Regulation of Gene Expression and Cell State in Embryonic Stem Cells

By

Jamie J. Newman
B.A. Biology
Amherst College, 2003

ARCHIVES

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 May 2010
[June 2010]

Signature of Author.....
Jamie Newman
May 3, 2010

Certified by.....
Dr. Richard A. Young
Professor of Biology
Thesis Supervisor

Accepted by.....
Dr. Steve Bell
Professor of Biology
Chairperson, Biology Graduate Committee

Regulation of Gene Expression and Cell State in Embryonic Stem Cells

By

Jamie J. Newman

Submitted to the Department of Biology on May 17, 2010
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy in Biology

Abstract

Cell state is established and maintained through the combined action of transcription factors, chromatin regulators and signaling pathways, which all contribute to a transcriptional regulatory circuitry. Embryonic stem (ES) cells are capable of self-renewal and can give rise to nearly all differentiated cell-types, making them an ideal system in which to address the challenges of understanding gene expression and cell state. Valuable insights into the control of cell state have been revealed by recent studies of the ES cell transcriptional regulatory circuitry. Here I present work contributing to the understanding of transcriptional regulatory mechanisms that control ES cell state, specifically signaling pathways and proteins that affect chromatin structure.

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

Acknowledgements

There are many people who have played a significant role in my getting to and getting through graduate school. First, I want to thank the people who have been with me through everything from the very beginning, my parents, Frank and Margaruite Newman. Thanks for supporting me and pushing me to keep going when I thought I wanted to give up. I appreciate everything you have done for me even when it didn't seem like I did.
To my sister, Coree, and my brother, Trevor, thanks for always encouraging me. I loved the times when we were on the east coast together and could enjoy Thanksgiving in Providence or watching Trevor row at the Head of the Charles. It hasn't been the same out here without you.

There are many other people who have helped me get to MIT, but I would especially like to thank Dr. Richard Goldsby. Dick, thank you so much for taking a chance on me my first summer at Amherst College and for supporting and guiding me these last ten years. I could never express how much you have influenced my life and I can honestly say that I would not be here if it were not for you.
I would also like to thank Dr. Tyler Jacks. Thank you for introducing me to MIT, high profile science and letting me work with you and your group; it was a truly wonderful experience.

Once at MIT, there were many people who helped me get to this point.
First, I want to thank the girls, Crystal, Kate and Gina. You girls have been there for me at different points in the last five years and I hope you know how much I have appreciated everything you've done, especially helping me laugh through some of the hardest parts.
I want to thank Dave Harris and Dave Shechner for helping me get through the first semester at MIT. Thank you for your patience and for being wonderful teachers.
Megan Cole and Lee Lawton, thank you for making me get out of the lab for ice cream and coffee breaks and for being great friends both in and out of the lab.
To my fiancé Brad, thank you for sticking out the long distance with me and supporting me in so many ways while I finished my degree.

Thank you to everyone in the Young lab for helping me to become a better scientist. I would especially like to thank everyone who helped put together the stories that are presented in this document.
Finally, thank you to my thesis advisor, Dr. Richard Young. I have learned so much from you about the different aspects of being a scientist. I appreciate your patience with me as I have learned to be a better writer, presenter and all around scientist. Thank you for supporting me these last few years.

Table of Contents

Chapter 1

Introduction

Regulation of Gene Expression and Cell State

Abstract

Regulation of gene expression in eukaryotes occurs in the context of a confined nuclear structure where DNA is tightly wrapped into nucleosomes and packaged into higher order chromatin. DNA-binding regulators must therefore work together with chromatin regulators to generate a gene expression program that is specific to each cell state. The gene expression program of each cell is influenced by the extracellular environment through signaling pathways that can connect directly to DNA-binding and chromatin regulators. Here I describe the key concepts that emerged from early studies of gene regulation and chromatin, discuss the use of embryonic stem (ES) cells as a model in vitro system to study control of cell state, and review our understanding of ES cell transcriptional regulatory circuitry. I highlight recent work that reveals how signaling pathways connect to the key DNA binding regulators of ES cells. I also highlight recent studies that have led to the model that Mediator and Cohesin physically and functionally connect the enhancers and core promoters of a key subset of active genes in ES cells, thus generating cell-type specific chromatin structure.

Key Concepts in Gene Regulation from Prokaryotes and Eukaryotes

The DNA fiber of the human genome is approximately two meters in length. The genome is packaged into a highly compacted form in order to fit inside the nucleus of a human cell with a diameter of only 10 μm (Mohd-Sarip and Verrijzer, 2004). Genome compaction also occurs in bacteria; the *E. coli* genome measures 2 mm in length and must fit into a space of only 0.5 μm^3. Accurate transcriptional regulation must occur in the context of these compacted genomes. In fact, evidence from early studies of gene regulation argues that gene regulators are involved in generating DNA loops that apparently contribute to genome compaction (Finch and Klug, 1976; Olins and Olins, 1974; Ris and Kubai, 1970).

Gene regulation by Transcription Factors

DNA-binding transcriptional regulators are key to specific gene control. Early studies into the transcriptional control of the *E. coli lac* operon created a framework for understanding gene control in all of biology (Jacob and Monod, 1961). In the absence of lactose, the *lac* operon is repressed by the Lac repressor, a DNA binding protein that binds specifically to a DNA element just downstream of the transcription start site called the *lac* operator. When the repressor is bound to the *lac* operator, transcription initiation by RNA polymerase is inhibited. If lactose is present, it is metabolized into an isomer, allolactose, which binds the *lac* repressor protein and alters its conformation, thus preventing it from binding to the *lac* operator, and allowing transcription to occur. A second control mechanism, which involves sensing of the nutrient environment and a second DNA binding transcriptional regulator, can contribute to further activation of the *lac* operon in the absence of glucose. Cyclic adenosine monophosphate (cAMP) is a signaling molecule whose prevalence is inversely proportional to that of glucose. It binds to Catabolite Activator Protein (CAP), which undergoes a conformational change that allows it to bind a DNA element called the CAP binding site, located upstream of the transcription start site, from which it recruits RNA polymerase, thus activating transcription. Specific DNA-binding repressors and activators, and the specific sequence elements they recognize in the genome, are the fundamental components of gene control in all of biology.

DNA-binding transcription factors make up the largest single class of proteins encoded in the human genome, representing approximately 10% of all protein-coding genes (Babu et al., 2004; Lander et al., 2001; Levine and Tjian, 2003). Most DNA-binding transcription factors that have been well-studied appear to function as transcriptional activators, but can be bound by other proteins that cause the multiprotein complex to act as a repressor. Transcription factors bind to both promoter-proximal DNA elements and to distal regions 1-100 kb away from the promoter (D'Alessio et al., 2009; Narlikar and Ovcharenko, 2009; Pan et al., 2010). The distal elements that are involved in positive gene regulation are generally called enhancers, and these elements are typically bound by multiple transcription factors. The best characterized of these enhancers is that of the INF-β promoter, where eight transcription factor molecules

occupy a 50 bp segment of DNA (Agalioti et al., 2000; Maniatis et al., 1998; Panne, 2008). Additional DNA elements exist that can prevent the activity of enhancers at nearby genes; elements called insulators fall into this class.

The activity of transcription factors in eukaryotes can be affected by the environment surrounding the cells, just as it is in bacteria. For example, metabolic ligands can bind specific transcription factors and alter their activity. In yeast, Leu3 can only initiate transcription of the amino acid biosynthesis genes it regulates after interacting with a leucine metabolic precursor which changes Leu3 from a repressor to an activator (Kirkpatrick and Schimmel, 1995). In addition, there are complex signaling pathways that employ both protein kinases and transcription factors to bring signals from the extracellular environment to specific genes throughout the genome (Browning and Busby, 2004; Martinez-Antonio and Collado-Vides, 2003). The gene expression program of any one cell is thus dependent on both the population of transcription factors expressed in the cell and the environment in which the cell resides.

Nucleosomes and Gene Regulation

Nucleosomes represent the fundamental unit of chromatin and are made up of an octamer of four core histone proteins around which 147 bp of DNA are wrapped (Davey et al., 2002; Finch and Klug, 1976; Kornberg and Klug, 1981; Luger et al., 1997; Olins and Olins, 1974; Ris and Kubai, 1970). Nucleosomes compact the mammalian genome by roughly 100,000-fold (Goetze et al., 2007). Genomic regions that are densely populated by nucleosomes and highly compacted are generally more transcriptionally silent than regions that have lower nucleosome density. Nucleosomes are distributed throughout the genome, but are depleted from active promoter regions (Bernstein et al., 2004; Gilbert et al., 2004; Jiang and Pugh, 2009; Mavrich et al., 2008; Schones et al., 2008; Segal et al., 2006; Tirosh and Barkai, 2008; Yuan et al., 2005). The presence of transcription factors and the transcription initiation apparatus at active promoters is thought to influence the local density and positioning of nucleosomes.

Two classes of nucleosome regulators have been described. ATP-dependent chromatin remodeling complexes are able to mobilize nucleosomes, which can enhance or reduce access to DNA sequences by transcription factors with consequent effects on

gene activity (de la Serna et al., 2006a; de la Serna et al., 2006b; Ho and Crabtree, 2010; Kingston and Narlikar, 1999; Narlikar et al., 2002; Saladi and de la Serna, 2010; Sif, 2004; Tsukiyama et al., 1995; Tsukiyama and Wu, 1995). There are also a large number of histone modifying enzymes whose activities contribute to local gene activity (Kouzarides, 2007). Transcription factors and components of the transcription apparatus can bind and recruit ATP-dependent chromatin remodeling complexes and histone modifying enzymes to specific sites in the genome (Cairns, 2009; Panne, 2008; Roeder, 2005; Segal and Widom, 2009). In this manner, most nucleosome regulators, which have no sequence-specific binding properties of their own, are brought to specific sites to facilitate gene activity or repression.

The histone modifying enzymes contribute to gene control by chemically modifying lysine, arginine, serine and other residues in the N-terminal "tails" of histone proteins, which then form binding sites for other proteins that contribute to positive or negative regulation of gene expression. For example, histone acetyltransferases (HATs) can acetylate specific lysine residues, which form sites for binding by regulatory proteins that contain bromodomains (Kouzarides, 2000; Phillips, 1963; Roth et al., 2001; Yang and Seto, 2007). Histone methyltransferases (HMTs) can methylate specific lysine and arginine residues, forming sites for binding by regulatory proteins that contain chromodomains (Berger, 2002; Gerber and Shilatifard, 2003; Trievel, 2004; Yeates, 2002). There are also enzymes that can remove these modifications from histones, thus having the opposite effect on gene control (Agger et al., 2008; Hassig and Schreiber, 1997; Kim et al., 2009; Thiagalingam et al., 2003; Verdin et al., 2003).

DNA Looping and Gene Regulation

In addition to DNA packaging at the level of the nucleosome, DNA looping contributes to further chromatin compaction (Finch and Klug, 1976; Olins and Olins, 1974; Ris and Kubai, 1970). Recent studies have identified a variety of looping interactions in the genomes of eukaryotes (Dostie et al., 2006; Hadjur et al., 2009; Kurukuti et al., 2006; Levasseur et al., 2008; Lieberman-Aiden et al., 2009; Nativio et al., 2009; Spilianakis and Flavell, 2004; Tolhuis et al., 2002; Vakoc et al., 2005), but the mechanisms involved in formation of loops are poorly understood. Some of this DNA looping is under the

control of transcription factors and the transcription apparatus at active genes and some involves proteins that bind insulators involved in gene repression.

Studies in bacteria first provided evidence that DNA looping is a key feature of transcriptional activation. Transcription factors called enhancer binding proteins (EBPs), which bind to the enhancer element, can interact with the σ^{54} RNA polymerase, which is bound to the transcription start site, thus creating a defined DNA loop (Popham et al., 1989; Sasse-Dwight and Gralla, 1988). The σ^{54} holoenzyme cannot actively transcribe genes until the enhancer elements are occupied and the enhancer-bound proteins physically interact with σ^{54}. These prokaryotic enhancers are typically located 70-150 bp from the promoter but can also act from distances as great as several kilobases. For example, the nitrogen regulatory protein C (NtrC) transcription factor creates a loop between its binding sites and the σ^{54}-RNA polymerase -occupied glnA promoter (Reitzer and Magasanik, 1986; Rippe et al., 1997; Su et al., 1990). The NtrC binding sites are located approximately 100 bp from the glnA promoter, but NtrC can bind enhancer elements and activate gene transcription from as far as 3 kb from the promoter. Thus, specific DNA loop formation between enhancers and core promoter sites is a consequence of the mechanism of transcriptional regulation and has been shown to be critical for activation of at least some genes transcribed by σ^{54}-RNA polymerase. (Figure 1A)

Studies with bacterial systems also first established that proteins dedicated to DNA bending play important accessory roles in DNA looping and gene activity. Bacterial proteins with DNA bending functions include histone-like nucleoid structure proteins (H-NS), integration host factors (IHF) and factors for inversion stimulation (Fis) (Luijsterburg et al., 2008). Each of these proteins can create a significant bend in the DNA, which range from a 90° bend created by the Fis proteins (Pan et al., 1996) to a 160° bend mediated by H-NS and IHF (Dame et al., 2000; Dorman et al., 1999; Ellenberger and Landy, 1997; Gerstel et al., 2003; Goodrich et al., 1990; Hochschild and Ptashne, 1988; Huo et al., 1988; Luijsterburg et al., 2006; Popham et al., 1989; Xu and Hoover, 2001). Some of these DNA bending proteins, such as IHF, have been shown to facilitate loop formation and gene activity at promoters where enhancer binding proteins interact with σ^{54}-RNA polymerase (Claverie-Martin and Magasanik,

1991; Hoover et al., 1990). In summary, the study of gene expression in these bacterial systems led to the concept that transcription factors bound at enhancers bind to RNA polymerase at the transcription start site, thus forming a specific DNA loop, and that this loop is facilitated and stabilized by the binding of DNA bending factors, which thereby contribute to gene expression. (Figure 1A)

In eukaryotes, enhancer elements are typically located at substantial distances from the core promoter where the transcription initiation apparatus is bound (Banerji et al., 1981; Benoist and Chambon, 1981; Levine and Tjian, 2003; Maniatis et al., 1987; Visel et al., 2009; Wasylyk et al., 1983). Transcription factors bind cofactors that can bind RNA polymerase II and are thus thought to bridge the enhancers to which they are bound and the transcriptional machinery located at the promoter (Fuda et al., 2009; Hampsey and Reinberg, 1999; McKenna and O'Malley, 2002; Naar et al., 2001; Spiegelman and Heinrich, 2004; Thomas and Chiang, 2006). Among the cofactors are Mediator and p300, which have been shown to bind both transcription factors and the transcription initiation apparatus (Conaway et al., 2005; Kornberg, 2005; Malik and Roeder, 2005, 2008; Roeder, 1998; Taatjes and Tjian, 2004; Visel et al., 2009). It thus seems likely that specific DNA loop formation is a natural consequence of the mechanism of transcriptional activation in eukaryotes. (Figure 1B)

Like prokaryotes, eukaryotes posses a family of proteins capable of bending DNA and facilitating interactions between proteins that are bound at distant DNA sites. This class of proteins includes the high mobility group box (HMGB) proteins. HMG box domain-containing proteins bend DNA 80-130° depending on the number of HMG box proteins present (Bustin and Reeves, 1996; Thomas and Travers, 2001). HMG box proteins do not recognize specific DNA sequences, but by altering DNA structure, these proteins aid in transcription either by stabilizing protein interactions or by promoting the recruitment of other proteins to a site of transcriptional activity (Bianchi and Agresti, 2005; Thomas and Chiang, 2006). Sox2, a key transcription factor in the transcriptional regulatory circuitry of embryonic stem cells, is a member of the HMG box domain protein family and functions in ES cells by forming a heterodimer with the POU-domain protein family member Oct4 (Pevny and Lovell-Badge, 1997; Wegner, 1999). This

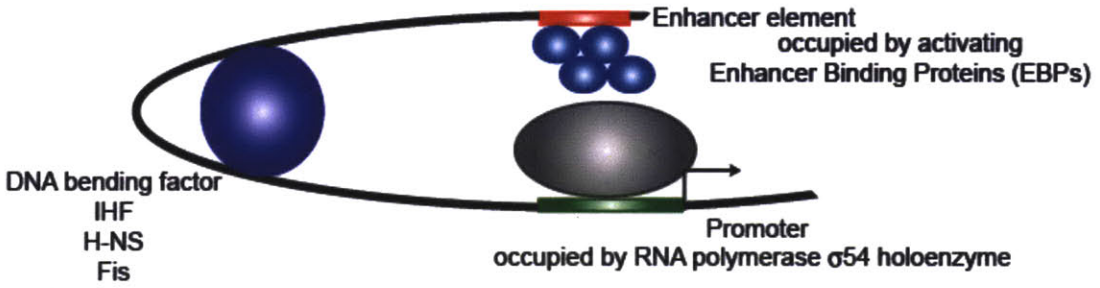
makes it tempting to speculate that Sox2 is essential for ES cell state both because of its interaction with Oct4 and its physical impact on DNA structure.

DNA loop formation has recently been implicated in insulator activity. The insulator binding protein CTCF (CCCTC-binding factor) is thought to block the interaction between enhancer-bound proteins and the transcriptional machinery to prevent inappropriate expression as well as the spread of repressive chromatin (Ohlsson et al.; Ohlsson et al.; Phillips and Corces, 2009). CTCF can dimerize, thus bringing together two CTCF-bound DNA sites. It has been implicated in the control of gene expression through knockout studies and may play a role in certain cancers (Dunn and Davie, 2003; Filippova et al., 2002; Kim et al., 2007). Whole genome studies have revealed that there are nearly 20,000 CTCF-bound sites across the genome of any given cell-type. Most CTCF binding sites are not cell-type specific, although the contribution of CTCF to the control of gene expression is specific (Barski et al., 2007; Kim et al., 2007; Xie et al., 2007). Several examples of CTCF's involvement in DNA looping at cell-type specific loci have recently been identified using chromosome conformation capture (3C) (Kurukuti et al., 2006; Lewis and Murrell, 2004; Li et al., 2008; Majumder et al., 2008; Splinter et al., 2006). The evidence for CTCF's involvement in cell-type specific transcriptional activity suggests it is interacting with other cell-type specific factors. The other factors that may be contributing to these long-range looping interactions remain unknown.

In summary, accurate transcriptional regulation must occur in the context of highly compacted genomes. In bacteria, transcription factors bound to regulatory DNA elements also bind to the transcription apparatus at the core promoter, thus forming a DNA loop, and this loop can be facilitated or stabilized by DNA bending proteins. In eukaryotes, it seems likely that a similar process takes place, although in the context of nucleosomal DNA. Transcription factors bound to regulatory DNA elements recruit chromatin regulators that mobilize and modify nucleosomes, thus creating a chromatin environment favorable to further regulation. The transcription factors also bind to cofactors, which in turn bind to the transcription apparatus at the core promoter, thus forming a DNA loop. We imagine that a variety of proteins may aid in the formation or stability of this loop, and thus contribute to gene regulation in eukaryotes.

Figure 1

A. Features of prokaryotic transcription initiation



B. Features of eukaryotic transcription initiation

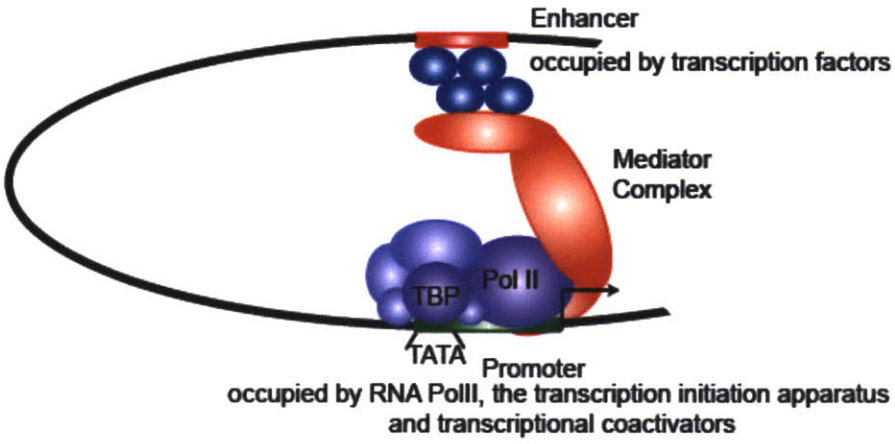


Figure 1. Features of prokaryotic and eukaryotic transcription initiation.

(A). Prokaryotic transcription requires DNA sequence-specific binding elements called enhancer binding proteins (EBPs) that occupy the enhancer elements located upstream of the promoter and transcriptional start site. RNA polymerase σ^{54} holoenzyme occupies the promoter and is activated upon interaction with EBPs. A DNA bending factor is often involved in facilitating this interaction. The integration host factor (IHF), histone-like nucleoid structure (H-NS) proteins and factors for inversion stimulation (Fis) are examples of proteins that bend the DNA. These proteins are responsible for altering the DNA in order to bring enhancer-bound proteins in proximity to the polymerase occupied promoter.

(B). In many ways, eukaryotic transcription resembles prokaryotic transcription. The size of eukaryotic genomes and the number of cell-type specific expression profiles required by multicellular organisms require more complex mechanisms to regulate gene expression. RNA polymerase II occupies the promoter along with the transcriptional apparatus and a number of cofactors including TATA binding protein (TBP). Upstream of the promoter are sequence-specific enhancer elements. These elements can be anywhere from 1-100 kb from the promoter and are commonly occupied by cell-type specific transcription factors. In order to activate transcription of the genes they regulate, transcription factors must interact with the transcriptional machinery at the promoter. Mediator is a multisubunit complex that helps to bridge this interaction by interacting at one end with enhancer-bound transcription factors and at the other with RNA polymerase II and the transcription apparatus.

ES Cells as a Model System to Study Transcriptional Control of Cell State

Embryonic stem cells posses the unique ability to self-renew, propagating almost indefinitely in culture under the appropriate conditions. ES cells are also pluripotent, capable of giving rise to any of the over 200 fully differentiated cell-types found in adult mammals (Pera and Trounson, 2004; Rossant, 2008; Silva and Smith, 2008; Wobus and Boheler, 2005; Yamanaka, 2008) (Figure 2). Before the isolation and culture of embryonic stem cells, cells derived from teratocarcinomas (P9 and F10) and embryonic carcinomas (EC) were used to study self-renewal and differentiation (Kahan and Ephrussi, 1970; Martin, 1980; Pardal et al., 2003; Stevens, 1978; Stevens et al., 1977). Cancer cells resemble ES cells in that they have the ability to self-renew and differentiate into a number of cells types, but differ in that they are not generally pluripotent and directed differentiation is not as cleanly controlled. The mutations that immortalize cancer cells also make cancer cell lines difficult in vitro systems from which to derive a reliable understanding of development (Chambers and Smith, 2004; Downing and Battey, 2004). Embryonic stem cells, on the other hand, provide a primary, pluripotent, self-renewing cell line that can be used to investigate many of the questions surrounding early mammalian development.

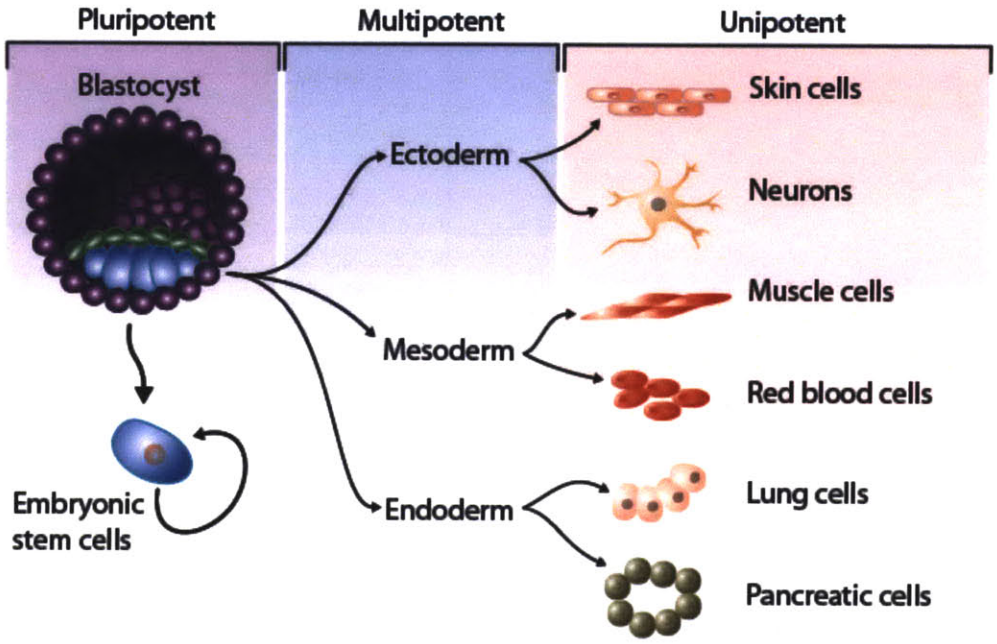
Waddington's model of the epigenetic landscape provides a useful concept to describe the progression of ES cells from their undifferentiated state to a more defined terminal cell state (Waddington, 1957) (Figure 2). A pluripotent stem cell is positioned at the top of the landscape, poised to begin traveling towards a differentiated cell-type. As the cell adopts a new cell identity, it loses those properties of pluripotency. The cell first passes through a stage of multipotency where it is committed to one of three cell lineages - endoderm, mesoderm or ectoderm - but maintains the ability to become any cell-type within that lineage. Ultimately, the cell reaches a state of unipotency where its identity is maintained as a fully differentiated cell. The study of embryonic stem cells has led to the identification of factors that promote self-renewal, pluripotency or direct cellular differentiation to any number of multipotent progenitor cells or fully differentiated cells. Several cell states can now be reproduced in culture, including beating cardiomyocytes, muscle cells, and neurons (D'Amour et al., 2006; Graichen et al., 2008;

Joannides et al., 2007; Kubo et al., 2004; Laflamme et al., 2007; Murry and Keller, 2008; Ng et al., 2005; Nostro et al., 2008; Wichterle et al., 2002; Ying and Smith, 2003).

The progression of normal cellular differentiation was long thought to be irreversible, progressing only in the direction of pluripotency to unipotency, but studies in the last 13 years have shown that this is not necessarily the case. Nuclear transplantation and the generation of Dolly the sheep provided powerful evidence that cell fate could be reversed (Wilmut et al., 1997). Dolly was a clone, created by the use of somatic cell nuclear transfer (SCNT). SCNT involves the introduction of a somatic nucleus into an enucleated oocyte. The genome contained in that somatic nucleus is reprogrammed by factors within the oocyte to resemble the gene expression program of an embryonic stem cell, which can give rise to all cell-types required for an adult organism (Gurdon and Byrne, 2003; McGrath and Solter, 1983). Cellular reprogramming techniques have since improved in efficiency and safety, providing greater value to the potential application of reprogrammed cells in areas of biomedical research.

To further our understanding of development, and to facilitate discoveries that advance the new field of regenerative medicine, it is important to understand the regulatory mechanisms that control pluripotency, self-renewal and differentiation. To gain this understanding, it is useful to discover how transcription factors, signaling pathways and chromatin regulators contribute to control of the gene expression program that is necessary to maintain ES cell state.

Figure 2



Embryonic stem cells are derived during the blastocyst stage of embryogenesis.

At this stage cells are pluripotent and can proliferate this way in culture indefinitely under the right conditions. As ES cells differentiate during development they are fated for particular lineages and take on a multipotent cell state. Within each lineage a cell can continue to differentiate into any of the over 200 defined cell-types of an adult mammal. Examples of these fully differentiated, unipotent cell states are depicted here.

Embryonic Stem Cell Transcriptional Regulatory Circuitry

Transcription factors, signaling pathways, chromatin regulators and noncoding RNAs play key roles in establishing and maintaining cell-type specific gene expression programs. Recent studies of these regulators have led to a model of the transcriptional regulatory network for embryonic stem cells and concepts that may provide a foundation for further understanding the control of cell state (Boyer et al., 2005; Boyer et al., 2006; Cole et al., 2008; Lee et al., 2006; Loh et al., 2006; Marson et al., 2008b). A version of this model is shown in Figure 3 and its key features are discussed below.

Transcription Factors

Oct4, Sox2 and Nanog are three core transcription factors required for determining and maintaining the state of embryonic stem cells. The critical role of the transcription factors Oct4 and Nanog was initially established based on their unique expression in ES cells and the impact that their loss of expression had on cell state (Chambers et al., 2003; Chambers and Smith, 2004; Hart et al., 2004; Mitsui et al., 2003; Nichols et al., 1998; Scholer et al., 1990). The high mobility group protein, Sox2, was added to the list of key regulators of ES cell state when it was discovered that Sox2 forms a heterodimer with Oct4 (Ambrosetti et al., 1997; Ambrosetti et al., 2000; Avilion et al., 2003; Masui et al., 2007).

The study of genes occupied and regulated by Oct4, Sox2 and Nanog has led to three key concepts in transcriptional control of ES cell state. First, the master regulators collaborate to regulate each of their own promoters, forming an interconnected autoregulatory loop (Boyer et al., 2005) (Fiugre 3A). This regulatory feature probably contributes to the ability to jump-start the ES cell gene expression program during reprogramming by forced expression of exogenous reprogramming factors (Jaenisch and Young, 2008). It is also thought to generate a bi-stable state for ES cells; a stable, positive feedback controlled gene expression program when the master transcription factors are adequately expressed, and a stimulus to enter a differentiation program when any one of the master transcription factors is no longer fully functional. A second concept is that the master regulators collaborate to regulate their target genes because

all three factors are found at these genes (Boyer et al., 2005). A third concept is that they function to activate expression of genes necessary to maintain ES cell state, while contributing to silencing of genes whose repression is essential to maintaining that state. These silenced genes encode developmental regulators whose expression is required for the appropriate differentiation of ES cells, and whose repression is critical for maintaining the self-renewing and pluripotent properties of ES cells. (Boyer et al., 2005; Lee et al., 2006; Loh et al., 2006) (Figure 3B)

Additional transcription factors contribute to the transcriptional control of ES cells. Sall4 and Tcf3 have recently been shown to occupy most of the same genes bound by Oct4, Sox2 and Nanog (Chen et al., 2008a; Cole et al., 2008; Wu et al., 2006; Zhang et al., 2006). The transcription factors c-Myc, Esrrb and Trim28 are also important for control of proliferation and maintenance of ES cell state (Chen et al., 2008b; Hu et al., 2009; van den Berg et al., 2008; Zhang et al., 2008). It is likely that there will be additional transcription factors among the hundreds that are expressed in ES cells that will emerge as making important contributions to the larger gene expression program of these cells.

Some transcription factors have the ability to reprogram cell fates. Weintraub and colleagues (Davis et al., 1987) first showed that a single transcription factor (MyoD) can reprogram fibroblasts into muscle-like cells. Takahashi and Yamanaka (2006) more recently showed that the forced expression of Oct4, Sox2, Klf4 and c-myc can reprogram fibroblasts into induced pluripotent stem (iPS) cells. Many groups have reproduced Yamanaka's findings and have demonstrated reprogramming to iPS cells with Oct4 and other ES cell transcription factors in a variety of cell-types (Maherali and Hochedlinger, 2008; Maherali et al., 2007; Park et al., 2008a; Park et al., 2008b; Silva et al., 2008; Wernig et al., 2007; Wernig et al., 2008). Reprogramming is now being used to generate patient-specific iPS cells. These patient-specific iPS cells have the ability to differentiate into any number of desired cell-types, which hold the potential to treat a number of diseases and eliminate the complications caused by the introduction of foreign donor cells to affected individuals (Dimos et al., 2008; Ebert et al., 2009; Kiskinis and Eggan; Soldner et al., 2009; Trounson, 2009). Further understanding of ES cell transcriptional regulatory mechanisms and circuitry should improve our knowledge of

the underlying mechanisms that control cell state and will almost certainly continue to improve reprogramming methods.

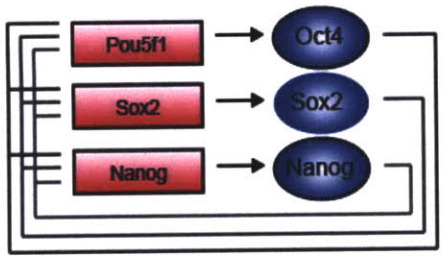
Chromatin Regulators

Polycomb group (PcG) proteins and SetDB1 are among the chromatin regulators that play roles in the maintenance of embryonic cell states. Genetic studies established that PcG genes are key regulators of early development in metazoans (Breiling et al., 2007; Faust et al., 1998; O'Carroll et al., 2001; Pasini et al., 2004). Later studies revealed that PcG proteins repress gene expression, in part by methylating and ubiquitylating histone tails (Orlando, 2003; Pirrotta, 1998; Ringrose and Paro, 2004; Schuettengruber et al., 2007; Schwartz and Pirrotta, 2007). In ES cells, PcG proteins occupy and silence genes encoding developmental regulators (Bernstein et al., 2006; Boyer et al., 2006; Lee et al., 2006; Pan et al., 2007; Yeap et al., 2009; Zhao et al., 2007).

SetDB1 knockdown causes loss of ES cell state (Bilodeau et al., 2009; Lohmann et al., 2010; Yuan et al., 2009). SetDB1 is among the histone H3 lysine 9 methytransferases, which have roles in gene repression (Ayyanathan et al., 2003; Schultz et al., 2002; Wang et al., 2003). SetDB1 occupies and methylates nucleosomes at many of the same silent developmental genes that are occupied by PcG proteins (Bilodeau et al., 2009; Yuan et al., 2009). It thus appears that both chromatin regulators contribute to repression of this key set of developmental regulators, whose expression leads to loss of ES cell state (Figure 3B).

Figure 3

A.


Pou5f1
Sox2
Nanog
Oct4
Sox2
Nanog

B.

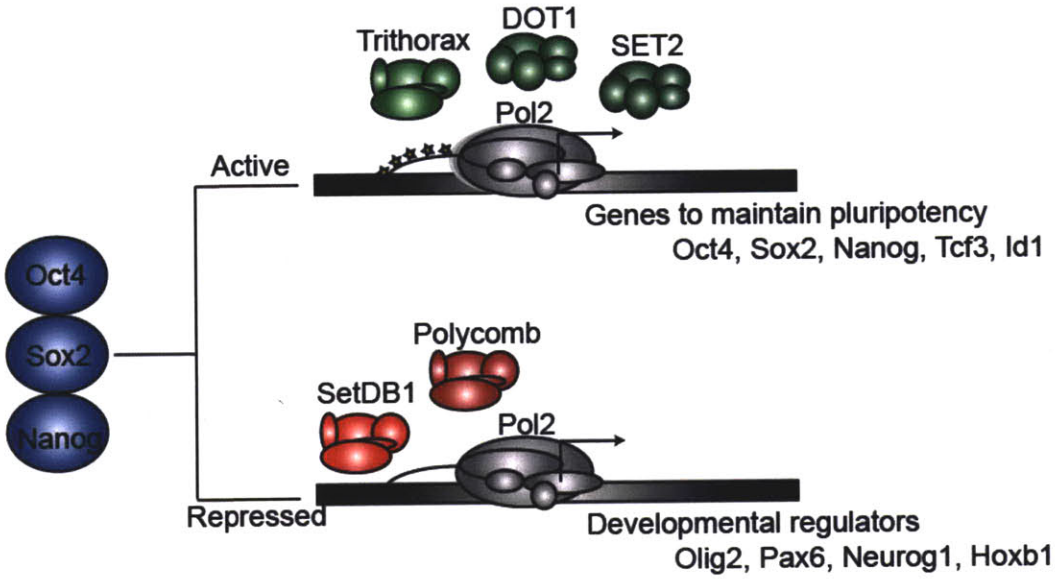

Trithorax
DOT1
SET2
Pol2
Active
Genes to maintain pluripotency
Oct4, Sox2, Nanog, Tcf3, Id1
Oct4
Sox2
Nanog
Polycomb
SetDB1
Pol2
Repressed
Developmental regulators
Olig2, Pax6, Neurog1, Hoxb1

Figure 3. Transcription factors and chromatin modifiers determine gene expression.

(A). Oct4, Sox2 and Nanog are part of an interconnected autoregulatory loop. Oct4, Sox2 and Nanog each regulate their own expression as well as the expression of the other two proteins.

(B). Oct4, Sox2 and Nanog are key transcription factors in determining embryonic stem cell state. Studies have demonstrated a pairing of these transcription factors with marks of activation including an elongating form of RNA polymerase II (Pol2), the chromatin modifying complexes trithorax, the enzyme complex responsible for H3K4me3, Dot1, the enzyme responsible for H3K36me3 and Set2, the enzyme responsible for H3K79me2. The genes occupied by this combination of factors encode proteins required for the maintenance of ES cell state including key ES cell-specific transcription factors and inhibitors of differentiation (Id1). Oct4, Sox2 and Nanog have also been found at silent genes with a nonproductive form of RNA polymerase II, the repressive chromatin modifying complex Polycomb, the enzymatic complex responsible for H3K27me3 and SetDB1, the enzyme responsible for H3K9me3. The genes occupied by this combination of factors encode proteins required for early differentiation and lineage commitment during development.

Signal Transduction Pathways

The culture of ES cells initially required a layer of irradiated fibroblasts in order to obtain the necessary factors for proliferation and pluripotency. These fibroblasts produce cytokines and growth factors necessary for ES cell pluripotency and self-renewal (Smith and Hooper, 1983). The key factors secreted by fibroblasts have been identified, so ES cells can now be grown in chemically defined medium in the absence of irradiated fibroblasts. LIF, Wnt, BMP4 and TGF-β were among the factors supplied by the fibroblasts and found to impact murine ES cell state (Okita and Yamanaka, 2006; Sato et al., 2004; Smith et al., 1998; Williams et al., 1988; Ying et al., 2003). Thus, the LIF, Wnt, BMP4 and TGF-β signaling pathways can contribute to maintenance of ES cell state (Figure 4).

Leukemia inhibitory factor (LIF), which was originally identified through its ability to inhibit growth of myeloid leukemia cells, has been implicated in growth promotion and differentiation of various cell types (Hilton, 1992). LIF, a member of the IL-6 family of cytokines, is expressed in the trophectoderm of the developing embryo, with its receptor LIFR expressed throughout the inner cell mass. ES cells are derived from the inner cell mass of blastocysts, so removing them from blastocysts also removes their source of this cytokine. LIF binding to LIFR leads to phosphorylation and activation of the JAK/STAT and MAPK cascades. In ES cells, activated STAT3 is translocated to the nucleus where it acts as a transcription factor (Auernhammer and Melmed, 2000; Stahl et al., 1995). STAT3 interacts with the core transcription factors Oct4, Sox2 and Nanog, explaining how LIF signaling connects to the core regulatory circuitry and thus contributes to maintainenance of murine ES cell state (Chen et al., 2008b). LIF has been demonstrated to activate STAT3 in human ES cells but is unable to maintain their pluripotency (Daheron et al., 2004; Humphrey et al., 2004). In mouse embryogenesis LIF is produced to allow a pregnant female to maintain her embryos in an embryonic stem cell state until her environment is appropriate for development and birth (Nichols et al., 2001). This is likely the reason why LIF has a profound impact on mES cell state but has no effect in the maintenance of hES cells.

Wnt signaling also contributes to the maintenance of ES cell state (Aubert et al., 2002; Kielman et al., 2002; Sato et al., 2004; Yoshikawa et al., 1997). Wnt signaling is

mediated by the intracellular signaling protein β-catenin. In the absence of Wnt signaling, β-catenin is phosphorylated and subsequently degraded by the axin/GSK-3/APC complex. When Wnt proteins bind to cell-surface receptors of the Frizzled family, the receptors activate Dishevelled proteins, which inhibit the axin/GSK-3/APC complex. Thus, when the Wnt signaling pathway is activated, unphosphorylated forms of β-catenin accumulate in the cytoplasm and are shuttled into the nucleus (Reya and Clevers, 2005). Within the nucleus, β-catenin interacts with DNA bound co-factors of the lymphoic enhancer factor (LEF)/T-cell factor (TCF) family of proteins. The key LEF/TCF factor expressed in ES cells is TCF3, which co-occupies promoters with Oct4, Sox2 and Nanog. In the absence of β-catenin, TCF3 acts as a transcriptional repressor, whereas when bound by β-catenin, it acts as an activator. Thus, Wnt signaling connects directly to the core regulatory circuitry of ES cells by converting TCF3 from a repressor to an activator (Cole et al., 2008). This study is the subject of Chapter 2.

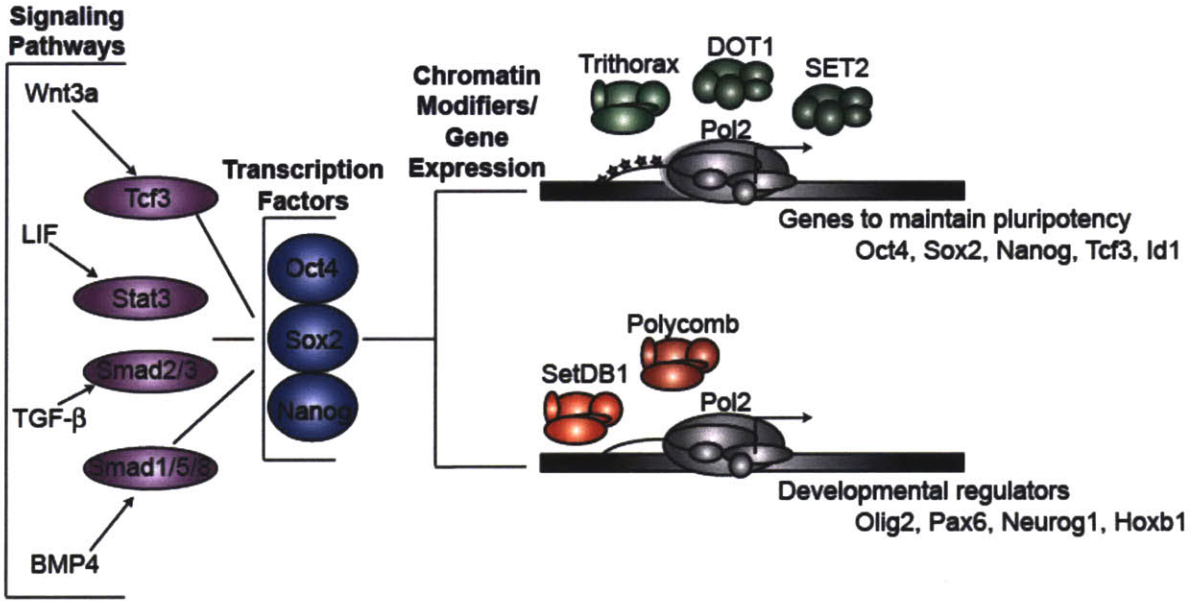
The bone morphogenic protein 4 (BMP4) signaling pathway is a member of the TGF-β superfamily of signaling pathways (Shi and Massague, 2003). Activation of the BMP4 pathway stimulates the phosphorylation of Smad proteins 1, 5 and 8. Phosphorylation of the Smads leads to an interaction with the co-Smad, Smad4. Once together in a complex with Smad4, the Smad proteins are shuttled to the nucleus where they act to regulate gene expression. BMP4 has been shown to maintain pluripotency of ES cells by activating genes encoding inhibitors of differentiation proteins (Graichen et al.) and can function together with LIF in the absence of serum to maintain ES cell state (Ying et al., 2003). As with the transcription factors targeted by the LIF and Wnt signaling pathways, Smad1 often occupies genes with the ES cell master transcription factors Oct4, Sox2 and Nanog (Chen et al., 2008b). In addition to BMP4 signaling, the TGF-β family of signaling pathways can also be activated by Activin/Nodal signaling. Here, signaling goes through Smad2 and Smad3 (James et al., 2005; Ross and Hill, 2008; Rossant, 2008). Upon activation, Smad2/3 is phosphorylated, interacts with Smad4 and is translocated to the nucleus. In mES cells the activation of this pathway can lead to differentiation, where as in hES cells this pathway has a role in maintaining ES cell state. A recent study identified the gene targets of Smad2/3 in both mouse and human ES cells as well as differentiated cell types. This study, presented here in

Chapter 3, highlights the specificity and interaction of signaling pathways with critical transcription factors. A high overlap is observed between the sites occupied by ES cell master transcription factors Oct4, Sox2 and Nanog with Smad3. Although similar genes are occupied in mouse and human ES cells their role at those genes is likely different. In addition, Smad3 was bound to a different set of genes in muscle cells and in proB cells but in each cell type Smad3 demonstrated a significant degree of co-occupancy with the master transcription factors, MyoD in muscle cells and PU.1 in proB cells, suggesting a relationship between transcription factors and signaling pathways that is critical for regulating cell-type specific gene expression (Mullen et al., 2010).

