and Olig2-V5 ChIP experiments. (**c**) The plots show a comparison of read enrichment from native and V5-tagged Olig2 ChIP-seq experiments at all detected peaks. Blue dots in the scatterplot represents peaks significantly differentially

enriched in one experiment over the other. Pie chart represents numbers of sites differentially enriched between native Olig2 and V5 tagged ChIP experiments as a percentage of total sites. (**d**) ChIP signal tracks over Hoxc5 genomic locus for V5- and Flag- tagged Hoxc9 experiments. Red peaks represent significant enrichment over control ($p<0.01$). (**e**) The most over-represented motifs discovered under ChIP-seq peaks for Hoxc9-V5 and FlagB-Hoxc9 ChIP. (**f**) The plots show a comparison of read enrichment from FlagB and V5-tagged Hoxc9 ChIP-seq experiments at all detected peaks. Blue points in the scatterplot represent peaks significantly differentially enriched in one experiment over the other. The pie chart shows numbers of sites differentially enriched between Hoxc9-V5 and FlagB-Hoxc9 ChIP experiments as a percentage of total sites.


a
Entry Vector
Destination Vector
P2Lox-40
or
P2Lox-FlagB
Transgene
Protein Product
TRE
Gene
V5
pA
PGK
neo
Flag
b
Patterning signals
RA and Hh
iOlig2
Dox
iHoxc9
Dox
Day of differentiation
0
2
3
4
5-7
Cell fate
ES
PE
NP
pMN
MN
Genes expressed
Oct4
Nanog
Fgf-5
Sox1
Olig2
Ngn2
Isl1/2
Lhx3
Hb9
c
Nkx2.2
Control
iOlig2
d
Pax6
Control
iOlig2
e
Hoxc4
V5
Flag
Control
iHoxc9-V5
iFlagB-Hoxc9


a
iOlig2-V5
250
endoOlig2
250
chr2:146995000
chr2:147025000
Nkx2-2
NR_030769
b
native Olig2
iOlig2-V5
c
native Olig2 read count
iOlig2-V5 read count
Percentage of total sites
Native
Shared
V5
d
FlagB-Hoxc9
1500
Hoxc9-V5
500
chr15:102830000
chr15:102860000
Hoxc6
Hoxc5
030526
e
iFlag-Hoxc9
iHoxc9-V5
f
iHoxc9-V5 read count
iHoxc9-FLAG read count
Percentage of total sites
V5
Shared
Flag

Appendix C:

Ronin/Hcf-1 Binds To A Hyperconserved Enhancer Element And Regulates Genes Involved In The Growth Of Embryonic Stem Cells

Marion Dejosez,[1,2] Stuart S. Levine,[3] Garrett M. Frampton,[3] Warren A. Whyte,[3] Sabrina A. Stratton,[4] Michelle C. Barton,[4] Preethi H. Gunaratne,[5] Richard A. Young,[3] Thomas P. Zwaka[1,2,†]

[1]Center for Cell and Gene Therapy, [2]Departments of Molecular and Cellular Biology and Human Genetics, Baylor College of Medicine, Houston, Texas, USA. [3]Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [4]Program in Genes and Development, Center for Stem Cell and Development Biology, Department of Biochemistry and Molecular Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [5]Department of Biology and Biochemistry, University of Houston, Houston, Texas, USA

Personal Contribution to the Project

This chapter is the result of a collaboration with Marion Dejosez on work that was published in 2010 in *Genes and Development*. Marion requested assistance in determining the genome-wide localization of the ES cell transcription factor Ronin in ES cells. I conducted ChIP-seq experiments, and performed some data analysis, which is described in this appendix.

Summary

Embryonic stem (ES) cells have an exceptional need for timely biomass production, yet the transcriptional control mechanisms responsible for meeting this requirement are largely unknown. We report here that the transcription factor Ronin (Thap11), which is essential for the self-renewal of ES cells, controls a key set of genes necessary for accommodating the biosynthetic and bioenergetic demands of self-renewing mouse ES cells. Genome-wide analysis revealed that Ronin occupies an exceptionally conserved regulatory motif in mammals, a DNA sequence that previously lacked a recognized binding factor. Ronin, together with the transcriptional coregulator Hcf-1, was generally observed to be an activator of genes essential to protein biosynthesis and energy production. We propose that Ronin activity is essential to the unimpeded growth of ES cells.

In contrast to all other mammalian cells, embryonic stem (ES) cells are characterized by a truncated cell cycle, relative autonomy from extracellular signal-regulated kinase (Erk) signaling and an unusually rapid growth rate, analogous to that of cancer cells and primitive unicellular organisms including bacteria (*1-3*). Thus, the task of replicating the genome, proteome and other cellular components (biomass) to keep pace with this rapid growth imposes stringent metabolic demands on ES cells. Emerging evidence indicates that such requirements are met not by a self-correcting, homeostatic system of housekeeping enzymes, but by a dynamically regulated genetic network (*4*). Although much has been learned about the factors governing the

pluripotency of ES cells (*5-10*), the molecular mechanisms that facilitate their efficient self-renewal are only beginning to be explored. A better understanding of the link between ES cell metabolism and pluripotency is fundamental to realizing the full potential of these cells in genetic engineering and regenerative medicine.

We recently discovered a novel zinc-finger transcriptional regulator, Ronin (also Thap11), that is essential for the self-renewal of ES cells (*11*). Conditional knockout of the *Ronin* gene induces ES cell death, while its forced expression enables the cells to proliferate transiently without differentiation under conditions that normally do not promote self-renewal. While studying *Ronin*-overexpressing mouse ES cells, we noticed that they possessed a strikingly enlarged nucleolus (Fig. 1A), the prime site of ribosome synthesis and assembly, suggesting that *Ronin* overexpression may alter the production of a key growth-related body in ES cells. This observation, coupled with our previous finding that Ronin can interact with the cell growth factor Hcf-1 (*11-13*), led us to investigate its contribution to the biosynthetic needs of mouse ES cells.

We first sought to identify the DNA-binding sites of Ronin using chromatin immunoprecipitation and DNA sequencing (ChIP-seq). Our mapping results revealed 866 Ronin-bound regions (Table S1), most of which were located at or immediately upstream of transcription start sites (Fig. 1B and C), suggesting that Ronin participates in transcriptional initiation. Alignment analysis identified a Ronin-binding motif (CTGGGARWTGTAGTY, designated here as RBM) in 844 of the target gene promoters (Fig 1D). This functional element was highly enriched compared with

random sequences ($P << 10^{-100}$). Intriguingly, the Ronin target sequence closely matched an orphan promoter sequence (ACTAYRNNNCCCR, so-called M4 motif; Fig. 1D), whose conservation rate in humans (61%) is the fourth highest among the top 50 conserved motifs described by Xie et al. (*14*). This sequence is notable for another reason: in contrast to most highly conserved regulatory motifs in the human genome, it lacked a recognized binding factor until this analysis. The RBM was present within 70% of all regions enriched for Ronin binding. When we examined the 22 sites that were bound by Ronin but lacked the full length RBM ($P>0.05$), we found that a substantial proportion of these targets contained at least half of the conserved motif (fig. S2A), suggesting that similar to other zinc-finger proteins, such as Rest (*15*), Ronin can occupy half-sites. Gel-shift experiments (Fig. S1A) and ChIP with a Ronin-specific antibody, followed by PCR analysis of selected target gene regions (fig. S1B and C), confirmed the identity of each site we had found to be specifically targeted by Ronin.

Because Ronin lacks a transactivation domain and can interact directly with a known transcriptional regulator, the Hcf-1 protein (*11*), we considered that both factors might be needed at the RBM to initiate gene transcription. We therefore performed ChIP-seq with an Hcf-1 antibody (*16*), identifying 743 genomic loci occupied by Hcf-1 at a high confidence level (see Table S2). These regions overlapped 176 target promoters that were also bound by Ronin (Table S2 and Fig. 1E). Even when the Hcf-1 signal did not attain significance by our conservative cut-off point, we were still able to detect a substantial binding peak, indicating that Hcf-1 generally co-occupies Ronin-bound

target sites. To study functional significance of the interaction between Ronin and Hcf-1 at a common regulatory motif we created ES cells overexpressing a Ronin mutant deficient in Hcf-1 binding (*EF1α-Ronin*^{DHSA}) (figs. S3 and S4A, B). As previously reported (*11*) Ronin made Lif nonessential for ES cells under our experimental conditions, while *EF1α-Ronin*^{DHSA} ES cells still differentiated (fig. S4C, D), suggesting that Ronin must interact with Hcf-1 to block differentiation.

Approximately 40% of the 866 promoters bound by Ronin were also occupied by one or more of the ES cell transcription factors Oct4, Sox2 and Nanog, which have central roles in pluripotency control (*5-10, 17*). It was interesting that Ronin consistently occupied sites within promoter regions that were only 50 to 100 bp upstream of transcription start sites, in contrast to the more distant sites occupied by Oct4 (Fig. 1F) and other core transcription factors. Most promoters bound by Ronin were not occupied by Oct4, Sox2, Nanog, and Tcf3. Indeed, Ronin and Hcf-1 clustered together rather than with the canonical factors in a hierarchical clustering analysis based on target similarity scores for genomic regions that were highly enriched in ChIP-seq experiments (*7, 17-19*) (Fig. 1G).

We next considered the possibility that the genes occupied by Ronin and Hcf-1, but not the canonical factors, are involved in biomass production supporting ES cell growth. To test this hypothesis, we focused on the subset of genes whose promoters were solely bound by Ronin and Hcf-1. Using the Panther tool for classifying biological processes, we determined the functional categories of all genes defined by

independent binding of Ronin/Hcf-1 to promoters. Transcription initiation, mRNA splicing and protein metabolism were among the most overrepresented categories, whereas cell signaling and cell development were underrepresented (Fig. 2A). Close inspection of the individual genes within these categories (Table S3) yielded a more informative functional portrait (Fig. 2B). Ronin bound as many as 30% of the ribosomal protein-encoding genes, including *Rplp2, Rps12* and *Rpl36*, and two key subunits of RNA polymerase I, *Rpo1-2* and *Rpo1-4* (all involved in protein biosynthesis); *Ctd*, *Cnot4/8* and *Med4* (transcription initiation); and *Rab1b, Nup133 and Timm22* (protein trafficking). These results are important because alterations in ribosomal biosynthesis, transcriptional initiation and protein transport can have profound effects on the metabolome of any cell, including stem cells (*20-22*). Ronin also bound specifically to a subset of genes encoding mitochondrial ribosomal proteins (*Mrpl19*, *32, 50* and *54*), mitochondrial translation factors (*Tufm*) and rate-limiting members of the oxidative phosphorylation cascade (*Atp5e, NADH dehydrogenase* and *Atp5e*), suggesting its involvement in the control of energy production in ES cells. Finally, Ronin occupied genes encoding threonine catabolic enzymes (*Mtr*), whose increased expression in ES cells facilitates a high-flux metabolic state characterized by enhanced threonine catabolism (*3*). Ronin did not bind to any of the classical cell cycle genes, suggesting that it influences self-renewal without affecting cell cycle progression. Thus, Ronin occupies a relatively small number of genes with specific functions in protein biosynthesis and energy production, but not cell development or mitosis (Fig. 2B).

To explore the nature of target gene regulation by Ronin/Hcf-1, we conducted gene set enrichment analysis (GSEA) of RNA from Ronin-targeted genes in ES cells compared with that from embryoid bodies (Fig. 3A, top). The result indicates that Ronin-occupied genes are generally highly transcribed, and the transcripts are significantly overrepresented in ES cells. We also compared the Ronin transcriptional program in wild-type ES cells with that in cells ectopically and stably overexpressing Ronin (*EF1α-Ronin* ES cells). Again, GSEA showed an overrepresentation of Ronin target genes (Fig. 3A, middle), suggesting that Ronin indeed positively affects the expression of many but not all of its target genes, consistent with the ability of Hcf-1 to either positively or negatively regulate transcription depending on the cellular context (*13*). To test the reverse prediction, we transfected *Ronin*$^{loxP/-}$ ES cells with the gene encoding Cre recombinase, sorted Cre-positive cells at 18 hours post-transfection, extracted the RNA and performed microarray analysis of gene expression. Interestingly, the entire subset of 133 genes found to be upregulated in *Ronin* knockout cells (Fig. 3B, right) were downregulated in our *Ronin*-overexpressing clones, whereas in the converse situation only 43 of 99 genes found to be downregulated after knockout (Fig. 3B, left) were upregulated in the *Ronin*-overexpressing clones. Additional evidence for direct transcriptional control of Ronin targets was obtained in experiments in which we cloned a set of the Ronin-targeted promoters and performed luciferase reporter assays. As shown in Fig. 3C, Ronin gain-of-function had a positive effect on gene expression, while loss of the RBM modulated the regulatory effects of Ronin on its targets. To track the expression of Ronin-controlled genes more closely, we analyzed the results of DNA microarrays over 14 days of ES cell differentiation

(Fig. 3D). As expected, the largest class of genes showed rapid downregulation after induction of differentiation, while the two remaining classes were either upregulated or demonstrated complex regulation. Taken together, our findings show that Ronin binding to target sites can lead to either gene repression or activation, with the latter being more common. These findings revise our earlier suggestion that Ronin is primarily a global repressor, based on the assumption that acute upregulation of Ronin under otherwise steady-state conditions would exert a dominant-negative effect on Ronin function, similar to observations on other proteins that harbor a Thap domain (*23*). Thus, Ronin is most frequently an activator of genes involved in protein biosynthesis and energy production.

To further validate the apparent role of Ronin in meeting the metabolic requirements of rapidly dividing ES cells, we focused our attention on cell size, reasoning that this property might be highly sensitive to fluctuations in Ronin activity. In experiments to determine the diameters of *EF1α-Ronin* and *EF1α-Ronin*^{DHSA} ES cells compared with wild-type controls, we found that *EF1α-Ronin* ES cells were significantly enlarged (> 6.4% increase in size, $P=5.63x10^{-4}$; Fig. 4A, top), a change that was clearly visible in cell pellets adjusted for cell number (Fig. 4A, bottom). Importantly, ES cells expressing *EF1α-Ronin*^{DHSA} did not show an increase in cell size, confirming the requirement for Hcf-1 in optimal Ronin function. We also measured the amount of protein per cell (Fig. 4B, top) as well as protein translation capacity per cell (Fig. 4B, bottom), demonstrating that *Ronin*-overexpressing cells have significantly more protein content and translational activity per cell than do Ronin^{DSHA} or wild-type cells

($P=1.64x10^{-2}$). Finally, we found that the amount of ATP per cell was significantly higher ($P=3.92x10^{-3}$) in *EF1α-Ronin* ES cells compared with wild-type or *EF1α-Ronin*^{DHSA} ES cells (Fig. 4C). This increase in cell size and protein metabolism was not due to a difference in cell cycle kinetics, as the proliferative rates of these cells did not differ over 96 hours of culture (Fig. 4D) and similar numbers of control and EF1α-Ronin ES cells were in S phase (fig. S5A and B).

