Mapping the Core Regulatory Circuitry of Embryonic Stem Cells

by

Sarah E. Johnstone
B.A., Biology
University of Chicago, 2001

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 2008

Signature of Author _____
Department of Biology
May 15, 2008

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

Accepted by _____
Dr. Stephen P. Bell
Professor of Biology and
Co-Chairperson, Biology Graduate Committee

Dedication

In honor of my father, Douglas Johnstone &
in memory of my mother, Carol Johnstone

Acknowledgements

I am indebted to many individuals for their support during my graduate training. Rick Young has been a tremendously supportive, encouraging and inspiring advisor. I am particularly grateful for his assistance as I learned how to work with genomic data. I have also derived much from him with regards to identifying important biological problems, designing hypothesis-driven research programs and interpreting complex data sets. I have also received a lot of helpful advice from my thesis committee, consisting of Rudolf Jaenisch and David Page. Numerous individuals in the Young lab have also been supportive of me during my thesis work. In particular, I would like to thank Duncan Odom, Elizabeth Jacobsen, Alla Sigova, Roshan Kumar, Lee Lawton and Michael Kagey who have been supportive mentors and good friends. I also owe much to Jamie Newman and Megan Cole for their years of collaboration and friendship. I have received critical technical support and advice from Jennifer Love and Sumeet Gupta. Many of us are grateful for the tremendous administrative support we receive while at MIT, and I am much obliged to Kercine Elie, Kathleen Blackett and Betsey Walsh for their help over the years. Laurie Boyer has been a tremendous support to me during my thesis work and I am grateful to her for planting the seeds that have inspired so many important experiments in the Young lab. Many other friends at MIT have greatly enriched my intellectual experience and been inspiring, especially Amy McCreath, Patricia Weinman, Kayvan Zainabadi, Luke Thompson, Sally Kwok, Hansen Bow and Conor Walsh. Roshan Kumar and his wife Radha have been wonderful friends and a great support to me. Kanchan Mirchandani has been a great friend to me over the years and has been a tremendous source of support and fun for me over the years. I am also deeply grateful to a number of special friends and family members who have been supportive to me during my graduate career despite living far away, especially Douglas Johnstone, Emily Johnstone, Jennifer Insley-Pruitt, Charles Taragin, Catherine Chung, Cynthia Hancock and Jennifer Sandy. Finally, I owe so much to Martin Aryee, who is my best friend and greatest support and inspiration. Much thanks to all of the wonderful people I list here.

Mapping the Core Regulatory Circuitry of Embryonic Stem Cells

by

Sarah E. Johnstone

Submitted to the Department of Biology on May 23, 2008
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy in Biology

Abstract

Embryonic stem (ES) cells are of tremendous biological interest because they have the capacity, termed pluripotency, to generate any cell type of the adult organism. Our lab is interested in understanding the genetic circuitry that governs pluripotency. For my thesis work I have contributed to a team effort to deduce the transcriptional regulatory circuitry of ES cells.

This collaborative effort first sought to define the genes that are regulated by the key pluripotency regulators, Oct4, Sox2 and Nanog. We then determined the genes targeted by the Polycomb Repressive Complex in ES cells. These datasets allowed us to define the core transcriptional regulatory circuitry for these cells and demonstrated that pluripotency is mediated through the repression of developmental regulators. Finally, an effort to understand how Wnt signaling modifies this circuitry led to the discovery that the Wnt signaling component Tcf3 is a core component of the transcriptional regulatory circuitry and serves to repress the pluripotency regulators, contributing to the balance between pluripotency and differentiation.

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

Table of Contents

Preface

Empedocles, then, was in error when he said that many of the characters presented by animals were merely the results of incidental occurrences during their development... In so saying he overlooked the fact that propagation implies a creative seed endowed with certain formative properties.

Aristotle
On the Parts of Animals, 350 BC

Elucidating the mechanisms that control differentiation will facilitate the efficient, directed differentiation of embryonic stem cells to specific cell types... Many diseases, such as Parkinson's disease and juvenile-onset diabetes mellitus, result from the death of dysfunction of just one or a few cell types. The replacement of those cells could offer lifelong treatment.

Thomson et al.
Science, 1998

In 350 BC Aristotle proposed that an embryo develops according to an ordered plan from a single *creative seed* (McKeon, 1973). The proposal that an embryo matures through a series of sequential steps is in accordance with a contemporary understanding of early embryonic development (Moody, 1999). The ordered steps of early mammalian development have been defined, as the zygote undergoes a series of highly coordinated cell divisions to progress through differentiation (Boiani and Scholer, 2005; Pera and Trounson, 2004; Ralston and Rossant, 2005; Spagnoli and Hemmati-Brivanlou, 2006). Embryonic stem (ES) cells derived from this *creative seed* can be grown in culture and facilitate the study of its properties. ES cells self-renew in culture and maintain the capacity, termed pluripotency, to generate into all cell types of the adult organism. The discovery of ES cells has permitted the earliest steps of development to be recapitulated

in vitro. Moreover, as suggested by Thomson and colleagues in their initial characterization of human ES cells, these cells may facilitate cell replacement therapies critical for the treatment of disease (Thomson et al., 1998). From the ancient proposal of a *creative seed* that can generate an adult organism to the isolation of human ES cells, the potential to develop cell replacement therapies has been a driving force behind the study of ES cell biology and will likely lead to new treatments for many human diseases.

A key challenge in understanding ES cell biology is deciphering how a single genome can direct the embryo to give rise to any adult cell type through the course of development. My interest in this challenge inspired me to join the Young lab, where I have worked to understand how regulation of gene expression impacts ES cell pluripotency. My thesis work has sought to characterize the transcriptional regulatory circuitry that governs ES cells. Shortly after I joined the lab, I became involved in a collaborative effort to unravel the genetic circuitry that defines ES cells. This effort involved using genome-wide location analysis to identify the complete set of target genes regulated by a number of factors critical for ES cell self-renewal and proliferation. These datasets enabled us to determine the core transcriptional regulatory circuitry that defines ES cell identity. During my third year in the lab, I became interested in understanding how signaling pathways regulate this circuitry. My efforts demonstrated that Tcf3, a component of the Wnt signaling pathway, directly regulates the core transcriptional regulatory circuitry. This thesis will describe the history and current understanding of ES cell biology that informed my research (Chapter 1), the contributions I have made to understand the regulatory circuitry of ES cells (Chapters 2-4) and remaining issues that must be addressed in order to realize the therapeutic potential of ES cells (Chapter 5).

Chapter 1

Introduction

This introduction will first provide background on the isolation and characterization of ES cells in order to contextualize my thesis work as part of a broader effort to understand and manipulate ES cells. Work over the last three decades has taken scientists from the initial isolation of ES cells to an advanced characterization of these cells. A proper framework for viewing this work is to consider how it has both been informed by and, reciprocally, informed our understanding of early mammalian development. Towards this end, this chapter will first review the critical steps of early embryonic development and consider the remarkable achievements of the late 20th century that established ES as a key model system for understanding early mammalian development. To do so, I will describe the discovery of both mouse and human ES cells and discuss the characteristics that make these cells so unique, including a discussion of their therapeutic potential.

I will then review the molecular characterization of ES cells that preceded and informed the work I have done to define the transcriptional regulatory circuitry of ES cells. Molecular and genetic tools have been applied to ES cell biology to elucidate a number of factors essential for ES cell identity. I will review the key transcription factors, chromatin regulators and signaling pathways that impact ES cells.

Finally, since this thesis will describe the transcriptional regulatory circuitry that governs ES cells, I will briefly describe how such circuitry can contribute to an understanding of cell state. An understanding of the transcriptional circuitry in ES cells will also enable a better understanding of the genetic mechanisms that govern pluripotency and methods to manipulate ES and somatic cells for therapeutic purposes.

Embryonic Stem Cell Isolation and Application

Embryonic stem cells are derived from the inner cell mass of the developing blastocyst and reflect its pluripotent state. The fertilized zygote is totipotent since it has the capacity to generate all tissues of the adult organism and extraembryonic tissues. However, as development proceeds the embryo's developmental potential is restricted. The coordinated differentiation events of early embryonic development include the formation of the trophectoderm, epiblast and primitive endoderm. Following fertilization of the oocyte, three rounds of cell division lead to the formation of the morula (Ralston and Rossant, 2007). Cells compact during the morula stage and the outer layer of the late morula forms an epithelium known as the trophectoderm, the first differentiation step in early embryonic development (Figure 1). The blastocoel, a cavity within this sphere of cells, forms, and the inner cell mass (ICM) becomes concentrated on one side of the blastocyst. The second differentiation step occurs when the ICM differentiates into epiblast and primitive endoderm, which happens in the late blastocyst. These three cell fate decisions are important because they represent the first restrictions on the totipotent embryo. Trophectoderm has the capacity to generate all of the non-maternal components of the placenta. The epiblast gives rise to the embryo, and the primitive endoderm gives rise to extraembryonic endoderm and the yolk sak. ES cells are isolated from the ICM (Kaufman and Evans, 1981; Martin, 1981). They possess the same potency as their source and have the capacity to give rise to the three primary germ layers, endoderm, ectoderm and mesoderm, as well as extraembryonic tissues (Spagnoli and Hemmati-Brivanlou, 2006).

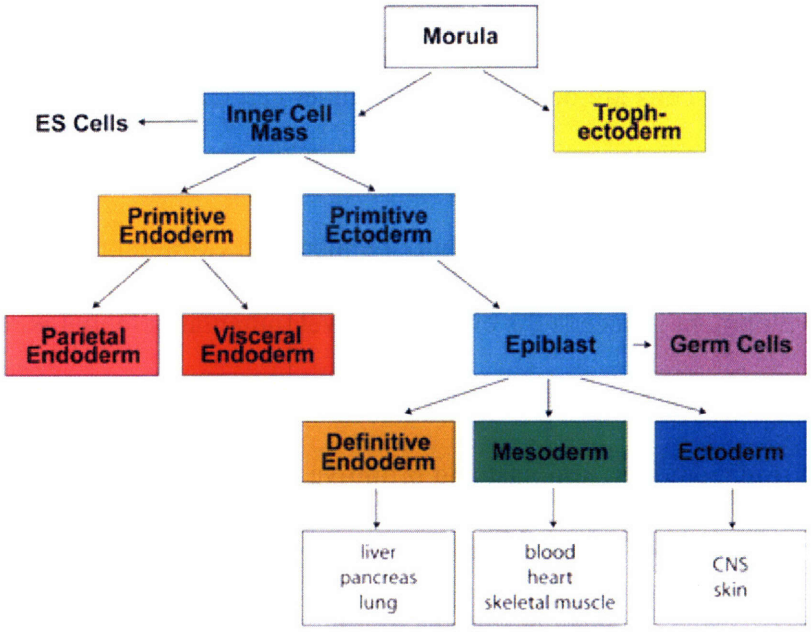


Figure 1. From Keller, 2005. Flowchart of early mouse development demonstrating the relationship between different tissue types.

Isolation of ES Cells

Two key observations in teratocarcinomas paved the way for the isolation of ES cells from the developing embryo (Damjanov, 2005; Friel et al., 2005; Smith, 2001). Teratocarcinomas are malignant tumors that contain a large population of undifferentiated cells, as well as differentiated cells of multiple lineages (Damjanov, 1990). First, it was reported that cells derived from teratocarcinomas, termed embryonic carcinoma (EC) cells, could be cultured *in vitro* and give rise to multiple lineages (Finch and Ephrussi, 1967; Friel et al., 2005). This provided evidence that teratocarcinomas contained a pluripotent stem cell. Then, in 1970, it was reported that grafting mouse

embryos onto an adult mouse led to teratocarcinoma formation (Solter et al., 1970; Stevens, 1970). The implication of this observation was the epiblast founder population of the early embryo was being maintained in the embryo-derived teratocarcinoma (Hogan et al., 1994). In the years following EC cell isolation, culture techniques for EC cells were improved (Smith, 2001). It was observed that EC cells proliferate best when in the presence of differentiated cells, and therefore co-culture with mitotically inactivated embryonic fibroblasts was introduced (Martin and Evans, 1975; Martin et al., 1977). It also became clear that EC cells are karyotypically abnormal, which was thought to be responsible for their capacity to self-renew indefinitely in culture (Friel et al., 2005; McBurney, 1976). The observation that EC cells represented a pluripotent mouse cell line and that such cells could be obtained from embryo-derived teratocarcinomas led to the effort to isolate self-renewing, pluripotent cells from the mouse embryo that were karyotypically normal.

In 1981, mouse ES cells were first isolated and shown to both self-renew and maintain pluripotency in culture (Evans and Kaufman, 1981; Martin, 1981). These cells were derived from cultures of preimplantation blastocysts and grown in medium conditioned by an established teratocarcinoma stem cell line (Martin, 1981) or on a co-culture with a mitotically inactivated mouse embryonic fibroblast (MEF) feeder population (Evans and Kaufman, 1981). At the time of their discovery, it was noted that these cells might contribute to the study of early mammalian development through the generation of embryo-derived cell lines with specific genetic alterations (Martin, 1981). It was also noted that these cells have a completely normal karyotype, unlike any previously described EC cell line (Evans and Kaufman, 1981). The mouse ES cells were

capable of forming teratocarcinomas when explanted onto an adult mouse, a phenotype similar to that observed in EC cells. They were also shown to be capable of differentiating into a wide variety of cell types, indicating that their pluripotency was similar to EC cells (Martin, 1981). The cells self-renewed indefinitely in culture when grown on mitotically inactivated embryonic fibroblasts, or feeder cells, which were thought to provide exogenous factors that maintained their capacity to proliferate in a pluripotent state (Keller, 2005). The remarkable identification of ES cells presented researchers with a tool to create transgenic models for mouse development and a challenge to identify a similar ES cell population from human embryos.

It was not until 1998 that scientists were first able to isolate human ES cells from the blastocyst, an achievement that has tremendous implications for therapeutic medicine and modeling human development since there are a multitude of differences between mouse and human ES cells (Thomson et al., 1998). An obvious challenge to the isolation of human ES cells was that the simple assay for pluripotency, which tests the capacity to contribute to germline and somatic chimerism, cannot be applied in humans (Smith, 2001). While there is now ample evidence that human ES cells can seed teratocarcinomas in mice and differentiate into multiple lineages *in vitro* (Amit et al., 2000; Itskovitz-Eldor et al., 2000; Reubinoff et al., 2000), it cannot be demonstrated that the cells can contribute to all cell types in a chimera. Additionally, there are known differences between the behavior of mouse and human ES cells (Pera and Trounson, 2004). Mouse ES cells grow in cohesive, rounded colonies while human ES cells flatten out and grow in less cohesive colonies (Friel et al., 2005). Also, for reasons that are not entirely clear, human ES cells grow more slowly and are more difficult to passage than

mouse ES cells. Interestingly, while mouse ES cell lines are incapable of differentiating into trophectoderm, human ES cell lines can (Boiani and Scholer, 2005; Friel et al., 2005). This observation, as well as the observation that the exogenous requirements for molecules like LIF and bFGF differ between the two species (Pera and Trounson, 2004), suggests that mouse ES cells may have limited utility for modeling human development. For this reason, great effort has been applied to improving the culture and manipulation of human ES cells (Andrews, 2006; Pera, 2001; Pera and Trounson, 2004; Pyle, 2006).

The isolation of ES cells from humans is subject to careful consideration and limitations on research due to the ethical implications of manipulating human embryos (Mendiola, 1999; Murry and Keller, 2008; Pera, 2001; Pera et al., 2001). A key criticism of human ES cell experimentation is that these cells are derived from human embryos. Some scientists have argued that these cells are not capable of growing into human embryos (Meyer, 2000; Pera, 2001). Nonetheless, the issue of the source for human ES cells has remained an ethical challenge for scientists. The principle source for human ES cells is surplus human embryos created by *in vitro* fertilization for fertility clinics (Meyer, 2000; Thomson et al., 1998). Currently, a large amount of effort is aimed at reprogramming somatic adult cells to become pluripotent cells that can be manipulated for therapeutic purposes with fewer ethical issues (Lewitzky and Yamanaka, 2007). While the functional qualities of ES cells make them ideal candidates for therapeutic development, these ethical issues will continue to be present throughout this effort.

Embryonic stem cells are defined by a number of unique qualities, one of which is their capacity for self-renewal. While their pluripotent counterparts, EC cells, are also capable of self-renewal in culture, these cells are also karyotypically abnormal (Evans

and Kaufman, 1981). In fact, most primary cell lines can only be passaged several times before they senesce (Smith, 2001). The ability for ES cells to proliferate in culture while retaining normal karyotypes has been demonstrated for both mouse and human ES cells for more than 100 passages (Keller, 2005). A key factor driving the self-renewal of embryonic stem cells is their maintained expression of the telomerase gene *TERT* (Armstrong et al., 2005; Hiyama, 2007). In fact, TERT overexpression in mouse ES cells leads to increased proliferation in culture (Armstrong et al., 2005), and *TERT* is downregulated and telomerase activity is reduced upon differentiation of human ES cells (Tzukerman, 2000).

A second defining feature of embryonic stem cells is their pluripotency, or capacity to generate all cell types of the adult organism. This aspect of ES cells is key to their therapeutic potential and utility as a model for development. As described previously, pluripotency was a quality first observed in EC cells isolated from teratocarcinomas (Finch and Ephrussi, 1967). The quintessential indicator of pluripotency is for the ES cells to contribute to all tissues in a chimeric animal. This was demonstrated for mouse ES cells shortly after their original isolation (Bradley et al., 1984). Mouse ES cells can contribute to all fetal lineages, yolk sac mesoderm, allanois and amnion, though they do not contribute to trophoblast (Beddington and Robertson, 1989; Bradley et al., 1984). Determining the potency of human ES cells is made more difficult by the impossibility of determining if cultured human ES cells can contribute to somatic and germ-line chimerism (Smith, 2001). Moreover, this observation would not get at the heart of the therapeutic value of these cells: their potential to be cultivated to give rise to specific cell lineages or tissue types for use in regenerative medicine (Murry

and Keller, 2008). Alternative assays for pluripotency, such as embryoid body formation, have confirmed the capacity of human ES cells to give rise to all cell types in the adult organism (Iskovitz-Eldor et al, 2000). Embryoid bodies are multicellular aggregates of ES cells that form in suspension culture and contain cells of various lineages (Pera, 2001). Embryoid bodies formed from human ES cells stain for markers of mesoderm, endoderm and ectoderm, indicating that human ES cells can give rise to all three germ cell layers.

Other qualities that have defined pluripotent cells, such as their origin in the ICM and the absence of X inactivation, are becoming less pertinent as it becomes possible to reprogram somatic cells into pluripotent cells that resemble ES cells (Smith, 2001; Takahashi and Yamanaka, 2006). In recent years somatic cells have been converted to pluripotent cells. This remarkable achievement has the potential to enable pluripotent cells derived from a somatic origin to be used for therapies rather than embryo-derived ES cells. In 2006 Takahashi and Yamanaka reported that four factors, Oct4, Sox2, c-Myc and Klf4, were capable of reprogramming mouse fibroblasts to adopt a pluripotent, ES cell-like identity. This landmark paper dramatically advanced the field of reprogramming of somatic cells, and a number of studies in the past two years have reported success with reprogramming differentiated cells to adopt an induced pluripotent state (iPS) (Aoi et al., 2008; Nakagawa et al., 2008; Park et al., 2008; Takahashi et al., 2007; Wernig et al., 2007; Yu et al., 2007). These important studies suggest that in the future human pluripotent stem cells may be able to be derived from sources other than human embryos. It also paves the way for personalized medical treatments, such that therapeutic cell types may be created for patients from their own sample tissue.

Applications of ES cells

Perhaps the most revolutionary application of ES cells to date has been the development of transgenic mice from genetically modified mouse ES cells. In fact, for contributing to the production of transgenic mice through the isolation of mouse ES cells, Martin Evans was awarded the Nobel Prize in Medicine in 2007 (Bradley et al., 1984). ES cells are capable of integrating into an embryo and producing viable chimeras (Smith, 2001). Moreover, ES cells have a diploid karyotype that is necessary for meiosis and so if they contribute to the germ line lineage in the chimera, they generate functional gametes (Smith, 2001). The ability to introduce alterations to the genome of mouse ES cells through homologous recombination and to create mice from these cells transformed developmental biology by enabling scientists to use whole animal genetics to understand gene function (Wobus and Boheler, 2005).

A second application of ES cells is their capacity to undergo directed differentiation in culture, which has enabled scientists to model specific events of early embryonic development through controlled differentiation experiments with ES cells. The default differentiation pathway of both mouse and human ES cells is believed to be neural (Evans and Kaufman, 1981). ES cells can be induced to differentiate *in vitro* through a number of strategies (Keller, 2005; Murry and Keller, 2008). Embryoid bodies can be formed from both mouse and human ES cells, which creates a heterogeneous aggregate of cell types (Doetschman et al., 1985; Evans and Kaufman, 1981; Itskovitz-Eldor et al., 2000). A second strategy involves growing the ES cells on a supportive layer of stromal cells that signal the ES cells to differentiate, which has been used to generate lymphohematopoietic cells from ES cells (Nakano et al., 1994). This strategy

can be used to induce highly specific differentiation programs by using stromal cells that express specific signaling molecules. For example, OP9 cells expressing the Notch ligand Delta-like 1 are capable of inducing ES cell differentiation into competent T cells (Schmitt et al., 2004). ES cells can also be stimulated to differentiate through the addition of exogenous factors such as extracellular matrix proteins or soluble signaling molecules (Cerdan et al., 2004; Laflamme et al., 2007; Nishikawa et al., 1998; Yao et al., 2006). Finally, genetic manipulation has been used to stimulate the differentiation of ES cells down specific lineages. Induction of Pax3 in mouse ES cell lines promotes myogenesis and the development of functional muscle progenitors (Darabi et al., 2008). The success of directing differentiation of ES cells *in vitro* has been encouraging for those who believe that ES cells hold great therapeutic value in regenerative medicine. Nonetheless, an improved knowledge of the regulatory circuitry that controls pluripotency and differentiation would better facilitate directed differentiation of ES cells.

While the above two applications of ES cells have already had considerable impact, these cells' capacity to revolutionize human medicine is yet to be fully developed for a number of reasons, including both technical and clinical obstacles. Several challenges make it uniquely difficult to manipulate human ES cells. First, these cells grow slowly and tend to spontaneously differentiate, making culture and manipulation of the cells difficult. Secondly, genetic modification of human ES cells is very difficult because of the inability to clone human ES cells (Trounson, 2006). Nonetheless, culture conditions for ES cells are improving and DNA has been introduced into the human ES cell genome through conventional transduction and lentiviral strategies (Eiges et al.,

2001; Gropp et al., 2003; Ma et al., 2003). Additionally, methods to introduce transgenes through homologous recombination have been demonstrated (Zwaka and Thomson, 2003). These advances will make the manipulation of human ES cells considerably easier. As studies in mouse and human ES cells begin to reveal factors that induce differentiation down specific lineages and unique markers for differentiated cells, manipulations of human ES cells towards particular lineages will become easier. Nonetheless, a number of clinical concerns will also need to be resolved before ES cell-based therapies will become feasible. These include risk of tumor formation, purity, insufficient graft size and immune rejection (Murry and Keller, 2008). Differentiated ES cells are unlikely to compose pure populations, and any remaining pluripotent cells could form teratocarcinomas upon transplant. This was a problem in the transplant of ES cell-derived skeletal muscle progenitors to dystrophic mice and was resolved by FACS isolation for cells expressing markers of skeletal muscle (Darabi et al., 2008). Issues of graft size and immune rejection have not yet been explored and will be significant challenges to the use of ES cell-derived therapeutic cells. The latter concern could be resolved by using reprogrammed somatic cells derived from the recipient patient as the source of pluripotent cells, though this has not yet been demonstrated.

Studies in mice have indicated that ES cells can be differentiated *in vitro* and reconstitute tissues or systems in the adult organism. As described, Darabi and colleagues demonstrated that skeletal muscle progenitors differentiated from mouse ES cells are capable of forming functional myofibers in dystrophic mice (Darabi et al., 2008). This study demonstrates the great potential that ES cells may play in treating muscular dystrophy. A second study recently demonstrated that pluripotent cells may

have significant therapeutic potential in treating sickle cell anemia (Hanna et al., 2007). Here, the authors converted induced pluripotent stem (iPS) cells to hematopoietic progenitors after correcting the human sickle cell hemoglobin by gene-specific targeting. Remarkably, a sickle cell anemia mouse model was rescued after transplantation with these hematopoietic progenitors. While this study did not use ES cells as the source for the differentiated cell type, they used a pluripotent cell type (iPS) that was developed to function like an ES cell. Recent work has also indicated that dopaminergic neurons can be generated from primate and, most recently, human ES cells (Cho et al., 2008; Takagi et al., 2005). These two studies used ES cells to generate neurospheres, which are composed of neural progenitors, and transplanted these cells into parkinsonian rat or monkey models to demonstrate that they can produce viable dopaminergic neurons in vivo. These initial studies that use pluripotent cells to produce therapeutically viable cells in vivo reinforces the hope that ES cells may have tremendous clinical significance for a number of diseases.

Molecular Characterization of ES Cells

Tremendous effort has been applied to define the molecular factors that govern ES cell identity. Characterization of genes expressed uniquely in ES cells has facilitated both an improved understanding of the regulation of ES cells and the manipulation of differentiated cell types to adopt a pluripotent state (Yamanaka, 2007). These factors include transcription factors, chromatin regulators and signaling pathway components. I will now review the large number of studies that have elucidated the roles of these

factors, indicating their critical contributions to ES cell pluripotency and self-renewal (Boiani and Scholer, 2005).

Transcription Factors

Perhaps the best characterized transcription factor associated with ES cell identity is the POU-domain transcription factor Oct4, which has been demonstrated to be a master regulator of ES cell state. This protein is expressed by all pluripotent cells during mouse embryogenesis and also by undifferentiated ES and EC cell lines of both mouse and human (Okamoto et al, 1990; Rosner et al., 1990; Scholer et al., 1989a; Scholer et al., 1989b). Oct4 is downregulated in differentiated tissues relative to pluripotent ES cells (Brandenberger et al., 2004; Yeom et al., 1996), suggesting it plays a critical role in pluripotency. In fact, Oct4 is essential for the formation of stem cells during embryogenesis (Nichols et al., 1998). Mice deficient for Oct4 develop until the blastocyst stage, but lack pluripotent cells in the ICM. In the absence of Oct4, ICM cells differentiate along the extraembryonic trophoblast lineage. A separate study demonstrated that depletion of Oct4 in mouse ES cells induces trophectoderm differentiation (Velkey and O'Shea, 2003). Its essential role has also been shown for human ES cells (Matin et al., 2004; Velkey et al., 2003). In the case of mouse ES cells, it has also been shown that upregulation of Oct4 can cause differentiation into primitive endoderm and mesoderm (Miwa et al., 2000). Despite the necessary contribution of Oct4 for pluripotency, the transcription factor is not sufficient for pluripotency since mouse ES cells that constitutively express Oct4 will differentiate in the absence of LIF (Niwa et al., 2000). Nonetheless, recent efforts to reprogram fibroblasts have shown that expression of Oct4 in differentiated cells can help drive the cell towards a pluripotent state

(Takahashi and Yamanaka, 2006). Taken together, these studies indicate that Oct4 is a pluripotency regulator since it can be characterized as a master regulator of the initiation and maintenance of pluripotency (Niwa, 2001).

The exact mechanism by which Oct4 regulates its target genes is not yet clear, though much is known about the proteins and DNA sequences with which it interacts. Oct4 encodes a homeodomain-containing protein that is capable of binding DNA at 'octamer' sequences found throughout the mammalian genome (Boiani and Scholer, 2005). Its DNA-binding domain, or POU domain, is conserved within the POU transcription factor family and consists of two structurally independent subdomains joined by a linker. Depending on adjacent flanking sequences and binding targets, Oct4 can act to either repress or activate target genes (Boiani and Scholer, 2005; Friel et al., 2005). Some of the earliest known binding targets for Oct4, such as *Fgf4*, *Utf-1*, *PDGFαR* and *Rex1*, include genes that show stem cell-specific expression (Ben-Shushan et al., 1998; Kraft et al., 1996; Nishimoto et al., 1999; Niwa, 2001; Yuan et al., 1995). Oct4 has both amino-terminal and carboxy-terminal transactivation domains, which are thought to share redundant functions such that one of them is sufficient to function with the POU domain to maintain self-renewal (Niwa, 2001; Okamoto et al., 1990; Rosner et al., 1990; Scholer et al., 1990a). A number of cofactors have been reported to associate with Oct4. One of the first characterized cofactors is the adenoviral protein E1A, which is thought to serve as a bridging factor between Oct4 and the basic transcriptional machinery (Scholer et al., 1991). A very important cofactor whose association with Oct4 is highly characterized is the Sry-related factor Sox-2, a protein that is highly expressed in the early embryo (Botquin et al., 1998).

Sox2 has been shown to cooperate with Oct4 to regulate target genes and contribute to pluripotency and is therefore characterized as a pluripotency regulator. Sox2 is an HMG-domain transcription factor that forms a complex with Oct4 (Boiani and Scholer, 2005). Expression of Sox proteins is highly specific for different cell types at different stages during animal development (Dailey and Basilico, 2001). While Sox2 is expressed in the early embryo, its expression is not restricted to pluripotent cells, and it can be detected in neural stem cells (Avilion et al., 2003). Mice lacking Sox2 have defective epiblast and extraembryonic ectoderm, though it has been suggested that earlier defects are avoided because of maternal stores of the protein. ES cells lacking Sox2 fail to self-renew. Numerous studies have shown that Octamer binding and Sox factors collaborate to potentiate transcriptional activation at specific subsets of genes (Ambrosetti et al., 1997; Ambrosetti et al., 2000; Dailey and Basilico, 2001). This co-occupancy was originally observed for the Octamer binding protein Oct2 (Zwilling et al., 1995). It was known that POU domain proteins collaborate with other factors to regulate transcription, and putative partners of Oct2 were identified through a mouse cDNA screen. HMG2 was isolated using this method and shown to interact with the homeodomain portion of the Oct2 POU domain through its HMG domain. This interaction was demonstrated for a number of HMG and homeodomain proteins, suggesting that the homeodomain of POU proteins preferentially interacts with HMG proteins. It was later shown that Oct4 and Sox2 form a regulatory complex at the promoters of target genes (Nishimoto et al., 1999; Remenyi et al., 2003; Yuan et al., 1995). Interestingly, Sox2 can complex with either Oct1 or Oct4, but only the Sox2/Oct4 complex is able to promote transcriptional activation. These studies indicate that Sox2 is

a necessary cofactor for Oct4 in the regulation of gene expression in ES cells and can be characterized as a pluripotency regulator.

A third transcription factor that plays a key role in ES cells and is classified as a pluripotency regulator is the homeodomain protein Nanog (Chambers et al., 2003; Mitsui et al., 2003). This factor was identified because its differential expression pattern was similar to Oct4, and it is upregulated in pluripotent cells and downregulated in differentiated tissues. Overexpression and gene disruption studies showed that Nanog is required for maintenance of pluripotency in the mouse epiblast and in ES cells. Overexpression of Nanog maintains mouse ES cells in their undifferentiated state independent of LIF. Disruption of Nanog in ES cells results in differentiation into extraembryonic endoderm. Nanog has a well conserved homeodomain, but the remainder of the protein has no known homology to other characterized proteins (Pan and Pei, 2003). Fusion of N- and C-terminal regions of Nanog to the Gal4 DNA binding domain revealed that each of these regions possesses a potent transactivation domain (Pan and Pei, 2003). Mass spectrometry and coimmunoprecipitations suggest that Nanog biochemically interacts with Oct4 (Wang et al., 2006; Zhang et al., 2007). It has recently been reported that Nanog expression fluctuates in ES cell populations, resulting in heterogeneous expression (Chambers et al., 2007; Singh et al., 2007). Moreover, ES cells can be maintained even upon disruption of Nanog, though the cells are predisposed towards differentiation (Chambers et al., 2007). This same group reported that germ cell formation is prevented in the absence of Nanog, suggesting that Nanog is dispensable for pluripotency in somatic cells but required for pluripotency in germ cells. It is possible that Nanog plays a critical role in balancing ES cells between pluripotency and

differentiation. However, given its critical role in maintaining pluripotency, Nanog is considered a pluripotency regulator.

It is clear that, in addition to these pluripotency regulators, there exist a number of other transcription factors that play key roles in pluripotency and self-renewal. For example, the transcription factor FoxD3 has been shown to be required for early embryonic development, and ES cells cannot be derived from mice deficient for this factor (Hanna et al., 2002). Zic3, a zinc finger transcription factor, is highly expressed in pluripotent cells and downregulated upon differentiation (Lim et al., 2007). Depletion of Zic3 leads to reduction of Nanog protein, and Zic3 is also directly regulated by the pluripotency regulators Oct4, Sox2 and Nanog. This suggests that Zic3 plays a critical role in regulating pluripotency. The homeodomain containing protein Pem is also important for maintaining pluripotency, and forced expression of Pem blocks differentiation of ES cells (Fan et al., 1999). Pem is expressed in the preimplantation embryo and undergoes lineage restriction upon implantation, similar to the pluripotency regulator Sox2. The zinc finger transcription factor Klf4 also promotes pluripotency of ES cells (Li et al., 2005; Nakatake et al., 2006). It is likely that in addition to FoxD3, Zic3, Pem and Klf4, other transcription factors help maintain the expression state that is important for pluripotency and self-renewal in ES cells. Recent studies reprogramming somatic cells towards a pluripotent state have demonstrated that in addition to Oct4, Sox2, Nanog and Klf4, c-myc and LIN28 are also capable of promoting pluripotency in somatic cells (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007). In addition to these putative pluripotency regulators, a number of transcription factors have been shown to be able to induce differentiation of ES cells towards a specific

lineage, including GATA6 and Cdx2 (Boiani and Scholer, 2005). Given the large set of transcription factors that can positively and negatively regulate ES cell identity, it is clear that there is a unique gene expression program that must be maintained for ES cells to remain pluripotent and be capable of self-renewal. Further characterization of the role of these additional factors in ES cells will help to clarify how these factors contribute to the gene expression program to influence pluripotency.

An important consideration when seeking to understand the role of the pluripotency regulators in initiating and maintaining ES cell identity is how these factors are themselves regulated. Early analyses of the 5' flanking sequence for the genes encoding Oct4, Sox2 and Nanog revealed a high degree of regulation by one another (Catena et al., 2004; Kuroda et al., 2005; Okumura-Nakanishi et al., 2005; Rodda et al., 2005). It has also been shown that p53 is able to bind and downregulate the *Nanog* gene (Lin et al., 2005) and that exposure to retinol (Vitamin A) leads to increased Nanog expression (Chen et al., 2007). The cancer-associated factor Tpt1 is capable of activating Oct4 and Nanog, and premature increases in levels of Tpt1 in oocytes leads to earlier activation of Oct4 (Koziol et al., 2007). PI3K signaling has been shown to be required for efficient self-renewal of ES cells and is known to maintain expression of Nanog (Storm et al., 2007). Interestingly, aggregation of ES cells has been shown to repress Nanog, which may be indicative of a mechanism used by the embryo to repress Nanog upon blastocyst growth (Hamazaki et al., 2004). A complete understanding of the mechanisms used to control expression of the pluripotency regulators will reveal how a pluripotent state is initiated and maintained, which will enhance our capacity to reprogram differentiated cells towards a pluripotent state.

In summary, a number of the transcription factors that regulate ES cell identity have been identified. Genetic and biochemical evidence indicates that these factors impact the ES cell transcriptome to either maintain pluripotency or direct differentiation. An important approach to characterizing the role of these factors in ES cells is to identify their complete set of downstream target genes. By clarifying which sets of genes are regulated by these factors, it will be possible to better understand how they modulate the ES expression program in order to govern ES cell identity. Additional components regulating transcription factor expression and ES cell identity, such as chromatin regulators and signaling pathways, will now be discussed.

Chromatin Regulators

When considering the mechanisms used to modulate transcriptional programs during development, it is necessary to consider the effect that chromatin structure has on gene expression. Chromatin consists of the complex of DNA and packaging histone proteins that allow the entire genome to compact and organize within the nucleus. Chromatin is organized into higher order structures to regulate transcription (Li et al., 2007; Misteli, 2007). A number of modifications of chromatin can occur in the mammalian genome, including phosphorylation, sumolation, ubiquitination, acetylation and methylation, and the exact location of these modifications relative to a gene influences the effect of the modification. These modifications are thought to modulate chromatin through their effect on higher order chromatin structure or by recruiting factors that bind to specific modifications on the chromatin template (Berger, 2007). In some cases these marks are known to mediate binding by specific factors. Methylated lysines are recognized by chromodomains (Bannister et al., 2001), and heterochromatin-like protein 1 (HP1)

associates with methylated histones through its chromodomain (Lachner et al., 2001). Alternatively, some modifications are simply associated with activation or repression. For example, histone acetylation is typically considered an activating mark while DNA methylation is often repressive. Nonetheless, the putative activating trimethylation of lysine 4 on histone H3 has recently been demonstrated to be associated with both silent and transcribed genes, revealing that a mark may not be associated with exclusively active or repressed genes (Guenther et al., 2007). Though our understanding of chromatin modifications is constantly evolving, it is clear that chromatin modifications affect gene expression. Therefore, these modifications have been explored as part of an effort to understand how the ES cell genome is regulated.

Epigenetic modifications are carefully regulated in the developing embryo and ES cells (Bibikova et al., 2008; Jaenisch and Bird, 2003). The term epigenetics has been applied to explain the chromatin-based heritable changes that do not reflect DNA sequence (Bird, 2002). Much effort has been applied to understanding the role of epigenetic gene regulation during early embryonic development, and has revealed that epigenetic modifications contribute to stable expression states in proliferating cells (Bantignies and Cavalli, 2006; Bird, 2002; Niwa, 2007; Orlando, 2003; Sasaki and Matsui, 2008). Epigenetic changes during development are precisely coordinated. Chromatin modifications are essential for pluripotency, and epigenetic marks on somatic cells must be reprogrammed to enable a pluripotent state. During embryonic development, genome-wide demethylation during cleavage is followed by genome-wide de novo methylation upon implantation (Jaenisch and Bird, 2003). This could explain why cloning by nuclear transfer frequently fails (Jaenisch and Bird, 2003). It is thought

that nuclear reprogramming of epigenetic marks takes place during germ cell development leading to the epigenetic state of a zygote as one with the capacity to generate into any cell type. This epigenetic state is maintained in ES cells in order to allow these cells to be pluripotent and self-renew. While the complete set of chromatin modifications that distinguishes a pluripotent cell from a differentiated somatic cell are unknown, the epigenetic signature of ES cells is beginning to be defined (Bernstein et al., 2007; Bibikova et al., 2006; Fulka et al., 2008; Nishino et al., 2006).

Polycomb group proteins (PcG) were originally identified for their role during Drosophila development in suppressing Hox gene expression (Orlando and Paro, 1995), and are also implicated in regulating pluripotent cells in the early embryo and ES cells (Faust et al., 1998; O'Carroll et al., 2001; Pasini et al., 2004). Polycomb proteins exist in one of two complexes (Bantignies and Cavalli, 2006). Polycomb Repressive Complex 2 (PRC2) is composed of Ezh2, Suz12 and Eed (Kuzmichev et al., 2004; Kuzmichev et al., 2005). Ezh2 is capable of trimethylation of lysine 27 of histone H3. Polycomb Repressive Complex 1 (PRC1) is composed of PC, PH, pSC and dRING. Embryos deficient for Eed, Ezh2, and Suz12 all display severe defects during gastrulation (Faust et al., 1998; O'Carroll et al., 2001; Pasini et al., 2004). ES cell lines cannot be established from Ezh2-deficient blastocysts, indicating that PRC2 is important for pluripotency and self-renewal (O'Carroll et al., 2001). Interestingly, Eed and Suz12 deficient ES cells can be derived and expanded in culture (Montgomery et al., 2005; Pasini et al., 2007). The PRC2 component Suz12 has also been shown to be important for differentiation (Pasini et al., 2007). Therefore, PRC2 components may have independent functions within ES cells that are distinct from the H3K27 trimethylation mark that this complex deposits on

the chromatin template. Nonetheless, each of the PRC2 components plays a key role in early embryogenesis and is thought to contribute to the unique epigenetic state that mediates pluripotency and self-renewal in ES cells. Given Polycomb proteins' critical role in repressing Hox genes during differentiation in Drosophila, it is possible that genes associated with differentiation are repressed by PRC2 in ES cells. In order to more fully understand the mechanisms that chromatin regulators use to regulate the ES cell transcriptional program, it is important to identify the downstream targets of these factors.

Chromatin modifications have been shown to regulate the expression of the pluripotency regulators (Hattori et al., 2004; Hattori et al., 2007). The promoter of the *Oct4* gene contains hypomethylated DNA and acetylated histones in undifferentiated ES cells (Hattori et al., 2004). Trophoblast stem (TS) cells do not normally express Oct4, and the *Oct4* promoter is hypermethylated in these cells. Treatment with an inhibitor of DNA methylation leads to upregulation of Oct4. An inhibitor of histone deacetylase also leads to upregulation of Oct4, indicating that both DNA methylation and hypoacetylation are responsible for repressing Oct4 in differentiated cell types. The *Nanog* promoter also displays DNA hypomethylation and histone acetylation in ES cells (Hattori et al., 2007). Like the promoter of Oct4, the Nanog promoter is heavily methylated in TS cells. The Nanog promoter is acetylated and displays H3K4 hypermethylation in ES cells, and all of these marks are lost in TS cells. TS cells display H3K9 and H3K27 hypermethylation at the Nanog promoter, a mark not observed on the Oct4 promoter in TS cells. Cumulatively, these results indicate that chromatin structure is an important and functional mechanism used by the cell to silence the pluripotency regulators upon

differentiation.

In summary, genetic and biochemical evidence indicates that chromatin modifications are important for ES cell viability and carefully controlled during development. Chromatin regulators and their downstream modifications govern the expression of pluripotency regulators in ES cells. Chromatin modifications are also likely to modify many additional genes to direct ES cell identity. An important approach to characterizing the role of chromatin regulators in ES cells is to identify their genomic targets. By clarifying which sets of genes are targeted by these regulators, it will be possible to better understand how they impact the ES cell expression program in order to control ES cell identity.

Signaling Pathways

A final aspect of molecular regulation of ES cells is mediated through external signaling pathways. External signals can promote ES cell pluripotency or cause these cells to differentiate. 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 (Li and Xie, 2005; Moore and Lemischka, 2006). For example, the IL-6 family cytokine leukemia inhibitory factor (LIF) that signals via gp130 and the LIF receptor (LIFR) is essential for the maintenance of mouse ES cells (Okita and Yamanaka, 2006; Smith et al., 1988; Williams et al., 1988). Other signal transduction pathways are also known to play essential roles in regulating the transcriptional response to external signals in ES cells. Recent studies have demonstrated that activation of the Wnt and Activin/Nodal signaling pathways promote pluripotency and self-renewal (Beattie et al., 2005; Sato et al., 2004; Valdimarsdottir and

Mummery, 2005), while Notch and BMP4 activation promote differentiation (Lowell et al., 2006; Ying et al., 2003). These signaling pathways, and others discussed below, influence the gene expression program of ES cells in order to affect ES cell identity. A comprehensive understanding of how signaling pathways regulate the genomic expression state and interact in combination will provide important clues to the means by which cell fate can be manipulated and facilitate new therapeutic approaches in regenerative medicine.

The first characterized signaling pathway associated with ES cells is the LIF/Stat3 signaling pathway, which is essential for mouse ES cell maintenance but dispensable for the maintenance of human ES cells. The signaling ligand, LIF, was demonstrated to be the molecule conferring the need to grow mouse ES cells on a mitotically inactivated mouse MEF feeder population (Smith et al., 1988; Williams et al., 1988). LIF signals through binding to a heterodimeric receptor consisting of LIFR and gp130 (Okita and Yamanaka, 2006). Binding promotes activation of the JAK kinase, which phosphorylates gp130 and LIFR causing recruitment of STAT3. STAT3 is then phosphorylated by JAK kinase, homodimerizes and translocates to the nucleus to act as a transcription factor. Constitutive activation of STAT3 replaces the need for LIF in culture in the presence of serum (Matsuda et al., 1999). STAT3 is known to regulate the transcription factor Myc, and constitutive Myc expression replaces the need for LIF (Cartwright et al., 2005). The connection between LIF/gp130 signaling and Myc represents an elegant example of external signals connecting directly to key transcription factors that regulate pluripotency and self-renewal. Nonetheless, the need for serum during culture of mouse ES cells in the presence of recombinant LIF implies that additional signaling molecules are

important for maintenance of pluripotency and self-renewal and directing gene expression.

Bone morphogenetic proteins (BMP) are known to influence the maintenance of ES cells in culture (Okita and Yamanaka, 2006). These proteins bind to type 1 and type 2 receptor tyrosine kinases and induce the heterodimerization of Smad transcription factors, which translocate to the nucleus to regulate target gene expression. BMP4 has been reported to cooperate with LIF in order to maintain mouse ES cells in culture (Ying et al., 2003). LIF and BMP4 are sufficient to maintain ES cells grown without a mitotically inactivated MEF feeder population and serum. BMP4 is known to signal through activation of the inhibitor of differentiation (Id) genes, and overexpression of Id proteins substitutes for BMP4 in culture. As in the case of LIF, the role of BMP4 is not conserved between mouse and human, and BMP4 induces mesodermal and ectodermal differentiation of human ES cells (Schuldiner et al., 2000). In contrast, inhibition of BMP4 signaling in combination with bFGF leads to self-renewal in the absence of serum or mitotically inactivated MEF feeder cells (Xu et al., 2005). BMP4, like LIF, has distinct functions in human and mouse ES cells and serves as an important link between the external environment and the transcriptional state of the ES cell.

The Wnt signaling pathway is a highly conserved pathway involved in development and has been demonstrated to play a key role in ES cell identity (Wang and Wynshaw-Boris, 2004). The canonical Wnt pathway activates ß-catenin, which translocates to the nucleus and, in conjunction with TCF/LEF transcription factors, activates gene expression (Giles, 2003). Both human and mouse embryonic stem cells are known to express components of this pathway (Walsh and Andrews, 2003; Wang et al., 2004), and

MEF feeder cells express Wnt ligands (Sato et al., 2004). Human embryonic stem cells treated with a drug that activates the canonical Wnt pathway remain undifferentiated in a feeder-free system (Sato et al., 2004). Mouse embryonic stem cells have an active Wnt pathway that is downregulated upon their differentiation (Sato et al., 2004) and cells with activating mutations in the Wnt pathway show reduced differentiation potential and remain undifferentiated longer than their wild type counterparts upon removal of self-renewal medium (Kielman et al., 2002). Nonetheless, while numerous studies have implicated Wnt signaling in the maintenance of ES cell identity (Hao et al., 2006; Miyabayashi et al., 2007; Ogawa et al., 2006; Singla et al., 2006; Takao et al., 2007), several studies have shown a role for Wnt signaling in differentiation (Lindsley et al., 2006; Otero et al., 2004). This dual function of Wnt signaling in ES cells may be due to crosstalk between Wnt signaling and other pathways that are active in ES cells under different conditions (Okita and Yamanaka, 2006).

The Activin/Nodal signaling pathway has also been shown to contribute to embryonic stem cell pluripotency (Valdimarsdottir and Mummery, 2005). Activin and Nodal are part of the TGFß superfamily, which includes ligands that play critical roles in a number of developmental processes. Activin/Nodal signaling through Smad2/3 activation and subsequent translocation to the nucleus is necessary to maintain expression of Oct4 and Nanog, two key markers of pluripotency, in human ES cells (James et al., 2005). Inhibition of this pathway induces differentiation and loss of Nanog and Oct4 expression. The role of Activin/Nodal in murine ES cells is less clear, but the pathway appears to be active as phosphorylated Smad2/3 is observed in the nucleus. In blastocyst outgrowths, Oct4 expression is lost upon treatment with an inhibitor of Activin/Nodal

signaling. Thus, it appears that this pathway contributes to pluripotency and self-renewal in both human and mouse ES cells. However, while the importance of Activin/Nodal signaling in ES cells has been demonstrated, it remains unclear how this signaling pathway impacts specific genes to create gene expression programs that promote pluripotency and self-renewal.

A number of other signaling pathways are also important for ES cell identity, including PI3K and ERK (Burdon et al., 1999; Jirmanova et al., 2002; Paling et al., 2004). Moreover, it is likely that these pathways function in combination with other pathways, posing a challenge to researchers who are best equipped to study them in isolation (Hayward et al., 2008; Hurlbut et al., 2007; Okita and Yamanaka, 2006; Sundaram, 2005). For example, it is known that LIF induces signaling pathways that are STAT3-independent. LIF activates the Ras/ERK signaling pathway through the effector protein SHP2 (Schiemann et al., 1997). Myc has not only been shown to be a target of the LIF signaling cascade, but it is also known to be degraded by GSK3β, indicating that Myc is regulated by both LIF and Wnt (Sears et al., 2000). Crosstalk between signaling pathways explains why a ligand may promote a particular state under one set of conditions and an alternative state under separate conditions. It will be important to develop reagents and culture conditions that permit pathways to be studied both in isolation and in combination. As ES cell culture conditions trend away from the use of serum and mitotically inactivated MEF feeders, it will become simpler to assay the effect of a particular signaling molecule (Adewumi et al., 2007; Bongso and Tan, 2005; Findikli et al., 2006; Loring and Rao, 2006; Mannello and Tonti, 2007).

An important challenge to understanding how signaling pathways impact ES cell identity is to identify the downstream transcriptional targets of these pathways. This effort will involve defining the complete set of terminal components for each pathway and the target genes that they regulate. This understanding will reveal the mechanisms through which these pathways directly impact the ES cell transcriptional program to modify cell state.

Transcriptional Regulatory Circuitry

Transcriptional regulatory circuitry describes the interactions between transcriptional regulatory components, such as transcription factors, chromatin regulators and terminal signaling components, and their target genes (Alon, 2007). This circuitry can respond to biological inputs to alter gene expression programs and thereby governs cell state. The goal of my thesis work has been to define the ES cell transcriptional regulatory circuitry and reveal how this circuitry controls ES cell identity. This effort has involved integrating the impact that transcription factors, chromatin regulators and signaling pathways have on the ES cell transcriptome using genomics approaches. The transcriptional regulatory circuitry that is produced from such an analysis is a model for how the cell maintains a stable cell state but allows for adaptive changes, such as differentiation, based on inputs from the environment (Figure 2).

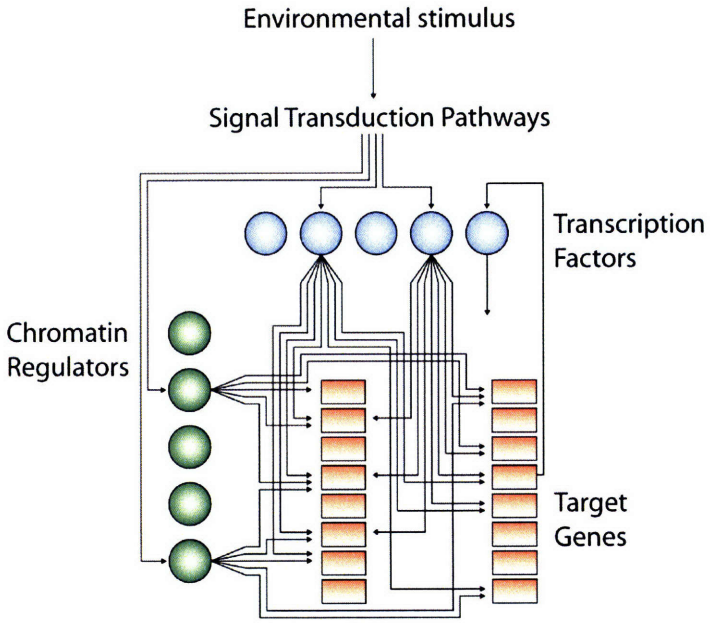


Figure 2. Model of transcriptional regulatory circuitry. Transcription factors, chromatin regulators and signal transduction pathways collaborate to regulate the expression of target genes and control cell state.