New knowledge of the contributions of signaling pathways to the control of ES cell state have led to advances in cellular reprogramming. A role for Wnt signaling in the core regulatory circuitry of ES cells led Marson and colleagues to show that Wnt3a conditioned media dramatically improves the efficiency of cellular reprogramming in the absence of the proto-oncogene c-Myc (Marson et al., 2008a). Similarly, by inhibiting TGF-β signaling using an inhibitor specific for part of the Activin/Nodal signaling pathway, a pathway known to promote differentiation of murine ES cells, reprogramming efficiency was improved in the absence of c-myc (Maherali and Hochedlinger, 2009). The inhibition of this signaling pathway likely blocked other signals of differentiation and so lowered the threshold for the expression of genes necessary to regain the pluripotent state. These results are important because of the interest in eliminating exogenous proto-oncogenes from reprogramming protocols.

Figure 4

A.


Signaling Pathways
Wnt3a
Tcf3
LIF
Stat3
Smad2/3
TGF-β
Smad1/5/8
BMP4
Transcription Factors
Oct4
Sox2
Nanog
Chromatin Modifiers/ Gene Expression
Trithorax
DOT1
SET2
Pol2
Genes to maintain pluripotency
Oct4, Sox2, Nanog, Tcf3, Id1
Polycomb
SetDB1
Pol2
Developmental regulators
Olig2, Pax6, Neurog1, Hoxb1

B.


HATs
HATs
Smad2/3
Smad1/5/8
Sox2
Sall4
Tcf3
Smad4
Oct4
Nanog
STAT3
Smad4

Figure 4. Signaling pathways interact with the transcriptional regulatory circuitry of embryonic stem cells.

(A). Signaling pathways have a critical role in the interaction with transcription factors and chromatin modifiers to direct gene expression in ES cells. The Wnt, LIF, TGF-β and BMP4 signaling pathways are known to contribute to ES cell state. Genome-wide profiling of the terminal component of each of these signaling pathways shows a direct involvement with Oct4, Sox2 and Nanog to control gene expression and cell state.

(B). Terminal signaling components co-occupy parts of the genome with transcription factors Oct4, Sox2, Nanog and Sall4. Nucleosome position is shifted and chromatin is opened by histone modifiers including histone acetyltransferases (HATs) that increase accessibility of DNA through the addition of acetyl groups to histones. Specific DNA sequences are subsequently recognized and bound by transcription factors and members of signaling pathways. The terminal components of signaling pathways are likely recruited to these sites by the transcription factors that control cell state.

Genetic Screens Reveal Novel Regulatory Factors of ES Cell State

The pioneering studies that identified Oct4 and Nanog as key regulators of embryonic stem cells came from studies of differential gene expression and genetic manipulation. The discovery that Oct4 was necessary for the maintenance of ES cell state came from a series of studies that first identified the differential expression of Oct4 in embryonic stem cells. These studies used cloning and sequencing techniques to identify Oct4 as a novel member of the POU domain family of proteins (Okamoto et al., 1990; Rosner et al., 1990; Scholer et al., 1990). Along with the identification of Oct4 as a novel protein, its expression was determined to be limited to embryonic cells and was undetectable in adult, differentiated cells (Okamoto et al., 1990). These studies were followed by experiments that revealed the significant impact on cell state following the loss of Oct4. The loss of Oct4 leads to a loss of pluripotency and a commitment to a trophectoderm lineage (Nichols et al., 1998). Overexpression of Oct4 has also been discovered to lead to differentiation, suggesting that the level of Oct4 expression is critical to the proper maintenance of ES cell state (Niwa et al., 2000).

More recently, large-scale genetic screens have identified additional transcription factors, chromatin modifiers, transcriptional coactivators, and chromosome scaffolding proteins that contribute to maintenance of ES cell state (Bilodeau et al., 2009; Fazzio et al., 2008; Hu et al., 2009; Ivanova et al., 2006; Zhang et al., 2006). Each of the screens identified known ES cell regulators such as Oct4, Nanog and Stat3, indicating that other genes identified in the screen were good candidates for novel ES cell regulators. The novel transcription factors include Trim28, Cnot3, Zfp42/Rex-1, Esrrb, Tbx3 and Tcl1 (Hu et al., 2009; Ivanova et al., 2006; Zhang et al., 2006). The novel chromatin regulators include Tip-60 and SetDB1 (Bilodeau et al., 2009; Fazzio et al., 2008). The novel transcriptional coactivators and chromosome scaffolding proteins include Mediator and Cohesin (Kagey et al., 2010).

The observation that reduced levels of Mediator and Cohesin cause loss of ES cell state led us to further investigate their functions in ES cells (Kagey et al., 2010), and this study is the subject of Chapter 4. Both Mediator and Cohesin were found to occupy the enhancers and core promoters of a key subset of actively transcribed genes and to

be necessary for normal transcriptional activity in ES cells. The Cohesin loading factor Nipbl was found at the sites co-occupied by Mediator and Cohesin, providing a mechanism for Cohesin loading at these sites. Mediator and Cohesin were found to physically interact, further explaining how the recruitment of Mediator by a transcription factor could lead to association with Cohesin. Chromosome conformation capture experiments revealed that the enhancer and core promoter sites occupied by Mediator and Cohesin are brought into close physical proximity, confirming DNA loop formation. Mediator and Cohesin co-occupancy of the genome was found to be cell-type specific due to cell-type specific gene activity. These and other results led us to propose that Mediator and Cohesin contribute to a looped and reinforced chromatin structure at active promoters genome-wide, thus generating a cell-type specific chromatin architecture associated with the transcriptional program of each cell.

Concluding Remarks

Embryonic stem cells provide a valuable system in which to study mechanisms of gene regulation, cell state, and differentiation. The ability to maintain ES cells and iPS cells indefinitely in a pluripotent state in culture are properties unique to these cells that have facilitated experimental investigation of the control of cell state. Consequently, much has been learned about the roles of specific transcription factors, signaling pathways, chromatin modifiers and microRNAs in controlling the gene expression program responsible for the pluripotent state of these cells (Bernstein et al., 2007; Boyer et al., 2005; Boyer et al., 2006; Chen et al., 2008b; Chew et al., 2005; Cole et al., 2008; Endoh et al., 2008; Jiang et al., 2008; Kim et al., 2008; Lee et al., 2006; Loh et al., 2006; Marson et al., 2008b; Tam et al., 2008).

The study of the key ES cell transcription factors has led to three key concepts in transcriptional control of ES cell state: 1) the master regulators collaborate to regulate each of their own promoters, forming an interconnected autoregulatory loop, 2) the master regulators co-occupy and collaborate to regulate their target genes, and 3) the master regulators function to activate expression of genes necessary to maintain ES cell state, while contributing to silencing of genes whose repression is essential to

maintaining that state (Boyer et al., 2005; Lee et al., 2006; Loh et al., 2006). The central theme that emerged from the study of ES cell signaling pathways is that the transcription factors that serve as effectors of these pathways can occupy and function at the genes regulated by the master transcription factors Oct4, Sox2 and Nanog. This suggests that master transcription factors –those that are key to control of cell state - have properties that facilitate binding of the signaling transcription factors to adjacent sites at enhancers.

Continued study of ES cell transcriptional regulatory mechanisms and circuitry is almost certain to continue to provide important new insights into the control of gene expression and cell state. Genetic screens have identified a large number of novel candidate regulators of ES cell state. Our own studies of Mediator and Cohesin, two complexes that emerged from these screens, have provided new insights into the mechanisms involved in gene control, DNA looping and control of cell state (Kagey et al., 2010). These studies argue that DNA loop formation between enhancers and core promoters occurs as a consequence of the interaction between enhancer-bound transcription activators, Mediator and promoter-bound RNA polymerase II. When the transcription activators bind Mediator, the Mediator complex undergoes a conformational change, and this form of Mediator binds Cohesin. The Cohesin loading factor Nipbl is located at the sites co-occupied by Mediator and Cohesin, providing a mechanism for Cohesin loading at these sites. The subset of genes occupied by Mediator and Cohesin is cell-type specific, thus indicating that cell-type specific loops exist in the chromosomes of vertebrate cells as a consequence of active gene regulation.

Further understanding the control of ES cell state will also provide new insights into human disease. For example, mutations in ES cell transcription factors (Myc) and signaling pathways (Wnt, TGFb) can lead to cancer. Mutations in the genes encoding Mediator and Cohesin components can cause an array of human developmental syndromes and diseases. Mediator mutations have been associated with Opitz-Kaveggia (FG) syndrome, Lujan syndrome, schizophrenia and some forms of congenital heart failure (Ding et al., 2008; Philibert et al., 2007; Philibert and Madan, 2007; Risheg et al., 2007; Schwartz et al., 2007). Mutations in Nipbl are responsible for most cases of

Cornelia de Lange syndrome (CdLS), which is characterized by developmental defects and mental retardation and appears to be the result of mis-regulation of gene expression rather than chromosome cohesion or mitotic abnormalities (Borck et al., 2004; Musio et al., 2006; Strachan, 2005; Zhang et al., 2007). More detailed understanding of these diseases and syndromes, and potential therapeutics, will almost certainly emerge from further study of ES cells and from iPS cells derived from patients with these diseases.

My Research

My work in graduate school has focused on understanding how transcriptional regulatory mechanisms function in the control of gene expression in embryonic stem cells. The chapters presented in this document reflect this work and my contributions to expanding the regulatory network of embryonic stem cells to include signaling pathways and structural proteins.

Chapter 2 describes how Wnt signaling is mediated through Tcf3 and directly impacts the core transcriptional regulatory circuitry of embryonic stem cells. I contributed to this work by establishing our lab's mouse ES cell culturing approach. I used knockdown studies to examine the functional role of Tcf3 and Wnt signaling in ES cells. I also studied expression changes following stimulation of ES cells with Wnt3a conditioned media.

Chapter 3 describes the discovery that the TGF-β terminal signaling factor Smad3 co-occupies genes with the master transcription factors of embryonic stem cells (Oct4), myotubes (Myod) and ProB cells (PU.1). My contributions to this project include the initiation of the culture and study of human embryonic stem cells in the lab. I also contributed to key concepts, experimental design and conducted experiments for this project, including the perturbations of signaling pathways. I also conducted experiments to demonstrate the active recruitment of Smad3 following the over-expression of specific transcription factors.

Chapter 4 describes a role for the Mediator and Cohesin protein complexes in maintaining embryonic stem cell state through the formation of DNA loops at active genes. I participated in screening the transcription factor short hairpin library and followed up on the role of Mediator in embryonic stem cells. I performed knockdown experiments and ChIP-seq following the screen, leading to the discovery that Mediator and Cohesin have a critical, cell-type specific role in DNA looping at active genes.

References

Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos, D. (2000). Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. Cell *103*, 667-678.

Agger, K., Christensen, J., Cloos, P.A., and Helin, K. (2008). The emerging functions of histone demethylases. Curr Opin Genet Dev *18*, 159-168.

Ambrosetti, D.C., Basilico, C., and Dailey, L. (1997). Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol Cell Biol *17*, 6321-6329.

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

Aubert, J., Dunstan, H., Chambers, I., and Smith, A. (2002). Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. Nat Biotechnol *20*, 1240-1245.

Auernhammer, C.J., and Melmed, S. (2000). Leukemia-inhibitory factor-neuroimmune modulator of endocrine function. Endocr Rev *21*, 313-345.

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

Ayyanathan, K., Lechner, M.S., Bell, P., Maul, G.G., Schultz, D.C., Yamada, Y., Tanaka, K., Torigoe, K., and Rauscher, F.J., 3rd (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes Dev *17*, 1855-1869.

Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., and Teichmann, S.A. (2004). Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol *14*, 283-291.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299-308.

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.

Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promotor region. Nature *290*, 304-310.

Berger, S.L. (2002). Histone modifications in transcriptional regulation. Curr Opin Genet Dev *12*, 142-148.

Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. (2004). Global nucleosome occupancy in yeast. Genome Biol *5*, R62.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669-681.

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.

Bianchi, M.E., and Agresti, A. (2005). HMG proteins: dynamic players in gene regulation and differentiation. Curr Opin Genet Dev *15*, 496-506.

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

Borck, G., Redon, R., Sanlaville, D., Rio, M., Prieur, M., Lyonnet, S., Vekemans, M., Carter, N.P., Munnich, A., Colleaux, L., *et al.* (2004). NIPBL mutations and genetic heterogeneity in Cornelia de Lange syndrome. J Med Genet *41*, e128.

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

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.

Breiling, A., Sessa, L., and Orlando, V. (2007). Biology of polycomb and trithorax group proteins. Int Rev Cytol *258*, 83-136.

Browning, D.F., and Busby, S.J. (2004). The regulation of bacterial transcription initiation. Nat Rev Microbiol *2*, 57-65.

Bustin, M., and Reeves, R. (1996). High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. Prog Nucleic Acid Res Mol Biol *54*, 35-100.

Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193-198.

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

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

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

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

Chew, J.L., Loh, Y.H., Zhang, W., Chen, X., Tam, W.L., Yeap, L.S., Li, P., Ang, Y.S., Lim, B., Robson, P., *et al.* (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Mol Cell Biol *25*, 6031-6046.

Claverie-Martin, F., and Magasanik, B. (1991). Role of integration host factor in the regulation of the glnHp2 promoter of Escherichia coli. Proc Natl Acad Sci U S A *88*, 1631-1635.

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

Conaway, R.C., Sato, S., Tomomori-Sato, C., Yao, T., and Conaway, J.W. (2005). The mammalian Mediator complex and its role in transcriptional regulation. Trends Biochem Sci *30*, 250-255.

D'Alessio, J.A., Wright, K.J., and Tjian, R. (2009). Shifting players and paradigms in cell-specific transcription. Mol Cell *36*, 924-931.

D'Amour, K.A., Bang, A.G., Eliazer, S., Kelly, O.G., Agulnick, A.D., Smart, N.G., Moorman, M.A., Kroon, E., Carpenter, M.K., and Baetge, E.E. (2006). Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. Nat Biotechnol *24*, 1392-1401.

Daheron, L., Opitz, S.L., Zaehres, H., Lensch, M.W., Andrews, P.W., Itskovitz-Eldor, J., and Daley, G.Q. (2004). LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. Stem Cells *22*, 770-778.

Dame, R.T., Wyman, C., and Goosen, N. (2000). H-NS mediated compaction of DNA visualised by atomic force microscopy. Nucleic Acids Res *28*, 3504-3510.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol *319*, 1097-1113.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.

de la Serna, I.L., Ohkawa, Y., Higashi, C., Dutta, C., Osias, J., Kommajosyula, N., Tachibana, T., and Imbalzano, A.N. (2006a). The microphthalmia-associated transcription factor requires SWI/SNF enzymes to activate melanocyte-specific genes. J Biol Chem *281*, 20233-20241.

de la Serna, I.L., Ohkawa, Y., and Imbalzano, A.N. (2006b). Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. Nat Rev Genet *7*, 461-473.

Dimos, J.T., Rodolfa, K.T., Niakan, K.K., Weisenthal, L.M., Mitsumoto, H., Chung, W., Croft, G.F., Saphier, G., Leibel, R., Goland, R., *et al.* (2008). Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science *321*, 1218-1221.

Ding, N., Zhou, H., Esteve, P.O., Chin, H.G., Kim, S., Xu, X., Joseph, S.M., Friez, M.J., Schwartz, C.E., Pradhan, S., *et al.* (2008). Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. Mol Cell *31*, 347-359.

Dorman, C.J., Hinton, J.C., and Free, A. (1999). Domain organization and oligomerization among H-NS-like nucleoid-associated proteins in bacteria. Trends Microbiol *7*, 124-128.

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

Downing, G.J., and Battey, J.F., Jr. (2004). Technical assessment of the first 20 years of research using mouse embryonic stem cell lines. Stem Cells *22*, 1168-1180.

Dunn, K.L., and Davie, J.R. (2003). The many roles of the transcriptional regulator CTCF. Biochem Cell Biol *81*, 161-167.

Ebert, A.D., Yu, J., Rose, F.F., Jr., Mattis, V.B., Lorson, C.L., Thomson, J.A., and Svendsen, C.N. (2009). Induced pluripotent stem cells from a spinal muscular atrophy patient. Nature *457*, 277-280.

Ellenberger, T., and Landy, A. (1997). A good turn for DNA: the structure of integration host factor bound to DNA. Structure *5*, 153-157.

Endoh, M., Endo, T.A., Endoh, T., Fujimura, Y., Ohara, O., Toyoda, T., Otte, A.P., Okano, M., Brockdorff, N., Vidal, M., *et al.* (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development *135*, 1513-1524.

Faust, C., Lawson, K.A., Schork, N.J., Thiel, B., and Magnuson, T. (1998). The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. Development *125*, 4495-4506.

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

Filippova, G.N., Qi, C.F., Ulmer, J.E., Moore, J.M., Ward, M.D., Hu, Y.J., Loukinov, D.I., Pugacheva, E.M., Klenova, E.M., Grundy, P.E., *et al.* (2002). Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter tts DNA-binding specificity. Cancer Res *62*, 48-52.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A *73*, 1897-1901.

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

Gerber, M., and Shilatifard, A. (2003). Transcriptional elongation by RNA polymerase II and histone methylation. J Biol Chem *278*, 26303-26306.

Gerstel, U., Park, C., and Romling, U. (2003). Complex regulation of csgD promoter activity by global regulatory proteins. Mol Microbiol *49*, 639-654.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell *118*, 555-566.

Goetze, S., Mateos-Langerak, J., and van Driel, R. (2007). Three-dimensional genome organization in interphase and its relation to genome function. Semin Cell Dev Biol *18*, 707-714.

Goodrich, J.A., Schwartz, M.L., and McClure, W.R. (1990). Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for Escherichia coli integration host factor (IHF). Nucleic Acids Res *18*, 4993-5000.

Graichen, R., Xu, X., Braam, S.R., Balakrishnan, T., Norfiza, S., Sieh, S., Soo, S.Y., Tham, S.C., Mummery, C., Colman, A., *et al.* (2008). Enhanced cardiomyogenesis of human embryonic stem cells by a small molecular inhibitor of p38 MAPK. Differentiation *76*, 357-370.

Gurdon, J.B., and Byrne, J.A. (2003). The first half-century of nuclear transplantation. Proc Natl Acad Sci U S A *100*, 8048-8052.

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature *460*, 410-413.

Hampsey, M., and Reinberg, D. (1999). RNA polymerase II as a control panel for multiple coactivator complexes. Curr Opin Genet Dev *9*, 132-139.

Hart, A.H., Hartley, L., Ibrahim, M., and Robb, L. (2004). Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. Dev Dyn *230*, 187-198.

Hassig, C.A., and Schreiber, S.L. (1997). Nuclear histone acetylases and deacetylases and transcriptional regulation: HATs off to HDACs. Curr Opin Chem Biol *1*, 300-308.

Hilton, D.J. (1992). LIF: lots of interesting functions. Trends Biochem Sci *17*, 72-76.

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

Hochschild, A., and Ptashne, M. (1988). Interaction at a distance between lambda repressors disrupts gene activation. Nature *336*, 353-357.

Hoover, T.R., Santero, E., Porter, S., and Kustu, S. (1990). The integration host factor stimulates interaction of RNA polymerase with NIFA, the transcriptional activator for nitrogen fixation operons. Cell *63*, 11-22.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev *23*, 837-848.

Humphrey, R.K., Beattie, G.M., Lopez, A.D., Bucay, N., King, C.C., Firpo, M.T., Rose-John, S., and Hayek, A. (2004). Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. Stem Cells *22*, 522-530.

Huo, L., Martin, K.J., and Schleif, R. (1988). Alternative DNA loops regulate the arabinose operon in Escherichia coli. Proc Natl Acad Sci U S A *85*, 5444-5448.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature *442*, 533-538.

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

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

James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. (2005). TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development *132*, 1273-1282.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet *10*, 161-172.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol *10*, 353-360.

Joannides, A.J., Fiore-Heriche, C., Battersby, A.A., Athauda-Arachchi, P., Bouhon, I.A., Williams, L., Westmore, K., Kemp, P.J., Compston, A., Allen, N.D., *et al.* (2007). A scaleable and defined system for generating neural stem cells from human embryonic stem cells. Stem Cells *25*, 731-737.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., van Berkum, N.L., Orlando, D.A., Ebmeier, C.C., Goossens, J., Rahl, P., Levine, S., *et al.* (2010). Mediator and Cohesin connect gene expression and chromatin architecture. Submitted for publication.

Kahan, B.W., and Ephrussi, B. (1970). Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. J Natl Cancer Inst *44*, 1015-1036.

Kielman, M.F., Rindapaa, M., Gaspar, C., van Poppel, N., Breukel, C., van Leeuwen, S., Taketo, M.M., Roberts, S., Smits, R., and Fodde, R. (2002). Apc modulates embryonic stem-cell differentiation by controlling the dosage of beta-catenin signaling. Nat Genet *32*, 594-605.

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

Kim, J.K., Samaranayake, M., and Pradhan, S. (2009). Epigenetic mechanisms in mammals. Cell Mol Life Sci *66*, 596-612.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231-1245.

Kingston, R.E., and Narlikar, G.J. (1999). ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. Genes Dev *13*, 2339-2352.

Kirkpatrick, C.R., and Schimmel, P. (1995). Detection of leucine-independent DNA site occupancy of the yeast Leu3p transcriptional activator in vivo. Mol Cell Biol *15*, 4021-4030.

Kiskinis, E., and Eggan, K. (2010). Progress toward the clinical application of patient-specific pluripotent stem cells. J Clin Invest *120*, 51-59.

Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. Trends Biochem Sci *30*, 235-239.

Kornberg, R.D., and Klug, A. (1981). The nucleosome. Sci Am *244*, 52-64.

Kouzarides, T. (2000). Acetylation: a regulatory modification to rival phosphorylation? EMBO J *19*, 1176-1179.

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

Kubo, A., Shinozaki, K., Shannon, J.M., Kouskoff, V., Kennedy, M., Woo, S., Fehling, H.J., and Keller, G. (2004). Development of definitive endoderm from embryonic stem cells in culture. Development *131*, 1651-1662.

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

Laflamme, M.A., Chen, K.Y., Naumova, A.V., Muskheli, V., Fugate, J.A., Dupras, S.K., Reinecke, H., Xu, C., Hassanipour, M., Police, S., *et al.* (2007). Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. Nat Biotechnol *25*, 1015-1024.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

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.

Levasseur, D.N., Wang, J., Dorschner, M.O., Stamatoyannopoulos, J.A., and Orkin, S.H. (2008). Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. Genes Dev *22*, 575-580.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147-151.

Lewis, A., and Murrell, A. (2004). Genomic imprinting: CTCF protects the boundaries. Curr Biol *14*, R284-286.

Li, T., Hu, J.F., Qiu, X., Ling, J., Chen, H., Wang, S., Hou, A., Vu, T.H., and Hoffman, A.R. (2008). CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. Mol Cell Biol *28*, 6473-6482.

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

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

Lohmann, F., Loureiro, J., Su, H., Fang, Q., Lei, H., Lewis, T., Yang, Y., Labow, M., Li, E., Chen, T., *et al.* (2010). KMT1E mediated H3K9 methylation is required for the maintenance of embryonic stem cells by repressing trophectoderm differentiation. Stem Cells *28*, 201-212.

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

Luijsterburg, M.S., Noom, M.C., Wuite, G.J., and Dame, R.T. (2006). The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. J Struct Biol *156*, 262-272.

Luijsterburg, M.S., White, M.F., van Driel, R., and Dame, R.T. (2008). The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. Crit Rev Biochem Mol Biol *43*, 393-418.

Maherali, N., and Hochedlinger, K. (2008). Guidelines and techniques for the generation of induced pluripotent stem cells. Cell Stem Cell *3*, 595-605.

Maherali, N., and Hochedlinger, K. (2009). Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. Curr Biol *19*, 1718-1723.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., *et al.* (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55-70.

Majumder, P., Gomez, J.A., Chadwick, B.P., and Boss, J.M. (2008). The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. J Exp Med *205*, 785-798.

Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci *30*, 256-263.

Malik, S., and Roeder, R.G. (2008). Epigenetics? Mediator does that too! Mol Cell *31*, 305-306.

Maniatis, T., Falvo, J.V., Kim, T.H., Kim, T.K., Lin, C.H., Parekh, B.S., and Wathelet, M.G. (1998). Structure and function of the interferon-beta enhanceosome. Cold Spring Harb Symp Quant Biol *63*, 609-620.

Maniatis, T., Goodbourn, S., and Fischer, J.A. (1987). Regulation of inducible and tissue-specific gene expression. Science *236*, 1237-1245.

Marson, A., Foreman, R., Chevalier, B., Bilodeau, S., Kahn, M., Young, R.A., and Jaenisch, R. (2008a). Wnt signaling promotes reprogramming of somatic cells to pluripotency. Cell Stem Cell *3*, 132-135.

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

Martin, G.R. (1980). Teratocarcinomas and mammalian embryogenesis. Science *209*, 768-776.

Martinez-Antonio, A., and Collado-Vides, J. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. Curr Opin Microbiol *6*, 482-489.

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

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., *et al.* (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358-362.

McGrath, J., and Solter, D. (1983). Nuclear transplantation in the mouse embryo by microsurgery and cell fusion. Science *220*, 1300-1302.

McKenna, N.J., and O'Malley, B.W. (2002). Combinatorial control of gene expression by nuclear receptors and coregulators. Cell *108*, 465-474.

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

Mohd-Sarip, A., and Verrijzer, C.P. (2004). Molecular biology. A higher order of silence. Science *306*, 1484-1485.

Mullen, A.C., Newman, J.J., Orlando, D.A., Kumar, R.M., and Young, R.A. (2010). Cell-type specific TGF-beta signaling is targeted to genes that control cell identity. Submitted for publication.

Murry, C.E., and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell *132*, 661-680.

Musio, A., Selicorni, A., Focarelli, M.L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P., and Larizza, L. (2006). X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat Genet *38*, 528-530.

Naar, A.M., Lemon, B.D., and Tjian, R. (2001). Transcriptional coactivator complexes. Annu Rev Biochem *70*, 475-501.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.

Narlikar, L., and Ovcharenko, I. (2009). Identifying regulatory elements in eukaryotic genomes. Brief Funct Genomic Proteomic *8*, 215-230.

Nativio, R., Wendt, K.S., Ito, Y., Huddleston, J.E., Uribe-Lewis, S., Woodfine, K., Krueger, C., Reik, W., Peters, J.M., and Murrell, A. (2009). Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. PLoS Genet *5*, e1000739.

Ng, E.S., Davis, R.P., Azzola, L., Stanley, E.G., and Elefanty, A.G. (2005). Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation. Blood *106*, 1601-1603.

Nichols, J., Chambers, I., Taga, T., and Smith, A. (2001). Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. Development *128*, 2333-2339.

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

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

Nostro, M.C., Cheng, X., Keller, G.M., and Gadue, P. (2008). Wnt, activin, and BMP signaling regulate distinct stages in the developmental pathway from embryonic stem cells to blood. Cell Stem Cell *2*, 60-71.

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.

Ohlsson, R., Bartkuhn, M., and Renkawitz, R. (2010). CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. Chromosoma.

Ohlsson, R., Lobanenkov, V., and Klenova, E. (2009). Does CTCF mediate between nuclear organization and gene expression? Bioessays *32*, 37-50.

Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M., and Hamada, H. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell *60*, 461-472.

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

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

Orlando, V. (2003). Polycomb, epigenomes, and control of cell identity. Cell *112*, 599-606.

Pan, C.Q., Finkel, S.E., Cramton, S.E., Feng, J.A., Sigman, D.S., and Johnson, R.C. (1996). Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. J Mol Biol *264*, 675-695.

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.

Pan, Y., Tsai, C.J., Ma, B., and Nussinov, R. (2010). Mechanisms of transcription factor selectivity. Trends Genet *26*, 75-83.

Panne, D. (2008). The enhanceosome. Curr Opin Struct Biol *18*, 236-242.

Pardal, R., Clarke, M.F., and Morrison, S.J. (2003). Applying the principles of stem-cell biology to cancer. Nat Rev Cancer *3*, 895-902.

Park, I.H., Lerou, P.H., Zhao, R., Huo, H., and Daley, G.Q. (2008a). Generation of human-induced pluripotent stem cells. Nat Protoc *3*, 1180-1186.

Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008b). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141-146.

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

Pera, M.F., and Trounson, A.O. (2004). Human embryonic stem cells: prospects for development. Development *131*, 5515-5525.

Pevny, L.H., and Lovell-Badge, R. (1997). Sox genes find their feet. Curr Opin Genet Dev *7*, 338-344.

Philibert, R.A., Bohle, P., Secrest, D., Deaderick, J., Sandhu, H., Crowe, R., and Black, D.W. (2007). The association of the HOPA(12bp) polymorphism with schizophrenia in the NIMH Genetics Initiative for Schizophrenia sample. Am J Med Genet B Neuropsychiatr Genet *144B*, 743-747.

Philibert, R.A., and Madan, A. (2007). Role of MED12 in transcription and human behavior. Pharmacogenomics *8*, 909-916.

Phillips, D.M. (1963). The presence of acetyl groups of histones. Biochem J *87*, 258-263.

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

Pirrotta, V. (1998). Polycombing the genome: PcG, trxG, and chromatin silencing. Cell *93*, 333-336.

Popham, D.L., Szeto, D., Keener, J., and Kustu, S. (1989). Function of a bacterial activator protein that binds to transcriptional enhancers. Science *243*, 629-635.

Reitzer, L.J., and Magasanik, B. (1986). Transcription of glnA in E. coli is stimulated by activator bound to sites far from the promoter. Cell *45*, 785-792.

Reya, T., and Clevers, H. (2005). Wnt signalling in stem cells and cancer. Nature *434*, 843-850.

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

Rippe, K., Guthold, M., von Hippel, P.H., and Bustamante, C. (1997). Transcriptional activation via DNA-looping: visualization of intermediates in the activation pathway of E. coli RNA polymerase x sigma 54 holoenzyme by scanning force microscopy. J Mol Biol *270*, 125-138.

Ris, H., and Kubai, D.F. (1970). Chromosome structure. Annu Rev Genet *4*, 263-294.

Risheg, H., Graham, J.M., Jr., Clark, R.D., Rogers, R.C., Opitz, J.M., Moeschler, J.B., Peiffer, A.P., May, M., Joseph, S.M., Jones, J.R., *et al.* (2007). A recurrent mutation in MED12 leading to R961W causes Opitz-Kaveggia syndrome. Nat Genet *39*, 451-453.

Roeder, R.G. (1998). Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. Cold Spring Harb Symp Quant Biol *63*, 201-218.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett *579*, 909-915.

Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier, F., Rigby, P.W., and Staudt, L.M. (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature *345*, 686-692.

Ross, S., and Hill, C.S. (2008). How the Smads regulate transcription. Int J Biochem Cell Biol *40*, 383-408.

Rossant, J. (2008). Stem cells and early lineage development. Cell *132*, 527-531.

Roth, S.Y., Denu, J.M., and Allis, C.D. (2001). Histone acetyltransferases. Annu Rev Biochem *70*, 81-120.

Saladi, S.V., and de la Serna, I.L. (2010). ATP dependent chromatin remodeling enzymes in embryonic stem cells. Stem Cell Rev *6*, 62-73.

Sasse-Dwight, S., and Gralla, J.D. (1988). Probing the Escherichia coli glnALG upstream activation mechanism in vivo. Proc Natl Acad Sci U S A *85*, 8934-8938.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004). 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*, 55-63.

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

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

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2002). 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*, 919-932.

Schwartz, C.E., Tarpey, P.S., Lubs, H.A., Verloes, A., May, M.M., Risheg, H., Friez, M.J., Futreal, P.A., Edkins, S., Teague, J., *et al.* (2007). The original Lujan syndrome family has a novel missense mutation (p.N1007S) in the MED12 gene. J Med Genet *44*, 472-477.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet *8*, 9-22.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772-778.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends Genet *25*, 335-343.

Shi, Y., and Massague, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell *113*, 685-700.

Sif, S. (2004). ATP-dependent nucleosome remodeling complexes: enzymes tailored to deal with chromatin. J Cell Biochem *91*, 1087-1098.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol *6*, e253.

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

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

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

Soldner, F., Hockemeyer, D., Beard, C., Gao, Q., Bell, G.W., Cook, E.G., Hargus, G., Blak, A., Cooper, O., Mitalipova, M., *et al.* (2009). Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. Cell *136*, 964-977.

Spiegelman, B.M., and Heinrich, R. (2004). Biological control through regulated transcriptional coactivators. Cell *119*, 157-167.

Spilianakis, C.G., and Flavell, R.A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nat Immunol *5*, 1017-1027.

Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev *20*, 2349-2354.

Stahl, N., Farruggella, T.J., Boulton, T.G., Zhong, Z., Darnell, J.E., Jr., and Yancopoulos, G.D. (1995). Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. Science *267*, 1349-1353.

Stevens, L.C. (1978). Totipotent cells of parthenogenetic origin in a chimaeric mouse. Nature *276*, 266-267.

Stevens, L.C., Varnum, D.S., and Eicher, E.M. (1977). Viable chimaeras produced from normal and parthenogenetic mouse embryos. Nature *269*, 515-517.

Strachan, T. (2005). Cornelia de Lange Syndrome and the link between chromosomal function, DNA repair and developmental gene regulation. Curr Opin Genet Dev *15*, 258-264.

Su, W., Porter, S., Kustu, S., and Echols, H. (1990). DNA-looping and enhancer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial glnA promoter. Proc Natl Acad Sci U S A *87*, 5504-5508.

Taatjes, D.J., and Tjian, R. (2004). Structure and function of CRSP/Med2; a promoter-selective transcriptional coactivator complex. Mol Cell *14*, 675-683.

Tam, W.L., Lim, C.Y., Han, J., Zhang, J., Ang, Y.S., Ng, H.H., Yang, H., and Lim, B. (2008). T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. Stem Cells *26*, 2019-2031.

Thiagalingam, S., Cheng, K.H., Lee, H.J., Mineva, N., Thiagalingam, A., and Ponte, J.F. (2003). Histone deacetylases: unique players in shaping the epigenetic histone code. Ann N Y Acad Sci *983*, 84-100.

Thomas, J.O., and Travers, A.A. (2001). HMG1 and 2, and related 'architectural' DNA-binding proteins. Trends Biochem Sci *26*, 167-174.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.

Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome Res *18*, 1084-1091.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell *10*, 1453-1465.

Trievel, R.C. (2004). Structure and function of histone methyltransferases. Crit Rev Eukaryot Gene Expr *14*, 147-169.

Trounson, A. (2009). New perspectives in human stem cell therapeutic research. BMC Med *7*, 29.

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

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

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol Cell *17*, 453-462.

van den Berg, D.L., Zhang, W., Yates, A., Engelen, E., Takacs, K., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2008). Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. Mol Cell Biol *28*, 5986-5995.

Verdin, E., Dequiedt, F., and Kasler, H.G. (2003). Class II histone deacetylases: versatile regulators. Trends Genet *19*, 286-293.

Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009). Genomic views of distant-acting enhancers. Nature *461*, 199-205.

Waddington, C.H. (1957). The Strategy of the Genes. London: Geo Allen and Unwin.

Wang, H., An, W., Cao, R., Xia, L., Erdjument-Bromage, H., Chatton, B., Tempst, P., Roeder, R.G., and Zhang, Y. (2003). mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. Mol Cell *12*, 475-487.

Wasylyk, B., Wasylyk, C., Augereau, P., and Chambon, P. (1983). The SV40 72 bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. Cell *32*, 503-514.

Wegner, M. (1999). From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res *27*, 1409-1420.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318-324.

Wernig, M., Zhao, J.P., Pruszak, J., Hedlund, E., Fu, D., Soldner, F., Broccoli, V., Constantine-Paton, M., Isacson, O., and Jaenisch, R. (2008). Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease. Proc Natl Acad Sci U S A *105*, 5856-5861.

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.

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

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810-813.

Wobus, A.M., and Boheler, K.R. (2005). Embryonic stem cells: prospects for developmental biology and cell therapy. Physiol Rev *85*, 635-678.

Wu, Q., Chen, X., Zhang, J., Loh, Y.H., Low, T.Y., Zhang, W., Sze, S.K., Lim, B., and Ng, H.H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J Biol Chem *281*, 24090-24094.

Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc Natl Acad Sci U S A *104*, 7145-7150.

Xu, H., and Hoover, T.R. (2001). Transcriptional regulation at a distance in bacteria. Curr Opin Microbiol *4*, 138-144.

Yamanaka, S. (2008). Pluripotency and nuclear reprogramming. Philos Trans R Soc Lond B Biol Sci *363*, 2079-2087.

Yang, X.J., and Seto, E. (2007). HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. Oncogene *26*, 5310-5318.

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

Yeates, T.O. (2002). Structures of SET domain proteins: protein lysine methyltransferases make their mark. Cell *111*, 5-7.

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

Ying, Q.L., and Smith, A.G. (2003). Defined conditions for neural commitment and differentiation. Methods Enzymol *365*, 327-341.

Yoshikawa, Y., Fujimori, T., McMahon, A.P., and Takada, S. (1997). Evidence that absence of Wnt-3a signaling promotes neuralization instead of paraxial mesoderm development in the mouse. Dev Biol *183*, 234-242.

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

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

Zhang, B., Jain, S., Song, H., Fu, M., Heuckeroth, R.O., Erlich, J.M., Jay, P.Y., and Milbrandt, J. (2007). Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome. Development *134*, 3191-3201.

Zhang, J., Tam, W.L., Tong, G.Q., Wu, Q., Chan, H.Y., Soh, B.S., Lou, Y., Yang, J., Ma, Y., Chai, L., *et al.* (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. Nat Cell Biol *8*, 1114-1123.

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

Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W.K., Shahab, A., Kuznetsov, V.A., *et al.* (2007). Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell *1*, 286-298.

Chapter 2

Tcf3 is an Integral Component of the Core Regulatory Circuitry of Embryonic Stem Cells

Published as: Megan F. Cole, Sarah E. Johnstone, Jamie J. Newman, Michael H. Kagey and Richard A. Young (2008). "Tcf3 is an Integral Component of the Core Regulatory Circuitry of Embryonic Stem Cells." Genes Dev. 22: 746-755.

My contribution on this project

The work to study the role of Tcf3 and Wnt signaling in embryonic stem cells was initiated by graduate student Megan Cole but quickly became a close, equal collaboration with graduate student Sarah Johnstone and myself. I led the efforts to establish the embryonic stem cell tissue culture room for the lab that was used to do a majority of this work. I worked out cell culture conditions for the embryonic stem cells and contributed a significant amount conceptually and experimentally to the cell culture aspects of this project, including Tcf3 knockdowns and Wnt3a stimulation.

Abstract

Embryonic stem cells have a unique regulatory circuitry, largely controlled by the transcription factors Oct4, Sox2 and Nanog, which generates a gene expression program necessary for pluripotency and self-renewal (Boyer et al. 2005; Loh et al. 2006; Chambers et al. 2003; Mitsui et al. 2003; Nichols et al. 1998). How external signals connect to this regulatory circuitry to influence embryonic stem cell fate is not known. We report here that a terminal component of the canonical Wnt pathway in embryonic stem cells, the transcription factor Tcf3, co-occupies promoters throughout the genome in association with the pluripotency regulators Oct4 and Nanog. Thus Tcf3 is an integral component of the core regulatory circuitry of ES cells, which includes an autoregulatory loop involving the pluripotency regulators. Both *Tcf3* depletion and Wnt pathway activation cause increased expression of Oct4, Nanog and other pluripotency factors and produce ES cells that are refractory to differentiation. Our results suggest that the Wnt pathway, through Tcf3, brings developmental signals directly to the core regulatory circuitry of ES cells to influence the balance between pluripotency and differentiation.

Introduction

Embryonic stem (ES) cells provide a unique opportunity to study early development and hold great promise for regenerative medicine (Thomson et al. 1998; Reubinoff et al. 2000; Pera and Trounson 2004). ES cells are derived from the inner cell mass of the developing blastocyst and can be propagated in culture in an undifferentiated state while maintaining the capacity to generate any cell type in the body. Discovering how signaling pathways and transcriptional regulatory circuitry contribute to self-renewal and pluripotency is essential for understanding early development and realizing the therapeutic potential of ES cells.

A model for the core transcriptional regulatory circuitry of ES cells has emerged from studying the target genes of the ES cell transcription factors Oct4, Sox2 and Nanog (Boyer et al. 2005; Loh et al. 2006). These master regulators occupy the promoters of active genes encoding transcription factors, signal transduction components and chromatin modifying enzymes that promote ES cell self-renewal. They also occupy the promoters of a large set of developmental transcription factors that are silent in ES cells, but whose expression is associated with lineage commitment and cellular differentiation. Polycomb Repressive Complexes co-occupy the genes encoding these developmental transcription factors to help maintain a silent transcriptional state in ES cells (Boyer et al. 2006; Lee et al. 2006; Wilkinson et al. 2006; Rajaskhar and Begemann 2007; Stock et al. 2007).

External signals can promote ES cell pluripotency or cause these cells to differentiate, but precisely how these pathways are connected to the ES cell regulatory network has not been determined. These signals are produced by the stem cell niche in the developing blastocyst or, for cultured ES cells, can be produced by added factors or serum to maintain stem cell identity or promote differentiation. Recent studies have demonstrated the importance of several signaling pathways in maintaining or modifying ES cell state, including the Activin/Nodal, Notch, BMP4 and Wnt pathways (Rao et al. 2004; Kristensen et al.

2005; Friel et al. 2005; Boiani and Scholer 2005; Valdimarsdottir and Mummery 2005; Dreesen and Brivanlou 2007; Pan and Thomson 2007). By understanding how these signaling pathways influence the gene expression program of ES cells, it should be possible to discover how they contribute to embryonic stem cell identity or promote specific differentiation programs.

The Wnt/β-catenin signaling pathway has multiple roles in embryonic stem cell biology, development and disease (Logan and Nusse 2004; Reya and Clevers 2005; Clevers 2006). Several studies have shown that activation of the Wnt pathway can cause ES cells to remain pluripotent under conditions that induce differentiation (Kielman et al. 2002; Sato et al. 2004; Singla et al. 2006; Hao et al. 2006; Ogawa et al. 2006; Miyabashi et al. 2007; Takao et al. 2007), while other studies have shown that the Wnt pathway has an important role in directing differentiation of ES cells (Otero et al. 2004; Lindsley et al. 2006). Recent studies have shown that T Cell Factor-3 (Tcf3), a terminal component of the Wnt pathway, acts to repress the *Nanog* gene in ES cells (Pereira et al. 2006), providing an important clue for at least one mechanism by which the Wnt pathway regulates stem cell state. Nonetheless, we have an incomplete understanding of how the pathway exerts its effects, in part because few target genes have been identified for its terminal components in ES cells.

Stimulation of the canonical Wnt signaling pathway causes the transcriptional co-activator β-catenin to translocate to the nucleus, where it interacts with constitutively DNA-bound Tcf/Lef proteins to activate target genes (Behrens et al. 1996; Brantjes et al. 2001; Cadigan 2002). Tcf3, a member of the Tcf/Lef family, is highly expressed in murine embryonic stem (mES) cells and is critical for early embryonic development (Korinek et al. 1998; Merrill et al. 2004; Pereira et al. 2006). To determine how the Wnt pathway is connected to the gene expression program of ES cells, we have determined the genome-wide binding profile of Tcf3 and examined how perturbations of the pathway affect the gene expression program. Remarkably, the genome-wide data reveal that Tcf3 co-occupies the ES cell genome with the pluripotency transcription factors Oct4 and Nanog. These and other results reveal that the Wnt pathway brings

developmental signals directly to the core regulatory circuitry of ES cells, which consists of the pluripotency transcription factors and Tcf3, together with their mutual target genes.

Results

Identification of Tcf3 Binding Sites Genome-wide

To determine how the Wnt pathway regulates the gene expression program of murine embryonic stem cells, we first identified genes occupied by Tcf3. Murine embryonic stem cells were grown under standard conditions (Supplemental Fig. S1) and DNA sequences occupied by Tcf3 were identified using chromatin immunoprecipitation (ChIP) combined with DNA microarrays (ChIP-Chip). For this purpose, DNA microarrays were designed with 60-mer oligonucleotide probes tiling the entire non-repeat portion of the mouse genome. The results revealed that Tcf3 occupies over 1000 murine promoters (Supplemental Table S1), including those of the known Wnt targets Axin2 and Myc (Fig. 1A)(He et al. 1998; Yan et al. 2001; Jho et al. 2002).

Tcf3 Co-occupies the Genome with ES Cell Master Regulators

Inspection of the genes occupied by Tcf3 revealed a large set that were previously shown to be bound by the homeodomain transcription factor Oct4 (Boyer et al. 2005; Loh et al. 2006), which is an essential regulator of early development and ES cell identity (Nichols et al. 1998; Hay et al. 2004). To examine the overlap of gene targets more precisely, we carried out ChIP-Chip experiments with antibodies directed against Oct4 in mES cells and used the same genome-wide microarray platform employed in the Tcf3 experiment. Remarkably, the binding profiles of Tcf3 and Oct4 revealed that they bind the same genomic regions and display identical spatial distribution patterns with regards to transcription start sites (Fig. 1B; Supplemental Fig. S2). These results identified a set of 1224 genes that are co-occupied by Tcf3 and Oct4 at high confidence (Supplemental Table S1) and suggested that the Wnt pathway connects directly to genes regulated by Oct4 through Tcf3.

