Molecular Control of Embryonic Stem Cell Identity

by

Divya Mathur

B.A., Biochemistry
Mount Holyoke College, 2003

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

September 2008



Signature of Author……
Department of Biology
July, 2008

Certified by……
Rudolf Jaenisch
Professor of Biology
Thesis Supervisor

Accepted by……
Stephen Bell
Professor of Biology
Chairman, Biology Graduate Committee

## ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Rudolf Jaenisch for giving me the opportunity to work in his lab, and for his support during my graduate career. Special thanks go to my committee members, Phil Sharp and Tyler Jacks, whose support and wise words were really appreciated throughout my time at MIT. I would also like to acknowledge my collaborators, Laurie Boyer, David Gifford, Rick Young and Steve Carr for their guidance, and especially Tim Danford and Betty Chang for their contributions to a lot of the work described here. I received invaluable support from the microarray and BaRC facilities at the Whitehead throughout my time in the lab.

I am highly indebted to Caroline Beard for her incredible guidance, advice and encouragement, without which reaching this stage in the Ph.D. career would have been very difficult. Ruth Foreman has been a great collaborator and friend, and it has been comforting to share the highs and lows of grad school life with her. I would like to thank Kathrin Plath, Emi Giacometti, Konrad Hochedlinger, Jessie Daussman, Sandra Luikenhuis, Lucas Dennis, Tobias Brambrink, Suzanne Nyguen, Chris Lengner, Ruth Flannery, Lea Medeiros, Mathias Pawlak, Alex Meissner, Marius Wernig, Eveline Steine, Grant Welstead, and a number of other current and former members of the lab for their support and friendship.

The MIT experience would not have been the same without my biograd2003 classmates. I will cherish every moment that I spent with this wonderful group of colleagues, right from hanging out in “The Pit” to the highly entertaining movie nights and birthday celebrations. Brett Tompson and Susan Cohen regularly provided me with sensational MIT gossip, and the fun times we spent together will always be very special to me. My former roommates, Lena Khibnik, Yasemin Sancak and Chia Wu added to the great times I had in the last few years, and I have fond memories of all the fun we had together.

The love and support of my family and friends has been invaluable. My grandmother, who was a woman of great courage and integrity, continues to be a source of inspiration that has given me the strength to get through every challenge that I have faced. My grandfather’s incredibly dreadful sense of humor makes even the most insurmountable problem miraculously disappear, and his zest for life is infectious. I am very grateful for the friendship and support of Swaroop, who has tolerated my daily drama with immeasurable patience, and has taught to me to appreciate the finer things in life. I do not have enough words to express my gratitude for the encouragement, love and affection that I have received from my parents and sister. I can only say that no success in life would have been possible, or worth having without them.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

MOLECULAR CONTROL OF EMBRYONIC STEM CELL IDENTITY

by

Divya Mathur

Submitted to the Department of Biology on August 28, 2008, in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

ABSTRACT

Embryonic Stem (ES) cells are the *in vitro* derivatives of the inner cell mass of a developing embryo, and exhibit the property of pluripotency, which is the ability of a cell to give rise to all cell lineages of an organism. Therefore, these cells hold great promise in the treatment of several degenerative diseases through patient-specific cell-based therapy. Consequently, a detailed knowledge of the factors regulating ES cell identity is required in order to exploit this therapeutic potential. In order to address this subject, genome-wide location analysis (or ChIP-chip) has been used to identify downstream genes that are bound, and potentially regulated by the key pluripotency transcription factors, Oct4 and Nanog. The data from this study have also been compared and integrated with Oct4 and Nanog DNA binding data obtained in a different study using the ChIP-PET technology. In order to gain further insight into the mechanisms by which the transcription factor Nanog regulates its downstream targets, an attempt at identifying proteins interacting with Nanog has also been described.

Research on ES cells has been plagued with ethical controversies since the creation of these cells requires the destruction of embryos. Recent studies have reported the reprogramming of somatic fibroblasts into an ES cell-like induced pluripotent state (iPS) by virus-mediated transduction of four transcription factors— Oct4, Sox2, c-Myc and Klf4, thereby circumventing the use of embryos in producing pluripotent cells. In these studies, selection for the activation of the markers Oct4 or Nanog led to completely reprogrammed cells, but selection for *fbx15*, a downstream target of Oct4, resulted in partially reprogrammed intermediates. An unresolved issue in the field was whether these intermediates were obtained due to early drug selection in the case of *fbx15* selection, or because Fbx15 expression is not relevant to pluripotency. Drug selection for *fbx15* activation at later time-points, and an examination of the methylation status of the Oct4 locus of Fbx15-iPS cells suggests that the intermediates were obtained due to early drug selection and not due to selection for *fbx15.* Therefore, these studies have begun to elucidate a framework that governs ES cell identity, and the mechanism by which a differentiated cell can be reprogrammed into a pluripotent state.

Thesis Supervisor: Rudolf Jaenisch
Title: Professor of Biology

# Chapter 1

# Introduction

## (I) MOLECULAR CONTROL OF PLURIPOTENCY

Parts of this section have been adapted from a review: Boyer, L.A., Mathur, D. and Jaenisch, R. (2006). "Molecular Control of Pluripotency." Current Opinion in Genetics and Development **16**(5): 455-62.

### Developmental Potency and the Isolation of Embryonic Stem Cells

For more than a century, biologists have been fascinated with the subject of developmental potency, which is the capacity of a cell to give rise to other cell types. The most versatile cell is the fertilized egg or zygote, which is totipotent and can generate an entire organism along with the extra-embryonic tissues necessary for development. In the late 1800s, experiments done by Hans Driesch on sea urchins demonstrated that this property of totipotency is also present in the individual blastomeres of the 2- and 4-cell stage embryos (Driesch 1892). As better techniques to manipulate embryos evolved, further work on mammalian embryos showed that such potency is retained even in 8-cell stage blastomeres (Tarkowski and Wroblewska 1967). After this stage, the developmental potential of the cells in the embryo becomes more restricted as the outer cells generate the trophectoderm, an extra-embryonic tissue, and cells on the inside contribute to the inner cell mass (ICM), which subsequently forms the embryo proper (Ziomek and Johnson 1982; Ziomek, Johnson et al. 1982). The cells of the ICM are therefore pluripotent since they can generate all the cell lineages of an organism except the extra-embryonic tissues (Figure 1(a)).

A remarkable discovery that had tremendous impact on developmental genetics and medicine was the isolation of embryonic stem (ES) cells (Evans and Kaufman 1981; Martin 1981; Thomson, Itskovitz-Eldor et al. 1998). These cells are derived from the ICM of a developing blastocyst, and mimic the pluripotent abilities of their founder cells (Figure 1(a)). Additionally, they have the capacity to self-renew indefinitely *in vitro*, thereby offering a convenient model for studying early development. As further differentiation occurs during development, other types of stem cells emerge in the organism. Some of these, such as hematopoietic stem cells, are multipotent and can generate all the cell types within a specific lineage (Orkin and Zon 2008). Others, like spermatogonial stem cells, are unipotent since they can only form a particular type of cell (Cinalli, Rangan et al. 2008).

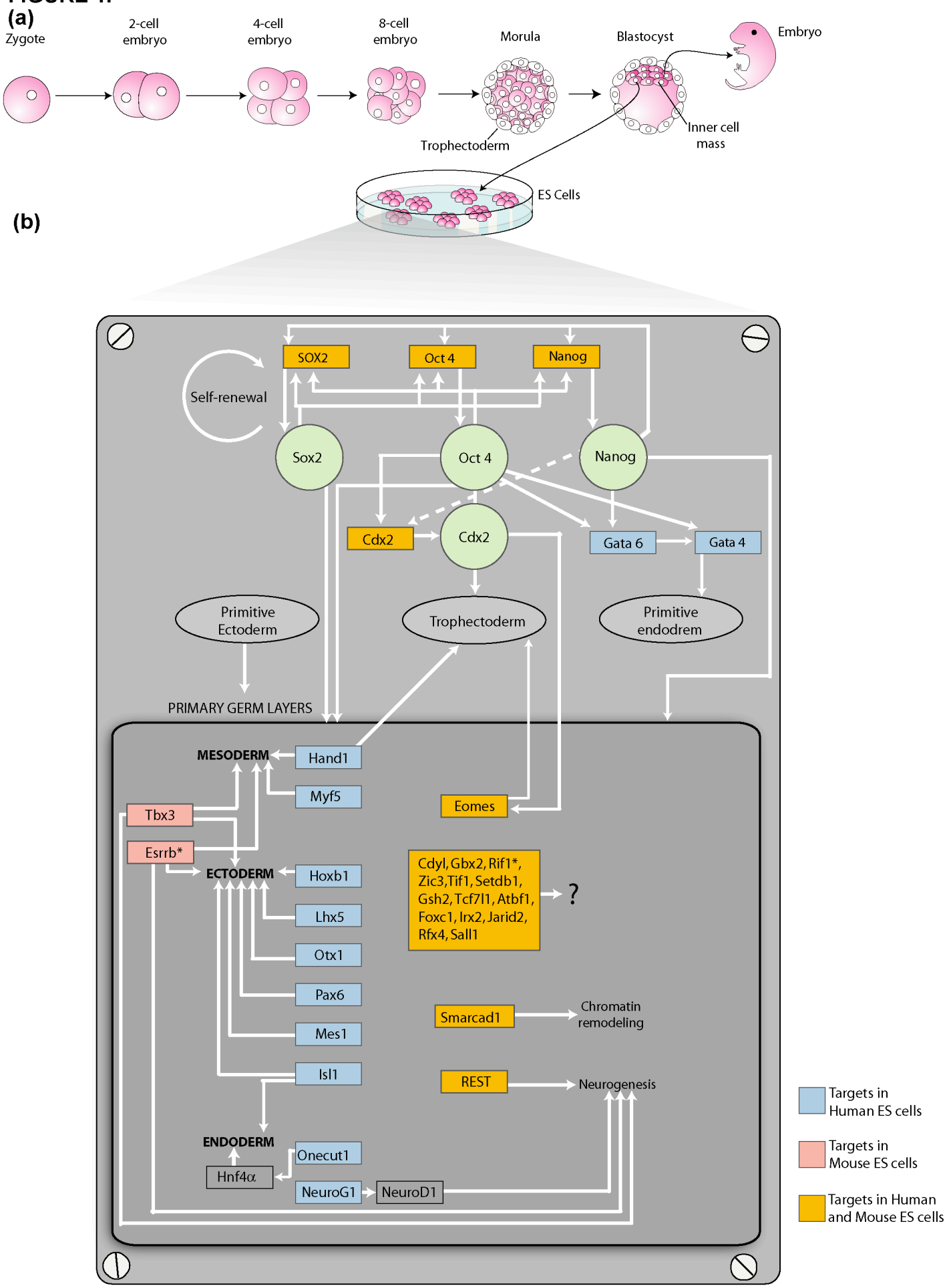
Even though specialized stem cells offer great benefits for therapeutic purposes, the pluripotent nature of ES cells gives them the great potential of being used in a wider range of regenerative therapies, along with being valuable tools in studying development and differentiation. For instance, ES cells have been used extensively in mammalian transgenics to study developmental phenomena and model human diseases. This technique involves targeting a DNA construct to specific genomic loci via homologous recombination in ES cells. These targeted ES cells can be injected into developing blastocysts, which are then implanted into a pseudo-pregnant female to generate chimeras. The chimeras can then be bred to wild-type strains to generate transgenic animals containing the targeted

construct in their germ line. This technique can also be exploited in the future to correct for various genetic disorders, such as sickle cell anemia. Additionally, since ES cells can give rise to any of the cell lineages in an organism, they may also be used in patient-specific cell-based therapies for treating degenerative disorders, while avoiding problems with immune-rejection. Aside from these applications, based on the similarity between these cells and those of the ICM, they hold great value as tools to study early development and lineage commitment *in vitro*.

Given the wide range of applications that ES cells can be used for, a detailed understanding of the mechanisms that enable propagation of these cells in a pluripotent state, poised to execute a broad range of developmental programs, is essential to realizing their therapeutic potential. In metazoans, the establishment and maintenance of lineage-specific gene expression programs are highly conserved throughout evolution and are vital for development (de la Serna, Ohkawa et al. 2006; Lin and Dent 2006). External environmental factors can also influence gene regulation (Smith 2001; Burdon, Smith et al. 2002; Boiani and Scholer 2005). It is of much interest to examine how genetic and epigenetic factors control ES cell identity and influence the balance between pluripotency and differentiation in mammals.

**Figure 1. Core Transcriptional Regulatory Circuitry in Pluripotent Mouse and Human ES Cell.** (a) Embryonic stem (ES) cells are derived from the pluripotent cells of the inner cell mass (ICM), which normally gives rise to the embryo. (b) Genomics studies have enabled the construction of a core transcriptional regulatory network in ES cells, initiated by Oct4, Sox2, and Nanog. This network reveals an integrated circuitry comprised of genes that specify the development of both the extraembryonic and embryonic lineages. Shown are a few examples of the circuitry components in the mouse and human studies. Boxes and circles indicate genes and proteins, respectively. Arrows represent interactions only, and not positive or negative effects. Genes for which the binding information with mouse Sox2 is available are marked with an asterisk.

**FIGURE 1.**


(a)
Zygote
2-cell embryo
4-cell embryo
8-cell embryo
Morula
Blastocyst
Embryo
Trophectoderm
Inner cell mass
ES Cells
(b)
SOX2
Oct 4
Nanog
Self-renewal
Sox2
Oct 4
Nanog
Cdx2
Cdx2
Gata 6
Gata 4
Primitive Ectoderm
Trophectoderm
Primitive endodrem
PRIMARY GERM LAYERS
MESODERM
Hand1
Myf5
Tbx3
Esrrb*
Eomes
ECTODERM
Hoxb1
Lhx5
Otx1
Pax6
Mes1
Isl1
Cdyl, Gbx2, Rif1*, Zic3, Tif1, Setdb1, Gsh2, Tcf7l1, Atbf1, Foxc1, Irx2, Jarid2, Rfx4, Sall1
?
Smarcad1
Chromatin remodeling
REST
Neurogenesis
ENDODERM
Onecut1
Hnf4α
NeuroG1
NeuroD1
Targets in Human ES cells
Targets in Mouse ES cells
Targets in Human and Mouse ES cells

## Genetic Control of Pluripotency in the Embryo and ES Cells

The homeodomain transcription factors Oct4 (also known as Pou5f1) and Nanog have been identified as crucial regulators of pluripotency and are predominantly expressed in pluripotent cell types. These factors also regulate preimplantation development in mammals, which is marked by two major differentiation events. The first of these occurs when the outer cells of the morula differentiate into the trophectoderm, which forms the chorion, the embryonic part of the placenta. This cell fate decision correlates with the expression of Oct4, since a loss of this transcription factor results in differentiation of all cells into the trophoblast lineage (Niwa, Miyazaki et al. 2000). The second differentiation event occurs in the early blastocyst, where certain cells from the ICM form the primitive endoderm, which forms the yolk sac. In addition to Oct4, the transcription factor Nanog plays a critical role in this event. The ICM of *nanog*-deficient embryos does not produce an epiblast, and only generates primitive endoderm (Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003).

Apart from the ICM, Oct4 and Nanog also play a vital role in maintaining the pluripotent state of ES cells. Loss of Oct4 causes inappropriate differentiation of ES cells into trophectoderm, whereas overexpression of Oct4 results in differentiation into primitive endoderm and mesoderm, suggesting that precise Oct4 levels are necessary for pluripotency (Nichols, Zevnik et al. 1998; Niwa, Miyazaki et al. 2000). Oct4 can regulate gene expression by interacting with other factors within the nucleus, including the high mobility group (HMG)-box

transcription factor, Sox2 (Boiani and Scholer 2005). Although Sox2 plays an important role in the maintenance of pluripotency and lineage specification, its expression is not restricted to pluripotent cells, because Sox2 is also found in early neural lineages (Avilion, Nicolis et al. 2003). ES cells lacking Nanog spontaneously differentiate into primitive endoderm (Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003). Conversely, overexpression of Nanog promotes self-renewal independently of the cytokine leukemia inhibitory factor (LIF), which functions by activating the transcription factor Stat3 (Matsuda, Nakamura et al. 1999). Although the LIF-Stat3 pathway is dispensable in human ES cells, recent functional analyses indicate an analogous role for Oct4 and Nanog in these cells (Hyslop, Stojkovic et al. 2005; Zaehres, Lensch et al. 2005). Thus, *Oct4*, *Nanog* and *Sox2* are the earliest-expressed set of genes known to maintain pluripotency. Together these studies suggest that Oct4, Nanog and Sox2 function in distinct pathways that might converge to regulate certain common genomic targets. It is likely that the interplay among these factors is critical for early cell fate decisions.

**The Balance between a Minimal Set of Lineage Specification transcription factors might drive early cell-fate decisions**

The simplest model for how Oct4, Nanog and Sox2 function is that they collaborate with other transcription factors to specify a pluripotent state and thus form the basis of a transcription factor hierarchy. Consistent with this, the balance between the levels of Oct4 and the Caudal-type homeodomain

transcription factor Cdx2 has recently been shown to influence the first overt lineage differentiation in the embryo (Niwa, Toyooka et al. 2005). Oct4 and Cdx2 expression patterns become mutually exclusive during embryogenesis, owing, in part, to their ability to reciprocally repress each other's expression. Oct4 is associated with the establishment of the ICM, whereas Cdx2 expression is necessary for trophectoderm development (Strumpf, Mao et al. 2005). Oct4 is lost from the outer cells of the morula that become fated for trophectoderm, whereas Cdx2 expression is restricted to these cells. Oct4 and Cdx2 also regulate the T-box transcription factor *eomesodermin* (*eomes*), which, like Cdx2 is necessary for trophectoderm maintenance (Niwa, Toyooka et al. 2005). These studies suggest that the interaction between these factors is essential for the segregation of the inner cell mass and trophectoderm lineages during development.

A similar balance between Nanog and the transcription factors, Gata4 and Gata6 might be necessary for differentiation into primitive endoderm, a derivative of the inner cell mass of the developing blastocyst. Forced expression of Gata4 or Gata6 in ES cells leads to differentiation into primitive endoderm, an effect similar to that caused by the loss of Nanog function (Fujikura, Yamato et al. 2002; Mitsui, Tokuzawa et al. 2003; Boyer, Lee et al. 2005). Moreover, Gata4 and Gata6 expression was upregulated in the absence of Nanog (Fujikura, Yamato et al. 2002), indicating that Nanog acts as a repressor of differentiation. Although there has been no *in vivo* evidence of Nanog acting as a transcriptional activator,

luciferase reporter assays indicate that Nanog can also activate transcription via its C-terminal domain (Pan and Pei 2005). Together, these studies suggest that a minimal set of lineage-specific factors can drive early cell fate decisions (Table 1). However, it is likely that other genetic, epigenetic and environmental factors play an important role in this process. It would be interesting, for instance, to identify the factors that proteins like Nanog and Oct4 interact with to allow them to act as transcriptional activators or repressors. One such study has identified a protein-interaction network for Nanog, although the relevance of individual binding events to pluripotency needs to be validated (Wang, Rao et al. 2006). A similar approach to identifying factors associated with Nanog is also described in the Appendix.