Thus, an understanding of the transcriptional regulatory circuitry will allow us to understand the gene expression changes that enable an organism to execute coordinated developmental events. Transcriptional regulatory circuitry provides the framework through which cells can specify over 200 lineages from an identical mammalian genome (Boiani and Scholer, 2005). These networks enable cells to generate complex organisms from a single totipotent zygote. As pluripotent cells differentiate, they restrict their potential. Yet, with the exception of plasma cells in the mammalian immune system, this change in potential is not mediated by genomic changes (Alt et al., 1992). Rather, stable expression states are defined by the transcriptional regulatory circuitry that governs the cell and is modified throughout development (Ferrell, 2002). Jacob and Monod first

suggested that gene regulatory circuits are wired in order to maintain changes in cell state during differentiation (Monod and Jacob, 1961). This assertion was precocious given the limited understanding of eukaryotic gene regulation at that point. Nonetheless, it is becoming clear that transcriptional networks are designed to create stable cell states that can respond to inputs and are key to understanding cellular identity (Alon, 2007).

Analysis of transcriptional regulatory circuitry for a number of organisms has revealed several network motifs that are conserved between species and play a functional role in the regulation of cell state (Alon, 2007). Autoregulation involves a gene regulating its own expression. This type of regulation can be positive or negative, with the former characterized by rapid response and minimal cell-cell variation and the latter displaying slow response time and enhanced cell-cell variation (Alon, 2007). Mathematical and experimental modeling of the regulatory motifs confirms these tendencies. Feedforward loops are an additional common regulatory motif that enable a cell to modulate response time to a stimulus (Mangan et al., 2003; Mangan and Alon, 2003; Mangan et al., 2006; Alon, 2007). Feedforward loops involve two transcription factors (X, Y) that regulate a shared gene target (Z) and X also regulates the gene encoding Y (Figure 3)(Mangan et al., 2003). These loops come in two varieties, coherent and incoherent, and can have positive or negative connections, such that there are 8 types of feedforward loops (Mangan et al., 2003). These different types are used by cells when response time to stimulation or loss of stimulation needs to be tightly controlled.

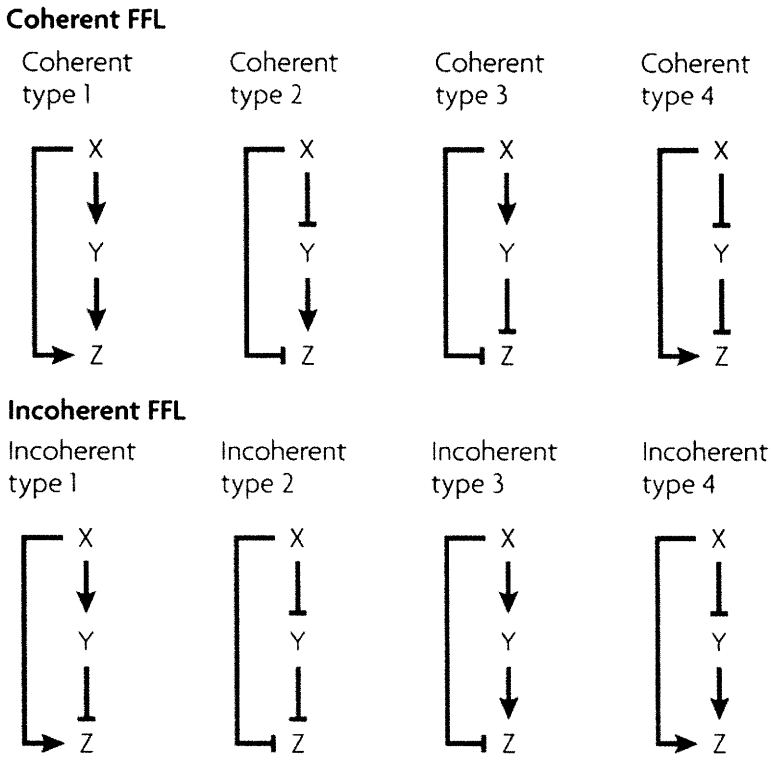


Figure 3. Adapted from Alon, 2007. Eight types of feedforward loops (FFL) exist. Coherent FFLs are such that the outcome of X signaling through Y to modify Z is identical to the direct impact of X on Z. Incoherent FFLs have conflicting outputs when X signals through Y compared to directly to Z.

Single-input modules are another type of regulatory mechanism used by cells. These modules involve a single transcription factor regulating a large set of genes with a shared function. This particular type of module is useful for inducing differentiation by temporal expression changes at target genes. Each of the above mechanisms permits a cell to respond to inputs with temporal and genetic specificity. Developmental regulatory circuitry can also involve additional motifs, such as dense overlapping regulons, feedback loops and transcription cascades, that allow for integrated and combinatorial signal response (Alon, 2007). These mechanisms reveal the complex genetic structures that

cells have evolved to be able to adapt to changes in cell state induced by environmental stimuli and differentiation. Defining the specific modules used by ES cells will reveal the mechanisms by which cell state is balanced between pluripotency and differentiation.

A clear understanding of the transcriptional regulatory circuitry that governs ES cell identity and pluripotency will also enable the manipulation of pluripotent cells for the purposes of regenerative medicine (Boiani and Scholer, 2005). By clearly defining the nature of the transcriptional circuitry that governs ES cells, it will be possible to perturb it in a targeted way to induce specific differentiation programs. Models for transcriptional regulatory circuitry provide for this in two ways. First, they reveal the regulatory motifs that ES cells use to balance between pluripotency and differentiation. For example, if a single-input module for Oct4 includes transcriptionally active genes encoding other pluripotency regulators, then removal of Oct4 may induce silencing of these regulators to permit differentiation. Alternatively, if Oct4 does not associate with the promoters of pluripotency regulators, then disrupting its expression is less likely to effect their expression. A second way the transcriptional regulatory circuitry is instructive for reprogramming is by identifying molecular targets. Identification of the transcriptional cascade that is downstream from Oct4 will reveal the set of genes whose regulation is key for pluripotency and self-renewal. This will generate a set of testable hypotheses to direct investigations of the role of Oct4 target genes in controlling ES cell identity. Modulating these genes may allow for the induction of distinct genetic programs.

It is clear that a mechanistic and molecular understanding of the transcriptional regulatory circuitry will afford scientists the tools necessary to manipulate pluripotent cells for the purposes of regenerative medicine. Such manipulations have been

accomplished in some instances and reveal that pluripotent cells have empirically demonstrated potential to be useful in human medicine for treatment of diseases such as Parkinson's disease, sickle cell anemia and muscular dystrophy (Cho et al., 2008; Darabi et al., 2008; Hanna et al., 2007; Redmond et al., 2007; Takagi et al., 2005). The tremendous efforts to understand ES cell identity that are described in this chapter indicate that we are fast approaching an era of regenerative medicine. The achievements that have brought us to this threshold support James Thomson's original aspiration for human ES cells—that they enable scientists to develop lifelong treatments for severe human disease.

References

Adewumi O, Aflatoonian B, Ahrlund-Richter L, Amit M, Andrews PW, Beighton G, Bello PA, Benvenisty N, Berry LS and Bevan S et al. (2007). "Characterization of human embryonic stem cell lines by the International Stem Cell Initiative." Nat Biotechnol **25**(7): 803-16.

Alon U (2007). "Network motifs: theory and experimental approaches." Nat Rev Genet **8**(6): 450-61.

Alt FW, Oltz EM, Young F, Gorman J, Taccioli G and Chen J (1992). "VDJ recombination." Immunol Today **13**(8): 306-14.

Ambrosetti DC, 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**(11): 6321-9.

Ambrosetti DC, Schöler HR, 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**(30): 23387-97.

Amit M, Carpenter MK, Inokuma MS, Chiu CP, Harris CP, Waknitz MA, Itskovitz-Eldor J and Thomson JA (2000). "Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture." Dev Biol **227**(2):271-8.

Aoi T, Yae K, Nakagawa M, Ichisaka T, Okita K, Takahashi K, Chiba T and Yamanaka S (2008). "Generation of Pluripotent Stem Cells from Adult Mouse Liver and Stomach Cells." Science

Armstrong L, Saretzki G, Peters H, Wappler I, Evans J, Hole N, von Zglinicki T and Lako M (2005). "Overexpression of telomerase confers growth advantage, stress resistance, and enhanced differentiation of ESCs toward the hematopoietic lineage." Stem Cells **23**: 516 – 529.

Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N and Lovell-Badge R (2003). "Multipotent cell lineages in early mouse development depend on SOX2 function." Genes Dev **17**(1): 126-40.

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC and Kouzarides T (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." Nature **410**(6824): 120-4.

Bantignies F and Cavalli G (2006). "Cellular memory and dynamic regulation of polycomb group proteins." Curr Opin Cell Biol **18**(3): 275-83.

Beattie GM, Lopez AD, Bucay N, Hinton A, Firpo MT, King CC and Hayek A (2005). "Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers." Stem Cells **23**(4): 489-95.

Beddington RS and EJ Robertson (1989). "An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo" Development **105**: 733-737.

Ben-Shushan E, Thompson JR, Gudas LJ and Bergman Y (1998). "Rex-1, a gene encoding a transcription factor expressed in the early embryo, is regulated via Oct-3/4 and Oct-6 binding to an octamer site and a novel protein, Rox-1, binding to an adjacent site." Mol Cell Biol **18**(4): 1866-78.

Berger SL (2007). "The complex language of chromatin regulation during transcription. Nature **447**(7143): 407-12.

Bernstein BE, Meissner A and Lander ES (2007). "The mammalian epigenome." Cell **128**(4): 669-81.

Bibikova M, Chudin E, Wu B, Zhou L, Garcia EW, Liu Y, Shin S, Plaia TW, Auerbach JM, Arking DE, Gonzalez R, Crook J, Davidson B, Schulz TC, Robins A, Khanna A, Sartipy P, Hyllner J, Vanguri P, Savant-Bhonsale S, Smith AK, Chakravarti A, Maitra A, Rao M, Barker DL, Loring JF and Fan JB (2006). "Human embryonic stem cells have a unique epigenetic signature." Genome Res **16**(9): 1075-83.

Bibikova M, Laurent L, Ren B, Loring JF and Fan J (2008). "Unraveling Epigenetic Regulation in Embryonic Stem Cells." Cell Stem Cell **2**: 123-134.

Bird A (2002). "DNA methylation patterns and epigenetic memory." Genes Dev **16**(1): 6-21.

Boiani M and Scholer H (2005). "Regulatory networks in embryo-derived pluripotent stem cells." Nat Rev Mol Cell Bio **6**: 872-884.

Bongso A and Tan S (2005). "Human blastocyst culture and derivation of embryonic stem cell lines." Stem Cell Rev **1**(2): 87-98.

Botquin V, Hess H, Fuhrmann G, Anastassiadis C, Gross MK, Vriend G and Schöler HR (1998). "New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2." Genes Dev **12**(13): 2073-90.

Bradley A, Evans M, Kaufman MH and Robertson E (1984). "Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines." Nature **309**(5965): 255-6.

Brandenberger R, Wei H, Zhang S, Lei S, Murage J, Fisk GJ, Li Y, Xu C, Fang R, Guegler K, Rao MS, Mandalam R, Lebkowski J and Stanton LW (2004). "Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation." Nat Biotechnol **22**(6): 707-16.

Burdon T, Stracey C, Chambers I, Nichols J and Smith A (1999). "Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells." Dev Biol **210**(1): 30-43.

Cartwright P, McLean C, Sheppard A, Rivett D, Jones K and Dalton S (2005). "LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism." Development **132**(5): 885-96.

Catena R, Tiveron C, Ronchi A, Porta S, Ferri A, Tatangelo L, Cavallaro M, Favaro R, Ottolenghi S, Reinbold R, Schöler H and Nicolis SK (2004). "Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells." J Biol Chem **279**(40): 41846-57.

Cerdan C, Rouleau A and Bhatia M (2004). "VEGF-A165 augments erythropoietic development from human embryonic stem cells." Blood **103**(7): 2504-12.

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**(5): 643-55.

Chen L, Yang M, Dawes J and Khillan JS (2007). "Suppression of ES cell differentiation by retinol (vitamin A) via the overexpression of Nanog." Differentiation **75**(8): 682-93.

Cho MS, Lee YE, Kim JY, Chung S, Cho YH, Kim DS, Kang SM, Lee H, Kim MH, Kim JH, Leem JW, Oh SK, Choi YM, Hwang DY, Chang JW and Kim DW (2008). "Highly efficient and large-scale generation of functional dopamine neurons from human embryonic stem cells." Proc Natl Acad Sci **105**(9): 3392-7.

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**(3): 315-28.

Damjanov I (1993). "Teratocarcinoma: neoplastic lessons about normal embryogenesis." Int J Dev Biol **37**: 39-46.

Damjanov E (2005). "The road from teratocarcinoma to human embryonic stem cells." Stem Cell Rev **1**: 273-6.

Darabi R, Gehlbach K, Bachoo RM, Kamath S, Osawa M, Kamm KE, Kyba M and Perlingeiro RC (2008). "Functional skeletal muscle regeneration from differentiating embryonic stem cells." Nat Med **14**(2): 134-43.

Doetschman TC, Eistetter H, Katz M, Schmidt W and Kemler R (1985). "The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium." J Embryol Exp Morphol **87**: 27-45.

Eiges R, Schuldiner M, Drukker M, Yanuka O, Itskovitz-Eldor J and Benvenisty N (2001). "Establishment of human embryonic stem cell- transfected clones carrying a marker for undifferentiated cells." Curr Biol **11**:514 –518.

Evans MJ and Kaufman MH (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-6.

Fan Y, Melhem MF and Chaillet JR (1999). "Forced expression of the homeobox-containing gene Pem blocks differentiation of embryonic stem cells." Dev Biol **210**(2): 481-96.

Faust C, Lawson KA, Schork NJ, 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**(22): 4495-506.

Ferrell JE (2002). "Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability." Curr Opin Cell Biol **14**(2): 140-8.

Finch BW and Ephrussi B (1967). Retention of multiple developmental potentialities by cells of a mouse testicular teratocarcinoma during prolonged culture in vitro and their extinction upon hybridization with cells of permanent lines. Proc. Nat. Acad. Sci. 57:615-621

Findikli N, Candan NZ and Kahraman S (2006). "Human embryonic stem cell culture: current limitations and novel strategies." Reprod Biomed Online **13**(4): 581-90.

Friel R, van der Sar S and Mee P (2005). "Embryonic stem cells: Understanding their history, cell biology and signaling." Adv Drug Delivery Rev **57**:1894-1903.

Fulka H, St John JC, Fulka J and Hozák P (2008). "Chromatin in early mammalian embryos: achieving the pluripotent state." Differentiation **76**(1): 3-14.

Giles RH, van Es JH and Clevers H (2003). "Caught up in a Wnt storm: Wnt signaling in cancer." Biochim Biophys Acta **1653**(1): 1-24.

Greely HT (2006). "Moving human embryonic stem cells from legislature to lab: remaining legal and ethical questions." PLoS Med **3**(5): e143.

Gropp M, Itsykson P, Singer O, Ben-Hur T, Reinhartz E, Galun E and Reubinoff BE (2003). "Stable genetic modification of human embryonic stem cells by lentiviral vectors." Mol Ther **7**(2): 281-7.

Guenther MG, Levine SS, Boyer LA, Jaenisch R and Young RA (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell **130**(1): 77-88.

Hamazaki T, Oka M, Yamanaka S and Terada N (2004). "Aggregation of embryonic stem cells induces Nanog repression and primitive endoderm differentiation." J Cell Sci **117**(23): 5681-6.

Hanna LA, Foreman RK, Tarasenko IA, Kessler DS and Labosky PA (2002). "Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo." Genes Dev **16**(20): 2650-61.

Hanna J, Wernig M, Markoulaki S, Sun CW, Meissner A, Cassady JP, Beard C, Brambrink T, Wu LC, Townes TM and Jaenisch R (2007). "Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin." Science **318**(5858): 1920-3.

Hao J, Li TG, Qi X, Zhao DF and Zhao GQ (2006). "WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells." Dev Biol **290**(1): 81-91.

Hattori N, Nishino K, Ko YG, Hattori N, Ohgane J, Tanaka S and Shiota K (2004). "Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells." J Biol Chem **279**(17): 17063-9.

Hattori N, Imao Y, Nishino K, Hattori N, Ohgane J, Yagi S, Tanaka S and Shiota K (2007). "Epigenetic regulation of Nanog gene in embryonic stem and trophoblast stem cells." Genes Cells **12**(3): 387-96.

Hayward P, Kalmar T and Arias AM (2008). "Wnt/Notch signalling and information processing during development." Development **135**(3): 411-24.

Hiyama E and Hiyama K (2007). "Telomere and telomerase in stem cells." Br J Cancer **96**(7): 1020-4

Hurlbut GD, Kankel MW, Lake RJ and Artavanis-Tsakonas S (2007). "Crossing paths with Notch in the hyper-network." Curr Opin Cell Biol **19**(2): 166-75.

Itskovitz-Eldor J, Schuldiner M, Karsenti D, Eden A, Yanuka O, Amit M, Soreq H and Benvenisty N (2000). "Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers." Mol Med **6**(2): 88-95.

Jaenisch R and Bird A (2003). "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." Nat Genet **33**: 245-54.

James D, Levine AJ, 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**(6): 1273-82.

Jirmanova L, Afanassieff M, Gobert-Gosse S, Markossian S and Savatier P (2002). "Differential contributions of ERK and PI3-kinase to the regulation of cyclin D1 expression and to the control of the G1/S transition in mouse embryonic stem cells." Oncogene **21**(36): 5515-28.

Keller, G (2005). "Embryonic stem cell differentiation: emergence of a new era in biology and medicine." Genes & Dev **19**:1129-1155.

Kielman MF, Rindapää M, Gaspar C, van Poppel N, Breukel C, van Leeuwen S, Taketo MM, 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**(4): 594-605.

Kouzarides T (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.

Koziol MJ, Garrett N and Gurdon JB (2007). "Tpt1 activates transcription of Oct4 and Nanog in transplanted somatic nuclei." Curr Biol **17**(9): 801-7.

Kraft HJ, Mosselman S, Smits HA, Hohenstein P, Piek E, Chen Q, Artzt K and van Zoelen EJ (1996). "Oct-4 regulates alternative platelet-derived growth factor alpha receptor gene promoter in human embryonal carcinoma cells." J Biol Chem **271**(22): 12873-8.

Kuroda T, Tada M, Kubota H, Kimura H, Hatano SY, Suemori H, Nakatsuji N and Tada T (2005). "Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression." Mol Cell Biol **25**(6): 2475-85.

Kuzmichev A, Jenuwein T, Tempst P and Reinberg D (2004). "Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3." Mol Cell **14**(2): 183-93.

Kuzmichev A, Margueron R, Vaquero A, Preissner TS, Scher M, Kirmizis A, Ouyang X, Brockdorff N, Abate-Shen C, Farnham P and Reinberg D (2005). "Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation." Proc Natl Acad Sci **102**(6): 1859-64.

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

Laflamme MA, Chen KY, Naumova AV, Muskheli V, Fugate JA, Dupras SK, Reinecke H, Xu C, Hassanipour M, Police S, O'Sullivan C, Collins L, Chen Y, Minami E, Gill EA, Ueno S, Yuan C, Gold J and Murry CE (2007). "Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts." Nat Biotechnol **25**(9): 1015-24.

Lewitzky M and Yamanaka S (2007). "Reprogramming somatic cells towards pluripotency by defined factors." Curr Opin Biotech **18**(5): 467-73

Li Y, McClintick J, Zhong L, Edenberg HJ, Yoder MC and Chan RJ (2005). "Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4." Blood **105**(2): 635-7.

Li L and Xie T (2005). "Stem cell niche: structure and function." Annu Rev Cell Dev Biol **21**: 605-31.

Li B, Carey M and Workman JL (2007). "The role of chromatin during transcription." Cell **128**(4): 707-19.

Lim LS, Loh YH, Zhang W, Li Y, Chen X, Wang Y, Bakre M, Ng HH and Stanton LW (2007). "Zic3 is required for maintenance of pluripotency in embryonic stem cells." Mol Biol Cell **18**(4): 1348-58.

Lin T, Chao C, Saito S, Mazur SJ, Murphy ME, Appella E and Xu Y (2005). "p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression." Nat Cell Biol **7**(2): 165-71.

Lindsley RC, Gill JG, Kyba M, Murphy TL and Murphy KM (2006). "Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm." Development **133**(19): 3787-96.

Loring, JF and Rao MS (2006). "Establishing standards for the characterization of human embryonic stem cell lines. Stem Cells 24: 145-150.

Lowell S, Benchoua A, Heavey B and Smith AG (2006). "Notch promotes neural lineage entry by pluripotent embryonic stem cells." PLoS Biol **4**(5):e121.

Ma Y, Ramezani A, Lewis R, Hawley RG and Thomson JA (2003). "High-level sustained transgene expression in human embryonic stem cells using lentiviral vectors." Stem Cells **21**(1): 111-7.

Mangan S and Alon U (2003). "Structure and function of the feed-forward loop network motif." Proc Natl Acad Sci **100**(21): 11980-5

Mangan S, Zaslaver A and Alon U (2003). "The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks." J Mol Biol **334**(2): 197-204.

Mangan S, Itzkovitz S, Zaslaver A and Alon U (2006). "The incoherent feed-forward loop accelerates the response-time of the gal system of Escherichia coli." J Mol Biol **356**(5): 1073-81.

Mannello F and Tonti GA (2007). "Concise review: no breakthroughs for human mesenchymal and embryonic stem cell culture: conditioned medium, feeder layer, or feeder-free; medium with fetal calf serum, human serum, or enriched plasma; serum-free, serum replacement nonconditioned medium, or ad hoc formula? All that glitters is not gold!" Stem Cells **25**(7): 1603-9.

Martin GR and Evans MJ (1975). The formation of embryoid bodies in vitro by homogeneous embryonal carcinoma cell cultures derived from isolated single cells. In Teratomas and Differentiation, ed. MI Sherman, D Solter, pp 169-87. New York: Academic.

Martin GR, Wiley LM and Damjanov I (1977). The developoment of cystic embryoid bodies in vitro from clonal teratocarcinoma stem cells. Dev. Biol. 61:230-44.

Martin GR (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Nat. Acad. Sci. 78:7634-38.

Matin MM, Walsh JR, Gokhale PJ, Draper JS, Bahrami AR, Morton I, Moore HD and Andrews PW (2004). "Specific knockdown of Oct4 and beta2-microglobulin expression by RNA interference in human embryonic stem cells and embryonic carcinoma cells." Stem Cells **22**(5): 659-68.

Matsuda T, Nakamura T, Nakao K, Arai T, Katsuki M, Heike T and Yokota T (1999). "STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells." EMBO J **18**(15): 4261-9.

McBurney MW (1976). Clonal lines of teratocarcinoma cells in vitro: differentiation and cytogenetic characteristics. J. Cell Physiol. 89:441-55.

McKeon, R (1973). "Introduction to Aristotle." The University of Chicago Press, Chicago.

Mendiola MM, Peters T, Young EW and Zoloth-Dorfman L (1999). "Research with human embryonic stem cells: ethical considerations." Hastings Cent Rep **29**(2): 31-6.

Meyer JR (2000). "Human embryonic stem cells and respect for life." J of Med Ethics **26**: 166–170.

Misteli T (2007). "Beyond the sequence: cellular organization of genome function." Cell **128**(4): 787-800.

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**(5): 631-42.

Miyabayashi T, Teo JL, 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 **104**(13): 5668-73.

Monod J and Jacob F (1961). "Teleonomic mechanisms in cellular metabolism, growth, and differentiation." Cold Spring Harb Symp Quant Biol **26**: 389-401.

Montgomery ND, Yee D, Chen A, Kalantry S, Chamberlain SJ, Otte AP and Magnuson T (2005). "The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation." Curr Biol **15**(10): 942-7

Moody S (Ed.). (1999). Cell Lineage and Fate Determination. San Diego: Academic Press.

Moore KA, Lemischka IR (2006). "Stem cells and their niches." Science **311**(5769): 1880-5.

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

Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N and Yamanaka S (2008). "Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts." Nat Biotechnol **26**(1): 101-6.

Nakano T, Kodama H and Honjo T (1994). "Generation of lymphohematopoietic cells from embryonic stem cells in culture." Science **265**(5175): 1098-101.

Nakatake Y, Fukui N, Iwamatsu Y, Masui S, Takahashi K, Yagi R, Yagi K, Miyazaki J, Matoba R, Ko MS and Niwa H (2006). "Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells." Mol Cell Biol **26**(20): 7772-82.

Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Schöler H and Smith A (1998). "Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4." Cell **95**(3): 379-91.

Nishikawa SI, Nishikawa S, Hirashima M, Matsuyoshi N and Kodama H (1998). "Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+

cells at a diverging point of endothelial and hemopoietic lineages." Development **125**(9): 1747-57.

Nishimoto M, Fukushima A, Okuda A and Muramatsu M (1999). "The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2." Mol Cell Biol **19**(8): 5453-65.

Nishino K, Ohgane J, Suzuki M, Hattori N and Shiota K (2006). "Methylation in embryonic stem cells in vitro." Methods Mol Biol **329**: 421-45.

Niwa H, Miyazaki J and Smith AG (2000). "Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells." Nat Genet **24**(4): 372-6.

Niwa H (2001). "Molecular mechanism to maintain stem cell renewal of ES cells." Cell Struct Funct **26**: 137-148.

Niwa H (2007). "Open conformation chromatin and pluripotency." Genes Dev **21**(21): 2671-6.

O'Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA and Jenuwein T (2001). "The polycomb-group gene Ezh2 is required for early mouse development." Mol Cell Bio **21**(13): 4330-6.

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**(1): 159-66.

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**(3): 461-72.

Okita K and Yamanaka S (2006). "Intracellular signaling pathways regulating pluripotency of embryonic stem cells." Curr Stem Cell Res Ther **1**(1): 103-11.

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**(7): 5307-17.

Orlando V and Paro R (1995). "Chromatin multiprotein complexes involved in the maintenance of transcription patterns." Curr Opin Genet Dev **5**(2): 174-9.

Orlando V (2003). "Polycomb, epigenomes, and control of cell identity. Cell **112**(5): 599-606.

Otero JJ, Fu W, Kan L, Cuadra AE and Kessler JA (2004). "Beta-catenin signaling is required for neural differentiation of embryonic stem cells." Development **131**(15): 3545-57.

Paling NR, Wheadon H, Bone HK and Welham MJ (2004). "Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling." J Biol Chem **279**(46): 48063-70.

Pan GJ and Pei DQ (2003). "Identification of two distinct transactivation domains in the pluripotency sustaining factor nanog." Cell Res **13**(6): 499-502.

Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW and Daley GQ (2008). "Reprogramming of human somatic cells to pluripotency with defined factors." Nature **451**(7175): 141-6.

Pasini D, Bracken AP, Jensen MR, Lazzerini Denchi E and Helin K (2004). "Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity." EMBO J **23**(20): 4061-71.

Pasini D, Bracken AP, Hansen JB, Capillo M and Helin K (2007). "The polycomb group protein Suz12 is required for embryonic stem cell differentiation." Mol Cell Biol **27**(10): 3769-79.

Pera MF (2001). "Scientific considerations relating to the ethics of the use of human embryonic stem cells in research and medicine." Reprod Fertil Dev **13**(1): 23-9.

Pera MF (2001). "Human pluripotent stem cells: a progress report." Curr Opin Genet Dev **11**(5): 595-9.

Pera MF and Trounson AO (2004). "Human embryonic stem cells: prospects for development." Development **131**(22): 5515-25.

Pereira L, Yi F and Merrill BJ (2006). "Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal." Mol Cell Biol **26**(20): 7479-91.

Ralston A and Rossant J (2005). "Genetic regulation of stem cell origins in the mouse embryo." Clin Genet **68**(2): 106-12

Redmond DE Jr, Bjugstad KB, Teng YD, Ourednik V, Ourednik J, Wakeman DR, Parsons XH, Gonzalez R, Blanchard BC, Kim SU, Gu Z, Lipton SA, Markakis EA, Roth RH, Elsworth JD, Sladek JR Jr, Sidman RL and Snyder EY (2007). "Behavioral improvement in a primate Parkinson's model is associated with multiple homeostatic effects of human neural stem cells." Proc Natl Acad Sci **104**(29): 12175-80.

Reményi A, Lins K, Nissen LJ, Reinbold R, Schöler HR and Wilmanns M (2003). "Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers." Genes Dev **17**(16): 2048-59.

Reubinoff BE, Pera MF, Fong CY, Trounson A and Bongso A (2000). "Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro." Nat Biotech **18**(4): 399-404.

Robertson E, Bradley A, Kuehn M and Evans M (1986). "Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector." Nature **323**(6087): 445-8.

Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH and Robson P (2005). "Transcriptional regulation of nanog by OCT4 and SOX2." J Biol Chem **280**(26): 24731-7.

Rosner MH, Vigano MA, Ozato K, Timmons PM, Poirier F, Rigby PW and Staudt LM (1990). "A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo." Nature **345**(6277): 686-92.

Sasaki H and Matsui Y (2008). "Epigenetic events in mammalian germ-cell development: reprogramming and beyond." Nat Rev Genet **9**(2): 129-40.

Sato N, Meijer L, Skaltsounis L, Greengard P and Brivanlou AH (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**(1): 55-63.

Schiemann WP, Bartoe JL and Nathanson NM (1997). "Box 3-independent signaling mechanisms are involved in leukemia inhibitory factor receptor alpha- and gp130-mediated stimulation of mitogen-activated protein kinase. Evidence for participation of multiple signaling pathways which converge at Ras." J Biol Chem **272**(26): 16631-6.

Schmitt TM, de Pooter RF, Gronski MA, Cho SK, Ohashi PS and Zúñiga-Pflücker JC (2004). "Induction of T cell development and establishment of T cell competence from embryonic stem cells differentiated in vitro." Nat Immunol **5**(4): 410-7.

Schöler HR, Hatzopoulos AK, Balling R, Suzuki N and Gruss P (1989a). "A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor." EMBO J **8**(9): 2543-50.

Schöler HR, Balling R, Hatzopoulos AK, Suzuki N and Gruss P (1989b). "Octamer binding proteins confer transcriptional activity in early mouse embryogenesis." EMBO J **8**(9): 2551-7.

Schöler HR, Ciesiolka T and Gruss P (1991). "A nexus between Oct-4 and E1A: implications for gene regulation in embryonic stem cells." Cell **66**(2): 291-304.

Schuldiner M, Yanuka O, Itskovitz-Eldor J, Melton DA and Benvenisty N (2000). "Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells." Proc Natl Acad Sci **97**(21): 11307-12.

Sears R, Nuckolls F, Haura E, Taya Y, Tamai K and Nevins JR (2000). "Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability." Genes Dev **14**(19): 2501-14.

Singh AM, Hamazaki T, Hankowski KE and Terada N (2007). "A heterogeneous expression pattern for Nanog in embryonic stem cells." Stem Cells **25**(10): 2534-42.

Singla DK, Schneider DJ, LeWinter MM and Sobel BE (2006). "Wnt3a but not wnt11 supports self-renewal of embryonic stem cells." Biochem Biophys Res Commun **345**(2): 789-95.

Smith AG, Heath JK, Donaldson DD, Wong GG, Moreau J, Stahl M and Rogers D (1988). "Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides." Nature **336**(6200): 688-90.

Smith, A (2001). "Embryo-Derived Stem Cells: Of Mice and Men." Annu. Rev. Cell Dev Biol **17**: 435-62.

Solter D, Skreb N and Damjanov I (1970). "Extrauterine growth of mouse egg cylinders results in malignant teratoma." Nature **227**: 503-4.

Spagnoli FM and Hemmati-Brivanlou A (2006). "Guiding embryonic stem cells towards differentiation: lessons from molecular embryology." Curr Opin Genet Dev **16**(5): 469-75.

Stevens LC. 1970. The development of transplantable teratocarcinomas from intratesticular grafts of pre-and postimplantation mouse embryos. Dev. Biol. 21:364-82.

Storm MP, Bone HK, Beck CG, Bourillot PY, Schreiber V, Damiano T, Nelson A, Savatier P and Welham MJ (2007). "Regulation of Nanog expression by phosphoinositide 3-kinase-dependent signaling in murine embryonic stem cells." J Biol Chem **282**(9): 6265-73.

Sundaram MV (2005). "The love-hate relationship between Ras and Notch." Genes Dev **19**(16): 1825-39.

Takahashi K and Yamanaka S (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-76.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K and Yamanaka S (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-72.

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**(3): 699-705.

Takagi Y, Takahashi J, Saiki H, Morizane A, Hayashi T, Kishi Y, Fukuda H, Okamoto Y, Koyanagi M, Ideguchi M, Hayashi H, Imazato T, Kawasaki H, Suemori H, Omachi S, Iida H, Itoh N, Nakatsuji N, Sasai Y and Hashimoto N (2005). "Dopaminergic neurons generated from monkey embryonic stem cells function in a Parkinson primate model." J Clin Invest **115**(1): 102-9.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS and Jones JM (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-7.

Trounson A (2006). "The production and directed differentiation of human embryonic stem cells." Endocr Rev **27**(2): 208-19.

Tzukerman M, Shachaf C, Ravel Y, Braunstein I, Cohen-Barak O, Yalon-Hacohen M and Skorecki KL (2000). "Identification of a novel transcription factor binding element involved in the regulation by differentiation of the human telomerase (hTERT) promoter." Mol Biol Cell **11**(12): 4381-91.

Valdimarsdottir G and Mummery C (2005). "Functions of the TGFbeta superfamily in human embryonic stem cells." APMIS **113**(11-12): 773-89.

Velkey JM and O'Shea KS (2003). "Oct4 RNA interference induces trophectoderm differentiation in mouse embryonic stem cells." Genesis **37**(1): 18-24.

Walsh J and Andrews PW (2003). "Expression of Wnt and Notch pathway genes in a pluripotent human embryonal carcinoma cell line and embryonic stem cell." APMIS **111**(1): 197-210.

Wang J and Wynshaw-Boris A (2004). "The canonical Wnt pathway in early mammalian embryogenesis and stem cell maintenance/differentiation." Curr Opin Genet Dev **14**(5): 533-9.

Wang QT, Piotrowska K, Ciemerych MA, Milenkovic L, Scott MP, Davis RW and Zernicka-Goetz M (2004). "A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo." Dev Cell **6**: 133-144.

Wang J, Rao S, Chu J, Shen X, Levasseur DN, Theunissen TW and Orkin SH (2006). "A protein interaction network for pluripotency of embryonic stem cells." Nature **444**(7117): 364-8.

Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE and Jaenisch R (2007). "In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state." Nature **448**(7151): 318-24.

Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, Wagner EF, Metcalf D, Nicola NA and Gough NM (1988). "Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells." Nature **336**(6200): 684-7.

Wobus AM and Boheler KR (2005). "Embryonic stem cells: prospects for developmental biology and cell therapy." Physiol Rev **85**: 635-678.

Xu RH, Peck RM, Li DS, Feng X, Ludwig T and Thomson JA (2005). "Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells." Nat Methods **2**(3): 185-90.

Yamanaka S (2007). "Strategies and New Developments in the Generation of Patient-Specific Pluripotent Stem Cells." Cell Stem Cell **1**: 39-49.

Yao S, Chen S, Clark J, Hao E, Beattie GM, Hayek A and Ding S (2006). "Long-term self-renewal and directed differentiation of human embryonic stem cells in chemically defined conditions." Proc Natl Acad Sci **103**(18): 6907-12.

Yeom YI, Fuhrmann G, Ovitt CE, Brehm A, Ohbo K, Gross M, Hübner K and Schöler HR (1996). "Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells." Development **122**(3): 881-94.

Ying QL, 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**(3): 281-92.

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II and Thomson JA (2007). "Induced pluripotent stem cell lines derived from human somatic cells." Science **318**(5858): 1917-20.

Yuan H, Corbi N, Basilico C and Dailey L (1995). "Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3." Genes Dev **9**(21): 2635-45.

Zhang L, Rayner S, Katoku-Kikyo N, Romanova L and Kikyo N (2007). "Successful co-immunoprecipitation of Oct4 and Nanog using cross-linking." Biochem Biophys Res Commun **361**(3): 611-4.

Zwaka TP and Thomson JA (2003). "Homologous recombination in human embryonic stem cells." Nat Biotechnol **21**(3): 319-21.

Zwilling S, König H and Wirth T (1995). "High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors." EMBO J **14**(6): 1198-208.

Chapter 2

Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells

Published as: Boyer* LA, Lee* TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R and Young RA (2005). "Core transcriptional regulatory circuitry in human embryonic stem cells." Cell **122**(6): 947-56.
* These authors contributed equally to this work.

My contributions to this work

The effort to profile transcription factors in human ES cells was initiated in the summer of 2004, shortly after I joined the Young lab. I was involved in this collaborative project from the beginning, participating in the many discussions during which we evaluated the best factors and cell types to use for these genome-wide experiments. Laurie Boyer's efforts to map binding events for the transcription factors Oct4, Sox2 and Nanog in ES cells were chosen for these genome-wide experiments because of their success experimentally and their value to the ES cell community. My contributions to this project span from the experimental effort to the analysis and writing of the results.

My initial efforts were to help improve the technical aspects of the lab's ChIP-chip protocol to adapt it for the purpose of genome-wide profiling of mammalian cells. Towards this end, I helped to scale the experimental method to enable genome-wide profiling of a key set of transcription factors in ES cells. I was a critical player in the effort to obtain binding data for the three factors described in this chapter, from the initial effort to amplify the material to hybridizing it to a set of microarrays representing approximately 18,000 annotated human promoters.

Upon obtaining high-quality genome-wide binding data for Oct4, Sox2 and Nanog, I was a key participant in analyzing these data. I was principally responsible for inspecting binding plots to determine the parameters and quality of binding events for the different ChIP samples, as well as their overlap. I also was responsible for generating many of the figures in the paper. Since much of the writing advanced simultaneously with the analysis, I was very involved with Laurie Boyer during the organizing and

editing stages of the manuscript.

Summary

The transcription factors Oct4, Sox2, and Nanog have essential roles in early development and are required for the propagation of undifferentiated embryonic stem (ES) cells in culture. To gain insights into transcriptional regulation of human ES cells, we have identified Oct4, Sox2, and Nanog target genes using genome-scale location analysis. We found, surprisingly, that Oct4, Sox2, and Nanog co-occupy a substantial portion of their target genes. These target genes frequently encode transcription factors, many of which are developmentally important homeodomain proteins. Our data also indicates that Oct4, Sox2, and Nanog collaborate to form regulatory circuitry in ES cells consisting of autoregulatory and feedforward loops. These results provide new insights into the transcriptional regulation of stem cells and reveal how Oct4, Sox2, and Nanog contribute to pluripotency and self-renewal.

Introduction

Mammalian development requires the specification of over 200 unique cell types from a single totipotent cell. Embryonic stem (ES) cells are derived from the inner cell mass (ICM) 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. As such, the recent derivation of human ES cells provides a unique opportunity to study early development and is thought to hold great promise for regenerative medicine (Pera and Trounson, 2004; Reubinoff et al., 2000; Thomson et al., 1998). An understanding of the transcriptional regulatory circuitry that is responsible for pluripotency and self-renewal in human ES cells is fundamental to understanding human development and realizing the therapeutic potential of these cells.

Homeodomain transcription factors are evolutionarily conserved and play key roles in cell fate specification in many organisms (Hombria and Lovegrove, 2003). Two such factors, Oct4 and Nanog, are essential regulators of early development and ES cell identity (Chambers et al., 2003; Hay et al., 2004; Matin et al., 2004; Mitsui et al., 2003; Nichols et al., 1998; Zaehres et al., 2005). Several genetic studies in mouse suggest that these regulators have distinct roles, but may function in related pathways to maintain the developmental potential of these cells (Chambers, 2004). For example, disruption of OCT4 and NANOG results in the inappropriate differentiation of ICM and ES cells to trophectoderm and extra-embryonic endoderm, respectively (Chambers et al., 2003; Mitsui et al., 2003; Nichols et al., 1998). However, functional over-expression of Oct4 in ES cells leads to a phenotype that is similar to loss of Nanog function (Chambers et al.,

2003; Mitsui et al., 2003; Nichols et al., 1998; Niwa et al., 2000). Thus, knowledge of the set of genes regulated by these two transcription factors might reveal why manipulation of Oct4 and Nanog results in these phenotypic consequences.

Oct4 is known to interact with other transcription factors to activate and repress gene expression in mouse ES cells (Pesce and Schöler, 2001). For example, Oct4, a member of the POU (Pit/Oct/Unc) class of homeodomain proteins, can heterodimerize with the HMG-box transcription factor, Sox2, to affect the expression of several genes in mouse ES cells (Botquin et al., 1998; Nishimoto et al., 1999; Yuan et al., 1995). The cooperative interaction of POU homeodomain and HMG factors is thought to be a fundamental mechanism for the developmental control of gene expression (Dailey and Basilico, 2001). The extent to which ES cell gene regulation is accomplished by Oct4 through an Oct4/Sox2 complex and whether Nanog has a role in this process is unknown.

Oct4, Sox2, and Nanog are thought to be central to the transcriptional regulatory hierarchy that specifies ES cell identity because of their unique expression patterns and their essential roles during early development (Avilion et al., 2003; Chambers et al., 2003; Hart et al., 2004; Lee et al., 2004; Mitsui et al., 2003; Nichols et al., 1998; Schöler et al., 1990). Studies in a broad range of eukaryotes have shown that transcriptional regulators that have key roles in cellular processes frequently regulate other regulators associated with that process (Guenther et al., 2005; Lee et al., 2002; Odom et al., 2004). It is likely that the key stem cell regulators bind and regulate genes encoding other transcriptional regulators, which in turn determine the developmental potential of these cells, but we currently lack substantial knowledge of the regulatory circuitry of ES cells and other vertebrate cells.

To further our understanding of the means by which Oct4, Sox2, and Nanog control the pluripotency and self-renewal of human ES cells, we have used genome-scale location analysis (chromatin immunoprecipitation coupled with DNA microarrays) to identify the target genes of all three regulators *in vivo*. The results reveal that Oct4, Sox2, and Nanog co-occupy the promoters of a large population of genes, that many of these target genes encode developmentally important homeodomain transcription factors, and that these regulators contribute to specialized regulatory circuits in ES cells.

Results and Discussion

Oct4 promoter occupancy in human ES cells

DNA sequences occupied by Oct4 in human H9 ES cells (NIH code WA09; Appendix A) were identified in a replicate set of experiments using chromatin-immunoprecipitation (ChIP) combined with DNA microarrays (Figure 1A and Appendix A). For this purpose, DNA microarrays were designed that contain 60-mer oligonucleotide probes covering the region from –8kb to +2kb relative to the transcript start sites for 17,917 annotated human genes. Although some transcription factors are known to regulate genes from distances greater than 8kb, 98% of known binding sites for human transcription factors occur within 8kb of target genes (Appendix A Figure S1). The sites occupied by Oct4 were identified as peaks of ChIP-enriched DNA that span closely neighboring probes (Figure 1B). Oct4 was associated with 623 (3%) of the promoter regions for known protein-coding genes and 5 (3%) of the promoters for known miRNA genes in human ES cells (Appendix A Table S2).

Figure 1

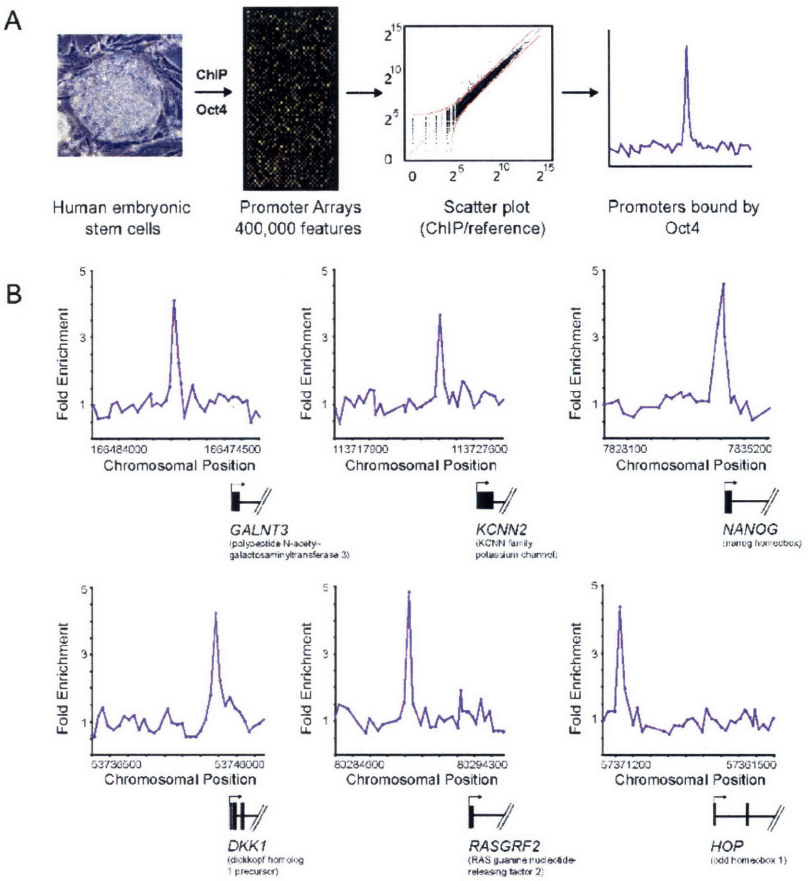

A
ChIP
Oct4
Human embryonic stem cells
Promoter Arrays 400,000 features
Scatter plot (ChIP/reference)
Promoters bound by Oct4
B
Fold Enrichment
Chromosomal Position
GALNT3
KCNN2
NANOG
DKK1
RASGRF2
HOP

Figure 1. Genome-wide ChIP-Chip in human embryonic stem cells.

A. DNA segments bound by transcriptional regulators were identified using chromatin-immunoprecipitation (ChIP) and identified with DNA microarrays containing 60-mer oligonucleotide probes covering the region from –8kb to +2kb for 17,917 annotated transcription start sites for human genes. ES cell growth and quality control, ChIP protocol, DNA microarray probe design and data analysis methods are described in detail in Experimental Procedures and Appendix A.

B. Examples of Oct4 bound regions. Plots display unprocessed ChIP-enrichment ratios for all probes within a genomic region. Genes are shown to scale below plots (exons and introns are represented by thick vertical and horizontal lines, respectively), and the genomic region represented is indicated beneath the plot. The transcription start site and transcript direction are denoted by arrows.

Two lines of evidence suggested that this protein-DNA interaction dataset is of high quality. First, the genes occupied by Oct4 in our analysis included many previously identified or supposed target genes in mouse ES cells or genes whose transcripts are highly enriched in ES cells, including POU5F1/OCT4, SOX2, NANOG, LEFTY2/EBAF, CDX2, HAND1, DPPA4, GJA1/CONNEXIN43, FOXO1A, TDGF1 and ZIC3 (Abeyta et al., 2004; Brandenberger et al., 2004; Catena et al., 2004; Kuroda et al., 2005; Niwa, 2001; Okumura-Nakanishi et al., 2005; Rodda et al., 2005; Sato et al., 2003; Wei et al., 2005) (Appendix A Table S2). Second, we have used improved protocols and DNA microarray technology in these experiments (Appendix A) that should reduce false positive rates relative to those obtained in previous genome-scale experiments (Odom et al., 2004). By using this new technology with yeast transcription factors, where considerable prior knowledge of transcription factor binding sites has been established, we estimated that this platform has a false positive rate of <1% and a false negative rate of 20% (Appendix A).

Oct4, Sox2, and Nanog co-occupy many target genes

We next identified protein-coding and miRNA genes targeted by the stem cell regulators Sox2 and Nanog using location analysis. Sox2 and Nanog were found associated with 1271 (7%) and 1687 (9%), respectively, of the promoter regions for known protein-coding genes in human ES cells (Appendix A Table S2 - S4). It was immediately evident that many of the target genes were shared by Oct4, Sox2, and Nanog (Figure 2A). Examples of protein-coding genes that are co-occupied by the three regulators are shown in Figure 2B (Appendix A Table S5). Control experiments showed

that the set of promoters bound by the cell cycle transcription factor E2F4 in these human ES cells did not overlap substantially with those bound by the three stem cell regulators (Appendix A Table S2 and S6). We found that Oct4, Sox2 and Nanog together occupy at least 353 genes in human ES cells.

Previous studies have shown that Sox2 and Oct4 can interact cooperatively to synergistically activate transcription of target genes in murine ES cells and that this activity is dependent upon the juxtaposition of Oct4 and Sox2 binding sites (Ambrosetti et al., 1997; Remenyi et al., 2004). Our results revealed that approximately half of the genes occupied by Oct4 were also bound by Sox2 in human ES cells (Figure 2A; Appendix A Table S2). It was surprising, however, to find that >90% of promoter regions bound by both Oct4 and Sox2 were also occupied by Nanog. Furthermore, we found that Oct4, Sox2, and Nanog binding sites occured in close proximity at nearly all of the genes that they co-occupied (Figure 2C). These data suggest that Oct4, Sox2 and Nanog function together to regulate a significant proportion of their target genes in human ES cells.

Figure 2

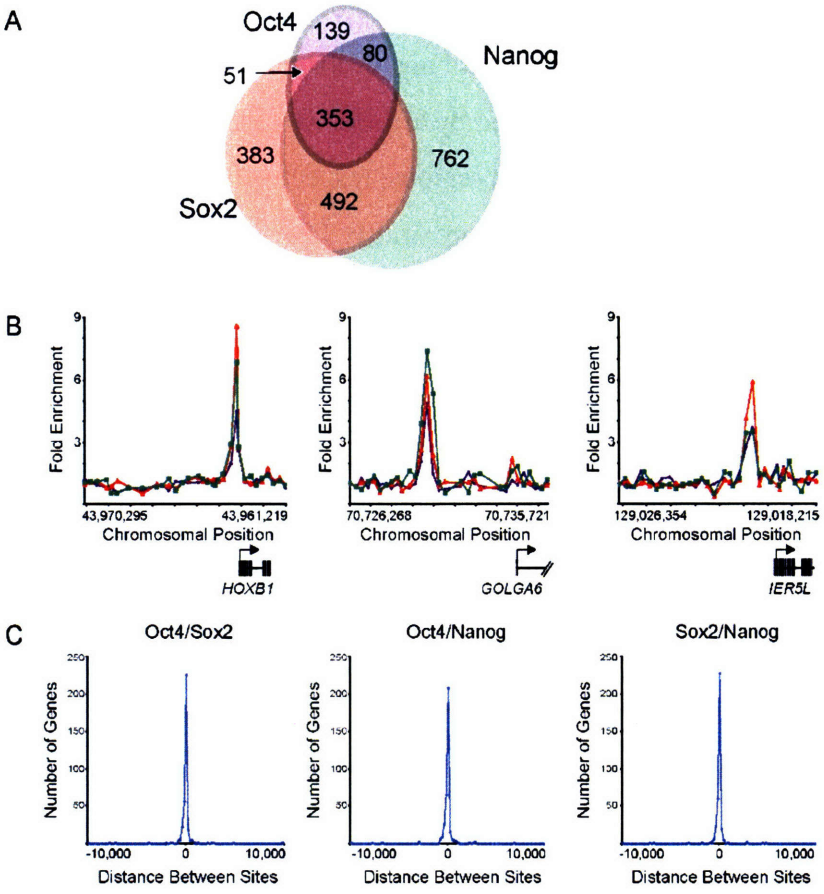

A
Oct4
139
80
Nanog
51
353
383
762
Sox2
492
B
Fold Enrichment
9
6
3
43,970,295
43,961,219
Chromosomal Position
HOXB1
70,726,268
70,735,721
GOLGA6
129,026,354
129,018,215
IER5L
C
Oct4/Sox2
Oct4/Nanog
Sox2/Nanog
Number of Genes
250
200
150
100
50
-10,000
0
10,000
Distance Between Sites

Figure 2. Oct4, Sox2 and Nanog target genes in human ES cells.

A. Venn diagram representing the overlap of Oct4, Sox2, and Nanog promoter bound regions.

B. Representative examples of protein-coding genes co-occupied by Oct4, Sox2, and Nanog. Plots display unprocessed ChIP enrichment ratios for all probes within a genomic region. Genes are shown to scale relative to their chromosomal position. Exons and introns are represented by thick vertical and horizontal lines, respectively. The start and direction of transcription are denoted by arrows. Green, red, and purple lines represent Nanog, Sox2, and Oct4 bound regions, respectively.

C. Oct4, Sox2, and Nanog bind in close proximity. The distances between the midpoint of bound regions for pairs of transcription factors was calculated for the 353 regions bound by all three transcription factors. Negative and positive values indicate whether the first factor is upstream or downstream of the second factor in relation to the gene. The frequency of different distances between the bound regions is plotted as a histogram.