Previous studies in human embryonic stem cells have shown that Oct4 shares target genes with the transcription factors Nanog and Sox2 (Boyer et al. 2005), suggesting that Tcf3-occupied genes in murine ES cells should also be

occupied by Nanog and Sox2. Additional genome-wide ChIP-Chip experiments with antibodies directed against Nanog revealed that it does indeed bind the same sites occupied by Oct4 and Tcf3 (Fig. 1B,C and 2, Supplemental Fig. S2). The fact that Oct4 and Sox2 form heterodimers in ES cells (Dailey and Bascilico 2001; Okumura-Nakanishi et al. 2005) and frequently co-occupy promoters in human ES cells (Boyer et al. 2005) makes it likely that Tcf3 co-occupies much of the genome with Oct4, Nanog and Sox2.

The observation that Tcf3 co-occupies much of the genome with the ES cell pluripotency transcription factors has a number of implications for the regulatory circuitry of these cells. Tcf3 binds its own promoter as well as the promoters of genes encoding Oct4, Sox2 and Nanog (Fig. 2). Thus Tcf3 is an integral component of an interconnected autoregulatory loop, where all four transcription factors together occupy each of their own promoters (Fig. 3A). This feature of ES cell regulatory circuitry was previously described for Oct4, Sox2 and Nanog alone (Boyer et al. 2005) and has been postulated to be a common regulatory motif for master regulators of cell state (Chambers et al. 2003; Okumura-Nakanishi et al. 2005; Rodda et al. 2005; Odom et al. 2004; Odom et al. 2006). Autoregulation is thought to provide several advantages to the control of cell state, including reduced response time to environmental stimuli and increased stability of gene expression (McAdams et al. 1997; Rosenfeld et al. 2002; Shenn-Orr et al. 2002; Thieffry et al. 1998). It is also notable that Tcf3 and the pluripotency transcription factors together occupy genes encoding many Wnt pathway components (Supplemental Fig. S3), suggesting that this transcription factor regulates much of its own signaling pathway apparatus together with the pluripotency factors.

Figure 1

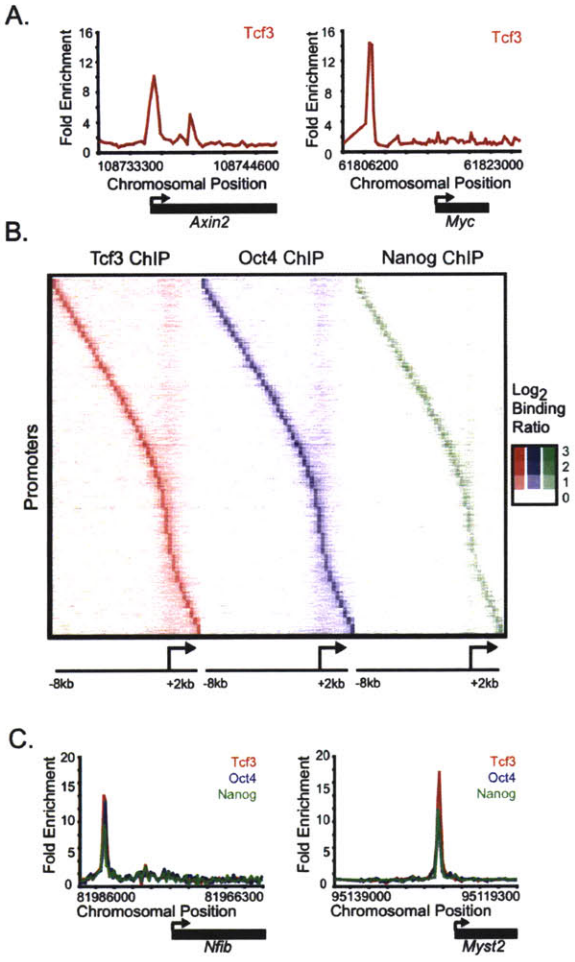

A.
Tcf3
Fold Enrichment
16
12
8
4
0
108733300
108744600
Chromosomal Position
Axin2
Tcf3
Fold Enrichment
16
12
8
4
0
61806200
61823000
Chromosomal Position
Myc
B.
Tcf3 ChIP
Oct4 ChIP
Nanog ChIP
Promoters
Log2 Binding Ratio
3
2
1
0
-8kb
+2kb
-8kb
+2kb
-8kb
+2kb
C.
Tcf3
Oct4
Nanog
Fold Enrichment
20
15
10
5
0
81986000
81966300
Chromosomal Position
Nfib
Tcf3
Oct4
Nanog
Fold Enrichment
20
15
10
5
0
95139000
95119300
Chromosomal Position
Myst2

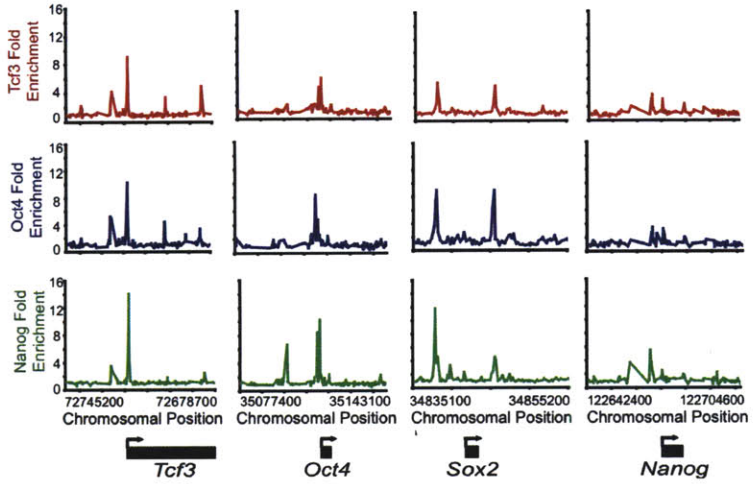
Figure 1. Tcf3, Oct4 and Nanog co-occupy the genome in mouse ES cells.

(A) Tcf3 binds to known target genes. Examples of previously known Tcf3 bound regions are displayed as unprocessed ChIP-enrichment ratios for all probes within the chromosomal region indicated beneath the plot. The gene is depicted below the plot, and the TSS and direction are denoted by an arrow.

(B) Tcf3, Oct4 and Nanog display nearly identical binding profiles. Analysis of ChIP-chip data from genes bound by Tcf3, Oct4, or Nanog reveals that the three factors bind to similar genomic regions at all promoters. Regions from –8kb to +2kb around each TSS were divided into bins of 250bp. The raw enrichment ratio for the probe closest to the center of the bin was used. If there was no probe within 250bp of the bin center then no value was assigned. For genes with multiple promoters, each promoter was used for analysis. The analysis was performed on 3764 genes, which represents 4086 promoters. Promoters are organized according to the distance between the maximum Tcf3 binding ratio and the TSS.

(C) Tcf3, Oct4 and Nanog bind in close proximity at target genes. Plots display unprocessed ChIP-enrichment ratios for all probes within the chromosomal region indicated beneath the plot. The gene is depicted below the plot, and the TSS and direction are denoted by an arrow.

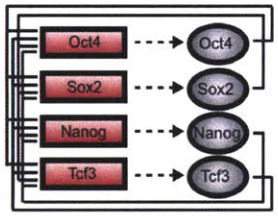
Figure 2



Tcf3, Oct4 and Nanog bind the promoters of *Tcf3*, *Oct4*, *Sox2* and *Nanog*. Plots display unprocessed ChIP-enrichment ratios for all probes within the chromosomal region indicated beneath the plot. The gene is depicted below the plot, and the TSS and direction are denoted by an arrow.

Figure 3

A.


Oct4
Sox2
Nanog
Tcf3
Oct4
Sox2
Nanog
Tcf3

B.

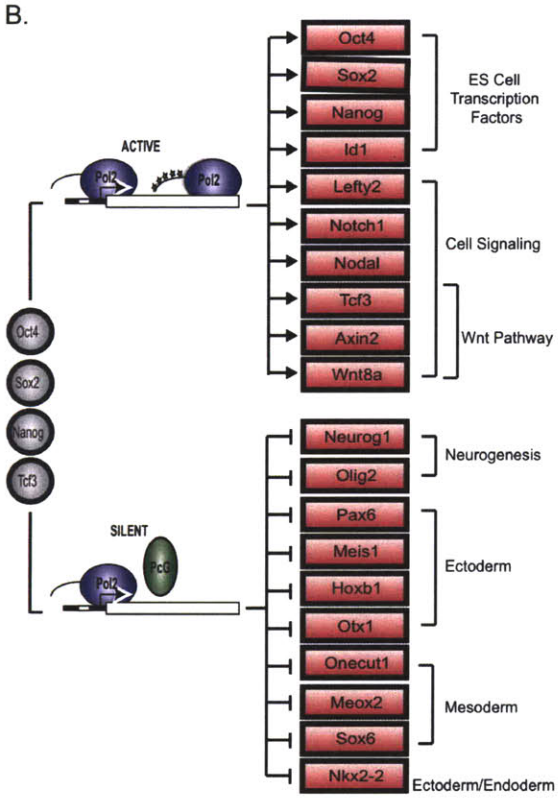

Oct4
Sox2
Nanog
Id1
ES Cell Transcription Factors
ACTIVE
Pol2
Pol2
Lefty2
Notch1
Nodal
Cell Signaling
Tcf3
Axin2
Wnt8a
Wnt Pathway
Oct4
Sox2
Nanog
Tcf3
Neurog1
Olig2
Neurogenesis
SILENT
Pol2
Pax6
Meis1
Hoxb1
Ectoderm
Otx1
Onecut1
Meox2
Sox6
Mesoderm
Nkx2-2
Ectoderm/Endoderm

Figure 3. Tcf3 is an integral component of the core regulatory circuitry of ES cells.

(A) Tcf3 forms an interconnected autoregulatory loop with Oct4, Sox2 and Nanog. Proteins are represented by ovals and genes by rectangles.

(B) Model showing a key portion of the regulatory circuitry of murine embryonic stem cells where Oct4, Sox2, Nanog and Tcf3 occupy both active and silent genes. The evidence that Oct4, Nanog and Tcf3 occupy these genes is described here; Sox2 occupancy is inferred from previous studies in human ES cells (Boyer et al. 2005). Evidence that the transcriptionally silent genes are occupied by Polycomb Repressive Complexes is from Boyer et al. (2006) and unpublished data and that these genes have stalled RNA polymerases is from Guenther et al. (2007) and Stock et al. (2007). Proteins are represented by ovals and genes by rectangles.

A model for the core regulatory circuitry of ES cells has been proposed in which the genes bound by the master regulators Oct4, Sox2 and Nanog fall into two classes: transcriptionally active genes encoding transcription factors, signaling components and other products that support the stem cell state, and transcriptionally inactive genes, consisting mostly of developmental regulators, where Polycomb is bound and RNA polymerase II is recruited, but transcription is stalled (Boyer et al. 2005; Boyer et al. 2006; Lee et al. 2006; Guenther et al. 2007; Stock et al. 2007; Zeitlinger et al. 2007). Our results reveal that Tcf3, together with the pluripotency regulators, is associated with both classes of genes, and thus provide a modified model of the core regulatory circuitry of ES cells (Fig. 3B). The association of Tcf3 with the set of genes encoding key transcription factors, signaling pathway components, and developmental regulators suggests that the Wnt signaling pathway contributes to the regulation of these genes, thereby impacting embryonic stem cell pluripotency and self-renewal.

Expression Analysis of Tcf3 Knockdown in mES Cells

Genes bound by Tcf/Lef proteins are thought to be repressed in the absence of Wnt/β-catenin signaling and to be activated upon Wnt pathway stimulation (Behrens et al. 1996; Brantjes et al. 2001; Miyabayashi et al. 2007; Daniels et al. 2005; Cavallo et al. 1998). Murine ES cells have low endogenous Wnt activity in standard culture conditions and the Wnt pathway can be further stimulated in culture (Dravid et al. 2005; Yamaguchi et al. 2005; Lindsley et al. 2006; Ogawa K et al. 2006; Anton et al. 2007; Takao et al. 2007) (Supplemental Fig. S4). Thus it is unclear whether Tcf3-occupied genes are being repressed or activated at the low level of Wnt activity characteristic of standard ES cell culture conditions. To investigate whether the effect of Tcf3 occupancy is to repress or to activate genes, RNAi constructs were used to deplete *Tcf3* mRNA in mES cells in two independent experiments (Supplemental Fig. S5) and changes in global mRNA levels were assayed with DNA microarrays (Fig. 4A). The ~3.5% of mouse genes whose mRNA levels changed by at least two-fold were significantly

enriched for Tcf3 targets relative to genes whose expression was unaltered by the *Tcf3* knockdown (p value < 2 x 10^{-10}; Supplemental Fig. S6; Supplemental Table S2). The genes whose expression increased upon loss of *Tcf3* included those encoding the master regulators Oct4, Sox2 and Nanog, other genes involved in pluripotency such as Lefty2 and Nodal, and the Wnt pathway component Dkk1 (Fig. 4A). The fact that upregulated genes are strongly enriched for Tcf3 binding suggests that Tcf3 mainly acts to repress genes. Upon loss of *Tcf3*, target genes are no longer repressed and can now be activated by other factors (such as Oct4, Sox2 and Nanog) present at their promoters.

While expression of Tcf3 target genes was often up-regulated upon loss of *Tcf3*, the expression of a substantial number of Tcf3-bound genes remained unchanged, and a relatively small number of Tcf3-bound genes showed reduced expression (Fig. 4A). Nearly half of the genes occupied by Tcf3, Oct4 and Nanog are co-occupied by Polycomb Repressive Complexes (Boyer et al. 2006; Lee et al. 2006; Wilkinson et al. 2006; Rajaskhar and Begemann 2007), and their transcriptional state would not be expected to change as Polycomb would prevent elongation of transcripts at these genes (Stock et al. 2007). Indeed, we find that expression of genes occupied by Tcf3 and Polycomb do not show a significant expression change upon loss of *Tcf3* (p value > 0.4). There were some Tcf3 target genes whose expression was down-regulated upon loss of *Tcf3*; because mES cells have a low level of Wnt pathway activation, it is possible that sufficient β-catenin enters the nucleus in order to associate with and activate this subset of genes. Indeed, we find that some amount of β-catenin does associate with Tcf3 and Oct4 as β-catenin can be detected in crosslinked chromatin extracts immunoprecipitated for either Tcf3 or Oct4 (Supplemental Fig. S7). It is also possible that the loss of expression of this set of Tcf3 target genes is a secondary consequence of the knockdown. The repressive activity of Tcf3 appears to be its dominant function for most genes under these conditions, as the set of Tcf3 bound genes were found to have a significantly higher increase in expression upon knockdown compared to all genes (Fig. 4A; p value < 7 x 10^{-5}).

Expression Analysis of Wnt Pathway Activation in mES Cells

We next studied the effect of increased stimulation of the Wnt pathway on Tcf3 target genes in murine ES cells. Cells were treated with Wnt3a conditioned media in two independent experiments, and changes in global mRNA levels were assayed with DNA microarrays (Fig. 4B). The <1% of mouse genes whose mRNA levels changed by at least two-fold in the Wnt treated cells were significantly enriched for Tcf3 targets relative to genes whose expression was unaltered by the addition of Wnt (p value < 1.5×10^{-5}; Supplemental Fig. S8; Supplemental Table S3). The genes whose expression most increased encode the pluripotency factors Oct4 and Nanog, Wnt pathway components such as Wnt8a and Dkk1, and known Wnt targets such as Brachury (Fig. 4B). These results are consistent with a model where Tcf3 acts to partially repress many of its target genes under standard mES cell culture conditions, yet contributes to increased expression of its target genes under conditions of increased Wnt stimulation. We would therefore expect a correlation between genes up-regulated upon loss of Tcf3 and genes up-regulated upon Wnt stimulation. Indeed, we do find these gene sets to be significantly correlated (p value < 1×10^{-8}; Supplemental Fig. S9). Although a significant portion of Tcf3 target genes undergo expression changes upon Wnt stimulation, it is possible that a second class of Tcf3 target genes are regulated independently of Wnt signaling and therefore are uninfluenced by changes in pathway activation (Yi and Merrill 2007). In fact, several studies have shown a β-catenin independent role for Tcf3 (Kim et al. 2000; Merrill et al. 2001; Roel et al. 2002). It should also be noted that ES cells express other mammalian Tcf/Lef proteins and that these factors may also mediate the functional consequences of Wnt signaling (Pereira et al., 2006).

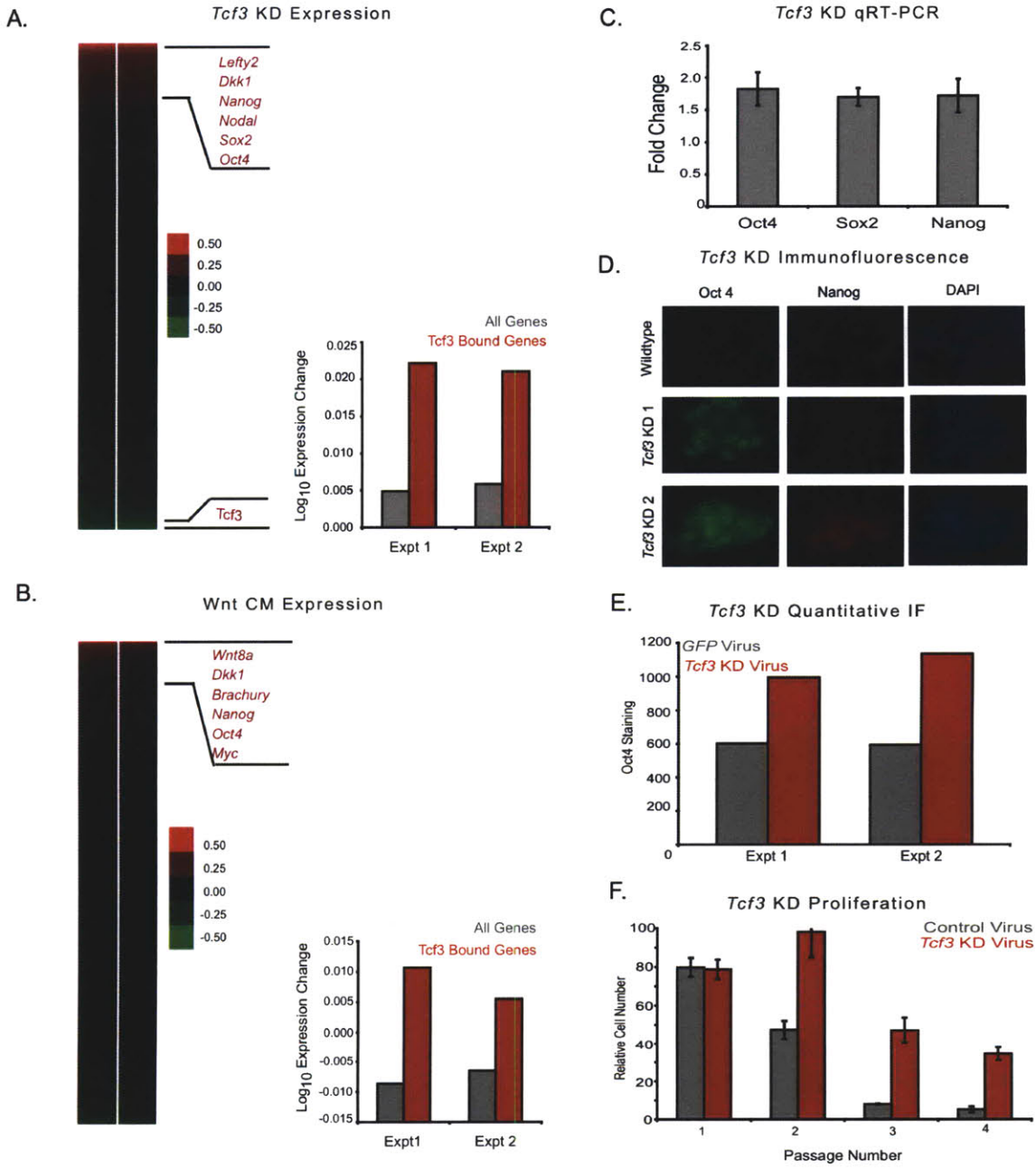
Influence of Tcf3 and Wnt on Pluripotency Regulators and ES Cell State

Evidence that Tcf3 is an integral component of the core transcriptional circuitry of ES cells that functions to partially repress transcription of pluripotency genes led us to examine whether *Tcf3* knockdown enhances features of ES cells associated with pluripotency and self-renewal. Quantitative real-time PCR

analysis demonstrated that *Tcf3* knockdown in mES cells results in higher transcript levels for the pluripotency genes *Oct4*, *Sox2* and *Nanog* (Fig. 4C). Upregulation of *Nanog* upon *Tcf3* depletion confirms a previous report that Tcf3 acts to repress this gene under normal ES cell growth conditions (Pereira et al. 2006). Thus the results of the *Tcf3* knockdown experiment indicate that under normal conditions Tcf3 functions to reduce expression of the three pluripotency regulators.

We next measured the levels of Oct4 and Nanog proteins in ES cells subjected to *Tcf3* knockdown. The results of immunofluorescence experiments show that there are substantial increases in the levels of Oct4 and Nanog transcription factors in the nucleus of such cells (Fig. 4D). There is a significant increase of Oct4 in *Tcf3* knockdown cells compared to control cells based on quantitative measurements of staining intensity using Cellomics software (Fig. 4E). Remarkably, *Tcf3* knockdown mES cells display enhanced proliferation and Oct4 staining in the absence of feeders and LIF compared to control cells, supporting previous results (Fig. 4F; Supplemental Fig. S10)(Pereira et al. 2006). Previous studies have demonstrated that activation of the Wnt/β-catenin pathway can have similar effects on ES cell pluripotency (Sato et al. 2003; Singla et al. 2006; Hao et al. 2006) and we also find that cells treated with Wnt conditioned media show increased staining of Oct4 (Supplemental Fig. S11). The observation that *Tcf3* knockdown and Wnt stimulation have similar functional consequences is consistent with the expression data described above for ES cells subjected to *Tcf3* knockdown and ES cells treated with Wnt3a CM. These studies demonstrate the functional importance of Tcf3 occupancy and Wnt pathway activation for a subset of target genes that includes the pluripotency regulators.

Figure 4


A.
Tcf3 KD Expression
Lefty2
Dkk1
Nanog
Nodal
Sox2
Oct4
Tcf3
0.50
0.25
0.00
-0.25
-0.50
All Genes
Tcf3 Bound Genes
Log10 Expression Change
0.025
0.020
0.015
0.010
0.005
0.000
Expt 1
Expt 2
B.
Wnt CM Expression
Wnt8a
Dkk1
Brachury
Nanog
Oct4
Myc
0.50
0.25
0.00
-0.25
-0.50
All Genes
Tcf3 Bound Genes
Log10 Expression Change
0.015
0.010
0.005
0.000
-0.005
-0.010
-0.015
Expt1
Expt 2
C.
Tcf3 KD qRT-PCR
Fold Change
2.5
2.0
1.5
1.0
0.5
0
Oct4
Sox2
Nanog
D.
Tcf3 KD Immunofluorescence
Oct 4
Nanog
DAPI
Wildtype
Tcf3 KD 1
Tcf3 KD 2
E.
Tcf3 KD Quantitative IF
GFP Virus
Tcf3 KD Virus
Oct4 Staining
1200
1000
800
600
400
200
0
Expt 1
Expt 2
F.
Tcf3 KD Proliferation
Control Virus
Tcf3 KD Virus
Relative Cell Number
100
80
60
40
20
0
1
2
3
4
Passage Number

Knockdown of *Tcf3* and activation of the Wnt pathway in mES cells reveals a role for Tcf3 in repression of target genes and a role in regulating pluripotency.

(A) *Tcf3* knockdown results in up-regulation of target genes. The effect of *Tcf3* knockdown on gene expression was measured by hybridization of labeled RNA prepared from *Tcf3* knockdown cells against RNA prepared from cells infected with non-silencing control lentivirus at 48 hours post infection. A heat map of biological replicate datasets of expression changes was generated where genes are ordered according to average expression change. Tcf3 target genes have a higher average expression change than the average for all genes upon knockdown of *Tcf3.*

(B) Wnt conditioned media (CM) results in up-regulation of Tcf3 target genes. The effect of Wnt activation on gene expression was measured by hybridization of labeled RNA prepared from mES cells grown in Wnt CM against RNA prepared from cells grown in mock CM. A heat map of biological replicate datasets of expression change upon addition of Wnt conditioned media where genes are ordered according to average expression change of replicates. Tcf3 target genes have a higher average expression change than the average for all genes upon addition of Wnt CM.

(C) *Tcf3* knockdown results in increased expression of *Oct4*, *Sox2* and *Nanog*. Real-time PCR demonstrates that *Oct4*, *Sox2* and *Nanog* have increased expression upon knockdown of *Tcf3*. Values are normalized to *Gapdh* transcript levels, and fold change is relative to cells transfected with a non-silencing hairpin.

(D) *Tcf3* knockdown results in increased staining for Oct4 and Nanog. Immunofluorecence was performed on mES cells grown one passage off of feeders that were either infected with *Tcf3* knockdown lentivirus or infected with non-silencing control lentivirus. Cells were fixed with 4% paraformaldehyde 96 hours post-infection. Cells were stained with Oct4, Nanog and DAPI. Images for Oct4 and Nanog staining were taken at 40x magnification and an exposure time

of 300 msec. *Tcf3* KD 1 and 2 represent different knockdown hairpin constructs. *Tcf3* KD 2 is the virus also used in panels 4A,C,E,F.

(E) *Tcf3* knockdown results in a significant increase of Oct4 staining. Quantification of Oct4 staining was performed in cells infected with *Tcf3* or *Gfp* knockdown virus.

(F) *Tcf3* knockdown cells proliferate over more passages in the absence of LIF. Relative cell numbers of ES cells transfected with *Tcf3* or control virus through multiple passages off of feeders in the presence or absence of LIF. Identical cell numbers were initially plated and cells were split 1:12 every 2-3 days. Cells were counted at each passage and values for cells grown in the absence of LIF were normalized to cells grown in the presence of LIF.

Discussion

It is fundamentally important to determine how signaling pathways control ES cell pluripotency and differentiation and how these pathways connect to discrete sets of target genes to affect such states. We have found that a terminal component of the Wnt signaling pathway, the transcription factor Tcf3, is physically associated with the same genomic sites as the pluripotency regulators Oct4 and Nanog in murine embryonic stem cells. This result reveals that the Wnt pathway is physically connected to the core regulatory circuitry of these cells. This core circuitry consists of two key features: an interconnected autoregulatory loop and the set of target genes that are mutually bound by the pluripotency transcription factors and Tcf3.

The genome-wide datasets we report here enhance our knowledge of the targets of Oct4, Nanog and Tcf3. These new datasets were generated using the same protocols and genome-wide tiling microarrays in ES cells grown under identical conditions, allowing more reliable conclusions about the overlap of these factors throughout the genome; previous datasets for these factors came from different murine ES cells grown in different settings, using different chromatin IP analysis platforms, and these data were not always genome-wide (Boyer et al. 2005; Boyer et al. 2006; Loh et al. 2006). The new data reveal, for example, the remarkable extent to which Oct4 and Nanog binding overlap throughout the ES cell genome and the striking association of Tcf3 with those sites (Fig. 1B). The new data also provide a revised model for the core regulatory circuitry of murine ES cells, which incorporates Tcf3 and high confidence target genes of key ES cell regulators (Fig. 3).

The revised model of core regulatory circuitry extends our knowledge of how extracellular signals from the Wnt pathway contribute to stem cell state. Pereira et al. (2006) demonstrated that Tcf3 binds the *Nanog* promoter and represses its mRNA expression in mES cells. Our data confirm Tcf3 binding and function at *Nanog* and extend our knowledge of Tcf3 targets to the other well-characterized pluripotency regulators Oct4 and Sox2, as well as most of their

target genes. Pereira et al. (2006) proposed a model wherein Tcf3-mediated control of Nanog levels allows stem cells to balance the creation of lineage-committed and undifferentiated cells. Our results also support this model, but argue that Tcf3 contributes to the balance through its functions in the core regulatory circuitry described here.

Our results suggest that the Wnt pathway, through Tcf3, influences the balance between pluripotency and differentiation in ES cells, as modeled in Figure 5. Under standard culture conditions, where there is a low-level of Wnt activation, ES cells are poised between the pluripotent state and any of a number of differentiated states. It is well established that Oct4, Sox2 and Nanog act to promote the pluripotent state, as depicted in the model where the influence of these factors is shown by an arrow. Under standard culture conditions, Tcf3 may exist in an activating or repressive complex, but is predominantly in a repressive complex promoting differentiation. The loss of Tcf3 in *Tcf3* knockdown cells, would, in this model, favor pluripotency. Wnt stimulation converts the repressive complex to an activating complex and thus promotes pluripotency. Our results suggest that the Wnt pathway, through Tcf3, influences the balance between pluripotency and differentiation by bringing developmental signals directly to the core regulatory circuitry of ES cells. The observation that the Wnt pathway can be manipulated to affect the balance between pluripotency and differentiation suggests that perturbation of this pathway may impact the efficiency of reprogramming somatic cells into pluripotent stem cells.

Figure 5

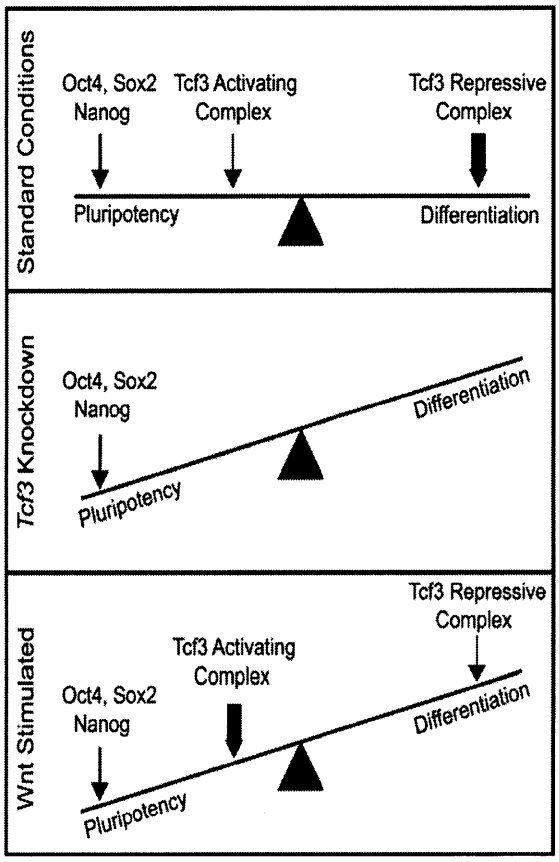

Standard Conditions
Oct4, Sox2
Nanog
Tcf3 Activating
Complex
Tcf3 Repressive
Complex
Pluripotency
Differentiation
Tcf3 Knockdown
Oct4, Sox2
Nanog
Differentiation
Pluripotency
Wnt Stimulated
Tcf3 Repressive
Complex
Tcf3 Activating
Complex
Oct4, Sox2
Nanog
Differentiation
Pluripotency

Figure 5. Model depicting the influence of Wnt pathway components on pluripotency and differentiation in ES cells.

ES cells are poised between the pluripotent state and any of a number of differentiated states. Oct4, Sox2 and Nanog act to promote the pluripotent state (depicted by an arrow). Tcf3 can exist in an activating complex with β-catenin or a repressive complex with Groucho (Reya and Clevers, 2005). Under standard growth conditions, the Wnt pathway is only active at low levels (Dravid et al. 2005; Yamaguchi et al. 2005; Lindsley et al. 2006; Ogawa K et al. 2006; Anton et al. 2007; Takao et al. 2007)(Supplemental Fig. S4). Therefore Tcf3 is mainly in a repressive complex promoting differentiation (depicted by a thick arrow), although some Tcf3 associates with β-catenin to activate target genes and promote pluripotency (depicted by a thin arrow). In *Tcf3* knockdown cells, there is no influence from Tcf3 on cell state. Thus the balance is tipped towards maintaining pluripotency. Upon Wnt stimulation, the balance again tips towards maintaining pluripotency as more Tcf3 associates with β-catenin in an activating complex (depicted by a thick arrow). This model is not meant to imply that Wnt or Tcf3 are themselves pluripotency factors, but rather that they can influence cell state in the presence of other pluripotency factors, such as Oct4, Sox2 and Nanog.

Materials and methods

Mouse embryonic stem cell culture conditions

V6.5 murine ES cells were grown on irradiated murine embryonic fibroblasts (MEFs) unless otherwise stated. Cells were grown under mES cell conditions as previously described in Boyer et al. (2005). Briefly, cells were grown on 0.2% gelatinized tissue culture plates in DMEM-KO (Invitrogen 10829-018) supplemented with 15 % fetal bovine serum (Hyclone, Characterized SH3007103), 1000 Units/mL leukemia inhibitory factor (LIF) (ESGRO ESG1106), 100μM nonessential amino acids (Invitrogen 11140-050), 2mM L-glutamine (Invitrogen 25030-081),100 Units/mL pennicillin and 100 μg/mL streptomycin (Invitrogen 15140-122), and 8 nL/ml 2-mercaptoethanol (Sigma M7522).

Genome-wide location analysis

Chromatin immunoprecipitation protocol

Protocols describing ChIP methods were downloaded from http://jura.wi.mit.edu/young_public/hESregulation/ChIP.html with slight modifications. Briefly, 10^8 mES cells were grown for one passage off of feeders and then crosslinked using formaldehyde. Cells were resuspended, lysed in lysis buffer and sonicated to solubilize and sheer crosslinked DNA. Triton X-100 and SDS were added to the lysate after sonication to final concentrations of 1% and 0.1% respectively. The resulting whole cell extract was incubated at 4°C overnight with 100 uL of Dynal Protein G magnetic beads that had been preincubated with 10 ug of the appropriate antibody overnight. After 16-18 hours, beads were washed with the following 4 buffers for 4 minutes per buffer: low salt buffer (20mM Tris pH 8.1, 150mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS), high salt buffer (20mM Tris pH 8.1, 500mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS), LiCl buffer (10mM Tris pH 8.1, 250mM LiCl, 1mM EDTA, 1% deoxycholate, 1% NP-40), and TE+ 50mM NaCl. Bound complexes were eluted from the beads in elution buffer by heating at 65°C with occasional vortexing, and crosslinks were reversed by overnight incubation at 65°C.

ChIP Antibodies

Cell extracts were immunoprecipitated using antibodies against Tcf3 (Santa Cruz sc-8635), Oct4 (Santa Cruz sc-8628) or Nanog (Bethyl Labs bl1662).

Array Design

The murine 244k whole genome array was purchased from Agilent Technology (www.agilent.com). The array consists of 25 slides each containing ~244,000 60mer oligos (slide ID 15310-3, 15317, 15319-21, 15323, 15325, 15327-30, 15332-7, 15339-41, 15343-44) covering the entire non-repeat portion of the mouse genome at a density of about 1 oligo per 250bp.

Data Normalization and Analysis

Data normalization and analyses were performed as previously described in Boyer et al. (2005).

Tcf3 Knockdown

Lentiviral Production

Lentivirus was produced according to Open Biosystems Trans-lentiviral™ shRNA Packaging System (TLP4614). The shRNA constructs targeting murine *Tcf3* were designed using an siRNA rules based algorithm consisting of sequence, specificity and position scoring for optimal hairpins that consist of a 21 base stem and a 6 base loop (RMM4534-NM-009332). Five hairpin constructs were used to produce virus targeting *Tcf3*. A negative control virus was made from the pLKO.1 empty vector (RHS4080).

Lentiviral Infection of mES Cells

Murine V6.5 ES cells were plated at approximately 30% confluence on the day of infection. Cells were seeded in 2x mES media with 6 ug/ml of polybrene (Sigma H9268-10G) and *Tcf3* knockdown or control (pLKO.1) virus was immediately added. After 24 hours, infection media was removed and replaced with mES

media with 2 ug/ml of Puromycin (Sigma P8833). RNA was harvested at 48 hours after infection.

Knockdown Efficiency

Knockdown efficiency was measured using real-time PCR to measure levels of *Tcf3* mRNA (Supplemental Fig. S5).

RNA Isolation, Real-time PCR and Analysis of Transcript Levels

To determine transcript levels by RT-PCR, RNA was isolated from approximately $10^6 - 10^7$ mES cells using TRIzol reagent following the protocol for cells grown in monolayer (Invitrogen 15596-026). Samples were treated with Dnase I (Invitrogen 18068-015) and cDNA was prepared using SuperScript III reverse transcriptase kit (Invitrogen 180808-051) using oligo dT primed first strand synthesis. Real-time PCR was carried out on the 7000 ABI Detection System using Taqman probes for the housekeeping gene *Gapdh* (Applied Biosystems Mm99999915_g1) as a control and genes of interest (Applied Biosystems; *Tcf3* Mm00493456_m1, *Oct4* Mm00658129_gH, *Sox2* Mm00488369_s1, *Nanog* Mm02384862_g1).

Expression Arrays

Genomic expression analysis was measured using Agilent Whole Mouse Genome Microarrays (Agilent G4122F). 2 ug of RNA was labeled for each sample using the Two-color Low RNA Input Linear Amplification Kit PLUS (Agilent 5188-5340). RNA from the treated sample (either *Tcf3* KD cells or cells treated with Wnt3a conditioned media) were labeled with Cy5 and RNA from control cells (infected with empty-vector virus or a mock conditioned media control, respectively) were labeled with Cy3. Labeled cRNA was hybridized overnight at 65°C. Slides were washed according to the Agilent protocol and scanned on an Agilent DNA microarray scanner BA. Data was analyzed using Agilent Feature Extraction Version 9.5.3 with default settings recommended by

Agilent. Flagged and low-intensity spots were then removed and spots representing a single gene were averaged.

Wnt Pathway Activation

Wnt pathway activity in mES cells was stimulated using Wnt3a conditioned media (ATCC CRL-2647) and mock conditioned media (ATCC CRL-2648) was used as a control. Preparation of conditioned medias was performed as per protocol provided with the cells. Conditioned media was diluted with mES media at a ratio of 1:1.

Immunohistochemical Analysis

Mouse ES cells were crosslinked for 10 minutes at room temperature with 4% paraformaldehyde. Cells were permeabilized with 0.2% Triton X-100 for 10 minutes and stained for Oct4 (Santa Cruz, sc-5279; 1:200 dilution), Nanog (Abcam, ab1603; 1:250 dilution), and DAPI Nucleic Acid Stain (Invitrogen D1306; 1:10000 dilution) overnight at 4°C. After several washes cells were incubated for 2 hours at room temperature with goat-anti mouse conjugated Alexa Fluor 488 (Invitrogen 1:200 dilution) or goat-anti rabbit conjugated Alexa Fluor 568 (Invitrogen 1:200 dilution).

Quantitative Image Acquisition and Data Analysis

Image acquisition and data analysis was performed essentially as described in Moffat et al. (2006). Five days post infection cells were fixed and stained with Oct4 and Hoechst 33342 (1:1000 dilution). Stained cells were imaged on an Arrayscan HCS Reader (Cellomics) using the standard acquisition camera mode (10x objective, 9 fields). Hoechst was used as the focus channel and intra-well focusing was done every 3 fields. The Apotome feature was applied to acquire all images. Objects selected for analysis were identified based on the Hoechst staining intensity using the Target Activation Protocol and the Isodata Threshold method. Parameters were established requiring that individual objects pass an intensity and size threshold. The Object Segmentation Assay Parameter was

adjusted for maximal resolution. Following object selection the average Oct4 intensity was determined and then a mean value for each well was calculated. All wells used for subsequent analysis contained at least 5000 selected objects.

Supplemental Data

Supplemental Data includes nine figures, three tables and Supplemental text and can be found with this article at http://www.genesdev.org/cgi/content/full/22/6/746/DC1.

Accession Numbers

All microarray data from this study are available at ArrayExpress at the EBI (http://www.ebi.ac.uk.arrayexpress) under the accession designation E-TABM-409.

Acknowledgements

We thank Stuart Levine, Alex Marson, Martin Aryee and Sumeet Gupta for experimental and analytical support, Warren Whyte for the *Gfp* lentivirus vector, Roshan Kumar for knockdown and microarray advice, Jennifer Love for microarray advice, Laurie Boyer and Mathias Pawlak for cell culture advice and Tony Lee, Scott McCuine, Brett Chevalier and Rudolph Jaenisch for helpful discussions. Images for immunofluorescence were collected using the W.M. Keck Foundation Biological Imagining Facility at the Whitehead Institute and Whitehead-MIT Bioimaging Center. The SSEA-1 monoclonal antibody developed by D. Solter and B.B. Knowles was obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICGH and maintained by the University of Iowa, Department of Biological Sciences, Iowa City, IA 52242. This work was supported by grants from the NIH and The Whitehead Institute. SJ was supported by an NSF Predoctoral Training Fellowship and MK was supported by an NIH NIGMS Postdoctural Fellowship.

References

Anton, R., Kestler, H.A., and Kuhl, M. 2007. beta-Catenin signaling contributes to stemness and regulates early differentiation in murine embryonic stem cells. FEBS Lett 581: 5247-5254.

Behrens, J., von Kries, J.P., Kuhl, M., Bruhn, L., Wedlich, D., Grosschedl, R., and Birchmeier, W. 1996. Functional interaction of beta-catenin with the transcription factor LEF-1. Nature 382: 638-642.

Boiani, M., and Scholer, H.R. 2005. Regulatory networks in embryo-derived pluripotent stem cells. Nat Rev Mol Cell Biol 6: 872-884.

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

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.

Brantjes, H., Roose, J., van De Wetering, M., and Clevers, H. 2001. All Tcf HMG box transcription factors interact with Groucho-related co-repressors. Nucleic Acids Res 29: 1410-1419.

Cadigan, K.M. 2002. Wnt signaling--20 years and counting. Trends Genet 18: 340-342.

Cavallo, R.A., Cox, R.T., Moline, M.M., Roose, J., Polevoy, G.A., Clevers, H., Peifer, M., and Bejsovec, A. 1998. Drosophila Tcf and Groucho interact to repress Wingless signalling activity. Nature 395: 604-608.

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

Clevers, H. 2006. Wnt/beta-catenin signaling in development and disease. Cell 127: 469-480.

Dailey, L., and Basilico, C. 2001. Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes. J Cell Physiol 186: 315-328.

Daniels, D.L., and Weis, W.I. 2005. Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. Nat Struct Mol Biol 12: 364-371.

Dravid, G., Ye, Z., Hammond, H., Chen, G., Pyle, A., Donovan, P., Yu, X., and Cheng, L. 2005. Defining the role of Wnt/beta-catenin signaling in the survival, proliferation, and self-renewal of human embryonic stem cells. Stem Cells 23: 1489-501.

Dreesen, O., and Brivanlou, A.H. 2007. Signaling pathways in cancer and embryonic stem cells. Stem Cell Rev 3: 7-17.

Friel, R., van der Sar, S., and Mee, P.J. 2005. Embryonic stem cells: understanding their history, cell biology and signalling. Adv Drug Deliv Rev 57: 1894-1903.

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

Hao, J., Li, T.G., Qi, X., Zhao, D.F., and Zhao, G.Q. 2006. WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. Dev Biol 290: 81-91.

Hay, D.C., Sutherland, L., Clark, J., and Burdon, T. 2004. Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem Cells 22: 225-235.

He, T.C., Sparks, A.B., Rago, C., Hermeking, H., Zawel, L., da Costa, L.T., Morin, P.J., Vogelstein, B., and Kinzler, K.W. 1998. Identification of c-MYC as a target of the APC pathway. Science 281: 1509-1512.

Jho, E.H., Zhang, T., Domon, C., Joo, C.K., Freund, J.N., and Costantini, F. 2002. Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. Mol Cell Biol 22: 1172-1183.

Kim CH, Oda T, Itoh M, Jiang D, Artinger KB, Chandrasekharappa SC, Driever W, Chitnis AB. 2000. Repressor activity of Headless/Tcf3 is essential for vertebrate head formation. Nature 407: 913-6.

Kielman, M.F., Rindapaa, M., Gaspar, C., van Poppel, N., Breukel, C., van Leeuwen, S., Taketo, M.M., Roberts, S., Smits, R., and Fodde, R. 2002. Apc modulates embryonic stem-cell differentiation by controlling the dosage of beta catenin signaling. Nat Genet 32: 594-605.

Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O., and Clevers, H. 1998. Two members of the Tcf family implicated in Wnt/beta-catenin signaling during embryogenesis in the mouse. Mol Cell Biol 18: 1248-1256.