**Transcriptional Regulatory Networks in Pluripotent ES Cells**

Given that the factors orchestrating early cell fate decisions also regulate ES cell pluripotency, Oct4, Nanog and Sox2 are thought to establish the initial genomic state from which all other gene expression patterns are derived during development. Recent genomics studies have enabled the construction of transcriptional regulatory networks in ES cells that provide a foundation for understanding how Oct4, Nanog and Sox2 control pluripotency and influence subsequent differentiation events. Two studies have used chromatin immunoprecipitation (ChIP) combined with genome-wide methodologies to map the binding sites for Oct4 and Nanog throughout human and mouse ES cell genomes (Boyer, Lee et al. 2005; Loh, Wu et al. 2006). In the case of human ES

**Table 1. Gene Expression Analyses of Transcription Factors in ES Cell Pluripotency and Embryonic Development.**

| Transcription Factor | Protein Family | Expression Pattern | Loss of Function Phenotype | | Gain of Function Phenotype in ES Cells |
|---|---|---|---|---|---|
| | | | Embryonic Development | ES Cells | |
| **Oct4** | Pit/Oct/Unc protein family | oocytes, fertilized embryo, ICM, epiblast, ES cells, EC cells, germ cells | Embryonic lethality (blastocyst stage), differentiation of epiblast into TE lineage | Loss of pluripotency, differentiation into TE lineage | Differentiation into primitive endoderm and mesoderm |
| **Nanog** | Novel homeodomain protein | Morula, ICM, epiblast, ES cells, EC cells, germ cells | Embryonic lethality (E5.5), lack of epiblast, differentiation of ICM into primitive endoderm | Loss of pluripotency, differentiation into primitive endoderm | LIF/Stat3-independent self-renewal, resistance to retionoic acid-induced differentiation |
| **Sox2** | SRY-related HMG box protein | Oocytes, ICM, epiblast, germ cells, multipotent cells of extra-embryonic ectoderm, cells of neural lineage, brachial arches, gut endoderm | Embryonic lethal (E6.5), failure to maintain epiblast | Unknown | Unknown |
| **Stat3** | Signal Tranducer and Activator of Transcription family protein | Wide ranges of cell types | Embryonic lethality (E6.5-7.5) | Differentiation into primitive endoderm and mesoderm (Stat3 signaling is dispensable in human ES cells) | LIF-independent self renewal |
| **Cdx2** | Caudal-type homeodomain protein | Outer morula cells, TE cell lineages | Embryonic lethality due to implantation failure (lack of functional TE) | Normal contribution to all cell lineages except TE and intestinal cells | Differentiation into trophoblast |
| **Gata6** | GATA-binding protein | Extraembryonic endoderm lineages | Embryonic lethality (E5.5-7.5), defects in visceral endoderm formation | Unknown | Differentiation into primitive endoderm |
| **Gata4** | GATA-binding protein | Extraembryonic endoderm lineages | Embryonic lethality (E8-9), defects in heart morphogenesis | Can generate cardiac myocytes, inability to generate visceral endoderm and definitive endoderm of foregut | Differentiation into primitive endoderm |

cells, ChIP DNA was combined with a microarray platform (ChIP-chip) whereas for mouse ES cells, the ChIP DNA was linked to concatenated paired-end ditags and sequenced (ChIP-PET). These studies identified a large number of target genes and revealed that Oct4, Nanog, and in the case of human ES cells, Sox2 share a substantial portion of their targets. Further work in mouse ES cells using ChIP-chip has also been done to identify genomic targets of Oct4 and Nanog, and this work is described in Chapter 2. The experiments in this study identify a different set of targets than the one described by Loh et al (2006), indicating that each set is a partial representation of the Oct4 and Nanog regulatory network. This work has, therefore, begun to reveal the circuitry that is responsible for the combined biological output of these ES cell regulators.

## Similarities and Differences Between Mouse and Human ES Cell Genomic Targets

Oct4, Sox2 and Nanog occupy both transcriptionally active and inactive genes in mouse and human ES cells. Active genes include the transcription factors *Oct4*, *Sox2* and *Nanog* themselves, as well as others that are highly expressed in ES cells, such as *Rif1*, *Jarid2* and *Smarcad1.* *Rif1* has been implicated in regulating telomere length and might be important for self-renewal (Adams and McLaren 2004). Although *Jarid2* and *Smarcad1* have important roles in development (Schoor, Schuster-Gossler et al. 1993; Jung, Mysliwiec et al. 2005), their contribution to pluripotency is unknown. Interestingly, a large portion of the inactive targets identified in mouse and human ES cells include transcription

factors involved in lineage-specification (Figure 1(b)). The developmental importance of these genes suggests that Oct4, Sox2 and Nanog act in concert to maintain pluripotency by directly controlling a transcriptional regulatory hierarchy that specifies differentiation into extra-embryonic lineages in addition to derivatives of the primary germ layers.

A comparison of Oct4- and Nanog-bound regions identified in these studies, however, revealed only modest similarity between the target genes in the two species. For instance, certain genes such as *Hand1* and *Myst3* were identified as targets of Oct4 and Nanog exclusively in human ES cells, whereas others such as *Esrrb* were observed only in mouse ES cells. It is interesting to note that although *Hand1* was not identified as a target in mouse ES cells, its expression was upregulated upon RNAi-mediated silencing of both *Esrrb* and *Rif1* in these cells (Loh, Wu et al. 2006). The lack of orthologous genomic targets could be due to genuine differences between the gene regulatory networks or a result of the dissimilarities in genomic platforms used in these studies. Detailed comparisons of Oct4, Sox2 and Nanog target genes between the two species will be imperative for determining the extent to which genetic regulatory information can be extrapolated from one species to the other.

Although these studies provide an initial framework for deciphering the mechanisms by which these key regulators elicit their effects, genetic manipulation of *Oct4*, *Sox2* and *Nanog* combined with gene expression analyses is necessary to elucidate which of their targets are important for the maintenance

of pluripotency or downstream differentiation events. Such analyses, reported in the same study that identified mouse Oct4 and Nanog targets (Loh, Wu et al. 2006), as well as in another study in which mouse ES cells gene expression patterns were profiled under a wide range of conditions, are critical steps in this direction (Ivanova, Dobrin et al. 2006). In addition to confirming a role for *Esrrb* in mouse, Ivanova and colleagues recognized *Tcl1* and *Tbx3* as being important factors for sustaining an undifferentiated state. Interestingly, Esrrb has been shown to be important for placental development and germ cell proliferation, and Tcl1, which is highly expressed in ES cells (Mitsui, Tokuzawa et al. 2003), enhances cell proliferation and survival through augmentation of PI3K-Akt signaling (Teitell 2005; Meshorer and Misteli 2006). Thus, how these factors contribute to ES cell self-renewal and pluripotency is of particular interest. Together, these genome-wide studies suggest that Oct4, Sox2 and Nanog form the basis for a specialized transcriptional regulatory circuitry that allows for consistent patterning of gene expression during ES cell propagation.

## Epigenetic Control of Pluripotency: Chromatin Dynamics and Epigenetic Profile of Pluripotent ES Cells

Chromatin reorganization is essential for the establishment of new heritable gene expression programs that accompany lineage specifications (Figure 2) (Meshorer, Yellajoshula et al. 2006). For example, ES cell chromatin displays characteristics of transcriptionally permissive euchromatin, such as an abundance of acetylated histone modifications and increased accessibility to

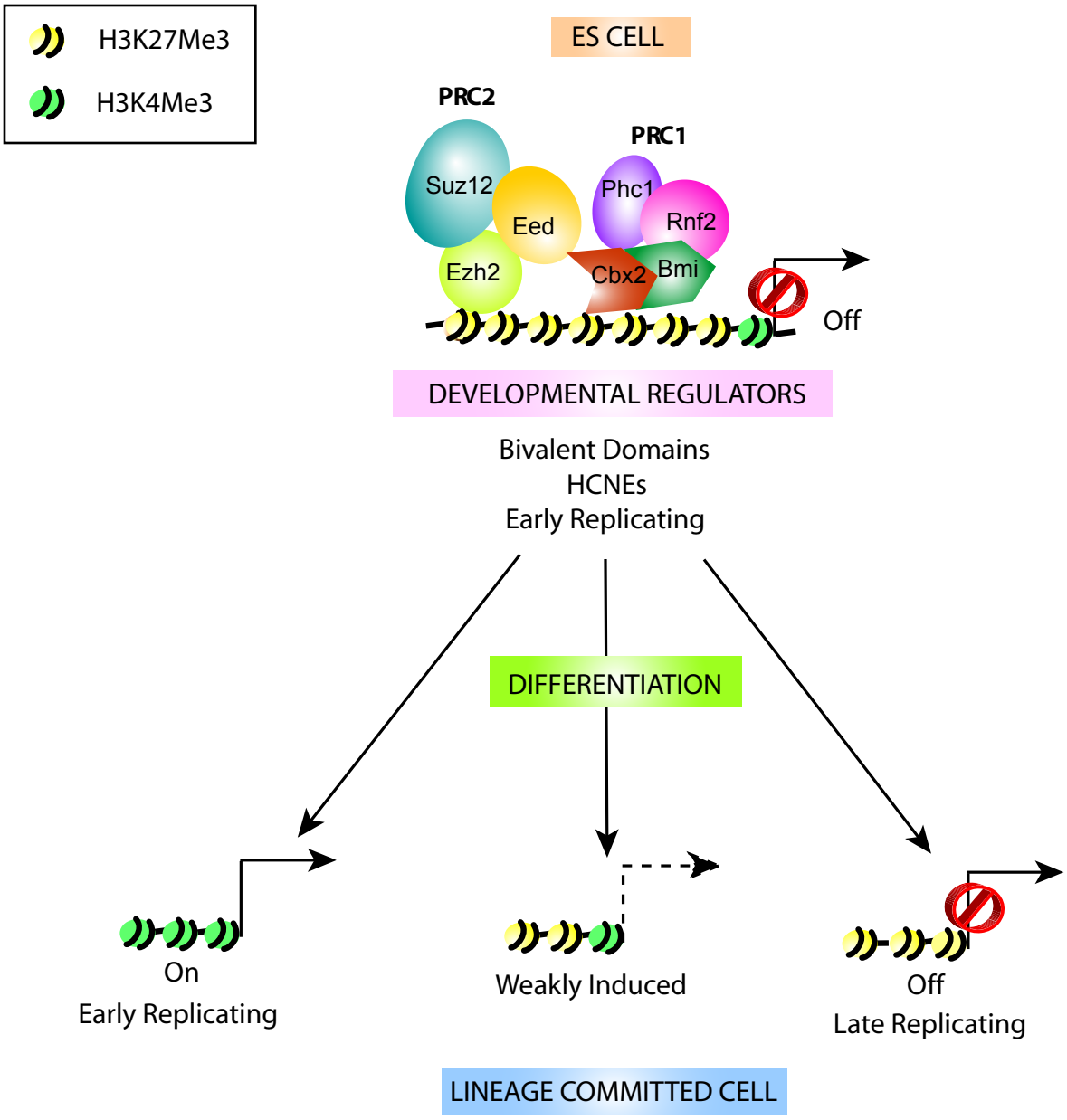
nucleases. Conversely, lineage specification is typified by a decrease in acetylation and concomitant increase in heterochromatin formation, indicating that restriction of developmental potential is associated with a decrease in genome plasticity. Recent studies have revealed additional unique properties of pluripotent chromatin that distinguish these cells from their differentiated progeny.

An analysis of global chromatin dynamics revealed a highly dynamic association of structural chromatin proteins (e.g. core and variant histones, the linker histone H1, and the heterochromatin associated proteins HP1$\alpha$) with the chromatin of pluripotent cells compared with that of differentiated cell types (Meshorer, Yellajoshula et al. 2006). This study also showed that replacement of histone H1 with a version that binds more tightly to chromatin inhibited ES cell differentiation. These data posit that structural proteins remain loosely associated with chromatin in pluripotent cells, thereby enabling the reorganization of chromatin structure during differentiation.

Consistent with the observation that the chromatin of pluripotent nuclei is in an 'open' conformation, recent studies have shown that tissue-specific genes that are expected to be silent in undifferentiated cells might be in a semi-permissive transcriptional state in ES cells (Szutorisz, Canzonetta et al. 2005; Levings, Zhou et al. 2006). For example, active epigenetic marks were noted in ES cells at discrete sites within the B-cell specific *λ5-VpreB1* locus prior to gene activation

**Figure 2. Epigenetic Characteristics of Pluripotent and Lineage Committed Cells.** PcG proteins have recently been shown to reversibly silence developmental regulators in ES cells, a process that might be necessary for the propagation of an undifferentiated state. These regulators, which are early replicating, contain highly conserved non-coding elements (HCNEs), which are rich in bivalent domains that consist of both H3K27me3 and H3K4me3 modifications. These domains might provide an epigenetic indexing mechanism to mark genes for expression at later developmental stages. During differentiation of ES cells, the bivalent marks resolve, because early-replicating genes that are expressed in the lineage-committed cells maintain or acquire activating H3K4me3 marks, and late-replicating genes that are turned off in these cells possess repressive H3K27me3 modifications. Notably, genes that are weakly induced still possess bivalent domains.

**FIGURE 2.**


H3K27Me3
H3K4Me3
ES CELL
PRC2
PRC1
Suz12
Eed
Ezh2
Phc1
Rnf2
Cbx2
Bmi
Off
DEVELOPMENTAL REGULATORS
Bivalent Domains
HCNEs
Early Replicating
DIFFERENTIATION
On
Early Replicating
Weakly Induced
Off
Late Replicating
LINEAGE COMMITTED CELL

during B-cell commitment (Szutorisz, Canzonetta et al. 2005). Two other reports (Azuara, Perry et al. 2006; Bernstein, Mikkelsen et al. 2006) support such an epigenetic indexing mechanism by revealing the existence of dual marks or 'bivalent' domains, consisting of repressive histone H3K27me3 and activating histone H3K4me3 modifications at a large set of developmentally important genes that are silent in ES cells but activated upon differentiation. These studies suggest that lineage-specific genes are cued in ES cells for subsequent activation during differentiation. Furthermore, bivalent domains coincide with the most highly conserved non-coding elements in the mammalian genome, suggesting an evolutionarily conserved role for these chromatin domains (Bernstein, Mikkelsen et al. 2006). The additional observation that Oct4, Nanog and Sox2 occupied a significant subset of genes that harbor bivalent domains supports a link between the repressions of developmental regulators and stem cell pluripotency (Boyer, Lee et al. 2005; Bernstein, Mikkelsen et al. 2006; Boyer, Plath et al. 2006; Lee, Jenner et al. 2006). It is important to note that not all tissue-specific genes appear to contain these bivalent marks and the underlying chromatin structure at these genes and their contributions to pluripotency await further characterization.

**A Role for Polycomb Group Proteins in Maintaining ES Cell Identity?**

Gene expression is influenced by enzymatic activities that can induce both global and local changes in chromatin structure. Polycomb group (PcG) proteins were first identified in *Drosophila* as transcriptional repressors of homeotic gene

expression during embryogenesis (Ringrose and Paro 2004). PcG proteins comprise at least two distinct repressor complexes (PRC1 and PRC2-PRC3), the core components of which are highly conserved between fly and human (Levine, Weiss et al. 2002). A role for PcG proteins in pluripotency in mammals was suggested on the basis that PcG components are required for early developmental gene expression patterning, the early establishment of pluripotent ES cell lines, and for adult stem cell maintenance.

Recently, the location of PcG components throughout the genome was mapped in Drosophila (Negre, Hennetin et al. 2006; Schwartz, Kahn et al. 2006; Tolhuis, Muijrers et al. 2006) and mammals (Boyer, Plath et al. 2006; Bracken, Dietrich et al. 2006; Lee, Jenner et al. 2006). These studies revealed that in human and mouse ES cells, the PRC1 and PRC2 complexes bind to a large set of genes comprised of transcriptional regulators and signaling factors with known roles in development. Genes occupied by PcG proteins also contained H3K27me3 in their promoter regions, a repressive histone modification catalyzed by PRC2. Many of the target genes were de-repressed in the absence of the PRC2 components Eed or Suz12, indicating a direct functional link between PRC2 and gene silencing in ES cells (Bernstein, Mikkelsen et al. 2006; Boyer, Plath et al. 2006). ES cells lacking Eed can contribute to most cell lineages, suggesting that PcG proteins are not necessary to maintain pluripotency (Morin-Kensicki, Faust et al. 2001). However, the observations that *Eed* mutant ES cells spontaneously differentiate (Boyer, Plath et al. 2006), and ES cells cannot be derived from

blastocysts deficient for the PRC2 component Ezh2 (O'Carroll, Erhardt et al. 2001) suggest that PcG proteins are necessary for ES cell identity.

PcG target genes were preferentially activated upon differentiation, indicating that they are poised for activation in ES cells (Boyer, Plath et al. 2006; Lee, Jenner et al. 2006). In flies, the maintenance of heritable epigenetic states requires the interplay between repression mediated by PcG proteins and activation by Trithorax group (Trx) proteins (Ringrose and Paro 2004). Trx proteins catalyze lysine 4 tri-methylation on histone H3 (H3K4me3) (Martin and Zhang 2005). Interestingly, many of the PcG target genes contained bivalent chromatin domains in their promoter regions (Azuara, Perry et al. 2006; Bernstein, Mikkelsen et al. 2006; Boyer, Plath et al. 2006; Lee, Jenner et al. 2006), consistent with the idea that chromatin accessibility is governed by the balance between positively and negatively acting factors (Dillon and Festenstein 2002). Additionally, PcG target genes are replicated early in ES cells, a property associated with transcriptionally active euchromatin (Perry, Sauer et al. 2004; Azuara, Perry et al. 2006). However, replication timing was not significantly altered in *Eed* mutant ES cells (Szutorisz, Canzonetta et al. 2005) suggesting that the presence of H3K4me3 or additional factors was required to maintain these genes in a semi-permissive transcriptional state. Many PcG target genes were also bound by Oct4, Sox2, and Nanog, indicating that these ES cell regulators may play a role in recruiting PcG complexes to catalyze the silencing of these genetic loci (Boyer, Plath et al. 2006; Lee, Jenner et al. 2006). The

identity of the components that catalyze the addition of the activating mark at these genes in ES cells, as well as identification of the factors that recruit PcG and Trx proteins will be important to better understand how these genes are regulated. A recent study also revealed a role for Mbd3, an essential component of the Nucleosome Remodeling and Histone Deacetylation (NuRD) complex, in ES cell differentiation (Kaji, Caballero et al. 2006). In *C. elegans*, germline-specific chromatin states specified through PcG-like activities are reorganized in somatic cells by a NuRD-like activity (Shin and Mello 2003). Thus, it is likely that the balance between pluripotency and lineage commitment is dependent upon the correct spatial and temporal expression of genes orchestrated by the action of both genetic and epigenetic factors.

## (II) REPROGRAMMING TO A PLURIPOTENT STATE

### Restriction of Developmental Potency & The Need for Nuclear Reprogramming

The promise that ES cells hold for therapeutic purposes has been countered with a number of practical and ethical dilemmas, since the production of these cells involves the destruction of embryos. Reprogramming somatic cells into an ES cell-like state would be a suitable alternative to circumvent these issues. The interest in reprogramming dates back to the 1950s, when Briggs and King showed through somatic cell nuclear transfer experiments in frog (*Rana pipiens*) embryos, a progressive loss of nuclear potency with increasing developmental

age (King and Briggs 1956). They isolated donor nuclei from cells at different stages of development, and transferred them into enucleated oocytes to examine the developmental potency of these nuclei. Their results showed that even though most nuclei from blastula cells could generate tadpoles, there was a substantial decrease in the ability of nuclei from later stages of development, such as gastrula and neurula, to produce offspring. Therefore, these observations were consistent with the notion that even though the genetic material of cells at different developmental stages is equivalent, genomic modifications restrict the nuclear potency of these cells as they undergo differentiation. However, the cloning of mammals such as Dolly has shown that even the potential of differentiated cell nuclei can be altered (Wilmut, Schnieke et al. 1997). Therefore, these modifications to the genome are reversible, even though the reprogramming process is extremely inefficient. The road to reprogramming has been one with major challenges, and some of the strategies that have been employed to overcome these hurdles are discussed here.

## Strategies for Reprogramming to a Pluripotent State

### (i) Nuclear Transfer

As described earlier, Somatic Cell Nuclear Transfer (SCNT) or Nuclear Transfer (NT) was the first method employed to reprogram the genome of a differentiated cell into a pluripotent one. Such a technique allows for the derivation of patient-specific ES cell lines, which have the potential to be used for therapeutic purposes. The process of nuclear cloning was successfully accomplished in

mammals, when the sheep Dolly was cloned by transplanting the nucleus of a differentiated mammary epithelium cell into an enucleated oocyte (Wilmut, Schnieke et al. 1997). Subsequently, this technique has also been used to clone other mammals, such as cattle (Kato, Tani et al. 1998), goats (Baguisi, Behboodi et al. 1999), pigs (Onishi, Iwamoto et al. 2000; Polejaeva, Chen et al. 2000) and mice (Wakayama, Perry et al. 1998). Moreover, generation of monoclonal mice by NT from mature lymphocytes (Hochedlinger and Jaenisch 2002), as well as the cloning of mice from post-mitotic olfactory neurons indicated that even the nuclei of terminally differentiated cells can be coaxed to re-enter the cell cycle and be reprogrammed to a totipotent state (Eggan, Baldwin et al. 2004). A proof-of-principle experiment in mice showed that disease-specific ES cells derived from NT blastocysts could be repaired by homologous recombination and used to treat an immunological disorder (Rideout, Hochedlinger et al. 2002).