Recently, the life cycle of ES cells was compared to that of yeast cells and other unicellular metazoans, in the sense that it follows a relatively primitive set of behavioral rules that differ from those of more mature cells (*2, 3, 24*). In yeast, changes in certain exogenous factors, such as temperature and the sugar composition of the environment, profoundly affect the ability of the cells to produce protein and therefore to grow and to metabolize different energy sources (*20*). Similarly, we suggest that fluctuations in the growth and metabolic capacity of ES cells may represent a previously unrecognized mechanism of pluripotency control. Thus, for example, a shortfall in protein biosynthesis could cause ES cells to lose their full self-renewal capacity, leading to apoptotic death or perhaps a rapid transition to differentiation. This model is supported by recent observations establishing a correlation between the translational capacity of ES cells and the likelihood of differentiation (*25*). Our results indicate that Ronin participates in pluripotency control by directly stimulating a genetic program that ensures sufficiently robust protein biosynthesis and energy production to sustain the rapid self-renewal of ES cells until they exit the undifferentiated state. A deeper understanding of the links between

Ronin- and canonical transcription factor-controlled regulatory systems may facilitate (i) the development of new methods to maintain ES cells in vitro, a key step toward using stem cell therapies to replace diseased or injured tissue, and perhaps (ii) the identification of prime targets for new cancer therapeutics in malignant cells constitutively expressing the *Ronin* gene.

References and Notes

1. K. W. Orford, D. T. Scadden, *Nat Rev Genet* **9**, 115 (Feb, 2008).
2. Q. L. Ying *et al.*, *Nature* **453**, 519 (May 22, 2008).
3. J. Wang *et al.*, *Science* **325**, 435 (Jul 24, 2009).
4. M. G. Vander Heiden, L. C. Cantley, C. B. Thompson, *Science* **324**, 1029 (May 22, 2009).
5. B. E. Bernstein *et al.*, *Cell* **125**, 315 (Apr 21, 2006).
6. L. A. Boyer *et al.*, *Cell* **122**, 947 (Sep 23, 2005).
7. X. Chen *et al.*, *Cell* **133**, 1106 (Jun 13, 2008).
8. M. F. Cole, S. E. Johnstone, J. J. Newman, M. H. Kagey, R. A. Young, *Genes Dev* **22**, 746 (Mar 15, 2008).
9. J. Kim, J. Chu, X. Shen, J. Wang, S. H. Orkin, *Cell* **132**, 1049 (Mar 21, 2008).
10. Y. H. Loh *et al.*, *Nat Genet* **38**, 431 (Apr, 2006).
11. M. Dejosez *et al.*, *Cell* **133**, 1162 (Jun 27, 2008).
12. E. Julien, W. Herr, *EMBO J* **22**, 2360 (May 15, 2003).
13. J. Wysocka, W. Herr, *Trends Biochem Sci* **28**, 294 (Jun, 2003).
14. X. Xie *et al.*, *Nature* **434**, 338 (Mar 17, 2005).
15. A. W. Bruce *et al.*, *Genome Res* **19**, 994 (Jun, 2009).
16. A. C. Wilson, K. LaMarco, M. G. Peterson, W. Herr, *Cell* **74**, 115 (Jul 16, 1993).
17. A. Marson *et al.*, *Cell* **134**, 521 (Aug 8, 2008).
18. A. C. Seila *et al.*, *Science* **322**, 1849 (Dec 19, 2008).
19. M. Ku *et al.*, *PLoS Genet* **4**, e1000242 (Oct, 2008).

20. J. R. Warner, *Trends Biochem Sci* **24**, 437 (Nov, 1999).

21. R. Y. Tsai, R. D. McKay, *Genes Dev* **16**, 2991 (Dec 1, 2002).

22. T. Moss, V. Y. Stefanovsky, *Cell* **109**, 545 (May 31, 2002).

23. C. Cayrol *et al*., *Blood* **109**, 584 (Jan 15, 2007).

24. J. Silva, A. Smith, *Cell* **132**, 532 (Feb 22, 2008).

25. P. Sampath *et al*., *Cell Stem Cell* **2**, 448 (May 8, 2008).

26. We thank Drs. J.A. Thomson and M.K. Brenner for critical reading of the manuscript and Laura Jo Zitur for help with some initial experiments. We also thank Dr. Winship Herr for providing us with the Hcf-1 antibody. This work was supported by the Diana Helis Henry Medical Research Foundation (T.P.Z.), the Huffington Foundation (T.P.Z.) and the National Institutes of Health (grants R01 EB005173-01, 1R01 GM077442-01, P20 EB007076 and P01 GM81627).

Figure Legends

Fig. 1. Ronin binds to a highly conserved promoter element in mouse ES cells that is shared with Hcf-1. (A) Ronin-overexpressing clones have a different morphology characterized by a rounder cell shape and a more prominent nucleolus. (B) Binding of Ronin at the promoter regions of four representative genes, *Rpo1-2* (RNA polymerase I), *Mrlp34* (mitochondrial ribosomal protein L34), *Cbx1* (chromobox homolog 1 [*Drosophila* HP1 beta]) and *Rplp2* (ribosomal protein LP2). Gene structure and chromosomal coordinates (mouse build NCBI36) are shown below the graphs; arrows indicate transcription start site. ChIP-seq results, shown on the y-axis are reads per million reads. (C) Histogram showing the distance of the midpoint of each Ronin binding event from the nearest transcription start site (arrow). There is a highly significant preference of Ronin binding for the region immediately upstream of the start site. (D) Identification of the consensus Ronin binding motif depicted as a bit matrix. The recently discovered M4 sequence in human promoters (*14*) is included for comparison (top). (E) ChIP-seq results obtained in R1 mouse ES cells after immunoprecipitation with Ronin and Hcf-1 antibodies. A representative region on mouse chromosome 8D3, containing three Ronin-bound genes (*Cenpt* [*G630055Po3Rik*], *Ronin* [*Thap11*] and *Nutf2*), shows substantial overlap between Ronin and Hcf-1 binding peaks in the promoter region of all three genes. (F) ChIP-seq results obtained by precipitation with antibodies against Ronin or Oct4. Both factors occupy the Max promoter, with the Ronin binding peak positioned closer to the transcription start site (arrow). (G) Hierarchical clustering analysis of 22 prominent

transcriptional regulators in mouse ES cells, based on target similarity scores calculated with a Pearson correlation similarity metric.

Fig. 2. Cellular functions of genes targeted by Ronin in mouse ES cells. (A) Panther analysis of Ronin target genes. Categories with enrichment values >1 are significantly overrepresented, while those with lower values are significantly underrepresented. (B) Examples of Ronin target genes representing six major functional classes.

Fig. 3. Ronin can either activate or repress its transcriptional targets in mouse ES cells. (A) GSEA analyses of Ronin target genes, showing enrichment of Ronin target genes in wild-type mouse ES cells compared with differentiated cells (top), in Ronin-overexpressing cell line compared with wild-type ES cells (middle), and in wild-type compared with Ronin-knockout ES cells. (B) Overlap between the numbers of Ronin-bound genes that are up- or downregulated in Ronin-overexpressing ES cells in relation to their status in *Ronin*-knockout ES cells. (C) Luciferase reporter assays with pGL3-based promoter constructs, showing effects on the expression of selected target genes (*Ywhag*, *Mdh1* and *Rplp2*) with or without the Ronin binding motif in Ronin-overexpressing cells. Values are means ±SD of triplicate experiments, asterisks indicate a significant difference in relative luciferase activity in Ronin-overexpressing versus control cells (* $P<0.05$, ** $P<0.01$). ΔRBM, mutant reporter construct with deleted Ronin binding motif. (D) Expression of Ronin target genes during the differentiation of mouse ES cells. Most target genes (Class I) are highly expressed in undifferentiated cells but are rapidly downregulated after induction of differentiation,

while a smaller set of genes (Class III) are upregulated. Class II genes appear to be regulated in a more complex fashion. Red, upregulated genes; blue, downregulated genes.

Fig. 4. Cell growth and metabolism correletate with Ronin expression in mouse ES cells. (A, top) Mean diameter of Ronin- versus Ronin^{DHSA}-overexpressing and control mouse ES cells as determined with an automated cell counter. Values are means ±SD of triplicate experiments; asterisk indicates $P=5.63x10^{-4}$. (A, bottom) Pellet size of $16x10^{6}$ control and Ronin- or Ronin^{DHSA}-overexpressing mouse ES cells after resuspension in PBS and centrifugation. (B, top) Protein amount per cell in lines shown in figs S3 and S4. Asterisk indicates $P=1.6410x10^{-2}$; values are means ±SD of triplicate experiments. (B, bottom) Analysis of protein synthesis rate in control cells (blue solid line) versus Ronin-overexpressing (red solid line) and Ronin^{DHSA}-overexpressing cells (green solid line) stained with AHA-Alexa Fluor488; additional controls were treated with cycloheximide (black dotted line) to block protein synthesis osr were not labeled with AHA (green dotted line). (C) ATP content of Ronin- versus Ronin^{DHSA}-overexpressing and control ES cells, quantified as relative luminescence units (RLUx106/$5x10^{4}$ cells). Asterisk indicates $P=3.92x10^{-3}$, values are means ±SD of triplicate experiments. (D) Proliferation rate of Ronin- versus Ronin^{DHSA}-overexpressing and control ES cells over 96 h.

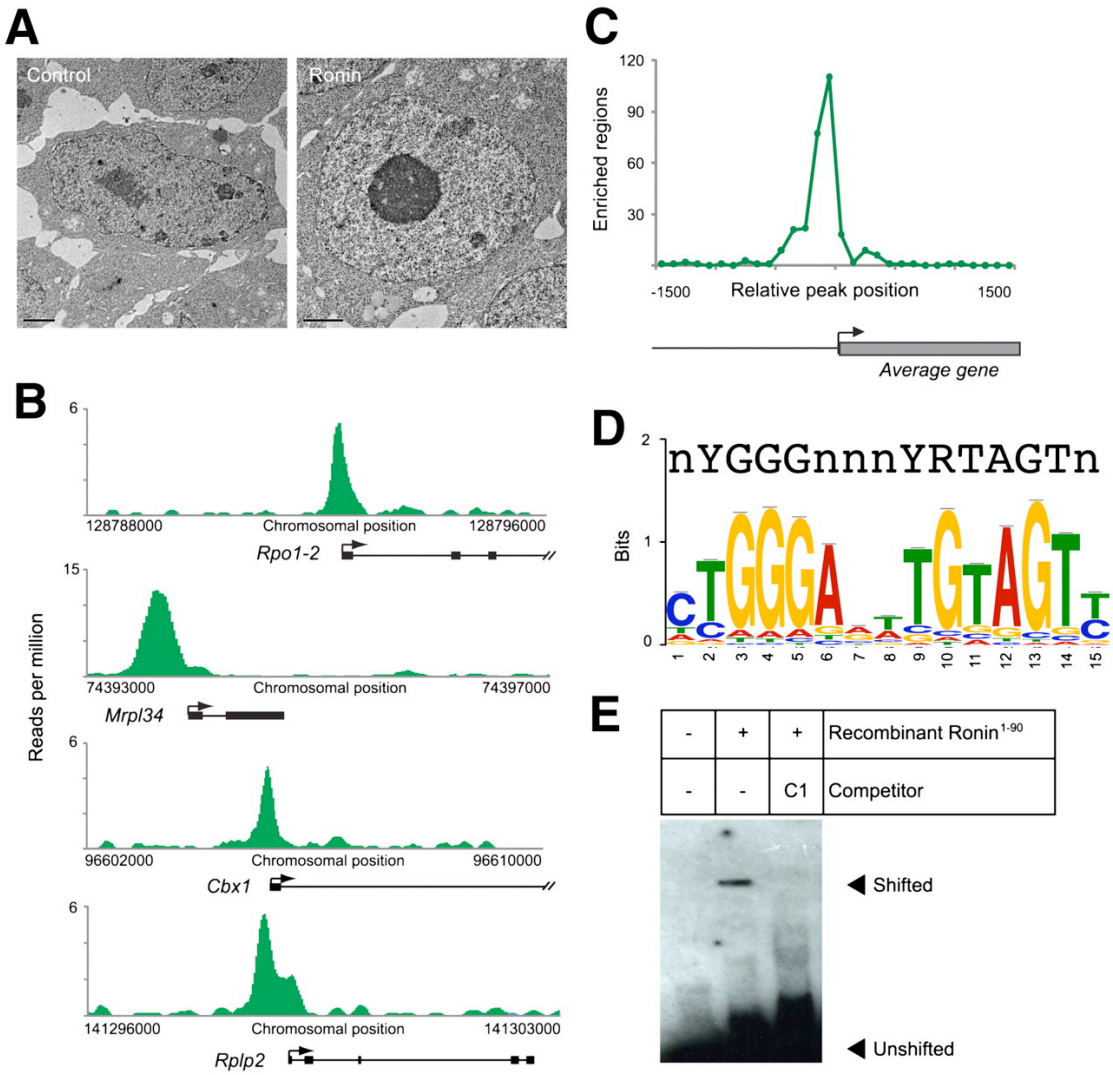

A
Control
Ronin
C
Enriched regions
120
90
60
30
-1500
Relative peak position
1500
Average gene
B
Reads per million
6
128788000
Chromosomal position
128796000
Rpo1-2
15
74393000
Chromosomal position
74397000
Mrpl34
6
96602000
Chromosomal position
96610000
Cbx1
6
141296000
Chromosomal position
141303000
Rplp2
D
nYGGGnnnYRTAGTn
Bits
2
1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
E
- + + Recombinant Ronin1-90
- - C1 Competitor
Shifted
Unshifted