Kristensen, D.M., Kalisz, M., and Nielsen, J.H. 2005. Cytokine signalling in embryonic stem cells. Apmis 113: 756-772.

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.

Lindsley, R.C., Gill, J.G., Kyba, M., Murphy, T.L., and Murphy, K.M. 2006. Canonical Wnt signaling is required for development of embryonic stem cell derived mesoderm. Development 133: 3787-3796.

Logan, C.Y., and Nusse, R. 2004. The Wnt signaling pathway in development and disease. Annu Rev Cell Dev Biol 20: 781-810.

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

McAdams, H.H., and Arkin, A. 1997. Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A 94: 814-819.

Merrill, B.J., Gat, U., DasGupta, R., and Fuchs, E. 2001. Lef-1 and Tcf-3 transcription factors mediate tissue specific Wnt signaling during Xenopus development. Genes & Development 15: 1688-1705.

Merrill, B.J., Pasolli, H.A., Polak, L., Rendl, M., Garcia-Garcia, M.J., Anderson, K.V., and Fuchs, E. 2004. Tcf3: a transcriptional regulator of axis induction in the early embryo. Development 131: 263-274.

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

Miyabayashi, T., Teo, J.L., Yamamoto, M., McMillan, M., Nguyen, C., and Kahn, M. 2007. Wnt/beta-catenin/CBP signaling maintains long-term murine embryonic stem cell pluripotency. Proc Natl Acad Sci U S A 104: 5668-5673.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K., et al. 2006. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell 124: 1283-1298.

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

Odom, D.T., Dowell, R.D., Jacobsen, E.S., Nekludova, L., Rolfe, P.A., Danford, T.W., Gifford, D.K., Fraenkel, E., Bell, G.I., and Young, R.A. 2006. Core transcriptional regulatory circuitry in human hepatocytes. Mol Syst Biol 2: 2006 - 0017.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., et al. 2004. Control of pancreas and liver gene expression by HNF transcription factors. Science 303: 1378-1381.

Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D., and Niwa, H. 2006. Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. Biochem Biophys Res Commun 343: 159-166.

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

Otero, J.J., Fu, W., Kan, L., Cuadra, A.E., and Kessler, J.A. 2004. Beta-catenin signaling is required for neural differentiation of embryonic stem cells. Development 131: 3545-3557.

Pan, G., and Thomson, J.A. 2007. Nanog and transcriptional networks in embryonic stem cell pluripotency. Cell Res 17: 42-49.

Pera, M.F., and Trounson, A.O. 2004. Human embryonic stem cells: prospects for development. Development 131: 5515-5525.

Pereira, L., Yi, F., and Merrill, B.J. 2006. Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. Mol Cell Biol 26: 7479-7491.

Rajasekhar, V.K., and Begemann, M. 2007. Concise review: roles of polycomb group proteins in development and disease: a stem cell perspective. Stem Cells 25: 2498-2510.

Rao, M. 2004. Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. Dev Biol 275: 269-286.

Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A., and Bongso, A. 2000. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nat Biotechnol 18: 399-404.

Reya, T., and Clevers, H. 2005. Wnt signalling in stem cells and cancer. Nature 434: 843-850.

Rodda, D.J., Chew, J.L., Lim, L.H., Loh, Y.H., Wang, B., Ng, H.H., and Robson, P. 2005. Transcriptional regulation of nanog by OCT4 and SOX2. J Biol Chem 280: 24731-24737.

Roël G., Hamilton F.S., Gent Y., Bain A.A., Destrée O., and Hoppler S. 2002. Lef-1 and Tcf-3 transcription factors mediate tissue-specific Wnt signaling during Xenopus development. Current Bio 12: 1941-1945.

Rosenfeld, N., Elowitz, M.B., and Alon, U. 2002. Negative autoregulation speeds the response times of transcription networks. J Mol Biol 323: 785-793.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. 2004. 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: 55-63.

Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. 2002. Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet 31: 64-68.

Singla, D.K., Schneider, D.J., LeWinter, M.M., and Sobel, B.E. 2006. wnt3a but not wnt11 supports self-renewal of embryonic stem cells. Biochem Biophys Res Commun 345: 789-795.

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.

Takao, Y., Yokota, T., and Koide, H. 2007. Beta-catenin up-regulates Nanog expression through interaction with Oct-3/4 in embryonic stem cells. Biochem Biophys Res Commun 353: 699-705.

Thieffry, D., Salgado, H., Huerta, A.M., and Collado-Vides, J. 1998. Prediction of transcriptional regulatory sites in the complete genome sequence of Escherichia coli K-12. Bioinformatics 14: 391-400.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. 1998. Embryonic stem cell lines derived from human blastocysts. Science 282: 1145-1147.

Valdimarsdottir, G., and Mummery, C. 2005. Functions of the TGFbeta superfamily in human embryonic stem cells. Apmis 113: 773-789.

Wilkinson, F.H., Park, K., and Atchison, M.L. 2006. Polycomb recruitment to DNA in vivo by the YY1 REPO domain. Proc Natl Acad Sci U S A 103: 19296-19301.

Yamaguchi, Y., Ogura, S., Ishida, M., Karasawa, M., and Takada, S. 2005. Gene trap screening as an effective approach for identifcation of Wnt-responsive genes in the mouse embryo. Dev Dyn 233: 484-95.

Yan, D., Wiesmann, M., Rohan, M., Chan, V., Jefferson, A.B., Guo, L., Sakamoto, D., Caothien, R.H., Fuller, J.H., Reinhard, C., et al. 2001. Elevated expression of axin2 and hnkd mRNA provides evidence that Wnt/beta –catenin signaling is activated in human colon tumors. Proc Natl Acad Sci U S A 98: 14973-14978.

Yi, F., and Merrill, B.J. 2007. Stem cells and TCF proteins: a role for beta catenin--independent functions. Stem Cell Rev 3: 39-48.

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.

Chapter 3

Cell-Type Specific TGF-β Signaling is Targeted to Genes that Control Cell Identity

Submitted for publication: Alan C. Mullen, Jamie J. Newman, David A. Orlando, Roshan M. Kumar, and Richard A. Young

My contributions on this project

This work was led by Alan Mullen, a postdoc in the lab, who I collaborated closely with both intellectually and experimentally. I initiated the culturing of human embryonic stem cells in the lab and worked closely with Alan on the validation and culturing of these cells. I have also contributed to experiments that are still in progress to investigate the shift in Smad3 target genes following the overexpression of MyoD in mouse embryonic stem cells. I worked closely with Alan from the beginning of this project contributing to many of the ideas and experiments conducted in this chapter.

Abstract

Transforming growth factor beta (TGF-β) signaling, which is involved in diverse developmental processes, is effected through the transcription factors Smad2 and Smad3, which require interaction with other transcription factors to stably bind DNA. Cell-type specific variation in the transcription factor milieu has been proposed to play a role in determining the cell-specific effects of TGF-β signaling, but the mechanism by which these Smad proteins exert their broad range of effects remains unclear. We report here that Smad3 co-occupies genomic sites with the master transcription factors that determine cellular identity and thereby transmits signals to the genes associated with these sites. Smad3 co-occupies sites with the master regulators Oct4, Sox2 and Nanog in ES cells, with Myod1 in myotubes, and with PU.1 in pro-B cells. In each cell type, genes bound by these master regulators of cell state are regulated by TGF-β signaling. We conclude that TGF-β signals are directed to unique genomic sites in different cell types through association of Smad proteins with the key transcription factors that determine cellular identity.

Introduction

TGF-β signaling directs essential cellular responses including differentiation, proliferation, cell cycle arrest and migration, and through these responses plays a central role in stem cell biology, development, autoimmunity, tumorigenesis and metastasis (Li and Flavell, 2008; Massague et al., 2005; Padua and Massague, 2009). Activation of the TGF-β receptor leads to phosphorylation of the transcription factors Smad2 and Smad3 (Smad2/3), allowing these proteins to accumulate in the nucleus in association with Smad4, where they regulate transcription of specific genes (Ross and Hill, 2008). Smad3 and the less common isoform of Smad2 can both bind DNA directly through interaction with the Smad binding element (SBE) (Dennler et al., 1998; Shi et al., 1998; Zawel et al., 1998). However, this short binding element is not sufficient for Smad proteins to bind to DNA alone, so Smad proteins must interact with additional transcription factors in order to stably bind DNA (Ross and Hill, 2008; Shi et al., 1998). Much progress has been made in the last decade in understanding how the Smad proteins can interact with a transcription factor to regulate expression of a specific gene in a specific cell type (Choy and Derynck, 2003; Gomis et al., 2006; Hanai et al., 1999; Ross and Hill, 2008; Seoane, 2004). Nonetheless, the mechanisms responsible for the pleiotropic effects of TGF-β signaling in various cell types is poorly understood, in part because only a fraction of Smad target genes have been identified in a limited number of cell types.

While cells express many different transcription factors, cellular identity can be determined by a small number of key transcription factors (Feng et al., 2008; Lassar et al., 1986; Seale et al., 2008; Takahashi and Yamanaka, 2006; Zhou et al., 2008). The concept that a few master transcription factors can control and even reprogram cell identity, and evidence that transcription factors can function to make DNA more accessible to other DNA-binding factors (Barrera and Ren, 2006; Cairns, 2009; Narlikar et al., 2002; Panne, 2008; Roeder, 2005; Segal and Widom, 2009), led us to investigate the hypothesis that Smad3 might

preferentially interact with sites occupied by master regulators in various cell types.

Results and Discussion

TGF-β signaling via Smad2 and Smad3 is required to maintain human embryonic stem (ES) cell identity (Beattie et al., 2005; James et al., 2005; Vallier et al., 2005; Vallier et al., 2009; Xu et al., 2008). If Smad2 and Smad3 preferentially interact with sites occupied by key regulators of cell state, then we would expect that either protein could be found at sites occupied by Oct4, which is a master regulator of ES cells (Chambers, 2004; Hay et al., 2004; Matin et al., 2004; Nichols et al., 1998; Zaehres et al., 2005). To identify the sites occupied by Smad3 and Oct4 in human ES cells, we performed chromatin immunoprecipitation combined with massively parallel DNA sequencing (ChIP-seq) using antibodies against both of these proteins (Fig. 1). The results showed a striking association of Smad3 and Oct4 at well-studied Oct4 target genes (Fig. 1A) and throughout the genome (Fig. 1B, table S1). Over 80% of the 1000 highest confidence Smad3-bound regions were also occupied by Oct4. De-novo DNA sequence motif analysis (Bailey and Elkan, 1994) revealed that the sites occupied by Smad3 were highly enriched in the Oct4 binding sequence (Fig. 1C), and inspection of each bound region revealed that the Smad binding element (SBE) (Ross and Hill, 2008) was also enriched at these sites (Fig. 1D). These results suggest that in human ES cells, Smad3 does not co-occupy DNA with a large variety of transcription factors, but instead predominantly binds DNA sites co-occupied by Oct4 that contain adjacent SBEs.

Human and mouse ES cells are both capable of responding to TGF-β signaling through activation of Smad2 and Smad3 (James et al., 2005; Rossant, 2008). To take advantage of the considerable knowledge of transcriptional circuitry in murine ES cells, where genome-wide binding data is available for many transcription factors (Chen et al., 2008; Kim et al., 2008; Loh et al., 2006; Marson et al., 2008), we performed ChIP-seq to determine if Smad3

Figure 1

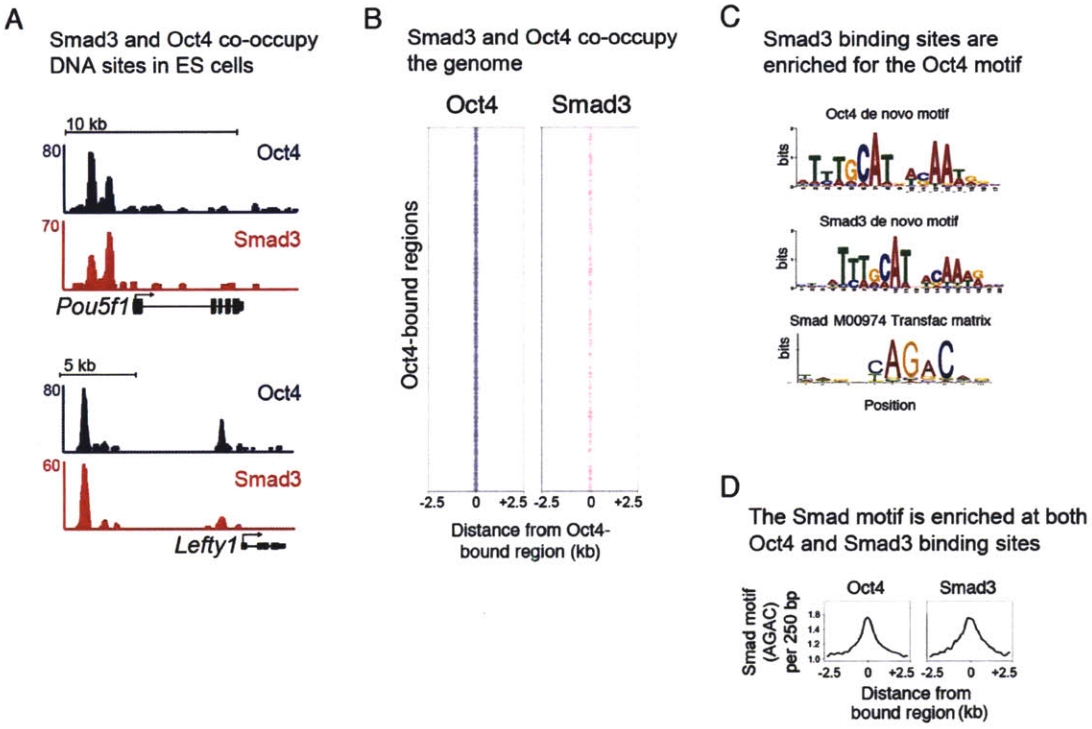

A
Smad3 and Oct4 co-occupy DNA sites in ES cells
10 kb
80
Oct4
70
Smad3
Pou5f1
5 kb
80
Oct4
60
Smad3
Lefty1
B
Smad3 and Oct4 co-occupy the genome
Oct4
Smad3
Oct4-bound regions
-2.5 0 +2.5 -2.5 0 +2.5
Distance from Oct4-bound region (kb)
C
Smad3 binding sites are enriched for the Oct4 motif
Oct4 de novo motif
bits
Smad3 de novo motif
bits
Smad M00974 Transfac matrix
bits
Position
D
The Smad motif is enriched at both Oct4 and Smad3 binding sites
Oct4
Smad3
Smad motif (AGAC) per 250 bp
1.8
1.4
1.2
1.0
-2.5 0 +2.5 -2.5 0 +2.5
Distance from bound region (kb)

Figure 1. Smad3 and Oct4 co-occupy DNA sites in human ES cells.

(A) Gene tracks represent binding of Oct4 (blue) and Smad3 (red) at *Pou5f1*, the gene encoding Oct4 (top) and *Lefty1* (bottom) in human ES cells (BG03) (online). The x-axis represents the linear sequence of genomic DNA and the y-axis in all tracks represents the total number of binding events. The black horizontal bar above the tracks indicate the genomic scale in kilobases (kb), and the gene map is located beneath the tracks. Black boxes represent exons and the arrow indicates the location and direction of transcription initiation.

(B) Smad3 and Oct4 co-occupy the genome. Region plots show the distribution of Oct4- (left) and Smad3- (Amsen et al.) bound regions relative to Oct4-bound regions. For each Oct4-bound region (y-axis) the presence of Oct4-bound regions (left) and Smad3-bound regions (Amsen et al.) are displayed within a 5 kb window centered on the Oct4-bound region. Intensity at the midpoint indicates that bound regions overlap. 7532 regions are occupied by Oct4 and 5282 are occupied by Smad3.

(C) Smad3 binding sites are enriched for the Oct4 motif. Motif discovery was performed using Oct4- and Smad3-bound regions identified by ChIP-seq. The motif enriched at Oct4-bound regions (top) and Smad3-bound regions (center) were identified using MEME (Bailey and Elkan, 1994).

(D) The Smad motif (Smad binding element, SBE) is enriched at both Oct4 and Smad3 binding sites. The histogram shows the frequency of the canonical SBE or the reverse complement (y-axis) relative to the distance from the peak (x-axis) of Oct4 for all Oct4-bound regions (left) and Smad3 for all Smad3-bound regions (Amsen et al.).

co-occupies DNA binding sites with Oct4, Sox2 and Nanog. These three transcription factors make up the core regulatory circuitry of transcription factors in ES cells and are required to maintain ES cell state (Boyer et al., 2005; Chen et al., 2008; Kim et al., 2008; Loh et al., 2006; Marson et al., 2008). We found that Smad2/3 (fig. S1) and Smad3 tend to bind at DNA sites occupied by these master transcription factors in murine ES cells, and this co-occupancy occurs at sites enriched for the SBE (Fig. 2A, B, table S1). Recent evidence of a biochemical interaction between Smad2/3 and Nanog (Vallier et al., 2009) suggests that Nanog may contribute to recruitment of Smad2/3 to these sites. To exclude the possibility that Smad3 was enriched at sites of all highly expressed transcription factors, we asked if Smad3 co-occupied DNA sites with Zfx, which does not co-occupy many sites bound by Oct4 (Chen et al., 2008). While Smad3 was enriched at DNA sites bound by Oct4, Sox2 and Nanog, Smad3 was not enriched at sites bound by Zfx (Fig. 2B). Furthermore, there were fewer SBEs at sites bound by Zfx (Fig. 2B, bottom).

If Smad3 co-occupies DNA with Oct4, Sox2 and Nanog, we would predict that genes bound by these master transcription factors are modulated by TGF-β signaling, while genes that are bound by Zfx are not affected. To test this we performed genome-wide expression analysis under normal culture conditions and in the presence of a TGF-β inhibitor and found that genes bound by Smad3, Oct4, Sox2 and Nanog were significantly affected by inhibition of TGF-β signaling while genes bound by Zfx were not (Fig. 2C, table S2,S3). These results suggest that TGF-β signaling, through activation of Smad2 and Smad3, is directed to the set of genes regulated by the master transcription factors, which are responsible for maintaining ES cell identity.

Figure 2

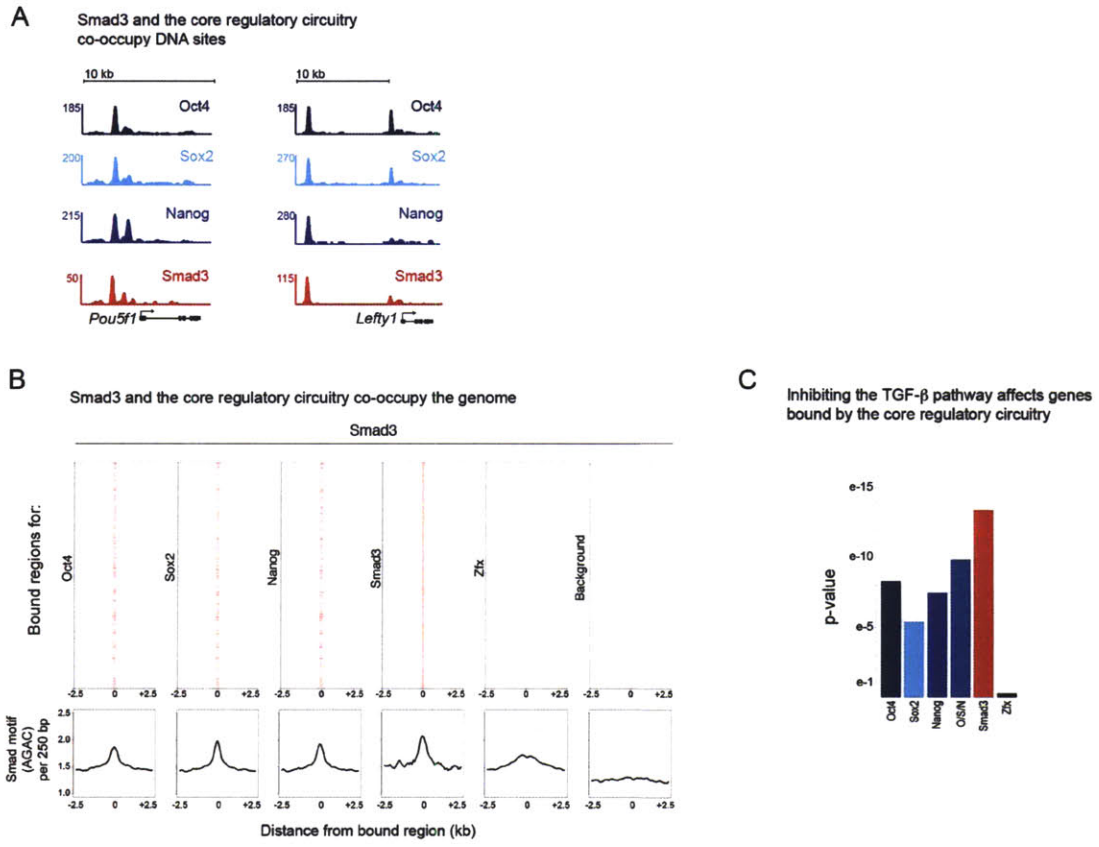

A
Smad3 and the core regulatory circuitry co-occupy DNA sites
10 kb
Oct4
Sox2
Nanog
Smad3
Pou5f1
Lefty1
B
Smad3 and the core regulatory circuitry co-occupy the genome
Smad3
Bound regions for:
Oct4
Sox2
Nanog
Smad3
Zfx
Background
Smad motif (AGAC) per 250 bp
Distance from bound region (kb)
C
Inhibiting the TGF-β pathway affects genes bound by the core regulatory circuitry
p-value
e-15
e-10
e-5
e-1
Oct4
Sox2
Nanog
Oct/N
Smad3
Zfx

Figure 2. Smad3 and the core regulatory circuitry of transcription factors co-occupy DNA binding sites in mouse ES cells.

(A) Gene tracks represent binding of Oct4, Sox2, Nanog (Marson et al., 2008) and Smad3 at *Pou5f1*, the gene encoding Oct4 (left) and at *Lefty1* (Amsen et al.). Similar results were found using an antibody against Smad2/3 (fig. S1).

(B) Smad3 and the core regulatory circuitry co-occupy the genome. Region plots show the distribution of Smad3-bound regions relative to regions bound by Oct4, Sox2, Nanog (Marson et al., 2008), Smad3, Zfx (Chen et al., 2008) and matched background control. The distribution of Smad3-bound regions (red) is shown relative to all bound regions for the indicated transcription factors (y-axis) in a 5 kb window centered on the bound region for each transcription factor. Background represents 10000 DNA regions with the same average distribution relative to the transcription start site (TSS) as the other transcription factors shown. Enrichment of the SBE (Smad motif) relative to the binding site of each transcription factor is indicated below each plot.

(C) Inhibiting the TGF-β pathway affects genes bound by the core regulatory circuitry of transcription factors. Mouse ES cells were cultured under normal conditions or in the presence of the TGF-β inhibitor, SB431542 (James et al., 2005) for 24 hours. The significance of the overlap between genes affected by SB431542 and genes bound by each transcription factor (x-axis) was calculated using the hypergeometric distribution and is plotted on the y-axis. O/S/N represents genes bound by Oct4, Sox2 and Nanog.

We next asked if Smad3 DNA-binding sites are conserved between different cell types or if Smad3 binding sites are cell-type specific. We performed ChIP-seq profiling to identify the genome-wide binding sites of Smad3 in mouse myotubes and pro-B cells and found that Smad3 occupies very different DNA binding sites in each cell type (Fig. 3A, table S1). Only 0.7% of Smad3-bound regions were conserved between ES cells and pro-B cells, 1.1% were conserved between ES cells and myotubes and 3.0% were conserved between myotubes and pro-B cells. These results indicate that there are remarkably few universal Smad3 binding sites and, instead, demonstrate that Smad3 binding is highly cell-type specific.

Motif discovery identified the Oct4 motif in Smad3-bound regions in human ES cells (Fig. 1C), and we used this method to identify transcription factors that might co-occupy Smad3 binding sites in mouse ES cells, myotubes and pro-B cells (Fig. 3B, S2). The Smad3-bound regions in mouse ES cells were most enriched for the Oct4 motif, while in myotubes the Smad3-bound regions were most enriched for an E-box motif (Tapscott, 2005) and in pro-B cells the Smad3-bound regions were most enriched for an Ets motif (Fig. 3B) (Kodandapani et al., 1996).

Myotube formation in C2C12 cells is directed by Myod1, an E-box protein that functions as a key determinant of muscle cell differentiation (Davis et al., 1987; Tapscott, 2005). TGF-β signaling in muscle cells causes atrophy (Zimmers et al., 2002) and has been suggested to function by inhibiting expression of muscle-specific genes (Liu et al., 2001). We used ChIP-seq profiling to identify genome-wide DNA binding targets of Myod1 in C2C12 myotubes to determine if Smad3 and Myod1 occupy the same DNA binding sites. The results demonstrate that Smad3 co-occupies sites bound by Myod1 in myotubes (Fig. 3C,E, table S1). Over 75% of the 1000 highest confidence Smad3-bound regions in myotubes are occupied by Myod1 (Fig. 3F), including the Myod1 enhancer (Fig. 3C) (Tapscott, 2005). Smad3 and Myod1 have been shown to interact biochemically (Liu et al., 2001), suggesting how Smad3 may be

Figure 3

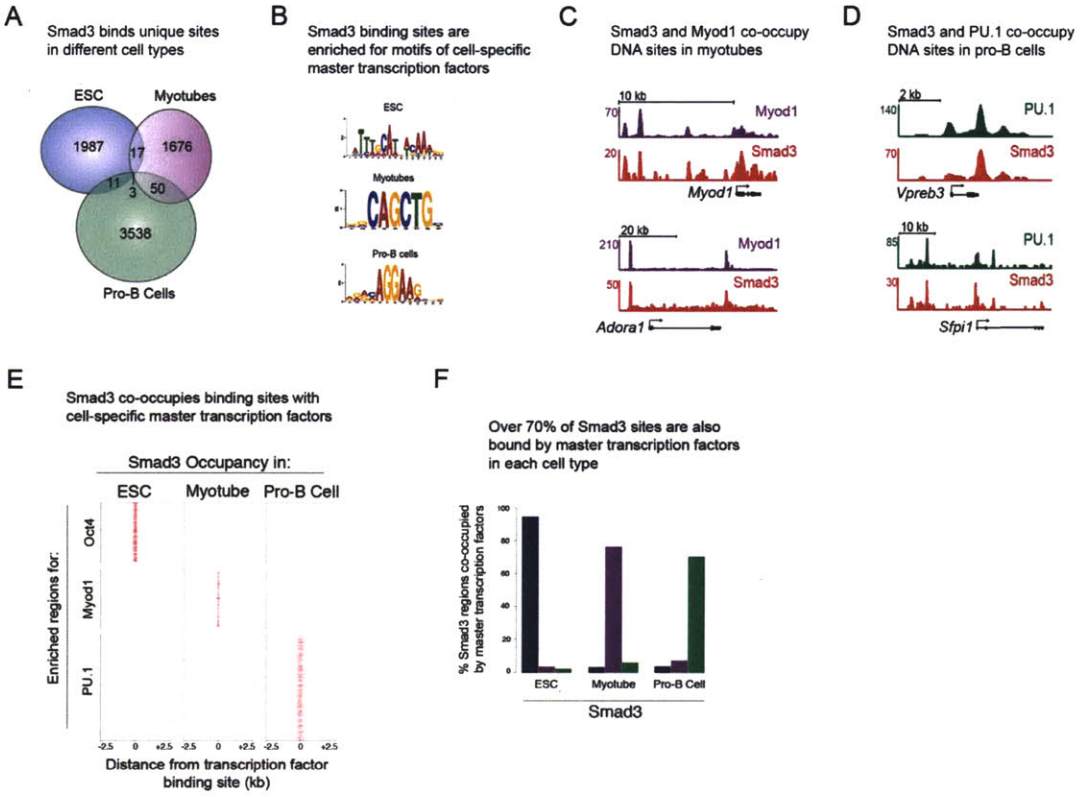

A
Smad3 binds unique sites in different cell types
ESC
Myotubes
1987
17
1676
11
3
50
3538
Pro-B Cells
B
Smad3 binding sites are enriched for motifs of cell-specific master transcription factors
ESC
Myotubes
Pro-B cells
C
Smad3 and Myod1 co-occupy DNA sites in myotubes
10 kb
70
Myod1
20
Smad3
Myod1
20 kb
210
Myod1
50
Smad3
Adora1
D
Smad3 and PU.1 co-occupy DNA sites in pro-B cells
2 kb
140
PU.1
70
Smad3
Vpreb3
10 kb
85
PU.1
30
Smad3
Sfpi1
E
Smad3 co-occupies binding sites with cell-specific master transcription factors
Smad3 Occupancy in:
ESC
Myotube
Pro-B Cell
Enriched regions for:
Oct4
Myod1
PU.1
-2.5
0
+2.5
Distance from transcription factor binding site (kb)
F
Over 70% of Smad3 sites are also bound by master transcription factors in each cell type
% Smad3 regions co-occupied by master transcription factors
100
80
60
40
20
0
ESC
Myotube
Pro-B Cell
Smad3

Figure 3. Smad3 co-occupies DNA with different master transcription factors in different cell types.

(A) Smad3 binds unique sites in different cell types. The Venn diagram shows overlap of Smad3-bound regions between three cell types. ChIP-seq was performed to identify genome-wide binding sites for Smad3 in mouse ES cells, C2C12 cells induced to form myotubes and the pro-B cell line, 38B9 (table S1). The numbers represent the total Smad3-bound regions in each shaded area. C2C12 myotubes and 38B9 cells were stimulated with TGF-β prior to analysis. All ChIP-seq binding data shown for myotubes and pro-B cells contain merged biological replicates.

(B) Smad3 binding sites are enriched for motifs of cell-specific master transcription factors. The top identifiable motifs enriched at the Smad3-bound regions in mouse ES cells (top), myotubes (center) and pro-B cells (bottom) are shown.

(C) Smad3 and Myod1 co-occupy DNA sites in myotubes. Gene tracks represent Myod1 (purple) and Smad3 (red) binding at *Myod1* (top) and *Adora1* (bottom).

(D) Smad3 and PU.1 co-occupy DNA sites in pro-B cells. Gene tracks represent PU.1 (green) and Smad3 (red) binding at *Vpreb3* (top) and *Spfi1* (bottom).

(E) Smad3 co-occupies binding sites with cell-specific master transcription factors. Region plots show the distribution of Smad3-bound regions in mouse ES cell (ESC, left), myotubes (center) and pro-B cells (Amsen et al.) relative to regions bound by Oct4 in ES cells (top), Myod1 in myotubes (middle) and PU.1 (bottom).

(F) Over 70% of Smad3 sites are also bound by the master transcription factors in each cell type. The percentage of Smad3-bound regions (y-axis) that are co-occupied by Oct4, Myod1 and PU.1 are indicated for ESCs, myotubes, and pro-B Cells. Percentages were calculated using the 1000 highest-confidence Smad3-bound regions for each cell-type.

recruited to these sites. These results suggest that the primary effect of TGF-β signaling in myotubes may be directed at genes bound by Myod1.

Pro-B cell development requires the Ets family member, PU.1, for specification and proliferation of early B cell fates (DeKoter and Singh, 2000; Nutt and Kee, 2007), and TGF-β signaling inhibits proliferation in early B cell development (Lee et al., 1989). We performed ChIP-seq profiling to identify DNA sites bound by PU.1 in the pro-B cell line 38B9 and found that Smad3 co-occupies sites bound by PU.1 (Fig. 3D-E, table S1). Over 70% of the 1000 highest confidence Smad3 binding sites in pro-B cells were also bound by PU.1 (Fig. 3F), including *Sfpi1*, the gene encoding PU.1 (Fig. 3D). Thus, Smad3 binds to different sites in three different cell types, and in each case these sites are co-occupied by the master transcription factors specific for that cell type (Fig. 3E). These results suggest that the primary effect of TGF-β signaling may be directed at genes bound by master transcription factors.

Our results indicate that TGF-β signaling regulates genes bound by the master transcription factors in ES cells (Fig. 2C), so we sought to determine whether TGF-β signaling regulates genes bound by the master transcription factors in other cell types. We performed genome-wide expression analysis before and after stimulation with TGF-β in both myotubes and pro-B cells. We found that TGF-β signaling tended to affect a unique set of genes in each cell type (Fig. 4A). If the TGF- β pathway targets genes bound by master transcription factors, we would expect that the genes affected in each cell type would tend to be bound by the cell-specific master transcription factors. Indeed, we found that TGF-β signaling preferentially affected genes bound by Oct4 in ES cells (Fig. 4B, table S2,S3), Myod1 in myotubes (Fig. 4C) and PU.1 in pro-B cells (Fig. 4D). These results demonstrate that Smad3 co-occupies DNA binding sites with transcription factors that control cell state in multiple cell types and directs TGF-β signaling to regulate expression of the genes bound by these master transcription factors.

Why might Smad3 preferentially co-occupy and function at sites with master regulators of cell state? Transcriptional regulators of cell state tend to be

expressed at high levels relative to other transcription factors. These transcription factors can nucleate cooperative binding as well as recruit ATP-dependent chromatin remodeling factors and histone acetyl transferases, which together mobilize and modify nucleosomes, thus providing more access to sequences for other DNA-binding factors (Cairns, 2009; Narlikar et al., 2002; Panne, 2008; Roeder, 2005; Segal and Widom, 2009). Smad binding elements present near the binding sites of master transcription factors may become more available through these mechanisms, providing Smad proteins with both the specific DNA sequences and the transcription factor partners, which are necessary for stable Smad binding.

TGF-β signaling, through activation of Smad2 and Smad3, plays an essential role in normal development and tissue homeostasis as well as in human diseases from cancer to autoimmunity (Li and Flavell, 2008; Massague et al., 2005; Padua and Massague, 2009). It is therefore critical to understand how activation of Smad2 and Smad3 can lead to such diverse cellular responses. Our findings reveal that the cell-type specific effects of TGF-β signaling are determined in large part by the interaction of Smad proteins with one or a small number of master transcription factors that maintain cellular identity. It is by this mechanism that TGF-β signaling is tailored to modulate genes that are most relevant to cell state, which may explain why aberrations in this pathway can have such profound effects in a range of human disease.

Figure 4

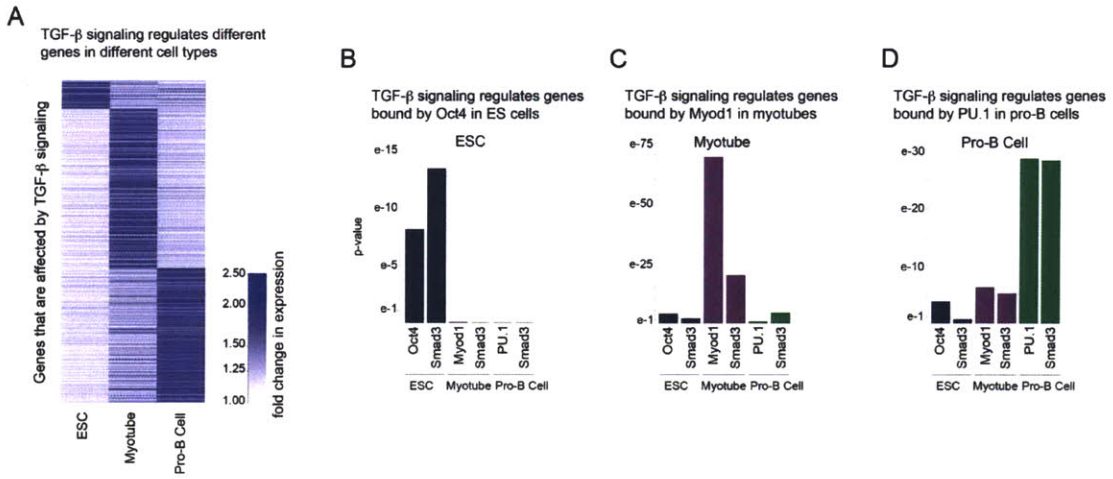

A
TGF-β signaling regulates different genes in different cell types
Genes that are affected by TGF-β signaling
ESC
Myotube
Pro-B Cell
2.50
2.00
1.50
1.25
1.00
fold change in expression
B
TGF-β signaling regulates genes bound by Oct4 in ES cells
ESC
p-value
e-15
e-10
e-5
e-1
Oct4
Smad3
Myod1
Smad3
PU.1
Smad3
ESC
Myotube
Pro-B Cell
C
TGF-β signaling regulates genes bound by Myod1 in myotubes
Myotube
e-75
e-50
e-25
e-1
Oct4
Smad3
Myod1
Smad3
PU.1
Smad3
ESC
Myotube
Pro-B Cell
D
TGF-β signaling regulates genes bound by PU.1 in pro-B cells
Pro-B Cell
e-30
e-20
e-10
e-1
Oct4
Smad3
Myod1
Smad3
PU.1
Smad3
ESC
Myotube
Pro-B Cell

Figure 4. TGF-β signaling regulates genes bound by Smad3 and master transcription factors.

(A) TGF-β signaling regulates different genes in different cell types. Genome-wide expression analysis was performed to compare changes in gene expression 24 hours after inhibition with TGF-β inhibitor (SB431542) in ES cells and 12 hours after activation of TGF- β signaling in myotubes and pro-B cells. All genes affected by TGF-β signaling are plotted on the y-axis. The change in expression for each affected gene is indicated by the intensity of blue color and is shown for ES cells (left), myotubes (center) and pro-B cells (Amsen et al.).

(B) TGF-β signaling regulates genes bound by Oct4 and Smad3 in ES cells. The significance of the overlap of genes affected by TGF- β signaling in ES cells and bound by each transcription factor (x-axis) was calculated using the hypergeometric distribution and is plotted on the y-axis. Genes affected by TGF-β signaling in ES cells tend to be bound by Oct4 and Smad3 in ES cells while genes bound by Smad3 or master transcription factors in other cell types are not affected.

(C) Genes affected by TGF- β signaling in myotubes tend to be bound by Myod1 and Smad3 in myotubes.

(D) Genes affected by TGF- β signaling in pro-B cells tend to be bound by PU.1 and Smad3 in pro-B cells.

Experimental Procedures

Located in the supplemental materials, Appendix A

Acknowledgements

We thank T. Lee for helpful discussions. We are grateful to D. Wotton for the gift of the Smad2/3 antibody and H. Singh for providing the 38B9 cells with permission from N. Rosenberg. We thank J-A. Kwon, J. Love, V. Dhanapal, S. Gupta and T.Volkert for help generating and analyzing ChIP-seq data. This work was supported by a Fellowship from the American Gastroenterological Association (AM) and by NIH grant HG002668 (RY).

References

Amsen, D., Antov, A., Jankovic, D., Sher, A., Radtke, F., Souabni, A., Busslinger, M., McCright, B., Gridley, T., and Flavell, R.A. (2007). Direct regulation of Gata3 expression determines the T helper differentiation potential of Notch. Immunity *27*, 89-99.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.

Barrera, L.O., and Ren, B. (2006). The transcriptional regulatory code of eukaryotic cells--insights from genome-wide analysis of chromatin organization and transcription factor binding. Curr Opin Cell Biol *18*, 291-298.

Beattie, G.M., Lopez, A.D., Bucay, N., Hinton, A., Firpo, M.T., King, C.C., and Hayek, A. (2005). Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. Stem Cells *23*, 489-495.

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

Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193-198.

Chambers, I. (2004). The molecular basis of pluripotency in mouse embryonic stem cells. Cloning Stem Cells *6*, 386-391.

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

Choy, L., and Derynck, R. (2003). Transforming growth factor-beta inhibits adipocyte differentiation by Smad3 interacting with CCAAT/enhancer-binding protein (C/EBP) and repressing C/EBP transactivation function. J Biol Chem *278*, 9609-9619.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.

DeKoter, R.P., and Singh, H. (2000). Regulation of B lymphocyte and macrophage development by graded expression of PU.1. Science *288*, 1439-1441.

Dennler, S., Itoh, S., Vivien, D., ten Dijke, P., Huet, S., and Gauthier, J.M. (1998). Direct binding of Smad3 and Smad4 to critical TGF beta-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene. Embo J *17*, 3091-3100.

Feng, R., Desbordes, S.C., Xie, H., Tillo, E.S., Pixley, F., Stanley, E.R., and Graf, T. (2008). PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. Proc Natl Acad Sci U S A *105*, 6057-6062.

Gomis, R.R., Alarcon, C., He, W., Wang, Q., Seoane, J., Lash, A., and Massague, J. (2006). A FoxO-Smad synexpression group in human keratinocytes. Proc Natl Acad Sci U S A *103*, 12747-12752.

Hanai, J., Chen, L.F., Kanno, T., Ohtani-Fujita, N., Kim, W.Y., Guo, W.H., Imamura, T., Ishidou, Y., Fukuchi, M., Shi, M.J., *et al.* (1999). Interaction and functional cooperation of PEBP2/CBF with Smads. Synergistic induction of the immunoglobulin germline Calpha promoter. J Biol Chem *274*, 31577-31582.

Hay, D.C., Sutherland, L., Clark, J., and Burdon, T. (2004). Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem Cells *22*, 225-235.

James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. (2005). TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development *132*, 1273-1282.

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

Kodandapani, R., Pio, F., Ni, C.Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R.A., and Ely, K.R. (1996). A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. Nature *380*, 456-460.

Lassar, A.B., Paterson, B.M., and Weintraub, H. (1986). Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. Cell *47*, 649-656.

Lee, G., Namen, A.E., Gillis, S., Ellingsworth, L.R., and Kincade, P.W. (1989). Normal B cell precursors responsive to recombinant murine IL-7 and inhibition of IL-7 activity by transforming growth factor-beta. J Immunol *142*, 3875-3883.

Li, M.O., and Flavell, R.A. (2008). TGF-beta: a master of all T cell trades. Cell *134*, 392-404.

Liu, D., Black, B.L., and Derynck, R. (2001). TGF-beta inhibits muscle differentiation through functional repression of myogenic transcription factors by Smad3. Genes Dev *15*, 2950-2966.

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

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

Massague, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes Dev *19*, 2783-2810.

Matin, M.M., Walsh, J.R., Gokhale, P.J., Draper, J.S., Bahrami, A.R., Morton, I., Moore, H.D., and Andrews, P.W. (2004). Specific knockdown of Oct4 and beta2-microglobulin expression by RNA interference in human embryonic stem cells and embryonic carcinoma cells. Stem Cells *22*, 659-668.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.

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

Nutt, S.L., and Kee, B.L. (2007). The transcriptional regulation of B cell lineage commitment. Immunity *26*, 715-725.

Padua, D., and Massague, J. (2009). Roles of TGFbeta in metastasis. Cell Res *19*, 89-102.

Panne, D. (2008). The enhanceosome. Curr Opin Struct Biol *18*, 236-242.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett *579*, 909-915.

Ross, S., and Hill, C.S. (2008). How the Smads regulate transcription. Int J Biochem Cell Biol *40*, 383-408.

Rossant, J. (2008). Stem cells and early lineage development. Cell *132*, 527-531.

Seale, P., Bjork, B., Yang, W., Kajimura, S., Chin, S., Kuang, S., Scime, A., Devarakonda, S., Conroe, H.M., Erdjument-Bromage, H., *et al.* (2008). PRDM16 controls a brown fat/skeletal muscle switch. Nature *454*, 961-967.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends Genet *25*, 335-343.

Seoane, J. (2004). p21(WAF1/CIP1) at the switch between the anti-oncogenic and oncogenic faces of TGFbeta. Cancer Biol Ther *3*, 226-227.

Shi, Y., Wang, Y.F., Jayaraman, L., Yang, H., Massague, J., and Pavletich, N.P. (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. Cell *94*, 585-594.

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

Tapscott, S.J. (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. Development *132*, 2685-2695.

Vallier, L., Alexander, M., and Pedersen, R.A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. J Cell Sci *118*, 4495-4509.

Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W., Cho, C.H., Martinez, A., Rugg-Gunn, P., *et al.* (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development *136*, 1339-1349.

Xu, R.H., Sampsell-Barron, T.L., Gu, F., Root, S., Peck, R.M., Pan, G., Yu, J., Antosiewicz-Bourget, J., Tian, S., Stewart, R., *et al.* (2008). NANOG is a direct

target of TGFbeta/activin-mediated SMAD signaling in human ESCs. Cell Stem Cell *3*, 196-206.

Zaehres, H., Lensch, M.W., Daheron, L., Stewart, S.A., Itskovitz-Eldor, J., and Daley, G.Q. (2005). High-efficiency RNA interference in human embryonic stem cells. Stem Cells *23*, 299-305.

Zawel, L., Dai, J.L., Buckhaults, P., Zhou, S., Kinzler, K.W., Vogelstein, B., and Kern, S.E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. Mol Cell *1*, 611-617.