In order to get around the issue of embryo destruction for ES cell derivation, William Hurlbut, a member of the United States President's Council on Bioethics, suggested a possible solution called Altered Nuclear Transfer (ANT) (Hurlbut 2005). This procedure was proposed as a variation on NT, since it would generate abnormal embryos that would fail to implant in the uterus and not develop into viable offspring. Such a technique was accomplished by disrupting the gene *Cdx2*, which is crucial for differentiation into the trophectodermal lineage (Meissner and Jaenisch 2006). Mouse embryos lacking this gene are unable to develop beyond the blastocyst stage since they do not form placentas

and cannot implant in the uterus. However, the blastocysts of these embryos have an ICM, which can be explanted in tissue culture to give rise to pluripotent ES cells (Chawengsaksophak, de Graaff et al. 2004; Strumpf, Mao et al. 2005).

Despite this success with NT, the major challenges plaguing this technology are that it is an extremely inefficient process, and a large number of cloned offspring have epigenetic instability that leads to abnormal gene expression and organismal growth (Humpherys, Eggan et al. 2002; Blelloch, Wang et al. 2006). A large number of cloned embryos die during gestation, exhibiting characteristics of large offspring syndrome (Young, Sinclair et al. 1998; Chavatte-Palmer, Heyman et al. 2000), frequently with respiratory and metabolic problems, and large and abnormal placentas (Hill, Roussel et al. 1999; Wakayama and Yanagimachi 1999; Hill, Burghardt et al. 2000). Therefore, even though patient-specific ES cells may be derived by reprogramming a somatic cell through NT, these cells are likely to have epigenetic abnormalities that could pose potential issues for therapeutic uses.

**(ii) Reprogramming by Fusion**

Another approach to reprogram somatic cells into a pluripotent state has been to fuse them with ES cells, thereby creating hybrids in which the differentiated nuclei get epigenetically reprogrammed and exhibit properties of ES cells. This technique has been demonstrated successfully in mice, where adult thymocytes were fused to ES cells, and the resulting hybrids could contribute to all three

germ layers, revealing the pluripotent features of the parent ES cells (Tada, Takahama et al. 2001). Another indication of reprogramming in these hybrids was that the inactive somatic X chromosome and repressed pluripotency markers, such as Oct4, were reactivated upon fusion with ES cells. Reprogramming through fusion has also been accomplished in human cells, where human fibroblasts as well as myeloid progenitors were reprogrammed by fusion with human ES cells (Cowan, Atienza et al. 2005; Yu, Vodyanik et al. 2006).

The key issue with using this cell fusion approach is the generation of tetraploid hybrid cells. If these reprogrammed hybrids are to be used for therapeutic purposes, it will be necessary to eliminate the ES cell genome used for the fusion procedure. Although the targeted elimination of a few chromosomes in these hybrid cells has been accomplished with the use of a chromosomal deletion cassette (CEC), this method of obtaining diploid reprogrammed cells poses major risks of creating genomic instability that would hamper cell-survival (Matsumura, Tada et al. 2007).

**(iii) *In Vitro* Reprogramming with Defined Transcription Factors**

A major feat in the field of reprogramming was accomplished in a recent study, where four transcription factors were used to reprogram somatic cells into an ES cell-like state (Takahashi and Yamanaka 2006). In this experiment, mouse embryonic fibroblasts (MEFs), as well as adult tail-tip fibroblasts were infected

with viruses carrying transgenes of the transcription factors Oct4, Sox2, C-myc and Klf4. The infected fibroblasts were selected for the activation of Fbx15, which is a direct target of Oct4 (Tokuzawa, Kaiho et al. 2003). Although Fbx15 is expressed predominantly in ES cells, it seems to be dispensable for the maintenance of pluripotency and mouse embryonic development. Infected cells in which Fbx15 had been activated were known as Fbx15-iPS (induced pluripotent stem) cells. These cells were shown to be pluripotent since they could form teratomas. However, they were unable to generate any live chimeras. Therefore, these Fbx15-iPS cells are thought to represent a partial state in reprogramming.

In subsequent studies, the activation of endogenous Oct4 or Nanog loci was used as a selection criterion for reprogramming. The Oct4- and Nanog-iPS cells obtained in these studies were completely reprogrammed, since not only could they give rise to teratomas, they could also generate live germline chimeras (Maherali, Sridharan et al. 2007; Okita, Ichisaka et al. 2007; Wernig, Meissner et al. 2007). Moreover, in contrast to the partially reprogrammed Fbx15-iPS cells, the Oct4- and Nanog-neo iPS cells had completely reprogrammed, unmethylated endogenous Oct4 and Nanog promoters. Additionally, the inactive somatic X chromosome was also activated in the Oct4- and Nanog-iPS cells (Maherali, Sridharan et al. 2007). Further work on these cells has also shown that the pluripotency markers, Alkaline Phosphatase (AP), Stage-specific embryonic antigen 1 (SSEA1), Oct4 and Nanog get activated during the course of

reprogramming (Wernig, Meissner et al. 2007; Brambrink, Foreman et al. 2008; Stadtfeld, Maherali et al. 2008). However, the major difference between the studies using Fbx15 and Oct4 or Nanog as selection markers for reprogramming was the time at which selection was started. In the first study, selection was started early (day 3 post infection) in order to obtain Fbx15-iPS cells, whereas in the case of Oct4 and Nanog-iPS cells, selection was started at later times (days 6 or 9 post infection). Therefore, it is still unclear whether the partial reprogramming in the former case is obtained due to the fact that Fbx15 is not relevant to pluripotency, or because of the difference in selection timing. This issue is addressed in further detail in Chapter 3 in an effort to shed more light onto the mechanisms by which a somatic cell can get reprogrammed partially or completely into a pluripotent state.

**Mechanism of Reprogramming Fibroblasts into iPS Cells**

Recent studies have started to explore the mechanisms by which a somatic cell can be reprogrammed *in vitro* into iPS cells by viral-mediated transduction of Oct4, Sox2, Klf4 and c-Myc transgenes (Brambrink, Foreman et al. 2008; Stadtfeld, Maherali et al. 2008). This work has been focused on defining the steps of reprogramming by determining the kinetics of pluripotency marker expression during this process (Figure 3). Both groups observed that the activation of such markers was a gradual and sequential process, and not a stochastic one. FACS analyses of cells isolated at different time points after infection with the four factors, revealed that the pluripotency marker, AP was the

first one to be reactivated. This was followed by the upregulation of the cell surface marker SSEA1, which is expressed in pluripotent cells, and a concomitant downregulation of Thy1, a cell surface antigen expressed in fibroblasts and differentiated cells (Rege and Hagood 2006). The pluripotency genes, Oct4, Nanog and Sox2 were upregulated later in the reprogramming process, and were accompanied by the reactivation of telomerase (mTert), as well as the silent X chromosome. Nearly all cells expressing Oct4 or Nanog also expressed the early markers, AP and SSEA1, but the reverse was not true for most AP and SSEA1 positive cells isolated at early time points. This observation supported the notion that the reactivation of pluripotency genes is a gradual and sequential process, and that SSEA1 marks an intermediate step in reprogramming. It will be of interest in the future to determine whether the occurrence of late events, such as Oct4 and Nanog expression is dependent on the early events in this process.

These studies also examined the minimal time of transgene expression that was required for reprogramming (Brambrink, Foreman et al. 2008; Stadtfeld, Maherali et al. 2008). In order to address this question, the four transcription factors were expressed using doxycycline-dependent inducible viruses carrying transgenes for these factors. Doxycycline was added to the fibroblasts immediately after they were infected with these viruses, and it was withdrawn at different time points after infection. The results of these studies suggested that transgene expression was required at least up to days 12-16 in order to obtain fully reprogrammed iPS

colonies (Figure 3). If Doxycycline was withdrawn earlier, then cells in the reprogramming process would revert to a fibroblast-like morphology. Both groups also noted that after stable iPS cells were obtained, downregulation of transgene expression was required in order for iPS cells to differentiate.

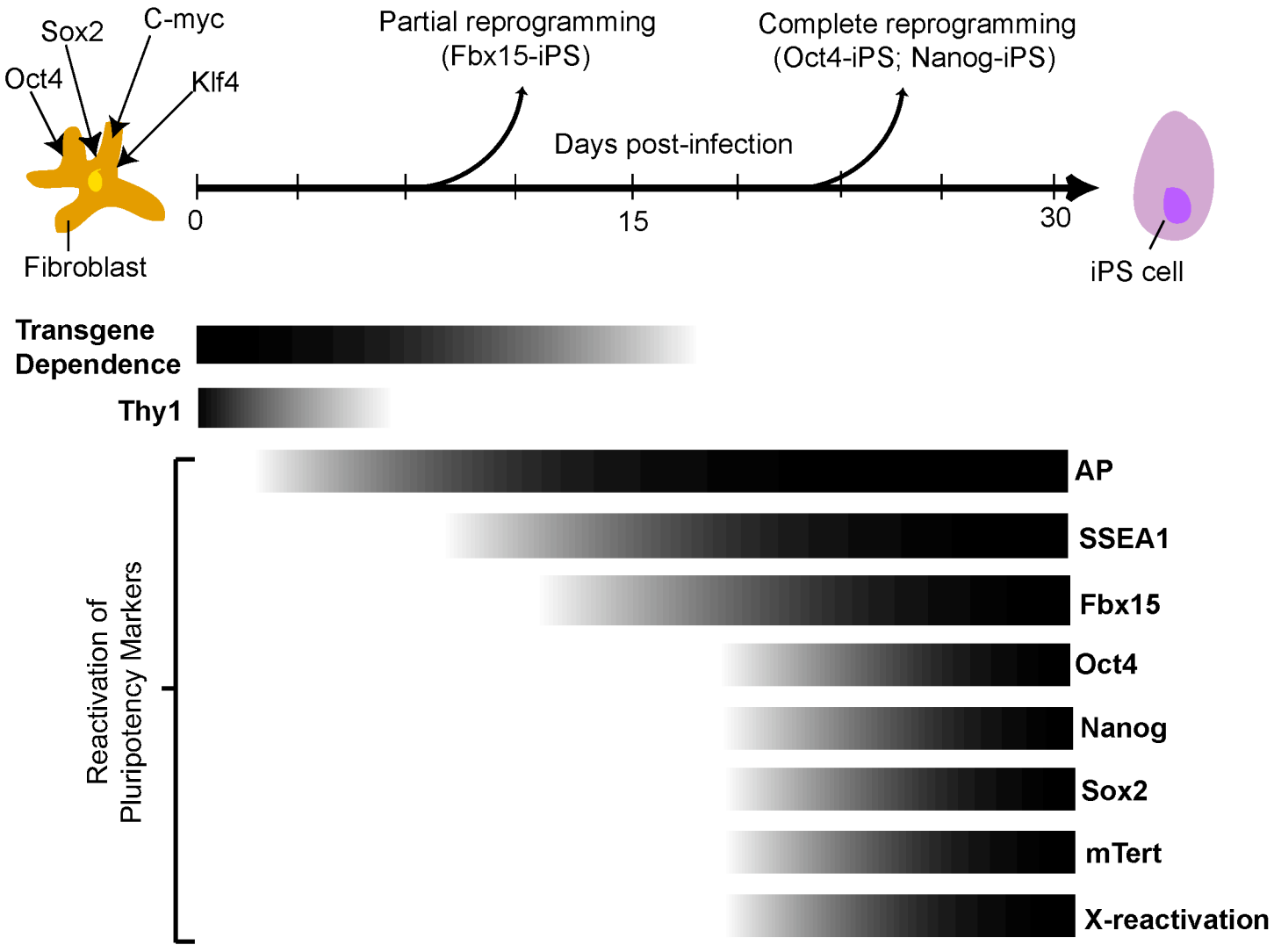
An interesting observation made by several groups has been that depending on the timing of selection for pluripotency markers, drug resistant colonies appear at a wide range of time points following infection with the four factors (Maherali, Sridharan et al. 2007; Okita, Ichisaka et al. 2007; Wernig, Meissner et al. 2007; Brambrink, Foreman et al. 2008; Stadtfeld, Maherali et al. 2008). However, in a number of drug resistant colonies that are obtained early, the endogenous Oct4 and Nanog loci do not seem to be reactivated. This discrepancy between the timing of drug resistance and reactivation of pluripotency genes is not well understood. One possible explanation for this observation could be that in such cells, a low level of Oct4 or Nanog expression may be present, which is sufficient to confer drug-resistance, but not enough for complete reprogramming, thereby supporting the idea that this is a process in which pluripotency genes are gradually reactivated.

## Applications of iPS cells in Therapeutic Medicine

The iPS cells obtained by viral transduction of the four transcription factors are morphologically and functionally indistinguishable from ES cells. Similar to ES cells, these iPS cells also hold great therapeutic potential, while circumventing the ethical dilemmas associated with ES derivation. In a recent proof-of-principle

**Figure 3. Sequence of marker expression during reprogramming.** Kinetics of fibroblast and pluripotency marker expression is represented by the rectangular gradients. The minimal amount of time for which transgenes expressing the four factors must be expressed for complete reprogramming, is also shown.

**FIGURE 3.**


Sox2
C-myc
Oct4
Klf4
Fibroblast
Partial reprogramming
(Fbx15-iPS)
Complete reprogramming
(Oct4-iPS; Nanog-iPS)
Days post-infection
0
15
30
iPS cell
Transgene Dependence
Thy1
Reactivation of Pluripotency Markers
AP
SSEA1
Fbx15
Oct4
Nanog
Sox2
mTert
X-reactivation

study, iPS cells generated from murine fibroblasts were used to successfully treat a mouse model of sickle-cell anemia (Hanna, Wernig et al. 2007). Furthermore, iPS cells could also be used to derive functional neuronal cell types *in vitro*, which when transplanted into the brain of a rat model of Parkinson's disease, could alleviate some symptoms associated with this disorder (Wernig, Zhao et al. 2008).

Even though this same combination of four transcription factors can be used to obtain iPS cells from human fibroblasts (Takahashi, Tanabe et al. 2007; Yu, Vodyanik et al. 2007; Park, Zhao et al. 2008), there are several issues that must be dealt with before such iPS cells can be used for medical purposes. The generation of iPS cells has thus far relied on virus-mediated transduction of the four transcription factors, as well as on drug-selection for the activation of Fbx15, Oct4 or Nanog markers. The use of viruses and drugs could potentially affect the eventual application of these cells in human therapies, and suitable alternatives for these experimental requirements need to be explored. The issue of drug selection has been addressed in a recent study, where completely reprogrammed, genetically unmodified iPS cells were obtained without drug selection, based solely on morphological similarity to ES cells (Meissner, Wernig et al. 2007). An additional problem with using these four factors for reprogramming is that iPS cell derived mice frequently develop tumors (Okita, Ichisaka et al. 2007). In an effort to address this issue, iPS cells were derived without the *c-myc* oncogene. Although reprogramming in this case was

significantly slower and less efficient, mice derived from these cells did not have *c-myc*-induced tumors (Nakagawa, Koyanagi et al. 2008; Wernig, Meissner et al. 2008). It is still unclear whether the introduction of the other three transcription factors will induce tumors at later stages of development. If this is the case, then safer substitutes for these reprogramming factors will have to be explored. Much of the current research in reprogramming is aimed at addressing these issues, in the hope that medical and scientific benefits will be reaped from this work, unhindered by the ethical and practical dilemmas that currently swamp this field.

## REFERENCES


Adams, I. R. and A. McLaren (2004). "Identification and characterisation of mRif1: a mouse telomere-associated protein highly expressed in germ cells and embryo-derived pluripotent stem cells." Dev Dyn **229**(4): 733-44.

Avilion, A. A., S. K. Nicolis, et al. (2003). "Multipotent cell lineages in early mouse development depend on SOX2 function." Genes Dev **17**(1): 126-40.

Azuara, V., P. Perry, et al. (2006). "Chromatin signatures of pluripotent cell lines." Nat Cell Biol **8**(5): 532-8.

Baguisi, A., E. Behboodi, et al. (1999). "Production of goats by somatic cell nuclear transfer." Nat Biotechnol **17**(5): 456-61.

Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-26.

Blelloch, R., Z. Wang, et al. (2006). "Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus." Stem Cells **24**(9): 2007-13.

Boiani, M. and H. R. Scholer (2005). "Regulatory networks in embryo-derived pluripotent stem cells." Nat Rev Mol Cell Biol **6**(11): 872-84.

Boyer, L. A., T. I. Lee, et al. (2005). "Core transcriptional regulatory circuitry in human embryonic stem cells." Cell **122**(6): 947-56.

Boyer, L. A., K. Plath, et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-53.

Bracken, A. P., N. Dietrich, et al. (2006). "Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions." Genes Dev **20**(9): 1123-36.

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

Burdon, T., A. Smith, et al. (2002). "Signalling, cell cycle and pluripotency in embryonic stem cells." Trends Cell Biol **12**(9): 432-8.

Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-55.

Chavatte-Palmer, P., Y. Heyman, et al. (2000). "[Cloning and associated physiopathology of gestation]." Gynecol Obstet Fertil **28**(9): 633-42.

Chawengsaksophak, K., W. de Graaff, et al. (2004). "Cdx2 is essential for axial elongation in mouse development." Proc Natl Acad Sci U S A **101**(20): 7641-5.

Cinalli, R. M., P. Rangan, et al. (2008). "Germ cells are forever." Cell **132**(4): 559-62.

Cowan, C. A., J. Atienza, et al. (2005). "Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells." Science **309**(5739): 1369-73.

de la Serna, I. L., Y. Ohkawa, et al. (2006). "Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers." Nat Rev Genet **7**(6): 461-73.

Dillon, N. and R. Festenstein (2002). "Unravelling heterochromatin: competition between positive and negative factors regulates accessibility." Trends Genet **18**(5): 252-8.

Driesch, H. (1892). "Entwicklungsmechanik Studien. I. Der Werth der beiden ersten Furchungszellen in der Echinodermen-entwiklung. ExperimentellenErzeugen von Theil-und-Doppelbildung." Zeit Fur Wissenschaft Zool 53: 160–178.

Eggan, K., K. Baldwin, et al. (2004). "Mice cloned from olfactory sensory neurons." Nature **428**(6978): 44-9.

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

Fujikura, J., E. Yamato, et al. (2002). "Differentiation of embryonic stem cells is induced by GATA factors." Genes Dev **16**(7): 784-9.

Hanna, J., M. Wernig, et al. (2007). "Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin." Science **318**(5858): 1920-3.

Hill, J. R., R. C. Burghardt, et al. (2000). "Evidence for placental abnormality as the major cause of mortality in first-trimester somatic cell cloned bovine fetuses." Biol Reprod **63**(6): 1787-94.

Hill, J. R., A. J. Roussel, et al. (1999). "Clinical and pathologic features of cloned transgenic calves and fetuses (13 case studies)." Theriogenology **51**(8): 1451-65.

Hochedlinger, K. and R. Jaenisch (2002). "Monoclonal mice generated by nuclear transfer from mature B and T donor cells." Nature **415**(6875): 1035-8.

Humpherys, D., K. Eggan, et al. (2002). "Abnormal gene expression in cloned mice derived from embryonic stem cell and cumulus cell nuclei." Proc Natl Acad Sci U S A **99**(20): 12889-94.

Hurlbut, W. B. (2005). "Altered nuclear transfer as a morally acceptable means for the procurement of human embryonic stem cells." Perspect Biol Med **48**(2): 211-28.

Hyslop, L., M. Stojkovic, et al. (2005). "Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages." Stem Cells **23**(8): 1035-43.

Ivanova, N., R. Dobrin, et al. (2006). "Dissecting self-renewal in stem cells with RNA interference." Nature **442**(7102): 533-8.

Jung, J., M. R. Mysliwiec, et al. (2005). "Roles of JUMONJI in mouse embryonic development." Dev Dyn **232**(1): 21-32.

Kaji, K., I. M. Caballero, et al. (2006). "The NuRD component Mbd3 is required for pluripotency of embryonic stem cells." Nat Cell Biol **8**(3): 285-92.

Kato, Y., T. Tani, et al. (1998). "Eight calves cloned from somatic cells of a single adult." Science **282**(5396): 2095-8.

King, T. J. and R. Briggs (1956). "Serial transplantation of embryonic nuclei." Cold Spring Harb Symp Quant Biol **21**: 271-90.

Lee, T. I., R. G. Jenner, et al. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-13.

Levine, S. S., A. Weiss, et al. (2002). "The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans." Mol Cell Biol **22**(17): 6070-8.

Levings, P. P., Z. Zhou, et al. (2006). "Recruitment of transcription complexes to the beta-globin locus control region and transcription of hypersensitive site 3 prior to erythroid differentiation of murine embryonic stem cells." Febs J **273**(4): 746-55.