A class of small non-coding RNAs known as microRNAs (miRNA) play vital roles in gene regulation and recent studies indicate that more than a third of mammalian protein-coding genes are conserved miRNA targets (Bartel, 2004; Lewis et al., 2005). ES cells lacking the machinery that processes miRNA transcripts are unable to differentiate (Kanellopoulou et al., 2005). Moreover, recent evidence indicates that microRNAs play an important role in organismal development through regulation of gene expression (Pasquinelli et al., 2005). Oct4, Sox2, and Nanog were found associated with 14 miRNA genes and co-occupied the promoters of at least two miRNA genes, mir-137 and mir-301 (Table 1). Our results suggest that miRNA genes which have been implicated in developmental processes are likely regulated by Oct4, Sox2 and Nanog in human ES cells and are important components of the transcriptional regulatory circuitry in these cells.

Table 1. miRNA Loci Near Oct4, Sox2 and Nanog Bound Regions

	Transcription Factor		
miRNA	**Oct4**	**Sox2**	**Nanog**
mir-7-1		+	
mir-10a	+		
mir-22		+	+
mir-32		+	+
mir-128a			+
mir-135b		+	+
mir-137	+	+	+
mir-196a-1			+
mir-196b	+		
mir-204		+	+
mir-205		+	+
mir-301	+	+	+
mir-361			+
mir-448	+		

Proximal binding of Oct4, Sox2 and Nanog to miRNAs from the RFAM database.
Transcription factors bound are indicated by a "+".

ES cell transcription factors occupy active and inactive genes

Oct4 and Sox2 are known to be involved in both gene activation and repression *in vivo* (Botquin et al., 1998; Nishimoto et al., 1999; Yuan et al., 1995), so we sought to identify the transcriptional state of genes occupied by the stem cell regulators. To this end, the set of genes bound by Oct4, Sox2, and Nanog were compared to gene expression datasets generated from multiple ES cell lines (Abeyta et al., 2004; Brandenberger et al., 2004; Sato et al., 2003; Wei et al., 2005) to identify transcriptionally active and inactive genes (Appendix A Table S2). The results showed that one or more of the stem cell transcription factors occupied 1303 actively transcribed genes and 957 inactive genes.

The importance of Oct4, Sox2, and Nanog for early development and ES cell identity led us to focus additional analyses on the set of 353 genes that are co-occupied by these regulators in human ES cells (Table S5). We first identified transcriptionally active genes. Transcripts were consistently detected in ES cells for approximately half of the genes co-bound by Oct4, Sox2, and Nanog. Among these active genes, several encoding transcription factors (e.g. POU5F1/OCT4, SOX2, NANOG, STAT3, ZIC3) and components of the Tgf-β (e.g. TDGF1, LEFTY2/EBAF) and Wnt (e.g. DKK1, FRAT2) signaling pathways were notable targets. Recent studies have shown that Tgf-β and Wnt signaling play a role in pluripotency and self-renewal in both mouse and human ES cells (James et al., 2005; Sato et al., 2004). These observations suggest that Oct4, Sox2, and Nanog promote pluripotency and self-renewal through positive regulation of their own genes and genes encoding components of these key signaling pathways.

Among transcriptionally inactive genes co-occupied by Oct4, Sox2, and Nanog, we noted a striking enrichment for transcription factor genes ($p<10^{-18}$; Appendix A Table

S7), many of which have been implicated in developmental processes. These included genes that specify transcription factors important for differentiation into extra-embryonic, endodermal, mesodermal, and ectodermal lineages (e.g. ESX1l, HOXB1, MEIS1, PAX6, LHX5, LBX1, MYF5, ONECUT1) (Appendix A Table S5). Moreover, nearly half of the transcription factor genes that were bound by the three regulators and transcriptionally inactive encoded developmentally important homeodomain proteins (Table 2). These results demonstrate that Oct4, Sox2, and Nanog occupy a set of repressed genes that are key to developmental processes.

Table 2. Examples of Inactive Homeodomain Genes Co-occupied by Oct4, Sox2, and Nanog

Gene Symbol	Entrez Gene ID	Gene Name
ATBF1	463	AT-binding transcription factor 1
DLX1	1745	distal-less homeo box 1
DLX4	1748	distal-less homeo box 4
DLX5	1749	distal-less homeo box 5
EN1	2019	engrailed homolog 1
ESX1L	80712	extraembryonic, spermatogenesis, homeobox 1-like
GBX2	2637	gastrulation brain homeobox 2
GSC	145258	goosecoid
HOP	84525	homeodomain-only protein
HOXB1	3211	homeo box B1
HOXB3	3213	homeo box B3
HOXC4	3221	homeo box C4
IPF2	3651	insulin promoter factor 2
ISL1	3670	ISL1 transcription factor, LIM/homeodomain (islet-1)
LBX1	10660	transcription factor similar to D. melanogaster homeodomain protein lady bird late
LHX2	9355	LIM homeobox 2
LHX5	64211	LIM homeobox 5
MEIS1	4211	myeloid ecotropic viral integration site 1 homolog (mouse)
NKX2-2	4821	NK2 transcription factor related, locus 2 (Drosophila)
NKX2-3	159296	NK2 transcription factor related, locus 3 (Drosophila)
ONECUT1	3175	one cut domain, family member 1
OTP	23440	orthopedia homolog (Drosophila)
OTX1	5013	orthodenticle homolog 1(Drosophila)
PAX6	5080	paired box gene 6
TITF1	7080	thyroid transcription factor 1

To determine which of the Oct4, Sox2, and Nanog bound genes were preferentially expressed in ES cells, we compared expression datasets (Abeyta et al., 2004; Sato et al., 2003) from ES cells and a compendium of differentiated tissues and cell types (Su et al., 2004) (Figure 3; Appendix A). It was notable that DPPA4, TDGF1, OCT4, NANOG, and LEFTY2 were at the top of the rank order list of genes that are bound and preferentially expressed in ES cells (Figure 3A). All five of these genes have been implicated in pluripotency (James et al., 2005; Mitsui et al., 2003; Chambers et al., 2003; Nichols et al., 1998; Bortvin et al., 2003). Moreover, several genes that encode developmentally important homeodomain proteins such as DLX5, HOXB1, LHX5, TITF1, LBX1, and HOP were at the bottom of this list, indicating that they are preferentially repressed in ES cells.

Figure 3

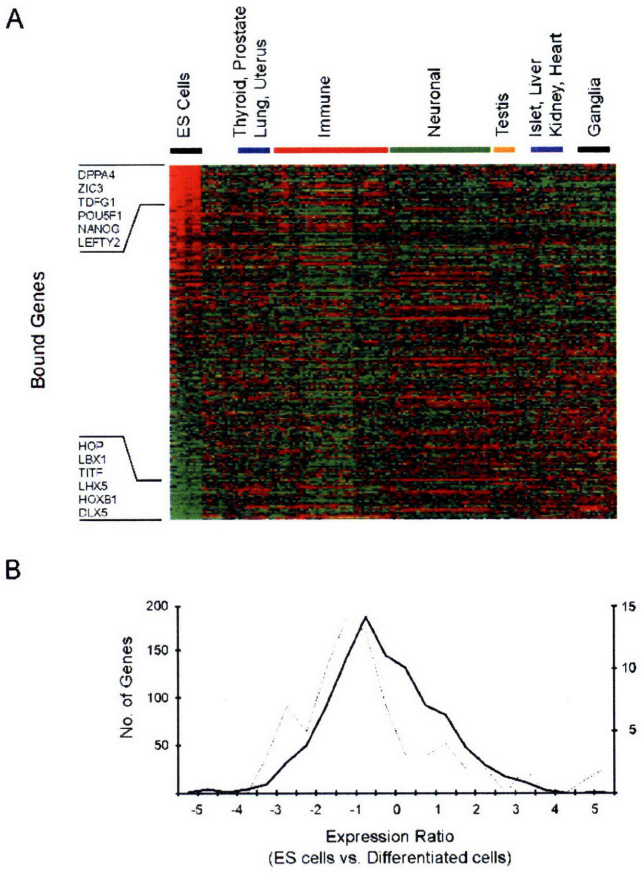

A
ES Cells
Thyroid, Prostate
Lung, Uterus
Immune
Neuronal
Testis
Islet, Liver
Kidney, Heart
Ganglia
DPPA4
ZIC3
TDFG1
POU5F1
NANOG
LEFTY2
HOP
LBX1
TITF
LHX5
HOXB1
DLX5
Bound Genes
B
No. of Genes
200
150
100
50
15
10
5
-5
-4
-3
-2
-1
0
1
2
3
4
5
Expression Ratio
(ES cells vs. Differentiated cells)

Figure 3. Expression of Oct4, Sox2, and Nanog co-occupied genes.

A. Affymetrix expression data for ES cells was compared to a compendium of expression data from 158 experiments representing 79 other differentiated tissues and cell types (Appendix A). Ratios were generated by comparing gene expression in ES cells to the median level of gene expression across all datasets for each individual gene. Genes were ordered by relative expression in ES cells and the results were clustered by expression experiment using hierarchical clustering. Each gene is represented as a separate row and individual expression experiments are in separate columns. Red indicates higher expression in ES cells relative to differentiated cells. Green indicates lower expression in ES cells relative to differentiated cells. Examples of bound genes that are at the top and bottom of the rank order list are shown.

B. Relative levels of gene expression in H9 ES cells compared to differentiated cells were generated and converted to log2 ratios. The distribution of these fold changes was calculated to derive a profile for different sets of genes. Data are shown for the distribution of expression changes between H9 ES cells and differentiated tissues for transcription factor genes that are not occupied by Oct4, Sox2 or Nanog (solid black line) and transcription factor genes occupied by all three (dotted line). The change in relative expression is indicated on the x axis and the numbers of genes in each bin are indicated on the y axes (left axis for unoccupied genes, right axis for occupied genes). The shift in distribution of expression changes for genes occupied by Oct4, Sox2 and Nanog is significant (p-value < 0.001 using a two-sampled Kolmogorov-Smirnov test), consistent with the model that Oct4, Sox2 and Nanog are contributing to the regulation of these genes.

The observation that Oct4, Sox2, and Nanog bound to transcriptionally active genes that have roles in pluripotency and transcriptionally inactive genes that promote development suggests that these binding events are regulatory. Two additional lines of evidence indicated that many of the binding events identified in this study contribute to regulation of their target genes. First, some of the genes identified here (e.g., OCT4, SOX2, and NANOG) were previously shown to be regulated by Oct4 and Sox2 in mouse ES cells (Catena et al., 2004; Kuroda et al., 2005, Okumura-Nakanishi et al., 2005; Rodda et al., 2005). Second, we further explored the hypothesis that bound genes are regulated by these transcription factors by taking advantage of the fact that Oct4 and Nanog are expressed in ES cells but their expression is rapidly downregulated upon differentiation. We compared the expression of Oct4, Sox2 and Nanog occupied genes in human ES cells with expression patterns in 79 differentiated cell types (Su et al., 2004) (Appendix A) and focused the analysis on transcription factor genes because these were the dominant functional class targeted by the ES cell regulators (Figure 3B). We expected that for any set of genes that there would be a characteristic change in expression levels between ES cells and differentiated cells. If Oct4, Sox2 and Nanog do not regulate the genes they occupy, then these genes should have the same general expression profile as the control population. We found, however, a significant shift in the distribution of expression changes for genes occupied by Oct4, Sox2 and Nanog (p-value < 0.001). Taken together, these data support the model that Oct4, Sox2 and Nanog functionally regulate the genes they occupy and suggest that loss of these regulators upon differentiation results in increased expression of genes necessary for development and reduced expression of a set of genes required for the maintenance of stem cell identity.

Our results suggest that Oct4, Sox2, and Nanog contribute to pluripotency and self-renewal by activating their own genes and genes encoding components of key signaling pathways and by repressing genes that are key to developmental processes. It is presently unclear how the three key regulators can activate some genes and repress others. It is likely that the activity of these key transcription factors is further controlled by additional cofactors, the precise levels of Oct4, Sox2, and Nanog, and by post-translational modifications.

Core transcriptional regulatory circuitry in ES cells

In order to identify regulatory network motifs associated with Oct4, Sox2 and Nanog, we assumed that regulator binding to a gene implies regulatory control and used algorithms that were previously devised to discover such regulatory circuits in yeast (Lee et al., 2002). The simplest units of commonly used transcriptional regulatory network architecture, or network motifs, provide specific regulatory capacities such as positive and negative feedback loops to control the levels of their components (Lee et al., 2002; Milo et al., 2002; Shen-Orr et al., 2002).

Our data indicated that Oct4, Sox2 and Nanog form feedforward loops that involve at least 353 protein coding and 2 miRNA genes (Figure 4A). Feedforward loop motifs contain a regulator that controls a second regulator, and have the additional feature that both regulators bind a common target gene. The feedforward loop has multiple regulatory capacities that may be especially useful for stem cells. When both regulators are positive, the feedforward loop can provide consistent activity that is relatively insensitive to transient changes in input (Mangan et al., 2003; Shen-Orr et al., 2002). If

the regulators have positive and negative functions, the feedforward loop can act as a switch that enables a rapid response to inputs by providing a time-sensitive delay where the downstream regulator acts to counter the effects of the upstream regulator in a delayed fashion (Mangan and Alon, 2003; Mangan et al., 2003). In ES cells, both regulatory capacities could be useful for maintaining the pluripotent state while retaining the ability to react appropriately to differentiation signals. Previous studies have shown that feedforward loop architecture has been highly favored during the evolution of transcriptional regulatory networks in less complex eukaryotes (Lee et al., 2002; Ma et al., 2004; Milo et al., 2002; Resendis-Antonio et al., 2005; Shen-Orr et al., 2002). Our data suggest that feedforward regulation is an important feature of human ES cells as well.

Our results also showed that Oct4, Sox2, and Nanog together bound to the promoters of their own genes, forming interconnected autoregulatory loops (Figure 4B; see also Appendix A, Figure S2). Transcriptional regulation of Oct4, Sox2, and Nanog by the Oct4-Sox2 complex was recently described in murine ES cells (Catena et al., 2004; Kuroda et al., 2005; Okumura-Nakanishi et al., 2005; Rodda et al., 2005). Our data indicate that this autoregulatory loop is conserved in human ES cells and more importantly, that Nanog is a component of the regulatory apparatus at these genes. Thus, it is likely that the expression and function of these three key stem cell factors are inextricably linked to one another. Autoregulation is thought to provide several advantages, including reduced response time to environmental stimuli and increased stability of gene expression (McAdams and Arkin, 1997; Rosenfeld et al., 2002; Shen-Orr et al., 2002; Thieffry et al., 1998).

Figure 4

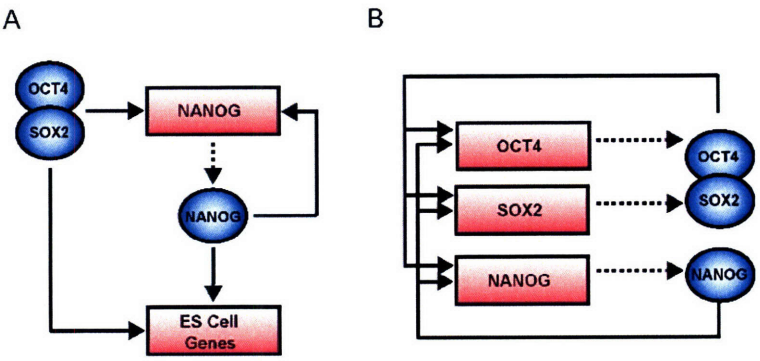

A
B
OCT4
SOX2
NANOG
NANOG
ES Cell Genes
OCT4
SOX2
NANOG
OCT4
SOX2
NANOG

Figure 4. Transcriptional regulatory motifs in human ES cells

A. Shown here is an example of feed-forward transcriptional regulatory circuitry in human ES cells. Regulators are represented by blue circles; gene promoters are represented by red rectangles. Binding of a regulator to a promoter is indicated by a solid arrow. Genes encoding regulators are linked to their respective regulators by dashed arrows. **B.** The interconnected autoregulatory loop formed by Oct4, Sox2, and Nanog.

The autoregulatory and feedforward circuitry described here may provide regulatory mechanisms by which stem cell identity can be robustly maintained, yet permit cells to respond appropriately to developmental cues. Modifying Oct4 and Nanog levels and function can change the developmental potential of murine ES cells (Avilion et al., 2003; Chambers et al., 2003; Mitsui et al., 2003; Nichols et al., 1998; Niwa et al., 2000) and this might be interpreted as being a consequence of perturbing independent regulatory pathways under the control of these two regulators. Our results argue that the levels and functions of these key stem cell regulators are tightly linked at both target genes and at their own promoters and thus provide an additional framework for interpreting the genetic studies. Changes in the relative stoichiometry of these factors would disturb the autoregulatory and feedforward circuitry, producing changes in global gene regulation and thus cell fate.

Expanded transcriptional regulatory circuitry

An initial model for ES cell transcriptional regulatory circuitry was constructed by identifying Oct4, Sox2, and Nanog target genes that encode transcription factors and chromatin regulators, and integrating knowledge of the functions of these downstream regulators in both human and mouse based on the available expression studies and literature (Figure 5). The model includes a subset of active and a subset of repressed target genes based on the extensive expression characterization of the 353 co-bound genes as described earlier. The active targets include genes encoding components of chromatin remodeling and histone modifying complexes (e.g. SMARCAD1, MYST3, and SET), which may have general roles in transcriptional regulation and genes encoding

transcription factors (e.g. REST, SKIL, HESX1, STAT3) which themselves are known to regulate specific genes. For instance, the REST protein has recently been shown to be present at high levels in ES cells and is required for the coordinate expression of a number of genes during neurogenesis, in part through repression of neuronal genes in the inappropriate cell type (Ballas et al., 2005). Previous studies have proposed that Nanog may function through the Tgf-β pathway in ES cells (Chambers, 2004). Our model suggests that this occurs through direct regulation of key components of this pathway (e.g. TDFG1, LEFTY2/EBAF) and through regulation of at least one transcription factor, SKIL, which controls the activity of downstream components of this pathway (SMAD2, SMAD4) (He et al., 2003). Our data also reveal that Oct4, Sox2, and Nanog co-occupy STAT3, a key regulator of self-renewal in mouse ES cells (Chambers, 2004), suggesting that Stat3 may also play a role in human ES cells.

The model described in Figure 5 also depicts a subset of the genes bound by Oct4, Sox2, and Nanog that are inactive and that encode transcription factors that have key roles in differentiation and development. These include regulators with demonstrated roles in development of all embryonic lineages. This initial model for ES cell transcriptional regulatory circuitry is consistent with previous genetic studies in mice that suggest that Oct4 and Nanog maintain pluripotency through repression of differentiation programs (Chambers et al., 2003; Mitsui et al., 2003; Niwa et al., 2000). This model also provides a mechanistic framework for understanding how this is accomplished through regulation of specific sets of genes that control cell fate specification.

Figure 5

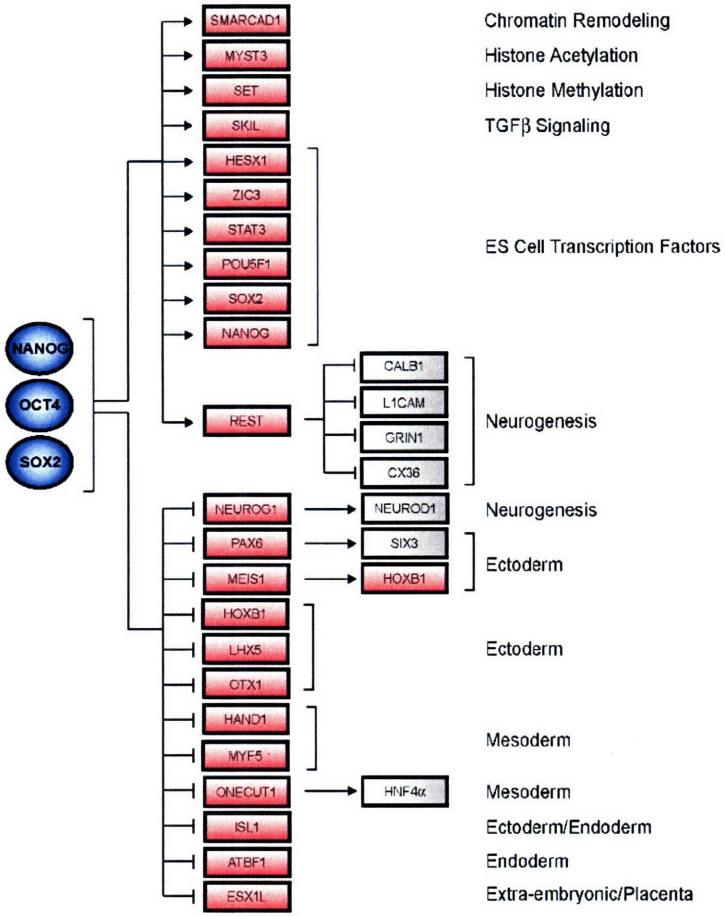

NANOG
OCT4
SOX2
SMARCAD1
MYST3
SET
SKIL
HESX1
ZIC3
STAT3
POU5F1
SOX2
NANOG
REST
CALB1
L1CAM
GRIN1
CX36
NEUROG1
PAX6
MEIS1
NEUROD1
SIX3
HOXB1
HOXB1
LHX5
OTX1
HAND1
MYF5
ONECUT1
HNF4α
ISL1
ATBF1
ESX1L
Chromatin Remodeling
Histone Acetylation
Histone Methylation
TGFβ Signaling
ES Cell Transcription Factors
Neurogenesis
Neurogenesis
Ectoderm
Ectoderm
Mesoderm
Mesoderm
Ectoderm/Endoderm
Endoderm
Extra-embryonic/Placenta

Figure 5. Core transcriptional regulatory network in human ES cells.

A model for the core transcriptional regulatory network was constructed by identifying Oct4, Sox2, and Nanog target genes that encode transcription factors and chromatin regulators, and integrating knowledge of the functions of these downstream regulators based on comparison to multiple expression datasets (Appendix A) and to the literature. A subset of active and inactive genes co-occupied by the three factors in human ES cells is shown here. Regulators are represented by blue circles; gene promoters are represented by red rectangles; grey boxes represent putative downstream target genes. Positive regulation was assumed if the target gene was expressed whereas negative regulation was assumed if the target gene was not transcribed.

Conclusions

Discovering how gene expression programs are controlled in living cells promises to improve our understanding of cell biology, development and human health. Identifying the target genes for key transcriptional regulators of human stem cells is a first critical step in the process of understanding these transcriptional regulatory networks and learning how they control cell identity. Mapping Oct4, Sox2, and Nanog to their binding sites within known promoters has revealed that these regulators collaborate to form regulatory circuitry in ES cells consisting of specialized autoregulatory and feedforward loops. Continued advances in our ability to culture and genetically manipulate human ES cells will allow us to test and manipulate this circuitry. Identification of the targets of additional transcription factors and chromatin regulators using the approaches described here should allow investigators to produce a more comprehensive map of transcriptional regulatory circuitry in these cells. Connecting signaling pathways to this circuit map may reveal how these pluripotent cells can be stimulated to differentiate into different cell types or how to reprogram differentiated cells back to a pluripotent state.

Experimental Procedures

Growth Conditions for Human Embryonic Stem Cells

Human embryonic stem (ES) cells were obtained from WiCell (Madison, WI; NIH Code WA09). Detailed protocol information on human ES cell growth conditions and culture reagents are available at http://www.mcb.harvard.edu/melton/hues. Briefly, passage 34 cells were grown in KO-DMEM medium supplemented with serum replacement, basic fibroblast growth factor (bFGF), recombinant human leukemia inhibitory factor (LIF) and a human plasma protein fraction. In order to minimize any MEF contribution in our analysis, H9 cells were cultured on a low density of irradiated murine embryonic fibroblasts (ICR MEFs) resulting in a ratio of approximately >8:1 H9 cell to MEF. The culture of H9 on low-density MEFs had no adverse affects on cell morphology, growth rate, or undifferentiated status as compared to cells grown under typical conditions. In addition, immunohistochemistry for pluripotency markers (e.g. Oct4, SSEA-3) indicated that H9 cells grown on a minimal feeder layer maintained the ability to generate derivates of ectoderm, mesoderm, and endoderm upon differentiation (Appendix A Figures S3 and S4).

Antibodies

The Nanog (AF1997) and Sox2 (AF2018) antibodies used in this study were immunoaffinity purified against the human proteins and shown to recognize their target proteins in Western blots and by immunocytochemistry (R&D Systems Minneapolis, MN). Multiple Oct4 antibodies directed against different portions of the protein, some of which were immunoaffinity purified (AF1759 R&D Systems, sc-8628 Santa Cruz, sc-

9081 Santa Cruz) were used in this study and have been shown to recognize their target protein in Western blots and by immunocytochemistry. The E2F4 antibody used in this study was obtained from Santa Cruz (sc-1082) and has been shown to recognize E2F4-responsive genes identified in previous ChIP studies (Table S2) (Ren et al., 2002; Weinmann et al., 2002).

Chromatin Immunoprecipitation

Protocols describing all materials and methods can be downloaded from http://jura.wi.mit.edu/young/hESRegulation/. Human embryonic stem cells were grown to a final count of $5x10^7 - 1x10^8$ cells for each location analysis reaction. 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 1xPBS and harvested using a silicon scraper and flash frozen in liquid nitrogen and 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. We used a Misonix Sonicator 3000 and sonicated at power 7 for 10 x 30 second pulses (90 second pause between pulses) at 4°C while samples were immersed in an ice bath. The resulting whole cell extract was incubated overnight at 4°C with 100 μl of Dynal Protein G magnetic beads that had been preincubated with 10 μg of the appropriate antibody. Beads were washed 5 times with RIPA buffer and 1 time with TE containing 50 mM NaCl. Bound complexes were eluted from the beads by heating at 65°C 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. Immunoprecipitated DNA and

whole cell extract DNA were then purified by treatment with RNAseA, proteinase K and multiple phenol:chloroform:isoamyl alcohol extractions. Purified DNA was blunted and ligated to linker and amplified using a two-stage PCR protocol. Amplified DNA was labeled and purified using Invitrogen Bioprime random primer labeling kits (immunoenriched DNA was labeled with Cy5 fluorophore, whole cell extract DNA was labeled with Cy3 fluorophore). Labeled DNA was combined (5 – 6 μg each of immunoenriched and whole cell extract DNA) and hybridized to arrays in Agilent hybridization chambers for 40 hours at 40°C. Arrays were then washed and scanned (Appendix A).

Array Design and Data Extraction

The design of the 10-slide oligo-based promoter arrays used in this study and data extraction methods are described in detail in Appendix A. Arrays were manufactured by Agilent Technologies (www.agilent.com). All microarray data is available at ArrayExpress under the following accession designation E-WMIT-5.

Acknowledgements

We would like to thank Biology and Research Computing (BaRC) and the Center for Microarray Technology (CMT) at the Whitehead Institute for computational and technical support. We would also like to thank members of the Young lab as well as Chad Cowan and Kevin Eggan for helpful discussions. L.A.B. was supported by NRSA postdoctoral fellowship CA094664 and H. L. M. by NRSA postdoctoral fellowship GM068273 This work was supported by NHGRI grant HG002668 to D.K.G. and R.A.Y. and NIH grant GM069400 to R.A.Y. T.I.L., D.K.G., and R.A.Y. consult for Agilent Technologies.

References

Abeyta MJ, Clark AT, Rodriguez RT, Bodnar, MS, Pera, RA and Firpo MT (2004). "Unique gene expression signatures of independently derived human embryonic stem cell lines." Hum Mol Genet **13:** 601-608.

Ambrosetti DC, 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.

Avilion AA, Nicolis SK, Pevny LH, 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.

Ballas N, Grunseich C, Lu DD, Speh JC and Mandel G (2005). "REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis." Cell **121**: 645-657.

Bartel DP (2004). "MicroRNAs: genomics, biogenesis, mechanism, and function." Cell **116**: 281-297.

Bortvin A, Eggan K, Skaletsky H, Akutsu H, Berr DL, Yanagimachi R, Page DC and Jaenisch R (2003). "Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei." Development **130**: 1673-1680.

Botquin V, Hess H, Fuhrmann G, Anastassiadis C, Gross MK, Vriend G and Scholer HR (1998). "New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2." Genes Dev 12: 2073-2090.

Brandenberger R, Khrebtukova I, Thies RS, Miura T, Jingli C, Puri R, Vasicek T, Lebkowski J and Rao M (2004). "MPSS profiling of human embryonic stem cells." BMC Dev Biol **4**: 10.

Catena R, Tiveron C, Ronchi A, Porta S, Ferri A, Tatangelo L, Cavallaro M, Favaro R, Ottolenghi S, Reinbold R, et al. (2004). "Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells." J Biol Chem **279**: 41846-41857.

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

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.

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.

Guenther MG, Jenner RG, Chevalier B, Nakamura T, Croce CM, Canaani E and Young RA (2005). "Global and Hox-specific roles for the MLL1 methyltransferase." Proc Natl Acad Sci **102**: 8603-8608.

Hart AH, 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.

Hay DC, 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 J, Tegen SB, Krawitz AR, Martin GS and Luo K (2003). "The transforming activity of Ski and SnoN is dependent on their ability to repress the activity of Smad proteins." J Biol Chem **278**: 30540-30547.

Hombria JC and Lovegrove B (2003). "Beyond homeosis--HOX function in morphogenesis and organogenesis." Differentiation **71**: 461-476.

James D, Levine AJ, 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.

Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM and Rajewsky K (2005). "Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing." Genes Dev **19**: 489-501.

Kuroda T, Tada M, Kubota H, Kimura H, Hatano SY, Suemori H, Nakatsuji N and Tada T (2005). "Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression." Mol Cell Biol **25**: 2475-2485.

Lee JH, Hart SR and Skalnik DG (2004). "Histone deacetylase activity is required for embryonic stem cell differentiation." Genesis **38**: 32-38.

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison, CT, Thompson, CM, Simon I, et al. (2002). "Transcriptional regulatory networks in Saccharomyces cerevisiae." Science **298**: 799-804.

Lewis BP, Burge CB and Bartel DP (2005). "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." Cell **120**: 15-20.

Ma HW, Kumar B, Ditges U, Gunzer F, Buer J and Zeng AP (2004). "An extended transcriptional regulatory network of Escherichia coli and analysis of its hierarchical structure and network motifs." Nucleic Acids Res **32**: 6643-6649.

Mangan S and Alon U (2003). "Structure and function of the feed-forward loop network motif." Proc Natl Acad Sci **100**: 11980-11985.

Mangan S, Zaslaver A and Alon U (2003). "The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks." J Mol Biol **334**: 197-204.

Matin MM, Walsh JR, Gokhale PJ, Draper JS, Bahrami AR, Morton I, Moore HD and Andrews PW (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.

McAdams HH and Arkin A (1997). "Stochastic mechanisms in gene expression." Proc Natl Acad Sci **94**: 814-819.

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D and Alon U (2002). "Network motifs: simple building blocks of complex networks." Science **298**: 824-827.

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.

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.

Nishimoto M, Fukushima A, Okuda A and Muramatsu M (1999). "The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2." Mol Cell Biol **19**: 5453-5465.

Niwa H (2001). "Molecular mechanism to maintain stem cell renewal of ES cells." Cell Struct Funct **26**: 137-148.

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

Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, Volkert TL, Schreiber J, Rolfe PA, Gifford DK, et al. (2004). "Control of pancreas and liver gene expression by HNF transcription factors." Science **303**: 1378-1381.

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.

Pasquinelli AE, Hunter S and Bracht J (2005). "MicroRNAs: a developing story." Curr Opin Genet Dev **15**: 200-205.

Pera MF and Trounson AO (2004). "Human embryonic stem cells: prospects for development." Development **131**: 5515-5525.

Pesce M and Scholer HR (2001). "Oct-4: gatekeeper in the beginnings of mammalian development." Stem Cells **19**: 271-278.

Remenyi A, Scholer HR and Wilmanns M (2004). "Combinatorial control of gene expression." Nat Struct Mol Biol **11**: 812-815.

Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, Young RA and Dynlacht BD (2002). "E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints." Genes Dev **16**: 245-256.

Resendis-Antonio O, Freyre-Gonzalez JA, Menchaca-Mendez R, Gutierrez-Rios RM, Martinez-Antonio A, Avila-Sanchez C and Collado-Vides J (2005). "Modular analysis of the transcriptional regulatory network of E. coli." Trends Genet **21**: 16-20.

Reubinoff BE, Pera MF, Fong CY, Trounson A and Bongso A (2000). "Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro." Nat Biotechnol **18**: 399-404.

Rodda DJ, 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**(26): 24731-7.

Rosenfeld N, Elowitz MB 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 AH (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.

Sato N, Sanjuan IM, Heke M, Uchida M, Naef F and Brivanlou AH (2003). "Molecular signature of human embryonic stem cells and its comparison with the mouse." Dev Biol **260**: 404-413.

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

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

Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR and Hogenesch JB (2004). "A gene atlas of the mouse and human protein-encoding transcriptomes." Proc Natl Acad Sci **101**: 6062-6067.

Thieffry D, Salgado H, Huerta AM 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 JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS and Jones JM (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**: 1145-1147.

Wei CL, Miura T, Robson P, Lim SK, Xu XQ, Lee MY, Gupta S, Stanton L, Luo Y, Schmitt J, Thies S, Wang W, Khrebtukova I, Zhou D, Liu ET, Ruan YJ, Rao M and Lim B (2005). "Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state." Stem Cells **23**: 166-185.

Weinmann AS, Yan PS. Oberley MJ, Huang TH and Farnham PJ (2002). "Isolating human transcription factor targets by coupling chromatin immunoprecipiation and CpG island microarray analysis." Genes Dev **16**: 235-244.

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

Zaehres H, Lensch MW, Daheron L, Stewart SA, Itskovitz-Eldor J and Daley GQ (2005). "High-Efficiency RNA Interference in Human Embryonic Stem Cells." Stem Cells **23**: 299-305.

Chapter 3

Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells

Published as: Lee* TI, Jenner* RG, Boyer* LA, Guenther* MG, Levine* SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-13.
* These authors contributed equally to this work.

My contributions to this work

The effort to profile chromatin regulators and RNA polymerase in ES cells was initiated in the summer of 2004. I was involved in this collaborative project from its inception, participating in the experimental efforts and the data analysis. I worked with Tony Lee to help improve the technical aspects of the lab's ChIP-chip protocol to adapt it for the purpose of genome-wide profiling. Towards this end, I helped to develop a high-throughput protocol for mammalian ChIP-chip (Appendix B; Lee et al., 2006). My principle contributions were developing the techniques necessary to hybridize 115 microarrays for individual experiments with minimal experimental variability. I was also responsible for helping lead the efforts to amplify, label and hybridize the three ChIP samples described in this chapter.

Upon obtaining high-quality genome-wide binding data for Suz12 and RNA Polymerase, I contributed to the analysis by inspecting binding plots to determine the parameters and quality of binding events for the different ChIP samples. Part of this effort included identifying patterns of binding in order to understand the role of Suz12 at its target genes. I also provided edits to the manuscript and contributed to figures of binding plot data.

Summary

Polycomb group proteins are essential for early development in metazoans but their contributions to human development are not yet well understood. We have mapped the Polycomb Repressive Complex 2 (PRC2) subunit Suz12 across the entire non-repeat portion of the genome in human embryonic stem (ES) cells. We found that Suz12 is distributed across large portions of over two hundred genes encoding key developmental regulators. These genes are occupied by nucleosomes trimethylated at histone H3K27, are transcriptionally repressed, and contain some of the most highly conserved non-coding elements in the vertebrate genome. We found that preferential activation of PRC2 target genes occurs during differentiation of ES cells into other cell types. The ES cell transcriptional regulators Oct4, Sox2 and Nanog co-occupied a significant subset of these genes, further supporting a link between repression of developmental regulators and stem cell pluripotency. These results indicate that PRC2 occupies a special set of developmental genes in ES cells that must be repressed to maintain pluripotency and that are poised for activation during ES cell differentiation.

Introduction

Embryonic stem (ES) cells are a unique self-renewing cell type that can give rise to the ectodermal, endodermal, and mesodermal germ layers during embryogenesis. Human ES cells, which can be propagated in culture in an undifferentiated state but selectively induced to differentiate into many specialized cell types, are thought to hold great promise for regenerative medicine (Thomson et al., 1998; Reubinoff et al., 2000; Mayhall et al., 2004; Pera and Trounson, 2004). The gene expression program of ES cells must allow these cells to maintain a pluripotent state but also allow for differentiation into more specialized states when signaled to do so. How the gene expression program of ES cells is regulated to accomplish this is not well understood but may be key to realizing the therapeutic potential of ES cells and further understanding early development.

Among regulators of development, the Polycomb group proteins (PcG) are of special interest. These regulators were first described in Drosophila, where they repress the homeotic genes controlling segment identity in the developing embryo (Lewis, 1978; Denell and Frederick, 1983; Simon et al., 1992; Orlando and Paro, 1995; Pirrotta, 1998; Kennison, 2004). The initial repression of these genes is carried out by DNA-binding transcriptional repressors, and PcG proteins modify chromatin to maintain these genes in a repressed state (Duncan, 1986; Bender et al., 1987; Strutt et al., 1997; Horard et al., 2000; Hodgson et al., 2001; Mulholland et al., 2003).

The PcG proteins form multiple Polycomb Repressive Complexes (PRCs), the components of which are conserved from Drosophila to humans (Franke et al., 1992;

Shao et al., 1999; Birve et al., 2001; Tie et al., 2001; Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Levine et al., 2002). The PRCs are brought to the site of initial repression and act through epigenetic modification of chromatin structure to promote gene silencing (Pirrotta, 1998; Levine et al., 2004; Lund and van Lohuizen, 2004; Ringrose and Paro, 2004). PRC2 catalyzes histone H3 lysine-27 (H3K27) methylation and this enzymatic activity is required for PRC2-mediated gene silencing (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Muller et al., 2002). H3K27 methylation is thought to provide a binding surface for PRC1, which facilitates oligomerization, condensation of chromatin structure and inhibition of chromatin remodeling activity in order to maintain silencing (Shao et al., 1999; Francis et al., 2001; Cao et al., 2002; Czermin et al., 2002).

Components of PRC2 are essential for the earliest stages of vertebrate development (Faust et al., 1998; O'Carroll et al., 2001; Pasini et al., 2004). PRC2 and its related complexes PRC3 and PRC4 contain the core components Ezh2, Suz12 and Eed (Kuzmichev et al., 2004; Kuzmichev et al., 2005). Ezh2 is a H3K27 methyltransferase and Suz12 (Suppressor of zeste 12) is required for this activity (Cao and Zhang, 2004; Pasini et al., 2004). ES cell lines cannot be established from Ezh2-deficient blastocysts (O'Carroll et al., 2001), suggesting that PRC2 is involved in regulating pluripotency and self-renewal. Although the PRCs are known to repress individual Hox genes (van der Lugt et al., 1996; Akasaka et al., 2001; Wang et al., 2002; Cao and Zhang, 2004), it is not clear how these important PcG regulators contribute to early development in vertebrates.

Because the nature of PRC2 target genes in ES cells might reveal why PRC2 is essential for early embryonic development, pluripotency and self-renewal, we have

mapped the sites occupied by the Suz12 subunit throughout the genome in human ES cells. This genome-wide map reveals that PRC2 is associated with a remarkable cadre of genes encoding key regulators of developmental processes that are repressed in ES cells. The genes occupied by PRC2 contain nucleosomes that are trimethylated at histone H3 lysine-27 (H3K27me3), a modification catalyzed by PRC2 and associated with the repressed chromatin state. Both PRC2 and nucleosomes with histone H3K27me3 occupy surprisingly large genomic domains around these developmental regulators and are frequently associated with highly conserved non-coding sequence elements previously identified by comparative genomic methods. The transcription factors Oct4, Sox2 and Nanog, which are also key regulators of ES cell pluripotency and self-renewal, occupy a significant subset of these genes. Thus, the model of epigenetic regulation of homeotic genes extends to a large set of developmental regulators whose repression in ES cells appears to be key to pluripotency. We suggest that PRC2 functions in ES cells to repress developmental genes that are preferentially activated during differentiation.

Results and Discussion

Mapping genome occupancy in ES cells

We mapped the location of both RNA polymerase II and the Suz12 subunit of PRC2 genome-wide in human ES cells (Figure 1). The initiating form of RNA polymerase II was mapped to test the accuracy of the method and provide a reference for comparison with sites occupied by PRC2. The Suz12 subunit of PRC2 is critical for the function of the complex and was selected for these genome-wide experiments. Human ES cells (H9, NIH code WA09) were cultured as described previously, analyzed by immunohistochemistry for characteristic stem cell markers, tested for their ability to generate cell types from all three germ layers upon differentiation into embryoid bodies and shown to form teratomas in immunocompromised mice (Appendix C; Figures S1-S3).

Figure 1. Genome-wide ChIP-Chip in human embryonic stem cells

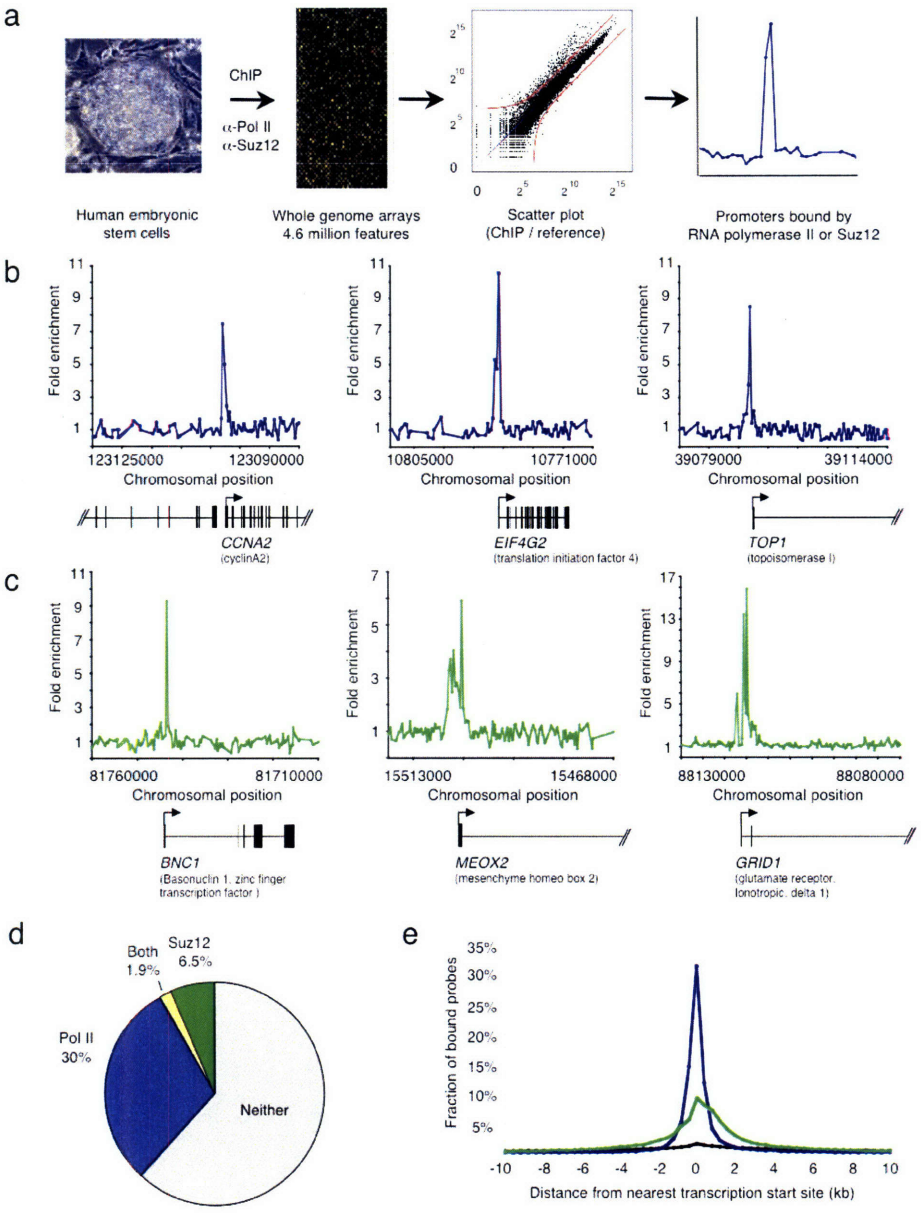


Figure 1. Genome-wide ChIP-Chip in human embryonic stem cells.

a. DNA segments bound by the initiation form of RNA polymerase II or Suz12 were isolated using chromatin-immunoprecipitation (ChIP) and identified with DNA microarrays containing over 4.6 million unique 60-mer oligonucleotide probes spanning the entire non-repeat portion of the human genome. ES cell growth and quality control, the antibodies, ChIP protocol, DNA microarray probe design and data analysis methods are described in detail in Appendix C.

b. Examples of RNA polymerase II ChIP signals from genome-wide ChIP-Chip. The plots show unprocessed enrichment ratios (blue) for all probes within a genomic region (ChIP vs. whole genomic DNA). Chromosomal positions are from NCBI build 35 of the human genome. Genes are shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.

c. Examples of Suz12 ChIP signals from genome-wide ChIP-Chip. The plots show unprocessed enrichment ratios (green) for all probes within a genomic region (ChIP vs. whole genomic DNA). Chromosomal positions, genes and notations are as described in **b**.

d. Chart showing percentage of all annotated genes bound by RNA polymerase II (blue), Suz12 (green), both (yellow) or neither (grey).

e. Distribution of the distance between bound probes and the closest transcription start sites from RefSeq, Ensembl, MGC, UCSC Known Genes and H-Inv databases for Suz12 (green line) and RNA polymerase II (blue line). The number of bound probes is given as the percentage of total probes and is calculated for 400 bp intervals from the start site. The null-distribution of the distance between all probes and the closest transcription is shown as a black line.

DNA sequences bound by RNA polymerase II were identified in replicate chromatin-immunoprecipitation (ChIP) experiments using DNA microarrays that contain over 4.6 million unique 60-mer oligonucleotide probes spanning the entire non-repeat portion of the human genome (Figure 1 and Appendix C). To obtain a probabilistic assessment of binding events genome-wide, an algorithm was implemented that incorporates information from multiple probes representing contiguous regions of the genome and threshold criteria were established to identify a dataset with minimal false positives and false negatives (Appendix C). RNA polymerase II was associated with the promoters of 7,106 of the approximately 22,500 annotated human genes, indicating that one-third of protein-coding genes are prepared to be transcribed in ES cells. Three lines of evidence suggest this dataset is high quality. First, eighty-seven percent of the RNA polymerase II sites occurred at promoters of known or predicted genes. Additionally, previously reported expression experiments in ES cells detected transcripts for 88% of the genes bound by RNA polymerase II. Finally, independent analysis using gene-specific PCR (Appendix C) indicated that the frequency of false positives was approximately 4% and the frequency of false negatives was approximately 30% in this dataset. A detailed analysis of the RNA polymerase II dataset can be found in Appendix C (Tables S1-S6 and Figures S4 and S5).

The sites occupied by Suz12 were then mapped throughout the entire non-repeat genome in H9 ES cells using the same approach described for RNA polymerase II (Figure 1C). Suz12 was associated with the promoters of 1,893 of the approximately 22,500 annotated human genes, indicating that ~8% of protein-coding genes are occupied

by Suz12 in ES cells (Appendix C; Tables S7 and S8). Independent site-specific analysis indicated that the frequency of false positives was approximately 3% and the frequency of false negatives was approximately 27% in this dataset.

Comparison of the genes occupied by Suz12 with those occupied by RNA polymerase II revealed that the two sets were largely exclusive (Figure 1D; Appendix C; Table S8). There were, however, genes where Suz12 and RNA polymerase II co-occupied promoters. At these genes, PRC complexes may fail to block assembly of the preinitiation complex (Dellino et al., 2004) consistent with the observation that Polycomb group proteins can associate with components of the general transcription apparatus (Breiling et al., 2001; Saurin et al., 2001).

It was notable that, like RNA polymerase II, the vast majority of Suz12 bound sequences were found at gene promoters (Figure 1E). 95% of the Suz12 bound regions were found within 1 kb of known or predicted transcription start sites (Appendix C; Table S7). This suggests that Suz12 functions in human ES cells primarily at promoters rather than at distal regulatory elements.

Global transcriptional repression by PRC2

PRC2 is composed of three core subunits, Suz12, Eed and Ezh2, and has been shown to mediate histone H3K27 methylation at specific genes in vivo (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Muller et al., 2002; Cao and Zhang, 2004; Kirmizis et al., 2004). To confirm that Suz12 is associated with active PRC2 at target genes, we used chromatin immunoprecipitation with antibodies against Eed and the histone H3K27me3 mark and analyzed the results with promoter microarrays. We found

that Eed and the histone H3K27me3 mark co-occurred with Suz12 at most genes (Figure 2A; Appendix C, Figure S6), as shown for NeuroD1 in Figure 2B.

Genetic and biochemical studies at selected genes indicate that PRC2-mediated H3K27 methylation represses gene expression, but it has not been established if it acts as a repressor genome-wide. If genes occupied by Suz12 are repressed by PRC2, then transcripts from these genes should generally be present at lower levels in ES cells than in differentiated cell types. To test this idea, we compared the expression levels of PRC2-occupied genes in four different ES cell lines with the expression level of these genes in 79 differentiated human cell and tissue types (Sato et al., 2003; Abeyta et al., 2004; Su et al., 2004). We found that PRC2 occupied genes were generally underexpressed in ES cells relative to other cell types (Figure 2C). A small fraction of the genes occupied by PRC2 were overexpressed in ES cells (Figure 2C) and these tended to show limited Suz12 occupancy (Appendix C). These results are consistent with the model that PRC2-mediated histone H3K27 methylation promotes gene silencing at the majority of its target genes throughout the genome in ES cells.

Figure 2. Suz12 is associated with Eed, histone H3K27me3 modification and transcriptional repression in ES cells

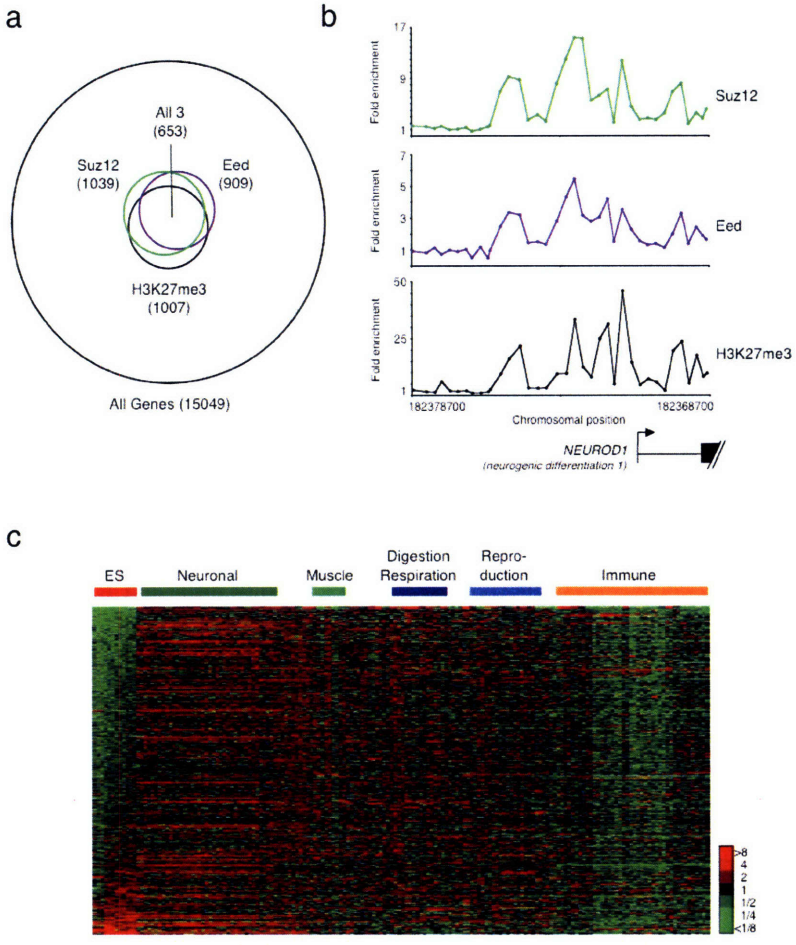


Figure 2. Suz12 is associated with Eed, histone H3K27me3 modification and transcriptional repression in ES cells.