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

Zimmers, T.A., Davies, M.V., Koniaris, L.G., Haynes, P., Esquela, A.F., Tomkinson, K.N., McPherron, A.C., Wolfman, N.M., and Lee, S.J. (2002). Induction of cachexia in mice by systemically administered myostatin. Science *296*, 1486-1488.

Chapter 4

Mediator and Cohesin Connect Gene Expression and Chromatin Architecture

Submitted for publication as: Michael H. Kagey, Jamie J. Newman, Steve Bilodeau, Ye Zhan, Nynke L. van Berkum, David A. Orlando, Christopher C. Ebmeier, Jesse Goossens, Peter Rahl, Stuart Levine, Dylan J. Taatjes, Job Dekker, and Richard A. Young

My contribution on this project

The work to study the role of Mediator and Cohesin in embryonic stem cells was a project worked on in close collaboration with three postdocs in the lab Michael Kagey, Steve Bilodeau and David Orlando. I worked with Michael Kagey to screen a short hairpin library in mouse embryonic stem cells and initiated the follow-up of Mediator as a result of the hits in that screen. My involvement in this project included ChIP-seq of various factors, knockdown experiments to validate and explore the affects of loss of both Mediator and Cohesin, and analysis of the data. I contributed intellectually to the communication of the concept that Mediator and Cohesin work together to regulate gene expression and to additional experiments presented in this paper.

Abstract

DNA loop formation has been implicated in gene activation and repression, but the processes that regulate DNA loop formation throughout the genome are poorly understood. We report here evidence that Mediator and Cohesin physically and functionally connect the enhancers and core promoters of a key subset of active genes in embryonic stem (ES) cells. Mediator and Cohesin were found to be essential for normal expression of the genes they occupy and for maintenance of ES cell state. The Mediator complex was found to interact with Cohesin and its loading factor Nipbl, which is found predominantly at promoters occupied by Mediator and Cohesin. DNA looping between the enhancers and promoters occupied by Mediator and Cohesin was confirmed using Chromosome Conformation Capture. We propose the interaction between Mediator and Cohesin creates a stable, looped chromatin structure at active promoters throughout the genome, thus generating cell-type specific chromatin architecture.

Introduction

Eukaryotic genomes are packaged into chromatin such that long DNA fibers can be compacted within the nucleus yet allow for replication, transcription and other processes. Much is known about DNA packaging at the level of the nucleosome, the fundamental unit of chromatin, which is composed of an octamer of four core histone proteins around which 147 base pairs of DNA are wrapped (Davey et al., 2002; Luger et al., 1997). Further chromatin compaction occurs, which involves looping of nucleosomal DNA (Finch and Klug, 1976; Olins and Olins, 1974; Ris and Kubai, 1970). Recent studies have identified a variety of looping interactions in the genomes of eukaryotes (Dostie et al., 2006; Fullwood and Ruan, 2009; Hadjur et al., 2009; Kurukuti et al., 2006; Levasseur et al., 2008; Lieberman-Aiden et al., 2009; Nativio et al., 2009; Spilianakis and Flavell, 2004; Tolhuis et al., 2002; Vakoc et al., 2005), but the mechanisms involved in formation of loops remain poorly understood.

Transcriptional regulatory proteins have been proposed to provide a mechanism by which specific loops can be formed in DNA (Adhya, 1989; Bulger and Groudine, 1999; Matthews, 1992; Ptashne, 1986; Saiz and Vilar, 2006; Schleif, 1992; Treisman and Maniatis, 1985). For example, bacteriophage lambda repressor proteins bind cis elements located 3kb apart and can multimerize, forming a DNA loop that has been visualized using electron microscopy (Revet et al., 1999). Bacterial DNA-binding regulators and some RNA polymerase molecules can bind elements located approximately 100 base pairs apart and cause looping of the intervening DNA when the two proteins bind to one another (Borowiec et al., 1987; Dunn et al., 1984; Hahn et al., 1986; Hochschild and Ptashne, 1988; Huo et al., 1988; Popham et al., 1989; Xu and Hoover, 2001). In addition, some bacterial DNA loops are facilitated or stabilized by proteins such as Integration Host Factor (IHF), which can bend DNA as much as 160°, and can thereby enhance gene expression (Xu and Hoover, 2001). These and other studies with bacteria and bacteriophage have thus established the concept that transcriptional regulation often involves DNA loop formation

between specific sites and that some regulatory proteins such as IHF contribute to this process by facilitating or stabilizing loops.

In eukaryotes, transcription factors bind to enhancer elements that can be located some distance (e.g., 1kb to 1Mb) from the core promoter elements where the transcription initiation apparatus is bound (Atchison, 1988; Banerji et al., 1981; Benoist and Chambon, 1981; Levine and Tjian, 2003; Maniatis et al., 1987; Visel et al., 2009; Wasylyk et al., 1983). The enhancer-bound transcription factors bind coactivators such as Mediator and p300, which in turn bind the transcription initiation apparatus (Conaway et al., 2005; Malik and Roeder, 2005, 2008; Roeder, 1998; Taatjes et al., 2004; Visel et al., 2009). This set of interactions, well established in vitro, implies that activation of gene expression is accompanied by DNA loop formation. Indeed, Chromosome Conformation Capture (3C) experiments have confirmed that some enhancers are brought into proximity of the promoter during active transcription (Amano et al., 2009; Carter et al., 2002; Drissen et al., 2004; Hatzis and Talianidis, 2002; Park et al., 2005; Tolhuis et al., 2002; Vernimmen et al., 2007; Wang et al., 2005). If DNA looping does occur between the enhancers and core promoters of active genes, it would be valuable to identify the proteins that play key roles in the formation and stability of such loops.

While screening for genes essential for maintenance of cell state in embryonic stem (ES) cells, we identified components of the Mediator and Cohesin complexes. Both Mediator and Cohesin were found to occupy the enhancers and core promoters of a key subset of actively transcribed genes and to be necessary for normal transcriptional activity in ES cells. The Cohesin loading factor Nipbl was found at the sites co-occupied by Mediator and Cohesin, providing a mechanism for Cohesin loading at these sites. Mediator and Cohesin physically interact, further explaining how the recruitment of Mediator by a transcription factor could lead to association with Cohesin. Chromosome Conformation Capture experiments revealed that the enhancer and core promoter sites occupied by Mediator and Cohesin are brought into close physical proximity, confirming DNA loop formation. Mediator and Cohesin co-occupancy

of the genome was found to be cell-type specific due to cell-type specific gene activity. These and other results lead us to propose that Mediator and Cohesin contribute to a looped and reinforced chromatin structure at active promoters genome-wide, thus generating a cell-type specific chromatin architecture associated with the transcriptional program of each cell. Furthermore, our results provide new insights into Cornelia de Lange syndrome, which affects 1/10,000 children and where most cases are due to mutations in Nipbl (Krantz et al., 2004; Liu and Krantz, 2009; Liu et al., 2009).

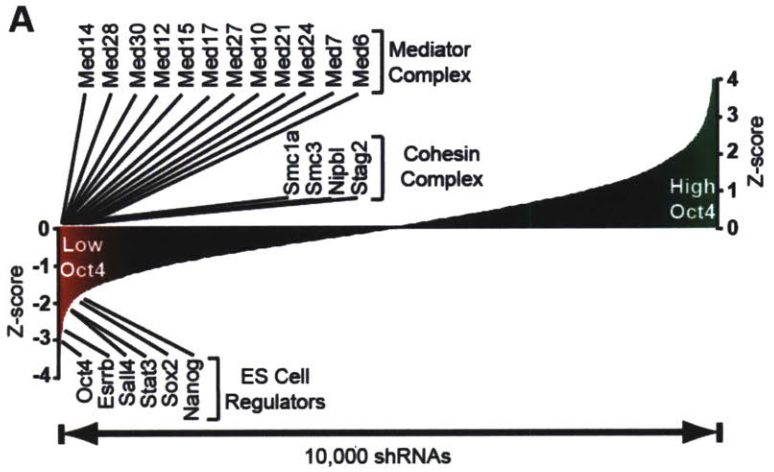
Results

Reduced levels of Mediator and Cohesin Cause Loss of ES Cell State

We used a small hairpin RNA (shRNA) library to screen for regulators of transcription and chromatin necessary for the maintenance of embryonic stem (ES) cell state (Figure 1A; Figure S1). The screen was designed to detect changes in the level of the ES cell transcription factor Oct4, which is a master regulator of the pluripotent state, in cells that remain viable during the course of the experiment. Most known regulators of ES cell state were identified in this screen, including Oct4, Sox2, Nanog, Esrrb, Sall4, and Stat3 (Figure 1A; Table S1 and S2), suggesting that other components identified in this screen may also be important for maintenance of ES cell state. It was particularly striking that many of the subunits of the Mediator complex (Med6, Med7, Med10, Med12, Med14, Med15, Med17, Med21, Med24, Med27, Med28 and Med30), the Cohesin complex (Smc1a, Smc3 and Stag2) and the Cohesin loading factor (Nipbl) emerged from the screen. Mediator and Cohesin are thought to play essential roles in gene expression and chromosome segregation in many different cell types, so their identification in this screen suggests that ES cell state may be highly sensitive to a reduction in the levels of these protein complexes.

The loss of ES cell state is characterized by reduced levels of Oct4 protein, a loss of ES cell colony morphology, reduced levels of mRNAs specifying transcription factors associated with ES cell pluripotency (e.g., Oct4, Sox2 and Nanog) and increased expression of mRNAs encoding developmentally important transcription factors (Bernstein et al., 2006; Boyer et al., 2006; Lee et al., 2006a; Niwa et al., 2000). We confirmed that shRNAs targeting Mediator and Cohesin subunits produced all these effects (Figure 1B and 1C; Figure S1; Table S3). Thus, reduced levels of Mediator and Cohesin have the same effect on these key characteristics of ES cell state as loss of Oct4 itself.

Figure 1


A
Med14
Med28
Med30
Med12
Med15
Med17
Med27
Med10
Med21
Med24
Med7
Med6
Mediator Complex
Smc1a
Smc3
Nipbl
Stag2
Cohesin Complex
High Oct4
Low Oct4
Z-score
Oct4
Esrrb
Sall4
Stat3
Sox2
Nanog
ES Cell Regulators
10,000 shRNAs

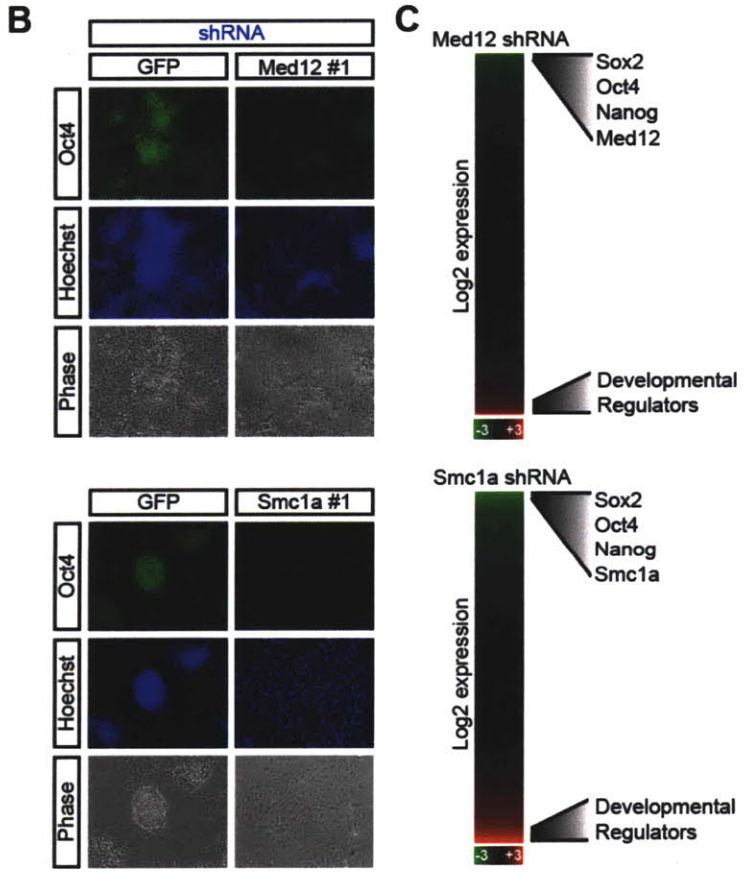

B
shRNA
GFP
Med12 #1
Oct4
Hoechst
Phase
GFP
Smc1a #1
Oct4
Hoechst
Phase
C
Med12 shRNA
Sox2
Oct4
Nanog
Med12
Log2 expression
Developmental Regulators
-3
+3
Smc1a shRNA
Sox2
Oct4
Nanog
Smc1a
Log2 expression
Developmental Regulators
-3
+3

Figure 1. Loss of Mediator and Cohesin Affects ES Cell State.

(A) Components of the Mediator and Cohesin complexes were highly represented in an shRNA screen designed to identify transcriptional and chromatin regulators of ES cell state. The shRNAs were designed against ~2000 transcriptional and chromatin regulators, with ~5 shRNAs/regulator. Murine ES cells were seeded in 384-well plates and infected with individual lentiviral shRNA constructs, fixed five days post infection and quantified for Oct4 staining intensity. To normalize for plate effects, Z-scores for every shRNA were calculated based on average Oct4 staining intensity and rank ordered. Details of the screening protocol are described in Extended Experimental Procedures.

(B) Knockdown of Mediator (top panel) and Cohesin (bottom panel) caused reduced Oct4 protein levels and changes in ES cell colony morphology. Murine ES cells were infected with GFP, Med12 or Smc1a shRNAs, and stained for Oct4 and with Hoechst.

(C) Global gene expression analysis indicates that Mediator and Cohesin knockdowns both cause a decrease in expression of key ES cell regulators and an increase in expression of developmental regulator genes. ES cells were infected with GFP control, Smc1a and Med12 shRNAs. Five days post-infection, gene expression levels relative to the control GFP infection from two biological replicates were determined with Agilent whole genome expression arrays. Log_2 fold expression changes were rank ordered from lowest to highest and are shown for every gene. A relative signal scale is shown at the bottom of both panels.

Mediator is Detected at Enhancers and Core Promoters of Active Genes

Transcription factors bound to enhancers are thought to bind coactivators such as the Mediator complex, which in turn can bind RNA polymerase II at the core promoter (Conaway et al., 2005; Malik and Roeder, 2005, 2008; Roeder, 1998; Taatjes et al., 2004; Visel et al., 2009). This concept suggests that Mediator should be detected at both enhancer and core promoter sites at active genes in vivo. We used chromatin immunoprecipitation coupled with massively parallel DNA sequencing (ChIP- Seq) to identify sites occupied by Mediator subunits Med1 and Med12 in the ES cell genome (Figure 2; Table S4). Med1 and Med12 were studied because they occupy different functional domains within the Mediator complex (Malik and Roeder, 2005; Taatjes and Tjian, 2004). Analysis of the results revealed that Mediator frequently occupied the promoter regions of actively transcribed genes (Figure 2A). We estimate that approximately 79% of genes occupied by Pol2 were co-occupied by Mediator (Table S5; Med12 data). Many of the Pol2 bound genes that had little or no Mediator occupancy play general and essential housekeeping roles in cells (Figure S3).

More detailed examination of the ChIP-Seq data for Mediator with that of key transcription factors (Oct4, Nanog and Sox2) and components of the transcription initiation apparatus (Pol2 and TBP) revealed that Mediator is found at both the enhancers and core promoters of actively transcribed genes (Figure 2B). For example, Mediator was detected at the well-characterized enhancers of the *Oct4 (Pou5f1)* and *Nanog* genes, which are bound by the ES cell master transcription factors Oct4, Sox2 and Nanog (Boyer et al., 2005; Marson et al., 2008; Okumura-Nakanishi et al., 2005; Seila et al., 2008; Wu et al., 2006; Yeom et al., 1996). Mediator was also detected at the *Oct4 (Pou5f1)* and *Nanog* core promoters together with RNA polymerase II and TATA-binding protein (TBP). These observations provide in vivo support for the model that Mediator bridges interactions between transcription factors at enhancers and the transcription initiation apparatus at core promoters.

Cohesin and Mediator Co-occupy Sites Independent of CTCF

Cohesin has been shown to occupy sites bound by CTCF and to contribute to DNA loop formation associated with gene repression or activation (Hadjur et al., 2009; Parelho et al., 2008; Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008). To gain insights into why Cohesin and Mediator knockdowns cause very similar phenotypes in ES cells, we used ChIP-Seq to determine the genome-wide occupancy of the two Cohesin core complex proteins, Smc1 and Smc3 (Table S4). The results show that Cohesin occupies sites bound by CTCF, as expected, but also occupies the enhancer and core promoter sites bound by Mediator (Figure 2B and 2C). Approximately 57% of Cohesin sites were coincident with CTCF sites and 31% were associated with Mediator sites (Figure 2C). The regions co-occupied by Cohesin and Mediator were associated with RNA polymerase II whereas those co-occupied by Cohesin and CTCF were not (Figure 2D). These results demonstrate that a portion of Cohesin is associated with the enhancer and core promoter sites occupied by Mediator in active promoters.

The Cohesin loading factor Nipbl, which was also identified in the shRNA screen, has been implicated in transcriptional regulation and is mutated in the majority of individuals afflicted with Cornelia de Lange Syndrome, a developmental disorder (Figure 1A; Figure S1D and S1F) (Krantz et al., 2004; Tonkin et al., 2004). ChIP-Seq data revealed that Nipbl generally occupies the enhancer and core promoter regions bound by Mediator and Cohesin, but is only rarely found at sites occupied by CTCF and Cohesin but not Mediator (Figure 2B, 2C and 2D; Figure S4A). The association between Nipbl and Mediator/Cohesin sites was highly significant (P-val $<10^{-300}$) whereas the association of Nipbl with CTCF/Cohesin sites was no greater than expected by chance (P-val =1). Thus, the Cohesin loading factor Nipbl is associated with Cohesin/Mediator sites but not with Cohesin/CTCF sites in ES cells.

Gene Regulation Depends on Both Mediator and Cohesin

The presence of Mediator and Cohesin in the promoter regions of *Oct4 (Pou5f1)* and other active ES cell genes (Figure 2A, 2B and 2D) suggests that these

complexes may both contribute to control of transcription. If Mediator and Cohesin function together to regulate the genes they occupy, then we would expect that knockdown of key components of these complexes would have similar effects on expression of these genes. Analysis of changes in mRNA levels in knockdown cells revealed that this is the case (Figure 2E). Of the approximately 4092 genes that are co-occupied by Mediator, Cohesin and Pol2 at high confidence, approximately 1792 showed significant expression changes (P-val <0.01) in both the Mediator and Cohesin knockdown datasets (Figure 2E; Table S3). The two knockdowns had strikingly similar effects at this set of genes (Pearson Correlation of 0.73), which may explain why Mediator and Cohesin knockdowns cause very similar ES cell phenotypes. Similar results were observed at an earlier time point (3 days post knockdown) (Figure S4B), and there was a significant enrichment (P-val $<10^{-87}$) for Mediator/Cohesin co-occupancy at genes with expression changes, indicating that most expression changes were likely a direct result of Mediator and Cohesin knockdown. These results indicate that actively transcribed genes occupied by Mediator and Cohesin typically depend on both complexes for normal expression.

Maintenance of ES cell state is dependent on several key transcription factors, including Oct4 (Nichols et al., 1998; Niwa et al., 2000; Scholer et al., 1990). If the presence of Mediator and Cohesin is dependent on Oct4, then loss of Oct4 should lead to loss of Mediator and Cohesin at sites co-occupied by Oct4, while Cohesin alone should continue to occupy CTCF sites. To test this notion, we utilized a doxycycline-inducible Oct4 shutdown mES cell line (Niwa et al., 2000) and monitored Mediator and Cohesin levels genome-wide by ChIP-Seq (Figure 3A; Table S4; Table S6). At Oct4, Mediator and Cohesin co-occupied sites, Mediator and Cohesin levels were reduced (Figure 3B and 3C). Cohesin continued to occupy most CTCF sites throughout the genome (Figure 3D). These results indicate that the presence of Mediator and Cohesin with Oct4 is dependent on Oct4, that Cohesin's association with CTCF is independent of Oct4

Figure 2

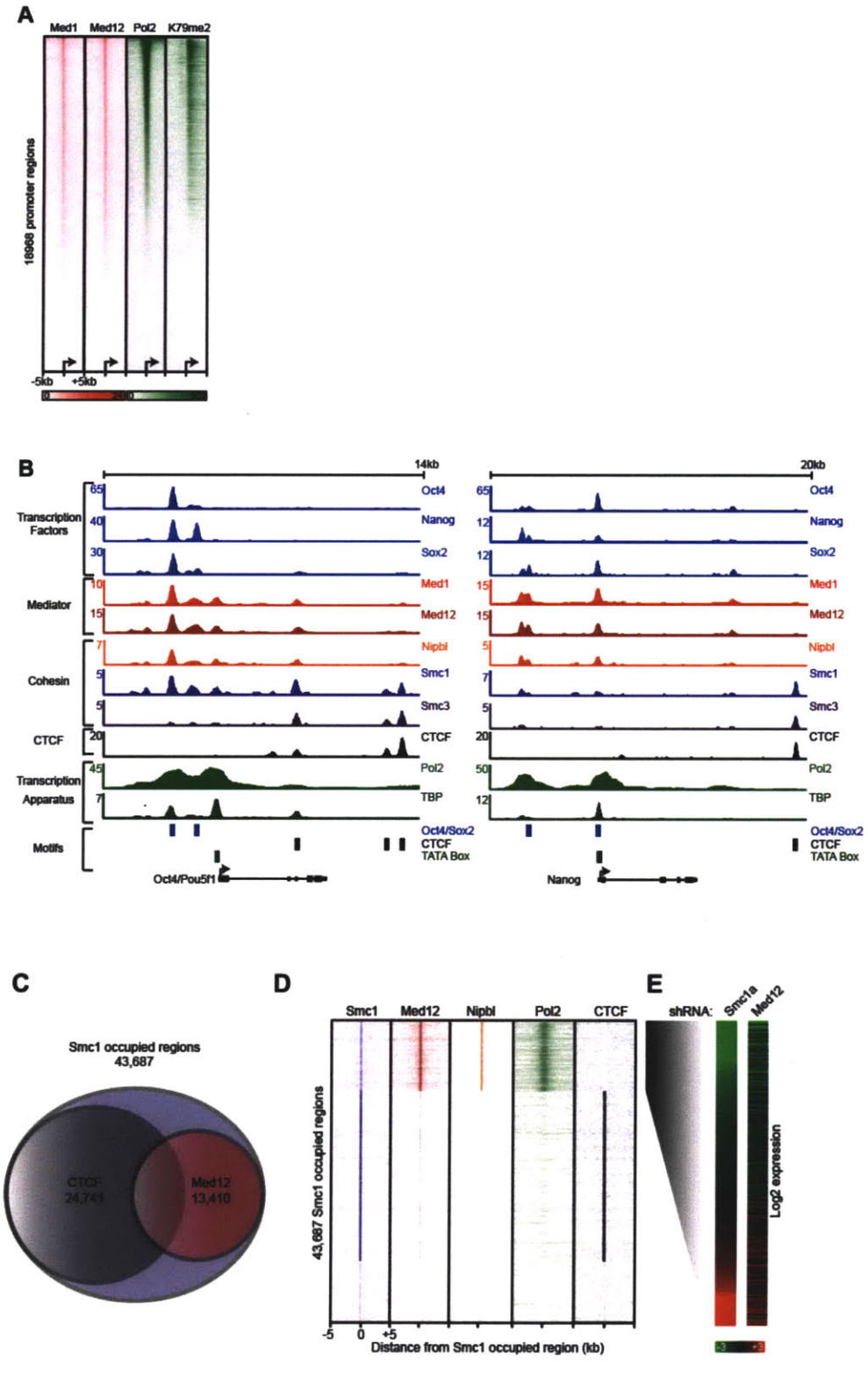

A
Med1
Med12
Pol2
K79me2
18988 promoter regions
-5kb
+5kb
B
14kb
20kb
Transcription Factors
Mediator
Cohesin
CTCF
Transcription Apparatus
Motifs
Oct4
Nanog
Sox2
Med1
Med12
Nipbl
Smc1
Smc3
CTCF
Pol2
TBP
Oct4/Sox2
CTCF
TATA Box
Oct4/Pou5f1
Nanog
C
Smc1 occupied regions
43,687
CTCF
24,741
Med12
13,410
D
Smc1
Med12
Nipbl
Pol2
CTCF
43,687 Smc1 occupied regions
-5
0
+5
Distance from Smc1 occupied region (kb)
E
shRNA:
Smc1a
Med12
Log2 expression

Figure 2. Mediator and Cohesin Genome-wide Occupancy in ES cells.

(A) Density map of ChIP-Seq results for Mediator (Med1 and Med12), RNA polymerase II (Pol2) and di-methylated histone H3 lysine 79 (K79me2) demonstrates Mediator occupancy at genes that are actively transcribed in ES cells. Normalized read counts are shown for 10kb surrounding 18,968 Refseq promoters (from -5kb to +5kb) sorted by maximum level of Pol2 enrichment. A relative signal scale (reads/million) and the position of the transcription start site are shown at the bottom of the panel. Reads from two biological replicates for the Med1 and Med12 ChIP-Seq datasets were combined. The specificity of the Mediator antibodies was validated as shown in Figure S2.

(B) Binding profiles for ES cell transcription factors (Oct4, Nanog and Sox2), Mediator (Med1 and Med12), Cohesin (Smc1, Smc3 and Nipbl), CTCF and components of the transcription apparatus (Pol2 and TBP) at the *Oct4 (Pou5f1)* and *Nanog* loci. ChIP-Seq data is shown in reads/million with the base of the y-axis set to 0.5 reads/million. Oct4/Sox2, CTCF and TBP (TATA Box) binding motifs are shown as blue, black and green boxes respectively. The transcription start site and direction of transcription are noted by an arrow. Reads from two biological replicates for the Smc1, Smc3, Med1, Med12, Nipbl, Pol2 and TBP ChIP-Seq datasets were combined. The specificity of the Mediator, Cohesin and Nipbl antibodies was validated as shown in Figure S2.

(C) Venn diagram showing the overlap of high confidence (P-val $<10^{-9}$) Cohesin occupied sites with CTCF and Mediator (Med12). Approximately 57% of the genomic sites occupied by the Cohesin protein Smc1 were co-occupied by CTCF, ~31% were co-occupied by Mediator.

(D) Region map demonstrating that Smc1, Nipbl and Med12 co-occupied sites generally occur in the absence of CTCF occupancy and within close proximity to RNA polymerase II (Pol2). For each Smc1 occupied region, the presence of Med12, Nipbl, Pol2 and CTCF occupancy is indicated within a 10kb window centered on the Smc1 occupied site.

(E) Heat map indicating that regions co-occupied by Smc1 and Med12 are associated with active genes that exhibit similar expression changes with either

Smc1a or Med12 knockdown (5 days post knockdown). Log_2 expression data from biological replicates is shown for all Smc1 and Med12 co-occupied regions that could be mapped to a gene. Mapped genes have evidence of a co-occupied Smc1/Med12 region within the gene body or within 10kb upstream of the transcriptional start site, evidence of Pol2 occupancy within the gene body and significant (P-val <0.01) expression changes in both an Smc1a and Med12 knockdown in independent experiments. The log_2 expression data was ordered based on the Smc1a knockdown and the corresponding expression change for each gene following a Med12 knockdown is shown. A relative signal scale for the expression data is shown at the bottom of both panels.

Figure 3

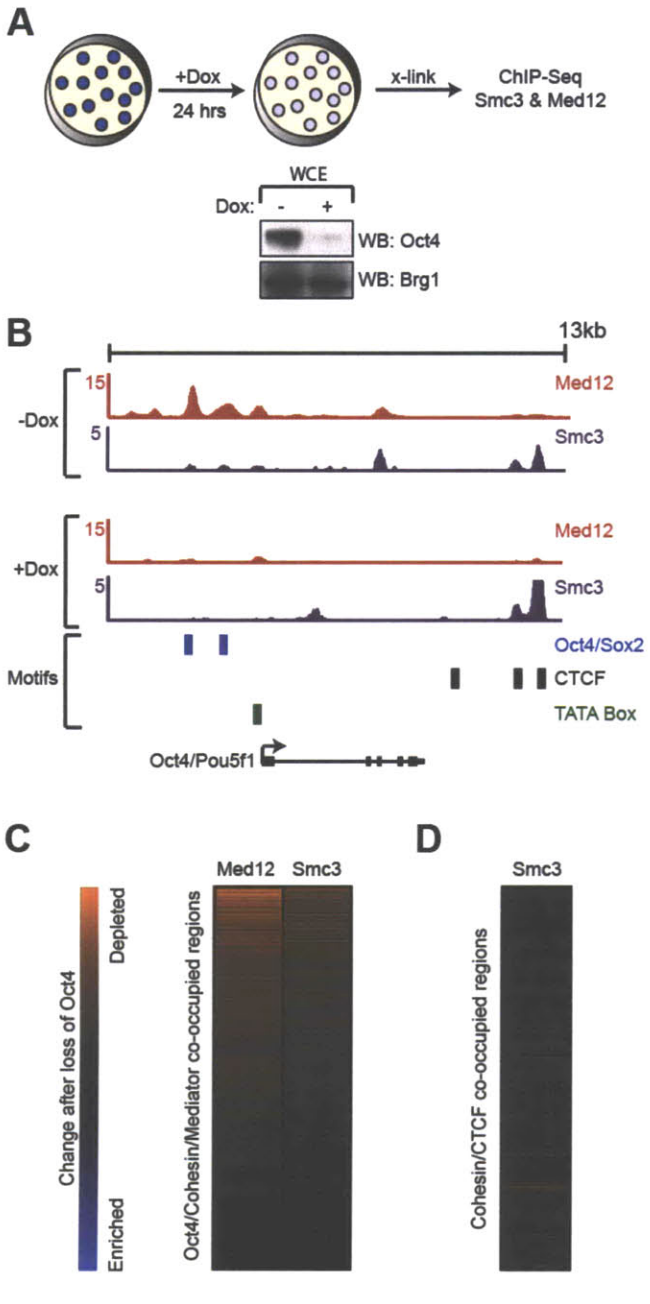

A
+Dox
24 hrs
x-link
ChIP-Seq
Smc3 & Med12
WCE
Dox: - +
WB: Oct4
WB: Brg1
B
13kb
-Dox
15
Med12
5
Smc3
+Dox
15
Med12
5
Smc3
Motifs
Oct4/Sox2
CTCF
TATA Box
Oct4/Pou5f1
C
Med12 Smc3
Change after loss of Oct4
Depleted
Enriched
Oct4/Cohesin/Mediator co-occupied regions
D
Smc3
Cohesin/CTCF co-occupied regions

Figure 3. Mediator and Cohesin Occupancy of Oct4 Regulated Promoters

(A) Schematic of inducible Oct4 shutdown experiment. ES cells were treated with doxycycline to induce shutdown of the *Oct4 (Pou5f1)* gene, cells were crosslinked 24 hours later and ChIP-Seq experiments were performed. A western blot shows a significant drop in Oct4 protein levels 24 hours after doxycycline treatment.

(B) ChIP-Seq data demonstrating that Mediator and Cohesin occupancy are not detectable at the *Oct4 (Pou5f1)* promoter following Oct4 shutdown by doxycycline treatment. ChIP-Seq data for Mediator (Med12) and Cohesin (Smc3) at the *Oct4 (Pou5f1)* locus is shown for wildtype ES cell (-Dox) and ES cells treated with doxycycline (+Dox). The data is displayed in reads/million with the base of the y-axis set to 0.5 reads/million. Oct4/Sox2, CTCF and TBP (TATA Box) binding motifs are shown as blue, black and green boxes respectively. The transcription start site and direction of transcription are noted by an arrow.

(C) Heat maps showing that Mediator and Cohesin occupancy is reduced at sites co-occupied with Oct4 following loss of Oct4 protein. Change in occupancy for Mediator and Cohesin was calculated as the $\log_2$ ratio of the normalized total read densities in each co-occupied region between untreated ES cells and ES cells 24 hours post doxycycline treatment. A relative signal scale for occupancy change is shown.

(D) Cohesin occupancy at Cohesin/CTCF co-occupied regions is not reduced following doxycycline treatment to deplete Oct4 protein. Heat maps were generated as described for (C), except that the regions examined were those co-occupied by Cohesin/CTCF in untreated ES cells.

and that Cohesin-CTCF co-occupancy continues to occur throughout the genome even when cells have lost all the key features of ES cell state.

Evidence for Interaction Between Mediator and Cohesin

The ChIP-Seq results show that Mediator and Cohesin co-occupy thousands of sites in the ES cell genome and thus suggest that these complexes may physically interact. To investigate this possibility, we crosslinked ES cells using the ChIP protocol, immunoprecipitated complexes using antibodies against Mediator (Med1 and Med12) and Cohesin (Smc1, Smc3) and determined whether the Mediator subunit Med23 could be detected in the immunoprecipitate. (Figure 4A). The results showed that Mediator and Cohesin components can coprecipitate with one another. Oct4 is present at many Mediator/Cohesin co-occupied sites in ES cells, so it might be expected to coprecipitate with Mediator and Cohesin. Indeed, Oct4 was detected in both Mediator and Cohesin immunoprecipitates (Figure 4A). Similar results were obtained when extracts were prepared from cells that had not been subjected to crosslinking (Figure 4B) or were subject to DNase I treatment (Figure S5A). Furthermore, an antibody against the Cohesin loading factor Nipbl co-precipitated both Cohesin and Mediator subunits (Figure 4C). These results suggest that Mediator and Cohesin interact.

We obtained additional evidence for a Mediator and Cohesin interaction by using the purification protocol outlined in Figure 4D. Mediator was affinity purified from ES cell nuclei using the activation domain of SREBP-1a, which is known to bind Mediator (Toth et al., 2004; Yang et al., 2006). Following elution from the SREBP-1a affinity resin, the Mediator-containing eluate was subjected to a second, orthogonal immunoprecipitation step, with an anti-CDK8 antibody resin (Figure 4D). CDK8 is a Mediator specific subunit, which ensured that Mediator and Mediator-associated factors would be specifically retained on the antibody column. After binding, the CDK8 antibody resin was subjected to a series of high-salt washes, bound proteins were then eluted and examined by silver stain and western blot. The results show that Cohesin and Nipbl co-

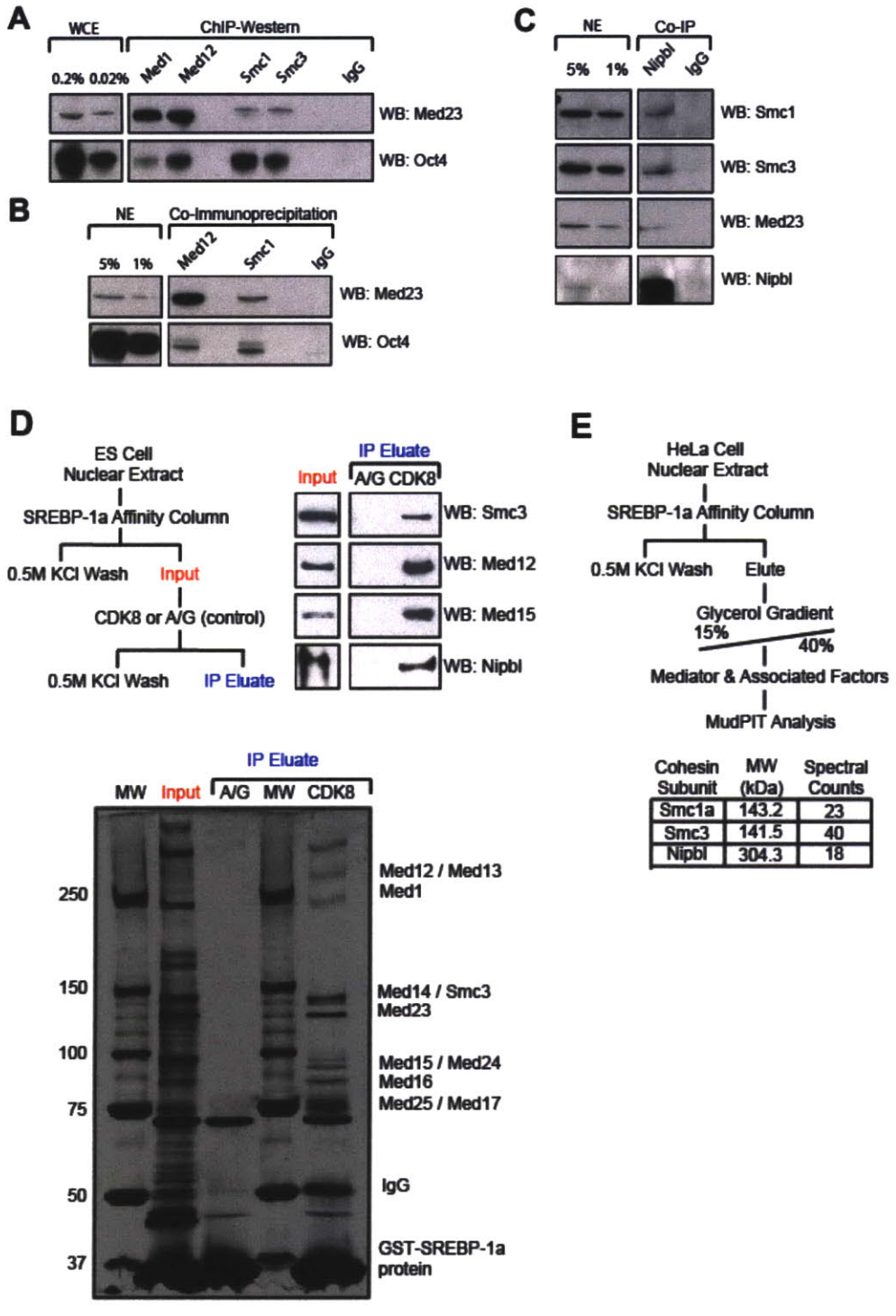
purified with Mediator following the second orthogonal purification step (Figure 4D). Similar results were obtained with an anti-Med1 antibody (Figure S5B and S5C). These results indicate that Mediator and Cohesin physically interact and suggest that this interaction accounts for their co-occupancy at active promoters in vivo.

Mediator and Cohesin are expressed in all cell types, so we purified the Mediator complex from HeLa cells using a different protocol and subjected the complex to analysis using Multidimensional Protein Identification Technology (MudPIT, Figure 4E), which combines multidimensional liquid chromatography with electrospray ionization tandem mass spectrometry to give an unbiased analysis of proteins within the sample (Wolters et al., 2001). In addition to the consensus Mediator subunits (Sato et al., 2004), all of which were identified in this preparation, the Cohesin subunits Smc1a, Smc3 and Nipbl were each well represented in the sample (Figure 4E). Collectively, the results in Figure 4 and Figure S5 confirm that Mediator can be purified in association with Cohesin and Nipbl and show that such complexes occur in multiple vertebrate cell types.

Mediator and Cohesin Binding Profiles Predict DNA Looping Events

Mediator bridges interactions between transcription factors and the transcription apparatus (Conaway et al., 2005; Malik and Roeder, 2005, 2008; Roeder, 1998; Taatjes et al., 2004; Visel et al., 2009). Cohesin has been shown to contribute to transcriptional regulation at specific loci by altering chromosomal architecture (Hadjur et al., 2009; Parelho et al., 2008; Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008). Our evidence shows that Mediator and Cohesin occupy the enhancer and core promoter regions of a set of active genes, suggesting that they contribute to DNA looping between the enhancer and core promoter of these genes. We selected four different loci, *Phc1, Nanog, Oct4 (Pou5f1)* and *Lefty1,* to test enhancer-promoter interaction frequencies in ES cells and in Murine Embryonic Fibroblasts (MEFs). These genes were selected because Mediator and Cohesin occupy their enhancer and core promoter regions in ES cells, where they play a positive role in their transcription, whereas

Figure 4



Cohesin Subunit	MW (kDa)	Spectral Counts
Smc1a	143.2	23
Smc3	141.5	40
Nipbl	304.3	18

Figure 4. Mediator and Cohesin Interact.

(A) Mediator (Med23) and Oct4 are detected by western blot analysis when crosslinked, sheared chromatin is subjected to immunoprecipitation with antibodies directed against Mediator (Med1 or Med12) or Cohesin (Smc1 or Smc3).

(B) Mediator (Med23) and Oct4 are detected by western blot analysis following immunoprecipitation of uncrosslinked ES cell nuclear extracts (NE) with Smc1 or Med12 antibodies.

(C) Cohesin (Smc1, Smc3) and Mediator (Med23) are detected by western blot analysis following immunoprecipitation of uncrosslinked ES cell nuclear extracts with a Nipbl antibody.

(D) Cohesin and Nipbl co-purifies with a Mediator complex. The Mediator complex was initially affinity purified from an ES nuclear extract utilizing a SREBP-1a activation domain. The eluted (Input) material was further purified over an anti-CDK8 antibody resin. The IP Elution was subjected to silver staining and western blot analysis.

(E) Multidimensional Protein Identification Technology (MudPIT) identifies co-purifying Cohesin subunits (Smc1 and Smc3) and Nipbl with the Mediator complex. The Mediator complex was affinity purified from a HeLa cell nuclear extract, separated by glycerol gradient and subjected to MudPIT analysis. Each consensus Mediator subunit was identified in this analysis, as were Smc1a, Smc3 and Nipbl. Spectral counts for Smc1A, Smc3 and Nipbl are shown, after applying 1% false discovery rate threshold are shown.

Mediator and Cohesin are not present at these genes in MEFs, where these genes are transcriptionally silent (see below).

We utilized Chromosome Conformation Capture (3C) technology (Dekker, 2006; Dekker et al., 2002) to determine whether a looping event could be detected between the enhancer and promoter of *Phc1, Nanog, Oct4 (Pou5f1)* and *Lefty1* loci in both ES cells and MEFs (Figure 5 and Figure S6). Crosslinked cells were restriction digested and religated, under conditions that favor intramolecular ligation events, to capture distal DNA fragments that interact within close proximity. We then used semi-quantitative PCR to detect the interaction frequency of distal fragments with a specific anchoring point (upstream enhancer for *Phc1, Oct4 and Lefty,* and core promoter for *Nanog*). The interaction frequency was normalized to a BAC template and control regions (Dekker, 2006). For all loci tested we observed an increased interaction frequency between the core promoter and the enhancer in ES cells, indicating the presence of a DNA loop (Figure 5 and Figure S6). Importantly, this interaction was not observed in MEFs where *Phc1, Nanog, Oct4 (Pou5f1)* and *Lefty1* are silent. Knockdown of either Cohesin subunit Smc1a or Mediator subunit Med12 caused a reduction in the interaction frequency at *Nanog* (Figure S6E). These 3C results are consistent with a model where the Mediator/Cohesin/Nipbl complex promotes cell-type specific gene activation through enhancer/promoter DNA looping.

Mediator and Cohesin in Embryonic Fibroblasts

Mediator and Cohesin are expressed in a broad range of cell types, so we sought to determine whether the key themes that emerged from their study in ES cells were maintained in other cell types. We first investigated whether Mediator occupies the promoters of actively transcribed genes in MEFs (Figure 6A and 6B). High confidence ChIP-Seq data showed that Mediator occupies at least 14% of genes bound by Pol2 (Table S5; Med1 data). We then compared sites occupied by Mediator, Cohesin and CTCF, and found Cohesin co-occupies most Mediator sites and most CTCF sites (Figure 6B and 6C). As in ES cells, the

Mediator/Cohesin co-occupied sites generally showed evidence of RNA polymerase II occupancy, whereas the CTCF/Cohesin sites typically lacked evidence for RNA polymerase II (Figure 6D).

We then investigated whether knockdowns of Mediator and Cohesin have similar effects on the expression of genes that they co-occupy in MEFs. Indeed, knockdown of Med12 and Smc1a had very similar effects on this group of active genes (Pearson Correlation of 0.65)(Figure 6E). Of the 443 genes that are co-occupied by Mediator, Cohesin and Pol2 at high confidence, approximately 300 showed significant expression changes (P-val <0.01) in both the Mediator and Cohesin knockdown datasets (Figure 6E). These results indicate that Mediator and Cohesin have similar effects on the expression of genes they occupy in MEFs, as was observed in ES cells.