Figure 1

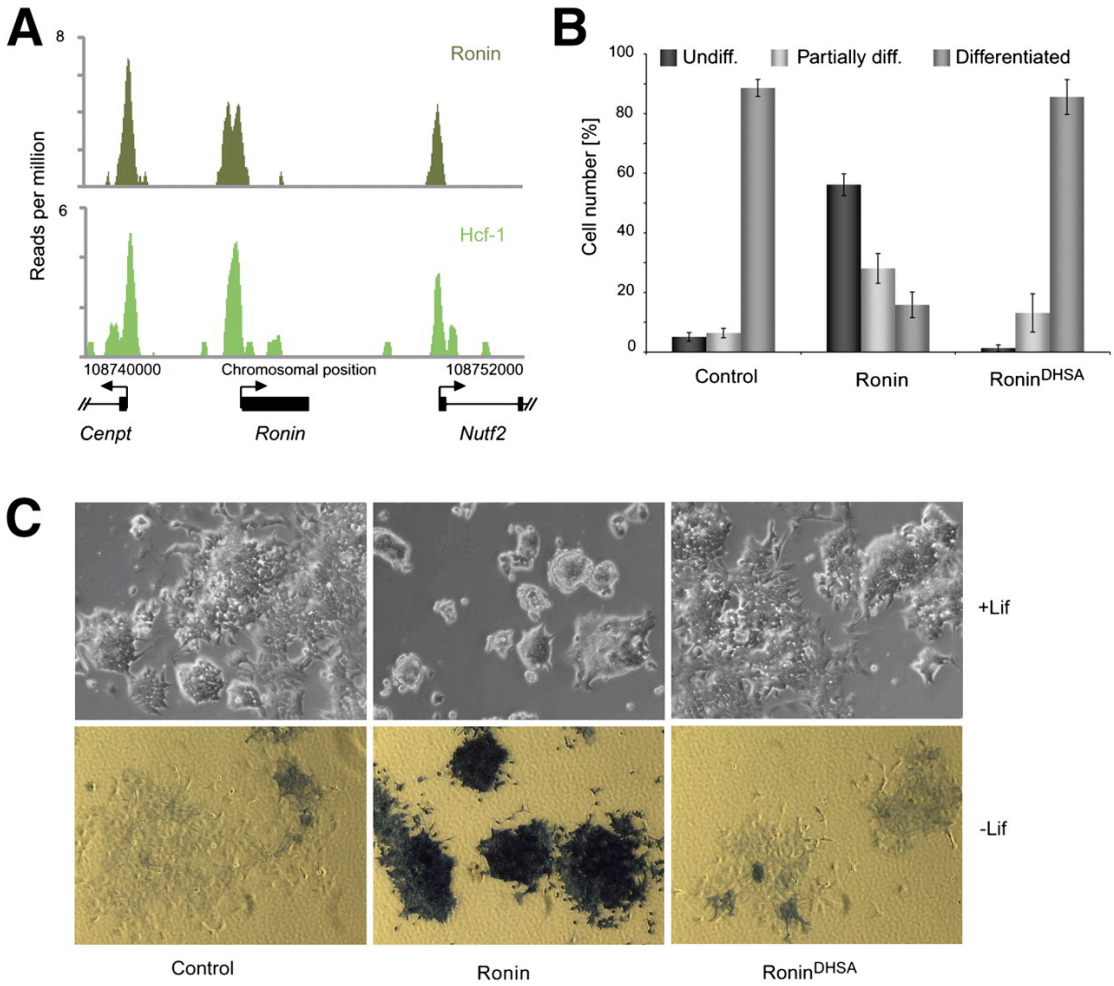
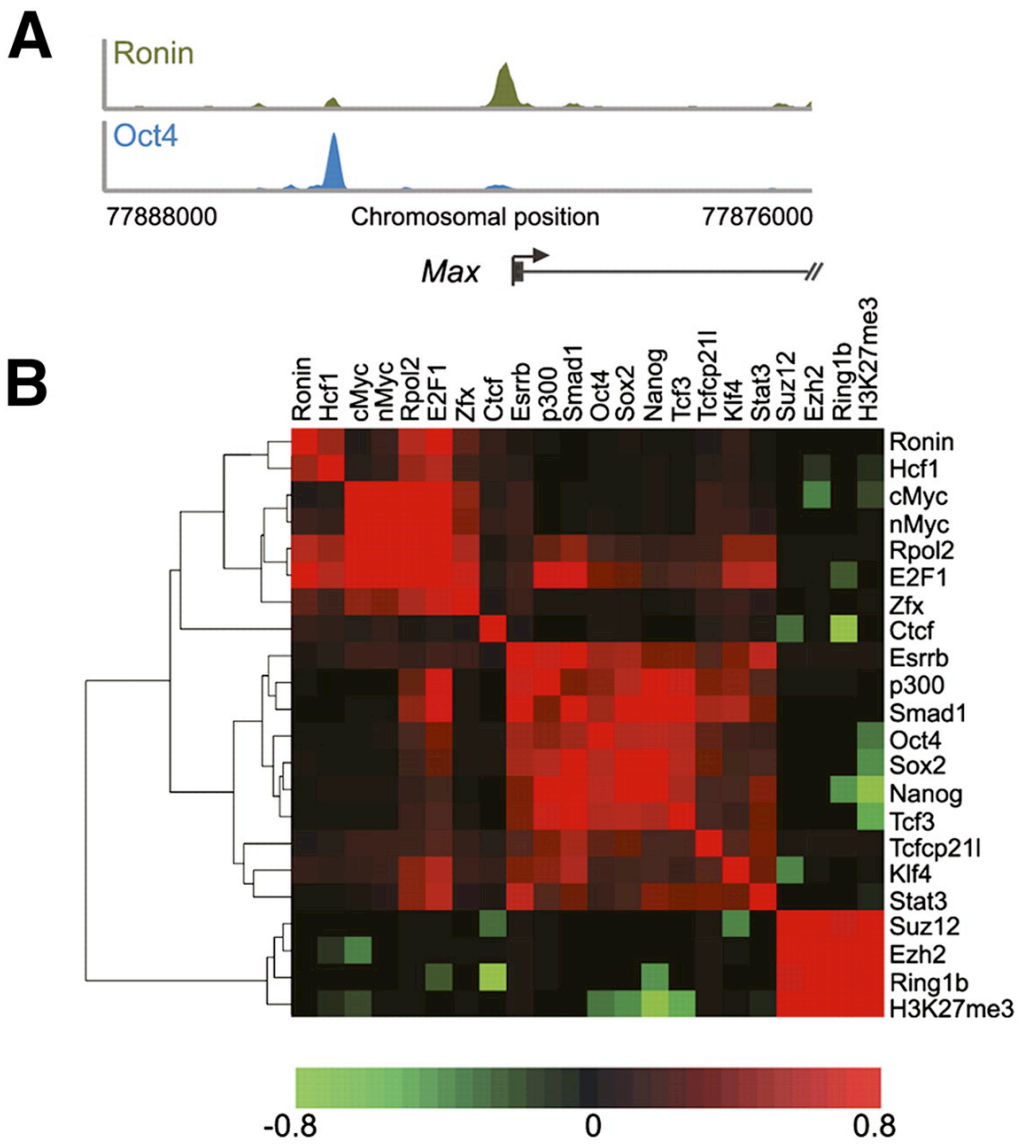

A
8
Ronin
Reads per million
6
Hcf-1
108740000
Chromosomal position
108752000
Cenpt
Ronin
Nutf2
B
Undiff.
Partially diff.
Differentiated
Cell number [%]
100
80
60
40
20
0
Control
Ronin
RoninDHSA
C
+Lif
-Lif
Control
Ronin
RoninDHSA

Figure 2


A
Ronin
Oct4
77888000
Chromosomal position
77876000
Max
B
Ronin
Hcf1
cMyc
nMyc
Rpol2
E2F1
Zfx
Ctcf
Esrrb
p300
Smad1
Oct4
Sox2
Nanog
Tcf3
Tcfcp21l
Klf4
Stat3
Suz12
Ezh2
Ring1b
H3K27me3
-0.8
0
0.8

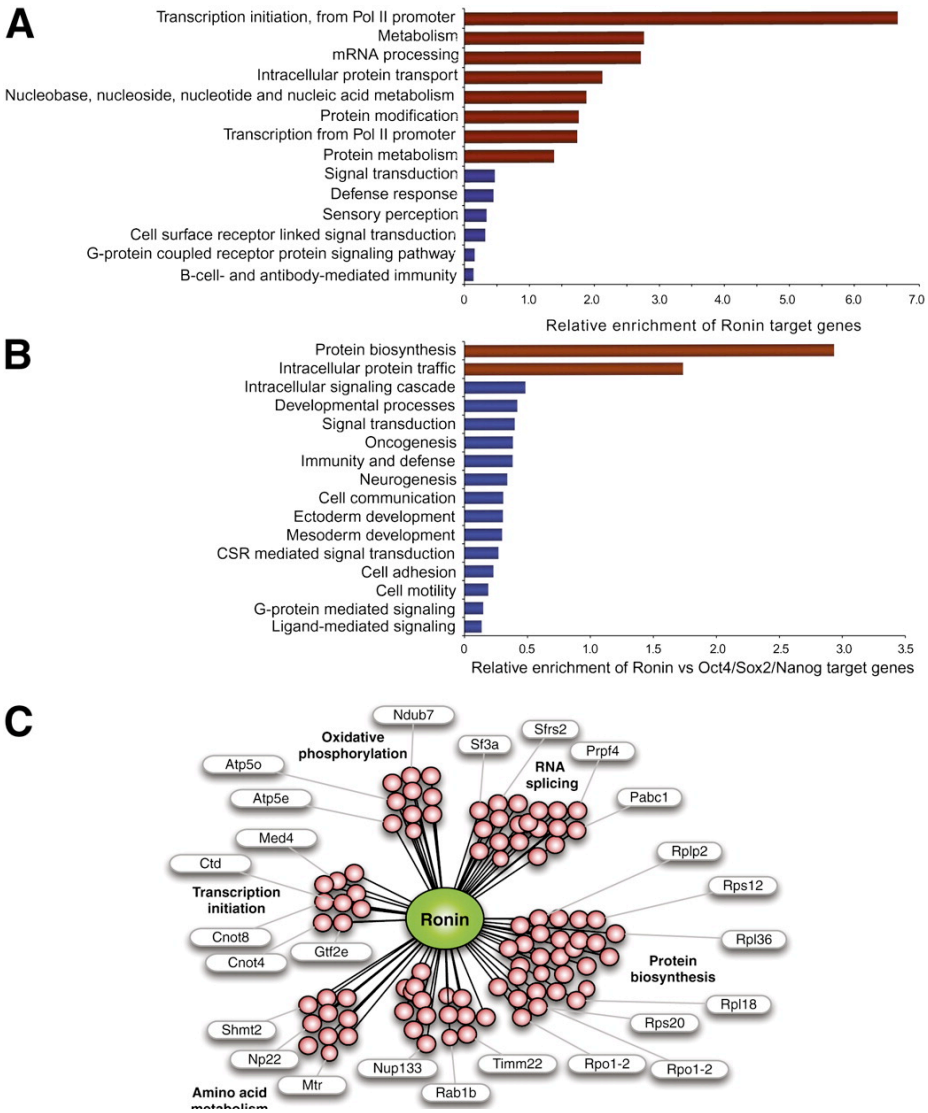

A
Transcription initiation, from Pol II promoter
Metabolism
mRNA processing
Intracellular protein transport
Nucleobase, nucleoside, nucleotide and nucleic acid metabolism
Protein modification
Transcription from Pol II promoter
Protein metabolism
Signal transduction
Defense response
Sensory perception
Cell surface receptor linked signal transduction
G-protein coupled receptor protein signaling pathway
B-cell- and antibody-mediated immunity
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
Relative enrichment of Ronin target genes
B
Protein biosynthesis
Intracellular protein traffic
Intracellular signaling cascade
Developmental processes
Signal transduction
Oncogenesis
Immunity and defense
Neurogenesis
Cell communication
Ectoderm development
Mesoderm development
CSR mediated signal transduction
Cell adhesion
Cell motility
G-protein mediated signaling
Ligand-mediated signaling
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Relative enrichment of Ronin vs Oct4/Sox2/Nanog target genes
C
Ronin
Oxidative phosphorylation
Ndub7
Atp5o
Atp5e
RNA splicing
Sf3a
Sfrs2
Prpf4
Pabc1
Transcription initiation
Med4
Ctd
Cnot8
Cnot4
Gtf2e
Protein biosynthesis
Rplp2
Rps12
Rpl36
Rpl18
Rps20
Rpo1-2
Rpo1-2
Timm22
Nup133
Rab1b
Shmt2
Np22
Mtr
Amino acid

Figure 4

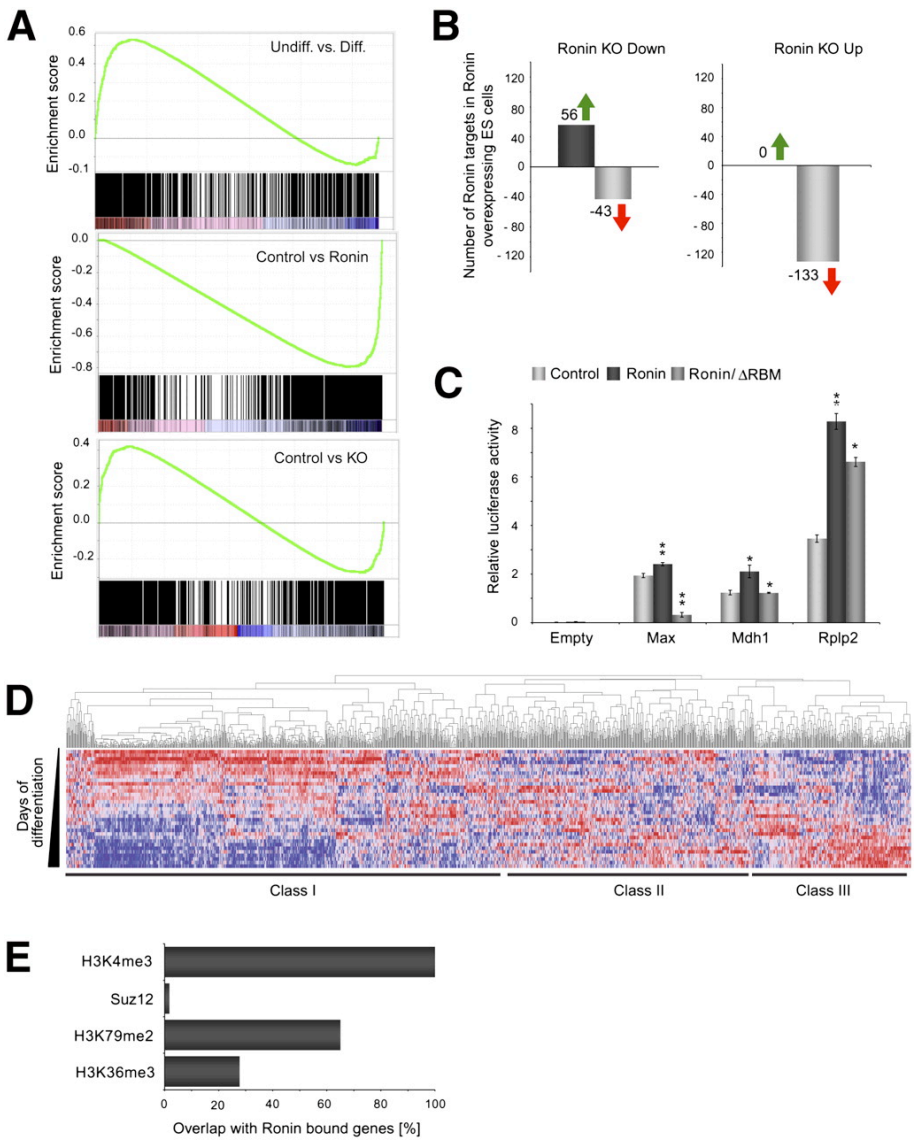

A
Undiff. vs. Diff.
Control vs Ronin
Control vs KO
Enrichment score
B
Ronin KO Down
Ronin KO Up
Number of Ronin targets in Ronin overexpressing ES cells
56
-43
0
-133
C
Control
Ronin
Ronin/ΔRBM
Relative luciferase activity
Empty
Max
Mdh1
Rplp2
D
Days of differentiation
Class I
Class II
Class III
E
H3K4me3
Suz12
H3K79me2
H3K36me3
Overlap with Ronin bound genes [%]

Figure 5

Appendix D:

Short RNAs Are Transcribed from Repressed Polycomb Target Genes And Interact With Polycomb Repressive Complex-2

Aditi S. Kanhere [1], Keijo Viiri [1], Carla C. Araújo [1], Jane Rasaiyaah [1], Russell D. Bouwman [1], Warren Whyte [2], C. Filipe Pereira [3], Emily Brookes [3], Kimberly Walker [2], George W. Bell [2], Ana Pombo [3], Amanda G. Fisher [3], Richard A. Young [2,4] and Richard G. Jenner [1,5]

1. Division of Infection and Immunity, University College London, London, W1T 4JF, United Kingdom. 2. Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA. 3. MRC Clinical Sciences Centre, Imperial College London, London W12 0NN, United Kingdom. 4. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02141, USA

Personal Contribution to the Project

This chapter is the result of a collaboration with Aditi Kanhere on work that was published in 2010 in *Molecular Cell*. Aditi requested assistance in determining the role for short RNAs during the differentiation of ES cells. I conducted in vitro differentiation experiments, and the results are described in this appendix.