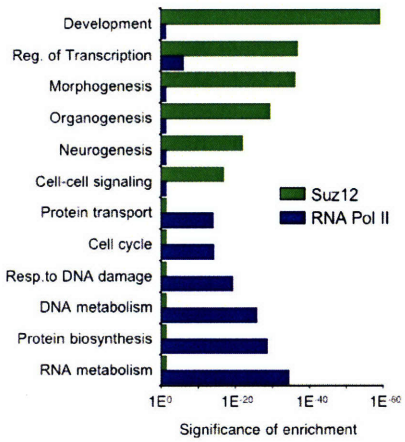
a. Venn diagram showing the overlap of genes bound by Suz12 at high-confidence, genes bound by Eed at high-confidence and genes trimethylated at H3K27 at high-confidence. The data are from promoter microarrays that contain probes tiling -8 kb and +2 kb around transcription start. 72% of the genes bound by Suz12 at high-confidence are also bound by Eed at high-confidence; others are bound by Eed at lower confidence (Figure S6).
b. Suz12 (top), Eed (middle) and H3K27me3 (bottom) occupancy at NeuroD1. The plots show unprocessed enrichment ratios for all probes within this genomic region (Suz12 ChIP vs. whole genomic DNA, Eed ChIP vs whole genomic DNA and H3K27me3 ChIP vs. total H3 ChIP). Chromosomal positions are from NCBI build 35 of the human genome. NeuroD1 is shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.
c. Relative expression levels of 604 genes occupied by PRC2 and trimethylated at H3K27 in ES cells. Comparisons were made across 4 ES cell lines and 79 differentiated cell types. Each row corresponds to a single gene that is bound by Suz12, associated with Eed and H3K27me3 and for which Affymetrix expression data is available. Each column corresponds to a single expression microarray. ES cells are in the following order; H1, H9, HSF6, HSF1. For each gene, expression is shown relative to the average expression level of that gene across all samples, with shades of red indicating higher than average expression and green lower than average expression according to the scale on the right. Cell types are grouped by tissue or organ function and genes are ranked according the significance of their relative level of gene expression in ES cells.

Key developmental regulators are targets of PRC2

Examination of the targets of Suz12 revealed that they were remarkably enriched for genes that control development and transcription (Figure 3 and examples in Figure 1C) and that Suz12 tended to occupy large domains at these genes (Figure 4). Although only 8% of all annotated genes were occupied by Suz12, ~50% of those encoding transcription factors associated with developmental processes were occupied by Suz12. By comparison, RNA polymerase II preferentially occupied genes involved in a broad spectrum of cell proliferation functions such as nucleic acid metabolism, protein synthesis and cell cycle (Figure 3A and examples in Figure 1B; Appendix C, Table S10).

Figure 3. Cellular functions of genes occupied by Suz12

a



b

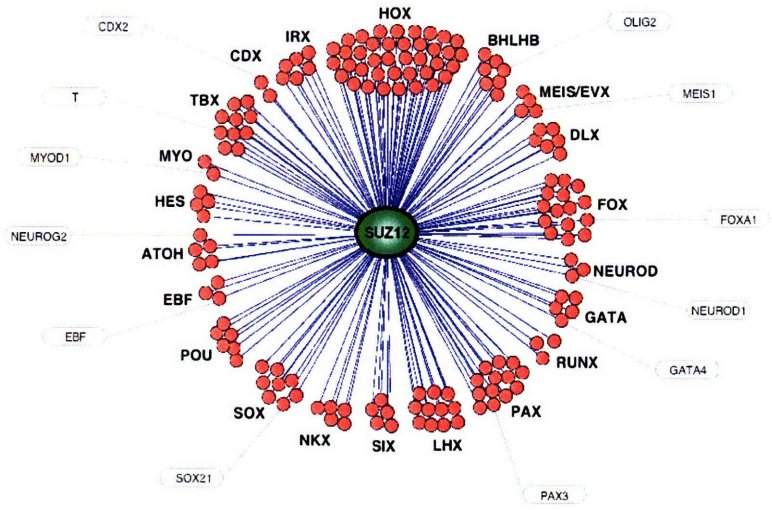


Figure 3. Cellular functions of genes occupied by Suz12

a. Genes bound by Suz12 or RNA polymerase II were compared to biological process gene ontology categories; highly represented categories are shown. Ontology terms are shown on the y-axis; p-values for the significance of enrichment are graphed along the x-axis (Suz12 in green, RNA polymerase II in blue).

b. Selected examples of developmental transcription factor families bound by Suz12. Suz12 is represented by the green oval; individual transcription factors are represented by circles and grouped by family as indicated. Examples of transcription factors with defined roles in development are labeled. Transcription factor families include homeobox protein (HOX: all Hox genes except HOXA11S), basic helix-loop-helix domain containing, class B (BHLHB: BHLHB1/OLIG2, BHLHB2, BHLHB3, BHLHB4, BHLHB5, BHLHB6/OLIG1, BHLHB7/OLIG3), Hox co-factors (MEIS/EVX: MEIS1, MEIS2, MEIS3, EVX1), distal-less homeobox (DLX: DLX1, DLX2, DLX3, DLX4, DLX5, DLX6), Forkhead box (FOX: FOXA1, FOXA2, FOXB1, FOXC2, FOXD1, FOXD2, FOXD3, FOXE1, FOXF1, FOXG1B, FOXJ1, FOXL1, FOXL2, FOXP2, FOXP4), NEUROD (NEUROD1, NEUROD2, NEUROD3/NEUROG1), GATA binding protein (GATA: GATA2, GATA3, GATA4, GATA5, GATA6), runt related transcription factor (RUNX: RUNX1, RUNX2, RUNX3), paired box and paired-like (PAX1, PAX3, PAX5, PAX6, PAX7, PAX8, PAX9, PRRX1, PRRXL1, PITX1, PITX2, PHOX2A, PHOX2B), LIM homeobox (LHX: LHX1, LHX2, LHX4, LHX5, LHX6, LHX8, LHX9, LMX1A, LMX1B, ISL1, ISL2), sine oculis homeobox homolog (SIX: SIX1, SIX2, SIX3, SIX5, SIX6), NK transcription factor related (NKX: NKX2-2, NKX2-6, NKX2-8, NKX6-1, NKX6-2), SRY box (SOX: SOX3, SOX6, SOX8, SOX9, SOX12, SOX14, SOX17, SOX21), POU domain containing, classes 3 and 4 (POU: POU3F1, POU3F2, POU3F4, POU4F1, POU4F2, POU4F3), early B-cell factor (EBF: EBF, EBF2, EBF3), atonal homolog (ATOH: ATOH1, ATOH4/NEUROG2, ATOH5/NEUROG3, ATOH7, ATOH8), hairy and enhancer of split protein (HES: HES2, HES3, HES4, HES7), myogenic basic domain (MYO: MYOD1, MYF6), T-box (TBX: TBX1, TBX2, TBX3, TBX4, TBX5, TBX15, TBX18, TBX20, TBX21, T), caudal type homeobox (CDX2, CDX4), and iroquois homeobox protein (IRX: IRX1, IRX2, IRX3, IRX4, IRX5, IRX6).

Figure 4. Suz12 occupies large portions of genes encoding transcription factors with roles in development

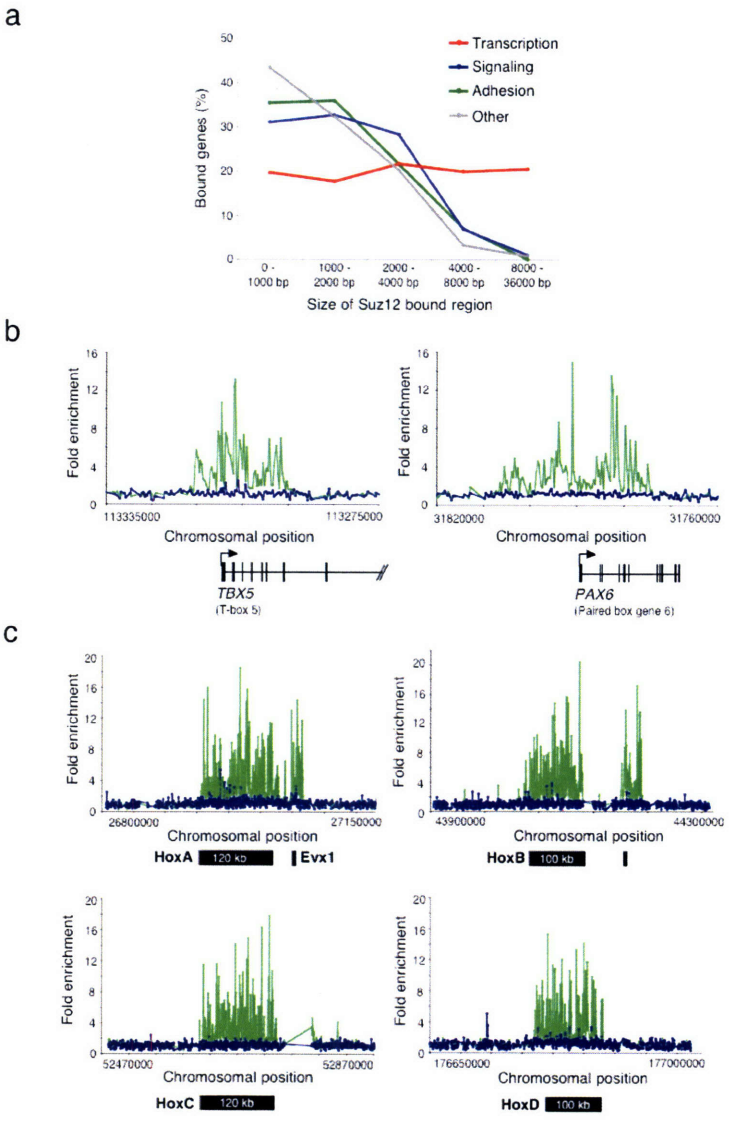


Figure 4. Suz12 occupies large portions of genes encoding transcription factors with roles in development

a. The fraction of Suz12 target genes associated with different sizes of binding domains. Genes are grouped into four categories according to their function: Signaling, Adhesion/migration, Transcription and Other.

b. Examples of Suz12 (green) and RNA polymerase II (blue) binding at the genes encoding developmental regulators TBX5 and PAX6. The plots show unprocessed enrichment ratios for all probes within a genomic region (ChIP vs. whole genomic DNA). Genes are shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.

c. Binding profiles of Suz12 (green) and RNA polymerase II (blue) across ~500 kb regions encompassing Hox clusters A-D. Unprocessed enrichment ratios for all probes within a genomic region are shown (ChIP vs. whole genomic DNA). Approximate Hox cluster region sizes are indicated within black bars.

It was striking that Suz12 occupied many families of genes that control development and transcription (Figure 3B; Appendix C, Figure S7 and Table S11). These included 39 of 40 of the homeotic genes found in the Hox clusters and the majority of homeodomain genes. Suz12-bound homeodomain genes included almost all members of the DLX, IRX, LHX and PAX gene families, which regulate early developmental steps in neurogenesis, hematopoiesis, axial patterning, tissue patterning, organogenesis and cell-fate specification. Suz12 also occupied promoters for large subsets of the FOX, SOX and TBX gene families. The forkhead family of FOX genes is involved in axial patterning and tissue development from all three germ layers (Carlsson and Mahlapuu, 2002; Lehmann et al., 2003). Mutations in members of the SOX gene family alter cell-fate specification and differentiation and are linked to several developmental diseases (Schepers et al., 2002). The TBX family of genes regulates a wide variety of developmental processes such as gastrulation, early pattern formation, organogenesis and limb formation (Logan, 2003; Showell et al., 2004). Thus, the genes preferentially bound by Suz12 have functions that, when expressed, promote differentiation. This is likely to explain, at least in part, why PRC2 is essential for early development and ES cell pluripotency.

A remarkable feature of Suz12 binding at most genes encoding developmental regulators was the extensive span over which the regulator occupied the locus (Figure 4; Appendix C, Figure S8). For the majority (72%) of bound sites across the genome, Suz12 occupied a small region in a promoter region similar in size to those bound by RNA polymerase II (Figure 1). For the remaining bound regions, Suz12 occupancy encompassed large domains spanning 2-35 kb and extending from the promoter into the

gene. A large portion of genes encoding developmental regulators (72%) exhibited these extended regions of Suz12 binding. In some cases, binding encompassed multiple contiguous genes. For instance, Suz12 binding extended ~100 kb across the entire HoxA, HoxB, HoxC and HoxD clusters but did not bind to adjacent genomic sequences, yielding a highly defined spatial pattern (Figure 4B). In contrast, clusters of unrelated genes, such as the interleukin 1-beta cluster, were not similarly bound by Suz12. Thus, genes encoding developmental regulators showed an unusual tendency to be occupied by Suz12 over much or all of their transcribed regions.

PRC2 and highly conserved elements

Previous studies have noted that many highly conserved non-coding elements of vertebrate genomes are associated with genes encoding developmental regulators (Bejerano et al., 2004; Siepel et al., 2005; Woolfe et al., 2005). Given Suz12's strong association with this class of genes, we investigated the possibility that Suz12 bound regions are associated with these highly conserved elements. Inspection of individual genes suggested that Suz12 occupancy was associated with regions of sequence conservation (Figure 5A). Eight percent of the approximately 1,400 highly conserved non-coding DNA elements described by Woolfe and colleagues (Woolfe et al., 2005) were found to be associated with the Suz12-bound developmental regulators (p-value 10^{-14}). Using entries from the PhastCons database of conserved elements (Siepel et al., 2005), we found that Suz12 occupancy of highly conserved elements was highly significant (using highly conserved elements with a LoD conservation score of 100 or better, the p-value for significances was less than 10^{-85}). Since PRC2 has not been shown

to directly bind DNA sequences, we expect that specific DNA-binding proteins occupy the highly conserved DNA sequences and may associate with PRC2, which spreads and occupies adjacent chromatin. Thus, the peaks of Suz12 occupancy might not be expected to precisely colocate with the highly conserved elements, even if these elements are associated with PRC2 recruitment.

Remarkably, the degree of the association between Suz12 binding and conserved sequences increases when considering sequences with an increasing degree of conservation (Figure 5B). By comparison, RNA polymerase II showed no such enrichment. These results suggest that the subset of highly conserved non-coding elements at genes encoding developmental regulators may be associated with PcG mediated silencing of these regulators.

Figure 5. Suz12 binding is associated with highly conserved regions

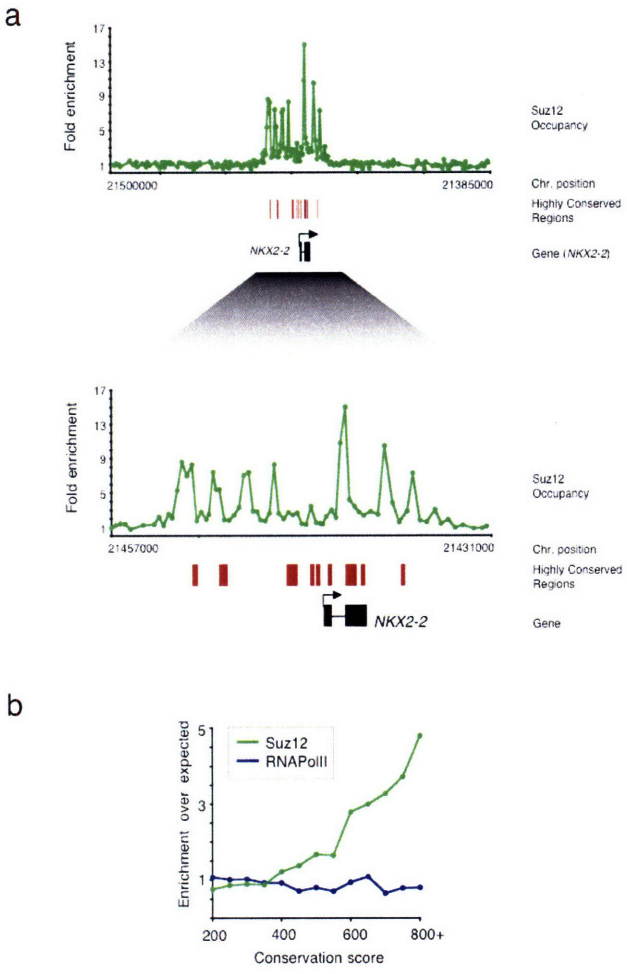


Figure 5. Suz12 binding is associated with highly conserved regions

a. Suz12 occupancy (green) and conserved elements are shown at NKX2-2 and adjacent genomic regions. The plots show unprocessed enrichment ratios for all probes within this genomic region (Suz12 ChIP vs. whole genomic DNA). Conserved elements (red) with LoD scores > 160 derived from the PhastCons program (Siepel et al., 2005) are shown to scale above the plot. Genes are shown to scale below plots (exons are represented by vertical bars). A higher resolution view is also shown below.

b. Enrichment of conserved non-coding elements within Suz12 (green) and RNA polymerase II (blue) bound regions. The maximum non-exonic PhastCons conservation score was determined for each bound region. For comparison, the same parameter was determined using a randomized set of genomic regions with the same size distribution. The graph displays the ratio of the number of bound regions with that score versus the number of randomized genomic regions with that score.

Signaling genes are among PRC2 targets

The targets of Suz12 were also enriched for genes that encode components of signaling pathways (Figure 3A and Table S12). Current studies suggest that the transforming growth factor-beta (TGFβ), bone morphogenic protein (BMP), wingless-type MMTV integration site (Wnt), and fibroblast growth factor (FGF) signaling pathways that are required for gastrulation and lineage differentiation in the embryo are also essential for self-renewal and differentiation of ES cells in culture (Loebel et al., 2003; Molofsky et al., 2004; Mishra et al., 2005; Varga and Wrana, 2005). We noted that Suz12 occupied the promoters of multiple components of these pathways. Although most of these genes were occupied by Suz12 over a small portion of their promoters, a few were more notable because Suz12 occupied large portions of their genes and these loci contained highly conserved elements, as observed for the developmental regulators. This group contained members of the Wnt family (WNT1, WNT2, WNT6) as well as components of the TGFβ superfamily (BMP2, GDF6). Recent studies have shown that Wnt signaling plays a role in pluripotency and self-renewal in both mouse and human ES cells (Sato et al., 2004) and our results suggest that it is important to maintain specific family members in a repressed state in ES cells.

Activation of PRC2 target genes during differentiation

PRC2 is associated with an important set of developmental regulators that must be silent in ES cells but activated during differentiation. This observation suggests that PRC2 ultimately functions to repress occupied genes in ES cells and that these genes may be especially poised for transcriptional activation during ES cell differentiation. We

reasoned that if this model is correct, genes bound by Suz12 should be preferentially activated upon ES cell differentiation or in cells that lack Suz12. Furthermore, in differentiated cells, Suz12 might continue to be observed at silent genes but must be removed from genes whose expression is essential for that cell type.

We first examined gene expression in ES cells stimulated to undergo differentiation (Sato et al., 2003). We found that genes occupied by Suz12 were more likely to be activated during ES cell differentiation than genes that were not occupied by Suz12 (Figure 6A; Appendix C, Table S13), indicating that Suz12-occupied genes show preferential activation during differentiation under these conditions. This effect was particularly striking at the set of developmental regulators (Figure 6B). Suz12 occupied most (83%) of the developmental regulators that were induced more than 10-fold during ES cell differentiation.

Figure 6. Suz12 occupies genes poised for expression during differentiation

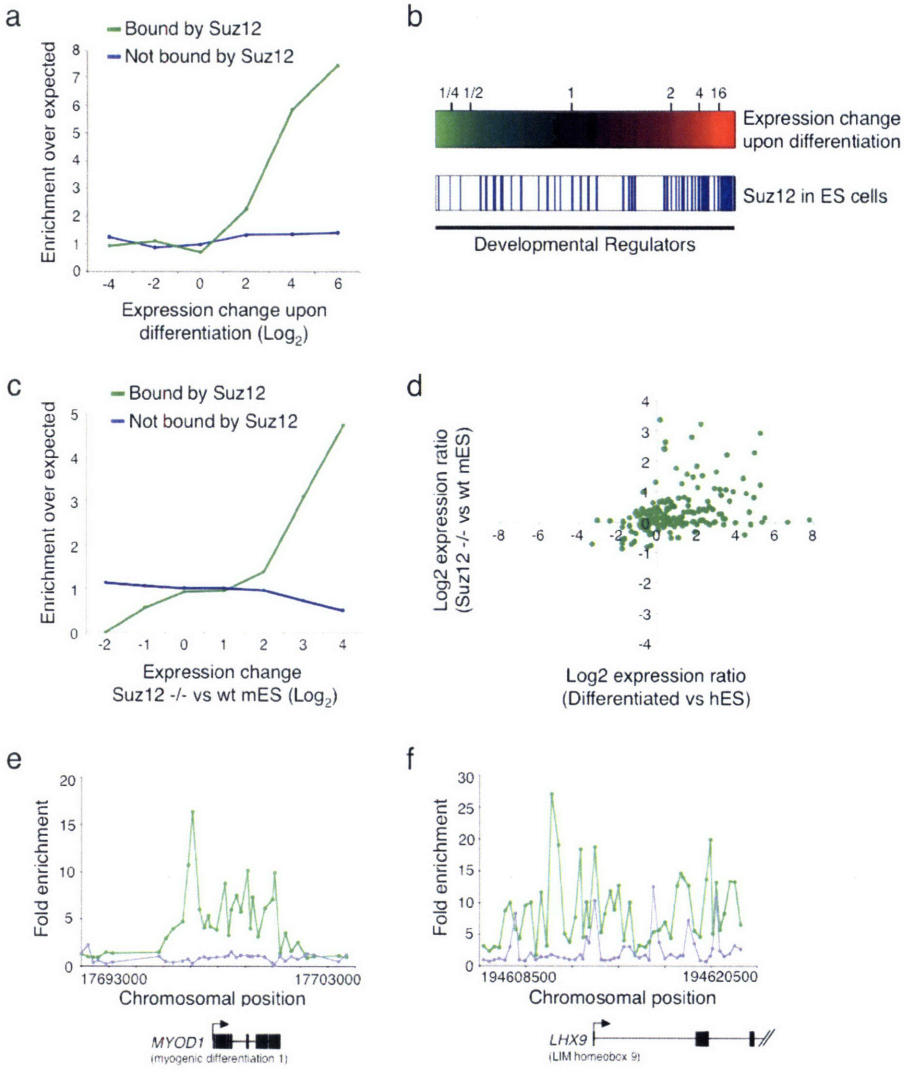


Figure 6. Preferential activation of PRC2 target genes during ES cell differentiation.

a. Fold enrichment in the number of genes induced or repressed during ES cell differentiation. The change in gene expression is given as the log(2) transformed ratio of the signals in differentiated cells versus pluripotent cells. The two lines show genes transcriptionally inactive in ES cells (absence of RNA polymerase II) and bound by Suz12 (green) and genes transcriptionally inactive in ES cells and repressed by other means (blue). In both cases, fold enrichment is calculated against the total population of genes.

b. Expression changes of genes encoding developmental regulators during ES cell differentiation. Expression ratio (differentiated / pluripotent) is represented by color, with shades of red indicating upregulation and shades of green downregulation according to the scale shown above. Genes are ordered according to change in gene expression, with genes exhibiting higher expression in pluripotent ES cells to the left and genes exhibiting higher expression in differentiated cells to the right. Genes bound by Suz12 in undifferentiated ES cells are indicated by blue lines in the lower panel.

c. Fold enrichment in the number of genes induced or repressed in Suz12-deficient mouse cells. The change in gene expression is given as the log(2) transformed ratio of the signals in Suz12-deficient cells versus wild-type ES cells. The two lines show genes transcriptionally inactive in human ES cells (absence of RNA polymerase II) and bound by Suz12 (green) and genes transcriptionally inactive in human ES cells and repressed by other means (blue). In both cases, fold enrichment is calculated against the total population of genes.

d. Gene expression ratios (log base 2) of Suz12 target genes in differentiated human H1 ES cells relative to pluripotent H1 ES cells (x-axis) and in Suz12-deficient mouse cells relative to wild-type mouse ES cells (y-axis).

e. Suz12 binding profiles across the gene encoding muscle regulator MYOD1 in H9 human ES cells (green) and primary human skeletal myotubes (grey). The plots show unprocessed enrichment ratios for all probes within a genomic region (ChIP vs. whole genomic DNA). Genes are shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.

f. Suz12 binding profiles across the gene encoding LHX9 in H9 human ES cells (green) and primary human skeletal myotubes (grey). The plots show unprocessed enrichment ratios for all probes within a genomic region (ChIP vs. whole genomic DNA). Genes are shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.

We next examined the expression of Suz12 target genes in Suz12-deficient cell lines derived from homozygous mutant blastocysts (Appendix C). We reasoned that genes bound by Suz12 in human ES cells have orthologs in mice that should be upregulated in Suz12-deficient mouse cells, although we expected the overlap in these sets of genes to be imperfect because of potential differences between human and mouse ES cells, the possible repression of PRC2 target genes by additional mechanisms, pleotropic effects of the Suz12 knockout on genes downstream of Suz12-target genes, and the false positive and negative rates in both binding and expression analysis. Differences in gene expression between Suz12 homozygous mutant cells and wild-type ES cells were measured using gene expression microarrays and the human Suz12 binding data mapped to orthologous mouse genes using Homologene (Wheeler et al., 2006). Among the orthologs in this comparison, we found that a significant portion of the genes bound by Suz12 in human ES cells were upregulated in Suz12-deficient mouse cells ($p = 6x10^{-4}$); these genes are listed in Table S14. The genes occupied by Suz12 in human ES cells were more likely to be activated in Suz12-deficient mouse cells than genes that were not occupied by Suz12 (Figure 6C). Furthermore, we found that the Suz12 target genes that were induced upon human ES cell differentiation were generally also induced upon loss of Suz12 in mouse cells (Figure 6D). Genes that were activated during ES cell differentiation and in Suz12-deficient cells included the transcriptional regulators GATA2, GATA3, GATA6, HAND1, MEIS2 and SOX17, the signaling proteins WNT5A, DKK1 (dickkopf homolog 1), DKK2, EFNA1 (ephrin A1), EFNB1, EPHA4 (ephrin receptor A4) and EPHB3 and the cell cycle inhibitor CDKN1A. These data

indicate that Suz12 is necessary to fully repress the genes that are occupied by PRC2 in wild-type ES cells.

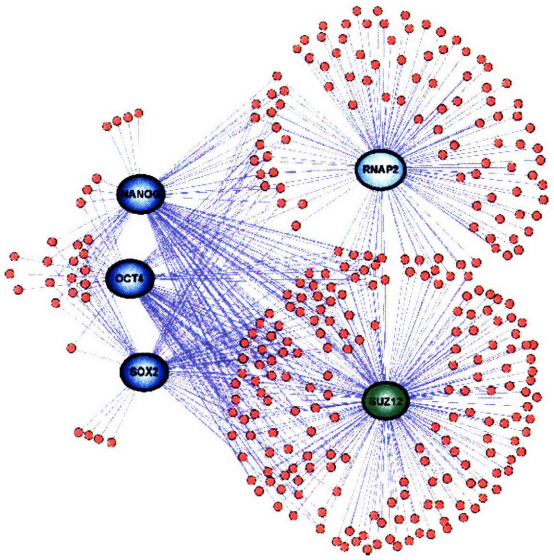
If PRC2 functions to repress genes in ES cells that are activated during differentiation, then in differentiated tissues Suz12 occupancy should be diminished at genes encoding developmental regulators that have a role in specifying the identity of that tissue, similar to results seen with Ezh2 at specific genes in mouse (Caretti et al., 2004). To test this, we designed an array focused on the promoters of developmental regulators and used ChIP-Chip to investigate Suz12 occupancy at these promoters in primary differentiated muscle cells. The results demonstrated that genes encoding key regulators of muscle differentiation, including MyoD, displayed greatly diminished Suz12 occupancy when compared to ES cells (Figure 6E). MyoD is a master regulator for muscle differentiation (Berkes and Tapscott, 2005; Tapscott, 2005), and the gene encoding this transcription factor displayed no significant Suz12 occupancy when compared to the levels of Suz12 occupancy observed in ES cells. Genes encoding other transcriptional regulators that play a central role in muscle development, such as Pax3 and Pax7 (Brand-Saberi, 2005), showed reduced levels of Suz12 occupancy in muscle cells relative to ES cells (Appendix C, Figure S11). In contrast, other developmental regulators important for differentiation of non-muscle tissues remained occupied by Suz12 in differentiated muscle cells (Figure 6F and Table S15). These data support a model where Suz12 binding in ES cells represses key developmental regulators that are later expressed during differentiation.

Targets of PRC2 are shared with key ES cell regulators

The transcription factors Oct4, Sox2 and Nanog have essential roles in early development and are required for the propagation of undifferentiated ES cells in culture (Nichols et al., 1998; Avilion et al., 2003; Chambers et al., 2003; Mitsui et al., 2003). We recently reported that these transcription factors occupied promoters for many important developmental regulators in human ES cells (Boyer et al., 2005). This led us to compare the set of genes encoding developmental regulators and occupied by Oct4, Sox2 and Nanog with those occupied by PRC2 (Figure 7; Appendix C). We found that each of the three DNA-binding transcription factors occupied approximately one-third of these PRC2-occupied developmental genes (Figure 7A, Table S11). Remarkably, we found that the subset of genes encoding developmental regulators that were occupied by Oct4, Sox2 and Nanog and repressed in the regulatory circuitry highlighted in Boyer et al. were almost all occupied by PRC2 (Figure 7B). These included transcription factors known to be important for differentiation into extra-embryonic, endodermal, mesodermal, and ectodermal lineages (e.g., ESX1L, ONECUT1, HAND1, HOXB1). As expected, active genes encoding ES cell transcription factors (e.g., ZIC3, STAT3, OCT4, NANOG) were occupied by Oct4, Sox2, Nanog and RNA polymerase II, but not PRC2 (Figure 7B).

Figure 7. Suz12 is localized to genes also bound by ES cell transcriptional regulators

a



b

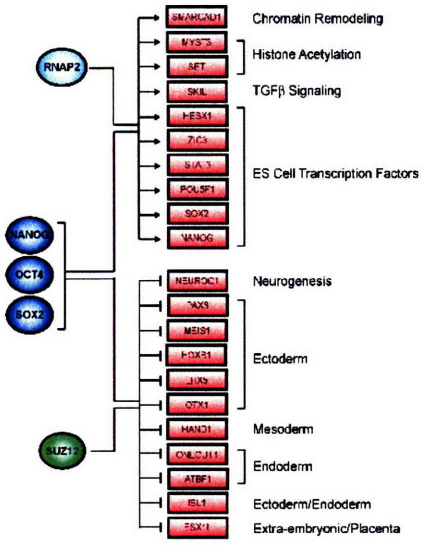


Figure 7. Suz12 is localized to genes also bound by ES cell transcriptional regulators.

a. Transcriptional regulatory network model of developmental regulators governed by Oct4, Sox2, Nanog, RNA polymerase II and Suz12 in human ES cells. The ES cell transcription factors each bound to approximately one-third of the PRC2-occupied, developmental genes. Developmental regulators were selected based on gene ontology. Regulators are represented by dark blue circles; RNA polymerase II is represented by a light blue circle; Suz12 is represented by a green circle; gene promoters for developmental regulators are represented by small red circles.
b. Suz12 occupies a set of repressed developmental regulators also bound by Oct4, Sox2 and Nanog in human ES cells. Genes annotated as bound by Oct4, Sox2 and Nanog previously and identified as active or repressed based on expression data (Boyer et al., 2005) were tested to see if they were bound by Suz12 or RNA polymerase II. Ten of eleven previously identified active genes were found to be bound by RNA polymerase II at known promoters while eleven of twelve previously identified repressed genes were bound by Suz12. Regulators are represented by dark blue circles; RNA polymerase II is represented by a light blue circle; Suz12 is represented by a green circle; gene promoters are represented by red rectangles.

The observation that Oct4, Sox2 and Nanog are bound to a significant subset of developmental genes occupied by PRC2 supports a link between repression of developmental regulators and stem cell pluripotency. Like PRC2, Oct4 and Nanog have been shown to be important for early development and ES cell identity. It is possible, therefore, that inappropriate regulation of developmental regulators that are common targets of Oct4, Nanog and PRC2 contributes to the inability to establish ES cell lines in OCT4, NANOG and EZH2 mutants (Nichols et al., 1998; O'Carroll et al., 2001; Chambers et al., 2003; Mitsui et al., 2003).

Conclusion

We have mapped the sites occupied by Suz12 throughout the genome to gain insights into how PRC2 contributes to pluripotency in human embryonic stem cells. ES cells proliferate in an undifferentiated state yet remain poised to respond to development cues. Genes encoding the transcriptional regulators that promote differentiation must therefore be repressed in ES cells but activated upon receiving signals to differentiate. We found that PRC2 occupies large domains at genes encoding a key set of repressed developmental regulators that are preferentially activated upon cellular differentiation, thus implicating this complex directly in the maintenance of the pluripotent state.

Transcription factors and chromatin regulators contribute to the transcriptional regulatory circuitry responsible for pluripotency and self-renewal in human ES cells. Understanding this circuitry is fundamental to understanding human development and realizing the therapeutic potential of these cells. In this context, we find it exciting that the outlines of the core transcriptional regulatory circuitry of human ES cells are emerging. The transcription factors Oct4, Sox2, and Nanog are associated with actively transcribed genes that contribute to growth and self-renewal (Boyer et al., 2005). These factors also occupy genes encoding key developmental regulators that are transcriptionally repressed, due at least in part to their association with PRC2 and nucleosomes modified at histone H3K27me3. Further study of transcription factors and chromatin regulators genome-wide will allow investigators to produce a more comprehensive map of transcriptional regulatory circuitry in ES cells and to test models

that emerge from the circuitry. This information may provide insights into approaches by which pluripotent cells can be stimulated to differentiate into different cell types.

Methods

See Appendix C.

References

Abeyta MJ, Clark AT, Rodriguez RT, Bodnar MS, Pera RA and Firpo MT (2004). "Unique gene expression signatures of independently-derived human embryonic stem cell lines." Hum Mol Genet **13**: 601-608.

Akasaka T, van Lohuizen M, van der Lugt N, Mizutani-Koseki Y, Kanno M, Taniguchi M, Vidal M, Alkema M, Berns A and Koseki H (2001). "Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression." Development **128**: 1587-1597.

Avilion AA, Nicolis SK, Pevny LH, 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.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS and Haussler D (2004). "Ultraconserved elements in the human genome." Science **304**, 1321-1325.

Bender M, Turner FR and Kaufman TC (1987). "A development genetic analysis of the gene regulator of postbithorax in Drosophila melanogaster." Dev Biol **119**: 418-432.

Berkes CA and Tapscott SJ (2005). "MyoD and the transcriptional control of myogenesis." Semin Cell Dev Biol **16**: 585-595.

Birve A, Sengupta AK, Beuchle D, Larsson J, Kennison JA, Rasmuson-Lestander A and Muller J (2001). "Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants." Development **128**: 3371-3379.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al. (2005). "Core transcriptional regulatory circuitry in human embryonic stem cells." Cell **122**: 947-956.

Brand-Saberi B (2005). "Genetic and epigenetic control of skeletal muscle development." Ann Anat **187**: 199-207.

Breiling A, Turner BM, Bianchi ME and Orlando V (2001). "General transcription factors bind promoters repressed by Polycomb group proteins." Nature **412**: 651-655.

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS and Zhang Y (2002). "Role of histone H3 lysine 27 methylation in Polycomb-group silencing." Science **298**: 1039-1043.

Cao R and Zhang Y (2004). "SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex." Mol Cell **15**: 57-67.

Caretti G, Di Padova M, Micales B, Lyons GE and Sartorelli V (2004). "The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation." Genes Dev **18**: 2627-2638.

Carlsson P and Mahlapuu M (2002). "Forkhead transcription factors: key players in development and metabolism." Dev Biol **250**: 1-23.

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.

Czermin B, Melfi R, McCabe D, Seitz V, Imhof A and Pirrotta V (2002). "Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites." Cell **111**: 185-196.

Dellino GI, Schwartz YB, Farkas G, McCabe D, Elgin SC and Pirrotta V (2004). "Polycomb silencing blocks transcription initiation." Mol Cell **13**: 887-893.

Denell RE and Frederick RD (1983). "Homoeosis in Drosophila: a description of the Polycomb lethal syndrome." Dev Biol **97**: 34-47.

Duncan I (1986). "Control of bithorax complex functions by the segmentation gene fushi tarazu of D. melanogaster." Cell **47**: 297-309.

Faust C, Lawson KA, Schork NJ, 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.

Francis NJ, Saurin AJ, Shao Z and Kingston RE (2001). "Reconstitution of a functional core polycomb repressive complex." Mol Cell **8**: 545-556.

Franke A, DeCamillis M, Zink D, Cheng N, Brock HW and Paro R (1992). "Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster." EMBO **11**: 2941-2950.

Hodgson JW, Argiropoulos B and Brock HW (2001). "Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing." Mol Cell Biol **21**: 4528-4543.

Horard B, Tatout C, Poux S and Pirrotta V (2000). "Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor." Mol Cell Biol **20**: 3187-3197.

Kennison JA (2004). "Introduction to Trx-G and Pc-G genes." Methods Enzymol **377**: 61-70.

Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, Green R and Farnham PJ (2004). "Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27." Genes Dev **18**: 1592-1605.

Kuzmichev A, Jenuwein T, Tempst P and Reinberg D (2004). "Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3." Mol Cell **14**: 183-193.

Kuzmichev A, Margueron R, Vaquero A, Preissner TS, Scher M, Kirmizis A, Ouyang X, Brockdorff N, Abate-Shen C, Farnham P and Reinberg D (2005). "Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation." Proc Natl Acad Sci **102**: 1859-1864.

Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P and Reinberg D (2002). "Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein." Genes Dev **16**: 2893-2905.

Lee TI, Johnstone SE, Young RA (2006). "Chromatin immunoprecipitation and microarray-based analysis of protein location." Nat Protoc **1**(2): 729-48.

Lehmann, OJ, Sowden JC, Carlsson P, Jordan T and Bhattacharya SS (2003). "Fox's in development and disease." Trends Genet **19**: 339-344.

Levine SS, King IF and Kingston RE (2004). "Division of labor in polycomb group repression." Trends Biochem Sci **29**: 478-485.

Levine SS, Weiss A, Erdjument-Bromage H, Shao Z, Tempst P and Kingston RE (2002). "The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans." Mol Cell Biol **22**: 6070-6078.

Lewis EB (1978). "A gene complex controlling segmentation in Drosophila." Nature **276**: 565-570.

Loebel DA, Watson CM, De Young RA and Tam PP (2003). "Lineage choice and differentiation in mouse embryos and embryonic stem cells." Dev Biol **264**: 1-14.

Logan M (2003). "Finger or toe: the molecular basis of limb identity." Development **130**: 6401-6410.

Lund AH and van Lohuizen M (2004). "Polycomb complexes and silencing mechanisms." Curr Opin Cell Biol **16**: 239-246.

Mayhall EA, Paffett-Lugassy N and Zon LI (2004). "The clinical potential of stem cells." Curr Opin Cell Biol **16**: 713-720.

Mishra L, Derynck R and Mishra B (2005). "Transforming Growth Factor-{beta} Signaling in Stem Cells and Cancer." Science **310**: 68-71.

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.

Molofsky AV, Pardal R and Morrison SJ (2004). "Diverse mechanisms regulate stem cell self-renewal." Curr Opin Cell Biol **16**: 700-707.

Mulholland NM, King IF and Kingston RE (2003). "Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA." Genes Dev **17**: 2741-2746.

Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE and Simon JA (2002). "Histone methyltransferase activity of a Drosophila Polycomb group repressor complex." Cell **111**: 197-208.

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.

O'Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA and Jenuwein T (2001). "The polycomb-group gene Ezh2 is required for early mouse development." Mol Cell Biol **21**: 4330-4336.

Orlando V and Paro R (1995). "Chromatin multiprotein complexes involved in the maintenance of transcription patterns." Curr Opin Genet Dev **5**: 174-179.

Pasini D, Bracken AP, Jensen MR, Denchi EL and Helin K (2004). "Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity." EMBO **23**: 4061-4071.

Pera MF and Trounson AO (2004). "Human embryonic stem cells: prospects for development." Development **131**: 5515-5525.

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

Reubinoff BE, Pera MF, Fong CY, Trounson A and Bongso A (2000). "Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro." Nat Biotechnol **18**: 399-404.

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

Sato N, Meijer L, Skaltsounis L, Greengard P and Brivanlou AH (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.

Sato N, Sanjuan IM, Heke M, Uchida M, Naef F and Brivanlou AH (2003). "Molecular signature of human embryonic stem cells and its comparison with the mouse." Dev Biol **260**: 404-413.

Saurin AJ, Shao Z, Erdjument-Bromage H, Tempst P and Kingston RE (2001). "A Drosophila Polycomb group complex includes Zeste and dTAFII proteins." Nature **412**: 655-660.

Schepers GE, Teasdale RD and Koopman P (2002). "Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families." Dev Cell **3**: 167-170.

Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W and Kingston RE (1999). "Stabilization of chromatin structure by PRC1, a Polycomb complex." Cell **98**: 37-46.

Showell C, Binder O and Conlon FL (2004). "T-box genes in early embryogenesis." Dev Dyn **229**: 201-218.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. (2005). "Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes." Genome Res **15**: 1034-1050.

Simon J, Chiang A and Bender W (1992). "Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products." Development **114**: 493-505.

Strutt H, Cavalli G and Paro R (1997). "Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression." EMBO **16**: 3621-3632.

Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, et al. (2004). "A gene atlas of the mouse and human protein-encoding transcriptomes." Proc Natl Acad Sci **101**: 6062-6067.

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

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS and Jones JM (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**: 1145-1147.

Tie F, Furuyama T, Prasad-Sinha J, Jane E and Harte PJ (2001). "The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3." Development **128**: 275-286.

van der Lugt NM, Alkema M, Berns A and Deschamps J (1996). "The Polycomb-group homolog Bmi-1 is a regulator of murine Hox gene expression." Mech Dev **58**: 153-164.

van Lohuizen M (1998). "Functional analysis of mouse Polycomb group genes." Cell Mol Life Sci **54**: 71-79.

Varga AC and Wrana JL (2005). "The disparate role of BMP in stem cell biology." Oncogene **24**: 5713-5721.

Wang J, Mager J, Schnedier E and Magnuson T (2002). "The mouse PcG gene eed is required for Hox gene repression and extraembryonic development." Mamm Genome **13**: 493-503.

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, et al. (2006). "Database resources of the National Center for Biotechnology Information." Nucleic Acids Res **34**: D173-180.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, et al. (2005). "Highly conserved non-coding sequences are associated with vertebrate development." PLoS Biol **3**: e7.

Chapter 4

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

Published as: Cole* MF, Johnstone* SE, Newman* JJ, Kagey MH, Young RA (2008). "Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells." Genes Dev **22**(6): 746-55.
* These authors contributed equally to this work.

My contributions to this work

Following the initial characterization of the core transcriptional regulatory circuitry of ES cells, I became interested in understanding how signaling pathways impact this circuitry. I felt that exploring this topic would be essential for understanding both how to cells maintain pluripotency in culture and how to direct cells to differentiate towards particular lineages. With this in mind, I worked with Megan Cole and Jamie Newman to explore the role of Wnt signaling in ES cells. We were interested in characterizing the role of Wnt in ES cells by understanding the target genes for this pathway.

I was highly involved in every aspect of this project. Megan, Jamie and I each contributed equally to the conceptualization of the project, the experimental design, executing the experiments, analyzing the data and writing the manuscript. To make our experimental efforts maximally efficient, we divided the experimental space. My responsibility was generating an ES cell line deficient in Tcf3 and analyzing the gene expression profile for these cells. I was also involved in much of the growing and manipulating of the ES cells that is described in this paper.

I was involved in analyzing both the genome-wide ChIP-chip data and expression data and exploring possible connections between Wnt signaling, Tcf3 and the core regulatory circuitry of ES cells. I contributed to each figure in the paper and the writing of the text.

Summary

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 (Appendix D 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 (Appendix D 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).

Figure 1

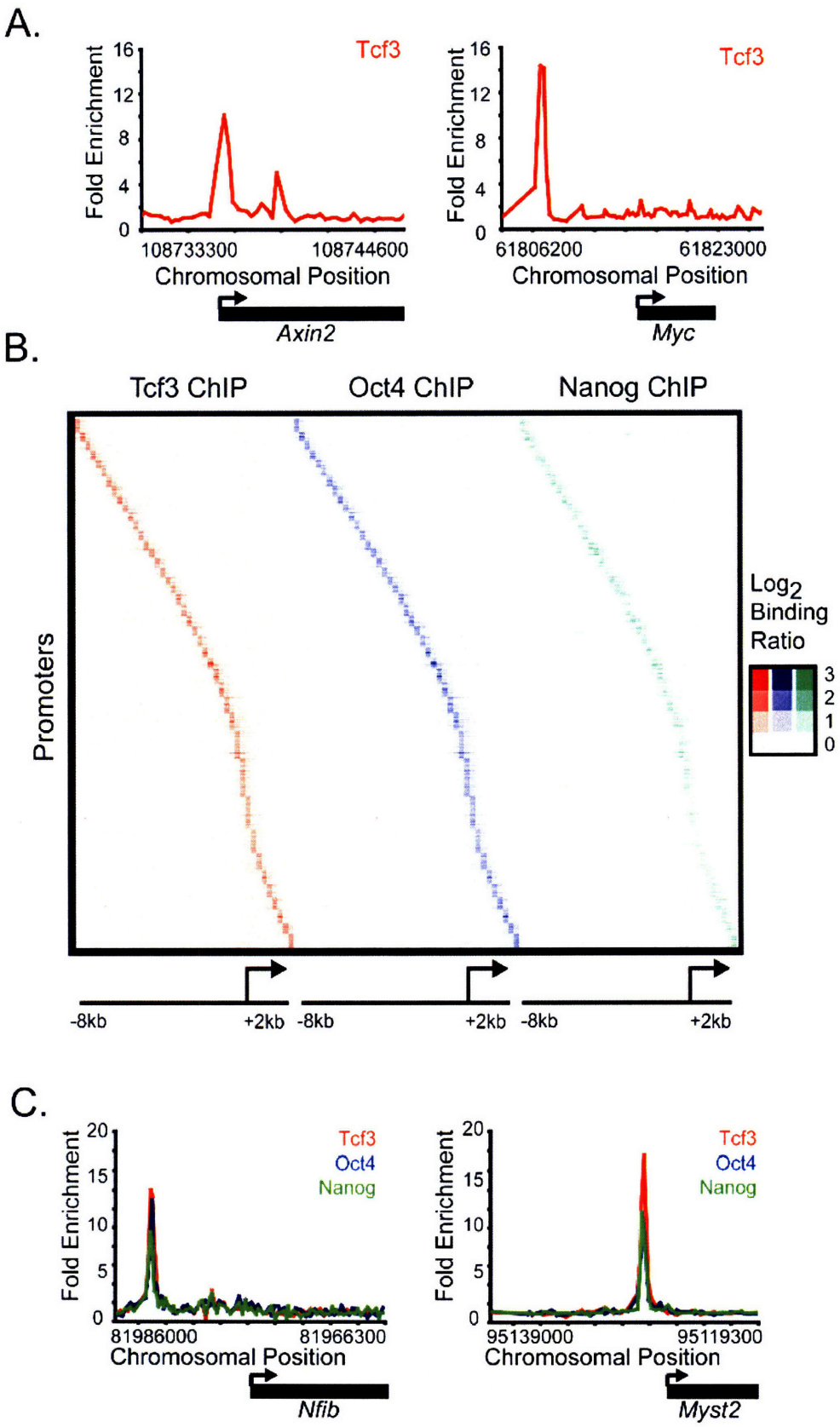

A.
Tcf3
Fold Enrichment
16
12
8
4
0
108733300
108744600
Chromosomal Position
Axin2
Tcf3
Fold Enrichment
16
12
8
4
0
61806200
61823000
Chromosomal Position
Myc
B.
Tcf3 ChIP
Oct4 ChIP
Nanog ChIP
Promoters
Log2 Binding Ratio
3
2
1
0
-8kb
+2kb
-8kb
+2kb
-8kb
+2kb
C.
Tcf3
Oct4
Nanog
Fold Enrichment
20
15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

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; Appendix D Fig. S2). These results identified a set of 1224 genes that are co-occupied by Tcf3 and Oct4 at high confidence (Appendix D 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, Appendix D 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.

Figure 2

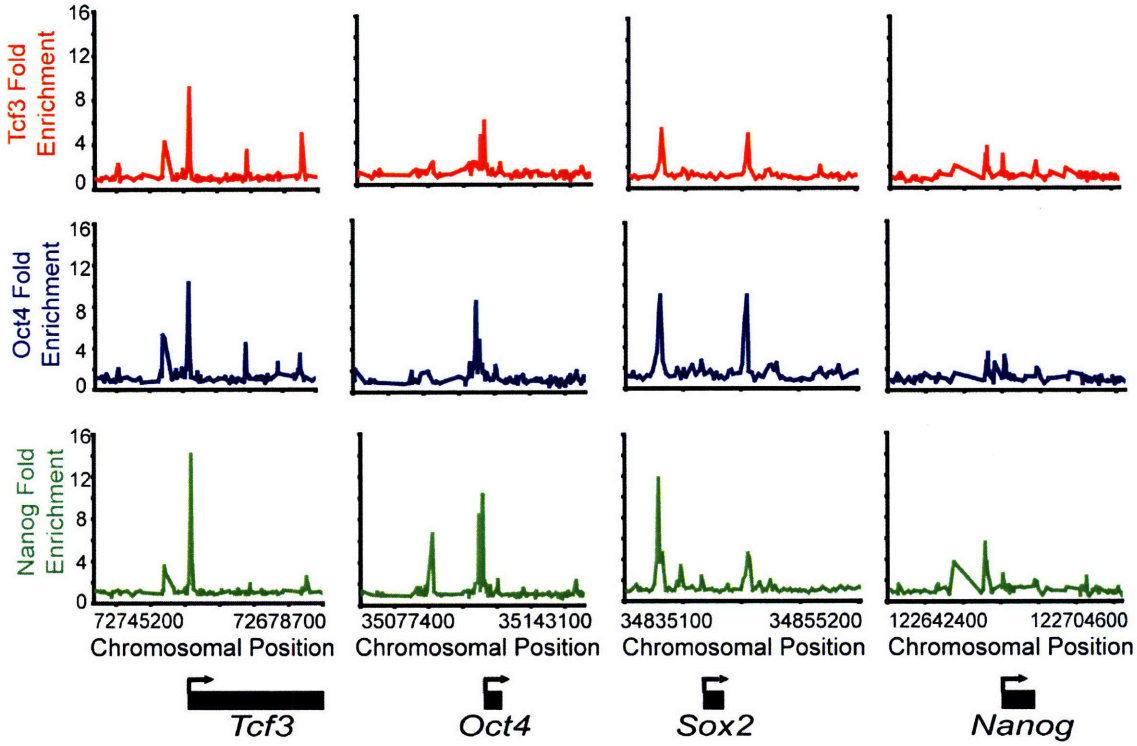

Tcf3 Fold Enrichment
Oct4 Fold Enrichment
Nanog Fold Enrichment
16
12
8
4
0
72745200
72678700
35077400
35143100
34835100
34855200
122642400
122704600
Chromosomal Position
Tcf3
Oct4
Sox2
Nanog

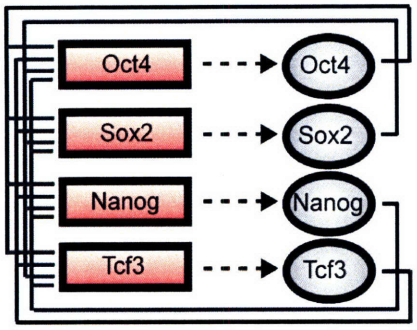
Figure 2

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.

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 (Appendix D Fig. S3), suggesting that this transcription factor regulates much of its own signaling pathway apparatus together with the pluripotency factors.

Figure 3

A.


Oct4
Oct4
Sox2
Sox2
Nanog
Nanog
Tcf3
Tcf3

B.