Figure 5

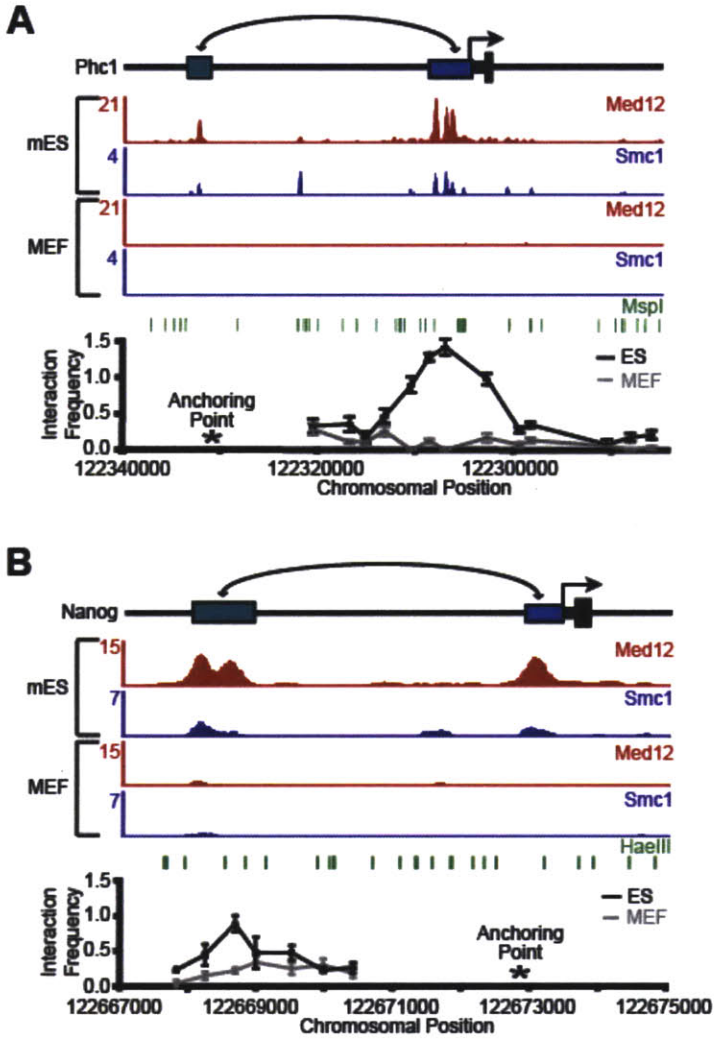

A
Phc1
mES
MEF
Med12
Smc1
MspI
Interaction Frequency
Anchoring Point
ES
MEF
122340000
122320000
122300000
Chromosomal Position
B
Nanog
HaeIII
122667000
122669000
122671000
122673000
122675000

Figure 5. Mediator and Cohesin Binding Profiles Predict Enhancer-Promoter Looping Events.

(A) A looping event between the upstream enhancer and the core promoter of *Phc1* was detected by Chromosome Conformation Capture (3C) in ES cells, but not in MEFs. ES cell and MEF crosslinked chromatin was digested by MspI and religated under conditions that favor intramolecular ligation events. The interaction frequency between the anchoring point and distal fragments was determined by PCR and normalized to BAC templates and control regions. The error bars represent the standard error of the average of 3 independent PCR reactions. The restriction enzyme sites are indicated above the 3C graph. The ChIP-Seq binding profiles for Med12 and Smc1 are shown in reads/million with the base of the y-axis set to 0.5 reads/million. Reads from two biological replicates for the Med12 and Smc1 (ES cells) ChIP-Seq datasets were combined. Biological replicates of the 3C experiments and the full 3C profile are presented in Figure S6.

(B) A looping event between an upstream Mediator/Cohesin co-occupied region and the core promoter was detected at the *Nanog* locus in ES cells but not in MEFs using 3C as described in (A).

Figure 6

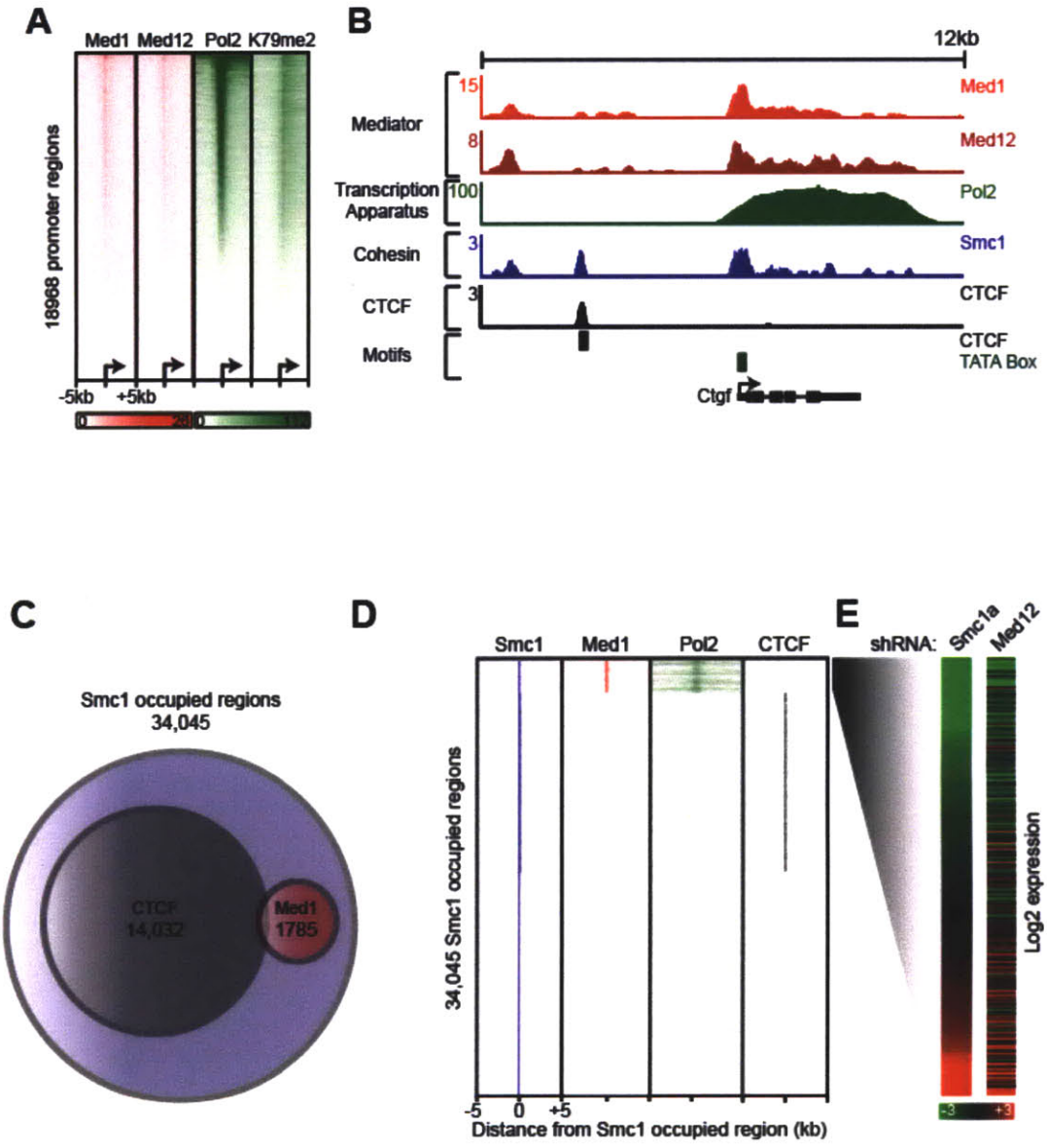

A
Med1 Med12 Pol2 K79me2
18968 promoter regions
-5kb +5kb
B
12kb
Mediator
Transcription Apparatus
Cohesin
CTCF
Motifs
15
Med1
8
Med12
100
Pol2
3
Smc1
3
CTCF
CTCF
TATA Box
Ctgf
C
Smc1 occupied regions
34,045
CTCF
14,032
Med1
1785
D
Smc1
Med1
Pol2
CTCF
34,045 Smc1 occupied regions
-5 0 +5
Distance from Smc1 occupied region (kb)
E
shRNA:
Smc1a
Med12
Log2 expression

Figure 6. Mediator and Cohesin Occupancy in Embryonic Fibroblasts.

(A) Density map of ChIP-Seq results for Mediator (Med1 and Med12), RNA polymerase II (Pol2) and di-methylated histone H3 lysine 79 (K79me2) demonstrates Mediator occupancy at genes that are actively transcribed in MEFs. Normalized read counts are shown for 10kb surrounding 18,968 Refseq promoters (from -5kb to +5kb) sorted by maximum level of Pol2 enrichment. A relative signal scale (reads/million) and the position of the transcription start site is shown at the bottom of the panel. Reads from two biological replicates for the Pol2 and H3K79me2 ChIP-Seq datasets were combined.

(B) Binding profiles for Mediator (Med1 and Med12), Cohesin (Smc1), CTCF and RNA polymerase II (Pol2) at *Ctgf*. ChIP-Seq data is shown in reads/million with the base of the y-axis set to 0.5 reads/million. CTCF and TBP (TATA Box) binding motifs are shown as black and green boxes respectively. The transcription start site and direction of transcription are noted by an arrow. Reads from two biological replicates for the Pol2, H3K79me2, Smc1 and CTCF ChIP-Seq datasets were combined.

(C) Venn diagram showing the overlap of high confidence (P-val $<10^{-9}$) Cohesin occupied sites with CTCF and Mediator. Approximately 41% of the genomic sites occupied by the Cohesin protein Smc1 were co-occupied by CTCF, 5% were co-occupied by Mediator.

(D) Region map demonstrating that Smc1 and Med1 co-occupied sites in MEFs generally occur in the absence of CTCF occupancy and within close proximity to RNA polymerase II (Pol2). For each Smc1 occupied site, the presence of Med1, Pol2 and CTCF occupancy is indicated within a 10kb window centered on the Smc1 occupied region.

(E) Heat map indicating that regions co-occupied by Smc1 and Med1 are associated with active genes that exhibit similar expression changes with either Smc1a or Med12 knockdown. Log_2 expression data from biological replicates is shown for all Smc1 and Med1 co-occupied regions that could be mapped to a gene. Mapped genes have evidence of a co-occupied Smc1/Med1 region within the gene body or within 10kb upstream of the transcriptional start site, evidence

of Pol2 occupancy within the gene body and significant (P-val <0.01) expression changes for both an Smc1a and Med12 knockdown in independent experiments. The log_2 expression data was ordered based on the Smc1a knockdown and the corresponding expression change for each gene following a Med12 knockdown is also shown. A relative signal scale for the expression data is shown at the bottom of both panels.

Cell-type Specific Behavior of Mediator and Cohesin

The observation that Mediator and Cohesin occupied the promoters of ES cell specific genes such as those encoding the pluripotency regulators Oct4 and Nanog, but had less occupancy at many housekeeping genes, led us to ask whether Mediator and Cohesin tend to occupy cell-type specific genes. Indeed, Mediator and Cohesin were found to occupy very different sets of promoters in ES cells and MEFs (Figure 7A). In contrast, Cohesin and CTCF occupied many of the same sites in the two cell types (Figure 7B). The levels of Mediator were found to be considerably higher in ES cells than in MEFs (Figure 7C; Figure S2C), accounting for the differences in the number of genes bound by Mediator and Cohesin in the two cell types. These observations suggest that Mediator and Cohesin play especially important roles in cell-type specific gene expression and thus, in cell-type specific chromosome structure.

Figure 7

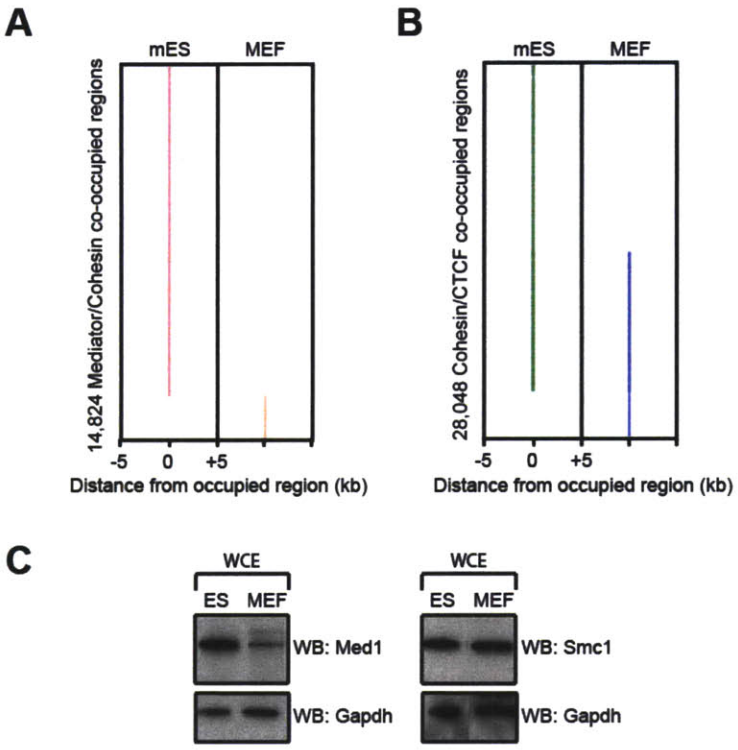

A
mES
MEF
14,824 Mediator/Cohesin co-occupied regions
-5
0
+5
Distance from occupied region (kb)
B
mES
MEF
28,048 Cohesin/CTCF co-occupied regions
-5
0
+5
Distance from occupied region (kb)
C
WCE
ES
MEF
WB: Med1
WB: Gapdh
WCE
ES
MEF
WB: Smc1
WB: Gapdh

Figure 7. Cell Type Specific Occupancy of Mediator and Cohesin.

(A) Region map of Mediator and Cohesin co-occupied regions for ES cells (Smc1 and Med12) and MEFs (Smc1 and Med1) indicates that co-occupied regions are largely different between the cell types. Mediator and Cohesin co-occupied regions are indicated for both cell types within a 10kb window. Reads from two biological replicates for the Smc1 and Med12 ChIP-Seq datasets were combined.

(B) Region map of Cohesin (Smc1) and CTCF co-occupied regions indicates that a large fraction of these regions are co-occupied in ES cells and in MEFs. Cohesin and CTCF co-occupied regions are indicated for both cell types within a 10kb window. Reads from two biological replicates for the Smc1 and CTCF (MEF) ChIP-Seq datasets were combined.

(C) Western blot analysis of ES and MEF cell extracts indicates that Cohesin protein levels are similar for both cell types, whereas Mediator protein levels are substantially lower in MEFs.

Discussion

DNA loop formation has been implicated in gene control in prokaryotes and eukaryotes, but the processes that connect DNA loop formation and gene regulation throughout the genome are not well understood. Our results reveal that Mediator and Cohesin contribute to the control of a subset of active genes, where they mediate the interaction between transcription factors at enhancers and the transcription apparatus at core promoters, and thus facilitate DNA loop formation. The subset of genes occupied by Mediator and Cohesin is cell-type specific, thus indicating that cell-type specific loops exist in the chromosomes of vertebrate cells as a consequence of active gene regulation.

Mediator Functions at a Key Subset of Active Promoters

At eukaryotic protein coding genes, transcription factors bind enhancers and other regulatory elements and recruit the transcription apparatus to core promoter sites where transcription begins (Fuda et al., 2009; Orphanides and Reinberg, 2002; Roeder, 1998). Enhancers are typically occupied by multiple transcription factors, which can have positive or negative influences on expression of their target genes (Maston et al., 2006; Panne, 2008; Panne et al., 2007; Visel et al., 2009). Coactivators such as Mediator provide an interface that transduces regulatory information from the transcription factors at regulatory elements to the transcription apparatus, thus producing finely calibrated output levels of gene activity (Conaway et al., 2005; Ding et al., 2008; Kornberg, 2005; Malik and Roeder, 2005; Taatjes et al., 2004).

In yeast, Mediator has been shown to be essential for expression of most, but not all, genes (Fan et al., 2006; Holstege et al., 1998; Thompson and Young, 1995). In vertebrates, Mediator is known to be essential, as null mutations produce embryonic lethality (Tudor et al., 1999), but the set of genes that require its function are not known in any cell type. We found that Mediator is associated with approximately 79% of promoters that are occupied by Pol2 in ES cells and approximately 14% of promoters that are occupied by Pol2 in embryonic

fibroblasts. Interestingly, Mediator tends to be associated with genes that are cell-type specific. Evidence that the ES cell specific transcription factors Oct4 and Nanog contribute to Mediator's recruitment may account, at least in part, for this observation (Figure 3)(Tutter et al., 2009).

Cohesin Occupies at Least Two Sites Associated with Gene Control

The Cohesin complex mediates cohesion of sister chromatids, which is essential for proper chromosome segregation and post-replicative DNA repair (Hirano, 2006; Nasmyth and Haering, 2005). Recent studies have shown that mammalian Cohesin complexes are also involved in the control of gene expression (Dorsett, 2007; Hagstrom and Meyer, 2003). Evidence that Cohesin associates with the insulator CCCTC-binding factor CTCF throughout the mammalian genome has led to the proposal that CTCF recruits Cohesin to these sites, where it can mediate chromosomal interactions between CTCF-bound sites (Hadjur et al., 2009; Parelho et al., 2008; Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008). The CTCF/Cohesin interaction has been implicated in repressing gene expression through enhancer blocking (Parelho et al., 2008; Wendt et al., 2008), but there is also evidence that Cohesin and CTCF are involved in organizing distal looping events that are associated with gene activity (Hadjur et al., 2009). Our results show that Cohesin occupies a large fraction of CTCF sites throughout the genomes of ES cells and MEFs, reveal that many CTCF/Cohesin co-occupied sites are conserved between these two cell types, and demonstrate that CTCF/Cohesin sites typically lack evidence of gene activity in these cells.

We find that Cohesin also occupies the set of enhancers and core promoters of active genes that are occupied by Mediator in ES cells and in MEFs. Multiple lines of evidence indicate that Mediator and Cohesin physically and functionally interact to control gene expression at these genes. ChIP-Seq data shows that the two protein complexes co-occupy nearly 14,000 sites in ES cells and 2000 sites in MEFs. Mediator and Cohesin co-precipitate in immunoprecipitation experiments and Cohesin is found associated with Mediator

complexes following a rigorous, multi-step affinity purification from ES cells. Knockdowns of Mediator and Cohesin components have strikingly similar effects on ES cell state and on gene expression at co-occupied genes. Thus, Cohesin appears to function together with CTCF at loci that are not generally transcribed and with Mediator at promoters that are generally transcriptionally active.

A Model for Gene Regulation and DNA Looping via Mediator and Cohesin

Our results and those of others suggest a model for the mechanistic contributions of Mediator and Cohesin to gene regulation and DNA looping in vertebrate cells. In this model, DNA loop formation between enhancers and core promoters occurs as a consequence of the interaction between enhancer-bound transcription activators, Mediator and promoter-bound RNA polymerase II. When the transcription activators bind Mediator, the Mediator complex undergoes a conformational change, and we show that this form of Mediator binds Cohesin (Taatjes et al., 2002). The Cohesin loading factor Nipbl is located at the sites co-occupied by Mediator and Cohesin, providing a mechanism for Cohesin loading at these sites.

Studies in bacteria first established the concept that transcriptional regulation can involve DNA loop formation due to interactions between regulators and the transcription apparatus when the two occupy distal DNA sites, and introduced the notion that certain proteins such as IHF are devoted to DNA loop stabilization (Borowiec et al., 1987; Dunn et al., 1984; Hahn et al., 1986; Hochschild and Ptashne, 1988; Huo et al., 1988; Popham et al., 1989; Xu and Hoover, 2001). Our results suggest that Mediator and Cohesin may contribute to these two types of regulatory processes in mammalian cells. While Mediator connects transcription factors and the transcription apparatus and thus contributes to DNA loop formation, Cohesin's contribution may be to stabilize the loop. Importantly, because different cells express different sets of genes, the DNA loops mediated by Mediator and Cohesin are cell-type specific.

Control of Cell State

ES cells provide an exceptional model system to study the regulatory components that contribute to control of cell state and early development (Chen et al., 2008; Cole and Young, 2008; Deato and Tjian, 2008; Enver et al., 2009; Jaenisch and Young, 2008; Kagalwala et al., 2008; MacArthur et al., 2008; Orkin et al., 2008; Yu and Thomson, 2008). The genetic screen described here, and those described previously (Fazzio et al., 2008; Hu et al., 2009; Ivanova et al., 2006; van den Berg et al., 2008), identified ES cell-specific regulators such as the transcription factors Oct4, Sox2, Nanog, Esrrb, Sall4, and Stat3 as key to maintenance of the ES cell state. These screens have also identified important regulators of global gene expression that are found more generally in cells, including a variety of chromatin regulators (e.g., Polycomb, Tip60-p400 and other histone modifying enzymes) and Mediator and Cohesin. This suggests that maintenance of the embryonic state is especially sensitive to the functional levels of certain general regulatory components, particularly those that contribute to chromatin structure.

Through their roles in DNA loop formation at a subset of active promoters, Mediator and Cohesin link gene expression with cell-type specific chromatin structure. In this context, it is interesting that mutations in the genes encoding Mediator and Cohesin components can cause an array of human developmental syndromes and diseases. Mediator mutations have been associated with Opitz-Kaveggia (FG) syndrome, Lujan syndrome, schizophrenia and some forms of congenital heart failure (Ding et al., 2008; Philibert and Madan, 2007; Risheg et al., 2007; Schwartz et al., 2007). Mutations in Nipbl are responsible for most cases of Cornelia de Lange syndrome (CdLS), which is characterized by developmental defects and mental retardation and appears to be the result of mis-regulation of gene expression rather than chromosome cohesion or mitotic abnormalities (Borck et al., 2004; Musio et al., 2006; Strachan, 2005; Zhang et al., 2007). It is possible that these disorders and diseases are due to deficiencies in the chromatin structure generated by Mediator and Cohesin, which we have shown is essential for normal transcriptional programs in mammalian cells.

Experimental Procedures

Additional information on materials and methods can be found in Supplementary Information.

Cell Culture Conditions

V6.5 murine embryonic stem (mES) and ZHBTc4 (Niwa et al., 2000) Oct4 shutdown cells were grown under standard mES cell conditions as described previously (Boyer et al., 2005). Murine embryonic Fibroblasts (MEFs) were cultured in DMEM supplemented with 10% FBS, 100 mM nonessential amino acids 2 mM L-glutamine,100 U/mL penicillin, 100 mg/mL streptomycin, and 8 nL/mL of 2-mercaptoethanol.

High Throughput shRNA Screening

Small hairpins targeting approximately 2000 chromatin regulators and transcription factors were designed and cloned into pLKO.1 lentiviral vectors (Moffat et al., 2006). Lentiviral supernatants were generated and arrayed in 384-well plates along with negative control supernatants targeting GFP, RFP, Luciferase and LacZ (Moffat et al., 2006). mES cells were seeded off of a MEF feeder layer into 384-well plates, infected the following day and placed under puromycin selection 24 hours post infection. Five days post infection cells were fixed and stained with Hoechst and for Oct4 (Santa Cruz sc-5279). Image acquisition and data analysis were performed essentially as described (Moffat et al., 2006). The average Oct4 pixel staining intensity was determined for all identified cells in a well and a mean value for each well was calculated. Z-scores were determined for each well based on average Oct4 staining and the values of the negative controls. Z-scores for replicate infections were averaged. A detailed description of the screening protocol and analysis is described in Extended Experimental Procedures.

Micorarray Hybridization and Analysis

RNA for expression analysis was extracted with TRIzol (Invitrogen) and purified with a RNAeasy column (Qiagen) before labeling. Samples were Cy3 (GFP shRNA control infected cells) or Cy5 (Med12 and Smc1a shRNA infected cells) labeled and hybridized to Agilent mouse 4x44K expression arrays. The details of hybridization, wash conditions and analysis are described in Extended Experimental Procedures.

ChIP-Seq

Chromatin immunoprecipitations (ChIPs) were performed as previously described (Lee et al., 2006b). For this paper, ChIPs were performed for Med1/TRAP220 (Bethyl A300-793A), Med12 (Bethyl A300-774A), Smc1 (Bethyl A300-055A), Smc3 (Abcam ab9263), Nipbl (Bethyl A301-778A), TBP (Abcam ab818), RNA Polymerase II (Covance 8WG16), CTCF (Upstate 07-729), and H3K79me2 (Abcam ab3594). RNA Polymerase II, CTCF, and H3K79me2 ChIP-Seq experiments for mES cells were performed previously (Chen et al., 2008; Marson et al., 2008; Seila et al., 2008). 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, minor protocol modifications and data analysis are described in Extended Experimental Procedures. A summary of the ChIP-Seq data generated for this paper can be found in Table S7.

ChIP-Western and Co-Immunoprecipitation

For the ChIP-Western experiments, the same conditions as for ChIP-Seq were used. Following immunoprecipitation and washes, proteins were separated by SDS-PAGE and analyzed by western blot. For the co-immunoprecipitation experiment, mES cells were harvested in PBS and proteins were extracted for 30 min at 4°C in TNEN250 buffer with protease inhibitors. After centrifugation, supernatant was increased to two volumes with TNENG buffer. Protein complexes were immunoprecipitated overnight at 4°C using antibodies against

Med12 (Bethyl A300-774A), Smc1 (Bethyl A300-055A), Nibpl (Bethyl A301-778A) and isotype-matched non-immune IgG (Upstate). Immunoprecipitates were washed three times with TNEN125 buffer, and proteins were separated by SDS-PAGE and analyzed by western blot. For both experiments, blots were probed with antibodies against Med23 (Bethyl A300-425A), Smc1 (Bethyl, A300-055A), Smc3 (Abcam, Ab9236), Nipbl (Bethyl A301-778A) and Oct4 (Santa Cruz sc-5279).

Mediator Complex Purification

The Mediator complex was purified from mES cell nuclear extracts using immobilized GST-SREBP-1a (residues 1-50). Bound material washed 4x with 20 column volumes of 0.5M KCl HEGN (20mM Hepes, 0.1mM EDTA, 10% Glycerol, 0.1% NP-40 & 0.5M KCl) buffer, 2x with 0.15M KCl HEGN buffer, and eluted. The eluted sample was further purified with a CDK8 antibody. After binding, this resin was washed 4x with 50 column volumes of 0.5M KCl HEGN buffer, 2x with 0.1M KCl HEGN buffer and eluted with 0.1M Glycine, pH 2.75.

Multidimensional Protein Identification Technology (MudPIT) assays

GST-SREBP-1a (residues 1-50) was immobilized to GSH-Sepharose beads (GE Lifesciences) and used as bait for overnight pull downs (4°C) from HeLa nuclear extract. Beads were washed with 5 x 20 column volumes 0.5M KCl HEGN (20mM Hepes, 0.1mM EDTA, 10% Glycerol, 0.1% NP-40 & 0.5M KCl) and 1 x 20 column volumes 0.15M KCl HEGN. Bound material was eluted with 30mM GSH in elution buffer (80mM Tris, 0.1mM EDTA, 10% Glycerol, 0.02% NP-40, 100mM KCl). GSH elutions were applied to the top of 15% to 40% linear glycerol gradients and centrifuged for 6h at 50,000rpm. Mediator-containing fractions were combined and TCA precipitated. The protein pellet was suspended with 4%(w/v) SDS, 0.1M Tris pH 8.5, 10mM TCEP then added to 8M Urea, 0.1M Tris pH 8.5 for iodoacetamide alkylation and trypsin digestion (37C, 1h) using a modified FASP protocol (Wisniewski et al., 2009). Digested peptides were separated using high pH (pH10 ammonium formate) /low pH (0.1%(v/v) formic

acid) reversed-phase two-dimensional liquid chromatography. An Agilent HP1100 nanoLC/MSD XCT ion trap mass spectrometer (Agilent Technologies) was used for all analyses. Peptide identifications (from a Mascot search of the human IPI_v3.65 database) were filtered at a 1% false discovery rate as determined by a search of the reversed database. Quantitation of proteins used the method of spectral counting, which is the total number of ms/ms assignments to peptides within a protein, described by Old et al. (Old et al., 2005).

Chromosome Conformation Capture (3C)

3C analysis was performed as previously described (Miele and Dekker, 2009). mES cells or MEFs were crosslinked, lysed and chromatin was digested with 1000 units HaeIII (NEB) for the *Nanog* and *Oct4 (Pou5f1)* loci or 2000 units MspI (NEB) for the *Phc1* and *Lefty1* loci. Crosslinked fragments were subsequently ligated with 50 units T4 DNA ligase (Invitrogen) for 4 hours at 16°C. Control templates were generated using BAC clones covering the *Nanog*, *Oct4 (Pou5f1)*, *Lefty1* or *Phc1* loci. 3C primers were designed for fragments upstream and surrounding the transcriptional start site. 3C analysis was done in triplicate and averaged for each primer pair. A complete list of primers, are available in Table S8. Details of the analysis are described in Extended Experimental Procedures.

Acknowledgements

We thank Tony Lee for helpful discussions, Jeong-Ah Kwon, Jennifer Love, Sumeet Gupta and Thomas Volkert for assistance with ChIP-Seq and David Root, Serena Silver, Thomas Neiland, and Hanh Le at the RNAi Screening Platform at the Broad Institute for screening advice and technical support. Immunofluorescence images were collected using the W.M. Keck Foundation Biological Imagining Facility at the Whitehead Institute and Whitehead-MIT Bioimaging Center. This work was supported by the Amcerican Cancer Society (DJT), a Fellowship from the Canadian Institutes of Health Research to SB, by an NIH NIGMS Post-doctoral Fellowship to MK and by NIH grant HG002668 (RY), HG003143 (JD) and a Keck Distinguished young scholar award (JD).

References

Adhya, S. (1989). Multipartite genetic control elements: communication by DNA loop. Annu Rev Genet 23, 227-250.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. Dev Cell 16, 47-57.

Atchison, M.L. (1988). Enhancers: mechanisms of action and cell specificity. Annu Rev Cell Biol 4, 127-153.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell 27, 299-308.

Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promotor region. Nature 290, 304-310.

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.

Borck, G., Redon, R., Sanlaville, D., Rio, M., Prieur, M., Lyonnet, S., Vekemans, M., Carter, N.P., Munnich, A., Colleaux, L., et al. (2004). NIPBL mutations and genetic heterogeneity in Cornelia de Lange syndrome. J Med Genet 41, e128.

Borowiec, J.A., Zhang, L., Sasse-Dwight, S., and Gralla, J.D. (1987). DNA supercoiling promotes formation of a bent repression loop in lac DNA. J Mol Biol 196, 101-111.

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

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.

Bulger, M., and Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. Genes Dev 13, 2465-2477.

Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F., and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. Nat Genet 32, 623-626.

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

Cole, M.F., and Young, R.A. (2008). Mapping key features of transcriptional regulatory circuitry in embryonic stem cells. Cold Spring Harb Symp Quant Biol 73, 183-193.

Conaway, R.C., Sato, S., Tomomori-Sato, C., Yao, T., and Conaway, J.W. (2005). The mammalian Mediator complex and its role in transcriptional regulation. Trends Biochem Sci 30, 250-255.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol 319, 1097-1113.

Deato, M.D., and Tjian, R. (2008). An unexpected role of TAFs and TRFs in skeletal muscle differentiation: switching core promoter complexes. Cold Spring Harb Symp Quant Biol 73, 217-225.

Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. Nat Methods 3, 17-21.

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

Ding, N., Zhou, H., Esteve, P.O., Chin, H.G., Kim, S., Xu, X., Joseph, S.M., Friez, M.J., Schwartz, C.E., Pradhan, S., et al. (2008). Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. Mol Cell 31, 347-359.

Dorsett, D. (2007). Roles of the sister chromatid cohesion apparatus in gene expression, development, and human syndromes. Chromosoma 116, 1-13.

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

Drissen, R., Palstra, R.J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., and de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev 18, 2485-2490.

Dunn, T.M., Hahn, S., Ogden, S., and Schleif, R.F. (1984). An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. Proc Natl Acad Sci U S A 81, 5017-5020.

Enver, T., Pera, M., Peterson, C., and Andrews, P.W. (2009). Stem cell states, fates, and the rules of attraction. Cell Stem Cell 4, 387-397.

Fan, X., Chou, D.M., and Struhl, K. (2006). Activator-specific recruitment of Mediator in vivo. Nat Struct Mol Biol 13, 117-120.

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

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A 73, 1897-1901.

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

Fullwood, M.J., and Ruan, Y. (2009). ChIP-based methods for the identification of long-range chromatin interactions. J Cell Biochem 107, 30-39.

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature 460, 410-413.

Hagstrom, K.A., and Meyer, B.J. (2003). Condensin and cohesin: more than chromosome compactor and glue. Nat Rev Genet 4, 520-534.

Hahn, S., Hendrickson, W., and Schleif, R. (1986). Transcription of Escherichia coli ara in vitro. The cyclic AMP receptor protein requirement for PBAD induction that depends on the presence and orientation of the araO2 site. J Mol Biol 188, 355-367.

Hatzis, P., and Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Mol Cell 10, 1467-1477.

Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. Nat Rev Mol Cell Biol 7, 311-322.

Hochschild, A., and Ptashne, M. (1988). Interaction at a distance between lambda repressors disrupts gene activation. Nature 336, 353-357.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell 95, 717-728.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev 23, 837-848.

Huo, L., Martin, K.J., and Schleif, R. (1988). Alternative DNA loops regulate the arabinose operon in Escherichia coli. Proc Natl Acad Sci U S A 85, 5444-5448.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature 442, 533-538.

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

Kagalwala, M.N., Singh, S.K., and Majumder, S. (2008). Stemness is only a state of the cell. Cold Spring Harb Symp Quant Biol 73, 227-234.

Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. Trends Biochem Sci 30, 235-239.

Krantz, I.D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L.A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C.A., et al. (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nat Genet 36, 631-635.

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

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. (2006a). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125, 301-313.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006b). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc 1, 729-748.

Levasseur, D.N., Wang, J., Dorschner, M.O., Stamatoyannopoulos, J.A., and Orkin, S.H. (2008). Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. Genes Dev 22, 575-580.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature 424, 147-151.

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

Liu, J., and Krantz, I.D. (2009). Cornelia de Lange syndrome, cohesin, and beyond. Clin Genet 76, 303-314.

Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M.A., Clark, D., Kaur, M., Tandy, S., Kondoh, T., Rappaport, E., et al. (2009). Transcriptional dysregulation in NIPBL and cohesin mutant human cells. PLoS Biol 7, e1000119.

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

MacArthur, B.D., Ma'ayan, A., and Lemischka, I.R. (2008). Toward stem cell systems biology: from molecules to networks and landscapes. Cold Spring Harb Symp Quant Biol 73, 211-215.

Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci 30, 256-263.

Malik, S., and Roeder, R.G. (2008). Epigenetics? Mediator does that too! Mol Cell 31, 305-306.

Maniatis, T., Goodbourn, S., and Fischer, J.A. (1987). Regulation of inducible and tissue-specific gene expression. Science 236, 1237-1245.

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

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet 7, 29-59.

Matthews, K.S. (1992). DNA looping. Microbiol Rev 56, 123-136.

Miele, A., and Dekker, J. (2009). Mapping Cis- and Trans- Chromatin Interaction Networks Using Chromosome Conformation Capture (3C). Methods Mol Biol 464, 105-121.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K., et al. (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell 124, 1283-1298.

Musio, A., Selicorni, A., Focarelli, M.L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P., and Larizza, L. (2006). X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat Genet 38, 528-530.

Nasmyth, K., and Haering, C.H. (2005). The structure and function of SMC and kleisin complexes. Annu Rev Biochem 74, 595-648.

Nativio, R., Wendt, K.S., Ito, Y., Huddleston, J.E., Uribe-Lewis, S., Woodfine, K., Krueger, C., Reik, W., Peters, J.M., and Murrell, A. (2009). Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. PLoS Genet 5, e1000739.

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

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

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

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., and Ahn, N.G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics 4, 1487-1502.

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

Orkin, S.H., Wang, J., Kim, J., Chu, J., Rao, S., Theunissen, T.W., Shen, X., and Levasseur, D.N. (2008). The Transcriptional Network Controlling Pluripotency in ES Cells. Cold Spring Harb Symp Quant Biol.

Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. Cell 108, 439-451.

Panne, D. (2008). The enhanceosome. Curr Opin Struct Biol 18, 236-242.

Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-beta enhanceosome. Cell 129, 1111-1123.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132, 422-433.

Park, S.W., Li, G., Lin, Y.P., Barrero, M.J., Ge, K., Roeder, R.G., and Wei, L.N. (2005). Thyroid hormone-induced juxtaposition of regulatory elements/factors and chromatin remodeling of Crabp1 dependent on MED1/TRAP220. Mol Cell 19, 643-653.

Philibert, R.A., and Madan, A. (2007). Role of MED12 in transcription and human behavior. Pharmacogenomics 8, 909-916.

Popham, D.L., Szeto, D., Keener, J., and Kustu, S. (1989). Function of a bacterial activator protein that binds to transcriptional enhancers. Science 243, 629-635.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature 322, 697-701.

Revet, B., von Wilcken-Bergmann, B., Bessert, H., Barker, A., and Muller-Hill, B. (1999). Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. Curr Biol 9, 151-154.

Ris, H., and Kubai, D.F. (1970). Chromosome structure. Annu Rev Genet 4, 263-294.

Risheg, H., Graham, J.M., Jr., Clark, R.D., Rogers, R.C., Opitz, J.M., Moeschler, J.B., Peiffer, A.P., May, M., Joseph, S.M., Jones, J.R., et al. (2007). A recurrent mutation in MED12 leading to R961W causes Opitz-Kaveggia syndrome. Nat Genet 39, 451-453.

Roeder, R.G. (1998). Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. Cold Spring Harb Symp Quant Biol 63, 201-218.

Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., and Krumm, A. (2008). CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A 105, 8309-8314.

Saiz, L., and Vilar, J.M. (2006). DNA looping: the consequences and its control. Curr Opin Struct Biol 16, 344-350.

Sato, S., Tomomori-Sato, C., Parmely, T.J., Florens, L., Zybailov, B., Swanson, S.K., Banks, C.A., Jin, J., Cai, Y., Washburn, M.P., et al. (2004). A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. Mol Cell 14, 685-691.

Schleif, R. (1992). DNA looping. Annu Rev Biochem 61, 199-223.

Scholer, H.R., Dressler, G.R., Balling, R., Rohdewohld, H., and Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. EMBO J 9, 2185-2195.

Schwartz, C.E., Tarpey, P.S., Lubs, H.A., Verloes, A., May, M.M., Risheg, H., Friez, M.J., Futreal, P.A., Edkins, S., Teague, J., et al. (2007). The original Lujan syndrome family has a novel missense mutation (p.N1007S) in the MED12 gene. J Med Genet 44, 472-477.

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.

Spilianakis, C.G., and Flavell, R.A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nat Immunol 5, 1017-1027.

Stedman, W., Kang, H., Lin, S., Kissil, J.L., Bartolomei, M.S., and Lieberman, P.M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. EMBO J 27, 654-666.

Strachan, T. (2005). Cornelia de Lange Syndrome and the link between chromosomal function, DNA repair and developmental gene regulation. Curr Opin Genet Dev 15, 258-264.

Taatjes, D.J., Marr, M.T., and Tjian, R. (2004). Regulatory diversity among metazoan co-activator complexes. Nat Rev Mol Cell Biol 5, 403-410.

Taatjes, D.J., Naar, A.M., Andel, F., 3rd, Nogales, E., and Tjian, R. (2002). Structure, function, and activator-induced conformations of the CRSP coactivator. Science 295, 1058-1062.

Taatjes, D.J., and Tjian, R. (2004). Structure and function of CRSP/Med2; a promoter-selective transcriptional coactivator complex. Mol Cell 14, 675-683.

Thompson, C.M., and Young, R.A. (1995). General requirement for RNA polymerase II holoenzymes in vivo. Proc Natl Acad Sci U S A 92, 4587-4590.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 10, 1453-1465.

Tonkin, E.T., Wang, T.J., Lisgo, S., Bamshad, M.J., and Strachan, T. (2004). NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. Nat Genet 36, 636-641.

Toth, J.I., Datta, S., Athanikar, J.N., Freedman, L.P., and Osborne, T.F. (2004). Selective coactivator interactions in gene activation by SREBP-1a and -1c. Mol Cell Biol 24, 8288-8300.

Treisman, R., and Maniatis, T. (1985). Simian virus 40 enhancer increases number of RNA polymerase II molecules on linked DNA. Nature 315, 73-75.

Tudor, M., Murray, P.J., Onufryk, C., Jaenisch, R., and Young, R.A. (1999). Ubiquitous expression and embryonic requirement for RNA polymerase II coactivator subunit Srb7 in mice. Genes Dev 13, 2365-2368.

Tutter, A.V., Kowalski, M.P., Baltus, G.A., Iourgenko, V., Labow, M., Li, E., and Kadam, S. (2009). Role for Med12 in regulation of Nanog and Nanog target genes. J Biol Chem 284, 3709-3718.

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol Cell 17, 453-462.

van den Berg, D.L., Zhang, W., Yates, A., Engelen, E., Takacs, K., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2008). Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. Mol Cell Biol 28, 5986-5995.

Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G., and Higgs, D.R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J 26, 2041-2051.

Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009). Genomic views of distant-acting enhancers. Nature 461, 199-205.

Wang, Q., Carroll, J.S., and Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. Mol Cell 19, 631-642.

Wasylyk, B., Wasylyk, C., Augereau, P., and Chambon, P. (1983). The SV40 72 bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. Cell 32, 503-514.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451, 796-801.

Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat Methods 6, 359-362.

Wolters, D.A., Washburn, M.P., and Yates, J.R., 3rd (2001). An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem 73, 5683-5690.

Wu, Q., Chen, X., Zhang, J., Loh, Y.H., Low, T.Y., Zhang, W., Sze, S.K., Lim, B., and Ng, H.H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J Biol Chem 281, 24090-24094.

Xu, H., and Hoover, T.R. (2001). Transcriptional regulation at a distance in bacteria. Curr Opin Microbiol 4, 138-144.

Yang, F., Vought, B.W., Satterlee, J.S., Walker, A.K., Jim Sun, Z.Y., Watts, J.L., DeBeaumont, R., Saito, R.M., Hyberts, S.G., Yang, S., et al. (2006). An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. Nature 442, 700-704.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development 122, 881-894.

Yu, J., and Thomson, J.A. (2008). Pluripotent stem cell lines. Genes Dev 22, 1987-1997.

Zhang, B., Jain, S., Song, H., Fu, M., Heuckeroth, R.O., Erlich, J.M., Jay, P.Y., and Milbrandt, J. (2007). Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome. Development 134, 3191-3201.

Chapter 5

Concluding Remarks

The work described in the previous chapters presents new contributions to the study of transcriptional regulatory mechanisms in embryonic stem cells. The connections of both Wnt and TGF- β signaling to the core regulatory circuitry of ES cells are novel findings that reveal how signaling pathways and transcription factors coordinately control gene expression. Discovery of the role that Mediator and Cohesin play in cell-type specific gene expression has highlighted a previously underappreciated mechanism of gene control in mammals. Together these studies add to our knowledge of the multiple layers of gene control required for mammalian cellular differentiation and development.

Signaling pathways are essential in the regulation of nearly all biological processes. Specific combinations and concentrations of signaling molecules stimulate the pathways, forcing the terminal factors into the nucleus where they often interact with transcription factors to influence gene epxression. The work presented here on Wnt and TGF-β signaling adds to the overall understanding of how various regulatory mechanisms work together to control transcription and embryonic stem cell state. Studies examining the consequences on cell state of a variety of signaling pathways, including LIF, Wnt, BMP4 and TGF-β signaling provided the functional data and a reason to further research the role of each of these pathways in ES cells. The data presented in chapters 2 and 3 identifies the specific genes controlled by Tcf3 and the Wnt signaling pathway and Smad3 and the TGF-β signaling pathway, respectively. The identification of interactions between terminal signaling components and transcription factors through these genome-wide ChIP-Seq studies has deepened our understanding of how groups of proteins might be acting together to regulate gene expression and cell state in ES cells. The finding that Tcf3 co-occupied the genome with Oct4, Sox2 and Nanog provided the first evidence that a signaling pathway was linked directly to the core regulatory circuitry and was in fact part of the interconnected autoregulatory loop of transcription factors that make up the foundation of this network. The study of TGF-β signaling not only provided additional evidence that signaling pathways act with key ES cell regulators to control cell state, but the data from myotubes and pro-b cells provided novel information on the

relationship between TGF-β signaling and cell-type specific transcription factors. Together, the studies presented here add novel genome-wide binding data and functional studies involving perturbations of signaling pathways to enhance our understanding of signaling pathways and their role in maintaining ES cell state.

The data collected from the high-throughput genetic screen in murine embryonic stem cells presented here in Chapter 4 adds to a growing list of novel factors that control ES cell state. The in-depth study of Mediator and Cohesin in ES cells led to the identification of a novel protein complex that is essential for maintaining a cell-type specific gene expression profile. This study provided the first genome-wide binding data for the Mediator complex, supporting its role as a coactivator. The genome-wide binding data for Cohesin subunits Smc1 and Smc3, supported previous studies that Cohesin predominantly associates with CTCF but also exists at a subset of regions across the genome in the absence of CTCF. This study provided novel data for the role of those Cohesin-independent occupied sites across the genome. The Cohesin-independent sites were identified at the promoters and enhancers of active genes along with Mediator and ES cell transcription factors Oct4, Sox2 and Nanog. In addition to determining occupancy across the genome, Mediator and Cohesin were indentified to be part of a protein complex regulating cell-type specific gene expression programs by participating in long-range looping events. The concept of DNA looping has been demonstrated previously in bacteria and has more recently been shown at specific loci in vertebrate cells. The co-occupancy of promoter and enhancer regions by Mediator and Cohesin allows for the accurate prediction of cell-type specific looping events that are responsible for maintaining gene expression profiles and cell state, something that had not yet been demonstrated in vertebrate cells.