Summary

Polycomb proteins maintain cell identity by repressing the expression of developmental regulators specific for other cell types. Polycomb repressive complex-2 (PRC2) catalyses trimethylation of histone H3 lysine-27 (H3K27me3). Although repressed, PRC2 targets are generally associated with the transcriptional initiation marker H3K4me3 but the significance of this remains unclear. Here, we identify a new class of short RNAs, ~50-200 nucleotides in length, transcribed from the 5'-end of polycomb target genes in primary T-cells and embryonic stem cells. Short RNA transcription is associated with RNA polymerase II and H3K4me3, occurs in the absence of mRNA transcription and is independent of polycomb activity. The short RNAs form stem-loop structures resembling PRC2 binding sites in Xist, interact with PRC2 through the SUZ12 subunit and, in neuronal cells, are specifically lost from genes no longer subjected to polycomb-mediated repression. We propose that short RNAs play a role in association of PRC2 with its target genes.

Introduction

Polycomb group (PcG) proteins are essential for embryogenesis and for maintaining embryonic stem (ES) cell pluripotency and differentiated cell states (Boyer et al., 2006; Cui et al., 2006; Faust et al., 1995; Kimura et al., 2001; Lee et al., 2006; Molofsky et al., 2003; O'Carroll et al., 2001; Pasini et al., 2007; Richie et al., 2002; van der Stoop et al., 2008). PcG proteins form at least two polycomb repressive complexes (PRCs) that are conserved across Metazoa (Schwartz and Pirrotta, 2007). PRC2 catalyses trimethylation of lysine 27 of histone H3 (H3K27me3), forming a binding site for PRC1 (Cao et al., 2002; Wang et al., 2004).

Genome-wide measurements of PRC2 localisation and H3K27 methylation have revealed that PRC2 represses the expression of hundreds of developmental regulators in ES cells that would otherwise induce cell differentiation (Azuara et al., 2006; Boyer et al., 2006; Lee et al., 2006). In differentiated cells, activated genes important for the identity of those cells lose H3K27me3 whereas genes that regulate alternate cellular identities remain methylated and repressed (Bernstein et al., 2006; Lee et al., 2006; Mikkelsen et al., 2007; Pan et al., 2007; Roh et al., 2006).

Although repressed, PRC2 target genes are thought to adopt a poised state that allows their rapid upregulation upon ES cell differentiation (Boyer et al., 2006; Lee et al., 2006). This poised state is reflected by the association of PRC2 target genes with histone H3K4me3, a marker of transcriptional initiation (Azuara et al., 2006; Bernstein

et al., 2006; Roh et al., 2006). The unphosphorylated and Ser-5 phosphorylated forms of RNA polymerase II have also been detected at some polycomb target genes but not the Ser-2 phosphorylated form that is associated with transcriptional elongation (Dellino et al., 2004; Lee et al., 2006; Stock et al., 2007). The block to RNA polymerase II elongation at these genes is due, at least in part, to the activity of PRC1, which binds to H3K27me3. PRC1 contains Ring1 that ubiquitinates H2A, blocking elongation by RNA polymerase II (Stock et al., 2007; Zhou et al., 2008).

PRC1 and PRC2 have no known DNA sequence binding specificity and it is not clear how PRC2 associates with its target genes. Recent results have shown that PRC2 can interact with long RNA transcripts. PRC2 interacts with the 1.6 kb ncRNA RepA transcribed from *Xist* and this is necessary for H3K27me3 and X-chromosome inactivation (Zhao et al., 2008). Similarly, the ncRNA HOTAIR associates with PRC2 and induces methylation of the *HOXD* locus in trans (Rinn et al., 2007) and the antisense transcript Kcnq1ot1 associates with PRC2 and induces methylation of *Kcnq1* (Pandey et al., 2008). Possibly related to such mechanisms, exogenous siRNAs designed against the promoters of the genes *EEF1A1* and *CCR5* induce H3K27 methylation and gene repression that requires the expression of long RNAs transcribed from alternative promoters upstream of the major mRNA transcription start site (TSS) (Han et al., 2007). These results suggest that ncRNAs may be important for PRC2 localisation across the genome.

Because polycomb target genes show evidence of stalled RNA polymerase II and do not transcribe appreciable amounts of mRNA, we hypothesised that short RNAs may be transcribed from these genes. We describe here the identification of such a set of short RNAs transcribed from the 5' ends of polycomb target genes in primary human CD4+ T cells and in murine ES cells. These RNAs are transcribed from sites enriched for markers of transcriptional initiation, but not transcriptional elongation, independently of polycomb function and are absent from genes that are derepressed in neuronal cells. Short RNAs interact with PRC2 and we propose that this may stabilize the association of PRC2 with chromatin.

Results

Detection of short RNAs associated with the 5' end of genes

We designed a DNA microarray to identify short RNAs at the 5' end of human genes (Figure 1A and Figure S1). RNA from primary human CD4+ T cells was fractionated by size into populations of short (<200nt) and large RNAs, labeled with fluorescent dyes and hybridized to the microarrays in replicate experiments (Figure S2). The arrays contained multiple probes allowing detection of sense-strand RNA at the promoter, 5'-most exons, 1st intron and 3'-most exons of known protein-coding genes (Supplemental Data and Table S1). These arrays also contained probes for known short nuclear RNA (snRNA) and short nucleolar RNA (snoRNA) that acted as positive controls and produced strong signals in the short RNA channel of the arrays compared to the mRNA and total RNA channels (Figure 1B and Figure S2).

We investigated whether we could detect short RNAs transcribed from protein-coding genes. Applying a threshold derived from our snRNA probes, we found thousands of short RNAs transcribed from the sense-strands of promoters, exons and introns of protein-coding genes in replicate experiments (Figure 1C, Table S2). The RNAs were concentrated at the 5' end of genes, within 700bp upstream and downstream of the mRNA TSS (Figure 1D). A substantial subset of the genes associated with short RNA at their 5' ends in primary human CD4+ T cells also produced short RNAs in cell lines and ES cells (Kapranov et al., 2007; Seila et al., 2008) (Figure 1E), although often from different locations (Figure S2). These results indicate that short RNAs are

generated in primary human T cells, are most abundant in promoter-proximal sequences, and are a general feature of the transcriptome of normal somatic cells.

Transcription of short RNAs from genes that are otherwise repressed

The short RNAs described in previous reports are typically associated with transcriptionally active genes but not with repressed polycomb target genes (Affymetrix/CSHL ENCODE, 2009; Core et al., 2008; Kapranov et al., 2007; Seila et al., 2008; Taft et al., 2009). We therefore examined the transcriptional status of human T cell genes that produce short RNAs. We found that short RNAs could be detected both at genes that produced mRNA transcripts and at genes that do not produce mRNA transcripts (Figure 2A) (Su et al., 2004) (Figure 2A). Although the mRNA signals are quite different between these two sets of genes, the short RNA signals are similar in the two classes (Figure 2B). This indicates that the transcription of short RNAs can occur in the absence of mRNA transcription.

We next used chromatin immunoprecipitation coupled with microarray analysis (ChIP-Chip) to determine whether RNA pol II was present at genes that transcribe short RNAs but do not produce mRNA (Figure 2C). The antibody used preferentially recognises the non-phosphorylated form of RNA pol II that initiates transcription (Thompson et al., 1989). We plotted average enrichment for RNA pol II across the TSS of genes that transcribe short RNA but not mRNA and compared this to genes that transcribe neither RNA type. We found that RNA pol II was enriched at genes that produce short RNAs but not mRNA. This enrichment of RNA pol II at these

genes was also apparent using data from independent ChIP-Seq experiments (Barski et al., 2007) (Figure S3).

At genes that produce short RNAs but not mRNA, the location of the short RNA correlates with the position of the initiating form of RNA pol II. At the set of genes where short RNA was detected at promoter regions, the RNA pol II peak was shifted upstream of the mRNA TSS. At the set of genes where short RNA was detected at exon or introns, RNA pol II occupancy was enriched downstream of the mRNA TSS (Figure 2D and Figure S3). The positioning of RNA pol II at the site of short RNA production, as opposed to the annotated mRNA TSS, can also be observed at individual genes (Figure 2E). At genes that produce both short RNA and mRNA, RNA pol II was concentrated at the mRNA TSS (Figure S3). These results argue that the RNA pol II observed at repressed genes is primarily involved in the transcription of short RNA rather than mRNA and that these short RNAs can be produced from TSS distinct from the mRNA TSS.

We next examined the transcriptional status of RNA pol II at short RNA loci. We used ChIP-Chip to map; H3K4me3, a marker of transcriptional initiation, and H3K79me2, a marker of transcription through a gene (Guenther et al., 2007; Steger et al., 2008). We first examined the average enrichment profile for these histone modifications across all human genes and found that H3K79me2 showed comparable enrichment to H3K4me3 (Figure 2F). In contrast, genes that produced short RNAs but not mRNAs were enriched for H3K4me3 but not H3K79me2 (Figure 2G). The presence of RNA pol II

and H3K4me3 in the absence of H3K79me2 was also apparent when examining genes individually (Figure S3). From these data we conclude that genes associated with short RNAs experience transcriptional initiation but the protein-coding portion of the gene is not transcribed.

Repressed genes producing short RNAs are associated with H3K27me3

Genes repressed by polycomb proteins are generally associated with nucleosomes containing H3K4me3, a marker of transcriptional initiation, and often RNA pol II (Azuara et al., 2006; Bernstein et al., 2006; Chopra et al., 2009; Roh et al., 2006; Stock et al., 2007). We considered the possibility that genes that generate short transcripts but not mRNA could be associated with polycomb proteins. To test this, we used ChIP-Chip to measure H3K27me3 in human CD4+ T cells and compared the results to our short RNA data. We found that genes that transcribed short RNA but not mRNA were enriched for H3K27me3 (Figure 3A and Figure S4). These results were also confirmed using independent measurements of H3K27 from ChIP-Seq experiments (Barski et al., 2007) (Figure 3B). H3K27me3 could be detected more frequently at genes that transcribe short RNA but not mRNA than at any other category of gene (Figure 3C), demonstrating the close association between short RNA transcription and H3K27 methylation.

Transcription of short RNAs from polycomb target genes would explain the association of these genes with H3K4me3. We therefore sought to verify that the transcriptional machinery was present at H3K27-methylated genes that express short

RNAs. Plotting enrichment of H3K27me3, H3K4me3 and RNA pol II across these genes confirms that transcriptional initiation occurs at polycomb target genes in CD4+ T cells (Figure 3D and E). Strikingly, H3K27-methylated nucleosomes flank the sites of RNA pol II and H3K4me3 occupancy, consistent with a block to transcriptional elongation previously associated with this modification (Stock et al., 2007; Zhou et al., 2008).

PRC2 targets developmental regulators that must be repressed to maintain cellular identity and thus plays a critical role in T cell development (Barski et al., 2007; Boyer et al., 2006; Chang and Aune, 2007; Koyanagi et al., 2005; Lee et al., 2006; Schoenborn et al., 2007). We therefore asked whether the set of genes that expressed short RNAs in the absence of mRNA were enriched for genes that play a role in development. Using Gene Ontology (GO) (Dennis et al., 2003) to annotate gene function, we found that functional categories such as multicellular development, nervous system development and cell-cell signalling were all significantly enriched in the set of genes from which short RNA was transcribed in the absence of mRNA (Figure 3F). These GO terms were not enriched in the set of repressed genes that are not associated with short RNA. Therefore, consistent with polycomb-mediated silencing, genes from which short RNA is transcribed in the absence of mRNA tend to have functions in development and cell differentiation.

Short RNAs transcribed from polycomb target genes are ~50-200 nt in length

We next used Northern blotting to verify the transcription of short RNAs from polycomb target genes and to perform a more accurate determination of their size. We selected array probes that detected short RNAs at genes for which no mRNA could be detected and were also associated with H3K27me3 (Figure S4). We then purified short RNA from peripheral blood mononuclear cells (PBMC), treated it with DNase and performed northern blotting for short RNAs (Figure 3G). 14 of the 16 probes we tested (88%) detected a short RNA species. Some probes identified a single RNA species of between ~50 and 200 nt, while others detected multiple RNA products. Short RNA transcripts could be detected from exons, introns and promoter regions, consistent with our array data. We did not detect RNA species between 20-30 nt, suggesting that these polycomb-associated short RNAs are not substrates for Dicer and are not involved in RNAi-mediated transcriptional gene silencing.

Short RNA transcription is not dependent on polycomb activity

Methylation of H3K27me3 is necessary for transcriptional repression by polycomb group proteins. We considered whether the production of short RNAs was dependent on H3K27 methylation. To test this, we made use of the murine ES cell line Ezh2-1.3, in which deletion of the Ezh2 SET domain can be induced by tamoxifen (derived from mice described by Su et al., 2003). We first tested whether short RNAs were transcribed from polycomb target genes in ES cells. We identified short RNA loci conserved between human and mouse and used histone methylation data (Boyer et al., 2006) to identify those targeted by polycomb in ES cells. Northern blotting detected a short RNA at each of these genes in murine ES cells (Figure 4A), indicating that the

transcription of short RNAs from repressed polycomb target genes is common to different cell types and conserved between different mammalian species.

We next measured short RNA transcription in Ezh2-1.3 cells at timepoints after the addition of tamoxifen. Loss of full-length Ezh2 protein (together with the appearance of a truncated form lacking the SET domain) and a reduction in H3K27me3 could be observed in these cells over the five-day timecourse (Figure 4B). Blotting for short RNAs at the genes Hes5, Msx1 and Ybx2 showed that loss of Ezh2 and H3K27me3 had no effect on the levels of short RNAs at these genes (Figure 4C). These results show that the production of short RNAs at polycomb target genes is not dependent on H3K27 methylation.

The block to mRNA transcription at bivalent genes is dependent on the PRC1 component Ring1 that catalyzes the ubiquitination of histone H2A (Stock et al., 2007). Deletion of Ring1 causes activation of PRC1 target genes, including Msx1. We blotted for short RNAs in the murine ES cell line ES-ERT2 with and without addition of tamoxifen that induces deletion of *Ring1b* and loss of H2AK119ub (Stock et al., 2007). As we found for deletion of *Ezh2*, loss of Ring1b had no effect on short RNA transcription (Figure 4D). These data show that transcription of short RNAs is not dependent on H2A ubiquitination and, taken together with the Ezh2 deletion experiments, indicate that short RNA transcription is independent of polycomb activity.

Short RNAs encode stem-loop structures and interact with PRC2

The long ncRNAs HOTAIR and Xist RepA interact with PRC2 and this interaction is necessary for H3K27me3 of *HOXD* and the X-chromosome, respectively. We therefore considered that PRC2 may interact with the short ncRNAs identified here.