Lin, W. and S. Y. Dent (2006). "Functions of histone-modifying enzymes in development." Curr Opin Genet Dev **16**(2): 137-42.

Loh, Y. H., Q. Wu, et al. (2006). "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells." Nat Genet **38**(4): 431-40.

Maherali, N., R. Sridharan, et al. (2007). "Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution." Cell Stem Cell **1**(1): 55-70.

Martin, C. and Y. Zhang (2005). "The diverse functions of histone lysine methylation." Nat Rev Mol Cell Biol **6**(11): 838-49.

Martin, G. R. (1981). "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." Proc Natl Acad Sci U S A **78**(12): 7634-8.

Matsuda, T., T. Nakamura, et al. (1999). "STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells." Embo J **18**(15): 4261-9.

Matsumura, H., M. Tada, et al. (2007). "Targeted chromosome elimination from ES-somatic hybrid cells." Nat Methods **4**(1): 23-5.

Meissner, A. and R. Jaenisch (2006). "Generation of nuclear transfer-derived pluripotent ES cells from cloned Cdx2-deficient blastocysts." Nature **439**(7073): 212-5.

Meissner, A., M. Wernig, et al. (2007). "Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells." Nat Biotechnol **25**(10): 1177-81.

Meshorer, E. and T. Misteli (2006). "Chromatin in pluripotent embryonic stem cells and differentiation." Nat Rev Mol Cell Biol **7**(7): 540-6.

Meshorer, E., D. Yellajoshula, et al. (2006). "Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells." Dev Cell **10**(1): 105-16.

Mitsui, K., Y. Tokuzawa, et al. (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-42.

Morin-Kensicki, E. M., C. Faust, et al. (2001). "Cell and tissue requirements for the gene eed during mouse gastrulation and organogenesis." Genesis **31**(4): 142-6.

Nakagawa, M., M. Koyanagi, et al. (2008). "Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts." Nat Biotechnol **26**(1): 101-6.

Negre, N., J. Hennetin, et al. (2006). "Chromosomal distribution of PcG proteins during Drosophila development." PLoS Biol **4**(6): e170.

Nichols, J., B. Zevnik, et al. (1998). "Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4." Cell **95**(3): 379-91.

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

Niwa, H., Y. Toyooka, et al. (2005). "Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation." Cell **123**(5): 917-29.

O'Carroll, D., S. Erhardt, et al. (2001). "The polycomb-group gene Ezh2 is required for early mouse development." Mol Cell Biol **21**(13): 4330-6.

Okita, K., T. Ichisaka, et al. (2007). "Generation of germline-competent induced pluripotent stem cells." Nature.

Onishi, A., M. Iwamoto, et al. (2000). "Pig cloning by microinjection of fetal fibroblast nuclei." Science **289**(5482): 1188-90.

Orkin, S. H. and L. I. Zon (2008). "Hematopoiesis: an evolving paradigm for stem cell biology." Cell **132**(4): 631-44.

Pan, G. and D. Pei (2005). "The stem cell pluripotency factor NANOG activates transcription with two unusually potent subdomains at its C terminus." J Biol Chem **280**(2): 1401-7.

Park, I. H., R. Zhao, et al. (2008). "Reprogramming of human somatic cells to pluripotency with defined factors." Nature **451**(7175): 141-6.

Perry, P., S. Sauer, et al. (2004). "A dynamic switch in the replication timing of key regulator genes in embryonic stem cells upon neural induction." Cell Cycle **3**(12): 1645-50.

Polejaeva, I. A., S. H. Chen, et al. (2000). "Cloned pigs produced by nuclear transfer from adult somatic cells." Nature **407**(6800): 86-90.

Rege, T. A. and J. S. Hagood (2006). "Thy-1, a versatile modulator of signaling affecting cellular adhesion, proliferation, survival, and cytokine/growth factor responses." Biochim Biophys Acta **1763**(10): 991-9.

Rideout, W. M., 3rd, K. Hochedlinger, et al. (2002). "Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy." Cell **109**(1): 17-27.

Ringrose, L. and R. Paro (2004). "Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins." Annu Rev Genet **38**: 413-43.

Schoor, M., K. Schuster-Gossler, et al. (1993). "The Etl-1 gene encodes a nuclear protein differentially expressed during early mouse development." Dev Dyn **197**(3): 227-37.

Schwartz, Y. B., T. G. Kahn, et al. (2006). "Genome-wide analysis of Polycomb targets in Drosophila melanogaster." Nat Genet **38**(6): 700-5.

Shin, T. H. and C. C. Mello (2003). "Chromatin regulation during C. elegans germline development." Curr Opin Genet Dev **13**(5): 455-62.

Smith, A. G. (2001). "Embryo-derived stem cells: of mice and men." Annu Rev Cell Dev Biol **17**: 435-62.

Stadtfeld, M., N. Maherali, et al. (2008). "Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse." Cell Stem Cell **2**(3): 230-40.

Strumpf, D., C. A. Mao, et al. (2005). "Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst." Development **132**(9): 2093-102.

Szutorisz, H., C. Canzonetta, et al. (2005). "Formation of an active tissue-specific chromatin domain initiated by epigenetic marking at the embryonic stem cell stage." Mol Cell Biol **25**(5): 1804-20.

Tada, M., Y. Takahama, et al. (2001). "Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells." Curr Biol **11**(19): 1553-8.

Takahashi, K., K. Tanabe, et al. (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-72.

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

Tarkowski, A. K. and J. Wroblewska (1967). "Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage." J Embryol Exp Morphol **18**(1): 155-80.

Teitell, M. A. (2005). "The TCL1 family of oncoproteins: co-activators of transformation." Nat Rev Cancer **5**(8): 640-8.

Thomson, J. A., J. Itskovitz-Eldor, et al. (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-7.

Tokuzawa, Y., E. Kaiho, et al. (2003). "Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development." Mol Cell Biol **23**(8): 2699-708.

Tolhuis, B., I. Muijrers, et al. (2006). "Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster." Nat Genet **38**(6): 694-9.

Wakayama, T., A. C. Perry, et al. (1998). "Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei." Nature **394**(6691): 369-74.

Wakayama, T. and R. Yanagimachi (1999). "Cloning of male mice from adult tail-tip cells." Nat Genet **22**(2): 127-8.

Wang, J., S. Rao, et al. (2006). "A protein interaction network for pluripotency of embryonic stem cells." Nature **444**(7117): 364-8.

Wernig, M., A. Meissner, et al. (2008). "c-Myc is dispensable for direct reprogramming of mouse fibroblasts." Cell Stem Cell **2**(1): 10-2.

Wernig, M., A. Meissner, et al. (2007). "In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state." Nature.

Wernig, M., J. P. Zhao, et al. (2008). "Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease." Proc Natl Acad Sci U S A **105**(15): 5856-61.

Wilmut, I., A. E. Schnieke, et al. (1997). "Viable offspring derived from fetal and adult mammalian cells." Nature **385**(6619): 810-3.

Young, L. E., K. D. Sinclair, et al. (1998). "Large offspring syndrome in cattle and sheep." Rev Reprod **3**(3): 155-63.

Yu, J., M. A. Vodyanik, et al. (2006). "Human embryonic stem cells reprogram myeloid precursors following cell-cell fusion." Stem Cells **24**(1): 168-76.

Yu, J., M. A. Vodyanik, et al. (2007). "Induced pluripotent stem cell lines derived from human somatic cells." Science **318**(5858): 1917-20.

Zaehres, H., M. W. Lensch, et al. (2005). "High-efficiency RNA interference in human embryonic stem cells." Stem Cells **23**(3): 299-305.

Ziomek, C. A. and M. H. Johnson (1982). "The roles of phenotype and position in guiding the fate of 16-cell mouse blastomeres." Dev Biol **91**(2): 440-7.

Ziomek, C. A., M. H. Johnson, et al. (1982). "The developmental potential of mouse 16-cell blastomeres." J Exp Zool **221**(3): 345-55.

# Chapter 2

## ANALYSIS OF THE MOUSE EMBRYONIC STEM CELL REGULATORY NETWORKS OBTAINED BY CHIP-CHIP AND CHIP-PET

Divya Mathur[1, 2, 4], Timothy W. Danford[3, 4], Laurie A. Boyer[1], Richard A. Young[1, 2], David K. Gifford[2, 3]* and Rudolf Jaenisch[1, 2]*.

[1]Department of Biology, Massachusetts Institute of Technology, 32 Ames Street, Cambridge, MA 02139, USA

[2]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

[3]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, MA 02139, USA

[4]These authors contributed equally to this work.

*Corresponding authors

### RESPECTIVE CONTRIBUTIONS

DM conducted all experiments for the paper and wrote the manuscript. TWD carried out data analyses and assisted with writing the manuscript. LAB conceived of the project and provided advice on the manuscript. RAY, DKG and RJ provided advice and funding for this work.

**Accession Numbers:**

This study contains array data, which has been submitted to ArrayExpress (http://www.ebi.ac.uk/arrayexpress). One whole chromosome array and a 2-slide promoter array set were used. The array design accession numbers and login information for reviewers is:

*Chromosome Array*: [A-MEXP-956]

Username: [Reviewer_A-MEXP-956]

Password: 1197968963026

*Promoter Array Slide 1*: [A-MEXP-957]

Username: [Reviewer_A-MEXP-957]

Password: 1197969012719

*Promoter Array Slide 2*: [A-MEXP-958]

Username: [Reviewer_A-MEXP-958]

Password: 1197969030574

All ChIP-chip experiments have also been deposited with Arrayexpress (accession code: [E-TABM-410]).

**ABSTRACT**

Genome-wide approaches have begun to reveal the transcriptional networks responsible for pluripotency in embryonic stem (ES) cells. Chromatin Immunoprecipitation (ChIP) followed either by hybridization to a microarray platform (ChIP-chip) or by DNA sequencing (ChIP-PET), has identified binding targets of the ES cell transcription factors Oct4 and Nanog in humans and mice, respectively. These studies have provided an outline of the transcriptional framework involved in maintaining pluripotency. Recent evidence with comparing multiple technologies suggests that expanding these datasets using different platforms would be a useful resource for examining the mechanisms underlying pluripotency regulation. We have now identified Oct4 and Nanog genomic targets in mouse ES cells by ChIP-chip and provided the means to compare these data with previously reported ChIP-PET results in mouse ES cells. We have mapped the sequences of Oct4 and Nanog binding events from each data set to genomic coordinates, providing a valuable resource to facilitate a better understanding of the ES cell regulatory circuitry. Interestingly, although considerable differences are observed in Oct4 and Nanog occupancy as identified by each method, a substantial number of targets in both data sets are enriched for genes with known roles in cell-fate specification and are differentially expressed upon *Oct4* or *Nanog* knockdown. This study suggests that each data set is a partial representation of the overall ES cell regulatory circuitry, and through integrating binding data obtained by ChIP-chip and ChIP-PET, the

methods presented here provide a useful means for integrating datasets obtained by different techniques in the future.

## INTRODUCTION

Embryonic stem (ES) cells are derived from the inner cell mass (ICM) of the embryo and possess the property of pluripotency, which is the ability to develop into any cell lineage of the organism (Evans and Kaufman 1981; Martin 1981; Thomson, Itskovitz-Eldor et al. 1998). The derivation of these cells has had significant impact on biomedical research and has important implications for regenerative medicine. Consequently, a detailed knowledge of the mechanisms governing pluripotency in ES cells is necessary to realize the potential of these cells. The homeodomain transcription factors Oct4 and Nanog are uniquely expressed in pluripotent cell types and have essential roles during development (Boyer, Mathur et al. 2006; Niwa 2007). For instance, *Oct4* knockout embryos and ES cells differentiate into trophectoderm, whereas overexpression of the gene leads to differentiation into primitive endoderm and mesoderm lineages (Nichols, Zevnik et al. 1998; Niwa, Miyazaki et al. 2000). Loss of *Nanog* in the early embryo and ES cells results in differentiation into primitive endoderm (Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003). Conversely, Nanog over-expression obviates the need for the cytokine, leukemia inhibitory factor (LIF) in ES cell self-renewal(Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003). Collectively, these studies suggest that Oct4 and Nanog function in

concert to regulate pluripotency in the early embryo, and similarly in ES cells to govern the transcriptional regulatory circuitry.

Recent genomic studies in ES cells have provided the foundation for understanding the genetic network that is the collective output of these pluripotency factors. Studies in both human and mouse ES cells have used Chromatin Immunoprecipitation (ChIP) combined with genome-wide technologies to uncover Oct4 and Nanog genomic binding events that may underlie transcriptional regulatory circuitries involved in maintaining a stem cell state (Boyer, Lee et al. 2005; Loh, Wu et al. 2006; Kim, Chu et al. 2008). Such investigations have shown that in both species, Oct4 and Nanog occupy a large number of transcriptionally active and silent genes, many of which are transcriptional regulators that have been implicated in lineage specification and cell fate determination. Moreover, a substantial overlap between the Oct4 and Nanog genomic targets exists within each data set, suggesting that these two factors act in concert to regulate a common set of downstream pathways. This has been further substantiated by gene-expression studies following shRNA-mediated knockdown of *Oct4* and *Nanog* (Ivanova, Dobrin et al. 2006).

ChIP coupled with a genome-wide DNA detection platform has been useful in studying protein-DNA interactions. The data obtained from these different platforms, however, are expected to exhibit variations due to the technical differences in the methods, as well as in data analysis. To date, ES cell binding

data have been collected using ChIP-PET (Paired End Ditags) (Loh, Wu et al. 2006) and ChIP-chip (Kim, Chu et al. 2008) for mouse ES cells and ChIP-chip for human ES cells (Boyer, Lee et al. 2005). However, comprehensive technological comparisons between ChIP-chip and ChIP-PET indicate that composite data sets that incorporate information from multiple platforms in a complementary fashion will be most useful in examining these networks in a comprehensive manner (Euskirchen, Rozowsky et al. 2007). Such analysis is necessary since the binding data obtained from different platforms can vary due to the differences in sample processing for each method. In the study by Kim et al (Kim, Chu et al. 2008), the authors provide a comparison between Oct4 and Nanog targets obtained from ChIP-chip and previously reported ChIP-PET data (Loh, Wu et al. 2006). However, such overlap can vary dramatically depending on the thresholds used for determining bound regions by each experimental method. Since these thresholds are to a large extent, arbitrary, it is important to examine how the binding data obtained by different platforms change under a wide range of threshold values.

To this end, we have employed ChIP-chip to identify the genomic binding targets of the pluripotency factors Oct4 and Nanog in mouse ES cells. Additionally, we have devised methods to examine these results along with previously published data for these factors using ChIP-PET under a wide range of binding thresholds (Loh, Wu et al. 2006). All data have been re-mapped to the same version of the mouse genome, and provide a resource for studying this expanded

transcriptional network obtained by integrating our ChIP-chip data and previously reported ChIP-PET results. Our analyses revealed substantially different sets of Oct4 and Nanog targets identified by each technique. However, a significant proportion of these targets included genes encoding transcription factors and other regulators of development in both data sets. Interestingly, many of the genes identified in both studies were differentially expressed upon *Oct4* or *Nanog* knockdown in ES cells, suggesting that these targets were regulated by Oct4 and Nanog. Importantly, an examination of multiple data sources provided in this study has revealed a more comprehensive framework for understanding the mouse ES cell regulatory network.

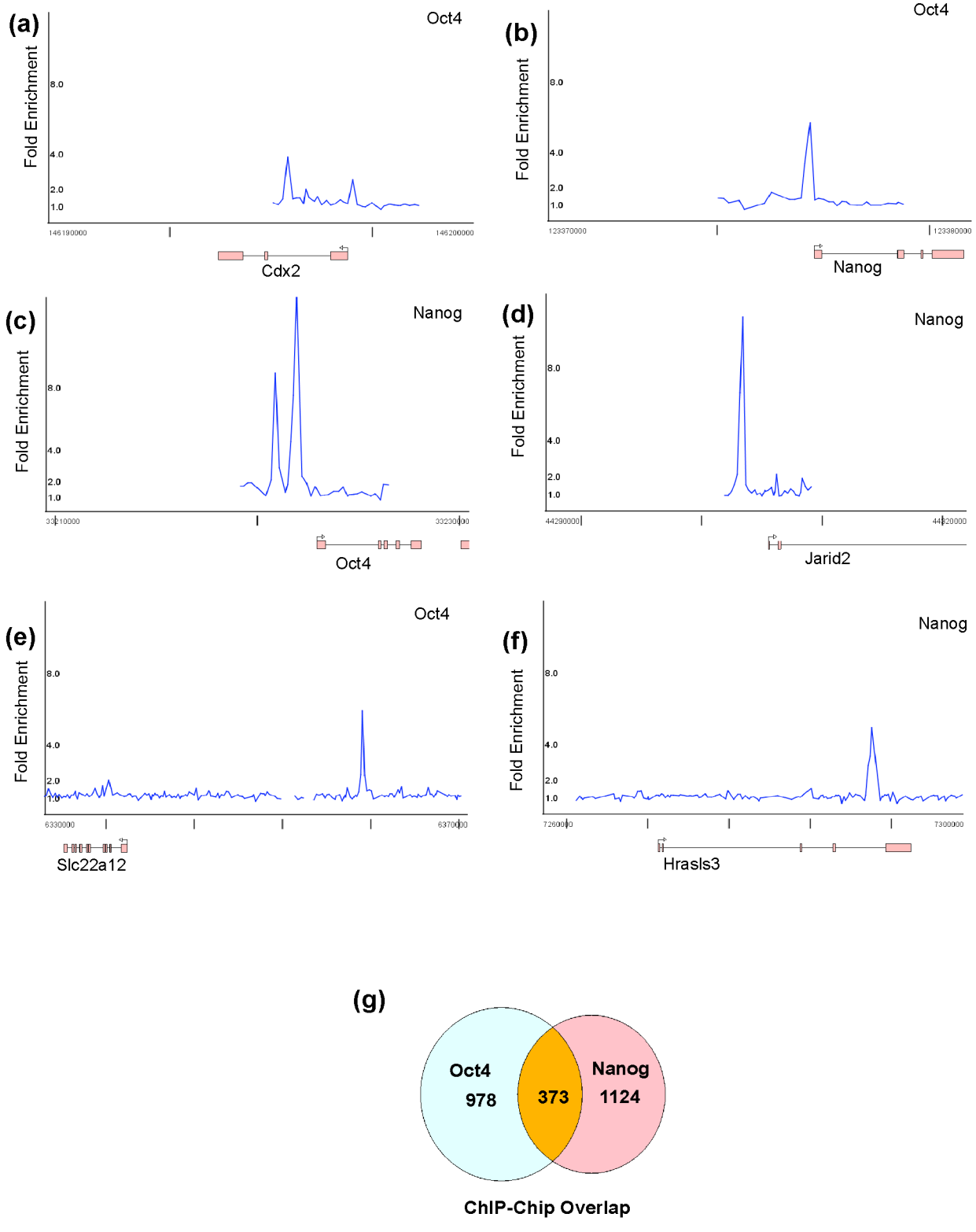
## RESULTS

### Oct4 and Nanog ChIP-Chip Binding Data

DNA sequences occupied by Oct4 and Nanog in mouse ES cells were identified in three independent biological replicates using ChIP as previously described (Boyer, Lee et al. 2005). Samples were hybridized to microarrays that contained oligonucleotide probes that span the region –4 to +4 kb relative to the transcriptional start sites (TSSs) for 19,993 annotated mouse genes and 258 miRNAs (see supplementary notes). Based on previously established criteria, bound regions were identified as peaks of ChIP-enriched DNA that span closely neighboring probes (Figure 1a, b, c, d; see supplementary note) (Boyer, Lee et al. 2005). Moreover, only those regions that were bound in all three replicates are represented in the final data set. Using these stringent parameters, we

**Figure 1.** Illustrative examples of ChIP-enrichment ratios of Oct4 (a, b, e) and Nanog (c, d, f) bound regions obtained from ChIP-Chip experiments with promoter arrays (a-d) and chromosome 19 arrays (e. f). The chromosomal position of the genes, as well as the genomic scale is represented along the x-axis. The fold enrichment of the probes is shown on the y-axis. These enrichment ratios represented the medians of the per-pixel ratios scanned at each spot on the microarray. Exons and introns are represented by boxes and horizontal lines, respectively. The transcription start site and direction of transcription are denoted by arrows. (g) Venn diagram depicting the overlap between gene whose promoters were bound by both Oct4 and Nanog in ChIP-Chip experiments (p-value< 0.001).