Together, the work presented in this thesis contributes to a growing area of biomedical research working to better understand embryonic stem cell biology and the control of a pluripotent cell state. I have provided novel evidence for the relationship of signaling pathways with master transcription factors in embryonic stem cells as well as more differentiated cell-types supporting a cell-type specific

role for terminal components of signaling pathways. In addition, I have identified a novel protein complex that also acts in cell-type specific manner to regulate a gene expression program to maintain cell state. As similar studies move forward, the knowledge of how embryonic stem cell state is maintained and changed will continue to grow, greatly impacting the use of ES cells in biomedical research.

Recent progress in cellular reprogramming and the potential use for reprogrammed cells in regenerative medicine highlight the importance of understanding the mechanisms through which cell state is altered during cellular differentiation. We are now in a position to examine in increased detail the changes that occur at multiple cellular levels, from extracellular signaling to the modifications of histones that regulate the binding of transcription factors and the transcription apparatus. Here I describe future studies that will (1) define the biochemical and molecular mechanisms of gene control downstream of signaling pathways, (2) elucidate the biochemical regulation of DNA looping and the functional roles that complex DNA architecture plays in the control of gene regulation, and (3) exploit improved genetic techniques to identify novel layers of transcriptional regulation. These studies of the multiple mechanisms of gene control that are crucial for normal human development have significant potential to increase our understanding of human disease and lead to novel therapeutics.

Mechanisms of gene control downstream of signaling pathways

The effect that different signaling pathways have on embryonic stem cell state is studied easily through the visualization of changes in cell morphology. The molecular events associated with pluripotency and differentiation and the ways that signaling pathways control these events are only now beginning to be understood. ChIP-Seq is a robust new technology that allows for the identification of genes occupied by components of signaling pathways. The combination of this data, along with genome-wide expression changes following perturbations of a given signaling pathway help to reveal the function of that signaling pathway. What still remains unknown is the precise mechanism through which signaling pathways affect gene expression. The studies presented here in Chapter 2 and

3 suggest that transcription factors are likely responsible for recruiting Tcf3 and Smad3 to defined genomic loci to regulate gene expression. In order to fully understand how signaling pathways influence gene expression and cell state, further studies must better define biochemical and molecular changes that occur in the nucleus when signaling pathways are activated. Some of the key studies that remain involve the identification of other factors being recruited along with transcription factors and signaling pathways both in embryonic stem cells and in more differentiated cells. In addition, an understanding of the dynamic changes that take place at the level of chromatin structure- changes required both to allow protein binding and to promote active transcription- will be necessary for a full understanding of how all levels of regulatory circuitry cooperate to control a cell-specific gene expression program.

A combination of techniques will be required for detailed mechanistic examinations of gene control, including small scale studies of genome structure and biochemical purifications to determine the complete list of proteins acting at a given loci. One of the current limitations of ChIP-Seq is the large number of cells required for a single experiment and the heterogeneous population that likely exists in a population of 10-100 million cells. As a result of this limitation, current studies of the dynamic changes in chromatin modifications that accompany differentiation are limited by uncertainty if all of the cells examined are at the same stage of differentiation. Smaller scale ChIP-Seq experiments may allow for better controlling cell population, although the challenges would not be eliminated. More likely, reporter assays will have to be designed to monitor changes in chromatin at specific loci. For example, knowing the profound impact that over-expression of MyoD has on cell state, the over-expression of MyoD and a reporter assay designed to monitor regulatory mechanisms at genes turned on early in muscle cell differentiation, like myogenin, could begin to answer some of these questions. By using ChIP-Seq to monitor changes in chromatin marks at specific loci we could begin to see what's happening at the level of chromatin structure during changes in gene expression. With the addition of a reporter assay these changes could be monitored as a particular gene is being

expressed. Similar experiments could be designed using iPS cells where reactivation of the endogenous Oct4 locus could be tracked and changes in chromatin structure surrounding this locus could be monitored.

Regulation of DNA looping

The study presented in Chapter 4 describing the role of Mediator and Cohesin in regulating cell-type specific gene expression demonstrates a link between the core transcriptional regulatory circuitry of embryonic stem cells and higher-order chromatin structure. A more complete understanding of how DNA architecture contributes to gene regulation will depend on further investigations of the factors that control DNA looping events. There appear to be looping events associated with both activation and repression of gene expression, but the factors that facilitate each of these looping events remains unknown. Future studies must not only identify what other proteins are involved in regulating these looping events, but should focus on the changes that accompany the formation of DNA loops; specifically the changes in chromatin structure and nucleosome positioning that facilitate long-range looping events.

Detailed, genome-wide studies of DNA architecture and related changes in chromatin are now possible due to several newly developed techniques that are derivatives of chromosome conformation capture including 4C, 5C, Hi-C and ChIA-PET (Dostie and Dekker, 2007; Dostie et al., 2006; Dostie et al., 2007; Fullwood and Ruan, 2009; Li et al.; Lieberman-Aiden et al., 2009; Vassetzky et al., 2009; Zhao et al., 2006). These techniques are geared towards identifying looping events across more of the genome. ChIA-PET and 4C specifically focus on identifying DNA loops, genome-wide, that are associated with a specific protein. Hi-C and 5C use next-generation sequencing technologies to identify looping events across entire genomes. Although providing useful information, the amount of sequencing required to adequately cover the genome is currently cost prohibitive. Microscopy will likely provide complentary data on DNA looping. High-resolution microscopy would be able to detect distinct fluorophores corresponding to the proteins of interest in the absence of a loop and would

detect an overlap of fluorophores during a looping event. Unlike the millions of cell required for 3C and ChIP-seq experiments, microscopy experiments require very few cells. Additionally, microscopy allows for the investigation of individual cells, rather than a large population of cells that inevitably contain some heterogeneous mix of cell states and cell cycle phases.

A genetic approach to identifying novel layers of transcription regulation

Genetic screens in mammalian cells are increasingly feasible and have demonstrated success in the identification of novel factors with essential roles in gene expression and maintenance of cell state. As screening technology improves in quality, the depth of libraries increases and new assays to measure the effects of perturbations on cell state are developed, more factors contributing to different branches of transcriptional regulatory circuitry are sure to be identified. Additionally, perturbations that direct ES cells down particular lineages will likely be identified by these screening approaches. Small molecules to activate or inhibit particular genes are likely to be identified that will improve methods of directed differentiation or even cellular reprogramming.

A recent short hairpin screen I conducted in human embryonic stem cells has already identified additional factors with a role in regulating cell state. This screen followed the same model used for the screens in mouse ES cells presented in Chapter 4. Human ES cells were plated in 384 well plates, infected with individual shRNA-containing lentiviral constructs and stained for Oct4 five days after infection. Some of the factors identified by the screen that reduce Oct4 levels include the histone methylase SetD8 (Fang J et al. 2002; Couture JF 2005) and members of the family of histone deacetylases. The follow-up of novel chromatin modifiers, by more closely examining the affects following knockdown and their role in regulating specific genes both in ES cells and during changes in ES cell state, will continue to add information to how chromatin remodeling affects gene expression and cell state. HDACs have already been demonstrated to have a role in certain forms of cancer. The information gathered from ES cells could aid in the study of HDACs, their role in cancer and the impact that HDAC

inhibitors have on altering cell state. In addition, proteins involved in DNA architecture, including Condensin and high mobility group box (HMGB) proteins appear to be necessary for the maintenance of cell state. Condensin has been shown to have a role in regulating gene expression, specifically a role in controlling sex determination in both *C. elegans* and *Drosophila,* but the regulatory role of this complex in mammals has not yet been demonstrated (Csankovszki et al., 2009; Ercan and Lieb, 2009; Grimaud and Becker, 2009; Meyer, 2005). HMG box proteins are involved in facilitating bends in the DNA to aid in the interactions of various proteins involved in transcription. It will be interesting to better understand where these shorter-range DNA bends occur in the genome and how they are involved in regulating gene expression and cell state. The follow-up study of the hits from this shRNA screen in human ES cells will add critical insight to novel elements of the circuitry and their roles in the maintenance of the ES cell gene expression profile.

Regulation of gene expression and human disease

The studies of transcriptional regulation in ES cells described here provide new information on the molecular underpinnings of pluripotency. More broadly, these studies also serve as a model for understanding the multiple levels of transcriptional control required for mammalian cellular differentiation. As we learn more about the transcriptional regulation required for normal human development, we will inevitably learn about regulatory dysfunctions that contribute to human disease and arrive at new approaches for the treatment of disease. Already the identification of transcription factors that are essential for ES cell state have led to the generation of patient-specific induced pluripotent stem (iPS) cells as a realistic therapeutic possibility. Combining iPS cells with studies of development will lead to significant insights into developmental disorders. For example, the work presented in Chapter 4 suggests a mechanism for how a mutation in the Cohesin loading factor, Nipbl, could lead to Cornelia de Lange syndrome. The generation of iPS cells with CdLs and the subsequent differentiation of those cells could illuminate where during the developmental

process this mutation causes a detrimental effect. ES cells share some characteristics with malignant cancers including the capability to self-renew, a property commonly acquired by mutations in critical regulatory mechanisms. The study of self-renewal in ES cells and perturbations including those that target specific genes, signaling pathways or transcription factors and limit the self-renewing capabilities of ES cells could lead to new therapeutic strategies for the treatment of various forms of cancer. This idea is being investigated in our own group, where the same protocol used to screen an shRNA library in human ES cells now is being used to screen a library of FDA approved drug compounds. The information collected from this screen will provide insights into how drugs target self-renewal and pluripotency and how these drugs may be useful in the clinic. The discoveries that result from continued study will only serve to increase our understanding and ability to treat a number of developmental diseases, cancer, and degenerative diseases.

References

Csankovszki, G., Petty, E.L., and Collette, K.S. (2009). The worm solution: a chromosome-full of condensin helps gene expression go down. Chromosome Res *17*, 621-635.

Dostie, J., and Dekker, J. (2007). Mapping networks of physical interactions between genomic elements using 5C technology. Nat Protoc *2*, 988-1002.

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

Dostie, J., Zhan, Y., and Dekker, J. (2007). Chromosome conformation capture carbon copy technology. Curr Protoc Mol Biol *Chapter 21*, Unit 21 14.

Ercan, S., and Lieb, J.D. (2009). C. elegans dosage compensation: a window into mechanisms of domain-scale gene regulation. Chromosome Res *17*, 215-227.

Fullwood, M.J., and Ruan, Y. (2009). ChIP-based methods for the identification of long-range chromatin interactions. J Cell Biochem *107*, 30-39.

Grimaud, C., and Becker, P.B. (2009). The dosage compensation complex shapes the conformation of the X chromosome in Drosophila. Genes Dev *23*, 2490-2495.

Li, G., Fullwood, M.J., Xu, H., Mulawadi, F.H., Velkov, S., Vega, V., Ariyaratne, P.N., Mohamed, Y.B., Ooi, H.S., Tennakoon, C., *et al.* (2010). ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biol *11*, R22.

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

Meyer, B.J. (2005). X-Chromosome dosage compensation. WormBook, 1-14.

Vassetzky, Y., Gavrilov, A., Eivazova, E., Priozhkova, I., Lipinski, M., and Razin, S. (2009). Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. Methods Mol Biol *567*, 171-188.

Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., *et al.* (2006). Circular

chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet *38*, 1341-1347.

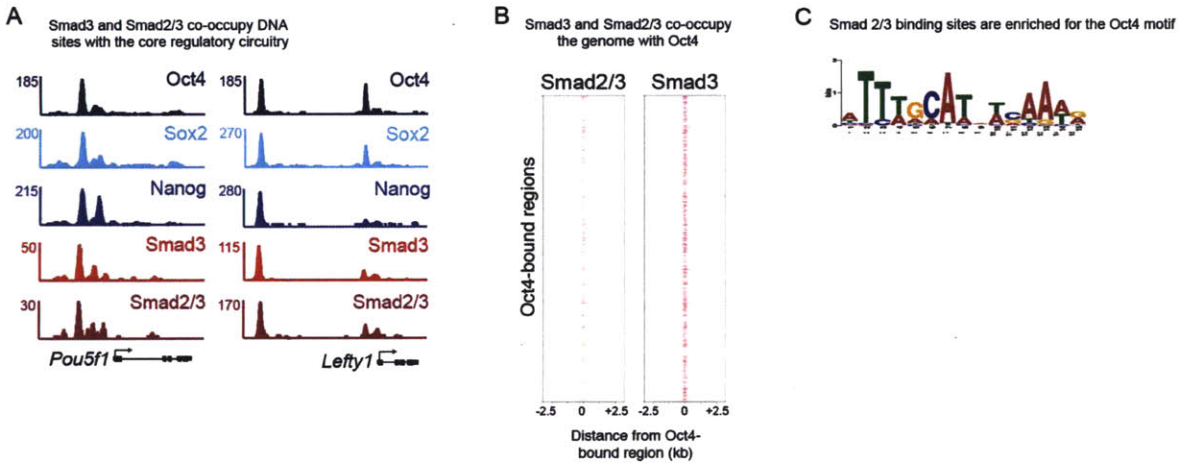
Appendix A

Supplementary Material for Chapter 3

Cell-Type Specific TGF-β Signaling is Targeted to Genes that Control Cell Identity

Table of Contents

Figure S1



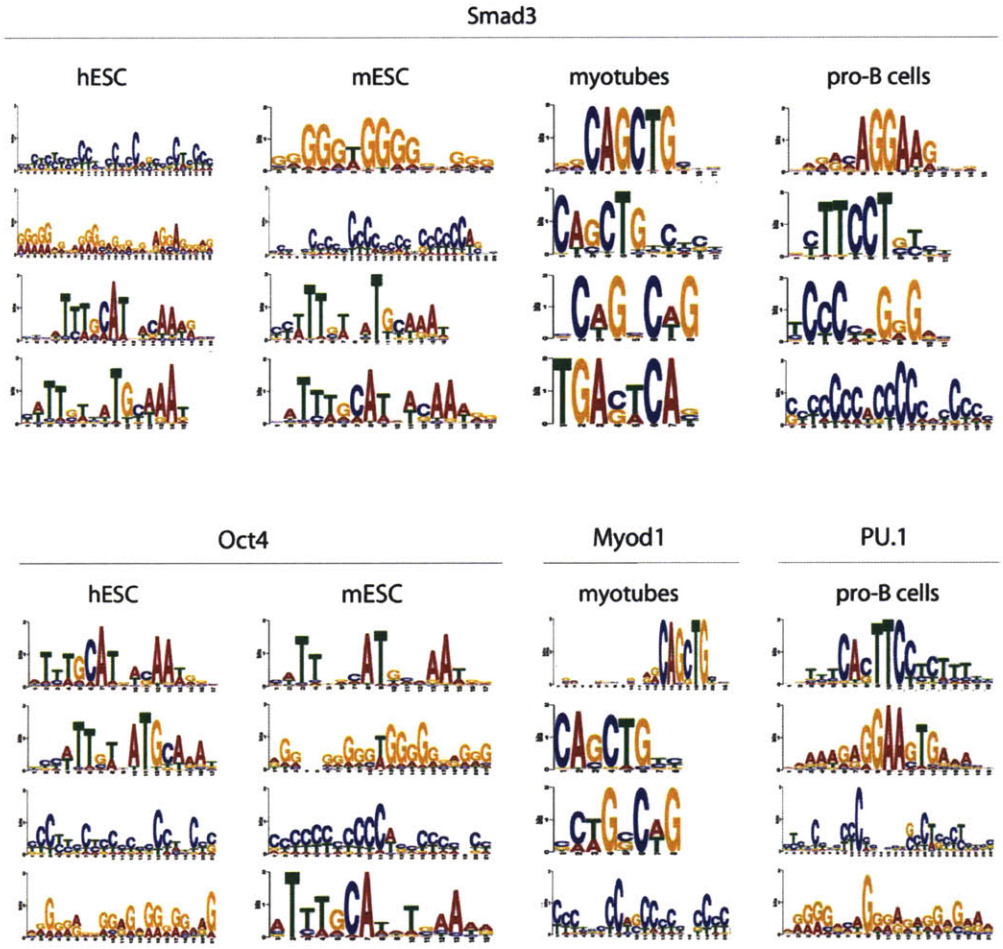
Antibodies against Smad3 and Smad2/3 show similar co-occupancy with Oct4 in mES cells.

(A) Smad3 and Smad 2/3 co-occupy DNA sites with the core regulatory circuitry. Gene tracks represent binding of Oct4, Sox2, Nanog (Marson et al., 2008), Smad3, and Smad2/3 at *Pou5f1*, the gene encoding Oct4 (left) and at *Lefty1* (Amsen et al.).

(B) Smad3 and Smad2/3 co-occupy the genome with Oct4. Region plots show the distribution of Smad2/3- (left) and Smad3- (Amsen et al.) bound regions relative to regions bound by Oct4. A 5 kb window centered on each regions bound by Oct4 is displayed on the y-axis. Red intensity at 0 indicates that Smad2/3- (left) or Smad3- (Amsen et al.) bound regions overlap with Oct4-bound regions.

(C) The Oct4 motif was the top identifiable motif enriched in sites bound by Smad2/3.

Figure S2



DNA motif discovery was performed using bound regions for Smad3 (top) and the associated master transcription factors (bottom) for human ES cells (Oct4), mouse ES cells (Oct4), myotubes (Myod1), and pro-B cells (PU.1). The top four motifs identified for each set of bound regions are displayed.

Materials and Methods

Growth Conditions for Cells

Mouse Embryonic Stem Cells

V6.5 murine embryonic stem (mES) cells were maintained on irradiated murine embryonic fibroblasts (MEFs) and expanded for two passages on gelatinized-tissue culture plates without MEFs prior to all experiments. Cells were grown under standard mES cell conditions as described previously (Marson et al., 2008). Briefly, cells were grown on 0.2% gelatinized (Sigma, G1890) tissue culture plates in ES cell media composed of 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). Cells were grown to 80-90% confluence before crosslinking. For expression analysis, cells were cultured in normal conditions and in the presence of 10 uM SB431542 (Tocris, 1614) solubulized in DMSO. Mouse ES cells were treated with 10 ng/ml Activin (R&D, 338-AC) to activate the TGF-β pathway for one hour prior to crosslinking cells used for the Smad2/3 ChIP-seq and for the second replicate of Smad3. This treatment was necessary to increase the Smad2/3 signal with a weaker antibody. The presence of Activin expanded the number of Smad3-bound regions compared to normal culture conditions (table S1).

Human Embryonic Stem Cells

The human ES cell line (BGO3) was maintained in feeder free conditions using defined media (Ludwig et al., 2006). Cells were grown in a monolayer on tissue culture plates coated with matrigel at the dilution recommended by the manufacturer (BD, 354277). Cells were maintained in mTESR1 media and supplement (Stemcell Technologies, 05850) with 100 U/mL penicillin, 100 μg/mL streptomycin and passaged as described in the Stemcell Techologies protocol. Briefly, cells were washed two times with DMEM/F12 (Invitrogen, 11320) before treatment with dispase (Stemcell Technologies, 07913) for 7 minutes at 37°C. Cells were then washed three times with DMEM/F12

before being resuspend in mTESR1 with a cell lifter (Corning, 3008). Cells for these experiments were between passage 40-60 and had been maintained off feeders for four passages. Cells were grown to 80% confluence before crosslinking.

C2C12 Myoblasts

C2C12 myoblasts (ATCC, CRL-1772) were expanded in C2C12 growth media and differentiated into myotubes as previously described (Caretti et al., 2004; Yaffe and Saxel, 1977). Briefly, cells were cultured in DMEM (Invitrogen, 11965) supplemented with 20% fetal bovine serum, 2 mM L-glutamine, 100 U/mL penicillin, 100 μg/mL streptomycin. To induce differentiation, C2C12 cells were grown to confluence and shifted to differentiation media containing DMEM, 2% horse serum (GIBCO, 26050-070), 1x transferrin/selenium/insulin (GIBCO, 51300-044), 2 mM L-glutamine, 100 U/mL penicillin, 100 μg/mL streptomycin. Multinucleated myotubes were visible after 48 hours of culture in differentiation media. Myotubes were treated with 2.5 ng/ml TGF-β (R&D Systems, 240-B) for 2 hours prior to crosslinking for ChIP-seq analysis or addition of TRIzol reagent (Invitrogen, 15596-026) for genome-wide expression analysis.

38B9 Pro-B Cells

38B9 cells (Ramakrishnan and Rosenberg, 1988) were grown in suspension in RPMI-1640 (Invitrogen, 22400), 10% fetal calf serum, 2 mM L-glutamine 100 U/mL penicillin, 100 μg/mL streptomycin, and 8 nL/mL of 2-mercaptoethanol. Cells were treated with 5 ng/ml of TGF- β for 1 hour prior to prior to crosslinking for ChIP-seq analysis or addition of TRIzol reagent for genome-wide expression analysis. Cells were crosslinked at a concentration of approximately $1x10^6$ per ml.

Chromatin Immunoprecipitation

A summary of the bound regions determined for all ChIP-seq data is contained within table S1.

For ChIP-seq experiments the following antibodies were used: Oct4 (Santa Cruz, sc8628), Smad3 (Abcam, ab28379), Smad2/3 (Gift from D. Wotton), Myod1 (Santa Cruz, sc760), PU.1 (Santa Cruz, sc352).

The protocol for Chromatin Immunoprecipitation (ChIP) was performed as previously described (Lee et al., 2006). Cells were chemically crosslinked by the addition of one-tenth volume of fresh 11% formaldehyde. BG03, V6.5 and C2C12 are adherent and were crosslinked for 10 minutes at room temperature. Cells were rinsed twice with 1X PBS and harvested using a silicon scraper and flash frozen in liquid nitrogen. 38B9 cells were grown in suspension and were crosslinked for 20 minutes at room temperature. Crosslinking was quenched with one-twentieth volume 2.5 M glycine. Cells were washed twice with 1X PBS and flash frozen in liquid nitrogen. Cells were stored at -80°C prior to use. Cells were resuspended, lysed and sonicated to shear and solubilize crosslinked DNA. Appropriate sonication conditions vary depending on cells, culture conditions, crosslinking and equipment.

Sonication was performed on approximately $1x10^8$ cells in sonication buffer (10mM Tris-HCl pH8, 100mM NaCl, 1mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate and 0.5% N-lauroylsarcosine) using a Misonix Sonicator 3000. BGO3 cells, C2C12 cells and 38B9 cells were sonicated at 21 watts for 8 x 20 second pulses (60 second pause between pulses). V6.5 murine ES cells were sonicated under the same conditions for 9 x 20 second pulses. After Sonication, samples were divided in half. 10% Triton-X was added for samples to be precipitated with Oct4, Myod1 and PU.1. No Triton-X was added for Smad3 or Smad2/3 antibodies. Sonicated samples were centrifuged at 20,000g for 10 minutes and the soluble whole cell extracts were incubated overnight with 50 μl of Dynal Protein G magnetic beads that had been pre-incubated with 5 μg of the appropriate antibody. Beads were washed 1X with 20mM Tris-HCl pH8, 150 mM NaCl, 2mM EDTA, 0.1% SDS, 1%Triton X-100, 1X with 20mM Tris-HCl pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1%Triton X-100, 1X, 1X with 10mM Tris-HCl pH8, 250nM LiCl, 2mM EDTA, 1% NP40 and 1X with TE containing 50 mM NaCl. A second wash of TE containing 50mM NaCl was performed for Smad3 and Smad2/3.

Bound complexes were eluted from the beads (50 mM Tris-Hcl, pH 8.0, 10 mM EDTA and 1-0.5% SDS) by heating at 65°C for 45 min and vortexing every 5 minutes. Crosslinking was reversed by incubating samples at 65°C for 6 hrs. Whole cell extract DNA reserved from the sonication step was treated in the same way to reverse crosslinking.

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

DNA was prepared for sequencing according to a modified version of the Illumina/Solexa Genomic DNA protocol. Fragmented 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 1:100 dilution of the Adaptor Oligo Mix (Illumina) was used in the ligation step. A subsequent PCR step with limited (18 cycle) amplification cycles added additional linker sequence to the fragments to prepare them for annealing to the Genome Analyzer flow-cell. After amplification, a narrow range of fragment sizes was selected by separation on a 2% agarose gel and excision of a band between 150-350 bp (representing shear fragments between 50 and 250 nucleotides in length and ~100bp of primer sequence). The DNA was purified from the agarose and diluted to 10 nM for loading on the flow cell. Human and mouse ES cell samples were prepared using the Illumina/Solexa Genomic DNA Kit. C2C12 and 38B9 samples were prepared in a similar manner using individually purchased reagents with the following differences. End repair of fragmented DNA was performed using the End-It-DNA Repair Kit (Epicentre #ER0720). DNA was purified with the Qiaquick PCR Purification Kit (#21806). Purified DNA was treated with Klenow fragment (NEB#M0212) and 1mM dATP for 37°C for 30 minutes to add an A tail. DNA was purified by Qiaquick MiniElute Purification Kit (#28006). Adapters from the Illumina/Solexa Kit were ligated onto the

fragmented DNA with DNA ligase (Promega #M8221) during a 15 minute incubation at room temperature. DNA was purified by MiniElute Purification Kit. PCR was performed using Phusion NZ (NEB#F531) and primers from the Illumina/Solexa Kit.

Polony Generation and Sequencing

The DNA library (2-4 pM) 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 1μm 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 1G. After the first base was incorporated in the Sequencing-by-Synthesis reaction the process was paused for a key 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 re-primed 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, 32, or 36 cycles.

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. Bowtie (version 0.12.2) (Langmead et al., 2009) was used to align sequences to NCBI Build 36 (UCSC mm8) of the mouse genome and NCBI Build 36 (UCSC hg18) of the human genome. Alignments were performed using the following criteria: -n2, -e 70, -m2, -k2, --best. Only sequences that

aligned to a unique location were used to determine enriched regions as described below. ChIP-seq quality score (FASTQ) files profiling the genomic occupancy of Oct4, Sox2, Nanog (Marson et al., 2008), and Zfx (Chen et al., 2008) in mES cells were aligned using Bowtie as described above.

Analysis methods were derived from previously published methods (Marson et al., 2008). Each read (reads from biological replicates were combined) was extended 200bp, towards the interior of the sequenced fragment, based on the strand of the alignment. Across the genome, in 25 bp bins, the number of ChIP-seq reads within a 1kb window surrounding each bin (+/- 500bp) 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 1kb window 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 (40) into which each read maps, divided by the total number of bins available. Enriched bins within 200bp of one another were combined into regions.

The Poissonian background model assumes a random distribution of background reads. However, significant deviations from this expectation have been observed. Some of these non-random events can be detected as sites of apparent enrichment in negative control DNA samples creating false positives. To remove these false positive regions, negative control DNA from whole cell extract (WCE) or IgG ChIPs were sequenced for each cell type. Enriched bins and enriched regions were defined as having greater than five-fold density in the experimental sample compared with the control sample when 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. For mouse, the complete set of RefSeq genes was downloaded from the UCSC website (http://hgdownload.cse.ucsc.edu/goldenPath/mm8/database/refGene.txt.gz) on March 5, 2010. For human, the complete set of RefSeq genes was downloaded from the UCSC table browser (http://genome.ucsc.edu/cgi-bin/hgTables?command=start) on

March 1, 2009. Genes with enriched regions within 10kb upstream of the transcription start site or within the body of the gene were called bound.

A summary of the bound regions (table S1) and bound genes (table S2) for each antibody is provided, and the totals are provided below. These data represent merged biological replicates for C2C12 and 38B9 cells. Data for Smad3 in murine ES cells contains separate analysis for biological replicates. These samples were not combined because the second sample was analyzed after treatment with Activin for 1 hour as described above. Data for BG03 human ES cells and Smad2/3 in mES cells represent single experiments.

Total enriched regions transcription factor:

Human			
Factor	**Cell Type**	**# Enriched Regions**	**#Bound Genes**
Oct4	ESC	7532	4387
Smad3	ESC	5282	4658

Mouse			
Oct4	ESC	15003	8024
Sox2	ESC	15699	7637
Nanog	ESC	16006	6949
Zfx	ESC	14561	12070
Smad3	ESC	2018	1311
Smad2/3	ESC	888	568
Smad3 – Activin	ESC	2269	1426
Myod	Myotube	14678	4387
Smad3	Myotube	1746	1429
PU.1	Pro-B	26349	12277
Smad3	Pro-B	3602	2709

Motif discovery

DNA motif discovery was performed as previously described (Marson et al., 2008). Briefly, a modified version of the ChIP-seq read mapping algorithm was used. The genomic bin size was reduced to 10 bp and the read extension placed greater weight towards the middle of the 200 pb extension. Greater weight was placed toward the middle of the extension to increase the precision of the peak for each region and worked by placing 1/3 count in the 8 bins from 0-40 and 160- 200 bp, 2/3 counts in the 8 bins from 40-80 and 120-160 bp and 1 count in the 4 bins from 80-120 bp. 250 bp of genomic sequence, centered at the 500 largest peaks of ChIP-seq density, were submitted to the motif discovery tool MEME (Bailey and Elkan, 1994) to search for over-represented DNA motifs. The top 4 motifs identified for each antibody is shown in fig. S2.

Background region generation

The distribution of distances from the center of each enriched region in the datasets shown in Fig. 2B (excluding background) to the closest transcription start site was calculated. A set of background regions (Fig. 2B, far right) was generated by selecting 10,000 random genomic locations such that the distribution of distances to the closest transcription start site within the random dataset was highly similar to the observed distribution.

Canonical motif scanning

The genomic sequence +/- 2.5kb from the center of each enriched region in the dataset indicated was downloaded from the UCSC website with repeats masked with "N". A window of 250 bp was shifted across each sequence at 50 bp intervals, and the number of occurrences of the complete Smad2/3 canonical motif "AGAC" or its reverse-complement "GTCT" within the window was counted. The plots in Fig. 1D and 2B show the number of occurrences within each window averaged across all sequences.

Expression arrays

Genomic expression analysis was measured using Agilent Whole-Mouse Genome

Microarrays (Agilent, G4122F). Total RNA was isolated from cells using TRIzol reagent following the protocol for cells grown in monolayer and suspension as appropriate (Invitrogen, 15596-026). RNA was further purified with RNeasy columns (Qiagen, 74104) after DNase treatment (Invitrogen, 18068-015) following the manufacturers' protocols. RNA samples from two biological replicates were used for duplicate microarray expression analysis. Two micrograms of RNA were labeled for each sample using the two-color low RNA Input Linear Amplification Kit PLUS (Agilent, 5188-5340). 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. Labeled cRNA was hybridized overnight at 65°C. 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°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°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 was calculated for each hybridization using Agilent's Feature Extraction Image Analysis software with the default two-color gene expression protocol.

Determination of change in gene expression

Biological replicates of each expression array were generated. To calculate an average dataset from the biological replicates the log10 ratio values for each feature was averaged and the log ratio p-values were multiplied. For each gene in the RefSeq dataset (see above), we used the feature with the median log10 ratio among all the

features annotated to that gene. If there were an even number of features associated to a gene, the middle two features were averaged. The feature(s) used for each gene in each experiment can be found in table S3. Genes without annotated features are reported as NA and were excluded from any expression analysis. A gene was classified as affected by TGF-β if its expression changed by over 50% (absolute log10 ratio of greater than 0.176) and had a p-value for that log ratio less than or equal to 0.05. Genes that changed in expression by more than 50%, but had a p-value greater than 0.05, were classified as unaffected and excluded from the analysis. Fig. 4A is an expression heatmap of the 3067 genes (one per row) affected by TGF-β signaling in at least one of the three cell-types. Genes affected in mESC cells were placed at the top, genes affected in myotubes in the middle and genes affected in pro-B cells at the bottom. Each gene is represented only once even if it was affected in multiple cell types.

Region plots

The visualization in Fig. 1B shows the location of Oct4 and Smad3 binding (x-axis) in relation to regions bound by Oct4 (y-axis). Fig. 2B and Fig.3E show the location of Smad3 binding (x-axis) in relation to regions bound by the identified transcription factor (y-axis). For each bound region in the base dataset (y-axis) the genomic interval +/- 2.5 kb from the center of that enriched region is shown, and any bound region in the query dataset (x-axis) within that 5 kb window is displayed.

Supplemental References

Amsen, D., Antov, A., Jankovic, D., Sher, A., Radtke, F., Souabni, A., Busslinger, M., McCright, B., Gridley, T., and Flavell, R.A. (2007). Direct regulation of Gata3 expression determines the T helper differentiation potential of Notch. Immunity *27*, 89-99.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.

Caretti, G., Di Padova, M., Micales, B., Lyons, G.E., and Sartorelli, V. (2004). The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. Genes Dev *18*, 2627-2638.

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

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

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc *1*, 729-748.

Ludwig, T.E., Bergendahl, V., Levenstein, M.E., Yu, J., Probasco, M.D., and Thomson, J.A. (2006). Feeder-independent culture of human embryonic stem cells. Nat Methods *3*, 637-646.

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

Ramakrishnan, L., and Rosenberg, N. (1988). Novel B-cell precursors blocked at the stage of DJH recombination. Mol Cell Biol *8*, 5216-5223.

Yaffe, D., and Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature *270*, 725-727.

Appendix B

Supplementary Material for Chapter 4

Mediator and Cohesin Connect Gene Expression and Chromatin Architecture

Table of Contents

Oct4_mES, Sox2_mES, Nanog_mES, Pol2_mES, H3K79me2_mES, CTCF_mES

Data File S1 contains data formatted (.WIG) for upload into the UCSC genome browser (Kent et al., 2002). To upload the file, first copy the files onto a computer with internet access. Then use a web browser to go to http://genome.ucsc.edu/cgi-bin/hgCustom?hgsid=105256378 for mouse. In the "Paste URLs or Data" section, select "Browse..." on the right of the screen. Use the pop-up window to select the copied files, then select "Submit". The upload process may take some time.

These files present ChIP-Seq data. 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 0.5 reads per million total reads. Subsequent tracks identify genomic regions identified as enriched.

Extended Experimental Procedures

Cell Culture Conditions

- Embryonic Stem Cells
- Mouse Embryonic Fibroblasts (MEFs)
- ZHBTc4 mES Cells

High-Throughput shRNA Screening

- Library Design and Lentiviral Production
- Lentiviral Infections
- Immunofluorescence
- Image Acquisition and Analysis
- Combining Screening Data (Table S1)
- Criteria for Identifying Screening Hits (Table S2)

Validation of shRNAs

- Lentiviral Production and Infection
- Immunofluorescence
- RNA Extraction, cDNA, and TaqMan Expression Analysis

<u>Supplemental References</u>

Figure S1

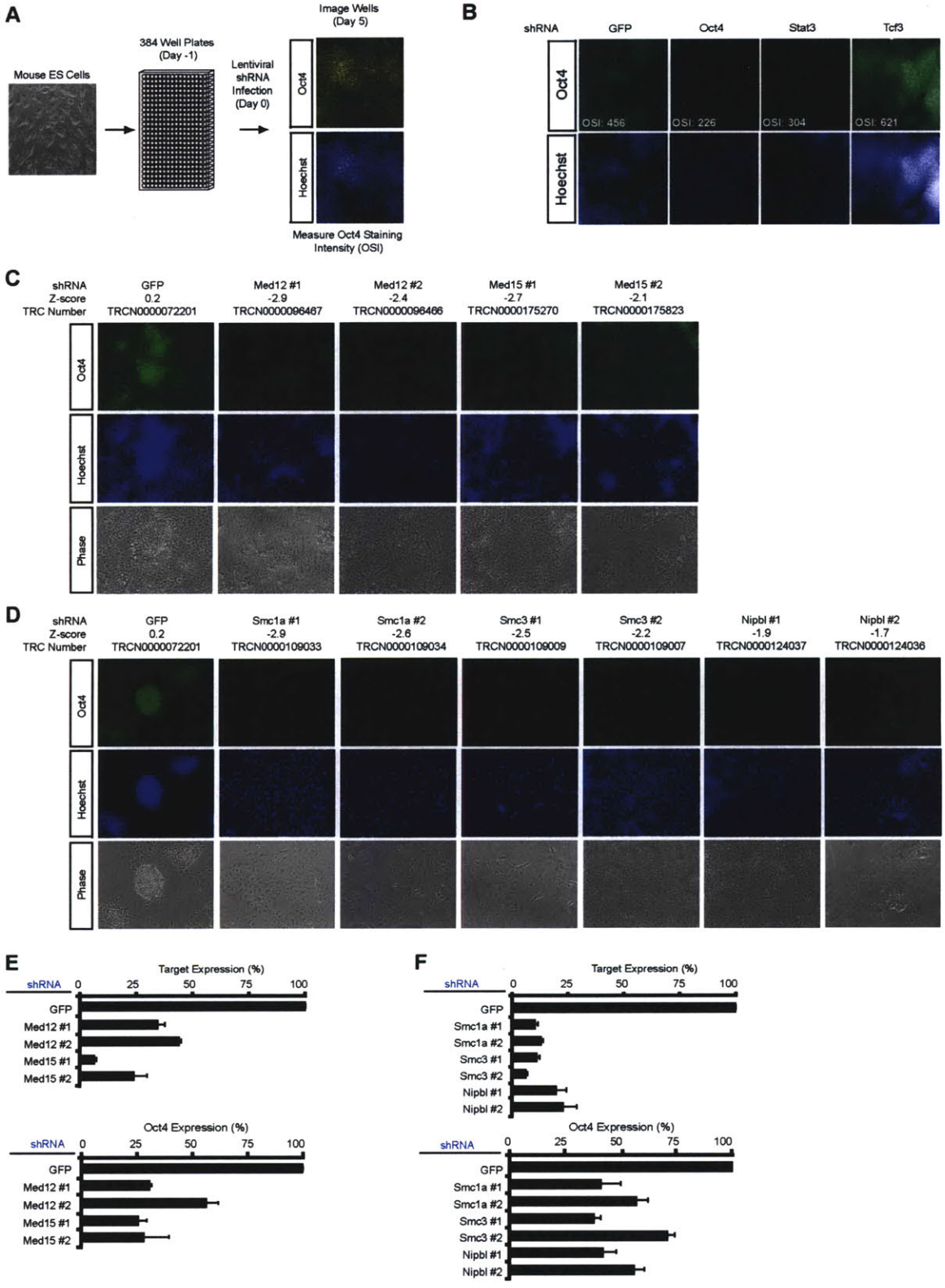


Figure S1

Figure S1. Screening Protocol and Validation of Mediator and Cohesin shRNAs, Related to Figure 1.

(A) Outline of the screening protocol. Murine embryonic stem cells were seeded without a MEF feeder layer into 384-well plates. The following day cells were infected with individual lentiviral shRNAs targeting chromatin regulators and transcription factors. Infections were done in quadruplicate (chromatin regulator set) or duplicate (transcription factor set) on separate plates (Table S1). Five days post-infection cells were fixed and stained with Hoechst and for Oct4. Cells were identified based on the Hoechst staining and the average Oct4 staining intensity was quantified using Cellomics software.

(B) Representative images from control wells on a 384-well plate infected with shRNAs targeting positive regulators of pluripotency (Oct4 and Stat3) and a negative regulator of pluripotency (Tcf3) (Borowiec et al., 1987; Cole et al., 2008; Hay et al., 2004; Nichols et al., 1998; Niwa et al., 2000; Pereira et al., 2006). OSI indicates the average Oct4 staining intensity of the cells in the well.

(C and D) Multiple shRNAs targeting Mediator (C) and Cohesin (D) components reduce Oct4 protein levels and result in changes in colony morphology. Murine ES cells were infected with the indicated shRNA and stained with Hoechst and for Oct4.

(E and F) Effect of multiple Mediator (E) and Cohesin (F) shRNAs on transcript levels for Med12, Med15, Smc1a, Smc3, Nipbl and Oct4. Murine ES cells were infected with the indicated shRNA and transcript levels were evaluated by real-time qPCR.

Figure S2

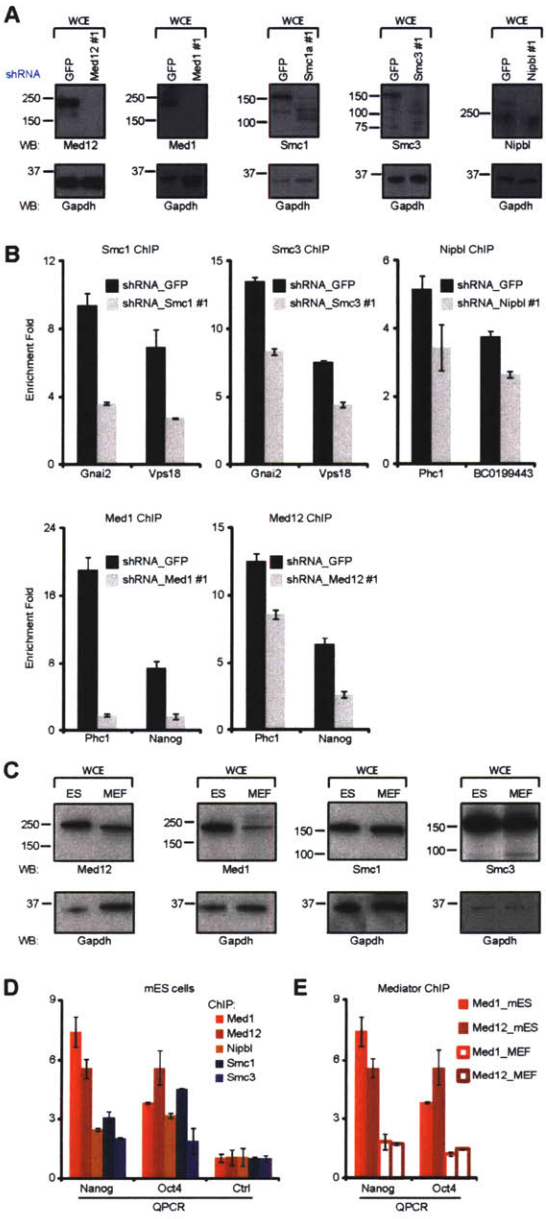


Figure S2

Figure S2. Validation of Mediator, Cohesin and Nipbl Antibodies Used for ChIP-Seq, Related to Figure 2, 3, 6 and 7.

(A) Antibodies against Med12, Med1, Smc1, Smc3 and Nipbl are specific and shRNAs targeting Med12, Med1, Smc1, Smc3 and Nipbl result in reduced levels of the target protein. Murine ES cells were infected with the indicated shRNA and protein levels were determined by western blot analysis.

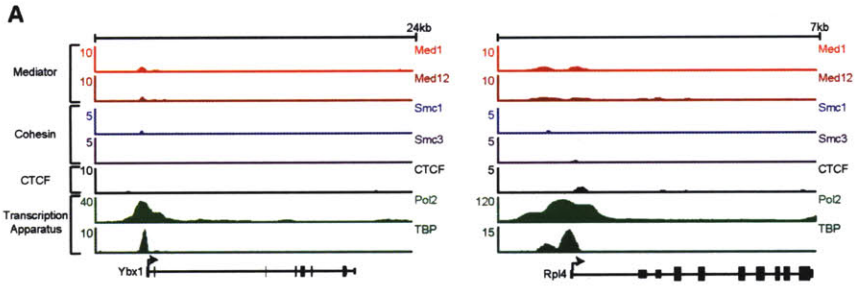
(B) Gene specific ChIPs demonstrating that a reduction in Smc1, Smc3, Nipbl, Med1 and Med12 protein levels by shRNA result in a decreased ChIP signal at the indicated gene. Murine ES cells were infected with the indicated shRNA, gene specific ChIP experiments were performed and analyzed by real-time qPCR.

(C) Antibodies against Med12, Med1, Smc1 and Smc3 are specific. Proteins were detected by western blot analysis from either ES or MEF cell lysates.

(D) Gene specific ChIPs verifying that Mediator, Cohesin and Nipbl occupy the promoter regions of Oct4 and Nanog in ES cells.

(E) Gene specific ChIPs demonstrating that Mediator does not occupy Oct4 and Nanog in MEFs.