The PRC2 binding site within mouse Xist RepA appears to be a double stem loop structure that is repeated 7 times (Zhao et al., 2008). We therefore first examined whether the short RNAs we have identified could form such a structure. We derived a general structural motif based on the RepA sequence (Figure 5A) and searched for the presence of this motif in the DNA sequences immediately surrounding probes that detect short RNAs from H3K27-methylated genes. We found that 71% of these sequences encode this PRC2-binding structure, compared to 36% of control sequences not associated with short RNAs but with an equal distribution around the mRNA TSS. 29% of short RNA sequences contained 2 or more consecutive structure motifs, compared with 7% for control sequences. Examples of these structures are given in Figure 5B. These data indicate that short RNAs transcribed from polycomb target genes have the potential to interact with PRC2. However, we found that these structures were not limited to short RNAs transcribed from polycomb-associated genes and that they were also present within short RNAs transcribed from genes not associated with H3K27me3 (Figure 6B). These RNAs may therefore also have the potential to bind PRC2.

To test for an interaction between short RNAs encoding the stem-loop structure and PRC2, we performed electromobility shift assays (EMSA). Consistent with recent

observations (Zhao et al., 2008), we found that incubation of T-cell lysate with radiolabeled RNA oligonucleotides encoding the Xist-RepA stem-loop produced a mobility shift indicative of binding by cellular proteins and that mutation of the RNA stem-loop structure abrogated this interaction (Figure 5C). We then repeated the experiment with RNA oligonucleotides corresponding to the stem-loops encoded by BSN, C20orf112, HEY1, MARK1 and PAX3 short RNAs. We observed a similar shift with RNA oligonucleotides corresponding to these short RNA stem-loop structures as we did for Xist-RepA and did not observe an interaction when the BSN short RNA structure was disrupted (Figure 5C). The BSN short RNA stem-loop was also able to compete with Xist-RepA for binding to the cellular complex, indicating that short RNAs bind to PRC2 (Figure 5D).

We next sought to identify which PRC2 component was responsible for the interaction with short RNAs. We purified GST-tagged recombinant SUZ12, EZH2, EED and RBBP4 from E. coli and incubated each protein with radiolabeled Xist-RepA and BSN oligonucleotides (Figure 5E). We found that SUZ12 interacted strongly with Xist-RepA and BSN short RNA stem-loops. SUZ12 also interacted with other short RNA stem-loops (Figure 5F), displayed a weaker interaction with mutated Xist-RepA RNA (Figure 5G) and did not bind to the BSN sequence encoded by single-stranded DNA, double-stranded DNA or DNA:RNA duplexes (Figure 5H). These data demonstrate that SUZ12 specifically interacts with RNA stem-loop structures encoded by Xist-RepA and short RNAs.

PRC2 interacts with short RNAs in cells

To test for an interaction between short RNAs and PRC2 in living cells, we immunoprecipitated the PRC2 subunit SUZ12 from a female T cell line, isolated co-purifying RNA and subjected this to quantitative reverse-transcription PCR (Figure 6A). In parallel, we performed RNA IP experiments with a non-specific rabbit control antibody. Immunoprecipitation with the Suz12 antibody, but not the control antibody, enriched for Xist RNA over Actin mRNA and mRNAs encoding GAPDH and HPRT, indicating we could specifically detect PRC2-RNA interactions in these cells. We then performed RT-PCR for short RNAs transcribed from polycomb target genes. We found that 4 of the 5 short RNAs we tested were enriched by Suz12 IP, although to a lesser extent than Xist. These short RNAs also possessed a stem-loop structure (Figure S5). Amplification of short RNAs was not observed in control reactions lacking reverse transcriptase, demonstrating that these results were not due to contaminating DNA (Figure 6B) and enrichment was maintained under stringent wash conditions (Figure S5). The snRNAs U1, U2 and U3 and the short structured RNAs 7SK and 5S ribosomal RNA, transcribed by RNA pol III, were not enriched by Suz12 IP, demonstrating that enrichment of short RNA was specific to the set transcribed from the 5' ends of genes.

We also examined short RNAs transcribed from genes that were not associated with H3K27me3. We found that these were also often enriched by Suz12 IP, consistent with the presence of double stem-loop structures in these transcripts (Figure 6A).

Therefore, PRC2 interacts with short promoter-associated RNA but this interaction does not necessarily lead to H3K27 methylation.

We next asked whether the PRC2-binding stem-loop structures could affect transcription. To test this, we modified the HIV LTR to encode short RNA sequences. The R portion of the LTR located immediately downstream of the TSS encodes the structured RNA transactivation response element (TAR) that is present in the 5' UTR of HIV transcripts. We replaced the R and U5 portions of the LTR with the Xist-RepA stem loop, the C20orf112 stem loop, which also interacts with PRC2, or the Xist-RepA mutant that does not interact with PRC2 and used these constructs to drive expression of luciferase in Hela cells (Figure 6C). We found that there was a significant drop ($p<0.05$) in luciferase activity in cells transfected with constructs containing the wild-type Xist-RepA and C20orf112 stem-loops but not in cells transfected with constructs containing the mutated Xist-RepA sequence. These data show that the stem-loop structures present in Xist and short RNAs can repress mRNA expression.

Short RNAs are lost from polycomb target genes active in other cell types

Differential H3K27 methylation allows for cell-type specific expression of developmental regulators and this underlies the differing identities of specialized cell types. We hypothesised that short RNAs may be specifically lost from polycomb target genes that are derepressed in other cell types. To address this, we examined short RNA expression in the neuronal cell line SH-SY5Y. We choose neuronal cells

because our Gene Ontology analysis (Figure 3F) revealed that many genes that express short RNAs in CD4+ T cells play roles in neuronal development. We used gene expression and functional data to select polycomb target genes that are repressed in CD4+ T cells and active in neuronal tissue (Figure S6). We then purified short RNAs from both PBMC and the neuronal cultures and blotted for short RNAs identified at these genes in CD4+ T cells (Figure 7A). We found that the increased expression of mRNA from the genes *FOXN4, HEY1, MARK1, NKX2-2, BSN* and *HES5* in the neuronal cells was accompanied by a reduction in the expression of short RNAs. In contrast, short RNAs transcribed from the 1st exon of *YBX2*, a gene expressed only in germ cells, or from the thyroid and lung-specific *NKX2-1*, were present equally in PBMC and neuronal cells. These results show that short RNAs are specific to polycomb target genes silent in a given cell type.

We next asked whether activation of polycomb target genes during differentiation of ES cells into specialised cell types is accompanied by loss of short RNAs. To test this, we differentiated murine ES cells through embryoid bodies and neural precursors to precursor motor neurons over 4 days (Wichterle et al., 2002). During this process there is an increase in the expression of mRNA encoding the neuronal proteins Hes5 and Pcdh8 (Figure 7B). Blotting for short RNAs at these genes reveals that activation of *Hes5* and *Pcdh8* is accompanied by a progressive loss of short RNAs (Figure 7C). As the levels of the ~190 nt Hes5 short RNA decrease, there is a concomitant increase in the levels of shorter species, implying that the longer RNA is degraded. We conclude

that derepression of polycomb target genes during ES cell differentiation is accompanied by the loss of short RNA transcripts.

Discussion

We report the identification of a novel class of short RNAs transcribed from the 5' end of polycomb target genes in primary CD4+ T cells and in ES cells. These short RNAs are approximately 50-200nt in length and are transcribed from promoters, introns and exons of protein-coding genes by RNA pol II independent of polycomb activity. The short RNAs interact with PRC2 through a stem-loop structure and are lost from polycomb target genes derepressed during cell differentiation.

Relationship to other short RNAs

Short RNAs of less than 200nt have previously been found to be transcribed from the 5' end of active genes in ES cells and cell lines (Affymetrix/ CSHL Laboratory ENCODE 2009; Kapranov et al., 2007, Core, 2008 #4; Seila et al., 2008; Taft et al., 2009) and our detection of short RNAs in primary differentiated human cells demonstrates that they are a core component of the transcriptome. Previous studies have not identified short RNAs at polycomb target genes and there are possible reasons for this. Transcription run-on techniques (Core et al., 2008) detect transcripts as they are being synthesised and the absence of polycomb-associated RNAs from this dataset may be because they are not being continuously transcribed. Analysis of 18-30nt RNAs (Seila et al., 2008; Taft et al., 2009) will also miss polycomb-associated short RNAs that northern blotting demonstrates to be primarily ~50 to 200 nt in size. The appearance of smaller RNAs during *Hes5* activation hints at how the different sizes of RNA may be related. Previously described promoter-associated short RNAs

have been suggested to be processed from longer transcripts (Affymetrix/CSHL Laboratory ENCODE, 2009) but the lack of H3K79me2 at sites of polycomb-associated short RNA transcription indicates that these RNAs can be produced independently of more extensive transcription.

Implications for bivalent chromatin states

The transcription of short RNAs from polycomb target genes explains why these loci are often associated with H3K4me3 (Azuara et al., 2006; Bernstein et al., 2006; Roh et al., 2006). The association of genes with both H3K27me3 and H3K4me3 was first described in ES cells and was hypothesised to poise genes for activation during subsequent stages of embryogenesis (Azuara et al., 2006; Bernstein et al., 2006; Lee et al., 2006). The detection of RNA pol II phosphorylated at ser-5, but not ser-2, at polycomb target genes in ES cells suggested that the poised state reflected the stalling of RNA pol II during mRNA transcription, with this block being released upon ES cell differentiation (Stock et al., 2007). We find that short RNAs are often transcribed from sites independent of the mRNA TSS. Furthermore, short RNA levels do not change upon loss of H3K27 methylation or H2A ubiquitination and therefore short RNAs do not appear to be a by-product of polycomb activity. These results argue that the bivalent state reflects short RNA production *per se* and that RNA pol II is not necessarily stalled during the initial stages of mRNA production. This would explain why genes are associated with H3K27me3 and H3K4me3 in differentiated cells even though such genes are unlikely to be poised in preparation for subsequent activation

(for example neuronal genes in T-cells). Stalled RNA pol II at developmental genes in Drosophila embryos may also be engaged in short RNA transcription independent of mRNA production (Muse et al., 2007; Zeitlinger et al., 2007).

Interaction with PRC2

We have shown that there is no change in short RNA levels upon loss of H3K27 methylation or H2A ubiquitination. The same short RNAs do however decline upon differentiation of ES cells to neurons, indicating that their constancy upon Ezh2 and Ring1 loss is not merely a reflection of their stability. These data instead indicated that short RNAs could function upstream of polycomb and our binding data indicate that this function could be to stabilize the association of PRC2 with chromatin. The interaction of PRC2 with ncRNAs was previously described at *HOXD*, the inactive X-chromosome and Kcnq1. The polycomb protein Cbx7 has also been shown to interact with chromatin in an RNA-dependent manner (Bernstein et al., 2006). A more recent study described a large set of long intergenic RNAs that associate with PRC2 (Khalil et al., 2009) but how these may effect polycomb association with protein-coding genes is unclear. The interaction of PRC2 with short ncRNAs transcribed from polycomb-target genes themselves presents a simple model for the association of PRC2 with its target genes.

Of all the PRC2 subunits, we found that SUZ12 interacted most strongly with short RNA stem-loops. Ezh2 has previously been shown to interact with the Xist-RepA stem-loop (Zhao et al., 2008) but we only observed a weak interaction between EZH2

purified from bacteria and short RNA stem-loops. It is possible that the use of baculovirus as a source of recombinant Ezh2 by Zhao and colleagues could have resulted in co-purification of *Drosophila* Suz12 or that bacteria produce EZH2 protein lacking an important post-translational modification.

Loss of short RNAs during gene activation

Short RNA levels are reduced at polycomb-target genes depressed in neurons and this decrease can be observed directly as ES cells differentiate. Given the interaction between short RNAs and PRC2, their loss would seem likely to destabilize the association of PRC2 with chromatin, thereby allowing demethylation of H3K27 and mRNA transcription. Although this provides a mechanism for how polycomb-target genes become activated, it does not explain how other genes expressing short RNAs capable of interacting with PRC2 remain free of H3K27 methylation. At this set of genes, it is possible that PRC2 is inactive or counteracted by H3K27 demethylases.

In summary, we have identified a novel class of short RNAs that are transcribed from the 5' end of repressed polycomb target genes, explaining why these genes are associated with markers of transcriptional activation and providing a mechanism for the interaction of PRC2 with its target loci across the genome.

Experimental Procedures

RNA purification and fractionation

Total RNA was purified with Trizol (Invitrogen) and checked for degradation using an Agilent Bioanalyzer. Short RNA (<200bp) was purified using Ambion's mirVana miRNA purification system and size fractionation confirmed using a Bioanalyzer.

Microarray analysis

Short RNA was poly-adenylated (Ambion), amplified and labeled with Agilent's Low RNA Input Linear Amplification protocol and hybridised with labeled mRNA or fractionated long RNA to custom microarrays containing probes for promoters, 5'-exons, first intron and 3'-exons of human RefSeq genes. The log ratio distribution of probes for known short RNAs was used to devise an algorithm for the prediction of novel short RNAs ($\log_2$ ratio >1.5, p-value <0.01, normalized signal >50 in both experiments).

ChIP-Chip

DNA associated with RNA polymerase II, H3K4me3, H3K79me2, H3K27me3 and total H3 was enriched from CD4+ T cells with the following antibodies anti-RNA pol II (8WG16, Abcam), anti-H3 (Abcam ab1791), anti-H3K4me3 (Abcam ab8580), anti-H3K79me2 (Abcam ab3594) and anti-H3K27me3 (Abcam ab6002) and analysed with DNA microarrays (Agilent) according to published protocols (Lee et al., 2006).

Northern blotting

Short RNA was resolved on 15% acrylamide-7M Urea TBE gels (Invitrogen) and electroblotted to Nytran membranes (Whatman). RNA was crosslinked to the membrane and exposed to probes designed from microarray elements and radioactively labeled using StarFire 3'-extension (IDT). Estimates of RNA length made in comparison to Century and Decade markers (Ambion).

Conditional deletion of Ezh2 and Ring1

Ezh2: ES cells were derived from conditional Ezh2 knock-out mice (Su et al., 2003). Clone Ezh2-1.3 carries a conditional *Ezh2* mutation on both alleles induced by 4-hydroxy-tamoxifen (800 nM, Sigma) without consequent changes in ES cell self-renewal. Ezh2-1.3 were maintained on 0.1% gelatin-coated surfaces using KO-DMEM medium, 10% FCS, 5% knock-out serum replacement (Invitrogen) and 1000 U/ml of leukaemia inhibitory factor (Chemicon).

Ring1: ES-ERT2 cells were cultured and Ring1b deletion induced with tamoxifen as described (Stock et al., 2007).

RNA structure prediction

RNA structures were identified within 200nt of sequence surrounding probes that detected short RNAs using RNAmotif (Macke et al., 2001) and free-energy structures predicted using RNAfold (Hofacker and Stadler, 2006).