**FIGURE 1.**


(a)
Oct4
Fold Enrichment
8.0
4.0
2.0
1.0
146190000
146200000
Cdx2
(b)
Oct4
Fold Enrichment
8.0
4.0
2.0
1.0
123370000
123380000
Nanog
(c)
Nanog
Fold Enrichment
8.0
4.0
2.0
1.0
33210000
33230000
Oct4
(d)
Nanog
Fold Enrichment
8.0
4.0
2.0
1.0
44290000
44320000
Jarid2
(e)
Oct4
Fold Enrichment
8.0
4.0
2.0
1.0
6330000
6370000
Slc22a12
(f)
Nanog
Fold Enrichment
8.0
4.0
2.0
1.0
7260000
7300000
Hrasls3
(g)
Oct4
978
373
Nanog
1124
ChIP-Chip Overlap

identified 1351 (6.8%) and 1124 (5.6%) known protein-coding genes (Table S1) and 22 (8.5%) and 23 (8.9%) microRNA genes (Table S2) that are occupied by Oct4 and Nanog respectively.

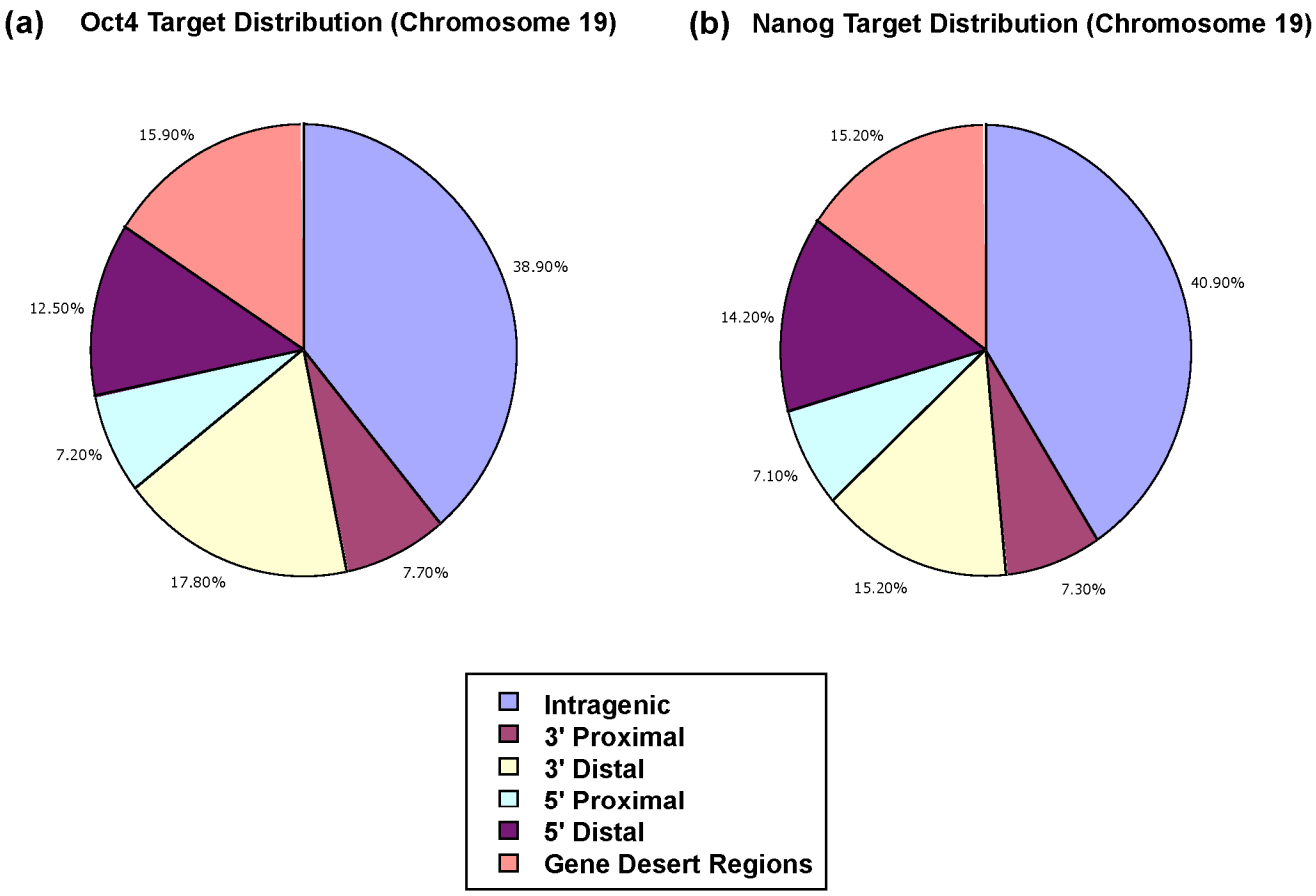
Several lines of evidence indicated that this ChIP-chip data set is of high quality. First, in accordance with previous findings in both mouse and human ES cells (Boyer, Lee et al. 2005; Loh, Wu et al. 2006), gene ontology (GO) analyses revealed that a significant number of promoters occupied by Oct4 and Nanog contained transcription factors and genes involved in developmental processes (Table S3). Some of these genes, such as *Jarid2, Cdx2 and Sox2* have been identified previously as Oct4 or Nanog targets (Loh, Wu et al. 2006). Additionally, as seen in both the human and mouse ES cell studies, Oct4 and Nanog bind to their own as well as each other's promoters (Boyer, Lee et al. 2005; Chew, Loh et al. 2005; Rodda, Chew et al. 2005; Loh, Wu et al. 2006). We also observed a substantial overlap between the Oct4 and Nanog-bound genes, where 373 gene promoters were occupied by both these factors (Figure 1g). Together, these binding data support prior models suggesting that Oct4 and Nanog act together to maintain ES cell pluripotency by promoting self-renewal and by regulating a number of developmentally important genes.

Given that it has been reported that a significant number of binding sites may be located outside of promoter regions (Loh et al., 2006), we next hybridized the Oct4 and Nanog ChIP samples to chromosome arrays that tiled the entire non-

repeat portion of mouse chromosome 19 (see supplementary note). Binding events were analyzed similarly to the promoter arrays and occupied regions were identified using the same criteria (Table S4). In addition to promoter regions bound by Oct4 and Nanog, this analysis revealed Oct4 and Nanog binding sites that were undetectable on the promoter arrays (Figure 1 e and f). Bound regions were classifed in relation to the nearest gene within 100 kb as: 5' proximal (0-10 kb upstream), 5' distal (10-100 kb upstream), 3' proximal (0-10 kb downstream), 3' distal (10-100 kb downstream) and intragenic (within the gene). Sites that were located >100 kb away from the nearest gene were classified as gene desert regions. 208 binding events for Oct4 and 381 for Nanog were identified using the chromosome array. For both factors we observed similar trends in distribution of these binding sites across chromosome 19 (Figure 2). Among Oct4 targets, 38.9% of the sites were in intragenic, 7.7% in the 3' proximal, 17.8% in the 3' distal, 7.2% in 5' proximal, 12.5% in 5' distal, and 15.9% in gene desert regions. Following a similar distribution, the Nanog data showed 40.9% of the binding sites in intragenic, 7.3% in the 3' proximal, 15.2% in the 3' distal, 7.1% in 5' proximal, 14.2% in 5' distal, and 15.2% in gene desert regions. These results show that Oct4 and Nanog targets are located across different genomic regions, and such extensively tiled arrays can be used to obtain more detailed binding data on a genome-wide scale. Additionally, the finding that approximately 40% of the binding sites were found in intragenic regions is also in concordance with earlier observations made in the ChIP-PET study for both Oct4 and Nanog,

**Figure 2.** Genomic distribution of (a) Oct4 and (b) Nanog binding sites on mouse chromosome 19, obtained by ChIP-chip analyses.

**FIGURE 2.**


(a) Oct4 Target Distribution (Chromosome 19)
15.90%
38.90%
12.50%
7.20%
17.80%
7.70%
(b) Nanog Target Distribution (Chromosome 19)
15.20%
40.90%
14.20%
7.10%
15.20%
7.30%
Intragenic
3' Proximal
3' Distal
5' Proximal
5' Distal
Gene Desert Regions

indicating that the Chromosome 19 array results are representative of the binding distribution in the genome.

**Oct4 and Nanog ChIP-PET Data**

In order to compare genomic targets across different platforms, we re-analyzed previously reported ChIP-PET experimental data for Oct4 and Nanog (Loh, Wu et al. 2006). In the ChIP-PET method, immuno-enriched DNA fragments are cloned into a plasmid library, which is then transformed into one containing concatenated signature paired-end ditag sequences. The DNA fragments or binding sites are subsequently sequenced and the reads are mapped to the mouse genome. All binding sites were first classified relative to the nearest gene (as intragenic, 5' distal, 5' proximal, 3' distal, 3' proximal and gene desert regions), according to the criteria described earlier. Next, we performed GO analyses on the ChIP-PET targets in each of these regions. As summarized in Table S3, we observed that similar to ChIP-chip data, both Oct4 and Nanog binding targets had a significant representation of genes encoding transcription factors and regulators of cell fate.

In order to analyze the ChIP-PET and ChIP-Chip data together, all raw ChIP-PET sequence reads were re-mapped to the same version of the mouse genome (*mm6*) used in the ChIP-chip experiments. The sequence reads were between 34-36 bp, and only those that had at least 34 matched base pairs and a gap-length of 10 bp were considered to be uniquely mapped to the mouse *mm6*

genome. Out of 951,437 Oct4 reads, 198,802 (20.9%) could be uniquely mapped. Similarly, among 624,237 Nanog reads, 333,248 reads (53.4%) could be mapped uniquely to the genome. Importantly, the methods and criteria used to remap data to a different genome version will provide a useful resource for performing such analyses with other sequencing based platforms that use other genome versions.

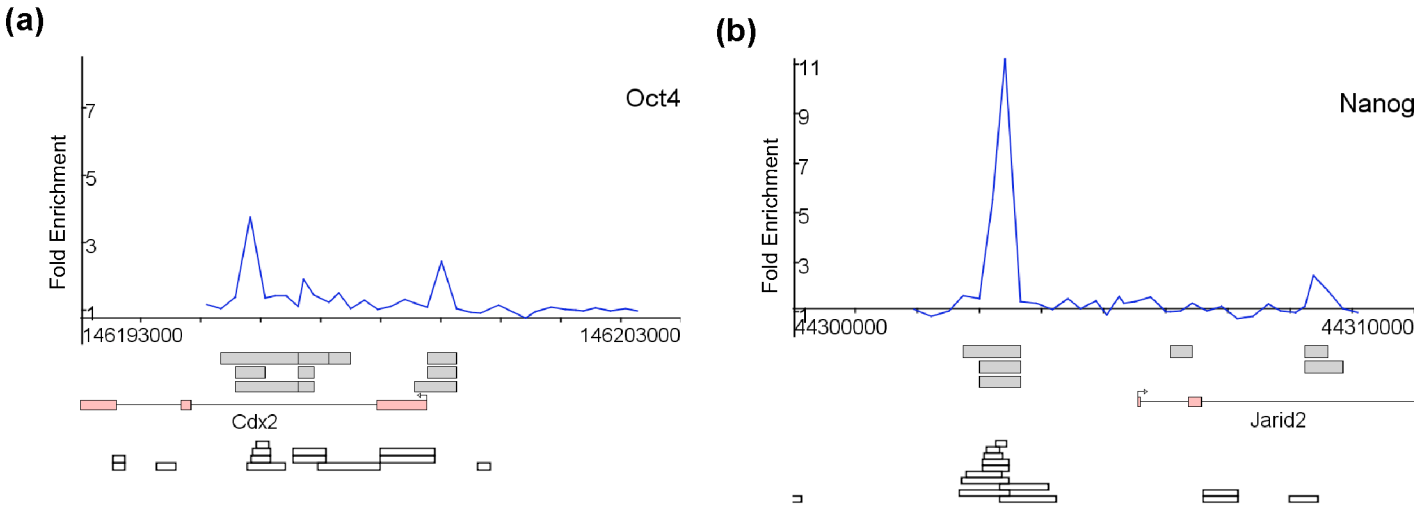
In ChIP-PET experiments, a minimum number of overlapping sequence reads was used as a criterion for identifying binding events. A region was considered occupied by Oct4 and Nanog if it had at least 4 or 3 overlapping sequence reads, respectively. In order to analyze our ChIP-chip findings in relation to these data, we examined only those ChIP-PET reads that had corresponding regions represented on the mouse promoter arrays (576 for Oct4 and 924 for Nanog). Additionally, for Chromosome 19, 90 Oct4 targets and 224 Nanog targets could be remapped for the ChIP-PET data.

**Examination of ChIP-chip and ChIP-PET Bound Regions**

To examine the binding events obtained by ChIP-chip and ChIP-PET, we used the 'Genomic Spatial Events' (GSE) Visualizer program (Danford TW 2007) (Figure 3, see also Supplementary Note). GSE is a Java software package, written to allow interactive browsing of ChIP-chip and ChIP-PET data, and genome annotations, from a remote database over a network connection (software for this program is available upon request). It handles datasets that are

**Figure 3.** GSE Spatial Visualizer snapshots showing illustrative examples of ChIP-Chip and ChIP-PET data for (A) Oct4 and (B) Nanog. The fold enrichment for a single ChIP-Chip replicate (for Oct4 or Nanog) is shown against the genomic coordinate scale for the gene (in base pairs). The grey boxes represent the locations of 'bound regions' from each of the factor's three ChIP-Chip replicates. The white boxes show the overlapping ChIP-PET reads for the displayed region. A ‘bound region’ in ChIP-PET experiments had 4 or more overlapping reads in Oct4, and 3 or more overlapping reads in the case of Nanog. Gene exons and introns are represented by pink boxes and solid horizontal lines respectively. For each visualized gene, the transcriptional start site, direction of transcription and RefSeq annotation derived from the UCSC database are also specified.

**FIGURE 3.**


(a)
Oct4
Fold Enrichment
7
5
3
1
146193000
146203000
Cdx2
(b)
Nanog
Fold Enrichment
11
9
7
5
3
1
44300000
44310000
Jarid2

simultaneously mapped against multiple genome builds, a requirement for any system that is to compare new experimental data against older datasets. The software is built to run on multiple platforms, and also provides a software interface for custom-written analysis modules. The locations of bound probes from replicate ChIP-chip experiments, as well as the overlapping ChIP-PET reads for the respective regions could be simultaneously visualized using the program. Therefore, this tool provides an important resource to compare data from multiple sources at a variety of genomic scales. It can also be utilized in the future for such purposes as data using other technologies become available for expanding the ES cell transcriptional circuitry.

In order to determine how the analysis methods and threshold criteria in ChIP-chip and Chip-PET experiments influenced the overall concordance between data sets, we examined the data by generating 'Recovery Curves' (see methods; Figures 4 and 5). A binding event in one experiment was considered 'recovered' by (or overlapping with) a similar event in a second experimental type if both events were within a fixed genomic distance (recovery distance) from each other. A typical p-value threshold of 0.001 was used initially to determine significant binding events in ChIP-chip experiments, and a minimum number of 'overlapping sequence reads' was used to establish bound regions in ChIP-PET experiments (4 or more overlapping reads for Oct4 targets, and 3 or more overlapping reads for Nanog). We generated two types of Recovery curves to analyze the ChIP-chip and ChIP-PET data. The ChIP-PET Recovery curve examined the fraction

of ChIP-PET regions overlapping with the ChIP-chip data at a wide range of p-value thresholds for the ChIP-chip experiments. In this instance, the threshold criteria were kept constant for the ChIP-PET experiments, and the ChIP-PET recovery (y-axis) was plotted against a range of ChIP-chip p-values (x-axis). Conversely, the other type of curve represented the ChIP-chip recovery at varying ChIP-PET 'overlapping read' threshold values. The ChIP-chip p-value threshold was kept constant at 0.001, and the ChIP-chip recovery (y-axis) was examined at different numbers of ChIP-PET sequence reads (x-axis). We examined each type of curve under a range of recovery distances, as binding events identified by both methods may not have exact overlaps due to differences in sample processing and technologies.

Not surprisingly, we observed that the recoveries of Oct4 and Nanog targets obtained by one experimental method increased as the threshold value for the other method was relaxed. The recoveries also increased as the distance permitted between a ChIP-chip peak and corresponding ChIP-PET peak was increased (Figures 4**)**. As an example of these results, converting the recoveries into percentages, among Oct4-bound regions, 24% of the peaks identified by ChIP-PET (>4 reads) were recovered in the ChIP-Chip data (p-value<0.001) within a distance of 1 kb. Conversely, using the same thresholds, 9.3% of the Oct4-bound peaks found by ChIP-chip were recovered in the ChIP-PET data. (Table S5). From the Nanog data we observed that 28.1% of the ChIP-PET peaks (>3 reads) were recovered in ChIP-chip bound regions (p-value<0.001)

**Figure 4. Oct4 and Nanog Promoter Array Recovery Curves.** (A) and (C) show the Oct4 and Nanog ChIP-PET recovery curves respectively for the promoter arrays. These represent the fraction of ChIP-PET recovery under a range of ChIP-Chip p-value cut-offs. (B) and (D) are Oct4 and Nanog ChIP-chip recovery curves respectively. These show the ChIP-Chip percent recovery at varying ChIP-PET read thresholds. In all cases, recovery curves are made for a variety of distances (0-8 kb) permitted between a ChIP-PET peak and ChIP-PET read for them to be considered 'overlapping'.