Figure S3



B

Pluripotency Genes

Gene Name	Maximum reads/million +/-5kb around TSS Pol2	Med12	Ratio (Pol2/Med12)
Pou5f1	39.4	11.6	3.40
Sox2	48	8.7	5.52
Nanog	48	10.3	4.66
Fgf4	39.6	7.4	5.35
Nodal	40.4	6.6	6.12

Housekeeping/Ribosomal Genes

Gene Name	Maximum reads/million +/-5kb around TSS Pol2	Med12	Ratio (Pol2/Med12)
Ybx1	33	1.7	19.41
Nol5	110.5	1.9	58.16
Rpl4	100.6	1.1	91.45
Rpl10	123.1	1.6	76.94
Rps6	81.4	1.2	67.83
Rps5	123.3	2.2	56.05
Rps7	61.7	1.5	41.13
Mrpl32	35.1	1.7	20.65
Rpl35a	68.8	2	34.40
Rpl11	56.8	1.9	29.89
Rpl37	79.3	2.9	27.34
Actg1	179.3	2.3	77.96
Hist1h1b	309.5	4.4	70.34
Actb	300.9	4.5	66.87
Hnrpa1	116.4	1.8	64.67
Ppia	119.7	1.7	70.41

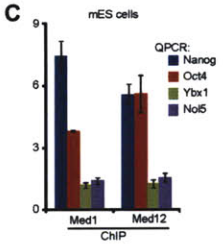


Figure S3

Figure S3. Mediator and Pol2 Occupancy of Ribosomal Protein Genes, Related to Figure 2.

(A) Gene tracks are displayed for a housekeeping (*Ybx1*) and a ribosomal protein gene (*Rpl4)* in ES cells, where there is a high level of Pol2 and TBP occupancy and a low level of Mediator occupancy. ChIP-Seq data is shown in reads/million with the base of the y-axis set 0.5 reads/million. The transcription start site and direction of transcription are noted by an arrow.

(B) Many actively transcribed housekeeping and ribosomal protein genes have low levels of Mediator (Med12) occupancy as opposed to pluripotency genes. The ratio of Pol2 occupancy to Mediator occupancy is significantly higher for the housekeeping and ribosomal genes, when compared to pluripotency genes. Maximum reads/million of Pol2 and Med12 observed within 5kb of the TSS for selected pluripotency, housekeeping and ribosomal protein genes are shown.

(C) Gene specific ChIPs demonstrate that Mediator (Med1 and Med12) does not strongly occupy the promoters of the house keeping genes (*Ybx1* and *Nol5*).

Figure S4

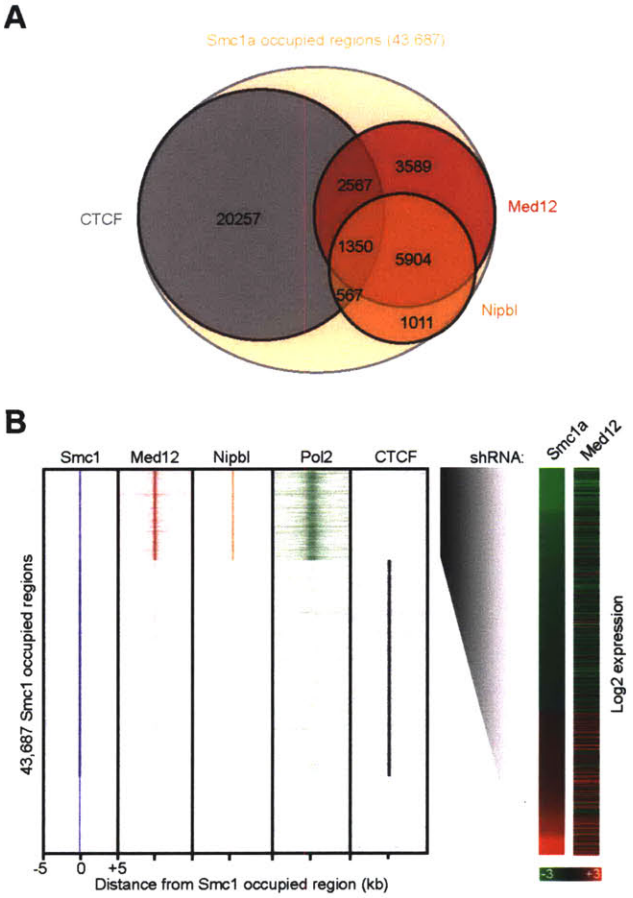


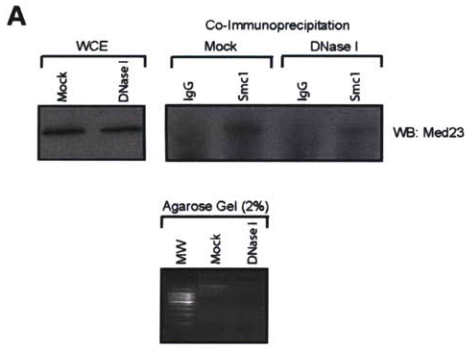
Figure S4

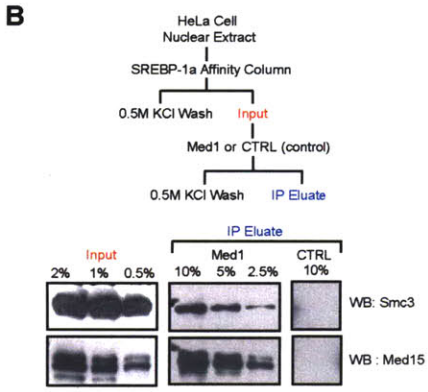
Figure S4. Mediator, Cohesin and Nipbl co-occupy at Mediator and Cohesin Regulated Genes, Related to Figure 2.

(A) Venn diagram demonstrating the overlap of high confidence Nipbl occupied sites with Mediator/Cohesin co-occupied sites. The overlap of high confidence (P-val 10^{-9}) Cohesin sites with Nipbl, Mediator and CTCF is shown. The overlap of Nipbl, Mediator and Cohesin sites is highly significant (P-val < 10^{-300}). Reads from two biological replicate for the Smc1, Med12 and Nipbl ChIP-Seq data sets were combined.

(B) Heat map indicating that regions co-occupied by Smc1 and Med12 are associated with active genes that exhibit similar expression changes with either Smc1a or Med12 knockdown (3 days post knockdown). Log_2 expression data is shown for all Smc1 and Med12 co-occupied regions that could be mapped to a gene. Mapped genes have evidence of a co-occupied Smc1 and Med12 region within the gene body or within 10kb upstream of the transcriptional start site, evidence of Pol2 occupancy within the gene body and significant (Pval <0.01) expression changes in both an Smc1a and Med12 knockdown in independent experiments. The two knockdowns had a strikingly similar effect at this set of genes (Pearson Correlation of 0.68). There is also significant enrichment (P-val = $5.3x10^{-26}$) for Mediator/Cohesin co-occupancy at the genes that change expression for both knockdowns in contrast to CTCF only occupied regions that are not enriched (P-val = 1). The log_2 expression data was ordered based on the Smc1a knockdown and the corresponding expression change for each gene following a Med12 knockdown is shown. A relative signal scale for the expression data is shown at the bottom of both panels.

Figure S5





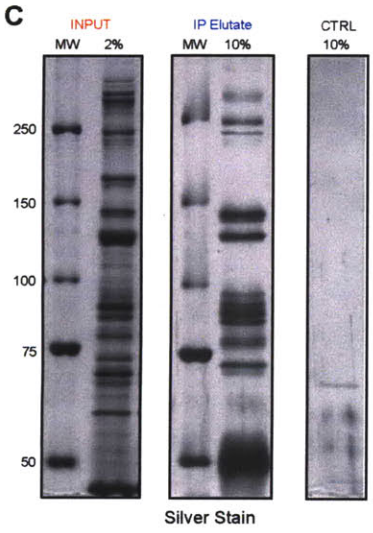


Figure S5

Figure S5. Mediator and Cohesin co-purify, Related to Figure 4.

(A) DNase I treatment does not effect Mediator and Cohesin interaction by co-immunoprecipitation. Cell lysates were Mock or DNase I treated prior to immunoprecipitation. DNase I treatment resulted in a reduction in the amount DNA detected in a whole cell extract (bottom panel).

(B) Cohesin (Smc3) interacts with the Mediator complex purified by Med1. The Mediator complex was initially affinity purified from a HeLa cell nuclear extract utilizing a SREBP-1a activation domain GST fusion. The eluted (Input) material was further purified by immunoprecipitation with a Med1 antibody. The IP Elution was subjected to western blot analysis for Smc3 and Med15.

(C) Silver stained gels from the purification scheme outlined in (B).

Figure S6

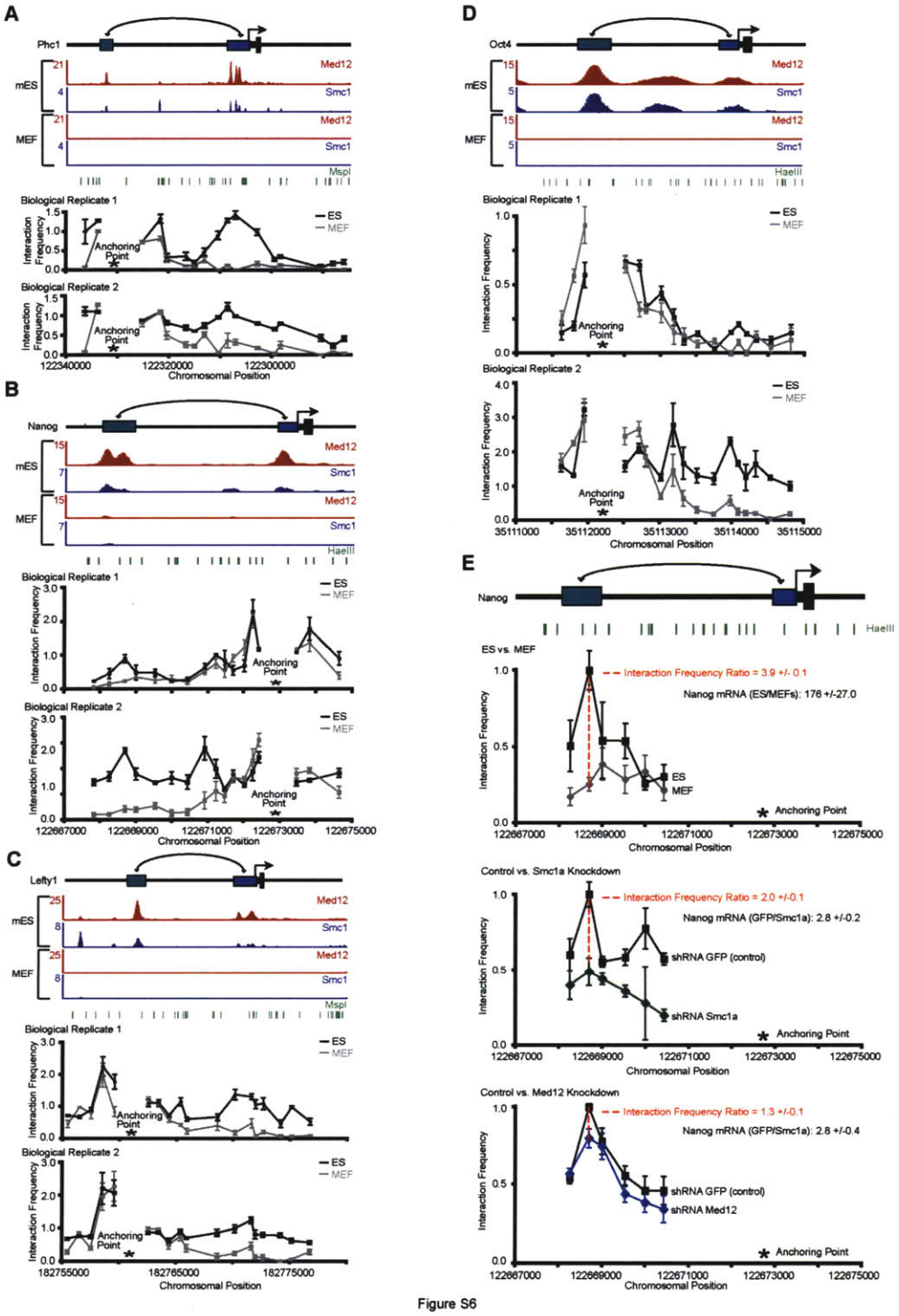


Figure S6

Figure S6. Mediator and Cohesin Binding Profiles Predict Enhancer-Promoter Looping Events, Related to Figure 5.

(A-D) A looping event between the upstream enhancer and the core promoter of *Phc1* (A), *Nanog* (B), *Lefty1* (C) and *Oct4 (Pou5f1)* (D) and was detected by Chromosome Conformation Capture (3C) in ES cells, but not in MEFs. Biological replicates are shown for each locus. ES cell and MEF crosslinked chromatin was digested by the indicated restriction enzyme and religated under conditions that favor intramolecular ligation events. The interaction frequency between the anchoring point and distal fragments was determined by PCR and normalized to BAC templates and control regions. The restriction enzyme sites are indicated above the 3C graph. The error bars represent the standard error of the average of 3 independent PCR reactions. The ChIP-Seq binding profiles for Med12 and Smc1 are shown in reads/million for both ES cells and MEF cells. Reads from two biological replicates for Smc1 (ES cells and MEFs) and Med12 (ES cells) ChIP-Seq datasets were combined.

(E) 3C data demonstrating that the interaction frequency between the promoter and enhancer of Nanog decrease for a Cohesin (Smc1a) or a Mediator (Med12) knockdown. The Interaction Frequency Ratio (red dash) was calculated for each graph using the interaction frequency between primer 4 (within the enhancer) and primer 20 (anchoring primer). A Fold Decrease in Nanog Expression (ES cells to MEFs or shRNA GFP to Knockdown cells) was determined by qPCR and is shown for each graph (See Nanog mRNA). For all graphs the interaction frequency between primer 4 (within the enhancer) and primer 20 (anchoring primer) was normalized to 1 for the ES and shRNA GFP cells. All other interaction frequencies were scaled accordingly.

Extended Experimental Procedures

Cell Culture Conditions

Embryonic Stem Cells

V6.5 murine embryonic stem (mES) cells were grown on irradiated murine embryonic fibroblasts (MEFs) unless otherwise stated. Cells were grown under standard mES cell conditions as described previously (Boyer et al., 2005). 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).

Mouse Embryonic Fibroblasts (MEFs)

Low passage MEFs were grown on tissue culture plates DMEM (Invitrogen, 11965) supplemented with 10% fetal bovine serum (Hyclone, characterized SH3007103), 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).

ZHBTc4 mES Cells

ZHBTc4 Oct4 shutdown cells (Niwa et al., 2000) were grown under standard mES cell conditions, expanded off of MEF feeders for two passages and treated with 2 μg/ml doxycycline for 24 hours prior to formaldehyde crosslinking or protein extraction.

High-Throughput shRNA Screening

Library Design and Lentiviral Production

Small hairpins targeting 197 chromatin regulators and 2021 transcription factors were designed and cloned into pLKO.1 lentiviral vectors as previously described (Moffat et al., 2006). On average 5 different shRNAs targeting each chromatin regulator or transcription factor were used. Lentiviral supernatants were arrayed in 384-well plates with negative control lentivirus (shRNAs targeting GFP, RFP, Luciferase and LacZ) (Moffat et al., 2006).

Lentiviral Infections

Murine ES cells were split off MEFs and placed in a tissue culture dish for 45 minutes to selectively remove the MEFs. Murine ES cells were counted with a Coulter Counter (Beckman, #1499) and seeded using a μFill (Bioteck) at a density of 1500 cells/well in 384-well plates (Costar 3712) treated with 0.2% gelatin (Sigma, G1890). An initial cell plating density of 1500 cells/well was established so that an adequate amount of cells would survive puromycin selection for analysis. However, the initial cell plating density was kept low enough to avoid wells reaching confluency during the timeframe of the assay. One day following cell plating the media was removed, replaced with ESC media containing 8 μg/ml of polybrene (Sigma, H9268-10G) and cells were infected with 2 μl of shRNA lentiviral supernatant. Infections were performed in duplicate (transcription factor set) or quadruplicate (chromatin regulator set) on separate plates. Table S1 denotes which screening set the shRNAs were in. Control wells on each plate were mock infected and designated as "Empty". Positive control wells on each plate were infected with 3 μl of validated control shRNA lentiviral supernatant targeting Oct4 (TRCN0000009613), Tcf3 (TRCN0000095454) and Stat3 (TRCN0000071454) that was generated independently of the screening sets (Lentiviral Production and Infection). Sequence and shRNAs are available from Open Bioystems. Plates were spun for 30 minutes at 2150 rpm following infection. Twenty-four hours post infection cells were treated with 3.5 μg/ml of puromycin (Sigma, P8833) in ESC media to select for stable integration of the shRNA construct. ESC media with puromycin

was changed daily. Five days post infection cells were crosslinked for 15 minutes with 4% paraformaldehyde (EMS Diasum, 15710).

Immunofluorescence

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°C for Oct4 (Santa Cruz Biotechnology, sc-5279; 1:100 dilution) and washed twice with blocking buffer. Cells were incubated for 4 hours at room temperature with goat anti-mouse-conjugated Alexa Fluor 488 (Invitrogen; 1:200 dilution) and Hoechst 33342 (Invitrogen; 1:1000 dilution). Finally, cells were washed twice with blocking buffer and twice with PBS before imaging.

Image Acquisition and Analysis

Image acquisition and data analysis were performed essentially as previously described (Moffat et al., 2006). Stained cells were imaged on an Arrayscan HCS Reader (Cellomics) using the standard acquisition camera mode (10X objective, 9 fields). Hoechst was used as the focus channel. Objects selected for analysis were identified based on the Hoechst staining intensity using the Target Activation Protocol and the Fixed Threshold Method. Parameters were established requiring that individual objects pass an intensity and size threshold. The Object Segmentation Assay Parameter was adjusted for maximal resolution between individual cells. Following object selection, the average Oct4 pixel staining intensity was determined per object and then a mean value for each well was calculated. Image acquisition for a well continued until at least 2500 objects were identified, the entire well (9 fields) was imaged or less than 20 objects were identified for three fields imaged in a row. To account for viability defects or low titer lentivirus for the chromatin regulator screening set an shRNA was excluded from subsequent analysis if less than 250 objects were identified for any one of the 4 replicates. The 250 identified objects threshold was determined based on

the average number of identified objects for the "Empty" (no virus) wells (mean: 53.4, standard deviation: 49.3). To account for viability defects or low titer lentivirus for the transcription factor screening set a shRNA was excluded from subsequent analysis if less than 300 objects were identified for any one of the 2 replicates. The 300 identified objects threshold was determined based on the average number of identified objects for the "Empty" (no virus) wells (mean: 39.2, standard deviation: 147.5).

To normalize for plate effects, a Z-score based on the Oct4 staining intensity was calculated for each well using the following negative control infections, 24 different shRNAs targeting GFP, 16 different shRNAs targeting RFP, 25 different shRNAs targeting Luciferase and 20 different shRNAs targeting LacZ. There were a total of between 16 and 22 wells infected with various negative control shRNAs on each 384-well plate, with the exception of one plate within the transcription factor set that contained 99 wells with control infections. The average Oct4 staining intensity for the negative control infected wells was calculated along with a standard deviation to give an estimation of the amount of the signal variability. The average Oct4 staining intensity for all the negative control infected wells on a plate and the standard deviation were utilized to calculated a Z-score for every well on the plate. The Z-scores for the four quadruplicate infections (chromatin regulator set) or two duplicate infections (transcription factor set) were averaged for a final Z-score for every shRNA. The Z-score data for both sets were combined (Table S1). Representative control 384-well plate images (shRNAs targeting Oct4, Stat3, Tcf3 and GFP) were exported (Cellomics Software), converted from DIBs to TIFs (CellProfiler, http://www.cellprofiler.org), and manipulated with Photoshop CS3 Extended (Figure S1B).

Combining Screening Data (Table S1)

We recently published the results of an ES screen where 197 chromatin regulators were selectively targeted for knockdown (Bilodeau et al., 2009). For the present study we screened an additional 2021 genes encoding transcription

factors and chromatin regulators. In order to generate a more complete picture of factors required for maintaining ES cells state we included the set of chromatin regulator results from the previous study. The shRNAs from each set are denoted in Table S1.

The same methodology was followed for screening with both the chromatin regulator and transcription factor sets with the following exception, infections for the chromatin regulator set were done in quadruplicate and infections for the transcription factor set were carried out in duplicate, due to the large size of the transcription factor screening set (30 x 384-well plates, 2021 genes). Because the average Z-scores of the added controls (Oct4 and Stat3) were within close proximity for both screening sets (Chromatin Regulator Set: -3.3 and –2.4 for Oct4 and Stat3 respectively; Transcription Factor Set: -3.0 and –2.1 for Oct4 and Stat3 respectively) we reasoned that Z-scores between the two screening sets were comparable.

Criteria for Identifying Screening Hits (Table S2)

We used multiple Z-score level thresholds to select chromatin regulators and transcription factors that had significantly reduced Oct4 levels for inclusion in Table S2. First, a chromatin regulator or transcription factor had to have at least two shRNA with a Z-score less than –1.5 and it was possible to classify the gene based on the literature. Second, a chromatin regulator or transcription factor with a single shRNA hit and a Z-score of less than –1.5 was also included if it could be classified with one of the multiple shRNA hits. Third, the following chromatin regulators (Cbx7, Cbx8/Pc3 and Ezh2) were included even though each was only a single shRNA hit, because all had strong negative Z-scores, all are polycomb proteins, and polycomb has been previously demonstrated to be important for regulating ES cell (Boyer et al., 2006). The -1.5 cut-off was chosen because it was within close proximity to the Z-score of the Stat3 controls (-2.4 and –2.1 for the chromatin regulator and the transcription factor sets respectively).

Validation of shRNAs

Lentiviral Production and Infection

Lentivirus was produced according to Open Biosystems *Trans*-lentiviral shRNA Packaging System (TLP4614). The shRNA constructs targeting Med12, Med15, Smc1a, Smc3, Nipbl, Oct4, Stat3 and Tcf3 are listed below. All are available, including sequences from Open Biosystems. The shRNA targeting GFP (TRCN0000072201, Hairpin Sequence: gtcgagctggacggcgacgta) was one of the negative controls included on all plates for the screen.

Smc1a #1	TRCN0000109033
Smc1a #2	TRCN0000109034
Smc3 #1	TRCN0000109009
Smc3 #2	TRCN0000109007
Nipbl #1	TRCN0000124037
Nipbl #2	TRCN0000124036
Med12 #1	TRCN0000096467
Med12 #2	TRCN0000096466
Med15 #1	TRCN0000175270
Med15 #2	TRCN0000175823
Oct4	TRCN0000009613
Stat3	TRCN0000071454
Tcf3	TRCN0000095454

For validation of the Mediator and Cohesin shRNAs, mES cells were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs and then plated in 6-well plates (200,000 cells/well). 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. Five days post infection RNA or proteins were extracted or the cells were crosslinked for immunofluorescence.

Immunofluorescence

Cells were crosslinked, permeabilized and stained as described for high-throughput screening. Images were acquired on a Nikon Inverted TE300 with a Hamamatsu Orca camera. Openlab (http://www.improvision.com/products/openlab/) was used for image acquisition. Openlab and Photoshop CS3 Extended were used for image manipulation.

RNA Extraction, cDNA, and TaqMan Expression Analysis

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
Med12	Mm00804032_m1
Med15	Mm01171155_m1
Smc1a	Mm01253647_m1
Smc3	Mm00484012_m1
Nipbl	Mm01297461_m1
Oct4	Mm00658129_gH
Nanog	Mm02384862_g1

Expression levels were normalized to Gapdh levels. All knockdowns are relative to control shRNA GFP infections.

Protein Extraction and Western Blot Analysis

Cells were lysed with CelLytic Reagent (Sigma, C2978-50ml) containing protease inhibitors (Roche), proteins were separated by SDS-PAGE and Western blots were revealed with antibodies against Med1 (Bethyl, A300-793A), Med12 (Bethyl, A300-774A), Smc1 (Bethyl, A300-055A), Smc3 (Abcam, ab9263), Nipbl (Bethyl, A301-778A) or Gapdh (Abcam, ab9484).

Chromatin Immunoprecipitation

A summary of the ChIP-Seq data is contained within Table S7.

For Med1 (CRSP1/TRAP220) occupied genomic regions, we performed ChIP-Seq experiments using Bethyl Laboratories (A300-793A) antibody. The affinity purified antibody was raised in rabbit against an epitope corresponding to amino acids 1523-1281 mapping at the C-terminus of human Med1.

For Med12 occupied genomic regions, we performed ChIP-Seq experiments using Bethyl Laboratories (A300-774A) antibody. The affinity purified antibody was raised in rabbit against an epitope corresponding to amino acids 2150-2212 mapping at the C-terminus of human Med12.

For Smc1 occupied genomic regions, we performed ChIP-Seq experiments using Bethyl Laboratories (A300-055A) affinity purified rabbit polyclonal antibody. The epitope recognized by A300-055A maps to a region between residue 1175 and the C-terminus of human Smc1.

For Smc3 occupied genomic regions, we performed ChIP-Seq experiments using Abcam (ab9263) antibody. The affinity purified antibody was raised in rabbit against an epitope corresponding to the last 100 amino acids of the human Smc3 protein.

For TBP occupied genomic regions, we performed ChIP-Seq experiments using Abcam (ab818) antibody. The antibody was raised with a synthetic peptide which represents amino acid residues 1-20 of human TBP.

For Pol2 occupied genomic regions, we performed ChIP-Seq experiments using Covance 8WG16 antibody. This mouse monoclonal antibody was raised against the C-terminal heptapeptide repeat region on the largest subunit of Pol2, purified from wheat germ extract.

For H3K79me2 occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab3594 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide that is within residues 50 to the C-terminus of Human Histone H3, di methylated at K79.

For CTCF occupied genomic regions, we performed ChIP-Seq experiments using an Upstate 07-729 rabbit polyclonal antibody.

For Nipbl occupied genomic regions, we performed ChIP-Seq experiments using a Bethyl A301-778A rabbit polyclonal antibody. The affinity purified antibody was raised in rabbit to a region between amino acid residues 550 and 600 of human Nipbl.

Protocols describing chromatin immunoprecipitation materials and methods have been previously described (Boyer et al., 2006). Embryonic stem cells 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 –80°C 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 Nipbl, Smc1, Smc3, Pol2, H3K79me2 and Med1 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°C with 100 μl of Dynal Protein G magnetic beads that had been pre-incubated with approximately 10 μg 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, 250nM LiCl, 2mM EDTA, 1% NP40 and 1X with TE containing 50 mM NaCl.

For Med12 and CTCF, the sonication buffer was 10mM Tris-HCl pH8, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine. We used the same sonication and wash conditions as described above.

For TBP, the sonication buffer was 10mM Tris-HCl pH8, 100mM NaCl, 1mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate and 0.5% N-lauroylsarcosine. We used a Misonix Sonicator 3000 and sonicated at approximately 24 watts for 10 x 30 second pulses (60 second pause between pulses). After Sonication, 10% Triton-X was added. After immunoprecipitation, beads were washed 4X with the RIPA buffer (50 mM Hepes-KOH pH 7.6, 500 mM LiCl, 1 mM EDTA, 1% NP40 and 0.7% Na-Deoxycholate) 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°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.

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

DNA was prepared for sequencing according to a modified version of the Illumina/Solexa Genomic DNA protocol. Fragmented 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 1:100 dilution of the Adaptor Oligo Mix (Illumina) was used in the ligation step. A subsequent PCR step with limited (18) amplification cycles added additional linker sequence to the fragments to prepare them for annealing to the Genome Analyzer flow-cell. After amplification, a narrow range of fragment sizes was selected by separation on a 2% agarose gel and excision of a band between 150-350 bp (representing shear fragments between 50 and 250nt in length and ~100bp of primer sequence). The DNA was purified from the agarose and diluted to 10 nM for loading on the flow cell.

Polony Generation and Sequencing

The DNA library (2-4 pM) 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 1μm 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 1G. After the first base was incorporated in the Sequencing-by-Synthesis reaction the process was paused for a key 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 re-primed 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 or 32 cycles.

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. Sequences were aligned using ELAND software to NCBI Build 36 (UCSC mm8) of the mouse genome. Only sequences that mapped uniquely to the genome with zero or one mismatch were used for further analysis. When multiple reads mapped to the same genomic position, a maximum of two reads mapping to the same position were used. A summary of the total number of ChIP-Seq reads that were used in each experiment is provided (Table S7). ChIP-Seq datasets profiling the genomic occupancy of H3K79me2 (Marson et al., 2008), Oct4 (Marson et al., 2008), Sox2 (Marson et al., 2008), Nanog (Marson et al., 2008), RNA polymerase II (Seila et al., 2008) and CTCF (Chen et al., 2008) in mES cells were obtained from previous publications and reanalyzed using the methods described below.

Analysis methods were derived from previously published methods (Guenther et al., 2008; Johnson et al., 2007; Marson et al., 2008; Mikkelsen et al., 2007). Sequence reads from multiple flow cells for each IP target and/or biological replicates were combined. For all datasets, excluding Pol2 and H3K79me2, each read was extended 200bp, towards the interior of the sequenced fragment, based on the strand of the alignment. For Pol2 and

H3K79me2 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 1kb window 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 (40) into which each read maps, divided by the total number of bins available (we estimated 70%). 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 (Table S4 and S5).

ChIP-Seq Density Maps (Figure 2A and 6A)

Selected genes were aligned with each other according to the position and direction of their transcription start site. For each experiment, the ChIP-Seq density profiles were normalized to the density per million total reads. Genes were sorted as indicated.

ChIP-Seq Enriched Region Maps (Figure 2D; Figure 6D; Figure 7A and 7B)

The visualizations in Figure 2D, Figure 6D, Figure 7A and Figure 7B show the location of enriched regions in a collection of datasets (query datasets, indicated on the top) in relation to the enriched regions of another dataset (base dataset, indicated on the y-axis). For each of the enriched regions in the base dataset, corresponding genomic regions were calculated as +/- 5kb from the center of that enriched region (one genomic region per enriched region, row). For each of these genomic regions, the location and length of any enriched regions in the query datasets were drawn.

Assigning ChIP-Seq Enriched Regions to Genes (Table S5)

The complete set of RefSeq genes was downloaded from the UCSC table browser (http://genome.ucsc.edu/cgi-bin/hgTables?command=start) on December 20, 2008. For all datasets, excluding Pol2 and H3K79me2, genes with enriched regions within 10kb of their transcription start site, or within the gene body were called bound. For Pol2 and H3K79me2 datasets, genes with enriched regions within the gene body were called bound (Table S5).

Calculations of Med12 and Smc3 Reduction at Oct4 co-occupied sites or Smc3 reduction at CTCF/Smc3 sites (Figure 3)

To determine if there was a reduction in occupancy of Med12 or Smc3 at Oct4, Med12 and Smc3 co-bound regions following the loss of Oct4, we compared the normalized peak binding density (in reads/million) of each factor in each co-occupied region across the genome. For each region co-occupied by Oct4, Med12 and Smc3 in ES cells, the log2 ratio of the normalized peak heights between ES cells and Oct4 shutdown cells was calculated. Peak heights were normalized for each factor (Med12 or Smc3) such that the average peak height (in reads/million) in enriched regions was the same in ES cells and Oct4 shutdown cells for that factor. This normalization was done to correct for any bias due to differing qualities of IPs which would effect the perceived density of

binding in reads/million. The calculation for change in Smc3 in CTCF/Smc3 co-occupied regions was calculated similarly, except the regions analyzed were those defined by the overlap between CTCF/Smc3 in ES cells. The log_2 ratio of Med12 and Smc3 for each Oct4/Med12/Smc3 co-occupied region, and the log_2 ratio of Smc3 for each CTCF/Smc3 co-occupied region is shown in Table S6.

Note regarding Bound Gene Table (Table S5)

Table S5 provides binding information on every entry in the RefSeq table downloaded on December 20, 2008 (See ChIP-Seq analysis above) and the bound gene numbers reflect counts of these entries. It should be noted however, that some of the gene names are not unique and thus the density maps in Figure 2A and 6A may have fewer rows than there are entries in Table S5.

Note Regarding Calculation of Co-occupied Regions

Table S4 contains the genomic coordinates of enriched regions co-occupied by the indicated pair of factors. These coordinates are the union of all overlapping enriched regions of the two factors. It is possible for an enriched region of one factor to span, or bridge a gap between, two separate enriched regions of the other factor, in those cases, only one enriched region would be reported and it would be the union of all three enriched regions. This will cause the number of reported co-bound regions to be less than the number of strictly overlapping sites reported in the Venn diagrams of Figure 2C and 6C. The Venn diagrams are strictly the number of Smc1 sites that are partially overlapped by either CTCF or Mediator.

Gene Specific ChIPs

Gene specific ChIPs were performed in the indicated cell type following the protocol outlined in ChIP-Seq Sample Preparation. For the Gene specific ChIPs carried out in the knockdown cells, approximately $8x10^6$ ES cells in 5 x 10cm tissue culture plates were infected with the indicated shRNA as described (Validation of shRNAs). Syber 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. Primers to the genes tested and control regions are listed below.

Gnai2
5'- ACAGAGCGATACGGCTCAGCAA-3'
5'-AAGTGGTAGCCGAAGGCAAGTGAA-3'

Vps18
5'-TCCTAGCGCCAACATGAGGAACT-3'
5'-TTTCAGCCGCGAGTGTTAACTGGA-3'

Phc1
5'-TTTGCTCTGCGTGACACTGAAGGT-3'
5'-AAATCCCAGCGCTTCTAGACGTAG-3'

BC0199443
5'-TGCCCACGTCGTAACAAGGTTT-3'
5'-AAGGCCGATCCTTTCTGGTTCA-3'

Nanog
5'-ATAGGGGGTGGGTAGGGTAG-3'
5'-CCCACAGAAAGAGCAAGACA-3'

Oct4
5'-TTGAACTGTGGTGGAGAGTGCT-3'
5'-TGCACCTTTGTTATGCATCTGCCG-3'

Ybx1
5'-AGATCCTGGACCGACTTCC-3'
5'-GTTCCCAAAACCTTCGTTG-3'

Nol5

5'-GGCTCCGAAAAGATGTGAA-3'

5'-AGCAGAGGTCGCCCTAAAT-3'

Ctrl

5'-TGGGTGCCGTATGCCACATTAT-3'

5'-TTTCTGGCCATCCGCACCTTAT-3'

ChIP-Western, Co-Immunoprecipitation and DNase I Treatment

ChIP-Western and Co-Immunoprecipitation

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°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°C using 5ug of Med1 (Bethyl, A300-793A), Med12 (Bethyl, A300-774A), Nipbl (Bethyl, A301-778A), Smc1 (Bethyl, A300-055A), Smc3 (Abcam, Ab9236) 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 Med23 (Bethyl, A300-425A), Oct4 (Santa Cruz, sc-5279), Smc1 (Bethyl, A300-055A), Smc3 (Abcam, Ab9236) and Nipbl (Bethyl, A301-778A).

Co-Immunoprecipitation Following DNase I treatment

Murine ES cells were harvested in cold PBS and lysed with 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% Glycerol, 1% NP-40, 2 mM $MgCl_2$ and protease

inhibitors. After centrifugation, lysates were mock treated or treated with DNase I (Sigma, AMP-D1) at room temperature for 45 minutes. Protein complexes were immunoprecipitated overnight at 4°C using 2ug of Smc1 (A300-055A, Bethly) or Rabbit IgG (Upstate, 12-370) bound to 20ul of Dynabeads®. Immunoprecipitates were washed four times with the lyses buffer, proteins were separated by SDS-PAGE and Western blots were revealed with antibodies against Med23 (Bethyl, A300-425A) and Smc1 (Bethyl, A300-055A). Post DNase I treatment, DNA from 5% of the mock and DNase I treated lysates were isolated by phenol:chloroform extraction following RNase A (0.2 mg/ml, 2 hours, 37°C) and Proteinase K (0.2 mg/ml, 2 hours, 55°C) treatment.

Protein Extraction and Western Blot Analysis

ES and MEF cells were lysed with CelLytic Reagent (Sigma, C2978-50ml) containing protease inhibitors (Roche). After SDS-PAGE, Western blots were revealed with antibodies against Med1 (Bethyl, A300-793A), Med12 (Bethyl, A300-774A), Smc1 (Bethyl, A300-055A), Smc3 (Abcam, ab9263) or Gapdh (Abcam, ab9484).

Mediator Affinity Purification

ES Cells

The Mediator complex was purified from mES cell nuclear extracts using immobilized GST-SREBP-1a (residues 1-50). Bound material washed 4x with 20 column volumes of 0.5M KCl HEGN (20mM Hepes, 0.1mM EDTA, 10% Glycerol, 0.1% NP-40 & 0.5M KCl) buffer, 2x with 0.15M KCl HEGN buffer, and eluted. The eluted sample was further purified with a CDK8 antibody. After binding, this resin was washed 4x with 50 column volumes of 0.5M KCl HEGN buffer, 2x with 0.1M KCl HEGN buffer and eluted with 0.1M Glycine, pH 2.75. Western blot analysis was conducted with Smc3 (Abcam ab9263-50), Med15 (Taatjes Lab stock), Med12 (Bethyl A300-77A) or Nipbl (Bethyl A301-778A) antibodies.

HeLa Cells

GST-SREBP-1a (residues 1-50) was immobilized to GSH-Sepharose beads (GE Lifesciences) and used as bait for overnight pull downs (4°C) from an ES cell or HeLa cell nuclear extract. Bound material was washed with 4x with 20 column volumes of 0.5M KCl HEGN (20mM Hepes, 0.1mM EDTA, 10% Glycerol, 0.1% NP-40 & 0.5M KCl) and 2x with 20 column volumes of 0.15M KCl HEGN. Bound material was eluted with 30mM GSH in elution buffer (80mM Tris, 0.1mM EDTA, 10% Glycerol, 0.02% NP-40, 100mM KCl). GSH elutions were immunoprecipitated with CKD8 (Santa Cruz Biotechnology C-19) or MED1 (Santa Cruz Biotechnology M-255) antibodies immobilized to Protein A/G-Sepharose beads (GE Lifesciences), washed with 3x 20 column volumes 0.5M KCl HEGN and eluted with 0.1M Glycine, pH 2.75. Western blot analysis was conducted with Smc3 (Abcam, ab9263-50), Med15 (Taatjes lab stock), Med12 (Bethyl, A300-774A) and Nipbl (Bethyl, A301-778A) antibodies.

Chromosome Conformation Capture (3C)

3C analysis was performed essentially as described by Miele *et al.* (Miele et al., 2009) with a few modifications. 10^8 mES or MEF cells were crosslinked as described (ChIP-Seq Sample Preparation and Analysis). Cells were lysed and chromatin was digested with 1000 units HaeIII (NEB) for the *Nanog* and *Oct4* loci or 2000 units MspI (NEB) for the *Phc1* and *Lefty1* loci. Crosslinked fragments were subsequently ligated with 50 units T4 DNA ligase (Invitrogen) for 4 hours at 16°C. A control template was generated using a BAC clone (RP23-474F18) covering the Nanog locus, a BAC clone (RP24-352O13) covering the Phc1 locus, a BAC clone (RP23-438H19) covering the Oct4 locus and a BAC clone (RP23-230B21) covering the Lefty1 locus. Ten μg of BAC DNA was digested with 2000 units HaeIII or 1800 units MspI. Random ligation of the fragments was done with 5 units T4 DNA ligase in a total volume of 60 microliters. 3C primers were designed for fragments both upstream and downstream of the transcription start site within HaeIII or MspI fragments. Primers Nanog 20, Phc1 48, Oct4 346 and Lefty1 5 were used as the anchor points (Table S8). 3C analysis was done, in which every PCR for a primer pair was done in triplicate and quantified. Each

data point was first corrected for PCR bias by dividing the average of three PCR signals by the average signal in the BAC control template. Data from ES cells and MEFs were normalized to each other using the interaction frequencies between fragments in the control regions with the following primer pairs for the Nanog locus (Biological Replicate 1 and 2); Acta2 11 and Acta2 16, Acta2 48 and Acta2 52, Gapdh 17 and Gapdh 19, Gapdh 17 and Gapdh 21, Gapdh 17 and Gapdh 32, Gapdh 21 and Gapdh 39, Gene Desert 5 and Gene Desert 6, Gene Desert 12 and Gene Desert 14, Gene Desert 25 and Gene Desert 26, Gene Desert 12 and Gene Desert 26. The following primer pairs were used for normalization between ES cells and MEFs for the Phc1 locus (Biological Replicate 1); Gene Desert 0 and Gene Desert 1, Gene Desert 0 and Gene Desert 2, Gene Desert 27 and Gene Desert 28, Phc1 47 and Phc1 48, Phc1 48 and Phc1 49. The following primer pairs were used for normalization between ES cells and MEFs for the Phc1 locus (Biological Replicate 2); Gene Desert 0 and Gene Desert 1, Gene Desert 0 and Gene Desert 2, Gene Desert 27 and Gene Desert 28, Acta2 0 and Acta2 1, Acta2 2 and Acta2 7, Acta2 8 and Acta2 9, Acta2 0 and Acta2 13, Gapdh 0 and Gapdh 2, Gapdh 7 and Gapdh 8, Gapdh 9 and Gapdh 12, Gapdh 4 and Gapdh 12. The following primer pairs were used for normalization between ES cells and MEFs for the Oct4 locus (Biological Replicate 1); Acta2 11 and Acta2 16, Gapdh 17 and Gapdh 19, Gapdh 17 and Gapdh 21, Gapdh 21 and Gapdh 39, Gene Desert 5 and Gene Desert 6, Gene Desert 12 and Gene Desert 14, Gene Desert 25 and Gene Desert 26, Oct4 346 and Oct4 344, Oct4 346 and Oct4 348. The following primer pairs were used for normalization between ES cells and MEFs for the Oct4 locus (Biological Replicate 2); Gapdh 17 and Gapdh 19, Gapdh 17 and Gapdh 21, Gapdh 21 and Gapdh 39, Gene Desert 5 and Gene Desert 6, Gene Desert 12 and Gene Desert 14, Gene Desert 25 and Gene Desert 26, Oct4 346 and Oct4 344, Oct4 346 and Oct4 348. The following primer pairs were used for normalization between ES cells and MEFs for the Lefty1 locus; Gene Desert 0 and Gene Desert 1, Gene Desert 0 and Gene Desert 2, Gene Desert 27 and Gene Desert 28, Acta2 0 and

Acta2 1, Acta2 8 and Acta2 9, Acta2 0 and Acta2 13, Gapdh 0 and Gapdh 2, Gapdh 7 and Gapdh 8, Gapdh 9 and Gapdh 12, Gapdh 4 and Gapdh 12.

A normalization factor was determined by calculating the log ratio of each interaction frequency within the control region in mES over MEFs, followed by calculating the average of all log ratios. The raw interaction frequencies in mES were subsequently normalized to MEFs using this factor.

Microarray Analysis

Cell Culture and RNA isolation

For ES cell knockdown expression analysis, ES cells 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 infected with lentiviral shRNAs targeting GFP, Smc1a #1 or Med12 #1 (See Validation of shRNAs) 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). Five days post infection RNA was isolated with TRIzol (Invitrogen, 15596-026), further purified with RNeasy columns (Qiagen, 74104) and DNase treated on column (Qiagen, 79254) following the manufacturer's protocols. RNA from two biological replicates were used for duplicate microarray expression analysis, except for the day 3 knockdown expression data (Figure S4B) which is a singlicate data set.

For MEF knockdown expression analysis, MEFs were cultured in 6-well plates and infected as described for the ES cells except that 2.0 μg/mL puromycin (Sigma, P8833) was used for selection. RNA was isolated and treated as described above.

Micorarray hybridization and Analysis

For microarray analysis, Cy3 and Cy5 labeled cRNA samples were prepared using Agilent's QuickAmp sample labeling kit starting with 1 μg 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°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°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. To calculate an average dataset from the biological replicates the log10 ratio values for each feature were averaged and the log ratio p-values were multiplied. For each gene in our RefSeq set (see ChIP-Seq analysis section), we selected the feature with the best average p-value that was annotated to that gene. Genes with no annotated features were reported as NA (Table S3). Heatmaps were generated using the log ratio values according to the provided color scale.

Supplemental References

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

Borowiec, J.A., Zhang, L., Sasse-Dwight, S., and Gralla, J.D. (1987). DNA supercoiling promotes formation of a bent repression loop in lac DNA. J Mol Biol 196, 101-111.

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

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.

Chen, X., Vega, V.B., and Ng, H.H. (2008). Transcriptional regulatory networks in embryonic stem cells. Cold Spring Harb Symp Quant Biol 73, 203-209.
Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev 22, 746-755.

Guenther, M.G., Lawton, L.N., Rozovskaia, T., Frampton, G.M., Levine, S.S., Volkert, T.L., Croce, C.M., Nakamura, T., Canaani, E., and Young, R.A. (2008). Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. Genes Dev 22, 3403-3408.

Hay, D.C., Sutherland, L., Clark, J., and Burdon, T. (2004). Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem Cells 22, 225-235.

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

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res 12, 996-1006.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008).

Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell 134, 521-533.

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.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K., et al. (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell 124, 1283-1298.

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

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

Pereira, L., Yi, F., and Merrill, B.J. (2006). Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. Mol Cell Biol 26, 7479-7491.

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.