EMSA

Full-length SUZ12 was amplified from the IMAGE clone and cloned into pGEX-4T1 (GE). Plasmids encoding GST-EED, GST-EZH2 and GST-RBBP4 were kindly provided by Dr.Y. Zhang. Recombinant proteins were purified by glutathione-agarose and 4mg used for EMSA with ssRNA, ssDNA, dsDNA and DNA-RNA duplex oligonucleotides end-labeled with [γ-^{32}P]ATP, as described (Zhao et al., 2008). Nuclear extracts were prepared from the T-cell line CEM using reagents from Active Motif.

RNA IP

RNA associated with PRC2 was enriched from CEM cells with an antibody to Suz12 (Abcam) or unspecific rabbit antibody (Santa Cruz) following published protocols (Keene et al., 2006; Zhao et al., 2008). RNA was treated with DNase-turbo (Ambion), and reverse transcribed with SuperScript-III (Invitrogen) using random priming. Enrichment of RNA species compared to input RNA was quantified using real-time PCR and normalized to housekeeping gene mRNA.

Luciferase assay

The R and U5 regions of the HIV LTR from SF2 were replaced with DNA encoding the murine Xist-RepA stem-loop, mutated Xist-RepA and the stem-loop from C20orf112 short RNA, cloned in place of the CMV promoter in pIRESneo3 (Clontech) and luciferase from pCSFLW inserted downstream. Plasmid DNA was

transfected into Hela cells in triplicate together with the Renilla luciferase plasmid phRL-null. 48 hours later, firefly luciferase activity was measured in relation to Renilla luciferase with the Dual-luciferase reporter system (Promega). The significance of differences were estimated with a one-sided paired T-test.

Neuronal cell culture

SH-SY5Y cells were cultured in DMEM with 10% FCS until 40-50% confluent and then differentiated in DMEM with 5% FCS and 10mM retinoic acid (Sigma) for 7 days.

ES cell differentiation

ES cell-derived motor neuron precursors were generated as described elsewhere (Wichterle et al., 2002). Briefly, v6.5 mouse ES cells were partially dissociated and cultured in ADFNK medium optimized for ES cell to HB9+ motor neurons. After two days, 1mM retinoic acid and 0.5mg/ml Hedgehog agonist Hh-Ag1.3 were added and cells cultured for an additional 24hrs (for NPCs) or 48hrs (for PMNs).

Acknowledgments

Thanks to Anne Palser for advice on SH-SY5Y cell culture. This work was funded by the Medical Research Council through an MRC Career Development Award to RGJ, an MRC Centre grant that supports AK and CA and a NIHR project grant and core funding from the UCL/UCLH Comprehensive Biomedical Research Centre.

References

Affymetrix/Cold Spring Harbor Laboratory ENCODE Transcriptome Project (2009). Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature *457*, 1028-1032.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

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.

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., Allis, C.D. (2006). Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. Mol. Cell. Biol. *26*, 2560-2569.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K.*, et al*. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.

Chang, S., and Aune, T.M. (2007). Dynamic changes in histone-methylation 'marks' across the locus encoding interferon-gamma during the differentiation of T helper type 2 cells. Nat. Immunol. *8*, 723-731.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Cui, H., Ma, J., Ding, J., Li, T., Alam, G., and Ding, H.F. (2006). Bmi-1 regulates the differentiation and clonogenic self-renewal of I-type neuroblastoma cells in a concentration-dependent manner. J. Bio. Chem. *281*, 34696-34704.

Dellino, G.I., Schwartz, Y.B., Farkas, G., McCabe, D., Elgin, S.C., and Pirrotta, V. (2004). Polycomb silencing blocks transcription initiation. Mol. Cell *13*, 887-893.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. *4*, P3.

Faust, C., Schumacher, A., Holdener, B., and Magnuson, T. (1995). The eed mutation disrupts anterior mesoderm production in mice. Development *121*, 273-285.

Han, J., Kim, D., and Morris, K.V. (2007). Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. Proc. Natl. Acad. Sci. USA *104*, 12422-12427.

Hofacker, I.L., and Stadler, P.F. (2006). Memory efficient folding algorithms for circular RNA secondary structures. Bioinformatics *22*, 1172-1176.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L., *et al*. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484-1488.

Keene, J.D., Komisarow, J.M., and Friedersdorf, M.B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. Nat. Protocols *1*, 302-307.

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. USA *106*, 11667-11672.

Kimura, M., Koseki, Y., Yamashita, M., Watanabe, N., Shimizu, C., Katsumoto, T., Kitamura, T., Taniguchi, M., Koseki, H., and Nakayama, T. (2001). Regulation of Th2 cell differentiation by mel-18, a mammalian polycomb group gene. Immunity *15*, 275-287.

Koyanagi, M., Baguet, A., Martens, J., Margueron, R., Jenuwein, T., and Bix, M. (2005). EZH2 and histone 3 trimethyl lysine 27 associated with Il4 and Il13 gene silencing in Th1 cells. J. Biol. Chem. *280*, 31470-31477.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., *et al*. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Macke, T.J., Ecker, D.J., Gutell, R.R., Gautheret, D., Case, D.A., and Sampath, R. (2001). RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res *29*, 4724-4735.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al*. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Molofsky, A.V., Pardal, R., Iwashita, T., Park, I.K., Clarke, M.F., and Morrison, S.J. (2003). Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. Nature *425*, 962-967.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat. Genet. *39*, 1507-1511.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., and Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. Mol. Cell. Biol *21*, 4330-4336.

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.

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol. Cell. Biol. *27*, 3769-3779.

Richie, E.R., Schumacher, A., Angel, J.M., Holloway, M., Rinchik, E.M., and Magnuson, T. (2002). The Polycomb-group gene eed regulates thymocyte differentiation and suppresses the development of carcinogen-induced T-cell lymphomas. Oncogene *21*, 299-306.

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.

Roh, T.Y., Cuddapah, S., Cui, K., and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. Proc. Natl. Acad. Sci. USA *103*, 15782-15787.

Schoenborn, J.R., Dorschner, M.O., Sekimata, M., Santer, D.M., Shnyreva, M., Fitzpatrick, D.R., Stamatoyannopoulos, J.A., and Wilson, C.B. (2007). Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-gamma. Nat. Immunol. *8*, 732-742.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat. Rev. Genet. *8*, 9-22.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat. Cell Biol. *9*, 1428-1435.

Su, I., Basavaraj, A., Krutchinsky, A.N., Hobert,O., Ullrich, A., Chait, B.T. & Tarakhovsky, A. (2002). Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. Nat. Immunol. *4*, 124-131.

Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R., Grimmond, S.M., Schroder, K., *et al*. (2009). Tiny RNAs associated with transcription start sites in animals. Nat. Genet. *41*, 572-578.

Thompson, N.E., Steinberg, T.H., Aronson, D.B., and Burgess, R.R. (1989). Inhibition of in vivo and in vitro transcription by monoclonal antibodies prepared against wheat

germ RNA polymerase II that react with the heptapeptide repeat of eukaryotic RNA polymerase II. The J. Biol. Chem. *264*, 11511-11520.

van der Stoop, P., Boutsma, E.A., Hulsman, D., Noback, S., Heimerikx, M., Kerkhoven, R.M., Voncken, J.W., Wessels, L.F., and van Lohuizen, M. (2008). Ubiquitin E3 ligase Ring1b/Rnf2 of polycomb repressive complex 1 contributes to stable maintenance of mouse embryonic stem cells. PloS One *3*, e2235.

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004). Hierarchical recruitment of polycomb group silencing complexes. Mol. Cell *14*, 637-646.

Wichterle, H., Lieberam, I., Porter, J.A., and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. Cell *110,* 385-397.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet. *39*, 1512-1516.

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.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol. Cell *29*, 69-80.

Figure Legends

Figure 1. Detection of promoter-associated short RNAs in primary human T cells

A. Experimental strategy. Short RNA was purified human CD4+ T cells, poly-adenylated amplified and labeled with Cy5 and hybridized together with Cy3-labeled mRNA (experiment 1) or total long RNA (experiment 2) to a custom microarray containing probes for the promoters, 5' exons, first intron and 3' exons of human RefSeq genes, together with control probes for snRNA and snoRNAs.

B. Background-subtracted array signals for control snRNA and snoRNA probes in the mRNA channel (x-axis) and the short RNA channel (y-axis).

C. Fraction of promoter regions (~1 kb upstream of mRNA transcription start site), 5' exons (up to 1kb downstream of mRNA 5' end) and first introns that give rise to short RNAs.

D. Distribution of short RNAs relative to the TSS of the average gene, plotted as a moving average with a window of 10 bp.

E. Venn diagram showing the overlap between genes with associated short RNAs detected in this study compared with two previous studies (Kapranov et al., 2007; Seila et al., 2008).

Figure 2. Transcriptional status of short RNA genes.

A. Percentage of genes with promoter-associated short RNAs that produce detectable mRNA (black bars) or no detectable RNA (grey bars) according to our mRNA expression data (data set 1) or published data (data set 2 (Su et al., 2004)).

B. Array signals (median and inter-quartile range) for short RNAs (grey) and mRNA (white) at genes that produce detectable mRNA or no detectable mRNA.

C. Composite enrichment profile of RNA pol II at genes for which no mRNA can be detected. The plot shows average fold-enrichment (normalized signal from RNA pol II ChIP divided by the signal from input DNA) and genes are divided into those that are associated with short RNA (green) and those not associated with short RNA (blue). The start and direction of transcription of the average gene is indicated by an arrow.

D. Composite enrichment profile of RNA pol II at genes that express short RNAs in the absence of detectable RNA. Genes are divided into those that detect a short RNA with probes positioned -100 to +100 (red), -700 to -500 (green) or +500 to +700 (blue) relative to the mRNA TSS.

E. Examples of RNA pol II ChIP signals at short RNA loci. The plots show unprocessed enrichment ratios for all probes within a genomic region (RNA Pol II ChIP vs. whole genomic DNA). Chromosomal positions are from NCBI build 35 of the human genome. Genes are shown to scale below and aligned with the plots by chromosomal position (exons are represented by vertical bars, the start and direction of transcription by an arrow). The positions of short RNAs are indicated by the vertical blue arrow.

F. Composite enrichment profile of H3K4me3 (green) and H3K79me2 (red) across all human RefSeq genes. The plot shows average fold-enrichment (normalized signal from H3 methylation ChIP divided by whole H3 ChIP).

G. As F., except at genes associated with short RNA but not mRNA.

Figure 3. Short RNA loci are associated with H3K27me3.

A. Composite enrichment profile of H3K27me3 at genes for which no mRNA can be detected. These genes are divided into those that are associated with short RNA (green) and those not associated with short RNA (blue). The plot shows average fold-enrichment (normalized signal from H3K37me3 ChIP divided by whole H3 ChIP).

B. As for A. except for ChIP-Seq sequence tag density from H3K27me3 ChIP material (data from (Barski et al., 2007)).

C. Percentage of genes in different transcriptional categories that are associated with H3K27me3. Genes are first divided into those that produce detectable mRNA and those that do not and then into those that produce detectable short RNA and those that do not.

D. Heat maps showing enrichment of H3K27me3, H3K4me3 and RNA polymerase II at genes associated with short RNAs and H3K27me3. Each row represents one gene and each column represents the data from one oligonucleotide probe, ordered by their position relative to the transcription start site. For the first three panels, fold enrichment is indicated by color, according to the scale on the right. The last panel

overlays the RNA polymerase II (green) and H3K27me3 (red) to show their relative locations.

E. P-values for the enrichment of GO categories in the set of genes that are not associated with mRNA and transcribe short RNA (black bars) compared with those that do not produce detectable short RNA (grey bars).

F. Northern blotting for short RNAs transcribed from polycomb target genes in PBMC. The position of single-stranded RNA size markers are shown to the left. The genes and relative locations from which the short RNAs are transcribed are indicated above each blot.

Figure 4. Transcription of short RNAs in murine ES cells deficient for Ezh2 and Ring1.

A. Northern blotting for short RNAs transcribed from polycomb target genes in murine ES cells.

B. Western blotting for Ezh2, histone H3K27me3, and total histone H3 in Ezh2-1.3 ES cells over a 5-day treatment with tamoxifen that induces genetic deletion of the Ezh2 SET domain.

C. Northern blotting for Hes5, Msx1 and Ybx2 short RNAs and 5S rRNA during tamoxifen treatment of Ezh2-1.3 cells.

D. Northern blotting for Hes5, Msx1 and Ybx2 short RNAs and 5S rRNA in ES-ERT cells, 2 and 3 days after the addition of tamoxifen that induces genetic deletion of *Ring1b,* compared with untreated cells at the same timepoints.

Figure 5. Short RNAs interact with PRC2 in vitro.

A. Left: A stem-loop structure motif based on the repeat region of Xist-RepA RNA that is known to bind PRC2. Right: Percentage of short RNAs transcribed from genes associated with H3K27me3 that contain one or more (black bars) or two or more (grey bars) of the RNA structure motifs. The same data for genomic sequences showing identical distributions around transcription start sites but that do not produce detectable short RNAs are shown as controls.

B. Example energy structures found in short RNAs transcribed from repressed genes. The gene proximal to each short RNA sequence and the free energies are shown alongside each structure.

C. EMSA for the interaction of ^{32}P-labeled RNA probes corresponding Xist-RepA, Xist-RepA mut (stem-loop disrupted), BSN short RNA, mutated BSN short RNA, and C20orf112, HEY1, MARK1 and PAX2 short RNA probes with T-cell nuclear extract (NE).

D. Competition for Xist-RepA binding with 400 or 800-fold excess of cold Xist-RepA or BSN short RNA probes.

E. EMSA for the interaction of PRC2 subunits expressed as GST-fusion proteins with Xist-RepA and BSN short RNA probes. GST alone was used a negative control.

F. EMSA for the interaction of GST-SUZ12 with short RNA probes.

G. Affinity of GST-SUZ12 for wild-type and mutant Xist-RepA probes. Increasing amounts of SUZ12 were added (0.25, 1.0 and 5μg).

H. Binding specificity of SUZ12 defined by incubating GST-fusion protein with the BSN stem-loop sequence encoded by ssRNA, ssDNA, dsDNA and a RNA-DNA duplex.

Figure 6. Short RNAs interact with PRC2 in cells.

A. Fold enrichment (mean and SD, n=3) of different RNA species in Suz12 immunoprecipitate (grey) and control immunoprecipitate (white) compared with input RNA and normalized to Actin measured by quantitative reverse-transcription PCR. The genes and relative locations from which the short RNAs are transcribed are indicated above (E=exon, I=intron, P=promoter).

B. Amplification products from Suz12 IP cDNA and control cDNA reactions lacking reverse transcriptase using primers specific for small RNAs (n=3 for each PCR).

C. Left: Schematic showing the wild-type HIV LTR-luciferase construct and modified forms in which R and U5, encoding the transactivation response element (TAR), are replaced with the Xist-RepA stem-loop, the Xist-RepA mutant stem-loop or the C20orf112 short RNA stem-loop. Right: Luciferase activity in Hela cells transfected with plasmids encoding firefly luciferase downstream of wild-type or modified HIV LTRs indicated in the schematic to the left. Firefly luciferase activity is plotted relative to co-transfected Renilla luciferase and normalized to the wild-type LTR. Error bars are standard deviations across 7 independent experiments (performed with 2 independent clones) with each experiment comprising 3 measurements. * indicates comparisons that gave significance at $p<0.05$.