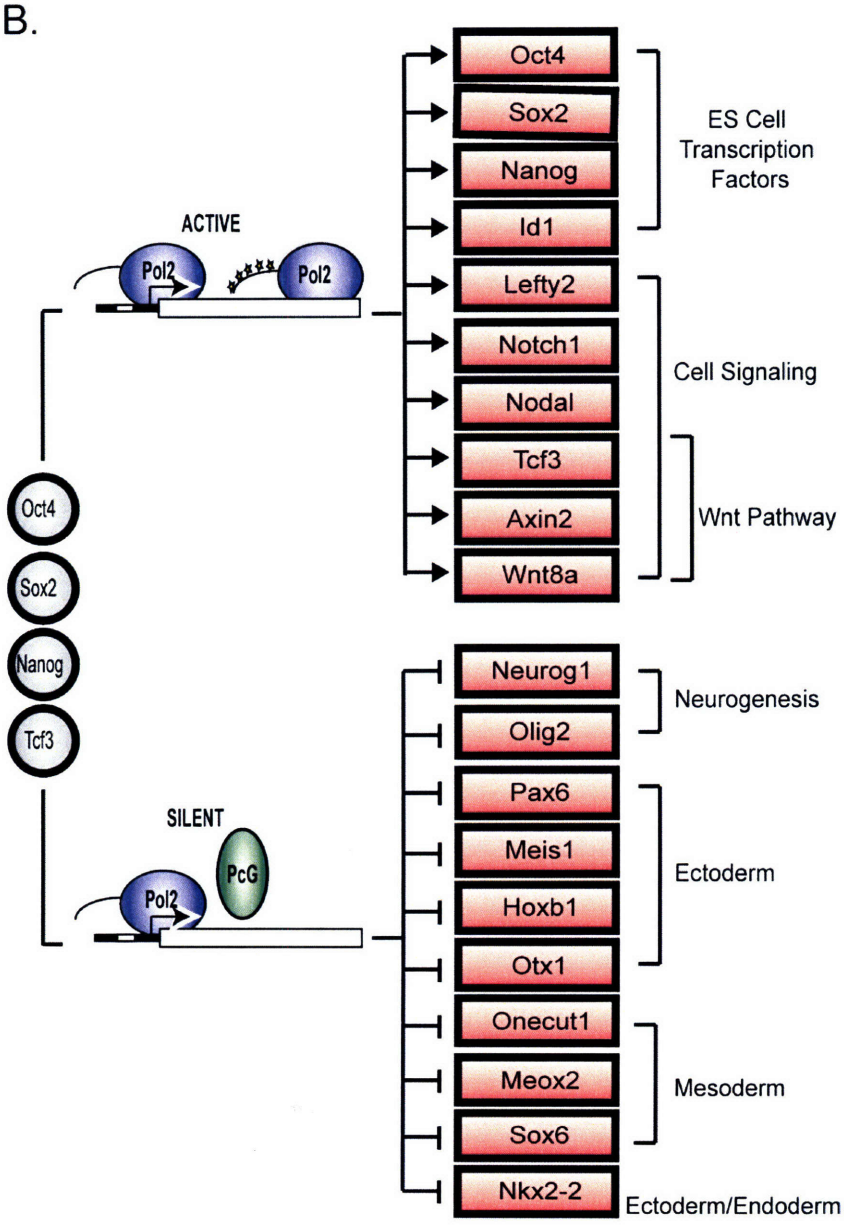

Oct4
Sox2
Nanog
Id1
ES Cell Transcription Factors
ACTIVE
Pol2
Pol2
Lefty2
Notch1
Nodal
Cell Signaling
Tcf3
Axin2
Wnt8a
Wnt Pathway
Oct4
Sox2
Nanog
Tcf3
Neurog1
Olig2
Neurogenesis
SILENT
PcG
Pol2
Pax6
Meis1
Hoxb1
Otx1
Ectoderm
Onecut1
Meox2
Sox6
Mesoderm
Nkx2-2
Ectoderm/Endoderm

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) (Appendix D 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 (Appendix D 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 $< 2 \times 10^{-10}$; Appendix D Fig. S6; Appendix D 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.

Figure 4

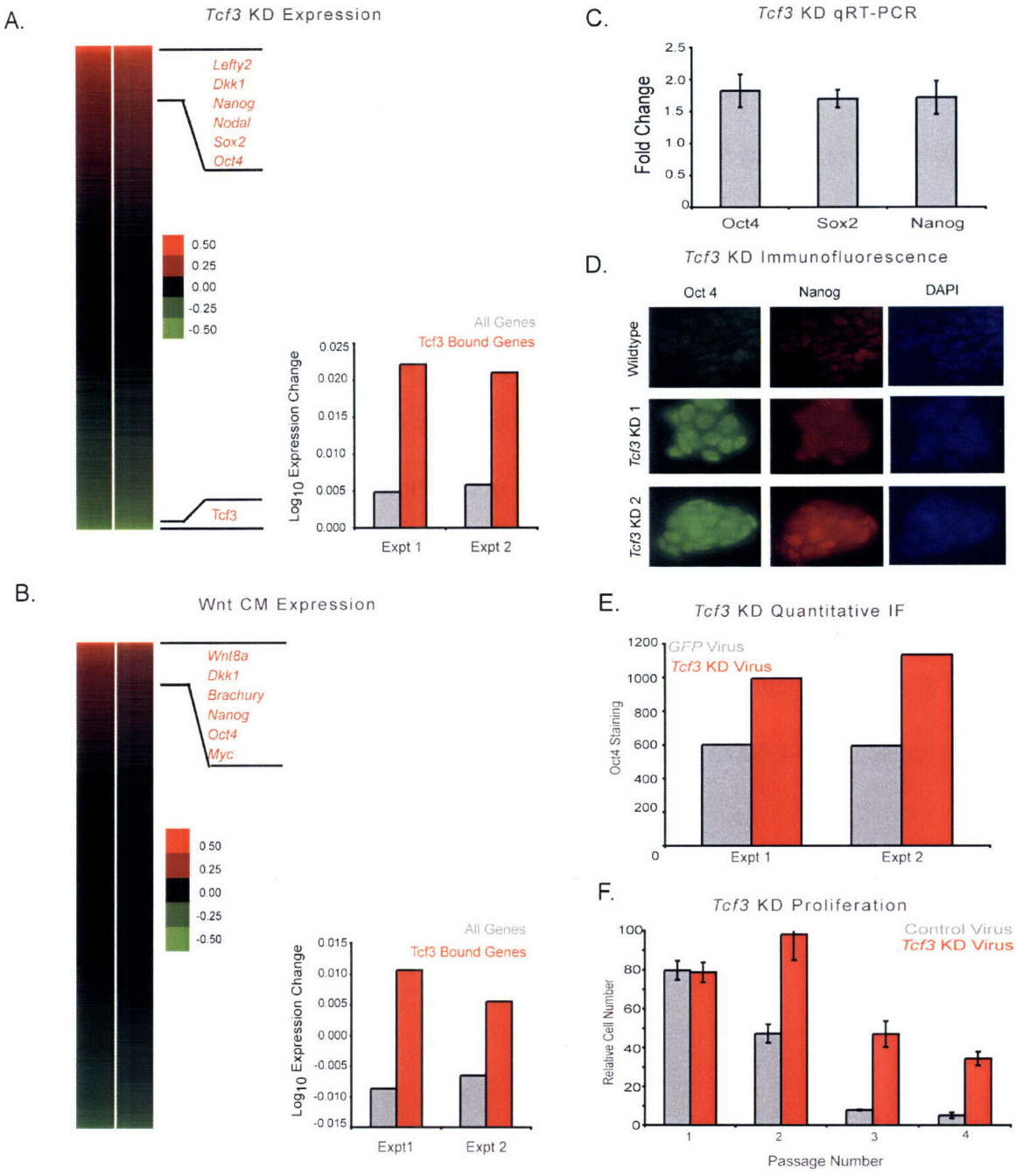

A.
Tcf3 KD Expression
Lefty2
Dkk1
Nanog
Nodal
Sox2
Oct4
Tcf3
0.50
0.25
0.00
-0.25
-0.50
All Genes
Tcf3 Bound Genes
Log10 Expression Change
0.025
0.020
0.015
0.010
0.005
0.000
Expt 1
Expt 2
B.
Wnt CM Expression
Wnt8a
Dkk1
Brachury
Nanog
Oct4
Myc
0.50
0.25
0.00
-0.25
-0.50
All Genes
Tcf3 Bound Genes
Log10 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

Figure 4

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.

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 b-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 (Appendix D 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 < 7×10^{-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}; Appendix D Fig. S8; Appendix D 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}; Appendix D 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; Appendix D 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 (Appendix D 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.

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

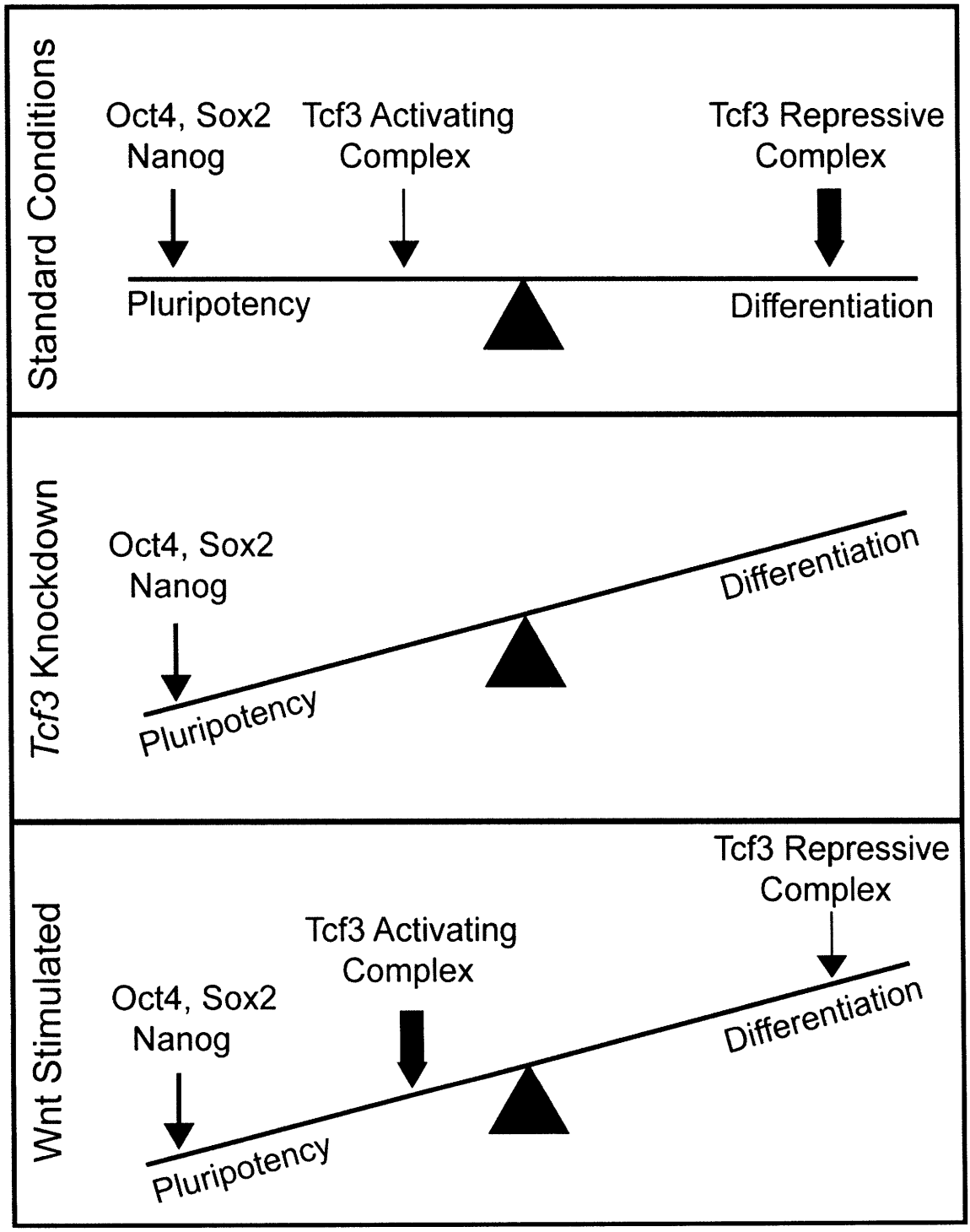

Standard Conditions
Oct4, Sox2
Nanog
Tcf3 Activating
Complex
Tcf3 Repressive
Complex
Pluripotency
Differentiation
Tcf3 Knockdown
Oct4, Sox2
Nanog
Differentiation
Pluripotency
Wnt Stimulated
Tcf3 Repressive
Complex
Tcf3 Activating
Complex
Oct4, Sox2
Nanog
Differentiation
Pluripotency

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)(Appendix D 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 (Appendix D 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.

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.

Acknowledgments

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 HA 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 JP, 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 HR (2005). "Regulatory networks in embryo-derived pluripotent stem cells." Nat Rev Mol Cell Biol **6:** 872-884.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al. (2005). "Core transcriptional regulatory circuitry in human embryonic stem cells." Cell **122:** 947-956.

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, 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 KM (2002). "Wnt signaling--20 years and counting." Trends Genet **18:** 340-342.

Cavallo RA, Cox RT, Moline MM, Roose J, Polevoy GA, 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 DL and Weis WI (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 AH (2007). "Signaling pathways in cancer and embryonic stem cells." Stem Cell Rev **3:** 7-17.

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

Guenther MG, Levine SS, Boyer LA, Jaenisch R and Young RA (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell **130:** 77-88.

Hao J, Li TG, Qi X, Zhao DF and Zhao GQ (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 DC, 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 TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B and Kinzler KW (1998). "Identification of c-MYC as a target of the APC pathway." Science **281:** 1509-1512.

Jho EH, Zhang T, Domon C, Joo CK, Freund JN 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 MF, Rindapaa M, Gaspar C, van Poppel N, Breukel C, van Leeuwen S, Taketo MM, 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 DM, Kalisz M and Nielsen JH (2005). "Cytokine signalling in embryonic stem cells." APMIS **113:** 756-772.

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, 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 RC, Gill JG, Kyba M, Murphy TL and Murphy KM (2006). "Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm." Development **133:** 3787-3796.

Logan CY and Nusse R (2004). "The Wnt signaling pathway in development and disease." Annu Rev Cell Dev Biol **20:** 781-810.

Loh YH, Wu Q, Chew JL, Vega VB, 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 HH and Arkin A (1997). "Stochastic mechanisms in gene expression." Proc Natl Acad Sci **94:** 814-819.

Merrill BJ, 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 BJ, Pasolli HA, Polak L, Rendl M, Garcia-Garcia MJ, Anderson KV 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 JL, 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 **104:** 5668-5673.

Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, Hinkle G, Piqani B, Eisenhaure TM, Luo B, Grenier JK, 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 DT, Dowell RD, Jacobsen ES, Nekludova L, Rolfe PA, Danford TW, Gifford DK, Fraenkel E, Bell GI and Young RA (2006). "Core transcriptional regulatory circuitry in human hepatocytes." Mol Syst Biol **2:** 2006-2017.

Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, Volkert TL, Schreiber J, Rolfe PA, Gifford DK, 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 JJ, Fu W, Kan L, Cuadra AE and Kessler JA (2004). "Beta-catenin signaling is required for neural differentiation of embryonic stem cells." Development **131:** 3545-3557.

Pan G and Thomson JA (2007). "Nanog and transcriptional networks in embryonic stem cell pluripotency." Cell Res **17:** 42-49.

Pera MF and Trounson AO (2004). "Human embryonic stem cells: prospects for development." Development **131:** 5515-5525.

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

Rajasekhar VK 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 BE, Pera MF, Fong CY, 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 DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH and Robson P (2005) "Transcriptional regulation of nanog by OCT4 and SOX2." J Biol Chem **280:** 24731-24737.

Roël G, Hamilton FS, Gent Y, Bain AA, 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 MB 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 AH (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 SS, Milo R, Mangan S and Alon U (2002). "Network motifs in the transcriptional regulation network of Escherichia coli." Nat Genet **31:** 64-68.

Singla DK, Schneider DJ, LeWinter MM and Sobel BE (2006). "wnt3a but not wnt11 supports self-renewal of embryonic stem cells." Biochem Biophys Res Commun **345:** 789-795.

Stock JK, 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 Re Commun **353:** 699-705.

Thieffry D, Salgado H, Huerta AM 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 JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ. Marshall VS and Jones JM (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 FH, Park K and Atchison ML (2006). "Polycomb recruitment to DNA in vivo by the YY1 REPO domain." Proc Natl Acad Sci **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 AB, Guo L, Sakamoto D, Caothien RH, Fuller JH, 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 **98:** 14973-14978.

Yi F and Merrill BJ (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 JW, Nechaev S, Adelman K, Levine M and Young RA (2007). "RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo." Nat Genet **39:** 1512-1516.

Chapter 5

Conclusions

A comprehensive understanding of the transcriptional regulatory circuitry that governs ES cells is of great value. It can enable these cells to be modulated to undergo directed differentiation and permit differentiated cells to be reprogrammed towards a pluripotent state. Two outstanding issues remain in understanding the transcriptional regulatory circuitry of ES cells. First, a number of factors and signaling pathways are known to have genetic importance for ES cells and should be integrated into this circuitry. Secondly, it is important to understand the nature of the connections between components within this circuitry and the combinatorial aspects of these factors. This latter effort will require integrating theories of circuitry into our understanding of the model, as well as defining new tests to evaluate the circuitry itself. I will briefly review additional components that should be integrated into the circuitry and then discuss how to approach understanding the combinatorial aspects of the network. Finally, I want to describe how the ES cell community is uniquely poised to revolutionize regenerative medicine through stem-cell based therapies given the understanding of pluripotency and differentiation that has evolved in recent years.

Integrating into the transcriptional regulatory circuitry the additional transcription factors and signaling pathways that impact ES cell pluripotency is important for understanding the full repertoire of ES cell components that contribute to the balance between maintaining pluripotency and initiating differentiation. The set of factors that should be integrated into the transcriptional regulatory circuitry were described in chapter 1 and include FoxD3, Zic3, Pem, c-Myc, Klf4, Gata6 and Cdx2 (Boiani and Scholer, 2005; Fan et al., 1999; Hanna et al., 2002; Li et al., 2005; Lim et al., 2007; Nakatake et al., 2006; Takahashi and Yamanaka, 2006). Signaling pathways that can impact ES cell

identity include LIF/gp130, BMP4, Activin/Nodal, PI3K and ERK (Okita and Yamanaka, 2006; Valdimarsdottir and Mummery, 2005). Additionally, the Notch signaling pathway plays a key role in differentiation of ES cells (Fox et al., 2008; Lowell et al., 2006; Nemir et al., 2006; Noggle et al., 2006; Schmitt et al., 2004; Schroeder et al., 2006). Since transcriptional regulatory circuitry governs cell state and ES cells are poised to differentiate, it will be important to understand how pro-differentiation factors and signaling pathways impact the core regulatory circuitry. For example, Tcf3 is a factor that is crucial for maintaining the balance between pluripotency and differentiation in ES cells since it functions to repress pluripotency regulators in ES cells (Cole et al., 2008). It is important to have the ES cell poised to differentiate, and pro-differentiation factors and pathways must be active in ES cells in order to permit differentiation. By integrating the entire set of factors and pathways that regulate ES cells into the transcriptional regulatory circuitry, we will be able to better understand how the ES cell is poised for differentiation. Perhaps more importantly, understanding how factors that promote differentiation modify the transcriptional regulatory circuitry may permit the manipulation of ES cells down specific lineages, a necessary step for the therapeutic potential of these cells to be fully developed.

In addition to the transcription factors and signaling pathways that impact ES cell identity, miRNAs have also been shown to play a key role in governing ES cell identity and are known targets of the pluripotency regulators (Boyer et al., 2005; Ivey et al., 2008; Li et al., 2008). Mice deficient in Dicer, a component necessary for miRNA processing, die at E7.5 and are incapable of producing ES cells (Bernstein et al., 2003; Li and Gregory, 2008). ES cells deficient in Dicer are incapable of differentiating

(Kanellopoulou et al., 2005). These studies suggest that miRNAs play a critical role in ES cell identity. Tissue-specific miRNAs can be isolated that confer lineage specificity during differentiation of ES cells. This has been shown for miR-1 and miR-133, which confer muscle-specific differentiation of ES cells (Ivey et al., 2008). Efforts to characterize the miRNA profile of human and mouse ES cells have revealed the bulk of miRNAs present in an ES cell are encoded on genomic miRNA clusters (Calabrese et al., 2007; Suh et al., 2004). Since miRNAs are predicted to fine-tune gene expression in cells, it is likely that these miRNAs impact the transcriptional regulatory circuitry that balances ES cells between pluripotency and differentiation. Further analysis of the miRNAs that are active in ES cells and upregulated in differentiated cell types will reveal the subsets that contribute to different cell states. It will also be important to integrate the miRNA targets into the transcriptional regulatory circuitry to understand how they impact ES cell state.

While many of the factors that may be important for ES cell identity are already characterized, it is likely that there are many unidentified factors and signaling pathways that impact ES cells. Differential gene expression can be used to identify factors that are upregulated in ES cells relative to other tissues that are likely to play a key role in ES cell identity. Genetic screens can also be conducted in order to further identify factors contributing to pluripotency and differentiation. It has been reported that screening factors implicated in ES cells by loss-of-function techniques using shRNAs can identify previously uncharacterized factors important for ES cell identity (Ivanova et al., 2006). Gain-of-function studies have also been used to identify factors whose overexpression is capable of maintaining ES cell self-renewal and pluripotency (Pritsker et al., 2006).

Additionally, it is possible to screen miRNAs for their role in ES cells. Finally, the identification of small molecules that can activate or inhibit signaling pathways will further reveal the role of signaling pathways on ES cells (Schugar et al., 2008; Schuldiner et al., 2000). These types of studies will be important for revealing the complete set of factors and pathways involved in ES cell identity and identifying an additional set of components that can be integrated into the transcriptional regulatory circuitry.

In addition to understanding the genomic targets of transcription factors, signaling pathways and miRNAs, it will be important to understand their combinatorial behavior. It is well characterized that POU and SOX domain proteins such as Oct4 and Sox2 interact to regulate gene expression (Reményi et al., 2004). Studies of transcriptional activation at promoters over the last two decades have revealed that many factors assemble at the transcriptional start site in order to regulate gene expression. New approaches to understand how multiple factors interact to govern regulation of gene expression will need to be developed to interpret the transcriptional outcome at target genes. As the transcriptional regulatory circuitry becomes further elucidated, it can be coupled to gene expression analyses in order to predict the combinatorial effect of different sets of factors at their target promoters.

Clarifying the role of signaling pathways on the transcriptional regulatory circuitry of ES cells will need to be particularly sensitive to combinatorial effects. The view that signaling pathways act in parallel to process information is becoming increasingly contested. It is clear, for example, that Wnt and Notch interact to regulate expression of target genes, and crosstalk is reported between Wnt and BMP4 (Hayward et al., 2008; Okita and Yamanaka, 2006). As the effect of different signaling pathways are

explored, it will be important to consider these combinatorial effects. Cell culture conditions will need to be improved so that all factors present in culture are known. Serum and feeder-dependent cultures will need to be replaced with medium that includes the necessary growth factors. Moreover, mapping of the downstream target genes for different signaling pathways to identify targets shared between different pathways is important. This will also allow signaling pathways to be integrated into the transcriptional regulatory circuitry.

A better understanding of how signaling pathways connect to the ES cell genome and impact pluripotency and self-renewal will be useful for developing tools to differentiate ES cells down specific developmental lineages. For example, it has been demonstrated that under certain conditions, Notch signaling promotes neural development of ES cells (Lowell et al., 2006). The identification of which signaling pathways promote which developmental pathways will be crucial for directing lineage differentiation and establish tools for promoting differentiation of ES cells for therapeutic purposes. Furthermore, signaling pathways may be useful for reprogramming of somatic cells. It may be possible to treat differentiated somatic cells with a set of signaling ligands to promote reprogramming towards a stem cell state. Since it has been shown that Wnt signaling leads to the upregulation of Oct4 and Nanog, Wnt may be a candidate factor for reprogramming somatic cells through the manipulation of signal transduction pathways. Screening combinations of ligands to identify a set that can reprogram ES cells will be important for exploring the possibility of reprogramming through signaling pathways. These external ligands could then substitute for genetic manipulation that is currently the best method for reprogramming (Yamanaka, 2007).

In order to direct differentiation of ES cells down specific lineages, it will be essential to upregulate the set of lineage specific developmental regulators. The repression of developmental regulators is a key mechanism that ES cells use to maintain pluripotency (Boyer et al., 2005; Lee et al., 2006). An outstanding question that pertains to the manipulation of ES cell fate is how these cells are able to upregulate a subset of the repressed developmental regulators to initiate development down a particular lineage. It is known that RNA polymerase II is paused at these developmental regulators, indicating that they are poised to be transcribed, and that this poised state is controlled by Polycomb Repressor Complex-mediated ubiquitination of H2A (Guenther et al., 2007; Stock et al., 2007). It will be essential to identify the signals that connect to particular subsets of developmental regulators to release poised polymerase and initiate gene expression.

Since the isolation of ES cells over 25 years ago, much has been learned about the molecular mechanisms that govern pluripotency and differentiation. ES cells have served as an ideal candidate system for the study of how transcriptional regulation governs cell state. The latter effort has revealed the mechanisms that the cells use to maintain the balance between pluripotency and differentiation. As more and more components are integrated into this circuitry, it will further elucidate the mechanisms the cell uses to achieve this balance. Efforts to understand how particular network motifs contribute to this balance, such as autoregulatory and feedforward loops, will reveal the switches that cells use to signal the cell to differentiate. Ideally, as the components of the circuitry become clearer, it will also become evident what modifications need to be made to the cell to promote specific differentiation program. Efforts to understand ES cells have allowed their fate to be directed and for differentiated cells to be directed towards a

pluripotent fate. These remarkable achievements are the results of tremendous effort in the research community to understand ES cells and are inspired by the hope that these cells will revolutionize regenerative medicine and aid the treatment of disease during the 21st century and beyond.

References

Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, Mills AA, Elledge SJ, Anderson KV and Hannon GJ (2003). "Dicer is essential for mouse development." Nat Genet **35**(3): 215-7.

Boiani M and Scholer H (2005). "Regulatory networks in embryo-derived pluripotent stem cells." Nat Rev Mol Cell Bio **6**: 872-884.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R and Young RA (2005). "Core transcriptional regulatory circuitry in human embryonic stem cells." Cell **122**(6): 947-56.

Calabrese JM, Seila AC, Yeo GW and Sharp PA (2007). "RNA sequence analysis defines Dicer's role in mouse embryonic stem cells." Proc Natl Acad Sci **104**(46): 18097-102.

Cole MF, Johnstone SE, Newman JJ, Kagey MH and Young RA (2008). "Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells." Genes Dev **22**(6): 746-55.

Fan Y, Melhem MF and Chaillet JR (1999). "Forced expression of the homeobox-containing gene Pem blocks differentiation of embryonic stem cells." Dev Biol **210**(2): 481-96.

Fox V, Gokhale PJ, Walsh JR, Matin M, Jones M and Andrews PW (2008). "Cell-cell signaling through NOTCH regulates human embryonic stem cell proliferation." Stem Cells **26**(3): 715-23.

Guenther MG, Levine SS, Boyer LA, Jaenisch R and Young RA (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell **130**(1): 77-88.

Hanna LA, Foreman RK, Tarasenko IA, Kessler DS and Labosky PA (2002). "Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo." Genes Dev **16**(20): 2650-61.

Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM and Rajewsky K (2005). "Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing." Genes Dev **19**(4): 489-501.

Hayward P, Kalmar T and Arias AM (2008). "Wnt/Notch signaling and information processing during development." Development **135**(3): 411-24.

Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y and Lemischka IR (2006). "Dissecting self-renewal in stem cells with RNA interference." Nature **442**(7102): 533-8.

Ivey KN, Muth A, Arnold J, King FW, Yeh RF, Fish JE, Hsiao EC, Schwartz RJ, Conklin BR, Bernstein HS and Srivastava D (2008). "MicroRNA regulation of cell lineages in mouse and human embryonic stem cells." Cell Stem Cell **2**(3): 219-29.

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R and Young RA (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-13.

Li Y, McClintick J, Zhong L, Edenberg HJ, Yoder MC and Chan RJ (2005). "Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4." Blood **105**(2): 635-7.

Li Q and Gregory RI (2008). "MicroRNA regulation of stem cell fate." Cell Stem Cell **2**(3): 195-6.

Lim LS, Loh YH, Zhang W, Li Y, Chen X, Wang Y, Bakre M, Ng HH and Stanton LW (2007). "Zic3 is required for maintenance of pluripotency in embryonic stem cells." Mol Biol Cell **18**(4): 1348-58.

Lowell S, Benchoua A, Heavey B and Smith AG (2006). "Notch Promotes Neural Lineage Entry by Pluripotent Embryonic Stem Cells." PLoS Biology **4**: 805-818.

Nakatake Y, Fukui N, Iwamatsu Y, Masui S, Takahashi K, Yagi R, Yagi K, Miyazaki J, Matoba R, Ko MS and Niwa H (2006). "Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells." Mol Cell Biol **26**(20): 7772-82.

Nemir M, Croquelois A, Pedrazzini T and Radtke F (2006). "Induction of Cardiogenesis in Embryonic Stem Cells via Downregulation of Notch1 Signaling." Circ Res **98**: 1471-1478.

Noggle SA, Weiler D and Condie BG (2006). "Notch Signaling Is Inactive but Inducible in Human Embryonic Stem Cells." Stem Cells **24**: 1646-1653.

Okita K and Yamanaka S (2006). "Intracellular signaling pathways regulating pluripotency of embryonic stem cells." Curr Stem Cell Res Ther **1**(1): 103-11.

Pritsker M, Ford NR, Jenq HT and Lemischka IR (2006). "Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells." Proc Natl Acad Sci **103**(18): 6946-51.

Reményi A, Schöler HR and Wilmanns M (2004). "Combinatorial control of gene expression. Nat Struct Mol Biol **11**(9): 812-5.

Schmitt TM, de Pooter RF, Gronski MA, Cho SK, Ohashi PS and Zúñiga-Pflücker JC (2004). "Induction of T cell development and establishment of T cell competence from embryonic stem cells differentiated in vitro." Nat Immunol **5**(4): 410-7.

Schroeder T, Meier-Stiegen F, Schwanbeck R, Eilken H, Nishikawa S, Häsler R, Schreiber S, Bornkamm GW, Nishikawa S and Just U (2006). Activated Notch1 alters differentiation of embryonic stem cells into mesodermal cell lineages at multiple stages of development. Mech Dev **123**: 570-579.

Schugar RC, Robbins PD and Deasy BM (2008). "Small molecules in stem cell self-renewal and differentiation." Gene Ther **15**(2): 126-35.

Schuldiner M, Yanuka O, Itskovitz-Eldor J, Melton DA and Benvenisty N (2000). "Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells." Proc Natl Acad Sci **97**(21): 11307-12.

Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M, Koseki H, Brockdorff N, Fisher AG 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**(12): 1428-35.

Suh MR, Lee Y, Kim JY, Kim SK, Moon SH, Lee JY, Cha KY, Chung HM, Yoon HS, Moon SY, Kim VN and Kim KS (2004). "Human embryonic stem cells express a unique set of microRNAs." Dev Biol **270**(2): 488-98.

Takahashi K and Yamanaka S (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-76.

Valdimarsdottir G and Mummery C (2005). "Functions of the TGFbeta superfamily in human embryonic stem cells." APMIS **113**(11-12): 773-89.

Yamanaka S (2007). "Strategies and New Developments in the Generation of Patient-Specific Pluripotent Stem Cells." Cell Stem Cell **1**: 39-49.

Appendix A

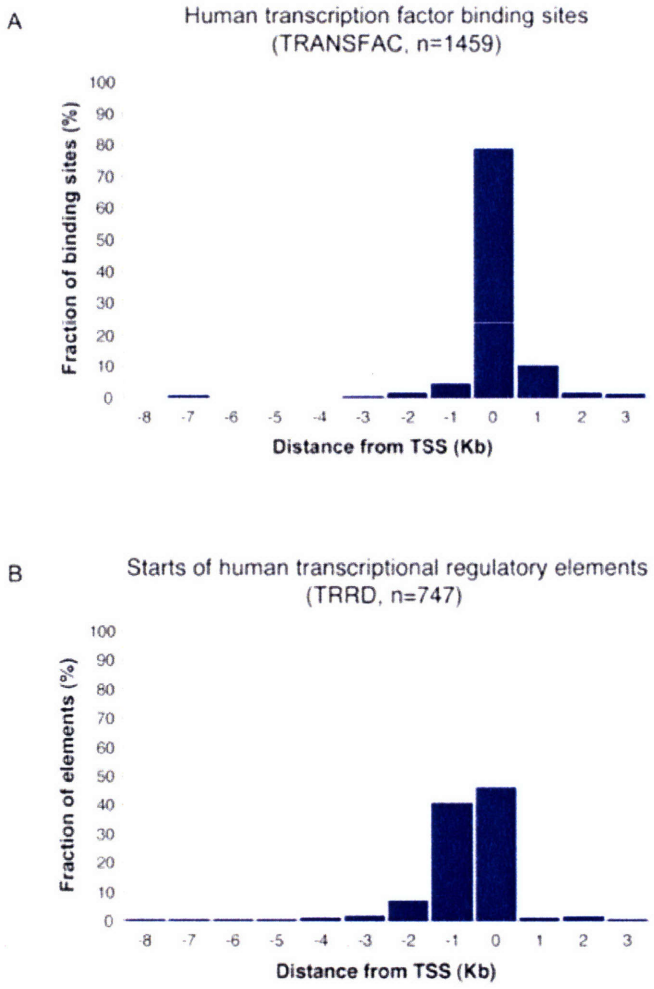
Supplementary Data for Chapter 2

Supplementary Data and Tables

All supplementary data and Tables S1-S7 can be found at the following website:

http://www.cell.com/cgi/content/full/122/6/947/DC1/

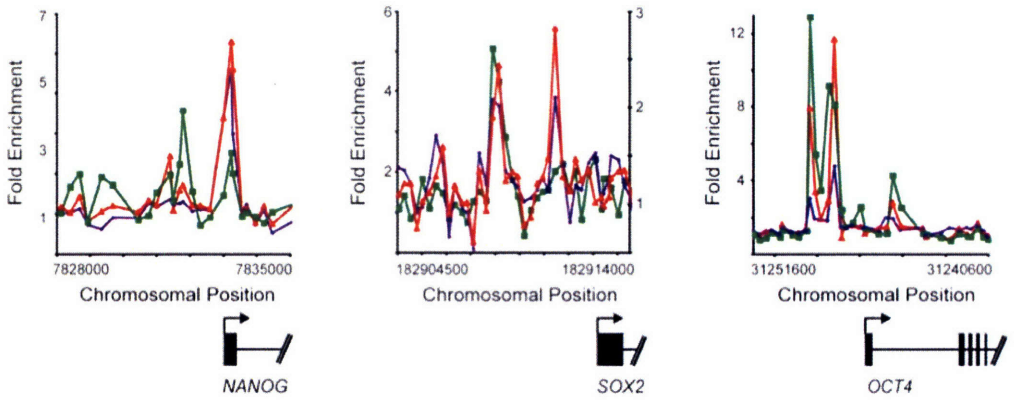
Figure S1. Distribution of Transcription Factor Binding Sites and Transcriptional Regulatory Elements Relative to Transcription Start Sites



A. Distribution of transcription factor binding sites from TRANSFAC from –8kb to +3kb around the transcription start site.

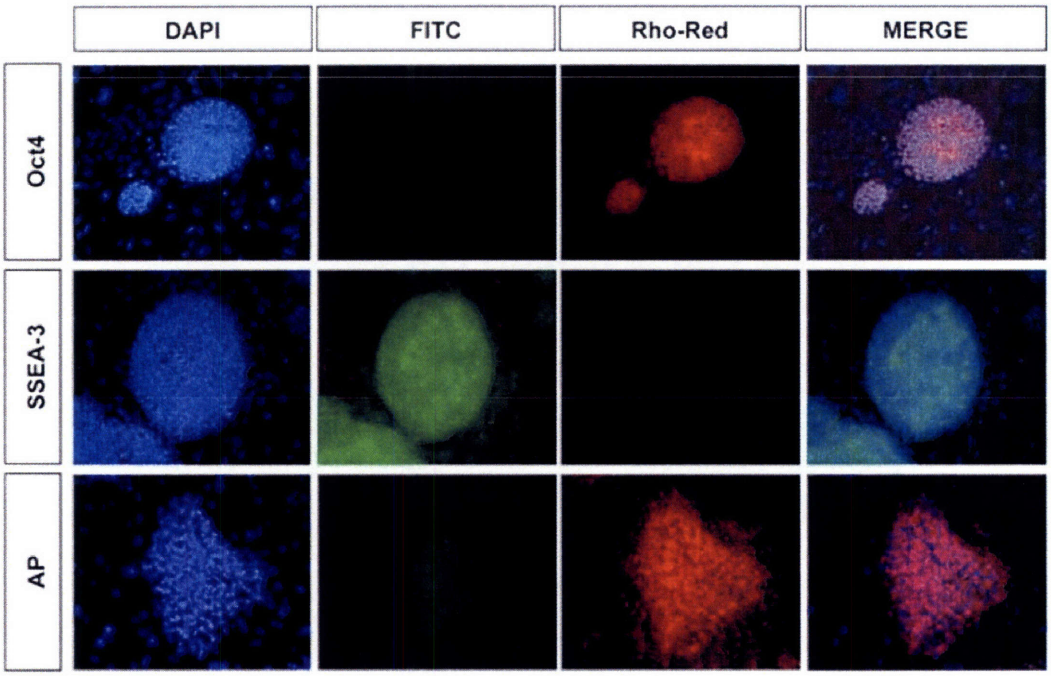
B. Distribution of functional regulatory elements from the TRRD (database of transcriptional regulatory regions, http://www.bionet.nsc.ru/trrd/34/) from –8kb to +3kb around the transcription start site.

Figure S2. Oct4, Sox2, and Nanog Cooccupy Each of Their Promoters



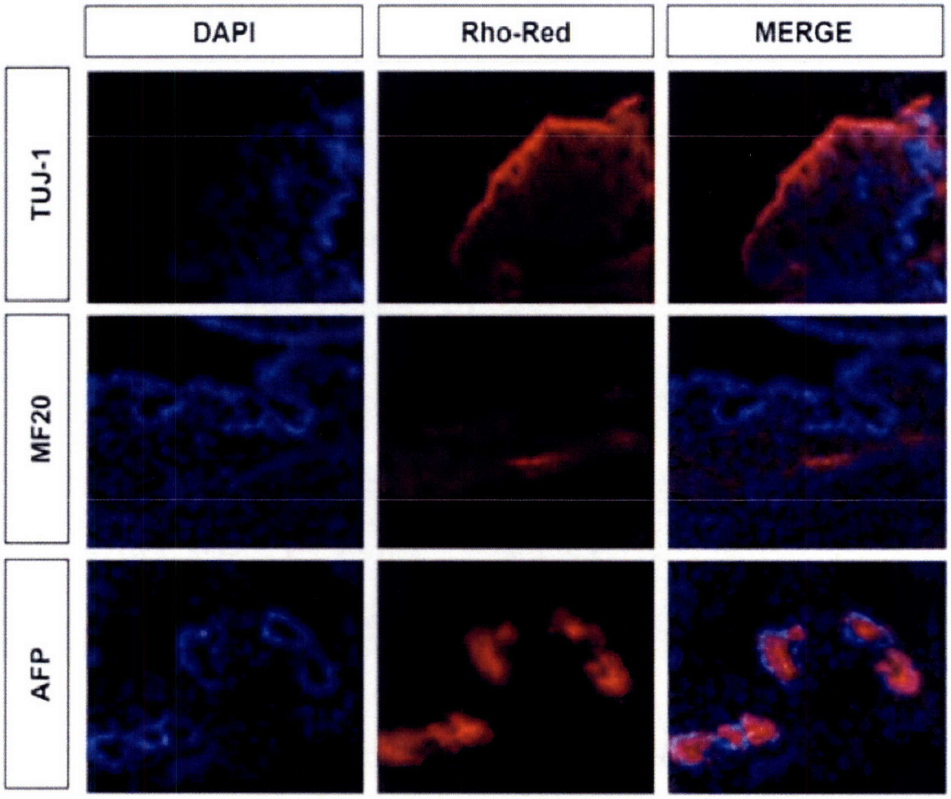
Plots display unprocessed ChIP enrichment ratios for all probes within a genomic region. Genes are shown to scale relative to their chromosomal position. Exons and introns are represented by thick vertical and horizontal lines, respectively. The start and direction of transcription are denoted by arrows. Green, red, and purple lines represent Nanog, Sox2, and Oct4 bound regions, respectively.

Figure S3. Immunuohistochemical Analysis of Pluripotency Markers



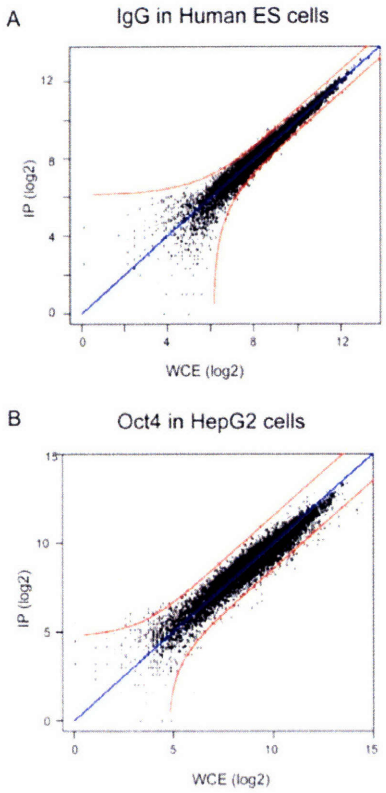
Human ES cells were analyzed by immunohistochemistry for the characteristic pluripotency markers Oct4 and SSEA-3. For reference, nuclei were stained with DAPI. Our analysis indicated that >>80% of the colonies were positive for Oct4 and SSEA-3. Alkaline phosphatase activity was also strongly detected in hES cells.

Figure S4. H9 Cells Maintain Differentiation Potential in Teratoma Assay



Teratomas were analyzed for the presence of markers for ectoderm (Tuj1), mesoderm (MF20) and endoderm (AFP). For reference, nuclei are stained with DAPI. Antibody reactivity was detected for derivatives of all three germ layers confirming that the human embryonic stem cells used in our analysis have maintained differentiation potential.

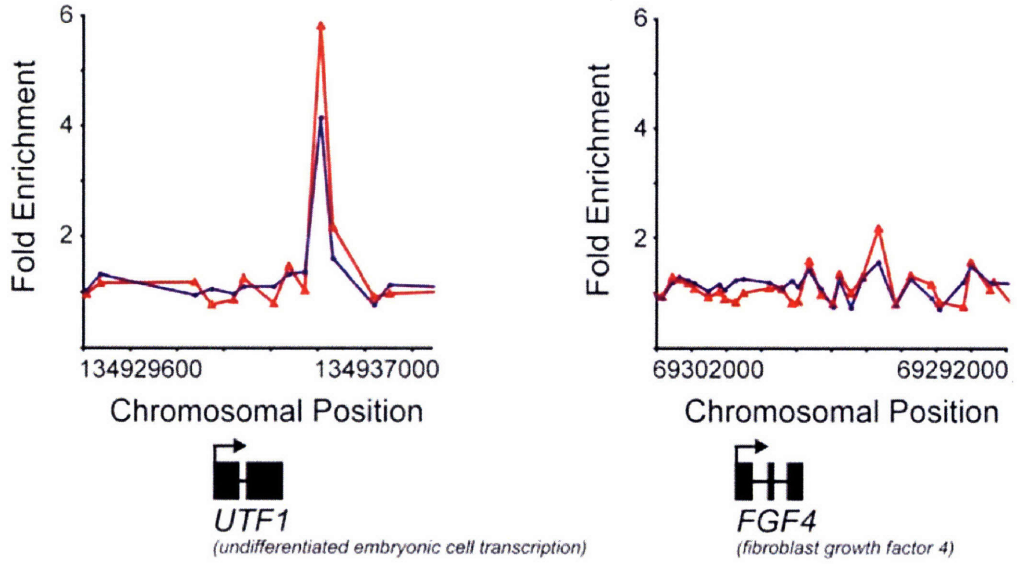
Figure S5. Control Chromatin Immunoprecipitations



A. Oct4, Sox2, and Nanog targets were not enriched using preimmune sera in human ES cells. ChIP was carried out using rabbit or goat IgG to assess antibody specificity. Labeled IP material and control DNA were hybridized to self-printed promoter arrays. Background subtracted normalized log2 intensities are plotted. Red lines represent enrichment / exclusion p-values of <10-3. Example shown is for the goat IgG control experiment.

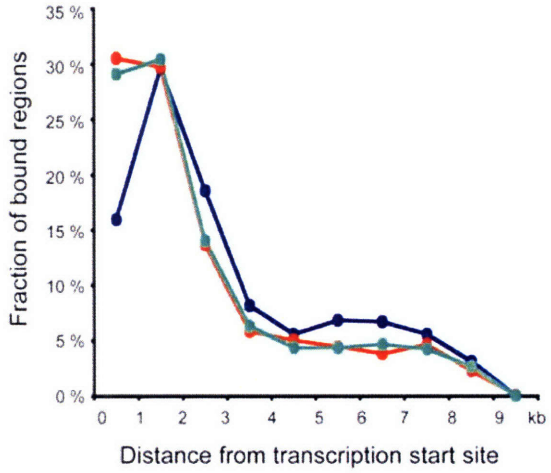
B. Potential antibody cross-reactivity with other family members was assessed by performing ChIP experiments in HepG2 cells. Data were analyzed as above. Example shown for Oct4 (sc-9081) in HepG2.

Figure S6. Oct4 and Sox2 Binding to UTF1 and FGF4



Plots display unprocessed ChIP enrichment ratios for all probes within a genomic region. Genes are shown to scale relative to their chromosomal position. Exons and introns are represented by thick vertical and horizontal lines, respectively. The start and direction of transcription are denoted by arrows. Green, red, and purple lines represent Nanog, Sox2, and Oct4 bound regions, respectively.

Figure S7. Distribution of Oct4, Sox2, and Nanog Bound Regions Relative to Transciption Start Sites



Histogram of the distance between transcription factor bound regions and the nearest transcription start site. Green, red, and purple lines represent Nanog, Sox2, and Oct4 bound regions, respectively. A distance of 0 refers to bound regions that overlap a transcription start site.

Appendix B

Chromatin Immunoprecipitation and Microarray-based Analysis of Protein Location

Published as: Lee TI, Johnstone SE and Young RA (2006). "Chromatin immunoprecipitation and microarray-based analysis of protein location." Nat Protoc **1**(2): 729-48.

Chromatin immunoprecipitation and microarray-based analysis of protein location

Tong Ihn Lee[1], Sarah E Johnstone[1,2] & Richard A Young[1,2]

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. Correspondence should be addressed to R.A.Y. (young@wi.mit.edu)

Published online 13 July 2006; doi:10.1038/nprot2006.98

Genome-wide location analysis, also known as ChIP-Chip, combines chromatin immunoprecipitation and DNA microarray analysis to identify protein-DNA interactions that occur in living cells. Protein-DNA interactions are captured *in vivo* by chemical crosslinking. Cell lysis, DNA fragmentation and immunoaffinity purification of the desired protein will co-purify DNA fragments that are associated with that protein. The enriched DNA population is then labeled, combined with a differentially labeled reference sample and applied to DNA microarrays to detect enriched signals. Various computational and bioinformatic approaches are then applied to normalize the enriched and reference channels, to connect signals to the portions of the genome that are represented on the DNA microarrays, to provide confidence metrics and to generate maps of protein-genome occupancy. Here, we describe the experimental protocols that we use from crosslinking of cells to hybridization of labeled material, together with insights into the aspects of these protocols that influence the results. These protocols require approximately 1 week to complete once sufficient numbers of cells have been obtained, and have been used to produce robust, high-quality ChIP-chip results in many different cell and tissue types.

INTRODUCTION

Chromatin immunoprecipitation (ChIP) has become a widely used method for studying how proteins interact with the genome. The ChIP technique identifies physical interactions between proteins and DNA that occur within living cells[1–3] and ChIP can provide a whole-genome view of protein-DNA interactions when combined with DNA microarray analysis[4–14]. Our goal here is to share our ChIP-Chip protocol, suggest useful quality-control tests, identify the most likely sources of trouble and suggest solutions to problems with experiments. The protocol presented here has been in use for several years and has been used by many different individuals using a broad range of different mammalian cell types and tissues.

Overview of the ChIP-Chip technique

In a typical experiment (**Figure 1**), protein-DNA interactions are first captured *in vivo* by treating cells with a small, reversible crosslinker that is capable of rapidly diffusing into cells. After cell lysis, the DNA is fragmented and protein-DNA complexes can be enriched by immunoaffinity capture of the desired protein. As the DNA is fragmented prior to the capture, the DNA fragments that are most proximal to the binding event are enriched. Often, an aliquot of the input DNA is saved prior to immunoprecipitation for use as a reference sample. The crosslinks for both the enriched and input DNA are then reversed, and the DNA is purified away from RNA and protein components. The enrichment relative to a reference sample can then be measured by one of several techniques, such as PCR, comparative sequencing or DNA microarray analysis. For most DNA microarray experiments, the enriched and reference samples must be amplified first to provide sufficient material. In this protocol, the ends of the DNA are blunted and ligated to small DNA linkers that are subsequently used in priming PCR. This ligation-mediated PCR (LM-PCR) results in expanded pools of the enriched and reference DNA that are then differentially labeled and hybridized to the microarrays. The arrays are washed and then scanned. Various algorithms are then used to normalize the enriched and reference channels, to calculate the degree of enrichment and confidence metrics for each feature, to incorporate the information gained from having multiple probes representing contiguous regions of the genome and to generate maps of protein-genome occupancy.

Applications of ChIP-Chip

The range of applications for ChIP-Chip is broad but still expanding. The method has recently been used to investigate DNA-binding transcription factors and the networks formed by these factors (reviewed in refs. 15–18), the distributions of chromatin-modifying machinery, histones, histone variants and histone modifications (reviewed in refs. 16, 19), factors involved in DNA replication, DNA repair and DNA methylation (reviewed in refs. 20–22), and the linkage between transcription and nuclear-export machinery[23–25]. Immunoprecipitation has been coupled with microarrays to identify interactions between proteins and RNA[26–30] and ChIP has recently been used to study RNA-binding proteins that become associated with the genome through their interactions with RNA[31]. We anticipate that many new biological insights will emerge from protein-DNA and protein-RNA interactions that are identified using ChIP.

Limitations of ChIP-Chip

There are several features of ChIP-Chip that currently limit its utility and make it a challenging technique[32–34]. The primary limitation is the quality of antibodies that are available for a protein of interest. The antibody, ideally, has high avidity and specificity for an epitope that is accessible in the crosslinked chromatin. Because this is not always the case, the investigator must be cautious about interpreting the lack of detection as reflecting the absence of an interaction. A second limitation is that a substantial number of cells (10^7–10^8) are generally needed to obtain a robust result. With current methods, using fewer cells can reduce signals, leading to an underestimation of protein-DNA interactions. High-quality, whole-genome

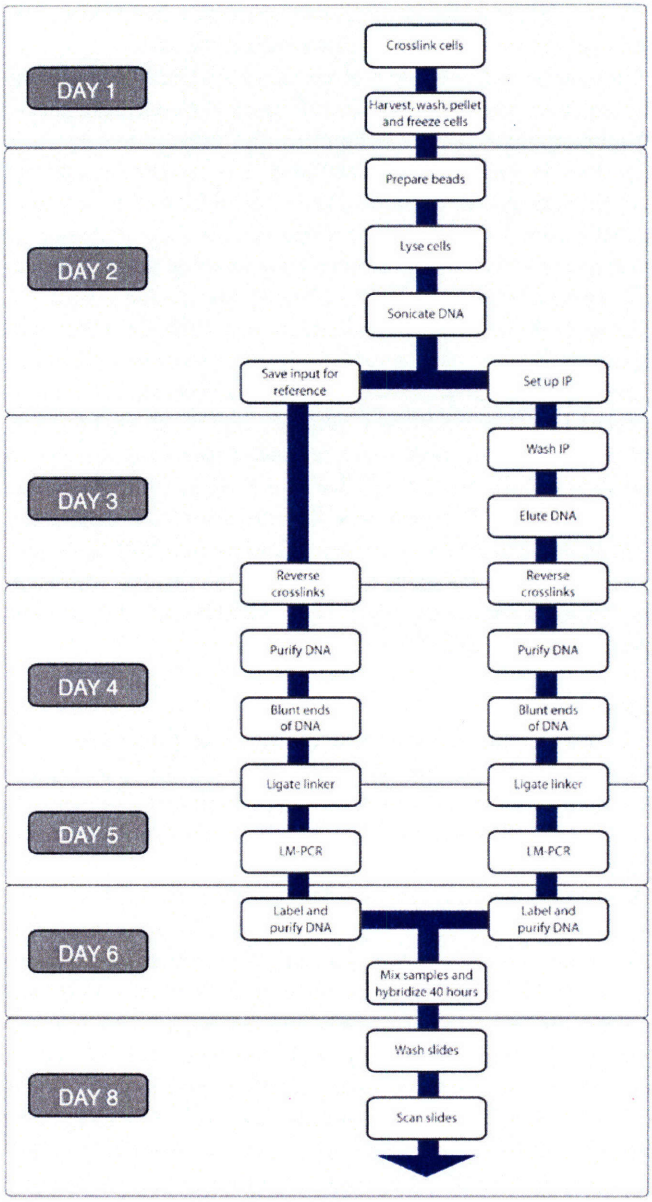


Figure 1 | A sample timeline for the ChIP-Chip protocol. Individual steps are shown in white boxes. Steps that are typically performed on the same day are grouped by day, which is indicated in gray boxes. IP, immunoprecipitation; LM-PCR, ligation-mediated PCR.