**FIGURE 4.**

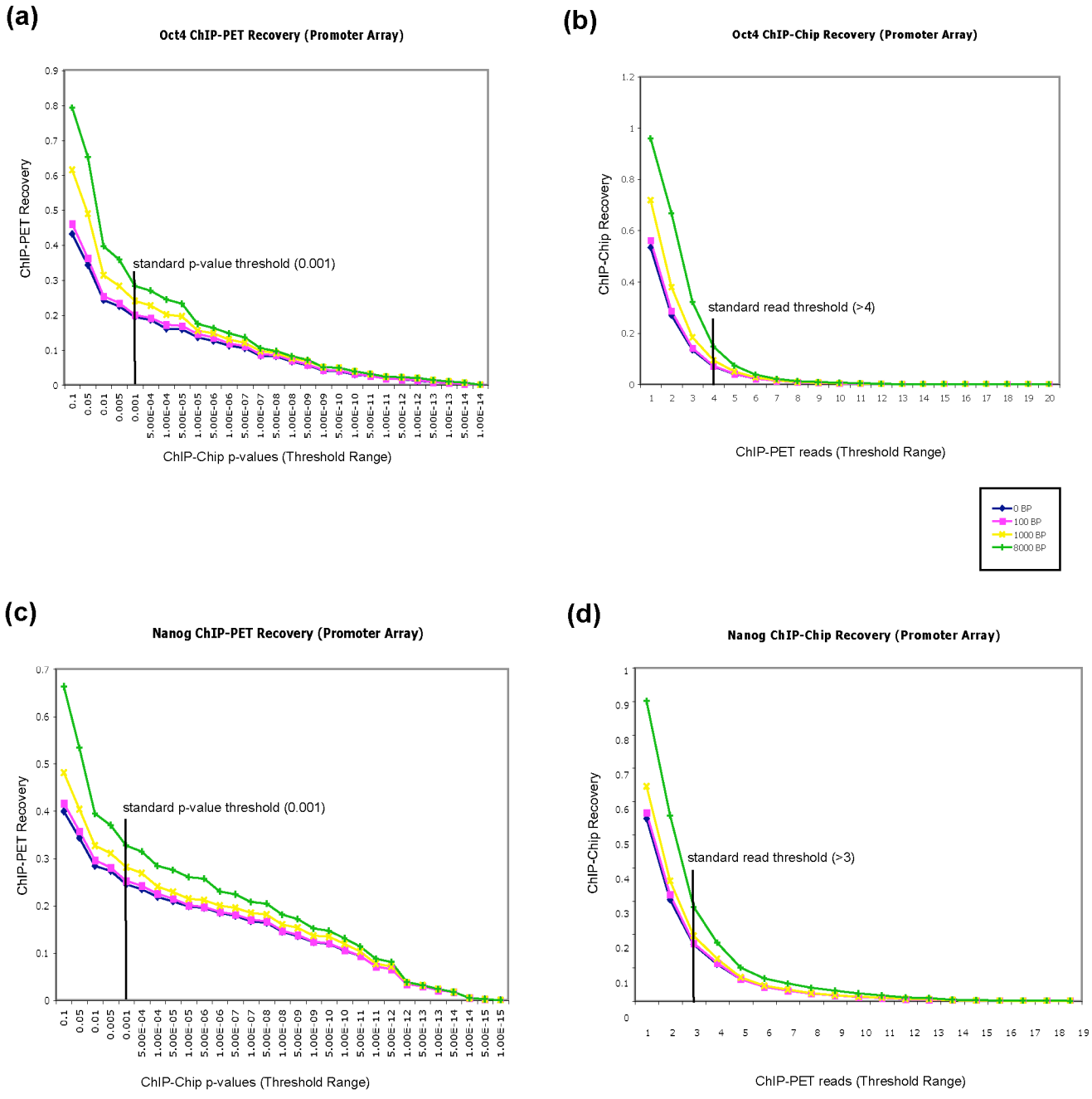
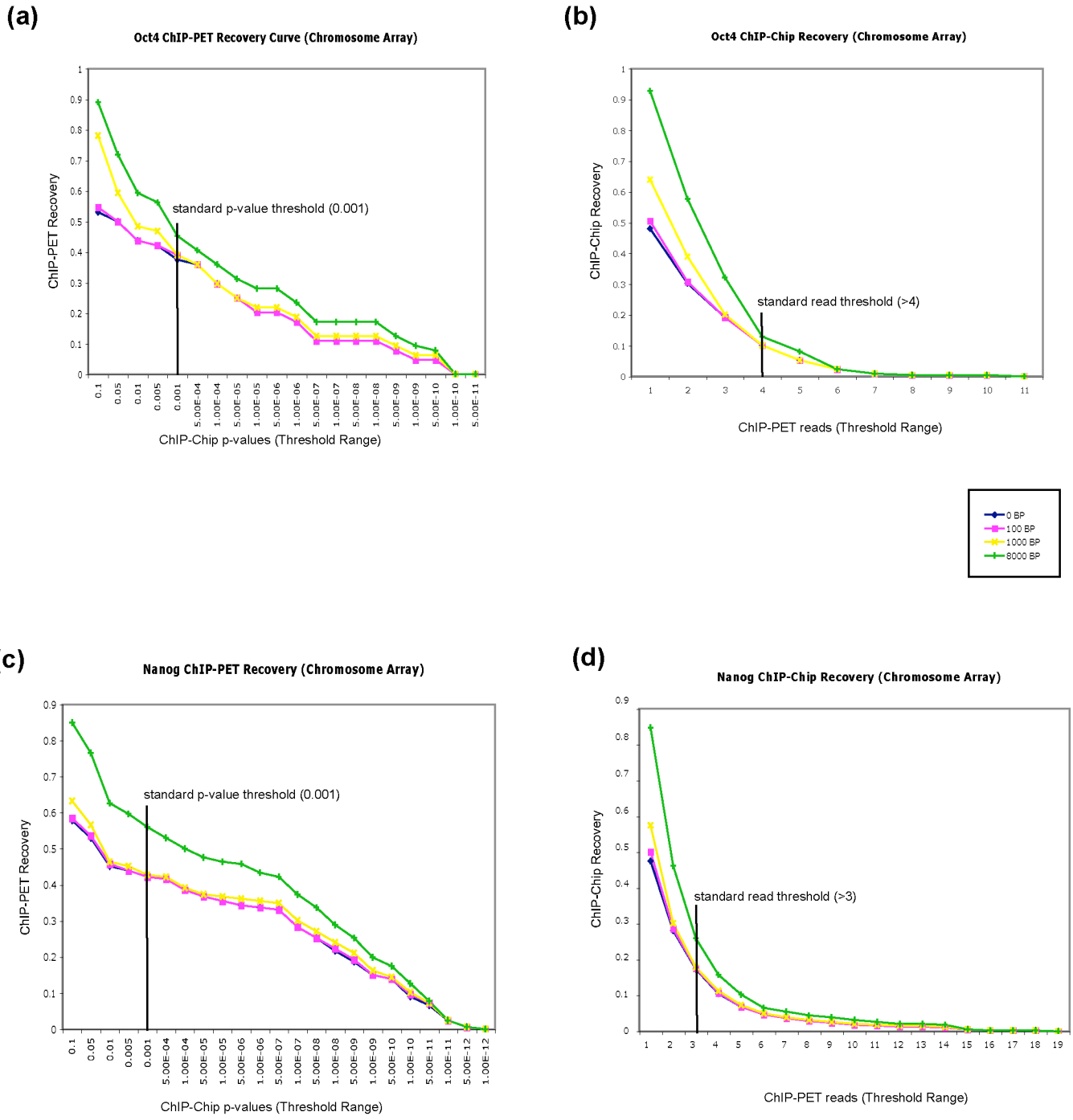

(a)
Oct4 ChIP-PET Recovery (Promoter Array)
ChIP-PET Recovery
standard p-value threshold (0.001)
ChIP-Chip p-values (Threshold Range)
(b)
Oct4 ChIP-Chip Recovery (Promoter Array)
ChIP-Chip Recovery
standard read threshold (>4)
ChIP-PET reads (Threshold Range)
0 BP
100 BP
1000 BP
8000 BP
(c)
Nanog ChIP-PET Recovery (Promoter Array)
ChIP-PET Recovery
standard p-value threshold (0.001)
ChIP-Chip p-values (Threshold Range)
(d)
Nanog ChIP-Chip Recovery (Promoter Array)
ChIP-Chip Recovery
standard read threshold (>3)
ChIP-PET reads (Threshold Range)

**Figure 5. Oct4 and Nanog Chromosome Array Recovery Curves.** (A) and (C) show the Oct4 and Nanog ChIP-PET recovery curves respectively for the chromosome array. These represent the fraction of ChIP-PET recovery under a range of ChIP-Chip p-value cut-offs. (B) and (D) are Oct4 and Nanog ChIP-chip recovery curves respectively. These show the ChIP-Chip percent recovery at varying ChIP-PET read thresholds. In all cases, recovery curves are made for a variety of distances (0-8 kb) permitted between a ChIP-PET peak and ChIP-PET read for them to be considered 'overlapping'.

# FIGURE 5.


(a)
Oct4 ChIP-PET Recovery Curve (Chromosome Array)
ChIP-PET Recovery
standard p-value threshold (0.001)
ChIP-Chip p-values (Threshold Range)
(b)
Oct4 ChIP-Chip Recovery (Chromosome Array)
ChIP-Chip Recovery
standard read threshold (>4)
ChIP-PET reads (Threshold Range)
0 BP
100 BP
1000 BP
8000 BP
(c)
Nanog ChIP-PET Recovery (Chromosome Array)
ChIP-PET Recovery
standard p-value threshold (0.001)
ChIP-Chip p-values (Threshold Range)
(d)
Nanog ChIP-Chip Recovery (Chromosome Array)
ChIP-Chip Recovery
standard read threshold (>3)
ChIP-PET reads (Threshold Range)

within a 1 kb area. Conversely, the ChIP-Chip percent recovery from ChIP-PET bound regions (>3 reads) was 19.5% (Table S6). Therefore, these recovery curves illustrated the importance of recovery distance and threshold calibration in examining data from different sources.

Similar analyses were performed using the mouse chromosome 19 data and corresponding ChIP-PET regions. We noted that the amount of overlap between ChIP-PET and ChIP-chip increased with the more extensively tiled arrays (Figure 5). This is because many targets identified in ChIP-PET experiments would not be identified by the promoter arrays, since regions outside of the promoter were not represented on these arrays. In summary, the Oct4 and Nanog ChIP-chip and ChIP-PET data sets show that the recovery among datasets varies as any of the threshold criteria for binding events are altered. Further evaluation of the binding events identified through both techniques, by genetic manipulation of the corresponding genes in ES cells, will lend better insight into the genes responsible for maintaining pluripotency.

Previous reports have suggested that a lack of concordance between array- and sequencing based technologies may also be due to the repeat-masking feature of tiled microarrays as well as a sequencing depth issue with ChIP-PET (Euskirchen, Rozowsky et al. 2007; Kim, Chu et al. 2008). Since 99% of the ChIP-chip probes on our promoter arrays do not have any major overlaps with repeat regions, and only 8.1% of all ChIP-PET sequences fall in repeat-masked

regions, we do not expect the results of this study to change by any significant degree if this small fraction of ChIP-PET sequences is removed from the analysis. In order to further examine the sequencing depth issue, the ChIP-Chip and ChIP-PET data on Chromosome 19 was used to performed a sequence-depth analysis to examine the changes in ChIP-chip recovery as increasing number of ChIP-PET sequences are randomly sampled (Figure S3 and Table S10). According to our observations for both Oct4 and Nanog, the number of ChIP-chip targets recovered increased with the number of ChIP-PET reads sampled, and did not approach a saturation point, even when all ChIP-PET reads for chromosome 19 were sampled. This result suggests that the lack of recovery of ChIP-chip targets in the ChIP-PET data can, at least in part, be explained by a lack of depth in sequencing.

**Differentially Regulated Targets of OCT4 and NANOG**

Since protein-DNA binding alone is not indicative of a regulatory event, the expression of Oct4 and Nanog binding targets obtained through ChIP-chip and ChIP-PET was compared by comparing binding data with previously published Oct4 and Nanog RNAi gene expression profiles in ES cells (Loh, Wu et al. 2006). The expression levels of targets determined exclusively by either technique, and those overlapping in both were examined in *Oct4* or *Nanog* knockdown ES cells (as summarized in Table 1. See also Tables S7 and S8 and supplementary note). We found that among the Oct4-bound targets (with corresponding

**Table 1.** Differential expression of Oct4 and Nanog targets in RNAi experiments.

| **Method for determining bound targets** | **Percent of differentially expressed Oct4 targets on Oct4 knockdown** | **Percent of differentially expressed Nanog targets on Nanog knockdown** |
|---|---|---|
| ChIP-Chip only | 33.9% (390/1151) | 21.4% (192/898) |
| ChIP-PET only | 29% (114/393) | 14.8% (91/616) |
| ChIP-chip+ChIP-PET | 70.3% (83/118) | 33.5% (73/218) |
| **% up (+) and down(-) regulated genes** | 39.4% (+)<br>60.6% (-) | 50.6% (+)<br>49.4% (-) |

Affymetrix probes) determined only by ChIP-chip, 33.9% were differentially expressed upon *Oct4* knockdown. Similarly, 29% of the Oct4 targets detected solely by ChIP-PET were differentially regulated. Interestingly, for Oct4 targets obtained by both ChIP-PET and ChIP-chip, 70.3% showed changes in gene expression upon downregulation of Oct4. In Nanog knockdown ES cells, 21.4% of the targets determined solely by ChIP-chip, and 14.8% identified only by ChIP-PET were differentially expressed compared to normal ES cells. This percentage increased for targets that were identified by both techniques, where 33.5% were differentially regulated upon *Nanog* knockdown. These analyses also showed that among the differentially regulated targets of Nanog, the distribution between up-and down-regulated genes upon *Nanog* knockdown was approximately equal. However, in the case of Oct4 regulated targets, a larger percentage of genes were downregulated (60.4%) upon *Oct4* knockdown. These results suggest that both Nanog and Oct4 can potentially activate or repress their binding targets. Therefore, these analyses have revealed a higher-value set Oct4 and Nanog-regulated genes, by collectively examining the targets identified by ChIP-chip and ChIP-PET.

The functional relevance of the ChIP-chip and ChIP-PET data, as examined by GO analyses, had revealed that the Oct4 and Nanog bound regions were significantly enriched for transcriptional and developmentally important regulators of gene expression. Similar observations had been made earlier for these factors in human ES cells as well (Boyer, Lee et al. 2005). For instance, among genes

that displayed changes in expression levels upon *Oct4* RNAi-mediated knockdown, certain genes including *Sox2* and *Rif1*, which have important roles in development, were bound by Oct4 in both ChIP-chip and ChIP-PET experiments. However, other genes that play a part in cell-fate determination, such as *Gdf3* and *Notch4* (Zhu, Zhang et al. 2006; Andersson, Bertolino et al. 2007) were bound by Oct4 only in the ChIP-chip experiments. A separate set of differentially expressed Oct4 targets, including *Yap1* and *Foxd3*, which have been shown to have developmentally important roles (Hanna, Foreman et al. 2002; Camargo, Gokhale et al. 2007), were obtained only in the ChIP-PET data. Similarly, observations were made in the *Nanog* RNAi knockdown data, which showed changes in expression of target genes identified both exclusively and collectively by the two technologies. Therefore, combining the binding data obtained by both techniques, along with gene expression data has provided a more detailed overview of the factors involved in the ES cell transcriptional circuitry. Further genetic studies of these regions will lend deeper insight into the mechanisms governing ES cell biology.

## DISCUSSION

ChIP-based technologies are being used extensively in identifying protein-DNA interaction networks in a variety of cell types and a number of varying conditions. In particular, ChIP-PET and ChIP-chip have been used to identify the mouse and human ES cell transcriptional circuitries, which are largely regulated by the key pluripotency factors, Oct4 and Nanog. Although each ChIP-based technology

used in the identification of these networks has its distinct advantages, we find substantial differences in the data derived through these different experimental methods. Recent technological comparisons have shown differences in the results obtained by these methods, and illustrated the need to use these data in a complementary manner (Euskirchen, Rozowsky et al. 2007). We have used ChIP-chip to uncover genomic regions bound by Oct4 and Nanog in mouse ES cells, and expanded on previously published ChIP-PET results, and find a large number of binding sites identified exclusively by each technique. Therefore, using these data in a complementary fashion provides a more detailed overview of the Oct4 and Nanog transcriptional networks.

We analyzed our ChIP-chip results for Oct4 and Nanog in relation to existing ChIP-PET data. Since the criteria for identification of genomic targets is different between platforms, the data sets obtained by the two methods was examined against each other under an exhaustive range of significance values. Recovery curves were used to measure the recovery of targets obtained by keeping the binding threshold for one technique constant and varying the threshold values for the other method. As expected, for both Oct4 and Nanog targets, the ChIP-PET recovery decreased as the ChIP-chip p-value threshold was made more stringent. A similar trend was observed for the ChIP-chip recovery when the ChIP-PET read stringency was increased. Additionally, at the same thresholds, this overlap decreased when the recovery distance permitted between a ChIP-chip peak and ChIP-PET peak was narrowed. Therefore, these recovery curves

revealed the necessity of recovery distance calibration in examining binding experiments from multiple sources. Interestingly, we also observed that the amount of recovery between ChIP-chip and ChIP-PET data increased when the whole chromosome arrays were used. Therefore, the criteria used to determine a binding event, as well as the extent of genome coverage had an effect on the overlap between the data obtained by the two methods. The recovery curves illuminated the sensitivity of recovery to distance threshold, and provided a useful means to examine the data sets relative to each other.

We combined the protein-DNA binding data with known *Oct4* and *Nanog* RNAi expression profiling data in order to analyze the targets that are differentially regulated upon *Oct4* or *Nanog* knockdown in ES cells. Oct4 and Nanog-bound regions uncovered by both technologies, as well as the ones obtained exclusively by each method contained a number of differentially regulated genes. Many of these genes comprised of transcription factors and regulators of gene expression, which are important in development. For instance, the expanded Oct4 and Nanog regulatory network contained genes such as, *Hoxa1*, *Foxd3*, *Msx2* and *Hexb*, which showed changes in expression upon *Oct4* or *Nanog* knockdown. These genes have been shown to be important in cell fate-specification, and are involved in developmentally important signaling pathways. Such additional targets identified by each technique can be used to expand the ES cell transcriptional regulatory framework, and thereby provide more detailed groundwork to understand pluripotency mechanisms. Further genetic

manipulations of each of these genes in ES cells would be necessary to independently validate their contributions to pluripotency.

Although both ChIP-chip and ChIP-PET technologies have been useful in studying protein-DNA interactions on a genome-wide scale, each method has its set of limitations. In ChIP-chip, our observations are restricted to regions tiled on the array platform, and the resolution is limited by the size of the probes, their spatial distribution, as well as the average fragment length of sonicated DNA hybridized to the arrays. In ChIP-PET experiments, the bacterial cloning and sequencing steps, as well as mapping issues introduce scope for error. We feel that a combination of more stringent mapping criteria and the inherent noise in the sequencing procedure may be responsible for the number of sequence reads which did not match perfectly to the genome. Moreover, as indicated by our sequence-depth analysis, the number of sequences obtained from ChIP-PET experiments can be a limiting factor, since more binding targets can be recovered through greater depth in sequencing. Additionally, as in the case of ChIP-chip experiments, the resolution of binding is limited by the average DNA fragment size used in the ChIP experiment. We observed some of these limitations in this study since there were a significant number of Oct4 and Nanog targets that had been identified by ChIP-PET, and did not have corresponding probes tiled on the arrays used in the ChIP-chip experiments. Apart from these limitations, it is also important to consider that binding sites may be differentially occupied at different times in the cell cycle since the chromatin state changes at

different times (Meshorer and Misteli 2006). However, since it is currently not feasible to culture ES cells in a synchronized manner, such genome-wide analyses should be done with this caveat in mind. In addition to this, another limitation to these studies is that the processing of ES cell samples can vary between different laboratories and also between different batches of serum used to culture these cells. Finally, different binding results may be obtained due to differences in ES cell strains. Therefore with the availability of binding information from different cell strains (Kim, Chu et al. 2008), can begin to address such issues.

Apart from ChIP-Chip and ChIP-PET, other ChIP based methodologies, such as ChIP-SACO (Impey, McCorkle et al. 2004) and STAGE (Bhinge, Kim et al. 2007) have been used to determine protein-DNA interactions on a genome-wide scale. Most recently, ChIP-Seq (Johnson, Mortazavi et al. 2007), a sequencing based technology has aimed to address many of the issues such as, genome coverage, sequencing-depth and binding resolution, which are encountered by other currently used techniques. With this rapid change in technologies, it will be important to investigate the results obtained from them techniques and incorporate them into our current understanding of regulatory networks. Importantly, the use of multiple techniques has been shown to produce variations in the information obtained through individual platforms (Euskirchen, Rozowsky et al. 2007). Using the data obtained through these different methodologies in a

complementary fashion provides a more thorough foundation for further investigating these networks.

The results of this study provide a useful way to integrate protein-DNA interaction data that are obtained by different techniques. We have used this to expand our current knowledge of the mouse ES cell regulatory network that is orchestrated by the transcription factors, Oct4 and Nanog. Although both the ChIP-chip and ChIP-PET technology platforms identified different sets of binding events, a considerable number of these events represented genes that were regulated by Oct4 and Nanog. Since a number of these genes have known roles in important developmental pathways and in cell-fate specification, it will be interesting to explore their biological roles with respect to ES cell pluripotency. Therefore, this expanded network provides a stronger foundation to further examine biochemical and genetic interactions that regulate stem cell properties. Moreover, the methods described to compare datasets from different platforms would be very useful as data from newer technologies, such as Chip-Seq becomes available. Since ES cells are a model system for studying developmental processes, and are thought to hold great promise in regenerative medicine, it will be important to gain a thorough understanding of the means by which a stem cell maintains its identity, and how it can be directed to form different cell types. Our work will allow for a more detailed examination of the components of this expanded stem cell circuitry and will lend better insight into the mechanisms of pluripotency.

## METHODS

### ES Cell Culture

V6.5 murine ES cells (genotype 129SvJae x C57BL/6; male) were grown at 5% $CO_2$ at 37°C on gelatinized tissue-culture plates. They were grown in DMEM (Gibco, 11965-118) with 15% Fetal Bovine Serum (FBS) (Hyclone, Lot No. ARC26080), Leukemia Inhibitory Factor, 1% penicillin/streptomycin (100X stock from Gibco, 15140-122), 1% L-glutamine (200 mM from Gibco, 25030-081), 1% non-essential amino acids (100X stock from Gibco, 11140-050) (Rideout, Wakayama et al. 2000). Since the replication time of ES cells can vary with different batches of serum, the doubling time of ES cells grown with this batch of serum was calculated to be approximately 16 hours. This doubling time was comparable to that obtained with other lots of FBS (Hyclone, Lot numbers ASJ30355 and ASB28896). Moreover, as an additional control, KH2 ES cells (Beard, Hochedlinger et al. 2006) were also cultured in these different batches of FBS, and showed similar doubling times as v6.5 ES cells The cells were grown without irradiated mouse embryonic fibroblasts, prior to the ChIP analyses in order to minimize contamination from feeders.

### Antibodies

For ChIP experiments, we used anti-Nanog (Bethyl, BL-1162) and anti-Oct4 (SantaCruz, sc-8628X), which have been previously characterized for immunoprecipitation.