Figure 7. Loss of short RNAs from polycomb target genes activated in neuronal cells.

A. Northern blotting for short RNAs transcribed from genes targeted by polycomb in CD4+ T cells in RNA from PBMC (P) and from differentiated SH-SY5Y neuronal cells (N).

B. Gene expression microarray data showing the expression of Hes5 (blue) and Pcdh8 (red) mRNA during the step-wise 4-day differentiation to precursor motor neurons (PMN). EB (day 2 of differentiation process), embryoid bodies; EB+RA, embryoid bodies treated with retinoic acid (8 hours later on day 2); NPC, neuronal precursor cells (day 3).

C. Northern blotting for Hes5 and Pcdh8 short RNAs and 5S rRNA in ES cells during the step-wise 4-day differentiation to precursor motor neurons described above.

Figure 1

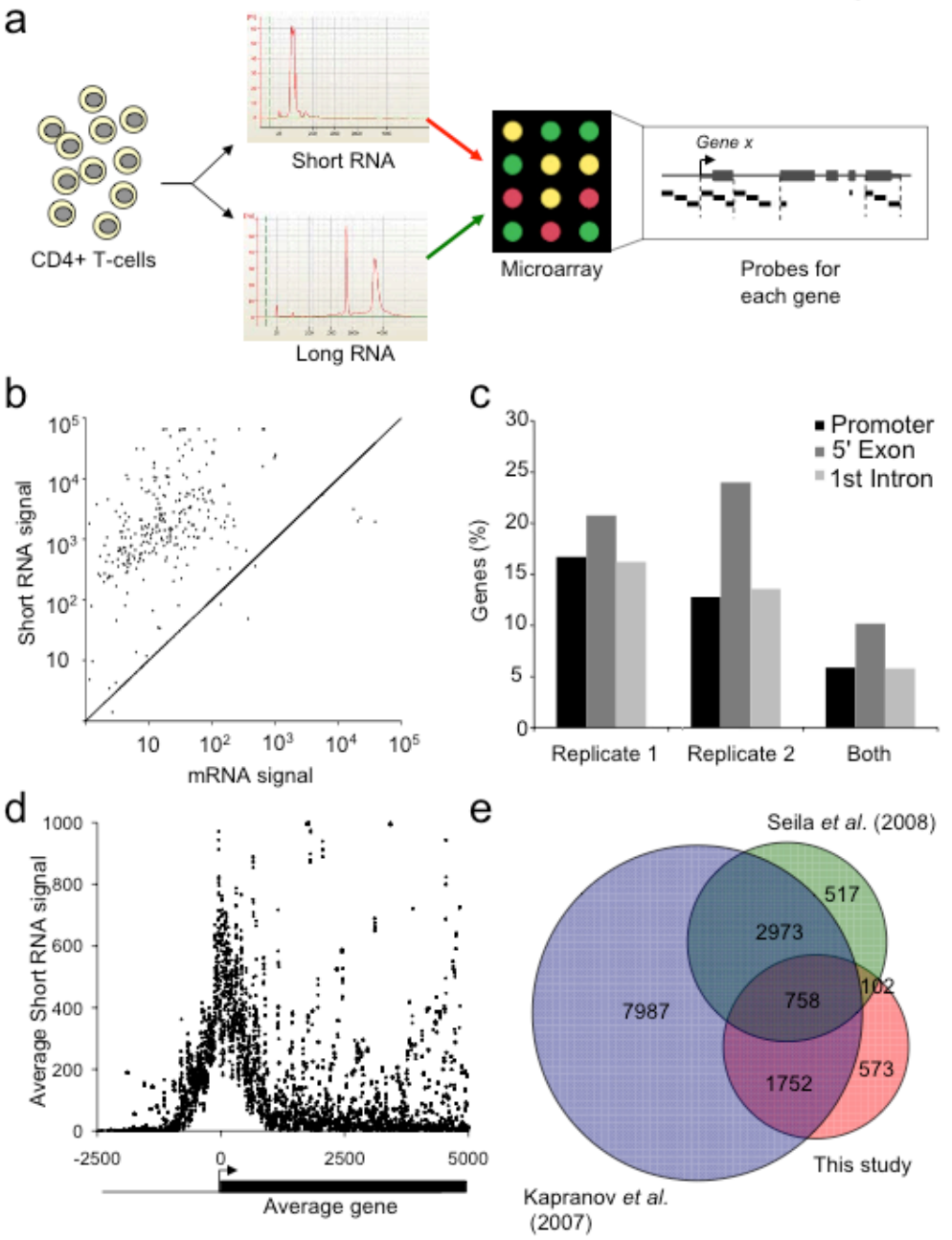

a
Short RNA
Long RNA
CD4+ T-cells
Microarray
Gene x
Probes for each gene
b
Short RNA signal
mRNA signal
c
Promoter
5' Exon
1st Intron
Genes (%)
Replicate 1
Replicate 2
Both
d
Average Short RNA signal
Average gene
e
Seila et al. (2008)
517
2973
102
758
7987
573
1752
This study
Kapranov et al. (2007)

Figure 2

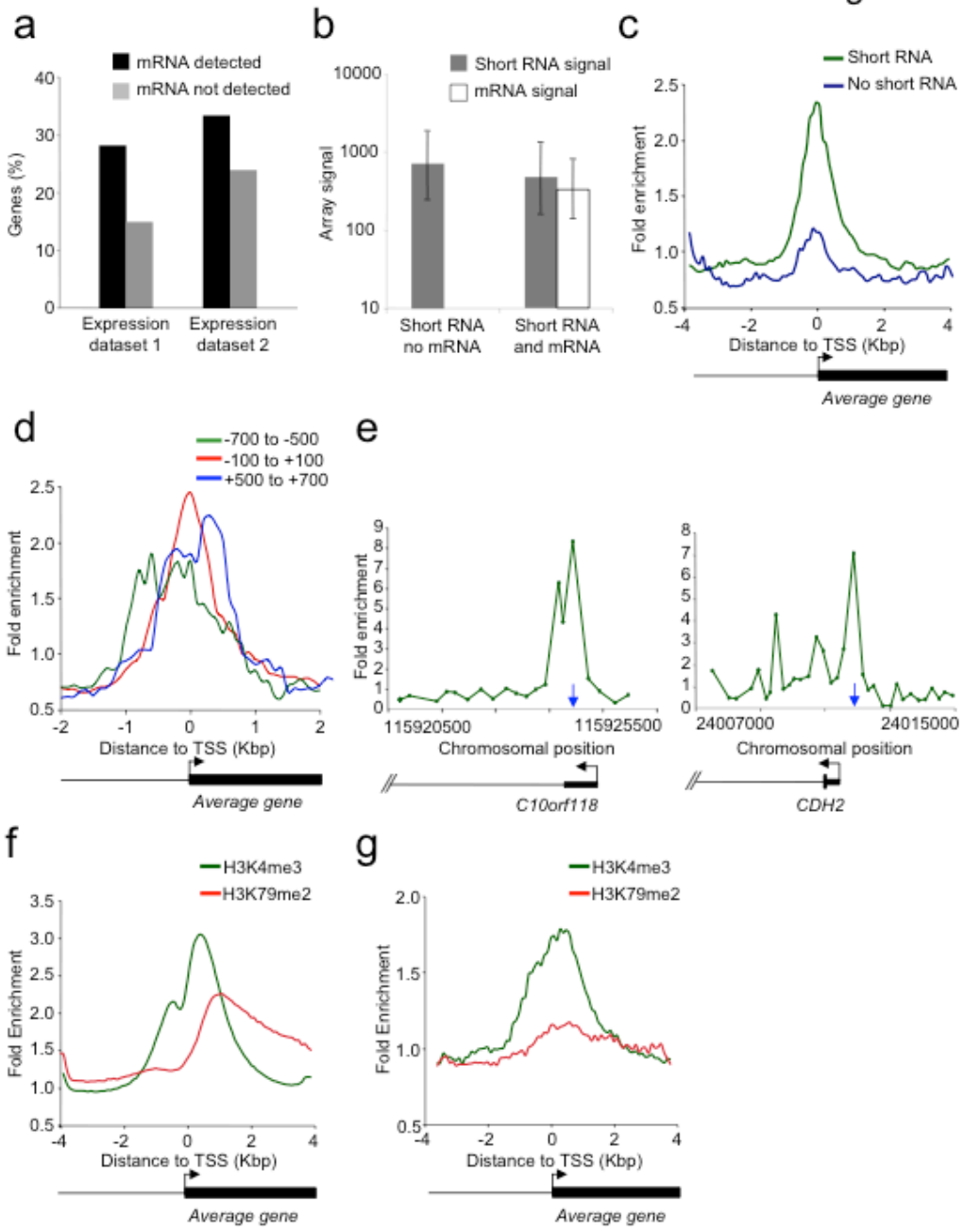

a
mRNA detected
mRNA not detected
Genes (%)
Expression dataset 1
Expression dataset 2
b
Short RNA signal
mRNA signal
Array signal
Short RNA no mRNA
Short RNA and mRNA
c
Short RNA
No short RNA
Fold enrichment
Distance to TSS (Kbp)
Average gene
d
-700 to -500
-100 to +100
+500 to +700
Fold enrichment
Distance to TSS (Kbp)
Average gene
e
Fold enrichment
115920500
115925500
Chromosomal position
C10orf118
24007000
24015000
Chromosomal position
CDH2
f
H3K4me3
H3K79me2
Fold Enrichment
Distance to TSS (Kbp)
Average gene
g
H3K4me3
H3K79me2
Fold Enrichment
Distance to TSS (Kbp)
Average gene

Figure 3

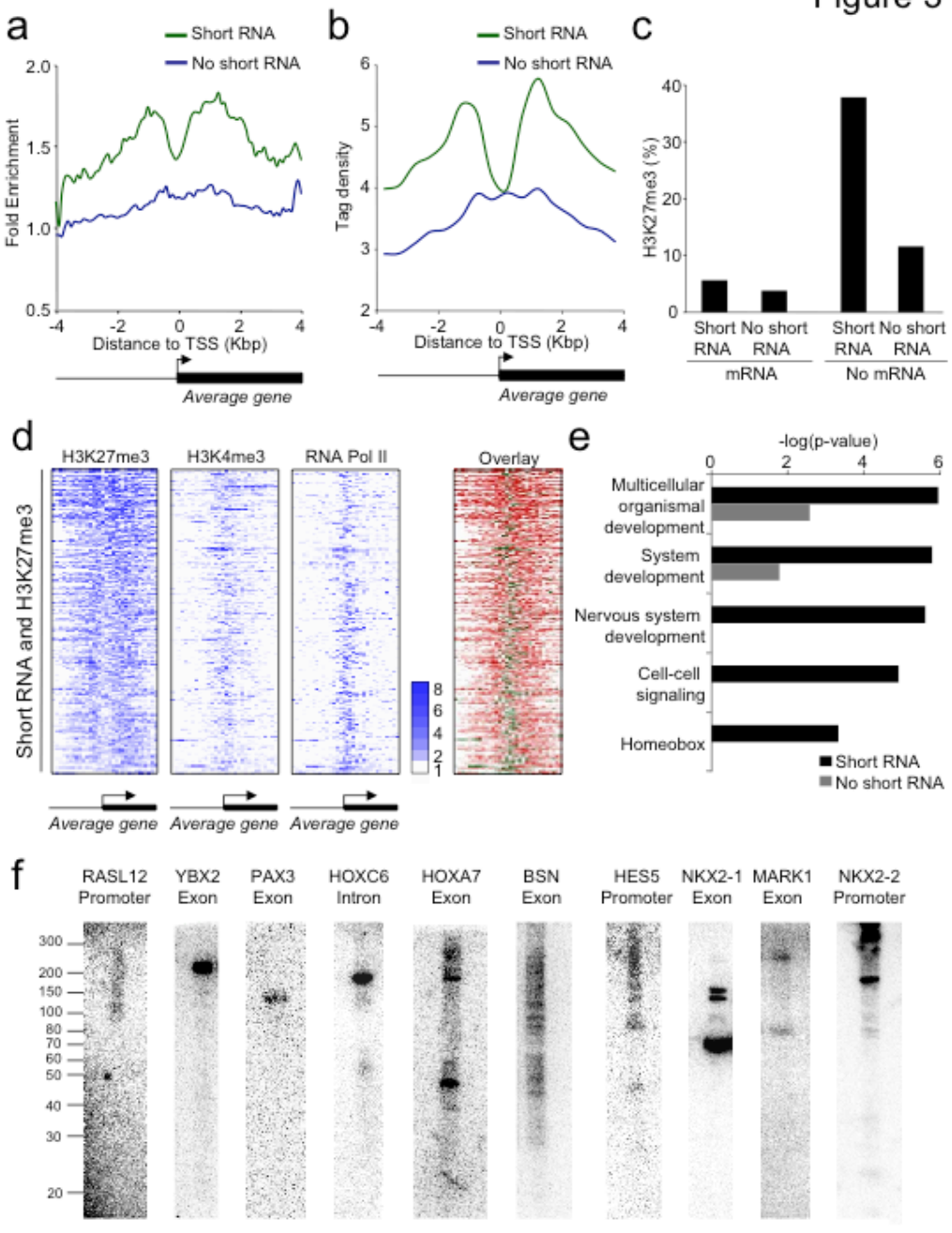
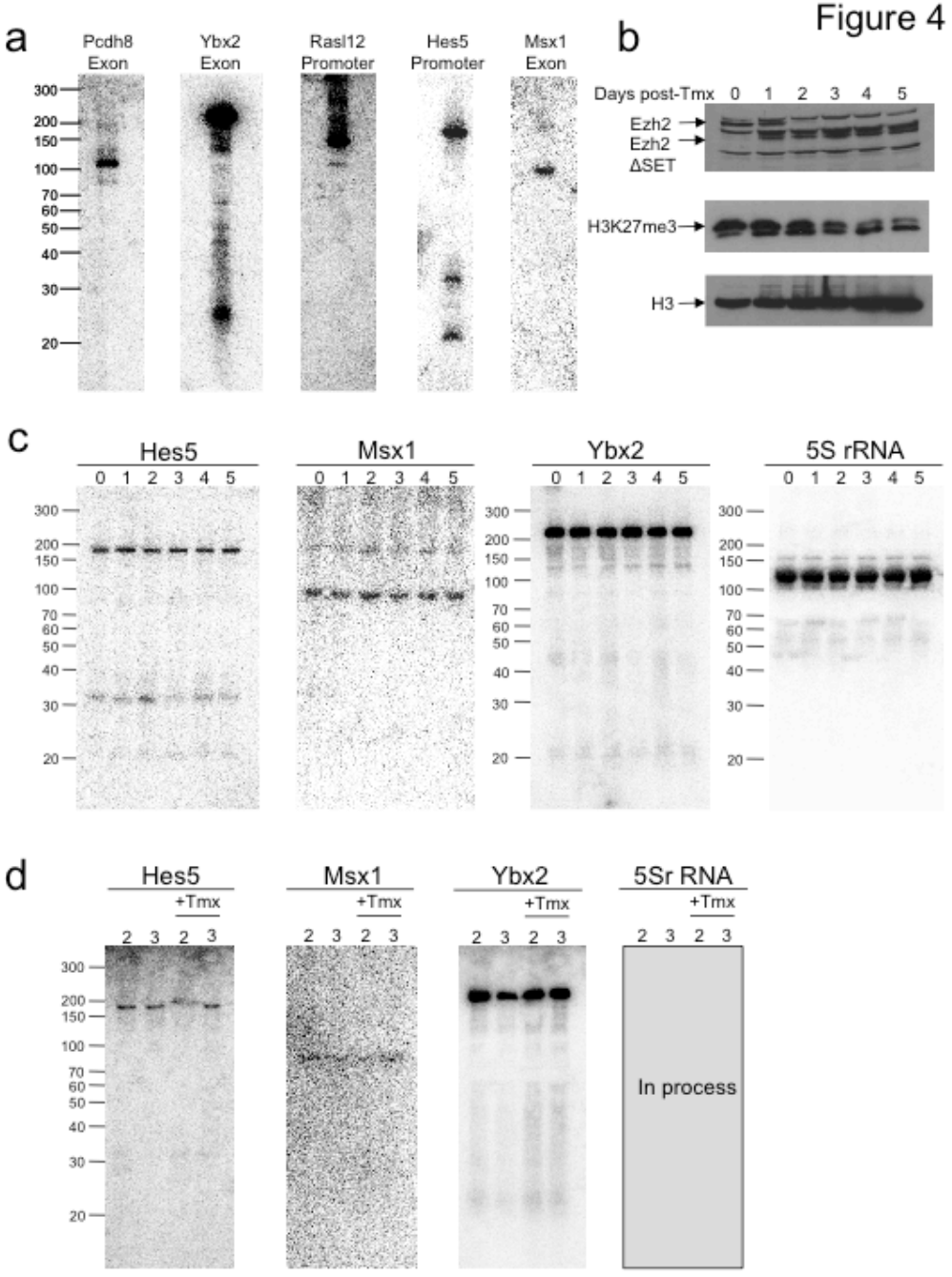