DNA microarray platforms are expensive. As array manufacturers continue to make improvements in designs and manufacturing, more cost-effective experiments will be possible. Finally, the experimental process is long and can be challenging to troubleshoot.

Experimental design

Background. The sections below describe several issues that we have found to be critical to consider in the early stages of planning and designing ChIP-Chip experiments.

Antibodies. ChIP-Chip is highly dependent on the antibody used for the immunoprecipitation. As individual antibodies can perform differently in the ChIP assay, it is helpful to screen a wide variety of available antibodies before launching a full-scale ChIP-Chip experiment. We generally start with antibodies, for which there is previous evidence for use in ChIP, or we use a variety of polyclonal and monoclonal antibodies from several sources.

The most reliable way to test the antibodies is to use them in small-scale ChIPs and to test for enrichment at a gene-specific level, although this approach depends entirely on the number of known binding sites, the strength of the evidence for the binding event and the comparability with your specific cell type, growth conditions and treatments. For small-scale ChIPs, we prepare material as described below, but instead of using all the material for a single reaction, we will split it into five to ten aliquots and use each aliquot in an immunoprecipitation. This lower amount of material seems adequate for gene-specific analysis, but does not usually work for microarray analysis. If no binding sites are known, it is much more difficult to screen for a useful antibody. Western blot analysis may be informative but is unreliable, as the antibody may recognize the denatured protein in a western blot but may still fail to recognize the native form *in vivo*.

If there are no antibodies for a specific protein, it is possible to use epitope-tagged versions of proteins, provided the tagged version does not radically affect protein function. We have used myc, TAP and FLAG epitope-tagged versions of various proteins with success.

Reference samples. Chromatin immunoprecipitation is an enrichment relative to a reference and not an absolute measurement, thus the choice of reference sample can be an important element of experimental design. Although it is most common to use unenriched, genomic DNA as the reference, it is sometimes more useful to hybridize an immunoprecipitation of one factor against an immunoprecipitation of a second factor to capture subtle effects. For instance, this approach can be useful when studying histone modifications. An immunoprecipitation against a core histone subunit is used as the reference to normalize the signal of the modification against the amount of histone. This can help to determine whether a perceived gain or loss of the modification is due to actual changes in modification or due to changes in histone density.

Replicates. ChIP-Chip, like many other forms of genome-wide analysis, requires the use of replicate samples to account for experimental noise that is inherent to high-throughput approaches and biological complexity due to the many factors influencing protein and nucleic-acid interactions. In particular, biological replicates (independently grown and treated pools of cells or tissue) or replicates with different antibodies against the same factor are valuable. The number of replicates needed depends on the noise that is inherent in the platform and the type of analysis to be performed, but triplicate experiments are preferred and duplicate experiments should be considered to be the minimum requirement whenever feasible.

Experimental controls. The use of a positive control and negative control for the experiment can be useful, both for troubleshooting and for calibrating the analysis. The positive control should be a high-quality antibody against a protein that is likely to be present in any cell type. For example, in mammalian cells, we regularly use antibodies against the ubiquitous cell-cycle regulator E2F4. The most common negative control is an immunoprecipitation using a non-specific antibody that matches the isotype of the experimental antibody.

Choice of arrays. There are a number of options available for array hybridization, including self-printed oligonucleotide or PCR amplicon arrays, and commercial arrays and services. The choice of array platform ultimately depends on the need to balance per-

formance, ease of use, resolution, cost of arrays and investment in equipment, and will vary from lab to lab. Due to the variety of options, a fully detailed protocol for our specific hybridization, washing, scanning and analysis steps is unlikely to be generally useful. However, we have included the relevant details for hybridization and washing and offer general guidelines that we try to follow for scanning and data analysis. This protocol was developed based on our work with both self-printed PCR amplicon arrays and commercially available oligonucleotide arrays, and reflects our most recent experience with oligonucleotide arrays[13,14,35,36]. A number of useful ChIP-Chip protocols have been published that describe protocols similar to this one, but which cover more specific aspects of dealing with different array platforms[37–44].

Normalization controls on the array. As ChIP measures enrichment in one sample compared with another, it is essential to define an appropriate baseline for a particular experiment — one that reflects the signal from the background of all sequences from the whole genome and reflects lack of enrichment in the immunoprecipitation. This requirement can affect the sets of controls that are needed on the array. When there are a small number of features expected to show enrichment, the enrichment ratios of the majority of features can serve to represent the baseline level; the large size of this latter set strengthens the statistical significance of the enrichment. In other situations, for example, when examining RNA polymerase II binding at promoter regions using arrays focused only on promoter regions, we found that it was more difficult to determine the baseline enrichment that represented 'no binding'. RNA polymerase II bound at a relatively high fraction of probes, making it more difficult to distinguish bound and unbound probes with as high a degree of statistical significance. In these cases, we found it useful to have a set of control probes designed against 'gene deserts', genomic regions that are devoid of known open reading frames and for which there is no evidence of noncoding transcripts. We assumed that these control probes would be generally unenriched for RNA polymerase II and could therefore serve as a normalization baseline. Similarly, we use control probes to subtract out a baseline level of non-specific hybridization signal and, for multislide arrays, control probes to normalize between slides.

Data analysis and bioinformatics. The ChIP-Chip analysis generates a large amount of data that requires some investment in statistical, computational and bioinformatics infrastructure.

Statistical analysis is required to account for experimental noise that is inherent to high-throughput approaches, but more importantly, to deal with the biological complexity of protein and nucleic-acid interactions. We expect that a protein of interest will be distributed across many different sites in a single cell in a manner that is dependent on protein concentration, its affinities for each site and other variables such as the presence of proteins that compete for similar sites. As a result, the output of ChIP-Chip is a distribution of enrichment ratios representing DNA or RNA occupancy, as opposed to two distinct sets of 'bound' and 'not bound'. Statistical methods are needed to identify the statistically significant set of enrichment ratios, to provide a measure of confidence at the level of individual probe features and to incorporate data from multiple features to build models for binding that are both sensitive and specific. As there are a wide variety of options for statistical analysis[7,12,45–51] and the needs of individual labs will vary, a fully detailed description of our analysis methods is unlikely to be generally useful. However, we have included a basic outline of the approach that we have used in the past.

The volume and complexity of the data from ChIP-Chip analysis may require access to resources that are capable of creating or adapting programming code to manipulate and analyze the data. Even if most of the early steps are handled by a commercially available analysis package, there is almost always a need for additional analyses that need to be customized in some way. The capacity to handle and display the data using common programming languages, such as Visual Basic, MATLAB, Perl, R, Java or Python, is highly useful.

Along these lines, the need to be able to map ChIP-Chip data back to genomic location and genomic features often requires the development of bioinformatics resources, such as access to genome annotation and transcription-factor databases, and the ability to handle and cross-reference genome-wide expression data. Commonly used databases include the UCSC Genome Browser (http://genome.ucsc.edu/); the NCBI databases (http://www.ncbi.nlm.nih.gov/), particularly RefSeq, Gene and GEO datasets; the EMBL-EBI databases (http://www.ebi.ac.uk/services/); GO functional annotation (http://www.geneontology.org/); and TRANSFAC (http://www.gene-regulation.com/pub/databases.html).

Protocol Tips

Background. The sections below describe a few key steps of the protocol for which we have some experience with variables that are particularly critical to the success of the experiment. The steps are presented in order, with those most likely to cause problems being presented first.

Sonication. Sonication is the most variable step in the process and will vary greatly depending on cell type, cell culture conditions used, quantity of cells, volume of sonication, degree of crosslinking and specifics of the sonicator being used. As a result, conditions must usually be optimized for each situation. Sonication fulfills two roles, solubilization of chromatin and shearing of chromatin, both of which are essential for successful ChIP analysis. Solubilization is achieved relatively easily. However, the degree of shearing requires more care. Undersonication results in a loss of resolution of binding events. Smaller DNA fragments allow for more precise localization of a specific binding event, as a smaller region of DNA will be pulled down in the immunoprecipitation. We have also found that oversonication must be avoided as it results in more noise in the microarray analysis and difficulty in identifying legitimate targets. As a guide, most protocols suggest sonication until reaching an average fragment size (different protocols suggest a value varying from 350 to 1,000 bp) or until the fragment sizes are within a certain range. Undersonication relative to these kinds of guidelines is easy to identify: the size of the reverse crosslinked, purified DNA fragments will be too large. Oversonication is more difficult to quantify. Most crosslinked material has a physical limit of sonication, and additional sonication past this point will not result in further visible shearing but will affect microarray results. Thus, it is possible that researchers trying to match previously published descriptions that were optimized for different conditions may inadvertently oversonicate their DNA. As an alternative, we suggest sonication using the lowest power and time settings that result in sheared DNA, of which at least a quarter of the total DNA is sized from 200 to 600 bp in size, with less emphasis on the average size or overall spread of fragment sizes. One way to identify these conditions is to run a small time course of sonications. We will set up a sample of cells, as

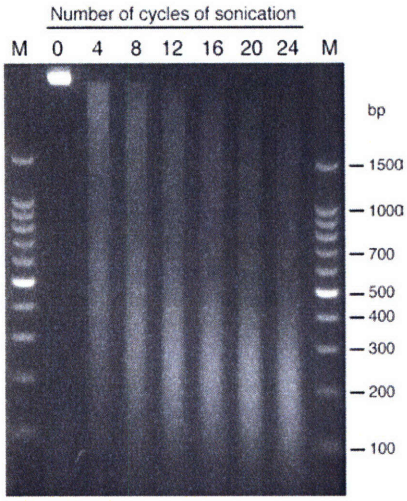


Figure 2 | Results of varying degrees of sonication on fragment size. A total of 2×10^8 crosslinked Jurkat cells were sonicated using the following conditions: Misonix 3000 sonicator with microtip; power 7; 24 cycles (30 s sonication, 90 s rest). Samples were removed at various times, crosslinks were reversed and DNA-purified, and run on a 2% agarose gel. Lanes with molecular weights are labeled M and lanes with sonicated material are labeled with the number of cycles of sonication. Sizes of molecular-mass markers are indicated. Twelve cycles of sonication provide a good degree of sonication. Four cycles of sonication results in undersonicated DNA. Note that sonication beyond 16 cycles results in little change in fragment size and is likely to result in oversonicated material.

described below, with a default power setting but with an extended time of sonication. We then remove small aliquots (100 µl out of 3.5 ml of material) at discrete time points. After crosslink reversal and purification, as described in the protocol below, the DNA can be run on a 2% agarose gel to estimate the degree of sonication. A sample sonication time course is shown in **Figure 2**.

Immunoprecipitation conditions. Antibody performance can be optimized by adjusting a number of variables, but we find that the simplest, most effective adjustment has been modifying the amount of salt in the immunoprecipitation. If gene-specific or microarray analysis indicates that antibody performance is poor (two-fold enrichment or less), but a positive control experiment indicates that the protocol is working well, it may be useful to perform a set of small-scale immunoprecipitations over a range of salt concentrations (typically starting with 0, 50, 100 and 250 mM sodium chloride). Prepare input DNA as described below but use only one-fifth to one-tenth of the material in an immunoprecipitation. This small-scale immunoprecipitation usually suffices for gene-specific analysis, but does not usually work for microarray analysis. If there is previous evidence indicating specific salt conditions that are optimal for a particular antibody, we will supplement the lysate with additional sodium chloride to approximate those conditions.

Washing. Following ChIP, it is possible to overwash the beads. This usually results in noise in the microarray analysis. In general, this occurs due to extensive wash times rather than too many washes. Five washes (as presented below) are adequate for most antibodies. If signal:noise ratios are low, increasing the number of washes (as high as 8) may help, although if you suspect low antibody affinity, a lower number of washes (as few as 3) may be more appropriate.

Crosslinking. In practice, we have found that the crosslinking times and formaldehyde concentrations provided in this protocol are generally applicable and have been used successfully with a variety of proteins in human, mouse, yeast, fruit fly and zebrafish model systems. But crosslinking times could theoretically be optimized for each protein. Insufficient crosslinking can result in the inability to capture protein-DNA contacts. Over-crosslinking can actually denature the protein of interest or cause over-aggregation, thus obscuring the epitope. Unfortunately, these problems are sometimes difficult to diagnose as the primary readout is the failure to detect enrichment, which could also result in problems at any number of other steps. One other sign would be unusual ease or difficulty in obtaining a range of sonicated material when testing a range of sonication conditions.

MATERIALS

REAGENTS

- 1 M HEPES-KOH, pKa 7.55 (Invitrogen/Gibco; Cat. no.: 15630-080)
- 5 M NaCl (Sigma S-5150)
- 500 mM EDTA (Invitrogen/Gibco; Cat. no: 15575-038)
- 500 mM EGTA (Sigma, Cat no: E3889)
- 37% formaldehyde (J.T. Baker; Cat. no: 2106-01) ! **CAUTION** Formaldehyde is flammable; highly toxic by inhalation, contact with skin or if swallowed; causes burns; and is potentially carcinogenic. Formaldehyde should be used with appropriate safety measures, such as protective gloves, glasses and clothing, and adequate ventilation. Formaldehyde waste should be disposed of according to regulations for hazardous waste.
- 2.5 M glycine (Sigma; Cat. no: G8790)
- 10× Dulbecco's phosphate buffered saline (PBS) (Invitrogen/Gibco; Cat. no.: 14200-075)
- Dynabeads (Dynal) ▲ **CRITICAL** Although it is possible to use other types of beads for the immunoprecipitation, we find that Dynabeads give us very low background and are easy to use. We use Dynabeads that come pre-coated with secondary antibody (e.g., PanMouse IgG, which is a monoclonal human anti-mouse IgG) or pre-coated with protein G. The exact type of bead used will depend on the primary antibody being used)
- Block Solution: 1× PBS, 0.5% bovine serum albumin (BSA) (Sigma; Cat. no: A7906) ▲ **CRITICAL** Should be made fresh and kept cold.
- Primary antibody of choice
- 25× solution Complete Protease Inhibitor Cocktail: 1 tablet dissolved in 2 ml of double distilled water (can store up to 12 weeks at –20 °C) (Roche; Cat. no.: 1 697 498)
- 50% glycerol (Sigma; Cat. no.: G5516)
- 10% Igepal CA-360 (Sigma; Cat. no.: I8896)
- 10% Triton X-100 (Sigma; Cat. no.: T8787)
- 1 M Tris-HCl, pH 8.0 (Invitrogen/Gibco; Cat. no.: 15568-025)
- 5% Na-Deoxycholate (Sigma; Cat. no.: D5760)
- 5% N-lauroylsarcosine (Fluka; Cat. no.: 61743) ▲ **CRITICAL** An ultrapure version of this reagent is not required for lysis steps, but is required for hybridizations. Lower grades can leave fluorescent residue on slides.
- 5 M LiCl (Sigma; Cat. no.: L7026)
- 10% SDS (Invitrogen/Gibco; Cat. no.: 15553-035)
- 10 mg ml^{-1} RNAseA (Sigma; Cat. no.: R6513)
- Proteinase K solution (Invitrogen; Cat. no.: 25530-049)
- Phenol:chloroform:isoamyl alcohol (Fluka; Cat. no.: 77617) ▲ **CRITICAL** If this solution is old or is at low pH, there will be degradation of DNA. ! **CAUTION** Phenol is toxic when in contact with skin or if swallowed; causes burns; and is irritating to eyes, the respiratory system and skin. Chloroform is harmful by inhalation or if swallowed; is irritating to skin; and is potentially carcinogenic. Isoamyl alcohol is flammable; and is irritating to eyes, the respiratory system and skin. The phenol:chloroform:isoamyl alcohol should be used with appropriate safety measures, such as protective gloves, glasses and clothing, and adequate ventilation.

The phenol:chloroform:isoamyl alcohol waste should be disposed of according to regulations for hazardous waste.

- 20 mg ml^{-1} glycogen (Roche; Cat. no.: 901 393)
- 100% ethanol (Aaper)
- 80% ethanol (diluted from 100% ethanol)
- T4 DNA polymerase, 3 U μl^{-1} (New England Biolabs; Cat. no.: M0203S)
- 10× NE Buffer 2 (supplied with T4 DNA Polymerase)
- 1 $\mu g\ \mu l^{-1}$ BSA (diluted from 10 mg ml^{-1} stock supplied with T4 DNA Polymerase)
- 2.5 mM dNTP mix (2.5 mM each dNTP) (GE Healthcare Life Sciences; Cat. no.: 27-2035-03)
- 3 M sodium acetate (Sigma; Cat. no.: S-7899)
- T4 DNA ligase, 400 U μl (New England Biolabs; Cat. no.: M0202S)
- 5× T4 DNA Ligase buffer (Invitrogen; Cat. no.: 46300-018)
- 40 μM solution oligo oJW102 (MWG Biotech; 5′-GCGGTGACCCGGGAGATCTGAATTC-3′)
- 40 μM solution oligo oJW103 (MWG Biotech; 5′-GAATTCAGATC-3′)
- 15 μM Linker (see REAGENT SETUP)
- 10× ThermoPol buffer (New England Biolabs; Cat. no.: B9004S)
- Double-distilled water
- AmpliTaq polymerase (Applied Biosystems)
- 7.5 M ammonium acetate (Sigma; Cat. no.: A2706)
- BioPrime Array CGH Genomic Labeling System (Invitrogen; Cat. no.: 18095-011; includes Labeling Module and Purification Module) ▲ **CRITICAL** We have experimented with other purification methods, but have found that these kits gave us the best yield while efficiently removing smaller DNA products. We have found that these small DNA products increased the background noise at lower signal intensity levels on our arrays.
- 1 mM Cy3-dUTP (PerkinElmer/NEN; Cat. no.: NEL578)
- 1 mM Cy5-dUTP (PerkinElmer/NEN; Cat. no.: NEL579)
- Formaldehyde Solution (see REAGENT SETUP)
- Lysis Buffer 1 (see REAGENT SETUP)
- Lysis Buffer 2 (see REAGENT SETUP)
- Lysis Buffer 3 (see REAGENT SETUP)
- Wash Buffer (RIPA) (see REAGENT SETUP)
- Elution Buffer: 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1.0% SDS
- TE: 10 mM Tris-HCl, pH 8.0, 1 mM EDTA
- TE + NaCl: 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 50 mM NaCl (keep cold)
- 2× Blunting Mix (See Step 44) ▲ **CRITICAL** Should be made fresh before use.
- 2× Ligase Mix (See Step 52) ▲ **CRITICAL** Should be made fresh before use. Linkers used in this mix should be thawed on ice, used immediately and not re-used.
- LMPCR Mix A (See Step 58)
- LMPCR Mix B (See Step 58)
- Label Mix (Nucleotide Mix and Klenow are from Invitrogen labeling kit) (See Step 70)
- The following reagents are specific to our hybridization and washing protocol; as different laboratories will select different array platforms, these are intended as a reference: 200 ng μl^{-1} sheared herring sperm DNA (Promega; Cat. no.: D1815); 8 $\mu g\ \mu l^{-1}$ yeast tRNA (Invitrogen; Cat. no.: 15401-029); 1 $\mu g\ \mu l^{-1}$ Cot-1 DNA (Invitrogen; Cat. no.: 15279-011; use species-specific Cot-1 DNA)
- 10× Control Targets (Agilent; Cat. no.: 5185-5976-P) (Reconstitute per manufacturer's directions. Used in protocol at 0.1×, not 1× as recommended by manufacturer)
- 500 mM Na-MES, pH 6.9 (Sigma; Cat. no.: M5287) (Adjust pH with 10 N sodium hydroxide) ▲ **CRITICAL** MES can go off. As a precaution, make small batches of the stock solution and use within 3 months. Store in the dark at 4 °C.
- 100% formamide (Sigma; Cat. no.: F9037) ! **CAUTION** Formamide is harmful by inhalation, in contact with skin or if swallowed; and causes burns. Formamide should be used with appropriate safety measures, such as protective gloves, glasses and clothing, and adequate ventilation. The formamide waste should be disposed of according to regulations.
- 100% acetonitrile (J.T. Baker; Cat. no.: 9017-03) ! **CAUTION** Acetonitrile is highly flammable; and is toxic by inhalation, in contact with skin or if swallowed. Acetonitrile should be used with appropriate safety measures, such as protective gloves, glasses and clothing, and adequate ventilation. The acetonitrile waste should be disposed of according to regulations.
- Hybridization Control Mix (See Step 80)
- Hybridization Buffer Master Mix (See Step 82) ▲ **CRITICAL** Should be made fresh before use.
- 20× SSPE (Invitrogen; Cat. no.: 15591-027)
- Wash I: 6× SSPE, 0.005% *N*-lauroylsarcosine
- Wash II: 6× SSPE
- Wash III (Stabilization and Drying Buffer; Agilent; Cat. no.: 5185-5979) ! **CAUTION** This solution contains acetonitrile, which is flammable and hazardous and should be used in a fume hood.

EQUIPMENT

- Cell strainer, 100 μm nylon (if using tissues) (e.g., BD Falcon)
- Rotator (e.g., Fisher Hematology/Chemistry Mixer)
- Magnetic particle collector (MPC; Dynal)
- Sonicator (e.g., Misonix 3000 sonicator equipped with microtip)
- Phase lock, heavy, pre-dispensed into 2.0 ml microfuge tubes (Eppendorf) (Although these could be replaced with standard liquid organic solvent extractions, the ease of use factor makes these ideal)
- NanoDrop ND-1000 spectrophotometer (This is highly useful as it allows you to assay DNA concentrations and dye incorporations by using low volumes (1.5 μl) of sample)
- Thermal cycler (e.g., MJ Research PTC-225) (For large-scale set-ups, it is useful to have a cycler that can handle 96-well plates)
- Swinging bucket centrifuge, variable temperature (e.g., Sorvall Legend RT)
- Rocking platform (e.g., Bellco Rocker Platform)
- Liquid nitrogen and appropriate container
- 65 degree oven (e.g., Techne Hybridiser HD-1B) (Ovens or warm rooms are preferred for longer incubations as the evenly applied heat limits condensation forming on tube lids)

REAGENT SETUP

15 μM Linker Mix 250 μl of 1 M Tris-HCl (pH 8.0), 375 μl of 40 μM oJW102 and 375 μl of 40 μM oJW103. Make 50 or 100 μl aliquots in PCR tubes. Place tubes in thermal cycler. Set up and run the following program: 95 °C for 5 min; 70 °C for 1 min; ramp down to 4 °C (0.4 °C/min); 4 °C Hold. Remove linkers and store at −20 °C. ▲ **CRITICAL** Linkers must be thawed on ice and used immediately. The linkers are labile and will dissociate, causing low-efficiency ligation. Aliquots of partially used linkers should not be re-frozen and re-used.

Formaldehyde Solution Should be made fresh. Consists of 50 mM HEPES-KOH, pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 11% formaldehyde.

Lysis Buffer 1 Add protease inhibitors just before use, filter and keep cold. Consists of 50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1× protease inhibitors

Lysis Buffer 2 Add protease inhibitors just before use, filter and keep cold. Consists of 10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1× protease inhibitors

Lysis Buffer 3 Add protease inhibitors just before use, filter and keep cold. Consists of 10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% *N*-lauroylsarcosine, 1× protease inhibitors

Wash Buffer (RIPA) Keep cold. Consists of 50 mM HEPES-KOH, pKa 7.55, 500 mM LiCl, 1 mM EDTA, 1.0% NP-40, 0.7% Na-Deoxycholate

EQUIPMENT SETUP

Sonicator We most commonly use a Misonix 3000 sonicator with microtip. The settings we typically begin with are power setting 7 (~35 W) and 12 cycles (where each cycle is a 30 s burst of sonication, followed by a 90 s pause). The range of settings will vary based on cell type, cell number, growth conditions and crosslinking. For comparison, the range of settings we have used for specific experiments is: power settings from 6 to 9; cycle number from 8 to 15; burst times of 20 to 30 s and pause times of 60 to 90 s.

Array scanning As the specifics of scanning depend on the platform chosen and the scanner available, it is less useful to provide a detailed protocol on how to scan arrays using any one specific set-up. However, there are a few guidelines that we have found to be useful for two-color hybridizations. In selecting scanner settings, the primary variable we change is photomultiplier tube (PMT) gain, which adjusts the sensitivity of the PMT that converts photons to a digitized electrical signal for the Cy3 and Cy5 channels.

A higher PMT gain creates a brighter image. We try to select PMT settings for each of the lasers in a two-color scanner that maximize signal intensity from features, without creating excessive background signal from the glass surface. Our current array platform has extremely low background signal, so our main concern is maximizing signal intensity. PMT gains are also set such that the distributions of signal intensities for each channel are similar. In addition, the settings should not result in a large number of signals that exceed, and are therefore thresholded at, the maximum detection range of the scanner. This can result in loss of detection of enrichment for features with high signal intensities.

PROCEDURE

Formaldehyde crosslinking cells ● TIMING Day 1, 1–4 h

1| For formaldehyde-crosslinking of cells, three procedures are described that can be used, depending on whether your cells are: adherent to culture apparatus (A); growing in suspension (B); or obtained fresh from tissues (C). The procedure given in (C) has been used with mouse and human tissues. Specifics vary from tissue to tissue, and this extraction procedure will therefore require some optimization — these steps are, therefore, general guidelines for extracting and crosslinking cells from tissues.

! CAUTION See REAGENTS for precautions when using formaldehyde.

? TROUBLESHOOTING

(A) For adherent cells:

(i) Use $5 \times 10^7 - 1 \times 10^8$ cells for each immunoprecipitation.

(ii) Add 1/10 volume of fresh 11% Formaldehyde Solution to plates. Swirl briefly.

(iii) Incubate cells with Formaldehyde Solution for 10 min at room temperature.

(iv) Add 1/20 volume of 2.5 M glycine to quench formaldehyde.

(v) Rinse cells twice with 10 ml of 1× PBS. Harvest cells using silicon scraper.

(vi) Pool cells in 50 ml conical tubes and spin at 700*g* for 5 min at 4 °C. Discard supernatant and resuspend pellet in 10 ml of 1× PBS per 10^8 cells with gentle inversion (cells may stick to pipettes at this stage).

(vii) Aliquot $5 \times 10^7 - 1 \times 10^8$ cells to individual 15 ml conical tubes and spin at 700*g* for 5 min at 4 °C. Discard supernatants.

(B) For suspension cells:

(i) Use $5 \times 10^7 - 1 \times 10^8$ cells for each immunoprecipitation.

(ii) Add 1/10 volume of fresh 11% Formaldehyde Solution to flasks. Swirl briefly.

(iii) Incubate cells with Formaldehyde Solution for 20 min at room temperature.

(iv) Add 1/20 volume of 2.5 M glycine to quench formaldehyde.

(v) Pool cells in required number of 50 ml conical tubes and spin at 700*g* for 5 min at 4 °C. Discard supernatant.

(vi) Resuspend cells in 50 ml of 1× PBS with gentle inversion (cells may stick to pipettes at this stage). Spin at 700*g* for 5 min at 4 °C to pellet cells. Discard supernatant. Repeat once. Resuspend final cell pellet in 10 ml of 1× PBS per 10^8 cells.

(vii) Aliquot $5 \times 10^7 - 1 \times 10^8$ cells to individual 15 ml conical tubes and spin at 700*g* for 5 min at 4 °C. Discard supernatants.

(C) For tissues:

(i) Harvest tissues. For human tissues, this involves an arrangement to obtain transplant-grade tissue that has been released for research and will depend on the tissue. Whenever possible, tissues are crosslinked as close to collection as possible and then shipped. If not possible, tissues are shipped on ice, allowed to recover briefly and then crosslinked and processed as described below. For mouse tissues, this entails euthanizing the mice and then immediately harvesting the desired tissues.

(ii) Mince tissues very finely. Transfer to 50 ml tube and add 2× volume of 1× PBS. Add 1/10 volume of fresh 11% Formaldehyde Solution.

(iii) Incubate cells with Formaldehyde Solution for 10 min at room temperature. Swirl tubes occasionally. Add 1/20 volume of 2.5 M glycine to quench formaldehyde. Immediately place cells on ice.

(iv) Homogenize cells. We typically use douncing or hand-held, mechanical homogenizers. Pass material through a 100-μm nylon cell strainer.

(v) Pool cells in required number of 50 ml conical tubes and spin at 1,100*g* for 5 min at 4 °C. Discard supernatant.

(vi) Resuspend cells in 50 ml 1× PBS with gentle inversion (cells may stick to pipettes at this stage). Spin at 1,100*g* for 5 min at 4 °C to pellet cells. Discard supernatant. Resuspend final cell pellet in 10 ml of 1× PBS per 10^8 cells.

(vii) Aliquot $5 \times 10^7 - 1 \times 10^8$ cells to individual 15 ml conical tubes and spin at 1,350*g* for 5 min at 4 °C. Discard supernatants.

2| Flash-freeze cells in liquid nitrogen and store pellets at –80 °C.

■ **PAUSE POINT** Once cells are crosslinked, they may be stored frozen at –80 °C indefinitely.

Preparing magnetic beads ● TIMING Day 2, 7–8 h

3| Add 100 μl of Dynal beads to a 1.5 ml microfuge tube. Set up 1 tube of beads per immunoprecipitation. Add 1 ml Block Solution.

▲ **CRITICAL STEP** Steps 3–9 are all performed at 4 °C.

4| Collect beads using Dynal MPC. Place tubes in rack. Allow beads to collect on side of tube. This should take approximately 15 s. Invert rack once or twice to help collect all beads. Remove supernatant with aspirator or pipettor.

5| Add 1.5 ml Block Solution and gently resuspend beads. This can be done by removing the tubes from the rack and inverting the tubes 10–20 times until the beads are evenly resuspended or by removing the magnetic strip from the rack and inverting the rack, with the tubes still in place — either 10–20 times or until the beads are evenly resuspended. Collect beads as above (Step 4). Remove supernatant with aspirator or pipettor.

6| Wash beads in 1.5 ml Block Solution, as in Step 5, one more time.

7| Resuspend beads in 250 µl Block Solution and add 10 µg of antibody. Incubate at 4 °C for a minimum of 6 h, or overnight, on a rotator.

■ **PAUSE POINT** This step can be extended to overnight at 4 °C. If so, beads should be prepared starting on Day 1.

? TROUBLESHOOTING

8| Wash beads three times in 1 ml Block Solution, as described in Step 5.

9| Resuspend each aliquot of beads in 100 µl Block Solution.

Cell sonication ● TIMING Day 2, 2 h

10| Remove frozen cell pellets from –80 °C and resuspend each pellet of ~10^8 cells in 10 ml of Lysis Buffer 1. Rock at 4 °C on platform rocker for 10 min.

11| Spin at 1,350*g* for 5 min at 4 °C in a table-top centrifuge. Discard supernatant.

12| Resuspend each pellet in 10 ml of Lysis Buffer 2. Rock gently at room temperature for 10 min.

13| Pellet nuclei in table-top centrifuge by spinning at 1,350*g* for 5 min at 4 °C. Discard supernatant.

14| Resuspend each pellet in each tube in 3.5 ml Lysis Buffer 3. Transfer cells to tubes for sonication. We currently prefer to use the bottom half of a standard polypropylene, 15 ml conical tube for sonication. We cut the tube into two pieces at the 7 ml mark and discard the upper half. Tubes can be covered with parafilm while setting up. You may also wish to save a 50 µl aliquot for use as a pre-sonication control.

15| Using a clamp, position tube so the sonicator probe sits approximately 0.5–1.0 cm above the bottom of the tube. Take care that the probe is centered and does not contact the sides of the tube.

▲ **CRITICAL STEP** Probe positioning can affect whether the solution foams or not during sonication. Typically, foaming indicates that the sonicated DNA will be poorly sheared.

? TROUBLESHOOTING

16| Immerse tube in an ice-water bath. This is most easily done by keeping the sonicator probe and tube fixed, placing the bath on a height-adjustable platform and raising it into position.

17| Sonicate suspension. Samples should be kept in an ice-water bath during sonication. To decrease foaming, initially set output power to 4 and increase manually to final power during first burst.

▲ **CRITICAL STEP** If there is significant foaming, we recommend removing all bubbles by centrifugation at 20,000*g* followed by gentle resuspension of all material, leaving no foam bubbles.

? TROUBLESHOOTING

18| Add 1/10 volume of 10% Triton X-100 to sonicated lysate. Split into two 1.5 ml centrifuge tubes. Spin at 20,000*g* for 10 min at 4 °C to pellet debris.

19| Combine supernatants from the two 1.5 ml centrifuge tubes in a new 15 ml conical tube for immunoprecipitation.

▲ **CRITICAL STEP** From here on, keep lysates on ice.

20| Save 50 µl of cell lysate from each sample as input DNA. Store at –20 °C. At least one input DNA aliquot should be kept per batch of sonicated lysate. Note that the effects of the sonication and the resulting distribution of fragment sizes can only be checked at Step 42 after crosslink reversal and purification of DNA.
■ **PAUSE POINT** Lysates can be frozen at –80 °C and used the next day.

Chromatin immunoprecipitation ● TIMING Day 2, 15 min and overnight incubation
21| Add 100 µl antibody/magnetic bead mix from Step 9 to cell lysates from Step 19. Incubate overnight on rotator at 4 °C.
▲ **CRITICAL STEP** Antibody performance can be affected by the amount of salt in the reaction. If there is previous evidence indicating specific salt conditions that are optimal for a particular antibody, we will supplement the lysate with additional sodium chloride to approximate those conditions. In this case, the salt concentrations of wash buffers used in Steps 25–26 should also be adjusted.
? TROUBLESHOOTING

Wash ● TIMING Day 3, 2 h
22| Set up 1.5 ml microfuge tubes on ice, one tube for each immunoprecipitation. Transfer half of each immunoprecipitation from the 15 ml conical tube to its assigned microfuge tube.
▲ **CRITICAL STEP** All wash steps are performed at 4 °C. Buffers should be kept cold.

23| Collect beads using Dynal MPC. Place tubes in rack. Allow beads to collect on side of tube. This should take approximately 15 s. Invert rack once or twice to help collect all beads. Remove supernatant with aspirator or pipettor, changing tips between samples.

24| Add second half of immunoprecipitation to tube. Let tubes sit again in magnetic stand to collect the beads. Remove supernatant with aspirator or pipettor, changing tips between samples.

25| Add 1 ml Wash Buffer (RIPA) to each tube and gently resuspend beads. This can be done by removing the tubes from the rack and inverting the tubes 10–20 times or by removing the magnetic strip from the rack and inverting the rack, with tubes still in place — 10–20 times or until the beads are evenly resuspended. Collect beads. Remove supernatant by aspirator or pipettor. Repeat this wash four more times, changing tips between washes.
? TROUBLESHOOTING

26| Wash once with 1 ml TE + 50 mM NaCl.

27| Spin at 960*g* for 3 min at 4 °C and remove any residual TE buffer.

Elution ● TIMING Day 3, 30–45 min
28| Add 210 µl of Elution Buffer and elute material from beads by incubating tubes in a 65 °C water bath for 15 min. Vortex briefly every 2 min. This incubation can be extended as long as 30 min, which can help improve recovery of the eluate.

29| Spin down beads at 16,000*g* for 1 min at room temperature.

30| Remove 200 µl of supernatant and transfer to new tube.
■ **PAUSE POINT** Material can be frozen at –20 °C and stored overnight.

Crosslink reversal ● TIMING Day 3, 6 h or overnight
31| Reverse crosslink the immunoprecipitation DNA from Step 30 by incubating at 65 °C for a minimum of 6 h. This incubation is usually done in an oven so that the tube is heated evenly and there is less condensation formed.
? TROUBLESHOOTING

32| Thaw 50 µl of input DNA reserved after sonication (Step 20), add 150 µl (3 volumes) of elution buffer and mix. Reverse crosslink this input DNA by incubating at 65 °C as in Step 31. From this point, every tube of immunoprecipitation or input DNA is considered to be a separate tube or sample for later processing steps.
■ **PAUSE POINT** If the wash was performed early in the day, the crosslink reversal can be stopped after a minimum of 6 h and the material can be frozen at –20 °C and stored overnight. If the wash was performed later in the day, the crosslink reversal can be extended up to 15 h and performed overnight with little effect on the reaction.

▲ CRITICAL STEP Longer times of crosslink reversal (18 h or more) usually result in increased noise in the microarray analysis.
? TROUBLESHOOTING

Purification of DNA ● TIMING **Day 4, 6 h**

33| Add 200 µl of TE to each tube to dilute SDS in Elution Buffer. We have found that high levels of SDS can inhibit RNAse activity.

34| Add 8 µl of 10 mg ml^{-1} RNaseA (0.2 mg ml^{-1} final concentration), mix by inverting the tube several times and incubate at 37 °C for 2 h.
■ PAUSE POINT Material can be frozen at −20 °C and stored overnight.

35| Add 4 µl of 20 mg ml^{-1} Proteinase K (0.2 µg ml^{-1} final concentration) and mix by inverting the tube several times and incubate at 55 °C for 2 h.

36| Add 400 µl phenol:chloroform:isoamyl alcohol (P:C:IA), vortex and separate phases with 2 ml Heavy Phaselock tube (follow instructions provided by Eppendorf).
▲ CRITICAL STEP If the P:C:IA solution is old or is at low pH, there will be degradation of DNA, causing noise in the microarray analysis and loss of detection of valid targets.
! CAUTION see REAGENTS for precautions when using phenol and chloroform.
? TROUBLESHOOTING

37| Transfer aqueous layer to new centrifuge tube containing 16 µl of 5M NaCl (200 mM final concentration) and 1.5 µl of 20 µg $µl^{-1}$ glycogen (30 µg total).

38| Add 800 µl EtOH. Incubate for 30 min at −20 °C.

39| Spin at 20,000*g* for 10 min at 4 °C to pellet DNA. Wash pellets by adding 500 µl of 80% EtOH, vortexing to resuspend pellet and spinning again at 20,000*g* for 5 min at 4 °C.

40| Remove any remaining 80% EtOH. Spin the tubes briefly to collect any remaining liquid and remove liquid with a pipetteman, avoiding the pellet. Let tubes air dry until pellets are just dry: pellets should still retain a moist appearance. Resuspend each pellet in 70 µl of 10 mM Tris-HCl, pH 8.0.
▲ CRITICAL STEP Overdrying of these pellets can make them difficult to resuspend, or liable to flake and peel away from the side of the tube.
? TROUBLESHOOTING

41| Save 15 µl of immunoprecipitation sample for future use. This material can be used to perform gene-specific PCR confirmation of microarray results.

42| Measure DNA concentration of input DNA with NanoDrop and dilute input DNA to 100 ng $µl^{-1}$. Run purified input DNA on a 2% agarose gel to check distribution of fragment sizes and digestion of RNA. The immunoprecipitation DNA will usually be too dilute to measure or visualize at this point. If possible, this is a good opportunity to check for enrichment in the immunoprecipitation sample using gene-specific PCR.
■ PAUSE POINT Material can be frozen at −80 °C and stored indefinitely.
? TROUBLESHOOTING

T4 polymerase blunting of ends ● TIMING **Day 4, 2 h**

43| Set up 1.5 ml microfuge tubes for each immunoprecipitation and input DNA sample. For each input DNA sample, mix 2 µl (200 ng) input DNA and 53 µl ddH_2O in an individual tube. For each immunoprecipitation, aliquot 55 µl into an individual tube.
▲ CRITICAL STEP Steps 43–45 should be performed on ice, unless otherwise stated.

44| Make 2× Blunting Master Mix according to the table below:

Component	Amount (per reaction)	Final amount/concentration
10× NEB buffer 2	11.0 µl	1×
1 µg µl^{-1} BSA	5.5 µl	50 µg ml^{-1}
2.5 mM dNTP mix	4.4 µl	100 µ M each dNTP
T4 DNA polymerase (3 U µl^{-1})	0.5 µl	1.5 U
ddH_2O	33.6 µl	
TOTAL volume	55.0 µl	

45| Add 55 µl of 2× Blunting Master Mix to all tubes. Incubate for 20 min at 12 °C in a water bath that is set up in a 4 °C cold room.

▲ **CRITICAL STEP** The reaction should be kept at 12 °C and promptly removed after incubation. Theoretically, the low temperature, short reaction time and high concentrations of nucleotides favor the polymerase activity of the T4 DNA polymerase and not the exonuclease activity, resulting in blunt ends and not 5′ overhangs. If no reliable water bath/cold room set-up exists, set up PCR tubes at Step 43 and use a thermal cycler to incubate reactions at 12 °C. Use of PCR tubes will require transferring the samples to microfuge tubes for Steps 46–47.

? TROUBLESHOOTING

46| Add 11.5 µl of 3 M sodium acetate and 0.5 µl of 20 µg µl glycogen (10 µg total).

47| Add 120 µl P:C:IA, vortex samples and transfer samples to room temperature Phaselock tubes. Extract once (follow instructions provided by Eppendorf).

▲ **CRITICAL STEP** If the P:C:IA solution is old or is at low pH, there will be degradation of DNA, causing noise in the microarray analysis and loss of detection of valid targets.

! CAUTION See REAGENTS for precautions when using phenol and chloroform.

? TROUBLESHOOTING

48| Transfer aqueous layer to new centrifuge tube containing 250 µl EtOH. Incubate for 30 min at −20 °C.

49| Spin at 20,000*g* for 10 min at 4 °C to pellet DNA. Wash pellets by adding 500 µl of 80% EtOH, vortexing to resuspend pellet and spinning again at 20,000*g* for 5 min at 4 °C.

50| Remove any remaining 80% EtOH. Spin the tubes briefly to collect any remaining liquid and remove liquid with a pipetteman, avoiding the pellet. Let tubes air dry until pellets are just dry: pellets should still retain a moist appearance.

▲ **CRITICAL STEP** Overdrying of these pellets can make them difficult to resuspend, or liable to flake and peel away from the side of the tube.

? TROUBLESHOOTING

51| Carefully resuspend each pellet in 25 µl H_2O with pipetting. Adding the water and then allowing the pellets to sit on ice for some time can help. Chill on ice after resuspending.

■ **PAUSE POINT** Material can be frozen at −80 °C and stored indefinitely.

Blunt-end ligation ● TIMING Day 4, 1 h and overnight incubation; Day 5, 2 h

52| Make 2× Ligase Master Mix according to the table below:

Component	Amount (per reaction)	Final amount/concentration
5× T4 DNA ligase buffer	10.0 µl	1×
15 µM linker	6.7 µl	2 µM
T4 DNA ligase (400 U µl^{-1})	0.5 µl	200 U
ddH_2O	7.8 µl	
TOTAL volume	25.0 µl	

▲ **CRITICAL STEP** Steps 52–57 should be done on ice, unless otherwise stated.

53| Add 25 µl Ligase Master Mix to 25 µl of sample and incubate for 16 h in a 16 °C water bath.

54| Add 6 µl of 3 M sodium acetate and 130 µl EtOH. Incubate for 30 min at –20 °C.

55| Spin at 20,000*g* for 10 min at 4 °C to pellet DNA. Wash pellets by adding 500 µl of 80% EtOH, vortexing to resuspend pellet and spinning again at 20,000*g* for 5 min at 4 °C.

56| Remove any remaining 80% EtOH. Spin the tubes briefly to collect any remaining liquid and remove liquid with a pipetteman, avoiding the pellet. Pellet may appear as a translucent smear on the side of the tube. Let tubes air dry until pellets are just dry: pellets should still retain a moist appearance.
▲ **CRITICAL STEP** Overdrying of these pellets can make them difficult to resuspend, or liable to flake and peel away from the side of the tube.
? TROUBLESHOOTING

57| Carefully resuspend each pellet in 25 µl H_2O with pipetting. Adding the water and then allowing the pellets to sit on ice for some time can help. Store on ice after resuspending.
■ **PAUSE POINT** Material can be frozen at –80 °C and stored indefinitely.

Ligation-mediated PCR ● TIMING Day 5, 3–6 h

58| Make LMPCR Mix A and LMPCR Mix B according to the tables below. Mixes should be prepared on ice.

LMPCR Mix A

Component	Amount (per reaction)	Final amount/concentration
10× ThermoPol buffer	4.0 µl	1×
2.5 mM dNTP mix	5.0 µl	250 µM each dNTP
40 µM oJW102 oligonucleotide	1.25 µl	1 µM
ddH_2O	4.75 µl	
TOTAL volume	15.0 µl	

LMPCR Mix B

Component	Amount (per reaction)	Final amount/concentration
10× ThermoPol buffer	1.0 µl	1×
Taq polymerase (5 U $µl^{-1}$)	0.5 µl	2.5 U
ddH_2O	8.5 µl	
TOTAL volume	10.0 µl	

59| Add 15 µl of LMPCR Mix A to each 25 µl sample, on ice. Transfer tubes to the thermal cycler.

60| If an additional expansion step is required to increase the amount of material for labeling, follow (**A**) below. If no expansion is required, run the program described in (**B**).

(A) Expansion required

(i) Program the thermocycler as follows:

Cycle	Denature	Anneal	Extend
1	55 °C, 4 min	72 °C, 3 min	
2	95 °C, 2 min		
3–15	95 °C, 30 s	60 °C, 30 s	72 °C, 1 min

(ii) Start program. Midway through the 55 °C step of cycle 1, add 10 µl Mix B to each tube to hot-start reactions. If necessary, pause the program so the tubes remain at 55 °C while adding Mix B.

(iii) After the 15-cycle program has finished, remove the samples and add 475 µl of double-distilled water. This diluted material can then be used as template for subsequent LMPCR reactions (Steps 58–60 (**B**)). A total of 5 µl of diluted material should be adequate for a new 50-µl reaction. Adjust the amount of water in the recipes listed above accordingly.
▲ **CRITICAL STEP** It is strongly recommended that you do not use this expansion until comfortable with the procedure. It is important to first calibrate the number of cycles in the first-round PCR to ensure that 15 cycles of amplification are still in the linear range. We used both real-time PCR and array hybridization of material expanded once with 25 cycles versus material expanded with the 15/25 protocol to ensure that our results were not affected by the first-round expansion.

(B) No expansion required

(i) If no expansion step is required, program the thermocycler as follows:

Cycle	Denature	Anneal	Extend
1	55 °C, 4 min	72 °C, 3 min	
2	95 °C, 2 min		
3–27	95 °C, 30 s	60 °C, 30 s	72 °C, 1 min
28	72 °C, 5 min		

The sample can be kept at 4 °C by adding a HOLD step at the end of the program.

(ii) Start program. Midway through the 55 °C step of cycle 1, add 10 µl of Mix B to each tube to hot-start reactions. If necessary, pause the program so the tubes remain at 55 °C while adding Mix B.

61| After PCR is completed, transfer reactions to individual 1.5 ml Eppendorf tubes. Pool samples if appropriate. For each 50 µl of PCR reaction, add 25 µl of 7.5 M ammonium acetate and 225 µl of cold EtOH. Incubate for 30 min at –20 °C.

62| Spin at 20,000*g* for 10 min at 4 °C to pellet DNA. Wash pellets with 500 µl of 80% EtOH. Wash pellets by adding 500 µl of 80% EtOH, vortexing to resuspend pellet and spinning again at 20,000*g* for 5 min at 4 °C.

63| Dry pellets as described above (Step 56) and resuspend each pellet in 50 µl H_2O.

? TROUBLESHOOTING

64| Measure DNA concentration with NanoDrop (use 10-fold dilutions, if necessary) and normalize all samples to 500 ng μl^{-1}. Run a 2% agarose gel to check size of LMPCR products.

■ **PAUSE POINT** Material can be frozen at –80 °C and stored indefinitely.

Cy3/Cy5 labeling of IP/WCE material ● **TIMING Day 6, 6 h**

65| Preheat thermal cycler to 95 °C.

66| For each labeling reaction, aliquot 2 µl (1 µg) of LMPCR-amplified immunoprecipitated or input DNA into a PCR tube.

67| Add 35 µl of 2.5× random primer solution and 38 µl of double-distilled water. This is a random-primed, Klenow-based extension protocol derived from Invitrogen's CGH kit. Our protocol varies from the instructions provided by Invitrogen in both reaction volume and reagent concentrations. We perform 20 reactions per '30 reaction' Invitrogen kit. A pair of reactions (one for immunoprecipitation and one for its reference input DNA) yields enough material for 1–2 hybridizations. To scale-up for more arrays, we increase number — not volume — of individual reactions.

68| Vortex samples for 30 s.

69| Place tubes in pre-heated thermal cycler. After 5 min at 95 °C, immediately transfer samples to an ice-water slurry bath for 10 min to flash-cool reagents.

70| While reagents are cooling, make up Label Mix. One mix is required for immunoprecipitation samples (usually Cy5). A second mix is required for input DNA samples (usually Cy3). If reagents permit, a dye swap can provide a useful comparison (a second set of replicates where the immunoprecipitate is labeled with Cy3 and the input DNA is labeled with Cy5).

Label Mix:

Component	Amount (per reaction)	Final amount/concentration
10× dUTP Nucleotide mix	8.2 µl	1×
Cy5 (or Cy3)	1.5 µl	17 µM
Klenow (40 U μl^{-1})	1.5 µl	60 U
ddH_2O	1.8 µl	
TOTAL volume	13.0 µl	

71| Vortex samples for 30 s.

72| Add 13 µl of Label Mix to each sample. Pipette up and down multiple times to mix reagents.

73| Incubate samples for 3 h at 37 °C.
▲ **CRITICAL STEP** Keep samples in dark as Cy dyes are sensitive to exposure to light. Other incubation times and conditions have been tried, but we find that these conditions offer the best yield, while limiting the accumulation of smaller DNA fragments (<100 bp) that can contribute to noise in the microarray analysis.

74| Add 9 µl stop buffer to each well and mix by pipetting gently.

75| Clean up DNA using the DNA purification columns provided with the CGH kit. Follow manufacturer's instructions.
? TROUBLESHOOTING

76| After eluting labeled DNA from the column into a microfuge tube, spin tubes at 15,000*g* for 2 min at room temperature to pellet any fine particulates. Carefully transfer supernatants to fresh tubes. Repeat if necessary to ensure that the solution is clear.
▲ **CRITICAL STEP** It is particularly important to thoroughly remove the particulates at this step. Particulates from the purification column can bind to some slide surfaces, making it difficult, if not impossible, to successfully scan the slide.
? TROUBLESHOOTING

77| Measure DNA concentration and dye incorporation with NanoDrop. Optional: Run a 2% agarose gel to check size of labeling products.
■ **PAUSE POINT** If this DNA is going to be used within the next day, it can be stored at –20 °C. If the DNA will be stored for longer, we recommend storing it as a pellet. Precipitate the DNA by adding 25 µl of ammonium acetate and 300 µl of ice-cold EtOH and placing the tubes at –80 °C for 10 min. Spin at 20,000*g* for 10 min at 4 °C to pellet DNA. Wash pellets with 500 µl of 80% EtOH. Remove excess ethanol. Allow pellets to air-dry briefly and store tubes at –80 °C in a light-proof box or wrapped in aluminum foil.
? TROUBLESHOOTING

Array hybridization ● TIMING Day 6, 1–2 h, plus 40-h incubation.
78| For each hybridization, combine 5 µg of the Cy5-labeled immunoprecipitated DNA and 5 µg of the appropriate Cy3-labeled reference DNA. Keep all tubes on ice. The hybridization can be performed with less DNA (we have used as little as 3 µg of each Cy5-labeled and Cy3-labeled), but better signal:noise ratios are seen with larger amounts of DNA. If working with frozen pellets, resuspend each pellet first in 50 µl of double-distilled water and then combine however much you need for the hybridization.

79| For each hybridization, bring the combined DNA up to a total volume of 120 µl. Keep all tubes on ice. If the DNA solutions are dilute, this volume can be extended as high as 170 µl and water can be removed from the hybridization buffer mix as necessary.

80| Make Hybridization Control Master Mix.