**Chromatin Immunoprecipitation**

The ChIP protocol was similar to previously published studies(Boyer, Lee et al. 2005; Boyer, Plath et al. 2006; Lee, Jenner et al. 2006; Lee, Johnstone et al. 2006). Briefly, for each location analysis reaction, approximately 1.5x $10^8$ ES cells were grown at 70-80% confluency. The cells were cross-linked by adding fresh 11% formaldehyde solution to the ES media for 10 minutes at room temperature. The cells were washed twice with 1x PBS and scraped off the plates, pelleted and stored at -80°C. They were then lysed and sonicated to solubilize chromatin and shear the cross-linked DNA. Sonications were performed with a Misonix Sonicator 3000 and sonicated at power 7 for 12 x 30 second pulses (60 second pause between pulses) at 4°C while the samples were immersed in a water bath. The whole cell extract was incubated overnight on a rotating platform at 4°C with 100 μl of Dyna1 Protein G magnetic beads, blocked with 0.05% BSA/PBS and preincubated for 6 hours to overnight with 10 μg of antibody of choice. The beads were washed 5 times with RIPA buffer and once with TE containing 50 mM NaCl. Bound protein-DNA complexes were eluted off the beads in elution buffer by occasional vortexing and heating at 65°C overnight. Whole cell extract (saved from the sonication step) was treated similarly for cross-link reversal. Following treatment with RNaseA and Proteinase K, the immunoprecipitated and whole cell extract DNA was purified by phenol:chloroform:isoamyl alcohol extraction. The DNA was blunted-end ligated to a universal linker and amplified using a two-stage expansion PCR protocol (3 reactions per sample were done for the second expansion and combined). The

amplified immunoprecipitated DNA and whole cell extract DNA were labeled with Cy5 and Cy3 fluorophores respectively, using Invitrogen random primer labeling kits. 1 ug of DNA was used in a labeling reaction and 3 labeling reactions were done per sample and combined after purification. The labeled DNA was purified using Invitrogen BioPrime Array CGH module purification kit. 5 μg each of immunoprecipitated and whole cell extract DNA was combined along with mouse Cot-1 DNA and hybridized to each of the arrays in Agilent hybridization chambers for 40 hours at 65°C using the Agilent hybridization protocol and reagents for 244K arrays. Arrays were then washed and scanned as previously described. Three biological replicates were done for each transcription factor in order to determine statistical significance for binding targets.

**Analysis of ChIP-Chip Data**

Three biological replicates were examined for both Oct4 and Nanog using mouse promoter arrays. The same samples were also hybridized to mouse chromosome arrays. A probe was marked 'bound' in a particular replicate if its Rosetta p-value was below a pre-determined cutoff (usually 0.001, although this was systematically varied for the Recovery curves). For each bound probe, a region around the probe was marked 'bound' extending to the nearest unbound probe (or 1kb, if the nearest probe was more than 1kb away). A region was marked as a ‘peak’ if it was bound in all three replicates.

The list of micro-RNAs to be analyzed was taken from version 8.0 of mirBASE, which were mapped against NCBI build 34 of the mouse genome (mm6). We retained for analysis only those micro-RNAs which had at least three probes from the promoter array design mapped to within 4kb of the mirBASE annotation. Of the original 267 mirBASE annotations, 258 satisfied this requirement.

**Mapping of ChIP-PET Sequences To The Mouse Genome**

Two sets of previously published ChIP-PET experiments for Oct4 and Nanog were examined. These reads were mapped to the Mouse mm6 genome sequence in order to be consistent with the genome version used for the arrays in the ChIP-Chip experiments. The mapping was done using the BLAT sequence alignment tool, with a step size of 5 and tile size of 11.

We received two FASTA files of sequences, each between 34-36 bp in length. Only those sequence hits reported by BLAT that had at least 34 matched bases (corresponding to 0, 1, or 2 mismatches), and a gap-length of at least 10bp, were retained. Any sequences with more than one retained hit to the mm6 genome were filtered out. The locations of the unique hits for the sequences that remained were stored in a database, and used for comparison with the ChIP-chip peaks.

## Comparison of ChIP-Chip to Chip-PET Data

In order to examine if a bound region from ChIP-Chip was ‘matched’ to another bound region from ChIP-PET, and vice versa, a simple genomic distance threshold was determined. If the two regions were on the same chromosome, and if the edge of one region was within the recovery distance from the edge of the other, then they were considered ‘matched’. The typical recovery distance used was 1kb, although this was varied systematically for the recovery curves. A recovery distance of '0' represents only strictly overlapping regions that were matched to each other.

## Recovery Curves

Recovery curves were produced for promoter and chromosome array data for Oct4 and Nanog. One curve was for the ChIP-Chip (p-value) cutoff, and the other for the ChIP-PET (overlapping region) cutoff. In the ChIP-PET Recovery curves, fixed cutoffs of 4 overlapping reads for Oct4 and 3 overlapping reads for Nanog were used to determine a 'background set' of ChIP-PET bound regions, as described above. At that ChIP-PET cut-off, the number of ChIP-Chip bound regions was calculated for a range of p-value cut-offs. The percentage of background ChIP-PET bound regions that were matched by at least one ChIP-Chip bound region, at each p-value threshold was graphed. This was repeated for several different matching distances.

For calculating the ChIP-Chip Recovery curve, this process was repeated by holding the ChIP-Chip binding threshold constant. A background set of bound regions using ChIP-Chip was calculated using a p-value threshold of 0.001. The threshold of overlapping reads used to call bound regions in ChIP-PET was varied from 1 to an upper limit where no matched regions were called. At each threshold, the fraction of the background ChIP-Chip bound regions matched by one of the ChIP-Pet regions was calculated. This process was also repeated for several different matching distances.

**Sequence Depth Analysis**

In order to carry out the sequence-depth analysis using data from chromosome 19, the total number of ChIP-PET reads for Oct4 and Nanog were calculated. For Oct4, 8675 reads, and for Nanog, 5233 reads were mapped to Chromosome 19. The number of ChIP-chip bound regions recovered for both proteins, was determined for differently sized subsets of these ChIP-PET reads. We started with a random subset that sampled 10% of the reads, and determined the ChIP-ChIP recovery for that sample. This analysis was done for increasingly large subsets of ChIP-PET reads in increments of 10%, until all ChIP-PET reads on chromosome 19 were sampled. For each subset size, the same number of reads was randomly sampled 10 times to calculate the average ChIP-chip recovery and standard deviation in each case.

## Expression Analysis of Oct4 and Nanog Targets

We wanted to determine the relevance to pluripotency for Oct4 and Nanog binding targets obtained by ChIP-Chip or ChIP-PET. Previously published *Oct* and *Nanog* RNAi gene expression profiles in ES cells were used for this analysis(Loh, Wu et al. 2006). These experiments produced sets of Affymetrix probes, which were differentially-regulated in one or more of the replicates. Those probes were matched to sets of gene names using an Affymetrix-provided probe annotation file. A gene was determined as differentially-regulated in either the *Oct4* or the *Nanog* knockdown experiments if it was associated with any differentially-regulated probe in the replicates of that factor's experiments. A gene annotation was called 'bound by ChIP-chip' if there was a ChIP-chip peak within +/-4kb of the gene annotation's start site. Equivalently, a gene annotation was also bound by ChIP-PET if a ChIP-PET peak fell within the same distance of the annotated start site. This allowed us to divide the set of differentially regulated gene annotations into four categories: bound in both experimental types, bound in neither, and bound in one (but not the other).

## Gene Ontology (GO) Analysis

A set of bound RefSeq gene identifiers was compiled for the Oct4 and Nanog bound genes according to a standard cutoffs (p-value< 0.001 for ChIP-chip, and a maximum distance of 4kb from the gene start site for any annotated gene). The sets of RefSeq identifiers were converted into sets of Known Gene gene symbols, using the kgXref table of the UCSC Genome database. A set of GO

categories was downloaded from the Gene Ontology Project (http://www.geneontology.org/GO.downloads.ontology.shtml), and an updated set of mouse annotations was downloaded from EBI (http://www.ebi.ac.uk/GOA/). Since the GO categories are structured as a directed graph, we propagated the annotations backwards through the graph; a gene symbol was marked as annotated to a GO category if the annotation was contained in the EBI dataset, or if the GO category was the ancestor of a GO category to which the gene symbol had already been assigned. A background set of gene symbols was assembled, from the corresponding Known Gene symbols for any RefSeq gene identifier that was tiled on the ChIP-chip array design. Given foreground sets of genes defined by the ChIP binding experiments, a background set of arrayed gene symbols, and a set of EBI-derived GO category annotations for each symbol, we calculated a list of GO categories ranked by their enrichment in each of the four foreground gene sets. Enrichment was calculated using the p-value of the hypergeometric probability for the overlap of each test set with each GO category.

**ACKNOWLEDGEMENTS**

We would like to thank Yijun Ruan (Genome Institute of Singapore, Singapore) for providing the Oct4 and Nanog ChIP-PET raw sequence data. We thank BingBing Yuan and George Bell of the Bioinformatics and Research Computing (BARC) core at the Whitehead Institute for assistance with primer design for gene-specific PCR, Sumeet Gupta at the Whitehead Center for Microarray

Technology (CMT) for help with creating the array design files and Yuchun Guo for assistance in setting BLAT parameters. This work was supported by NIH grants RO1-HD045022 and R37-CA084198 to RJ, and HG002688 to RAY and DKG.

## SUPPLEMENTARY INFORMATION

### Supplementary Note

We have summarized the ChIP-chip promoter (supplementary file A) and chromosome array data (supplementary file B) and ChIP-PET raw data for Oct4 and Nanog. The ChIP-chip data for each region contains the location and value of each probe in that region (the first number is the locations, and the second number, marked with a '+', is the offset of that probe within that region). Each probe is followed with its binding ratio in each of the three replicates. A bound probe is marked with a '*'. A second list represents a set of coordinates for each distinct ChIP-PET read that falls within that window. The numbers marked with a '+' represent the offset within that region.

### Oligo Array Design

Two kinds of array platforms were employed in this study. One was a 2-slide mouse promoter array that is based on the 10-slide mouse promoter array set described in previous studies (Boyer, Lee et al. 2005; Boyer, Plath et al. 2006). The arrays were manufactured at Agilent Technologies (www.agilent.com). The arrays include 19.993 features that include the promoters of all annotated genes

in the NCBI Refseq database and miRNAs in miRBase. They also include promoters of alternate transcription start sites (TSSs). The oligos are 60-mers and span from 4kb upstream to 4kb downstream of the TSS at a density of one probe every 250 bp. Additionally the entire mouse *HoxA* cluster is also tiled on these arrays.

The second type of array was a whole chromosome array for the mouse chromosome 19. It also tiles the non-repeat portion of the ENCODE Design. Since ENCODE is a human project, analogous regions in mouse were mapped for this array. The oligo length and probe density is similar to the promoter arrays.

**Gene Specific PCR For Oct4 and Nanog Bound Regions**

Gene Specific PCR was performed for three different sets of targets for both Oct4 and Nanog: targets identified solely by ChIP-Chip, ones identified only by ChIP-PET and another set uncovered using both methods. PCR was performed on the Ligation-mediated PCR products for two independent biological replicates for each protein. 10 ng of immuno-enriched DNA and 10, 30 and 90 ng of whole-cell extract DNA were used per reaction. The PCR was run for 23 cycles, and products were quantified on an agarose gel stained with SYBR Gold (Invitrogen). Primers for each target amplified a 200-300 bp region around the genomic locations of the probes showing enrichment. The primer sequences and PCR product coordinates can be found in Table S9. Enrichment was calculated as a

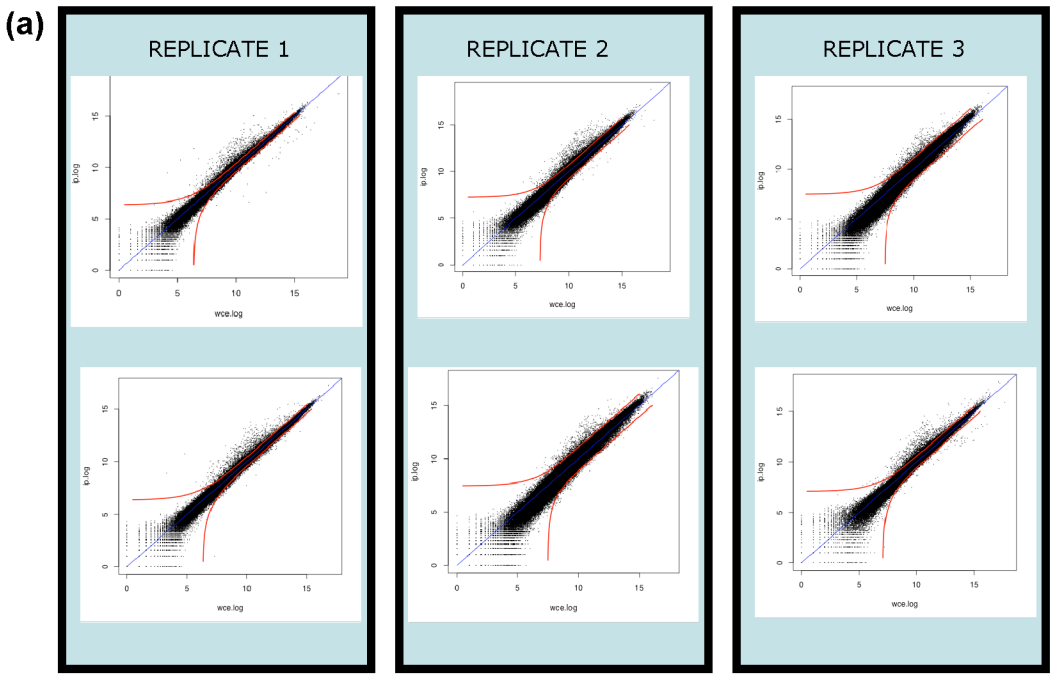
ratio of intensity of PCR product from 10 ng of IP DNA to the product intensity from 90ng or 30 ng of WCE DNA. For a product to be considered enriched, the IP DNA product intensity had to be atleast that of 90 ng of WCE DNA or 1.5 times that of 30 ng of WCE DNA. Enrichment ratios were normalized against the ratio for un-enriched *β-Actin* DNA. Among Nanog bound targets found exclusively by ChIP-chip, 26 of 33 regions were confirmed by gene-specific PCR experiments. Similarly, 28 out of 31 Oct4-bound regions, identified solely by ChIP-chip were unenriched (Table S9).

**Gene Expression Analysis Note**

A caveat in this analysis was that the gene-expression comparison to binding data was limited to –4 to +4 kb surrounding the transcriptional start site for both ChIP-Chip and ChIP-PET data. Such a restriction was necessary for this comparison since probes on the promoter arrays used in ChIP-chip experiments are limited to these regions. Therefore, a binding event outside of this 8kb region in the ChIP-PET data would not be captured in our analysis. An example of this is demonstrated in Supplementary Figure S3, where the *Rest* gene, which is differentially expressed on *Oct4* knockdown, is bound by Oct4 in both experiments. However, in our analysis of the expression data, it is observed as an Oct4 target only by ChIP-chip and not by ChIP-PET since the binding event in the latter case is outside the promoter array tiled region. Visualization of the two experimental types on the 'GSE Visualizer' reveals the utility of careful calibration of the thresholds at which binding events are called significant.

**Figure S1. Oct4 ChIP-Chip Enrichment plots**. IP vs. WCE enrichment in three biological replicate samples is shown for (a) the 2-slide set promoter arrays and (b) chromosome 19 arrays.

**FIGURE S1.**


(a)
REPLICATE 1
REPLICATE 2
REPLICATE 3
ip.log
wce.log

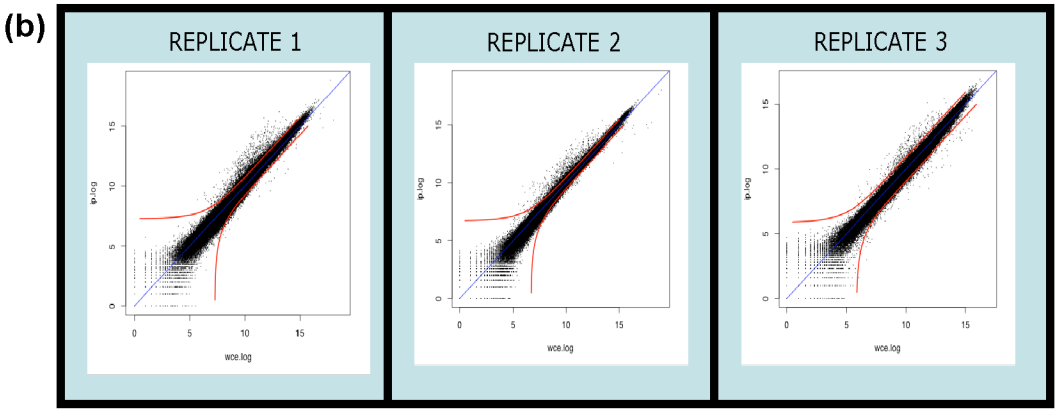

(b)
REPLICATE 1
REPLICATE 2
REPLICATE 3
ip.log
wce.log

**Figure S2. Nanog ChIP-Chip Enrichment plots**. IP vs. WCE enrichment in three biological replicate samples is shown for (a) the 2-slide set promoter arrays and (b) chromosome 19 arrays.

**FIGURE S2.**

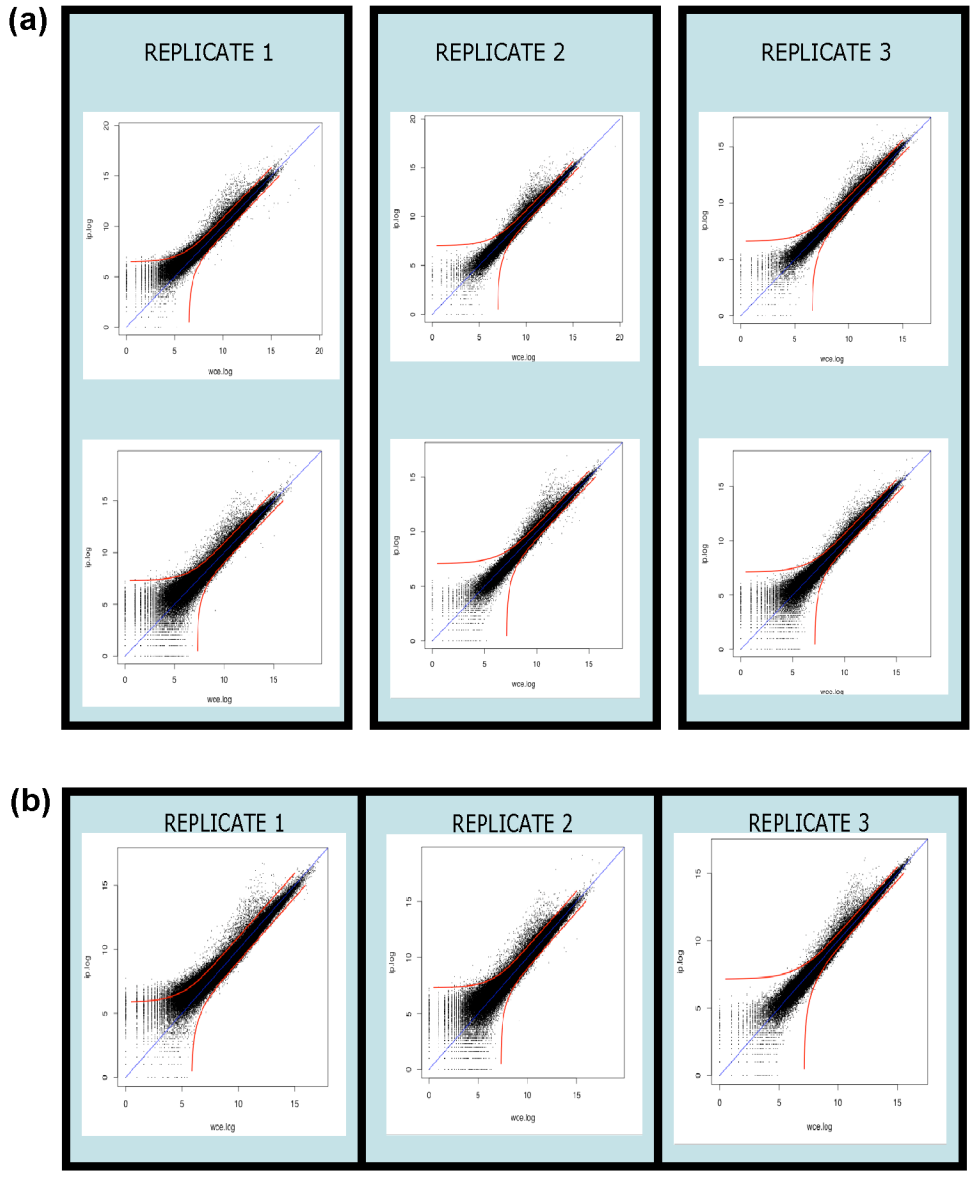
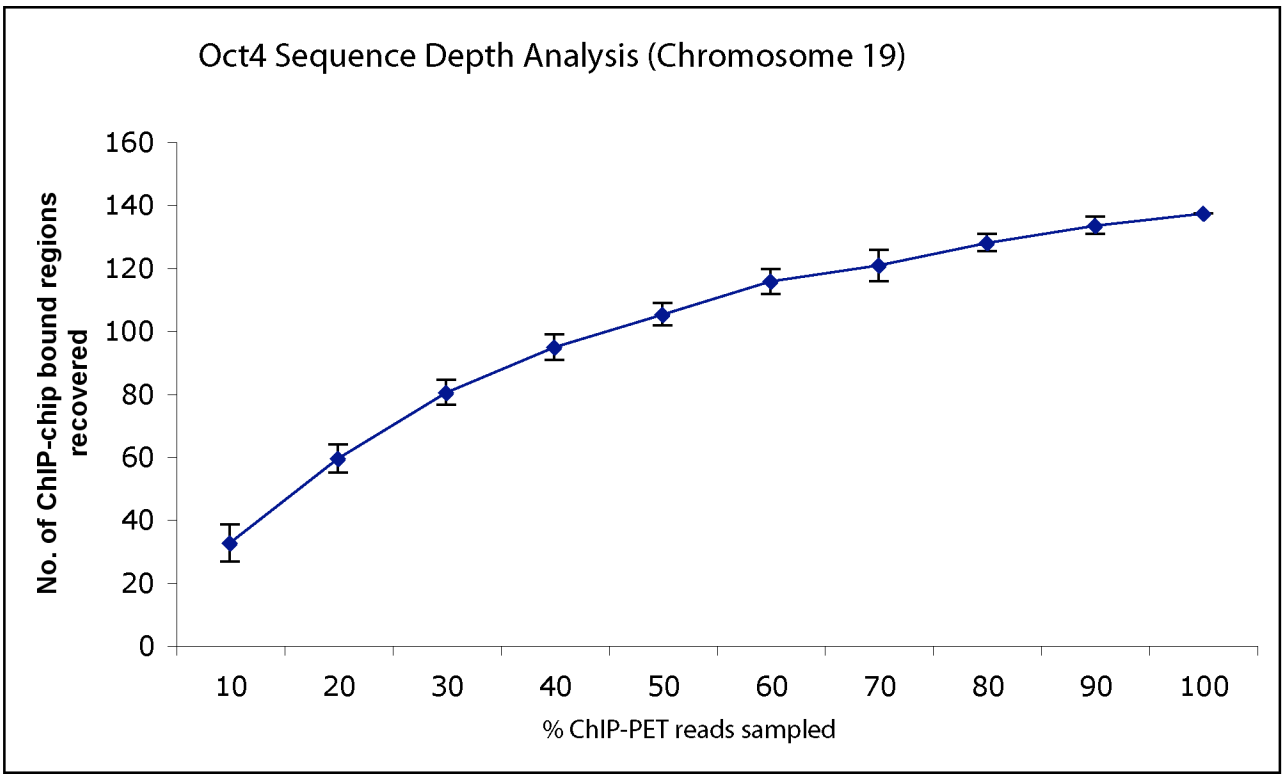