a
Short RNA
No short RNA
Fold Enrichment
Distance to TSS (Kbp)
Average gene
b
Short RNA
No short RNA
Tag density
Distance to TSS (Kbp)
Average gene
c
H3K27me3 (%)
Short RNA
No short RNA
Short RNA
No short RNA
mRNA
No mRNA
d
H3K27me3
H3K4me3
RNA Pol II
Overlay
Short RNA and H3K27me3
Average gene
Average gene
Average gene
e
-log(p-value)
Multicellular organismal development
System development
Nervous system development
Cell-cell signaling
Homeobox
Short RNA
No short RNA
f
RASL12 Promoter
YBX2 Exon
PAX3 Exon
HOXC6 Intron
HOXA7 Exon
BSN Exon
HES5 Promoter
NKX2-1 Exon
MARK1 Exon
NKX2-2 Promoter

Figure 4


a
Pcdh8 Exon
Ybx2 Exon
Rasl12 Promoter
Hes5 Promoter
Msx1 Exon
300
200
150
100
70
60
50
40
30
20
b
Days post-Tmx 0 1 2 3 4 5
Ezh2
Ezh2 ΔSET
H3K27me3
H3
c
Hes5
Msx1
Ybx2
5S rRNA
0 1 2 3 4 5
d
Hes5
Msx1
Ybx2
5Sr RNA
+Tmx
2 3 2 3
In process

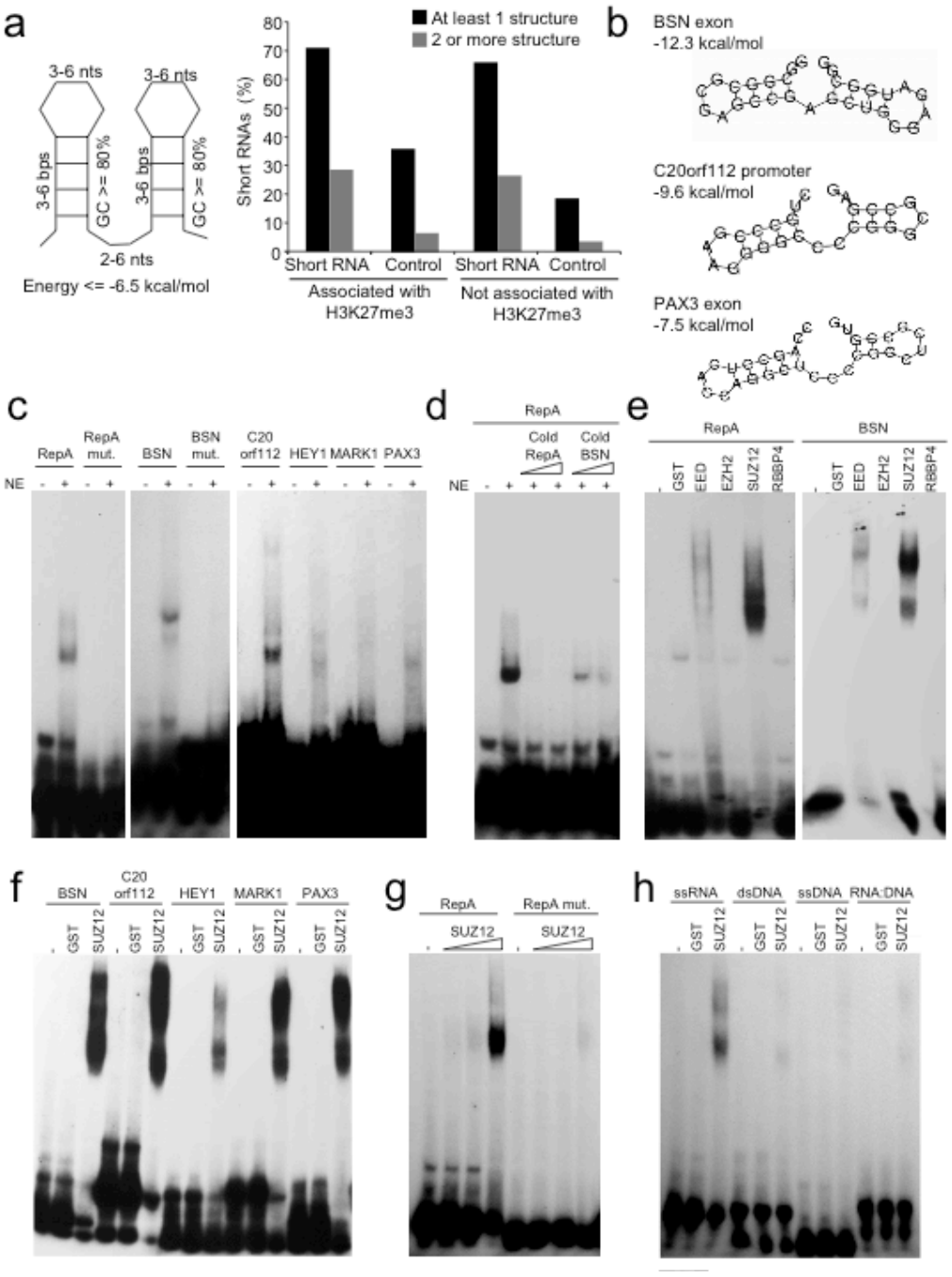

a
3-6 nts
3-6 nts
3-6 bps
GC >= 80%
3-6 bps
GC >= 80%
2-6 nts
Energy <= -6.5 kcal/mol
At least 1 structure
2 or more structure
Short RNAs (%)
Short RNA
Control
Short RNA
Control
Associated with H3K27me3
Not associated with H3K27me3
b
BSN exon
-12.3 kcal/mol
C20orf112 promoter
-9.6 kcal/mol
PAX3 exon
-7.5 kcal/mol
c
RepA
RepA mut.
BSN
BSN mut.
C20 orf112
HEY1
MARK1
PAX3
NE
d
RepA
Cold RepA
Cold BSN
NE
e
RepA
BSN
GST
EED
EZH2
SUZ12
RBBP4
f
BSN
C20 orf112
HEY1
MARK1
PAX3
GST
SUZ12
g
RepA
RepA mut.
SUZ12
h
ssRNA
dsDNA
ssDNA
RNA:DNA
GST
SUZ12

Figure 6

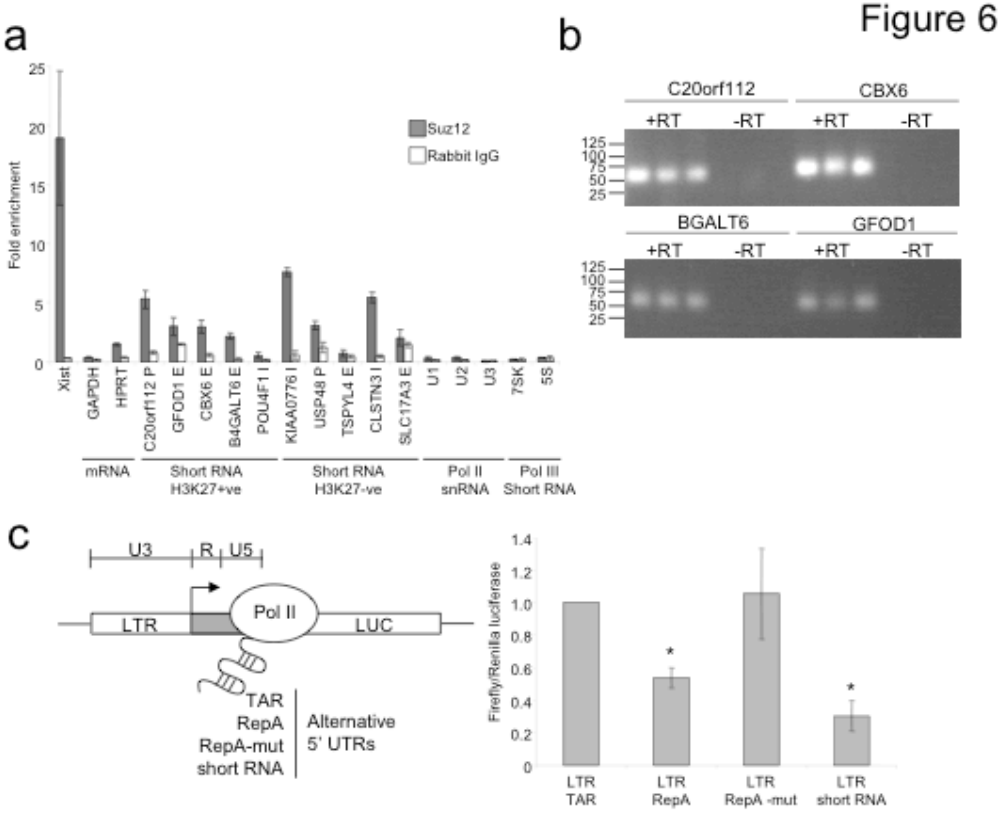
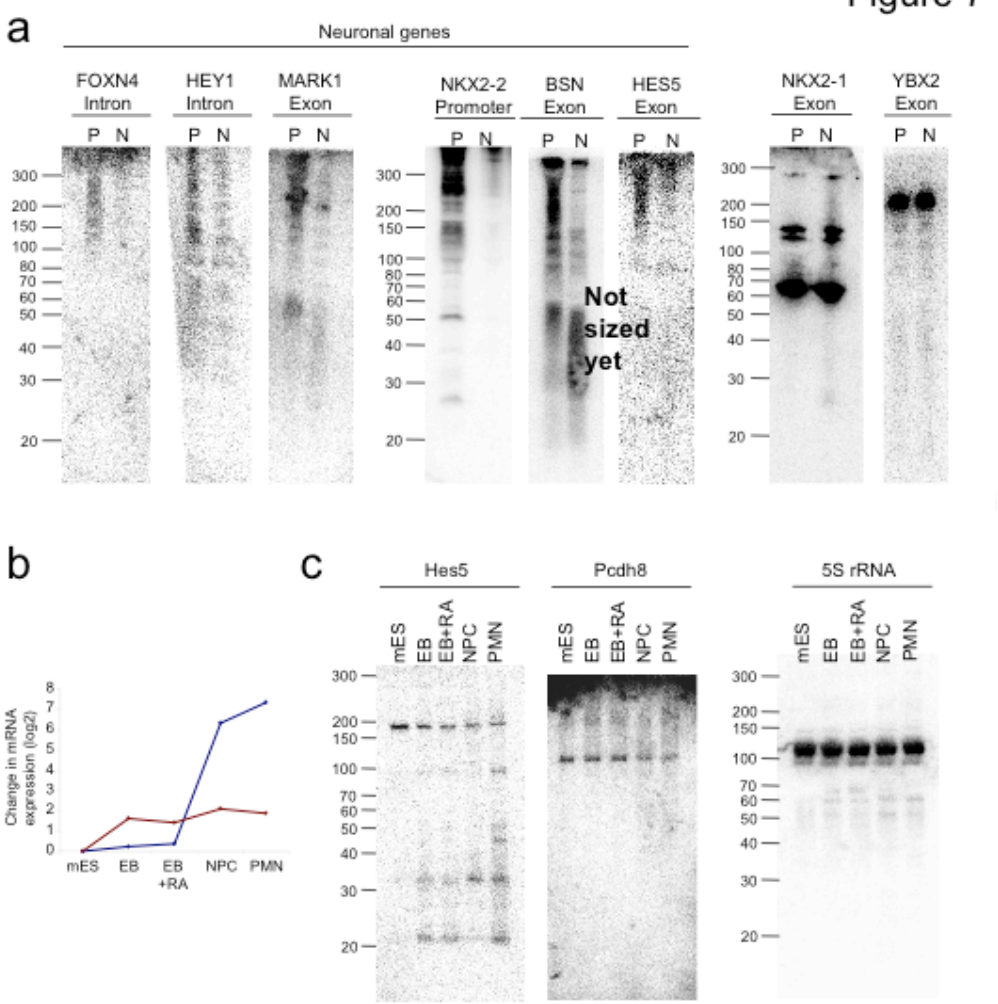

a
b
c
Fold enrichment
Suz12
Rabbit IgG
Xist
GAPDH
HPRT
C20orf112 P
GFOD1 E
CBX6 E
B4GALT6 E
POU4F1 I
KIAA0776 I
USP48 P
TSPYL4 E
CLSTN3 I
SLC17A3 E
U1
U2
U3
7SK
5S
mRNA
Short RNA H3K27+ve
Short RNA H3K27-ve
Pol II snRNA
Pol III Short RNA
C20orf112
CBX6
BGALT6
GFOD1
+RT
-RT
U3
R
U5
LTR
Pol II
LUC
TAR
RepA
RepA-mut
short RNA
Alternative 5' UTRs
Firefly/Renilla luciferase
LTR TAR
LTR RepA
LTR RepA -mut
LTR short RNA

Figure 7


a
Neuronal genes
FOXN4 Intron
HEY1 Intron
MARK1 Exon
NKX2-2 Promoter
BSN Exon
HES5 Exon
NKX2-1 Exon
YBX2 Exon
P N
300
200
150
100
80
70
60
50
40
30
20
Not sized yet
b
Change in mRNA expression (log2)
mES
EB
EB +RA
NPC
PMN
c
Hes5
Pcdh8
5S rRNA
mES
EB
EB+RA
NPC
PMN