Hybridization Control Master Mix

Component	Amount (per reaction)	Final amount/concentration
Sheared herring sperm DNA (200 ng µl^{-1})	4.0 µl	1.6 µg ml^{-1}
Yeast tRNA (8 µg µl^{-1})	5.0 µl	80 µg ml^{-1}
Cot-1 DNA (1.6 µg µl^{-1})	10.0 µl	20 µg ml^{-1}
Control oligonucleotides (1×)	5.0 µl	0.1×
TOTAL volume	24.0 µl	

81| Add 24 µl of Hybridization Control Master Mix to each hybridization. Mix by pipetting gently. Keep all tubes on ice.

82| Make Hybridization Buffer Master Mix. This solution should be made fresh before use and kept at room temperature.

Hybridization Buffer Master Mix

Component	Amount (per reaction)	Final amount/concentration
ddH_2O	50.0 µl	
500 mM Na-MES, pH 6.9	50.0 µl	50 mM
5 M NaCl	50.0 µl	500 mM
500 mM EDTA	6.0 µl	6 mM
5 % *N*-lauroylsarcosine	50.0 µl	0.5%
Formamide	150.0 µl	30%
TOTAL volume	356.0 µl	

! CAUTION See REAGENTS for precautions when working with formamide.

83| Add 356 µl of Hybridization Buffer Master Mix to each hybridization. Mix by pipetting gently. After mixing, do not return samples to ice. Instead, leave tubes at room temperature until all samples have been processed. Keep all samples protected from light.

84| Heat samples for 3 min in a heat block that has been set to 95 °C. Cover with foil to prevent exposure to light.

85| Transfer tubes to a 40 °C heat block for 15 min. Cover with foil to prevent exposure to light. At this point, start preparing hybridization chambers for assembly, aiming to handle 4–6 slides at a time. If there are more slides than this, samples can sit at this 40 °C step for at least 1 h.

86| After 15 min, spin tubes at 14,000*g* for 45 s to pellet any particulates.
? TROUBLESHOOTING

87| Assemble hybridizations according to manufacturer's instructions. Use 490 µl of hybridization mix per hybridization.

88| Incubate at 40 °C in rotating oven for 40 h. Ensure free rotation of liquid throughout the chamber. We find the best results are obtained with an incubation time of 38–42 h.

Array washing ● TIMING Day 8, 1–2 h

89| Before disassembling hybridization chambers, prepare Wash I (6× SSPE, 0.005% H-lauroylsarcosine) and Wash II (6× SSPE). Filter Wash I and Wash II before use. Place Wash III at 37 °C to dissolve solute. Swirling the bottle of Wash III occasionally will help to dissolve the solute. Allow the solution to cool to room temperature before use.
! CAUTION As Wash III is volatile, avoid temperatures that are higher than 45 °C.
? TROUBLESHOOTING

90| Prepare and wash all glassware and equipment that is required for slide disassembly (follow manufacturer's instructions) and for washes described in Steps 91–95. We usually use standard glass 10-slide microscope slide racks and dishes.
? TROUBLESHOOTING

91| After hybridization, disassemble hybridization chambers individually following the manufacturer's instructions. Batches of 4–6 slides are convenient. Slides are placed in slide racks in dishes containing a magnetic stir bar (thin enough to spin freely when slides and slide rack are added) and sufficient Wash I to cover the slides.
? TROUBLESHOOTING

92| Once all slides for this batch are in the dish, place dish on stir plate and gently stir for 5 min at room temperature. On a typical stir plate with settings from 1 to 10, use setting 3 or 4. The surface of the liquid should barely dimple. During this wash, prepare two dishes with 270 ml of Wash II, one with a stir bar.
? TROUBLESHOOTING

93| After the 5-min wash in Wash I, remove rack from Wash I, quickly dip it into a dish containing Wash II and immediately transfer to the second dish containing 270 ml of Wash II and stir bar.
? TROUBLESHOOTING

94| Place this dish on a stir plate and gently stir for 5 min at room temperature. Use stir plate settings as described in Step 92. During this wash, prepare the washes for the next step in the fume hood. Pour 270 ml of 100% acetonitrile (FPLC grade) into a small rectangular dish. Pour 250 ml (one half bottle) of room temperature Wash III into a second, small rectangular dish, add a stir bar and place on a stir plate. Cover both dishes when not in use to prevent evaporation.
! CAUTION See REAGENTS for precautions when using acetonitrile.
? TROUBLESHOOTING

95| Turn off stir plate and carry dish containing the slides to the hood. Turn on stir plate in the hood to begin agitating Wash III. Remove rack, quickly dip it into dish containing 100% acetonitrile, and immediately transfer to a final dish containing Wash III and a rotating stir bar.
? TROUBLESHOOTING

96| After 15 s, turn off the stir plate. This allows particulates to settle, and helps prevent them from sticking to your slide as you pull the slides out. Wait another 15 s.

97| Slowly and smoothly remove rack and slides from Wash III — it should take approximately 10 s to lift the rack from the solution (the slides should be dry at this point). If you move the rack abruptly as you are removing it, you may create visible lines of Wash III precipitate. If this happens, re-dip the slide momentarily and try to pull it out again.
? TROUBLESHOOTING

98| Scan immediately. As fluorescent dyes are sensitive to exposure to atmosphere, it is best to scan slides as soon as possible after washing. Various ozone-scavenging treatments, such as Wash III, extend the life of hybridized slides, but still eventually fail. Slides can be stored in vacuum-sealed bags or a high nitrogen environment for short times (2–5 h) until needed.
? TROUBLESHOOTING

Array analysis
99| As described above, each lab is likely to have specific preferences for analysis of the array. A very brief, general outline is provided here as an example and a more detailed explanation is available in Lee *et al.*[14]. Scan slides (we use an Agilent DNA microarray scanner) at wavelengths of 532 nm for input (Cy3) and 635 nm for enriched (Cy5).

100| Use GenePix feature-identification software to automatically identify features and to extract data for fluorescence at each wavelength by calculating the average intensities for all pixels within the feature boundaries.

101| Subtract local background signal for each feature. Normalize the signal intensities of the slide across the entire set of slides using signals from a set of common features that are printed on every slide in the set. Then, subtract the signal intensities from a set of negative controls for hybridization.

102| Median normalize the data from the two channels and generate a log ratio of signal intensities for each feature.

103| Process the log ratios using a whole chip error model[52], which incorporates signal intensity and background noise to calculate a *P*-value for the statistical significance of the enrichment ratio that is seen at each feature. We also calculate the *P*-value for enrichment for sets of three adjacent features and incorporate information about the enrichment at both individual features and sets of features to identify genomic regions that qualify as being bound by the protein in question.
? TROUBLESHOOTING

104| Identify distinct binding events by collapsing adjacent bound regions.

● TIMING
Day 1: Steps 1–2 (approximately 1–4 h, depending on numbers of cells or tissues being collected)
Day 2: Steps 3–9 (approximately 8 h, should be started first)
Day 2: Steps 10–21 (approximately 2 h and overnight incubation)
Day 3: Steps 22–32 (approximately 8–9 h, or approximately 2 h and overnight incubation)
Day 4: Steps 33–53 (approximately 9 h and overnight incubation)
Day 5: Steps 54–64 (approximately 6–7 h)
Day 6: Steps 65–88 (approximately 7–8 h and 40-h hybridization)
Day 8: Steps 89–98 (approximately 1–2 h and scanning time)

TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1.**

TABLE 1 | Troubleshooting table.

PROBLEM	POSSIBLE REASON	SOLUTION
Insufficiently sheared chromatin (detected at Step 42)	Over-crosslinked material (Step 1)	Reduce crosslinking time (Step 1)
	Solution foamed (Step 17)	Keep sonication tip centered and within 1 cm of bottom of tube (Step 15)
	Solution foamed (Step 17)	If the solution has foamed before, make sure that all bubbles are removed (Step 17)
	Under-sonicated (Step 17)	Increase sonication (either power settings or number of rounds of sonication) (Step 17)
Low signal:noise or failure to immunoprecipitated DNA (detected at Step 42 or Step 103)	Over-crosslinked material (epitope obscured or denatured) (Step 1)	Reduce crosslinking time (Step 1)
	Under-crosslinked (protein insufficiently crosslinked to DNA) (Step 1)	Increase crosslinking time. We have gone as long as 16 h at 4 °C (Step 1)
	Ineffective antibody (Step 7)	If possible, try a different antibody (Step 7)
	Poor conditions for immunoprecipitation (Step 21)	Change salt concentrations of immunoprecipitation and subsequent washes (Step 21 and Step 25)
Debris on slides (detected at Step 98)	Particulates not removed from labeled material (Step 76, Step 86)	Spin labeled DNA at high speed to remove particulates (Step 76, Step 86)
	Particulates during wash (Step 91, Step 92, Step 93, Step 94, Step 95)	Use clean glassware and filter necessary wash solutions. If using an ozone-scavenger solution, dissolve solute completely (Step 89, Step 90)
	Lines/streaks on slides (Step 98)	Transfer slides quickly between washes to minimize drying. When removing slides from final wash, avoid sudden movements (Step 93, Step 95, Step 97)
Noise in microarray analysis (abnormally wide distribution of immunoprecipitated/input DNA signal ratios) (detected at Step 103)	Over-sonicated material (Step 17)	Reduce sonication (either power settings or number of rounds of sonication) (Step 17)
	Beads overwashed (Step 25)	Minimize wash times; reduce number of washes (Step 25)
	Incubation period too long for crosslink reversal (Step 31, Step 32)	Reduce incubation for crosslink reversal to 6–8 h (Step 31, Step 32)
	DNA degraded by phenol (Step 36, Step 47)	Replace P:C:IA solution (Step 36, Step 47)
	Poor recovery of DNA (Step 40, Step 50, Step 56, Step 63)	Take care when resuspending pellets (Step 40, Step 50, Step 56, Step 63)
	Poor blunting (Step 45)	Keep reaction temperature low and incubation time short (Step 45)
	Small DNA fragments post-labeling (Step 77)	Use a clean-up method that limits purification of small (60 bp or less) fragments (Step 75)

ANTICIPATED RESULTS

Step 42 (i)

Measure DNA concentration of input DNA (from Step 20) with NanoDrop. Input DNA samples should be approximately 70 µl total volume with a DNA concentration of 300–500 ng μl^{-1}. Immunoprecipitation samples are usually too dilute to detect reliably and measurements are sometimes skewed due to the glycogen that is added to the sample, but we have pooled and precipitated several immunoprecipitates and, after adjusting for changes in volume, expect a DNA concentration of 2–6 ng μl^{-1}. The immunoprecipitated DNA concentration may be higher if the immunoprecipitation is directed against an abundant protein.

Step 42 (ii)

Run a 2% agarose gel of the input DNA that is saved prior to immunoprecipitation (Step 20). This allows you to visually check the distribution of fragment sizes that result from sonication, and to confirm digestion of RNA. Examples of different distributions of fragment sizes resulting from different degrees of sonication are shown in **Figure 2.** We would first try ChIP with the fragment distribution that is seen with 12 cycles of sonication, as this is the lowest amount of

sonication that results in a majority of fragments, ranging from 200–600 bp. The fragment distribution that is seen with eight cycles of sonication is also likely to yield reasonable ChIP results, but this material has generally larger fragments that could reduce the ability to resolve binding events. The fragment distribution that is seen with 16 cycles of sonication is similar to the 12-cycle material, but we would default to the lower amount of sonication. Note that different cell types, growth conditions and crosslinking conditions will change the appearance of similar sonication time courses. However, the basic principle of finding a setting that minimizes sonication while maximizing fragments in the range of 200–600 bp should still be a useful starting point.

It can also be useful to calculate the fragment-size distribution more precisely with laser densitometry or a device such as the Agilent 2100 Bioanalyzer. These values are useful for computational algorithms that combine knowledge of the size distribution with genomic probe locations, probe characteristics and enrichment ratios to develop more sophisticated models of protein-DNA binding.

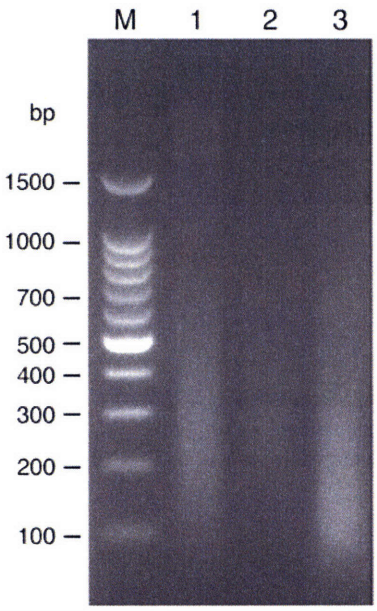


Figure 3| 2% agarose gel showing an example of input DNA, DNA after LMPCR amplification and labeled DNA. The lane with molecular-mass markers is labeled M. Input DNA is in Lane 1, LMPCR DNA is in Lane 2, labeled DNA is in Lane 3. Approximately 200–300 ng of DNA are loaded. Note that labeled DNA will fluoresce (and thus appear more abundant) when photographed with UV illumination due to the presence of dye.

Step 42 (iii)

Whenever possible, you should check that the immunoprecipitation was successful by looking for enrichment of known positive controls in the immunoprecipitated DNA sample. This is most easily done by performing PCR using primers for known positive control and negative control regions and running the products on an agarose gel. Compare the ratio of DNA for the positive control versus the negative control for both the immunoprecipitated DNA (use 2 μl) and the input DNA (use dilution series). The immunoprecipitation positive:negative ratio should be at least twice the input DNA positive:negative ratio. It is important to perform a serial dilution of the input DNA PCR reactions to ensure that all of the reactions are in the linear range of detection. Ideally, we run the reactions as a multiplex PCR so the expected positive and the negative control are being tested in the same PCR reaction (adjust primer concentrations so the input DNA positive:negative ratio is close to 1). This is not always possible with every combination of primers.

Step 64 (i)

Measure DNA concentration with NanoDrop. DNA samples should be approximately 50 μl total volume with a DNA concentration of 100–350 ng μl^{-1}. All samples, immunoprecipitated and input DNA should be comparable in DNA amounts.

Step 64 (ii)

Run a 2% agarose gel of the LMPCR DNA. DNA products should be centered around 200–400 bp. All samples, immunoprecipitated and input DNA should be comparable in distribution of fragment sizes and DNA amounts. An example of LMPCR DNA run on an agarose gel is shown in **Figure 3**.

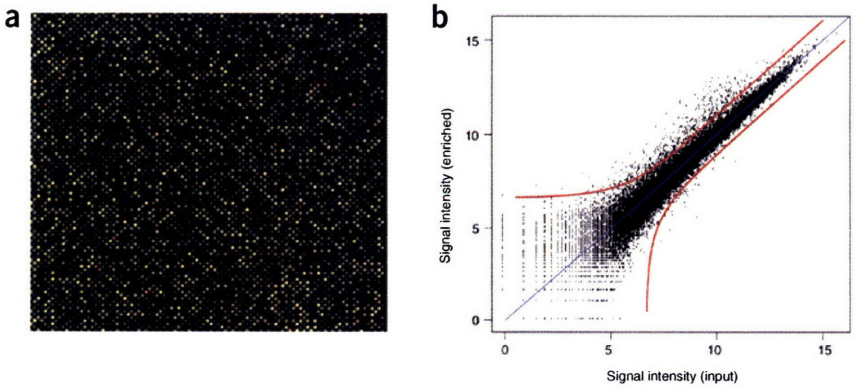


Figure 4| Samples of hybridized arrays and scatterplots. (**a**) A portion of a hybridized and scanned array. The material is enriched for RNA polymerase II binding. Red features indicate enrichment. Bright green features are hybridization controls. Yellow and yellow-green features indicate no enrichment. (**b**) A plot showing processed results of the hybridization and scan shown in **a**. Signal intensities (log base 2) for the RNA polymerase II-enriched DNA are shown on the *y*-axis. Signal intensities (log base 2) for the input DNA are shown on the *x*-axis. Features outside the boundaries, indicated by the red lines, are significantly enriched at a threshold of $P<0.001$. Features that are below the diagonal and appear significantly enriched are usually control features or dust particles that have been mistakenly identified as features. Features that are above the diagonal and appear significantly enriched usually represent immuno-enriched features and indicate protein binding.

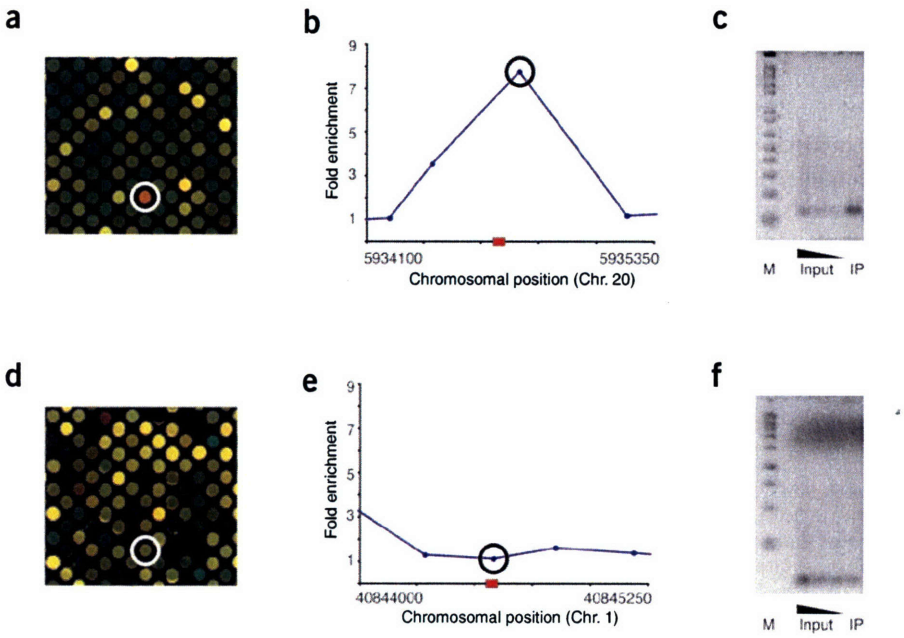


Figure 5| Examples of data from an array that are processed to identify genomic regions that are enriched for binding and verification by gene-specific PCR. (**a**) Another portion of the hybridized and scanned array shown in **Figure 4**. The material is enriched for RNA polymerase II binding. The circled feature is red, indicating enrichment. Yellow and yellow-green features indicate no enrichment. (**b**) A plot showing ratios of the RNA polymerase II enrichment. Enrichment is plotted on the *y*-axis. Genomic location is plotted on the *x*-axis. Probe location is shown by the blue circles. The circled probe represents the ratio seen for the circled feature shown in **a**. The red box indicates the region checked using gene-specific PCR to confirm enrichment in **c**. (**c**) Gene-specific PCR, confirming enrichment. Primers were used to PCR the genomic region, indicated by the red box in **b** for immunoprecipitated DNA and dilutions of input DNA. Molecular-mass markers are shown in the left-most lane, which is labeled M. Three-fold dilutions of input DNA are shown in lanes 1–3, as indicated. The immunoprecipitation (IP) is shown in the last lane. The amount of immunoprecipitate that is added to the PCR is equivalent to the last dilution of input DNA. (**d**) A portion of a hybridized and scanned array. The hybridized material is enriched for RNA polymerase II binding. The circled feature shows no enrichment. Yellow and yellow-green features indicate no enrichment. (**e**) A plot showing ratios of the RNA polymerase II enrichment. Enrichment is plotted on the *y*-axis. Genomic location is plotted on the *x*-axis. Probe location is shown by the blue circles. The circled probe represents the ratio seen for the circled feature shown in **d**. The red box indicates the region checked using gene-specific PCR to confirm enrichment in **f**. (**f**) Gene-specific PCR confirming lack of enrichment. Primers were used to PCR the genomic region, indicated by the red box in **d** for immunoprecipitated DNA and dilutions of input DNA. Molecular-mass markers are shown in the left-most lane, which is labeled M. Three-fold dilutions of input DNA are shown in lanes 1–3, as indicated. The immunoprecipitation is shown in the last lane. The amount of immunoprecipitate added to the PCR is equivalent to the last dilution of input DNA.

Step 77 (i)

Measure DNA concentration and dye incorporation with NanoDrop. DNA samples should be approximately 50 μl total volume with a DNA concentration of 100–150 ng μl^{-1} and 4–10 pmol μl^{-1} of incorporated dye.

Step 77 (ii)

Run a 2% agarose gel of the labeled, purified DNA. DNA products should be centered around 200–300 bp. There should not be significant amounts of DNA fragments that are smaller than 100 bp in the purified sample. An example of labeled DNA run on an agarose gel is shown in **Figure 3**.

Step 98

A sample region of a scanned array is shown in **Figure 4**. Arrays should have a clean and uniform background, with little particulate debris or dust. Features should be evenly arranged and have a uniform shape. Typically, we label immunoprecipitated material with Cy5 and input DNA material with Cy3, and set array visualization so that Cy5 signals appear red and Cy3 signals appear green. Thus, for a successful immunoprecipitation, we expect a subpopulation of red spots, indicating features that have been enriched in the immunoprecipitation channel. We expect few, if any, non-control features that appear bright green, indicating features enriched in the input DNA channel. We expect the majority of spots to be yellow or yellowish-green, indicating no enrichment. Note that this assumes there are relatively few binding events and that the array has been scanned to balance the distributions of signal intensities in the two channels.

Step 104

Figure 5 shows a progression of data from array signal to identification of genomic regions that have been enriched for binding to verification by gene-specific PCR. The array images (**Fig. 5a,d**) are from slides that were part of a multi-slide, whole-genome array that detects binding of the initiating form of RNA polymerase II. The slides were scanned and processed, as described, to generate ratios of signal intensities. The subsequent ratios were processed to identify significantly enriched features and binding events.

When we plot enrichment ratios against chromosomal position of array features, we expect binding events to have two basic characteristics (**Fig. 5b**). Features identifying a binding event should show significantly higher enrichment ratios than background. Observed enrichment ratios generally range from 2-fold to 15-fold, although ratios will vary based on antibody performance and can be as high as 80-fold in some cases. Binding events should be supported by multiple features: two or more

adjacent features should both indicate enrichment above background. The number of probes that are expected to identify a binding event will depend on the degree of sonication, which influences the size of DNA fragments in the ChIP and the tiling density of the array, which affects the average number of probes that could reasonably detect immuno-enriched fragments, given the distribution of fragment sizes and whether the protein binds over large or small regions.

Legitimate binding events can also be confirmed by PCR (**Fig. 5c**). PCR should be performed using primers that are designed to amplify the genomic regions that are bound and run the products should be run on an agarose gel. The signals that are generated using the immunoprecipitated DNA sample should be compared to the signal generated using the input DNA sample. There should be more signal in the immunoprecipitated DNA sample when equivalent amounts of immunoprecipitated DNA and input DNA are used in the PCR. A serial dilution of the input DNA PCR reactions should be performed to ensure that all of the reactions are in the linear range of detection.

ACKNOWLEDGMENTS The protocol and suggestions presented here represent the experience accumulated over several years of experimentation by a number of researchers in the lab. In particular, we would like to acknowledge Dmitry Pokholok and Brett Chevalier for their contributions to streamlining the protocol. Other particularly useful contributions or observations were offered by Duncan Odom, Beth Jacobsen, Richard Jenner, Laurie Boyer, Matthew Guenther, Roshan Kumar, Heather Plasterer, David Reynolds and Jane Yoo. We thank current members of the lab, particularly Megan Cole, Matthew Guenther, Beth Jacobsen, Richard Jenner and Michael Kagey, for critical reading of this manuscript.

COMPETING INTERESTS STATEMENTS The authors declare competing financial interests (see the HTML version of this article for details).

Published online at http://www.natureprotocols.com/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Solomon, M.J., Larsen, P.L. & Varshavsky, A. Mapping protein-DNA interactions *in vivo* with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* **53**, 937–947 (1988).
2. Dedon, P.C., Soults, J.A., Allis, C.D. & Gorovsky, M.A. Formaldehyde cross-linking and immunoprecipitation demonstrate developmental changes in H1 association with transcriptionally active genes. *Mol. Cell. Biol* **11**, 1729–1733 (1991).
3. Orlando, V. & Paro, R. Mapping Polycomb-repressed domains in the bithorax complex using *in vivo* formaldehyde cross-linked chromatin. *Cell* **75**, 1187–1198 (1993).
4. Reid, J.L., Iyer, V.R., Brown, P.O. & Struhl, K. Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. *Mol. Cell* **6**, 1297–1307 (2000).
5. Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
6. Iyer, V.R. *et al.* Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–538 (2001).
7. Lieb, J.D. Liu, X., Botstein, D. & Brown, P.O. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nature Genet.* **28**, 327–334 (2001).
8. Kurdistani, S.K., Robyr, D., Tavazoie, S. & Grunstein, M. Genome-wide binding map of the histone deacetylase Rpd3 in yeast. *Nature Genet.* **31**, 248–254 (2002).
9. Lee, T.I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
10. Ren, B. *et al.* E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev.* **16**, 245–256 (2002).
11. Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H. & Farnham, P.J. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev.* **16**, 235–244 (2002).
12. Kim, T.H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).
13. Pokholok, D.K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517–527 (2005).
14. Lee, T.I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
15. Nal, B., Mohr, E. & Ferrier, P. Location analysis of DNA-bound proteins at the whole-genome level: untangling transcriptional regulatory networks. *Bioessays* **23**, 473–476 (2001).
16. Hanlon, S.E. & Lieb, J.D. Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. *Curr. Opin. Genet. Dev.* **14**, 697–705 (2004).
17. Blais, A. & Dynlacht, B.D. Constructing transcriptional regulatory networks. *Genes Dev.* **19**, 1499–1511 (2005).
18. Sikder, D. & Kodadek, T. Genomic studies of transcription factor-DNA interactions. *Curr. Opin. Chem. Biol.* **9**, 38–45 (2005).
19. Carter, N.P. & Vetrie, D. Applications of genomic microarrays to explore human chromosome structure and function. *Hum. Mol. Genet.* **13** (Spec No 2), R297–R302 (2004).
20. van Steensel, B. & Henikoff, S. Epigenomic profiling using microarrays. *Biotechniques* **35**, 346–357 (2003).
21. MacAlpine, D.M. & Bell, S.P. A genomic view of eukaryotic DNA replication. *Chromosome Res.* **13**, 309–326 (2005).
22. Woodfine, K., Carter, N.P., Dunham, I. & Fiegler, H. Investigating chromosome organization with genomic microarrays. *Chromosome Res.* **13**, 249–257 (2005).
23. Lei, E.P., Krebber, H. & Silver, P.A. Messenger RNAs are recruited for nuclear export during transcription. *Genes Dev.* **15**, 1771–1782 (2001).
24. Ishii, K., Arib, G., Lin, C., Van Houwe, G. & Laemmli, U.K. Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell* **109**, 551–562 (2002).
25. Rodriguez-Navarro, S. *et al.* Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. *Cell* **116**, 75–86 (2004).
26. Takizawa, P.A., DeRisi, J.L., Wilhelm, J.E. & Vale, R.D. Plasma membrane compartmentalization in yeast by messenger RNA transport and a septin diffusion barrier. *Science* **290**, 341–344 (2000).
27. Tenenbaum, S.A., Carson, C.C., Lager, P.J. & Keene, J.D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl. Acad. Sci. USA* **97**, 14085–14090 (2000).
28. Brown, V. *et al.* Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107**, 477–487 (2001).
29. Hieronymus, H. & Silver, P.A. Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. *Nature Genet.* **33**, 155–161 (2003).
30. Shepard, K.A. *et al.* Widespread cytoplasmic mRNA transport in yeast: identification of 22 bud-localized transcripts using DNA microarray analysis. *Proc. Natl Acad. Sci. USA* **100**, 11429–11434 (2003).
31. Sanchez-Elsner, T., Gou, D., Kremmer, E. & Sauer, F. Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax. *Science* **311**, 1118–1123 (2006).
32. Orlando, V., Strutt, H. & Paro, R. Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *Methods* **11**, 205–214 (1997).
33. Kuo, M.H. & Allis, C.D. *In vivo* cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* **19**, 425–433 (1999).
34. Buck, M.J. & Lieb, J.D. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**, 349–360 (2004).
35. Boyer, L.A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
36. Boyer, L.A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
37. Lieb, J.D. Genome-wide mapping of protein-DNA interactions by chromatin immunoprecipitation and DNA microarray hybridization. *Methods Mol. Biol.* **224**, 99–109 (2003).
38. Oberley, M.J. & Farnham, P.J. Probing chromatin immunoprecipitates with CpG-island microarrays to identify genomic sites occupied by DNA-binding proteins. *Methods Enzymol.* **371**, 577–596 (2003).
39. Bannister, A.J. & Kouzarides, T. Histone methylation: recognizing the methyl

mark. *Methods Enzymol.* **376**, 269–288 (2004).
40. Bernstein, B.E., Humphrey, E.L., Liu, C.L. & Schreiber, S.L. The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. *Methods Enzymol.* **376**, 349–360 (2004).
41. Chaya, D. & Zaret, K.S. Sequential chromatin immunoprecipitation from animal tissues. *Methods Enzymol.* **376**, 361–372 (2004).
42. Ciccone, D.N., Morshead, K.B. & Oettinger, M.A. Chromatin immunoprecipitation in the analysis of large chromatin domains across murine antigen receptor loci. *Methods Enzymol.* **376**, 334–348 (2004).
43. Oberley, M.J., Tsao, J., Yau, P. & Farnham, P.J. High-throughput screening of chromatin immunoprecipitates using CpG-island microarrays. *Methods Enzymol.* **376**, 315–334 (2004).
44. Ren, B. & Dynlacht, B.D. Use of chromatin immunoprecipitation assays in genome-wide location analysis of mammalian transcription factors. *Methods Enzymol.* **376**, 304–315 (2004).
45. Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O. & Schreiber, S.L. Global nucleosome occupancy in yeast. *Genome Biol.* **5**, R62 (2004).
46. Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
47. Liu, C.L. *et al.* Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol* **3**, e328 (2005).
48. Rada-Iglesias, A. *et al.* Binding sites for metabolic disease related transcription factors inferred at base pair resolution by chromatin immunoprecipitation and genomic microarrays. *Hum. Mol. Genet.* **14**, 3435–3447 (2005).
49. Rao, B., Shibata, Y., Strahl, B.D. & Lieb, J.D. Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol. Cell. Biol.* **25**, 9447–9459 (2005).
50. Yuan, G.C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
51. Bieda, M., Xu, X., Singer, M.A., Green, R. & Farnham, P.J. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res.* **16**, 595–605 (2006).
52. Hughes, T.R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).

Appendix C

Supplementary Data for Chapter 3

Supplementary Data and Tables

All supplementary data can be found at the following website:

http://www.cell.com/cgi/content/full/125/2/301/DC1/

All tables can be found on the supporting website; the URLs below can be used to download the appropriate table.

Table S1. Regions bound by RNA polymerase II and their relationship to known and predicted genes. http://web.wi.mit.edu/young/hES_PRC/TableS1.xls

Table S2. HUGO/EntrezGene identifiers for RNA Pol II bound, annotated genes. http://web.wi.mit.edu/young/hES_PRC/TableS2.xls

Table S3. RNA polymerase II-bound regions that predict novel gene candidates. http://web.wi.mit.edu/young/hES_PRC/TableS3.xls

Table S4. Gene models bound by RNA polymerase II. http://web.wi.mit.edu/young/hES_PRC/TableS4.xls

Table S5. MicroRNA genes bound by RNA polymerase II and Suz12 in ES cells. http://web.wi.mit.edu/young/hES_PRC/TableS5.xls

Table S6. Expression of genes bound by RNA polymerase II in ES cells. http://web.wi.mit.edu/young/hES_PRC/TableS6.xls

Table S7. Regions bound by Suz12 and their relationship to known and predicted genes. http://web.wi.mit.edu/young/hES_PRC/TableS7.xls

Table S8. HUGO/EntrezGene identifiers for Suz12-bound, annotated genes. http://web.wi.mit.edu/young/hES_PRC/TableS8.xls

Table S9. Detection of Suz12, Eed and H3K27me3 occupancy using promoter arrays. http://web.wi.mit.edu/young/hES_PRC/TableS9.xls

Table S10. Enriched gene ontologies among RNA Pol II-bound and Suz12-bound genes. http://web.wi.mit.edu/young/hES_PRC/TableS10.xls

Table S11. Developmental transcription factors bound by Suz12. http://web.wi.mit.edu/young/hES_PRC/TableS11.xls

Table S12. Developmental signaling proteins bound by Suz12. http://web.wi.mit.edu/young/hES_PRC/TableS12.xls

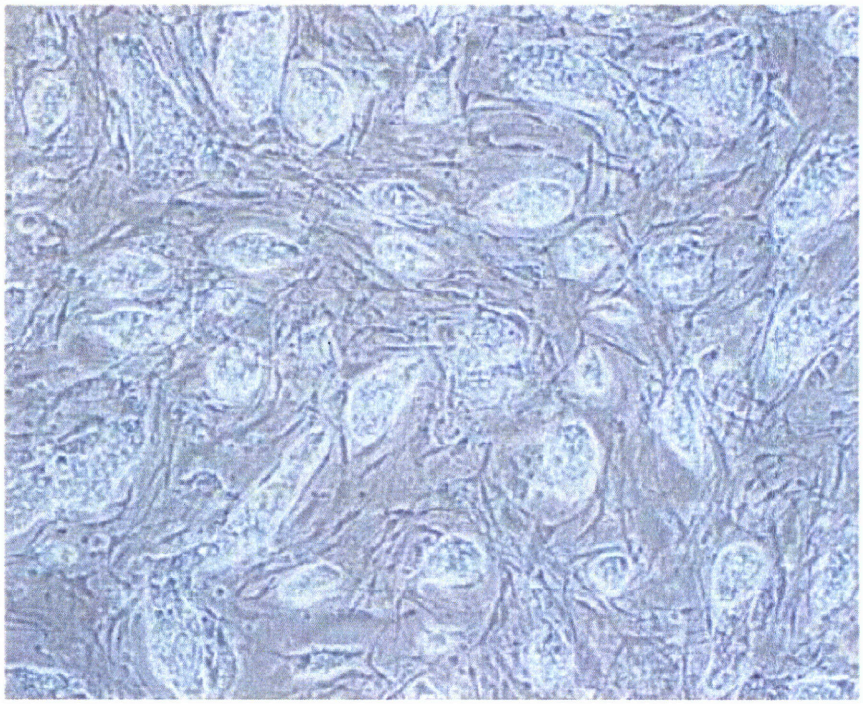
Table S13. Expression of Suz12-bound genes during ES cell differentiation.

http://web.wi.mit.edu/young/hES_PRC/TableS13.xls

Table S14. Genes bound by Suz12 in ES cells and upregulated in Suz12 -/- mouse cells.
http://web.wi.mit.edu/young/hES_PRC/TableS14.xls

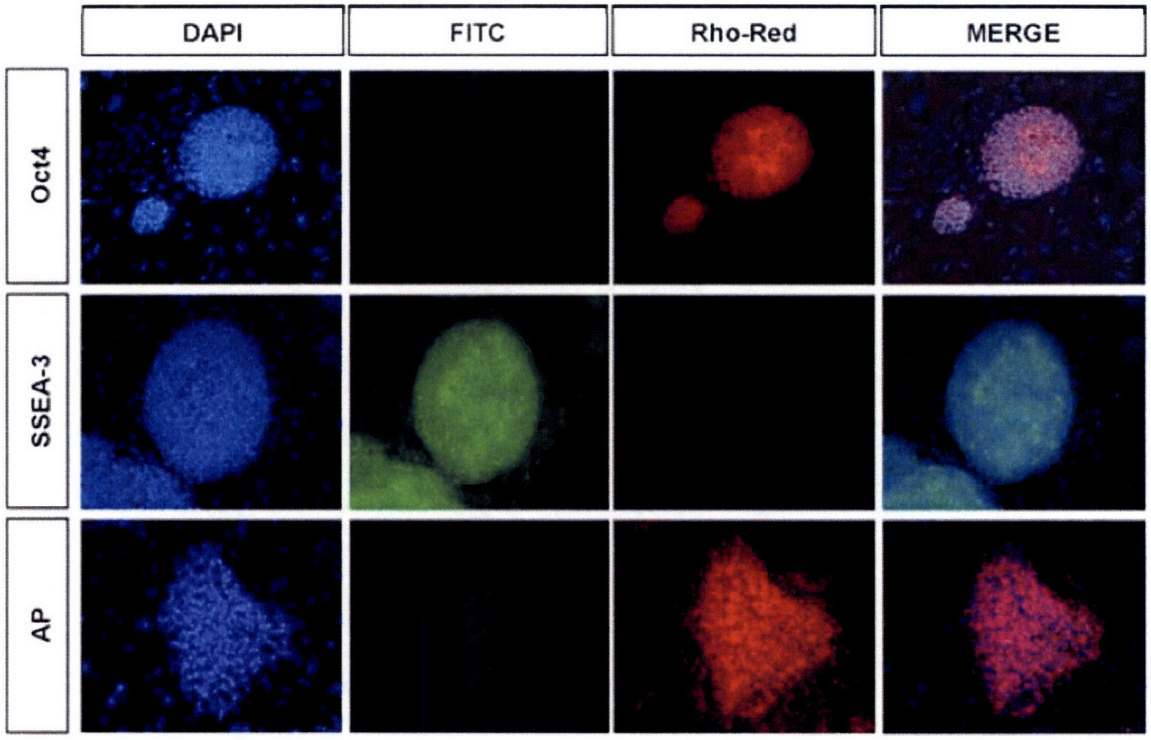
Table S15. Developmental regulators associated with PRC2 in ES cells and muscle.
http://web.wi.mit.edu/young/hES_PRC/TableS15.xls

Figure S1. Human H9 ES cells cultured on a low density of irradiated murine embryonic fibroblasts



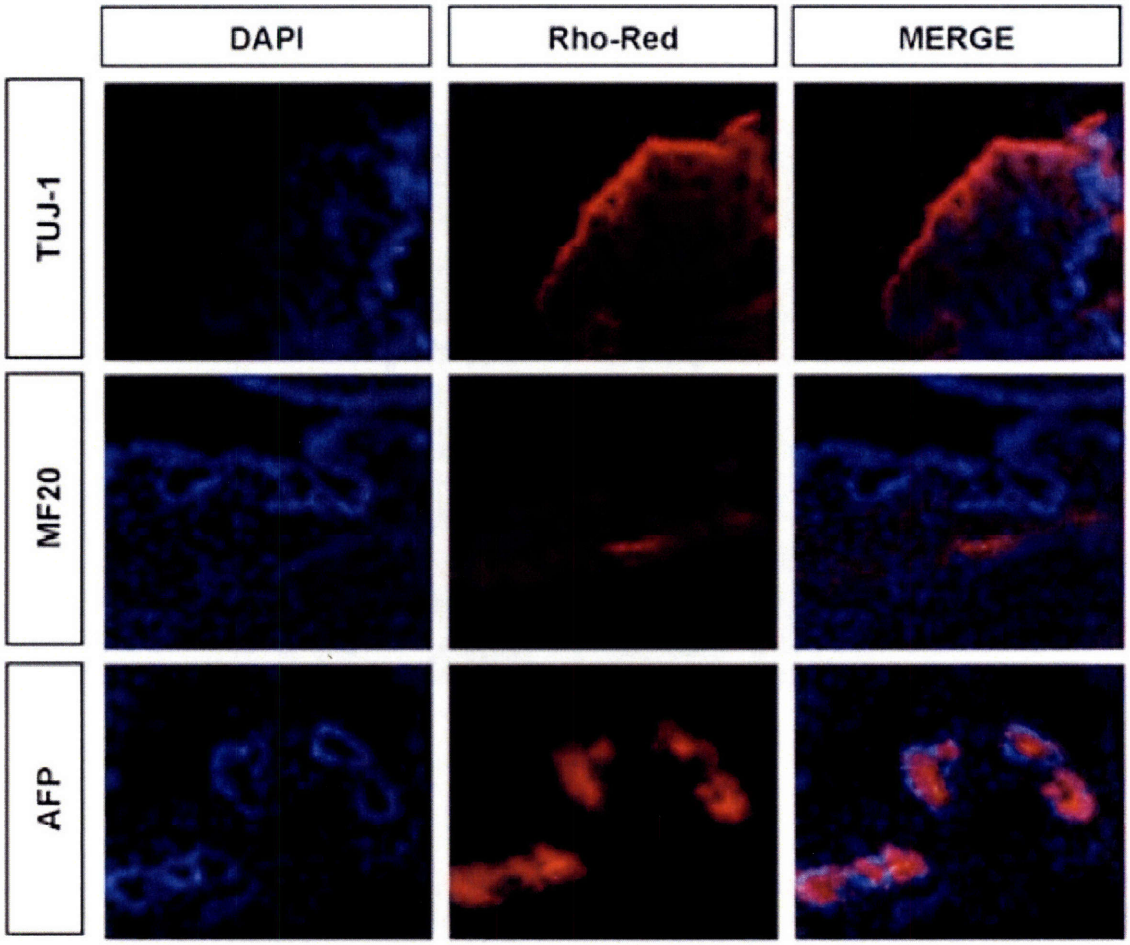
Bright-field image of H9 cell culture.

Figure S2. Analysis of human ES cells for markers of pluripotency



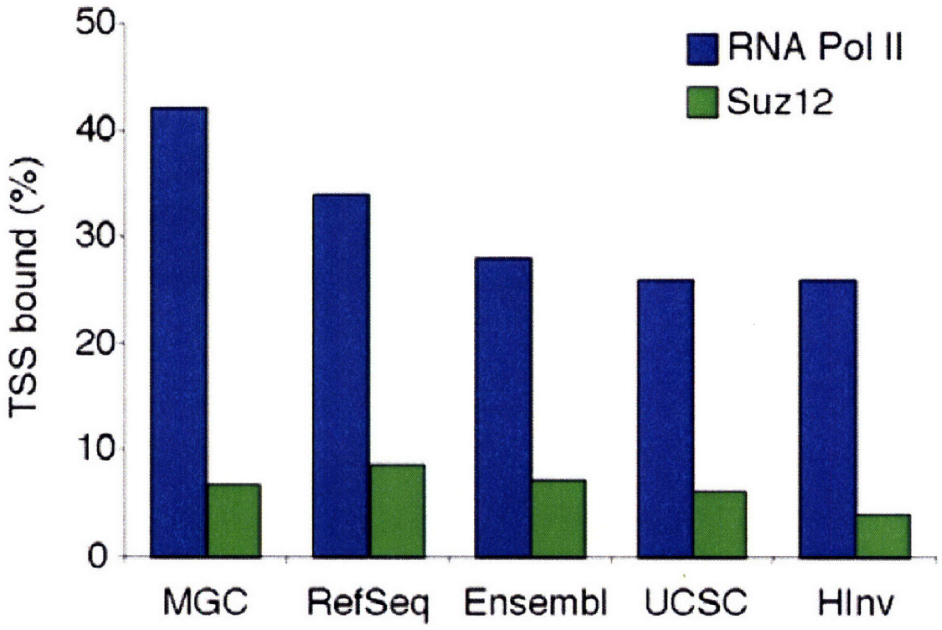
Human embryonic stem cells were analyzed by immunohistochemistry for the characteristic pluripotency markers Oct4 and SSEA-3. For reference, nuclei were stained with DAPI. Our analysis indicated that >90% of the ES cell colonies were positive for Oct4 and SSEA-3. Alkaline phosphatase activity was also strongly detected in human ES cells.

Figure S3. Analysis of human ES cells for differentiation potential



Teratomas were analyzed for the presence of markers for ectoderm (Tuj1), mesoderm (MF20) and endoderm (AFP). For reference, nuclei are stained with DAPI. Antibody reactivity was detected for derivatives of all three germ layers confirming that the human embryonic stem cells used in our analysis have maintained differentiation potential.

Figure S4. The fraction of annotated promoters bound by RNA polymerase II or Suz12



The fraction of unique gene transcription start sites that lie within 1 kb of a genomic region bound by RNA polymerase II and Suz12. The total number of start sites in each database is as follows: MGC n=17,188; RefSeq n=19,349; Ensembl n=30,121; UCSC Known Genes n=42,160; H-Inv n=42,777.

Figure S5. Estimating error rates

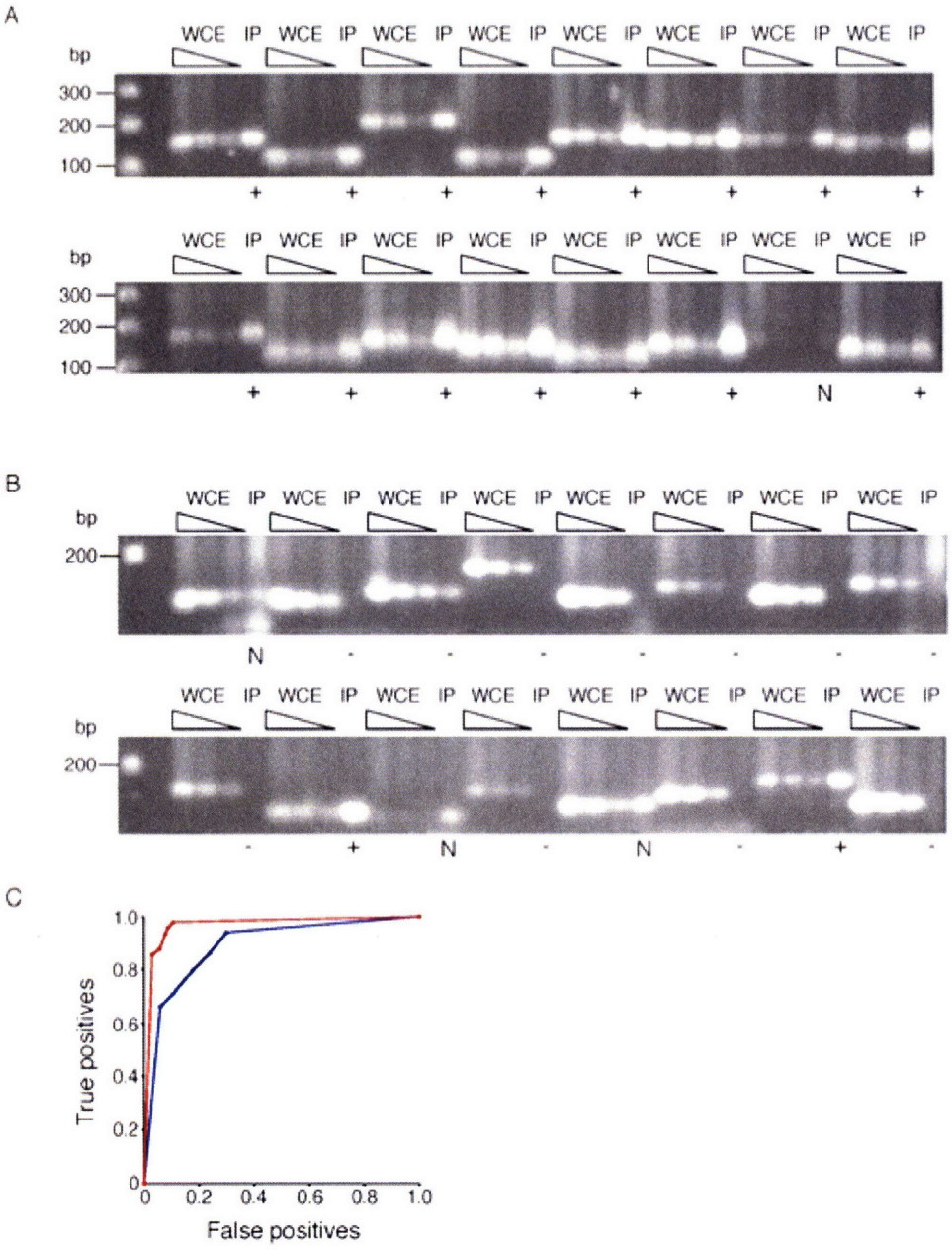


Figure S5

A. Example gel images showing PCR products amplified from 16 genomic regions judged to be bound by RNA polymerase II using the whole-genome arrays. Each primer-pair was used to amplify unenriched, whole cell extract (WCE) DNA (90, 30 and 10 ng) and immunoenriched (IP) DNA (10 ng). Enrichment in the IP DNA is indicated by a "+" and a lack of enrichment by a "-". PCR reactions judged to be inconclusive were labeled with an "N".

B. Example gel images showing PCR products amplified from 16 genomic regions judged not to be bound by RNA polymerase II using the whole-genome arrays. Each genomic region represents an annotated transcription start site.

C. Receiver-operator curve for RNA polymerase II binding in human ES cells. Curve compares percentage of true positives and false positives in binding events called from ChIP/chip compared to RT-PCR amplifications of anti-Pol II ChIP DNA. ROC curves were determined for all regions of the genome (blue) and for the subset of regions located within 1 kb of known transcription start sites (red).

Figure S6. Co-occupation of gene promoters by Suz12, Eed and H3K37me3

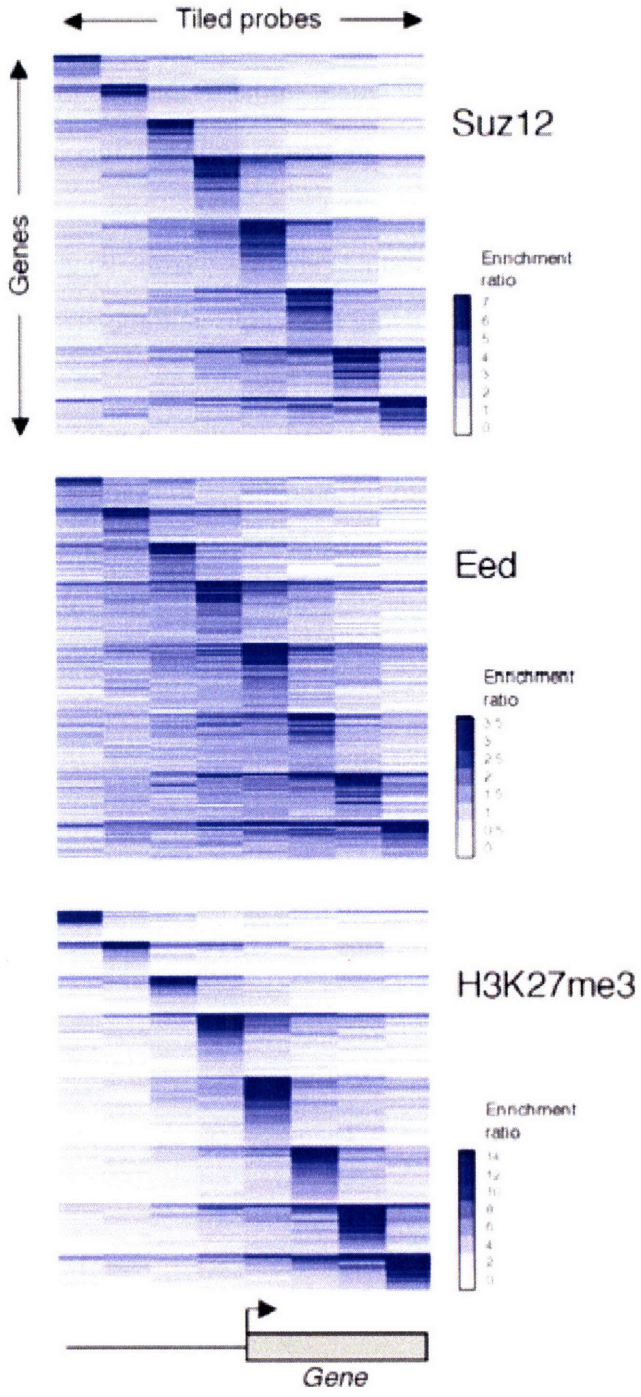
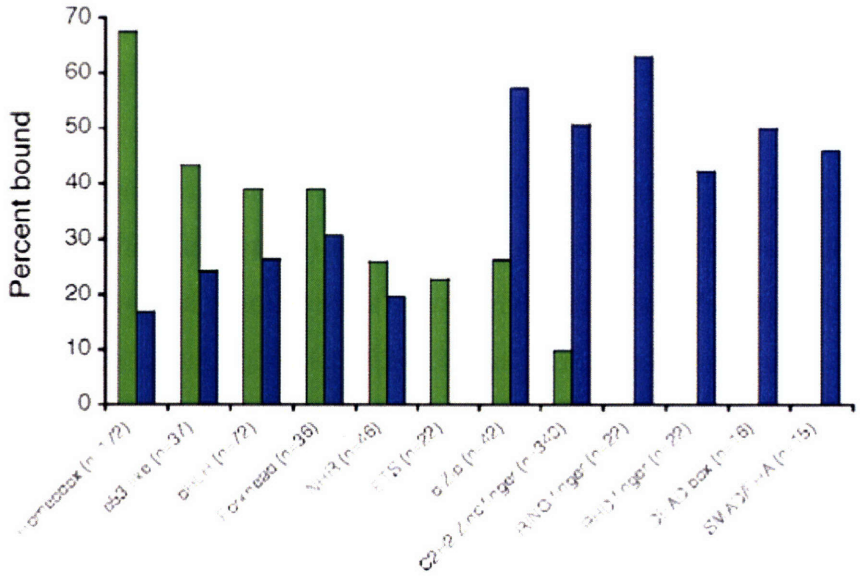


Figure S6

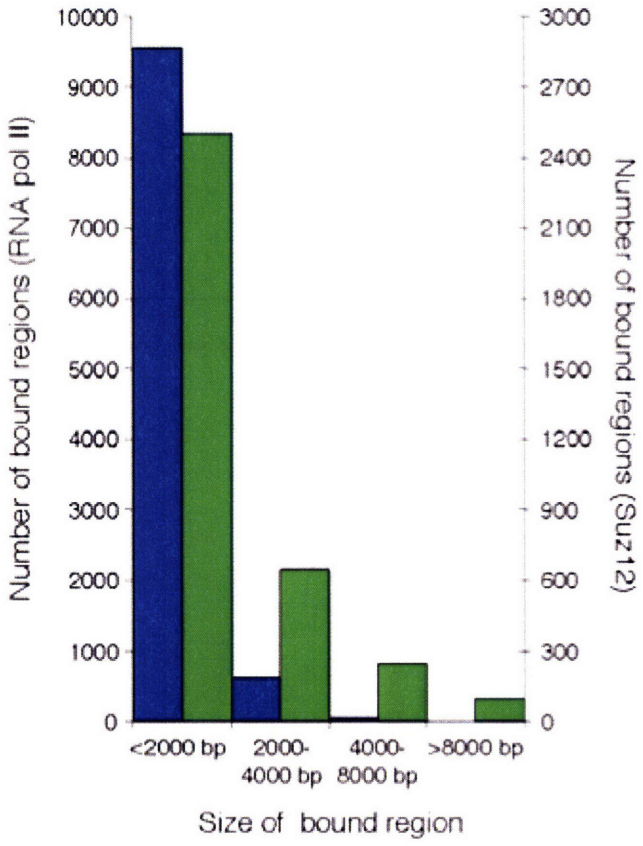
Suz12 occupancy (top panel), Eed occupancy (middle panel) and H3K27me3 occupancy (bottom panel) at transcription start sites. Each row represents a gene considered occupied by either Suz12, Eed or H3K27me3 using our high-confidence gene calling algorithm (see sections on Data Normalization and Analysis and Identification of Bound Regions). The same genes are illustrated in each of the three panels. Each column represents the data from an oligonucleotide probe positioned relative to the start site as indicated by the gene diagram below. The log binding ratios for each oligo are plotted for each protein; blue indicates enrichment of the immunoprecipitated factor (enrichment ratio >1). A scale for the binding ratios for each panel is shown. Each factor follows the same binding pattern. From this we conclude that Suz12, Eed and H3K27me3 are present at essentially the same set of genes and that our stringent gene calling algorithm sometimes calls a gene bound by one factor but not another factor because of the inherent false negative rate of ~30% (see Estimating Error Rates section).

Figure S7. Protein domain classification of Suz12- and Pol II-bound transcription factors



Fraction of transcription factor categories bound by Suz12 (green) or RNA Pol II (blue). The percentage is expressed relative to all transcription factor genes assigned to that category by InterPro domain (PANDORA) annotation at the default resolution. Abbreviations are b-HLH (basic helix-loop-helix), NHR (nuclear hormone receptor), ETS (erythroblast transformation specific), b-Zip (basic leucine zipper), PHD finger (plant homeodomain finger), SMAD (Sma- and Mad-related) and FHA (forkhead-associated). n indicates the number of transcription factor genes assigned to a given category.

Figure S8. Suz12 occupies large regions of DNA



Number of RNA polymerase II (blue bars, left hand axis) and Suz12 (green bars, right hand axis) bound regions of certain sizes (x axis). Unlike RNA polymerase II, Suz12 occupies over 2 kb of sequence at a significant number of genes.

Figure S9. H3K27me3 co-occupies large domains with Suz12

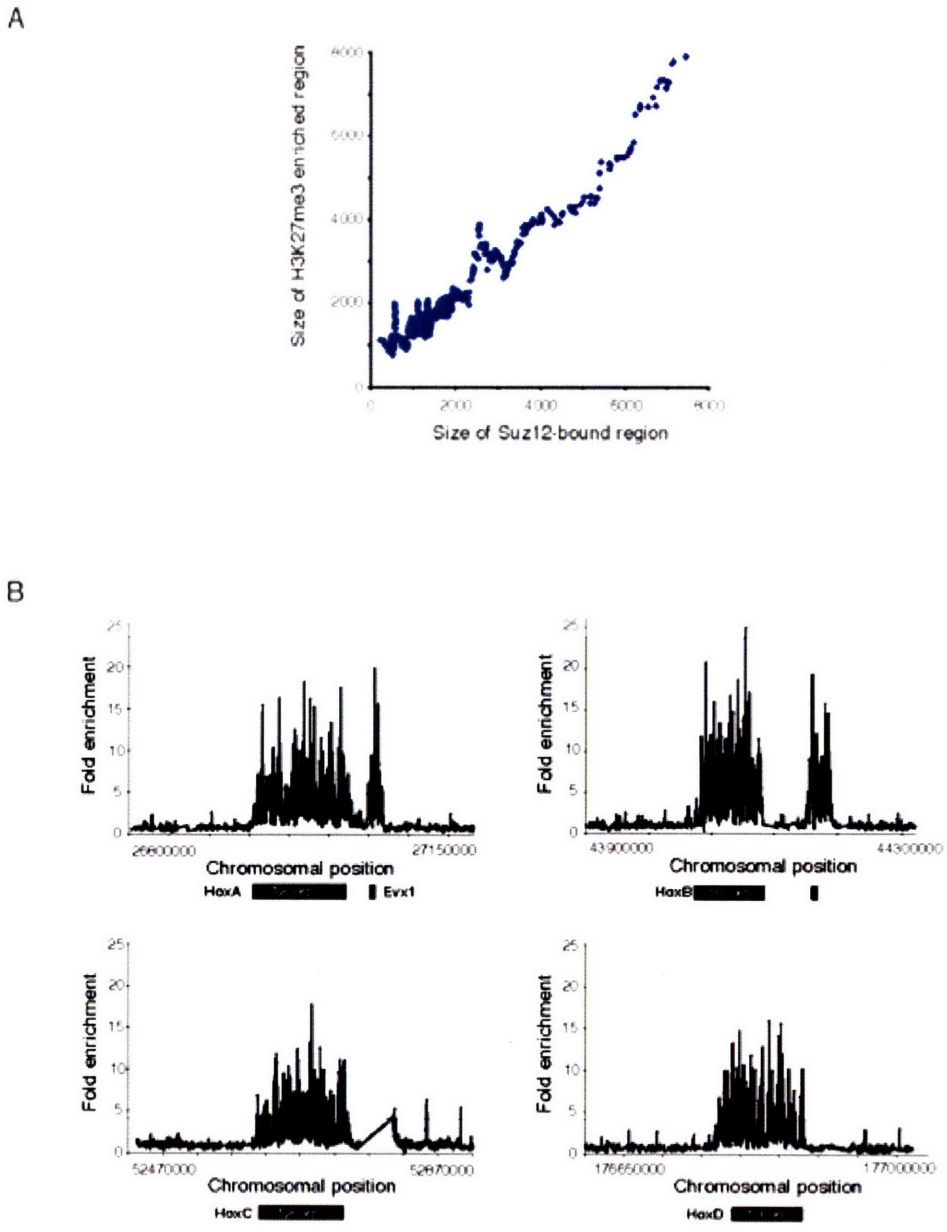
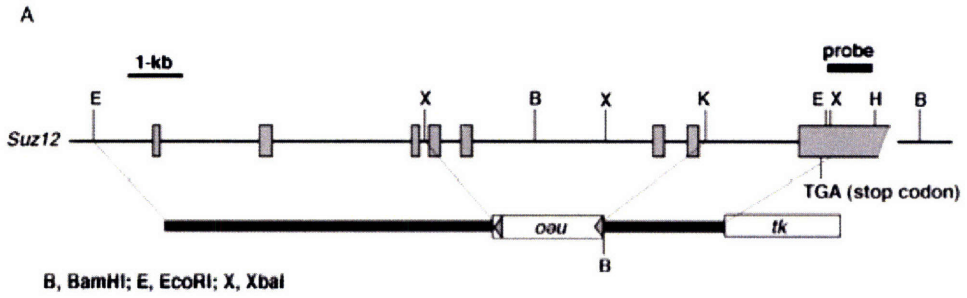


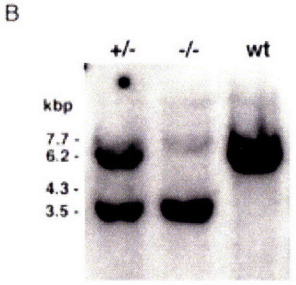
Figure S9

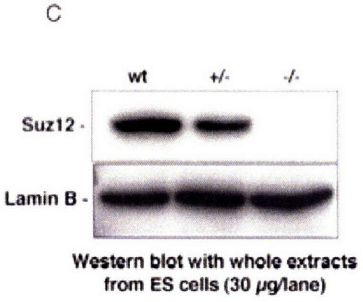
A. Correlation between size of domains of Suz12 binding and H3K27me3 binding. The trend was calculated by computing the moving average of the size of H3K27me3 regions using a sliding window of 20 genes across the set of genes bound by Suz12 and H3K27 and ordered by size of Suz12 bound region. Sizes of bound regions were calculated from promoter arrays.

B. Binding profile of H3K27me3 (black) across ~500 kb regions encompassing Hox clusters A-D. Unprocessed enrichment ratios for all probes within a genomic region are shown (ChIP vs. whole genomic DNA). Approximate Hox cluster region sizes are indicated within black bars.

Figure S10. Generation of Suz12 -/- cells







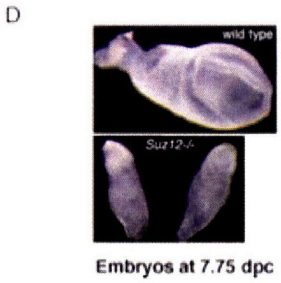


Figure S10

A. Targeted deletion of the Suz12 locus. Homologous recombination was used to replace the 5' portion of Suz12 with a neo cassette. Location of probe used for southern blot verification in (b) is shown. Restriction enzymes are denoted B, BamH1; E, EcoR1; X, Xba1.

B. Southern blot analysis of BamH1 digested genomic DNA from each genotype.

C. Western blot analysis of whole cell extracts derived from each genotype. Immunoblots were probed with anti-Suz12 (top) or anti-Lamin B (bottom).

D. Embryos generated from Suz12 heterozygous crosses were analyzed at different stages of development. At 7.75 dpc, normal as well as morphologically smaller embryos were evident. Genotyping analysis indicated that the abnormal embryos were homozygous for the Suz12 null allele confirming that Suz12 is required for early development.

Figure S11. Binding of Suz12 in differentiated muscle

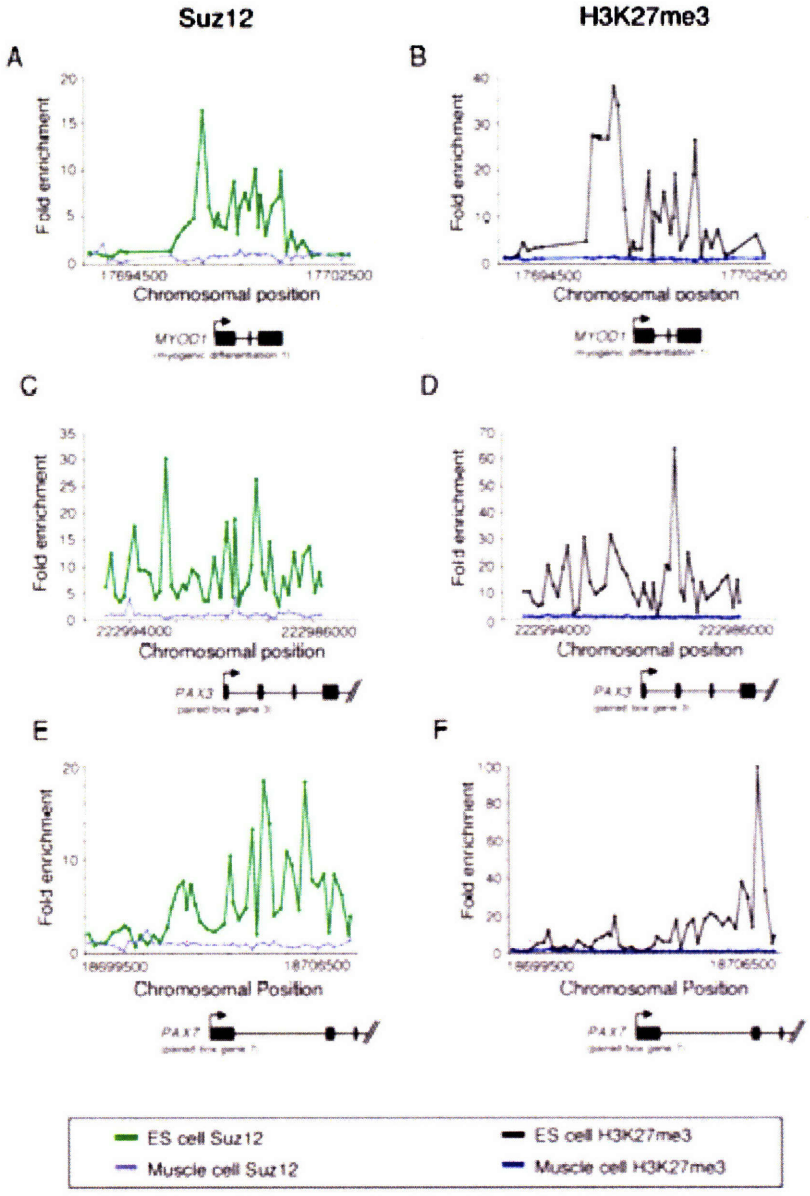


Figure S11

A. Suz12 binding profiles across the muscle regulator MYOD1 gene in H9 human ES cells (green) and differentiated myotubes (grey). The plots show unprocessed enrichment ratios for all probes within a genomic region (ChIP vs. whole genomic DNA). Genes are shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.

B. H3K27me3 profiles across the muscle regulator MYOD1 gene in H9 human ES cells (black) and differentiated myotubes (blue). The plots show unprocessed enrichment ratios for all probes within a genomic region (ChIP vs. whole genomic DNA). Genes are shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.

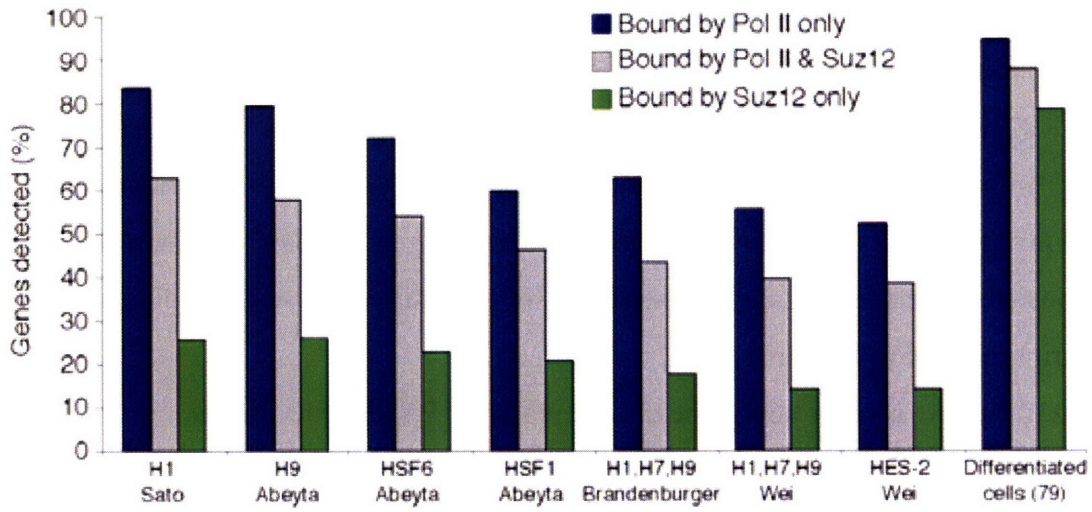
C. Suz12 binding profiles across the muscle regulator PAX3 gene, as in **a.**

D. H3K27me3 profiles across the muscle regulator PAX3 gene, as in **b.**

E. Suz12 binding profiles across the muscle regulator PAX7 gene, as in **a.**

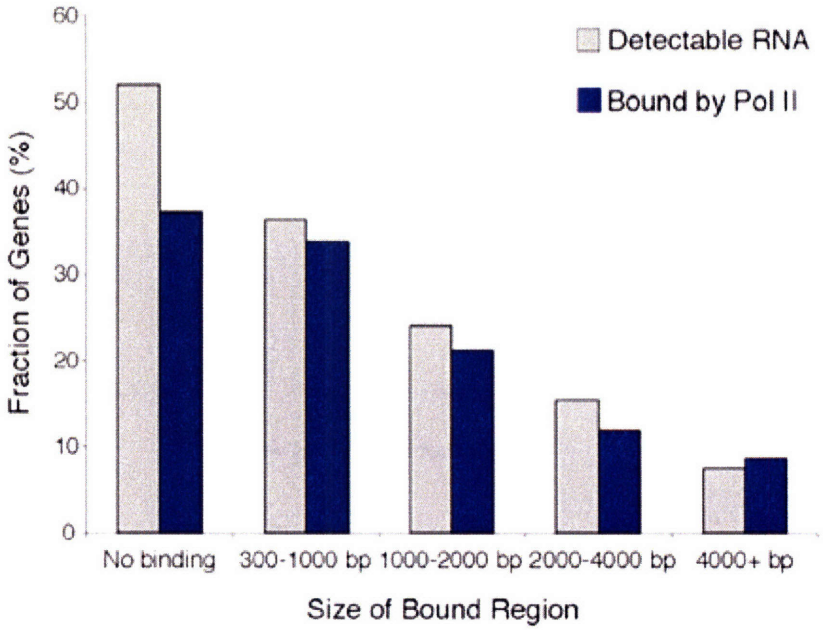
F. H3K27me3 profiles across the muscle regulator PAX7 gene, as in **b.**

Figure S12. Detection of genes bound by RNA polymerase II and Suz12 in human ES cell expression datasets



Percentages of genes that are bound by RNA polymerase II only, RNA polymerase II and Suz12, and Suz12 only that are detected in 7 ES cell expression datasets and one differentiated cell expression dataset. The first four ES cell datasets and the differentiated cell dataset were generated using gene expression arrays H1: U133A arrays; H9, HSF1 and HSF6: U133A+B arrays; differentiated tissues: U133A array. The percentages are relative to the fraction of bound genes that are represented on the arrays. The last three ES cell datasets were generated using MPSS.

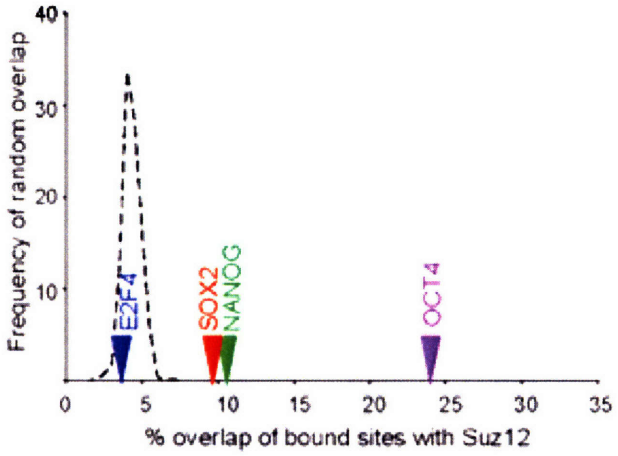
Figure S13. Relationship between size of Suz12 and RNA polymerase II co-occupancy and gene expression



The percentage of genes with detectable RNA (grey bars) and associated with RNA polymerase II (blue bars) as a function of the extent of Suz12 binding. The frequencies for genes not bound by Suz12 are indicated on the left as controls.

Figure S14. Association of Oct4, Sox2 or Nanog with Suz12-bound regions

A



B

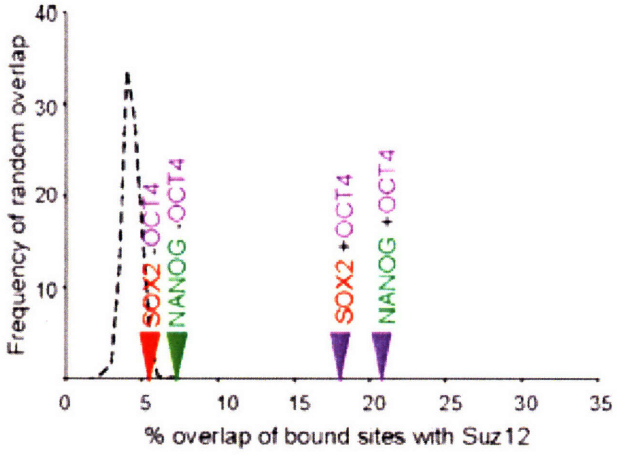


Figure S14

A. The percentage of Oct4-bound regions (purple arrow), Sox2-bound regions (red arrow) or Nanog-bound regions (green arrow) that overlap with Suz12-bound regions are shown along the x-axis. Comparisons were made between promoter array data from Boyer et al., 2005 and whole genome Suz12 data presented here. The dashed line indicates the distribution of the expected overlap based on randomized data. For comparison, we also show the results for a fourth transcription factor, E2F4 (blue arrow).

B. The percentage of Sox2 and Oct4-cobound regions or Nanog and Oct4-bound regions (purple arrows) that overlap with Suz12-bound regions are shown along the x-axis. Comparisons were made between promoter array data from Boyer et al., 2005 and whole genome Suz12 data presented here. The dashed line indicates the distribution of the expected overlap based on randomized data. For comparison, we also show the results for Sox2-bound regions that are not bound by Oct4 (red arrow) and Nanog-bound regions that are not bound by Oct4 (green arrow).

Figure S15. Motifs associated with DNA regions that are bound by Oct4, Sox2, Nanog and Suz12 or bound by Oct4, Sox2 and Nanog

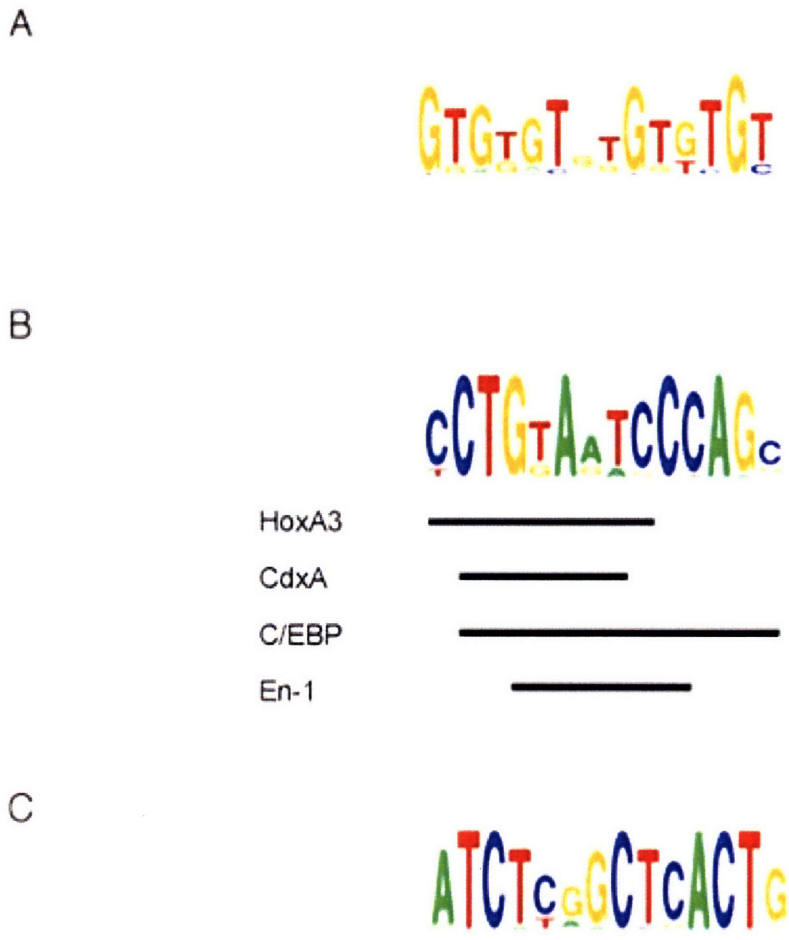


Figure S15

A. Consensus sequence of a motif associated with DNA regions bound by Oct4, Sox2, Nanog and Suz12. This motif was found in approximately 50% of the regions bound by Oct4, Sox2, Nanog and Suz12 and enriched 4.8-fold compared to regions bound by Oct4, Sox2 and Nanog but not Suz12.

B. Consensus sequence of a motif associated with DNA regions bound by Oct4, Sox2 and Nanog and not bound by Suz12. This motif was found in approximately 20% of the regions bound by Oct4, Sox2, Nanog and enriched 3.0-fold compared to regions bound by Oct4, Sox2, Nanog and Suz12. Putative transcription factor binding sites are labeled and indicated by black lines. Binding sites were identified with P-Match (http://www.gene-regulation.com) using the input sequence CCTGTAATCCCAGC and cut-off selection for matrix group to minimize the sum of false positives and negatives.

C. Consensus sequence of a motif associated with DNA regions bound by Oct4, Sox2 and Nanog and not bound by Suz12. This motif was found in approximately 15% of the regions bound by Oct4, Sox2, Nanog and enriched 2.4-fold compared to regions bound by Oct4, Sox2, Nanog and Suz12. No putative transcription factor binding sites were identified when examined as described in **b.** using the input sequence ATCTCGGCTCACTG. More lenient selections for the cut-off selection indicate potential binding sites for C/EBP, HoxA3, CdxA, Msx-1 and v-Myb.

Appendix D

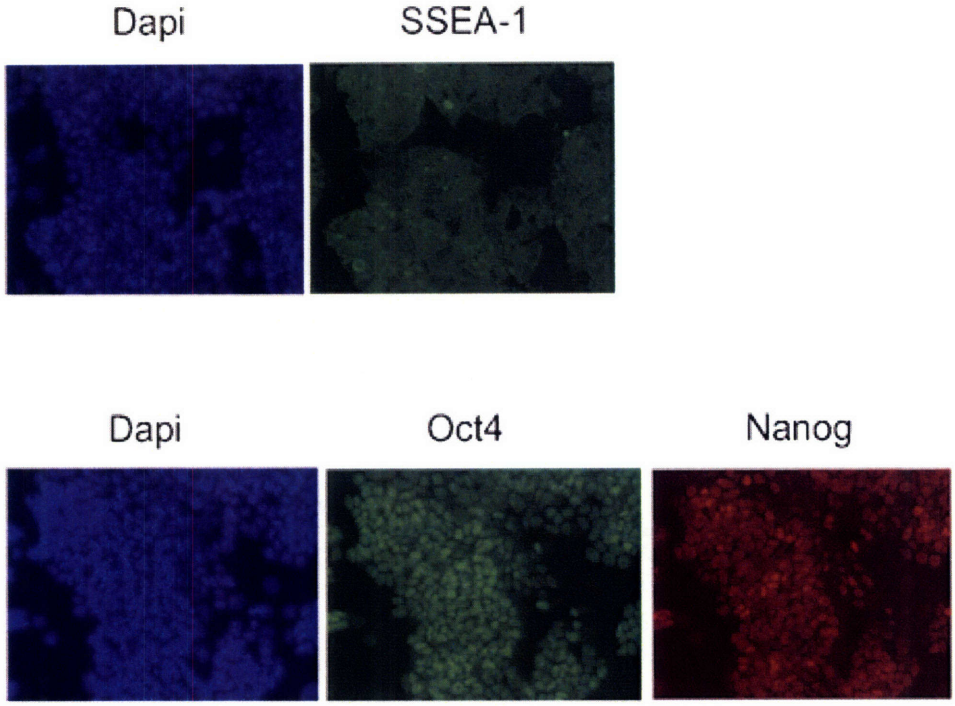
Supplementary Data for Chapter 4

Supplemental Data and Tables

All supplementary data and Tables S1-S3 can be found at the following website:

http://www.genesdev.org/cgi/content/full/22/6/746/DC1

Figure S1. Staining of v6.5 murine embryonic stem cells for pluripotency markers SSEA-1, Oct4 and Nanog



Cells were fixed with 4% paraformaldehyde and stained overnight at 4°C with SSEA-1 (DSHB, The University of Iowa, MC-480,1:10), Oct4 (Santa Cruz, SC- 5279, 1:200), Nanog (Abcam, ab1603, 1:250 dilution) and DAPI Nucleic Acid Stain (Invitrogen D1306; 1:10000 dilution). After several washes cells were incubated with goat anti-mouse conjugated Alexa Fluor 488 (Invitrogen 1:200 dilution) or goat anti-rabbit conjugated Alexa Fluor 568 (Invitrogen1:200 dilution) for 2 hours at room temperature. Stained cells were visualized at 10x magnification.

Figure S2. Tcf3 bound regions are in close proximity to Oct4 and Nanog bound regions.

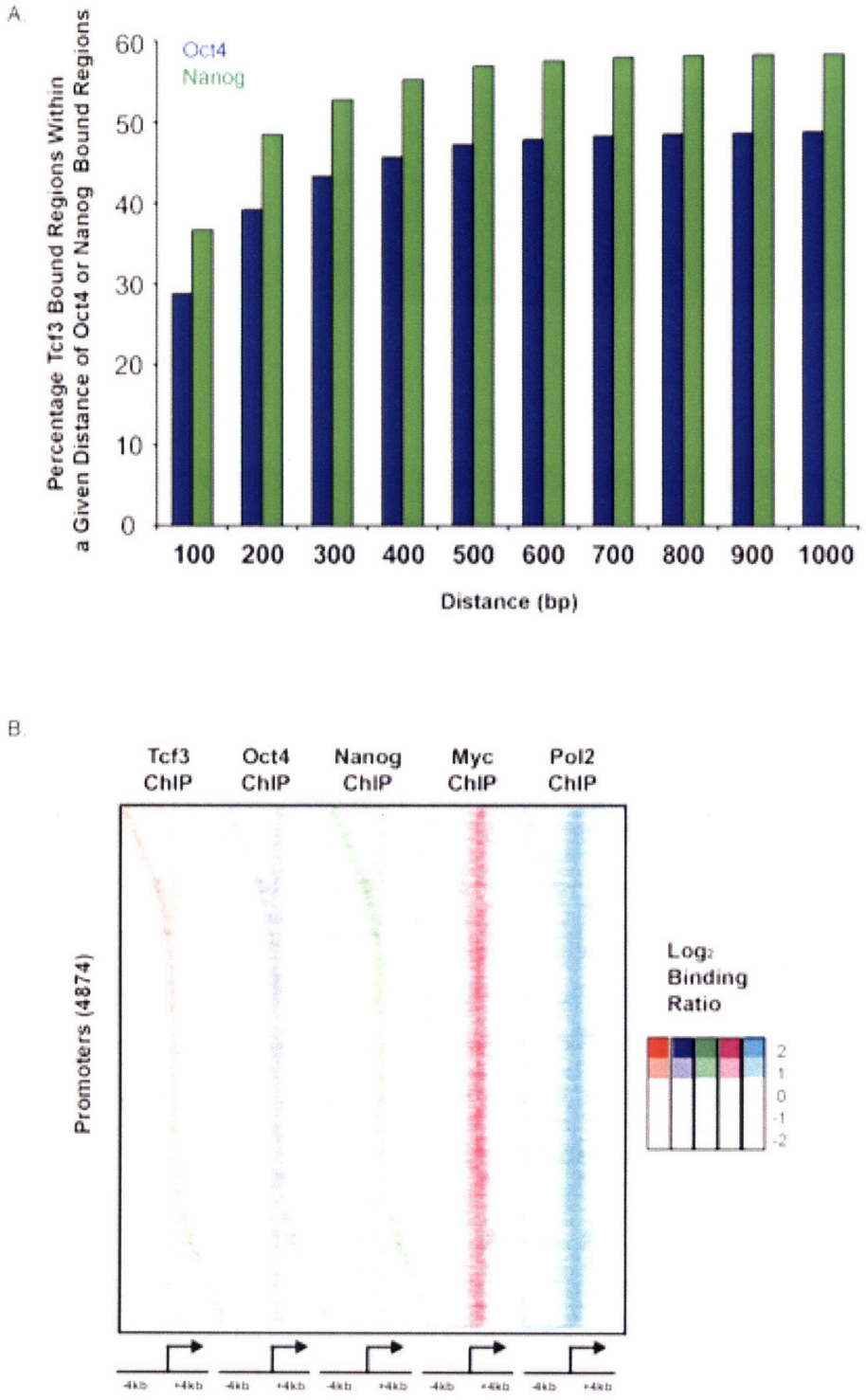
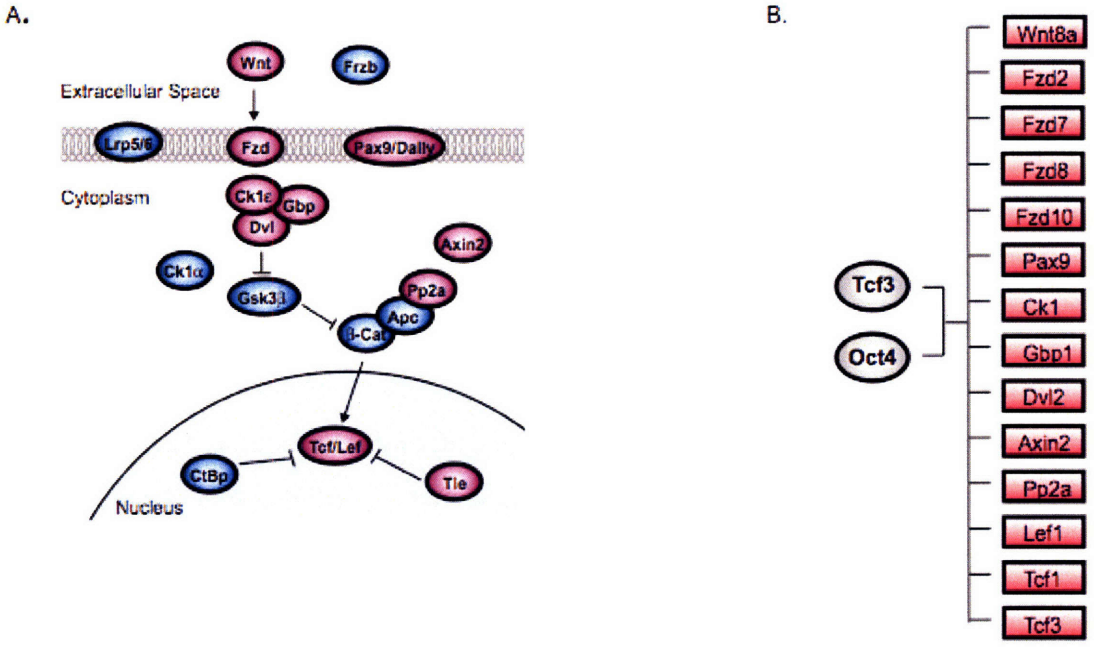


Figure S2

A. A histogram of the distance between each Tcf3 bound region to the nearest Oct4 or Nanog bound region. Distances were calculated as the distance in base pairs between the midpoints of bound regions.

B. Tcf3, Oct4 and Nanog display nearly identical binding profiles. Binding profiles of ChIP-chip data are shown for Tcf3 (Santa Cruz sc-8635), Oct4 (Santa Cruz sc-8628), Nanog (Bethyl Labs A300-398A), Myc (Santa Cruz sc-764) and PolII (Abcam ab817). Analysis of ChIP-chip data from genes bound by Tcf3, Oct4, Nanog or Myc reveals that Tcf3, Oct4 and Nanog have a distinct binding profile from Myc. Regions from –4kb to +4kb 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 bind center then no value was assigned. For genes with multiple promoters, each promoter was used for analysis. Promoters are organized according to the distance between the maximum Tcf3 binding ratio and the maximum PolII binding ratio. Array data used was from Agilent mouse promoter arrays that cover +/- 4kb from the TSS at a density of approximately 1 probe every 250bp.

Figure S3. Tcf3 and Oct4 bind components of the Wnt signaling pathway



A. Wnt signaling pathway diagram. Genes bound by both Tcf3 and Oct4 are shown in purple.

B. Genes encoding Wnt pathway components that are bound by both Tcf3 and Oct4. Proteins are represented by ovals, and genes are represented by rectangles.

Figure S4. Stimulation of the Wnt pathway

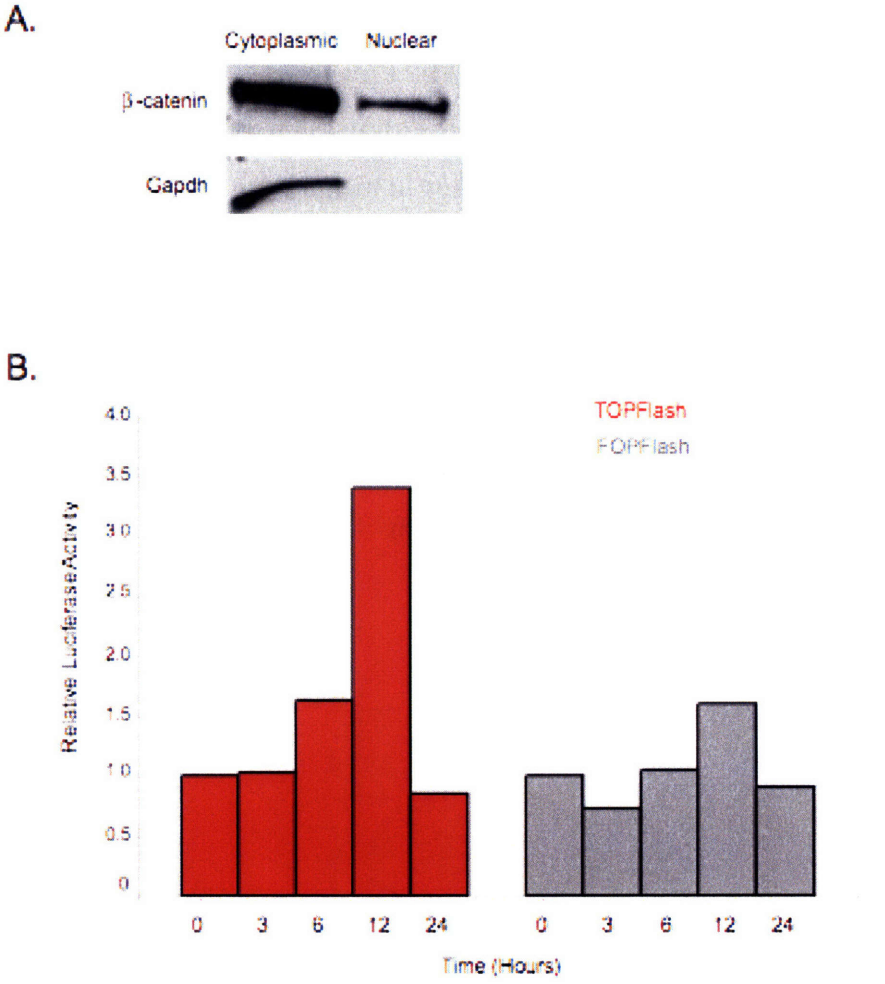
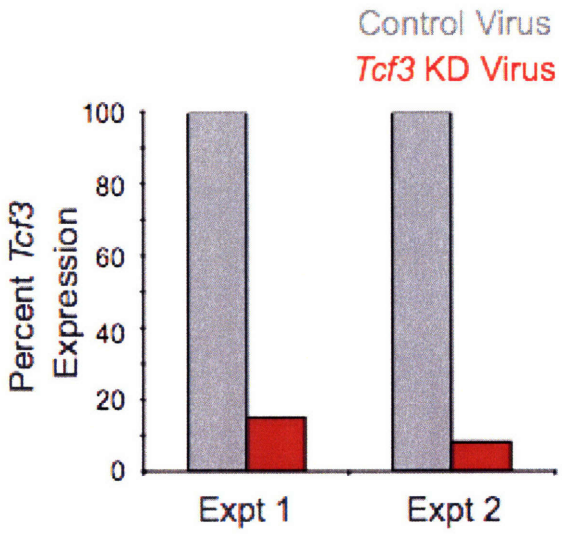


Figure S4

A. There is active β-catenin in mES cells under standard culture conditions. β-catenin was detected by western in both the cytoplasm and the nucleus of mES cells. Gapdh was used as a control to ensure the purity of the nuclear protein prep (β-catenin Cell Signaling 9562 1:1000; Gapdh Santa Cruz 25778 1:1000). Protein was collected using Sigma Cell Lytic C2978 and Sigma Cell Lytic NuCLEAR Extraction Kit NXTRACT.

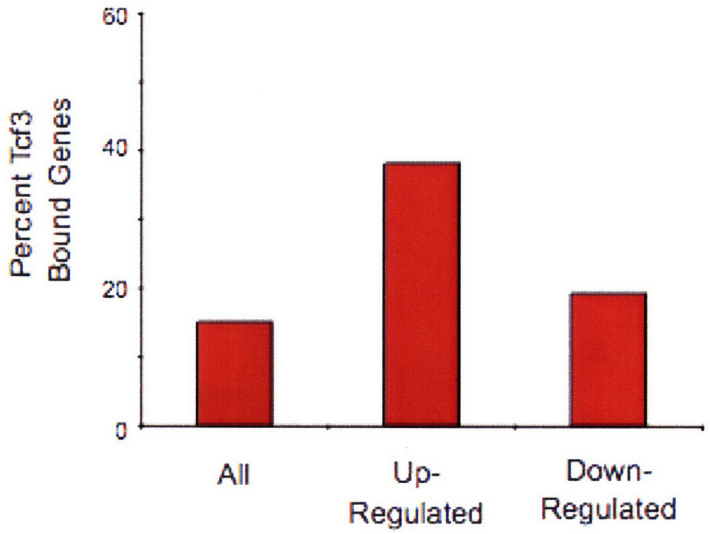
B. Wnt3a conditioned media or mock conditioned media were used to stimulate the Wnt pathway. Pathway activity was measured using a TOPFlash reporter plasmid kit (Upstate 17-285). Murine ES cells were transfected with either TOPFlash or FOPFlash, a negative control plasmid, along with a Renilla control plasmid using FuGENE HD Transfection Reagent (Roche 04709691001). Luciferase activity was determined using the Dual-Luciferase Reporter Assay System (Promega E1910) and measurements were normalized to the Renilla control and a mock control medium.

Figure S5. *Tcf3* knockdown virus reduces *Tcf3* transcript



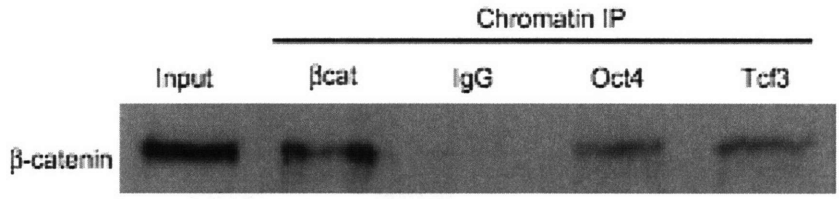
mES cells were infected with lentivirus carrying either a hairpin to knockdown *Tcf3* or a control non-silencing hairpin. Infected cells were allowed to grow for 48 hours and cells were harvested to collect RNA. Real-time PCR demonstrated that cells infected with *Tcf3* KD virus have decreased levels of *Tcf3* expression. Values were normalized to *Gapdh* transcript levels.

Figure S6. Genes with significant expression changes upon *Tcf3* knockdown are enriched for Tcf3 binding



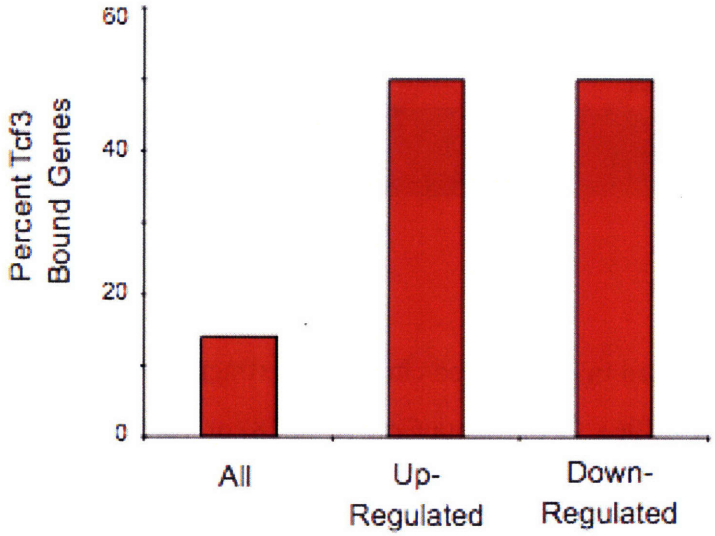
The graph represents the percent of genes with a greater than two-fold increase or decrease in expression that are bound by Tcf3.

Figure S7. β-catenin associates with Tcf3 and Oct4 in mES cells



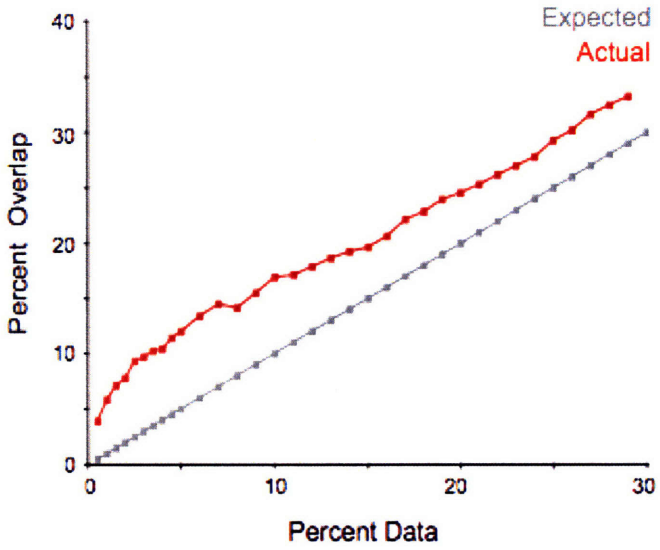
β-catenin could be detected in crosslinked chromatin extracts immunoprecipitated for β-catenin (Santa Cruz sc-1496), Oct4 (Santa Cruz sc-8628) or Tcf3 (Santa Cruz sc8635). Non-immunoprecipitated extract was used as a positive control and an immunoprecipitation against IgG (Upstate 12-371) was used as a negative control. Samples were prepared following the ChIP-chip protocol through reversal of crosslinks and then loaded onto a 7.5% acrylamide gel. Immunoblot was probed with β-catenin (Upstate 06-734 1:1000).

Figure S8. Genes with significant expression changes upon Wnt induction are enriched for Tcf3 binding



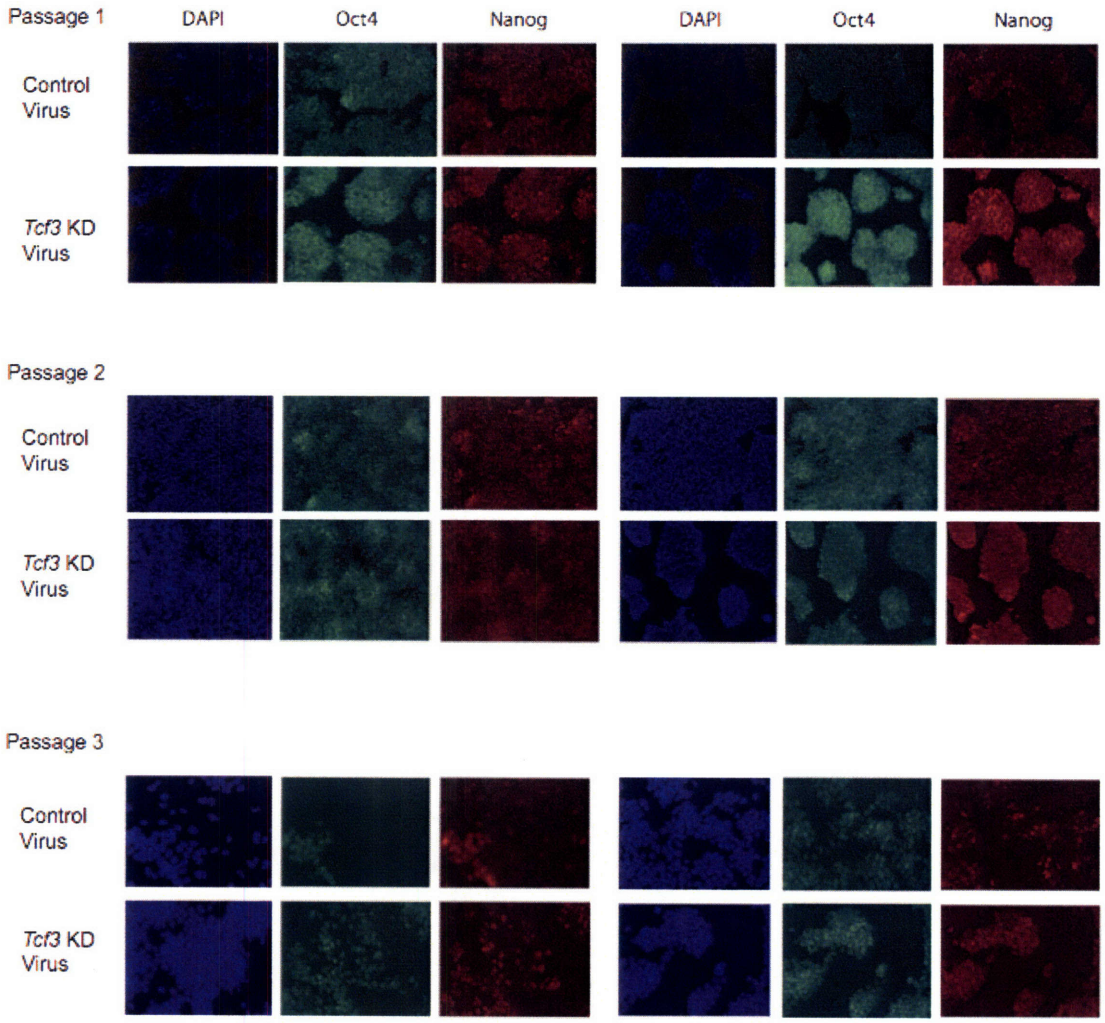
The graph represents the percent of genes with a greater than two-fold increase or decrease in expression that are bound by Tcf3.

Figure S9. Expression datasets from *Tcf3* KD cells and cells treated with Wnt CM significantly correlate



Expression data was rank ordered according to fold expression change. Various cutoffs for the percentage of genes to include for analysis were used, and results from each cutoff are graphed along the x-axis. The top ranking up-regulated genes from each dataset were compared, and the percent of genes overlapping is graphed on the y-axis. The expected line represents the expected overlap percentages assuming no correlation between datasets.

Figure S10. ***Tcf3*** **knockdown cells display stronger staining for Oct4 and Nanog through several passages**



Cells were passaged off of feeders in the presence or absence of LIF. Cells were fixed and stained as described in experimental procedures.

Figure S11. Cells grown 1 passage off of feeders treated with Wnt3a CM display increased Oct4 staining

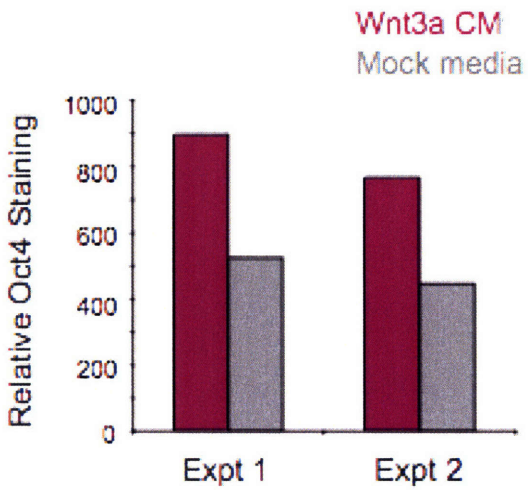


Image acquisition and data analysis was performed essentially as described in Moffat et al. (2006). Cells were grown for 3 days off of feeders in mES cell media supplemented with either Wnt3a CM or mock media. Cells were fixed and stained with Oct4 and Hoechst 33342 (1:1000 dilution). Stained cells were 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. 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.