(a)
REPLICATE 1
REPLICATE 2
REPLICATE 3
(b)
REPLICATE 1
REPLICATE 2
REPLICATE 3

**Figure S3.** Sequence Depth Analysis for OCT4 and NANOG Targets on Chromosome 19. Plots indicate the number of ChIP-chip targets recovered (y-axis) when different percentages of ChIP-PET sequences are randomly sampled. Error bars represent the standard deviation.

**FIGURE S3.**


Oct4 Sequence Depth Analysis (Chromosome 19)
No. of ChIP-chip bound regions recovered
160
140
120
100
80
60
40
20
0
10
20
30
40
50
60
70
80
90
100
% ChIP-PET reads sampled

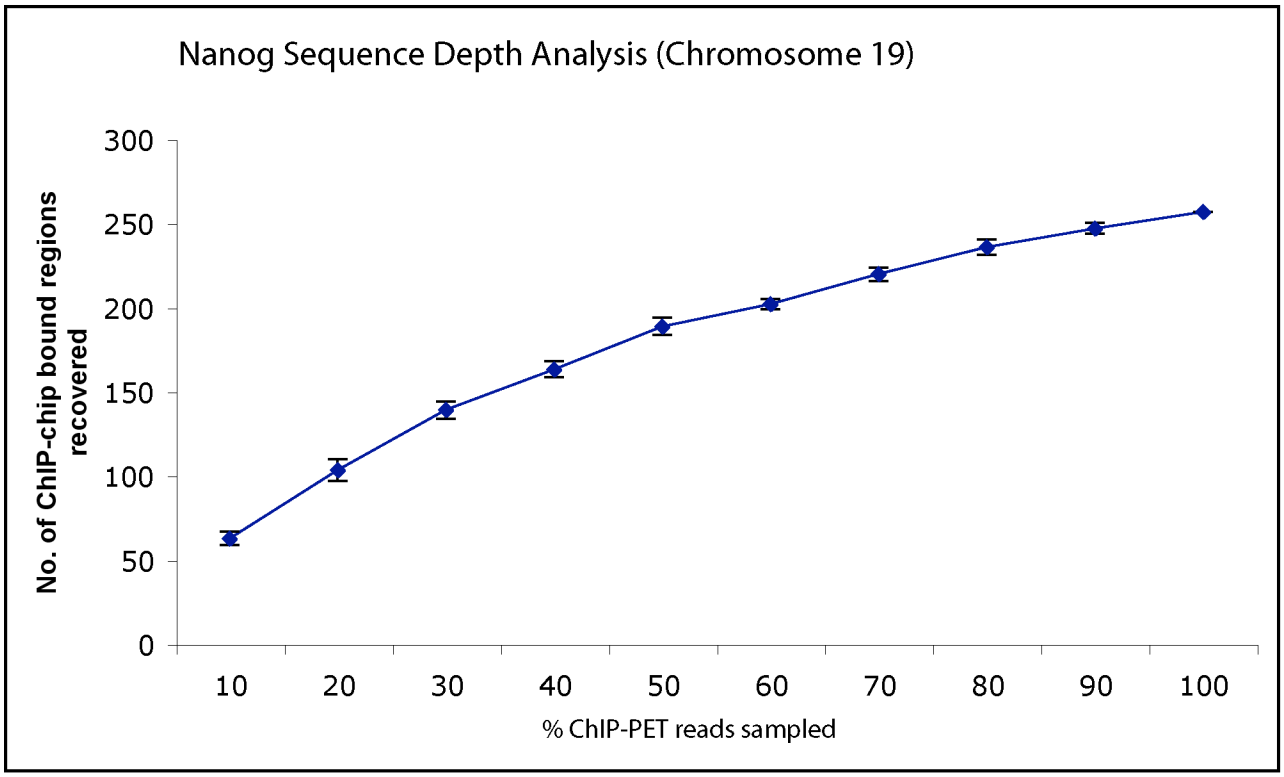
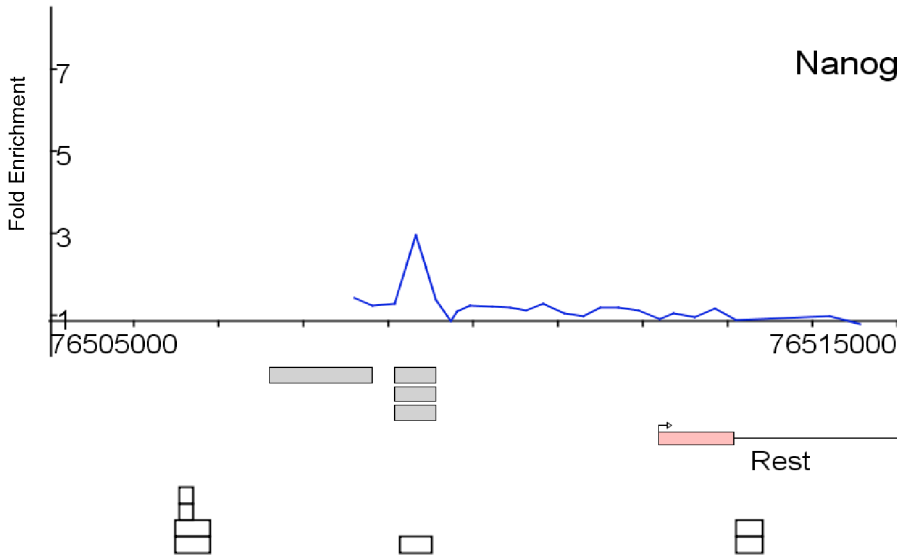

Nanog Sequence Depth Analysis (Chromosome 19)
No. of ChIP-chip bound regions recovered
300
250
200
150
100
50
0
10
20
30
40
50
60
70
80
90
100
% ChIP-PET reads sampled

**Figure S4. Inconsistency in combining expression-profiling data with binding information from ChIP-Chip and ChIP-PET.** The binding of Nanog at the REST gene promoter is identified by both ChIP-Chip and ChIP-PET. However, the binding event detected in the ChIP-PET experiment is not within the region used for combining the expression profiling information (+/- 4kb around the transcription start site). Consequently, the changed expression of REST after Nanog knockdown is associated with Nanog binding detected by ChIP-Chip only and not ChIP-PET.

**FIGURE S4.**



**Additional Data Files on CD**

**Table S1.** Protein Coding Gene List (Excel format)

**Table S2.** miRNA Targets

**Table S3.** Gene Ontology Analysis

**Table S4.** Chromosome 19 Bound Regions for Oct4 and Nanog

**Table S5.** Oct4 Recovery Curve Tables

**Table S6.** Nanog Recovery Curve Tables

**Table S7.** Oct4 RNAi Differentially Regulated Targets

**Table S8.** Nanog RNAi Differentially Regulated Targets

**Table S9.** Gene-Specific PCR Regions for Oct4 and Nanog Genomic Targets

**Table S10.** Sequence Depth Analysis of Oct4 and Nanog ChIP-PET data

**Supplementary File A.** Promoter Array and ChIP-PET Data for Oct4 and Nanog

**Supplementary File B.** Chromosome Array and ChIP-PET Data for Oct4 and Nanog

## REFERENCES


Andersson, O., P. Bertolino, et al. (2007). "Distinct and cooperative roles of mammalian Vg1 homologs GDF1 and GDF3 during early embryonic development." Dev Biol **311**(2): 500-11.

Beard, C., K. Hochedlinger, et al. (2006). "Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells." Genesis **44**(1): 23-8.

Bhinge, A. A., J. Kim, et al. (2007). "Mapping the chromosomal targets of STAT1 by Sequence Tag Analysis of Genomic Enrichment (STAGE)." Genome Res **17**(6): 910-6.

Boyer, L. A., T. I. Lee, et al. (2005). "Core transcriptional regulatory circuitry in human embryonic stem cells." Cell **122**(6): 947-56.

Boyer, L. A., D. Mathur, et al. (2006). "Molecular control of pluripotency." Curr Opin Genet Dev **16**(5): 455-62.

Boyer, L. A., K. Plath, et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-53.

Camargo, F. D., S. Gokhale, et al. (2007). "YAP1 Increases Organ Size and Expands Undifferentiated Progenitor Cells." Curr Biol.

Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-55.

Chew, J. L., Y. H. Loh, et al. (2005). "Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells." Mol Cell Biol **25**(14): 6031-46.

Danford TW, R. P., and Gifford DK (2007). GSE: A Comprehensive Database System for

the Representation, Retrieval, and Analysis of Microarray Data. Pacific Symposium of Biocomputing.

Euskirchen, G. M., J. S. Rozowsky, et al. (2007). "Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies." Genome Res **17**(6): 898-909.

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

Hanna, L. A., R. K. Foreman, et al. (2002). "Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo." Genes Dev **16**(20): 2650-61.

Impey, S., S. R. McCorkle, et al. (2004). "Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions." Cell **119**(7): 1041-54.

Ivanova, N., R. Dobrin, et al. (2006). "Dissecting self-renewal in stem cells with RNA interference." Nature **442**(7102): 533-8.

Johnson, D. S., A. Mortazavi, et al. (2007). "Genome-wide mapping of in vivo protein-DNA interactions." Science **316**(5830): 1497-502.

Kim, J., J. Chu, et al. (2008). "An extended transcriptional network for pluripotency of embryonic stem cells." Cell **132**(6): 1049-61.

Lee, T. I., R. G. Jenner, et al. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-13.

Lee, T. I., S. E. Johnstone, et al. (2006). "Chromatin immunoprecipitation and microarray-based analysis of protein location." Nat Protoc **1**(2): 729-48.

Loh, Y. H., Q. Wu, et al. (2006). "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells." Nat Genet **38**(4): 431-40.

Martin, G. R. (1981). "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." Proc Natl Acad Sci U S A **78**(12): 7634-8.

Meshorer, E. and T. Misteli (2006). "Chromatin in pluripotent embryonic stem cells and differentiation." Nat Rev Mol Cell Biol **7**(7): 540-6.

Mitsui, K., Y. Tokuzawa, et al. (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-42.

Nichols, J., B. Zevnik, et al. (1998). "Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4." Cell **95**(3): 379-91.

Niwa, H. (2007). "How is pluripotency determined and maintained?" Development **134**(4): 635-46.

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

Rideout, W. M., 3rd, T. Wakayama, et al. (2000). "Generation of mice from wild-type and targeted ES cells by nuclear cloning." Nat Genet **24**(2): 109-10.

Rodda, D. J., J. L. Chew, et al. (2005). "Transcriptional regulation of nanog by OCT4 and SOX2." J Biol Chem **280**(26): 24731-7.

Thomson, J. A., J. Itskovitz-Eldor, et al. (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-7.

Zhu, X., J. Zhang, et al. (2006). "Sustained Notch signaling in progenitors is required for sequential emergence of distinct cell lineages during organogenesis." Genes Dev **20**(19): 2739-53.

# Chapter 3

## EFFECTS OF SELECTION MARKER CHOICE AND DRUG SELECTION TIMING ON REPROGRAMMING FIBROBLASTS TO A PLURIPOTENT STATE

Divya Mathur[1,2], Ruth K. Foreman[1,2], Tobias Brambrink[2] and Rudolf Jaenisch[1,2]

[1]Department of Biology, Massachusetts Institute of Technology, 32 Ames Street, Cambridge, MA 02139, USA

[2]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

### RESPECTIVE CONTRIBUTIONS

DM performed the experiments described in this section. RF and DM did the viral infections for reprogramming. TB made the Fbx15 targeting construct.

**ABSTRACT**

The reprogramming of somatic cells into a pluripotent state through virus-mediated transduction of four transcription factors has been a major feat for the scientific and medical communities. However, before these induced pluripotent (iPS) cells are employed in medical treatments, issues such as drug selection, use of viruses and introduction of potentially oncogenic transcription factors need to be resolved. Therefore, in order to design screening procedures for finding safer alternatives, iPS cells obtained from different screens must be better characterized. One issue with earlier experiments that used activation of Fbx15 as a selection criterion for reprogramming was that the Fbx15-iPS cells were not truly pluripotent and did not give rise to chimeras. In contrast, selection for Oct4 or Nanog activation, albeit with later drug initiation, led to iPS cells that could generate germline chimeras. Therefore, it was unclear whether the partially-reprogrammed Fbx15-iPS cells were obtained due to early drug selection, or because selection for Fbx15 activation led them down this path. We examined Fbx15-iPS cells, obtained from early and late drug selection, by examining the methylation status of the Oct4 locus. Since this locus is methylated in the former case and completely unmethylated in the latter, our results support the notion that early drug selection leads to the partially-reprogrammed intermediates. Therefore, Fbx15 activation may be employed as a useful marker in reprogramming screens in the future.

## INTRODUCTION

Embryonic Stem (ES) cells possess the valuable property of pluripotency, and therefore hold great potential in medical applications. However, since the derivation of these cells involves the destruction of an embryo, research involving these cells had been plagued with ethical controversies. Recently, after many attempts at circumventing the use of embryos to obtain ES cells, scientists were successful in reprogramming the genomic state of differentiated skin fibroblasts back into an ES cell-like state (Takahashi and Yamanaka 2006; Maherali, Sridharan et al. 2007; Okita, Ichisaka et al. 2007; Wernig, Meissner et al. 2007). These reprogrammed cells are known as induced pluripotent stem (iPS) cells.

Reprogrammed iPS cells were generated by introducing four transcription factors into fibroblasts through virus-mediated transduction. These pluripotency factors were Oct4, Sox2, c-Myc and Klf4. In the initial reprogramming experiments, the activation of the *fbx15* gene locus was used as a marker to select for reprogrammed ES cells (Takahashi and Yamanaka 2006). The iPS cells generated in this experiment were termed Fbx15-iPS cells, and were thought to be pluripotent since they could form teratomas when injected into mice. However, these cells were unable to generate any chimeras, indicating that they were not truly pluripotent, and may represent an intermediate stage in the reprogramming process.

Activation of the *fbx15* locus was used as a selection criterion in initial reprogramming experiments since this gene is expressed predominantly in ES cells and is a direct target of Oct4 (Tokuzawa, Kaiho et al. 2003). *Fbx15* belongs to a family of F-box containing proteins, which are components of E3 Ubiquitin ligases (Winston, Koepp et al. 1999). The promoter of this gene contains an enhancer element that has an octamer-like motif and a sox-binding motif. Oct4 and Sox2 can bind to these motifs, respectively, and activate the enhancer. In fact, Oct4 is required for the maintenance of *fbx15* gene expression. However, *fbx15* knockout mice and ES cells are normal, indicating that the gene is dispensable for pluripotency, development and fertility (Tokuzawa, Kaiho et al. 2003).

Since Fbx15-iPS cells failed to produce chimeras, later reprogramming experiments used activation of Oct4 or Nanog loci as selection criteria for reprogramming. As described in earlier sections, these genes are not only expressed exclusively in pluripotent cells, but are also key regulators of pluripotency (Niwa, Miyazaki et al. 2000; Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003). In contrast to the Fbx15-iPS cells, when Oct4 or Nanog were used as markers to select for reprogramming, completely reprogrammed iPS cells were generated, which could give rise to germline chimeras (Okita, Ichisaka et al. 2007; Wernig, Meissner et al. 2007). Moreover, the Oct4 and Nanog loci were completely unmethylated in these cells and ES cells, but were still partially methylated in Fbx15-iPS cells. The expression levels of key

pluripotency markers in Nanog-iPS cells were higher than in Fbx15-iPS cells, and more similar to those in ES cells. These studies suggested that activation of Oct4 or Nanog was a more stringent criterion that Fbx15 selection, in order to get fully reprogrammed cells.

A key difference between the initial reprogramming experiments with Fbx15 activation (Takahashi, Tanabe et al. 2007), and later ones with Oct4 or Nanog activation (Okita, Ichisaka et al. 2007; Wernig, Meissner et al. 2007) was that in the former case, fibroblasts infected with viruses carrying the *Oct4, Sox2, Klf4* and *c-Myc* transgenes were selected for Fbx15 expression at an earlier time point than in the latter case, where drug selection for Oct4 or Nanog expression was started later. Further studies on Oct4- and Nanog-iPS cells have shown that pluripotency markers, such as Alkaline Phosphatase (AP), Stage-specific embryonic antigen 1 (SSEA1), Oct4 and Nanog are sequentially reactivated during the reprogramming process (Brambrink, Foreman et al. 2008; Stadtfeld, Maherali et al. 2008). Moreover, there is a discrepancy in the timing between the appearance of drug resistance cells and fully reprogrammed ones, suggesting that early colonies may have low levels of pluripotency gene expression, which is sufficient for antibiotic resistance but not for epigenetic reprogramming. Additionally, fewer reprogrammed cells are obtained when drug selection is initiated early, but many more iPS colonies are observed with later or no drug selection (Meissner, Wernig et al. 2007; Wernig, Meissner et al. 2007). All these observations raise the possibility that the partially reprogrammed Fbx15-iPS cells

were obtained due to early drug selection and incomplete epigenetic reprogramming, and not due to the fact that Fbx15 is not relevant to pluripotency. Therefore, we tested this hypothesis by selecting for *fbx15* at early and late time points, and examining the epigenetic state of the *Oct4* locus. Indeed, we observed that in Fbx15-iPS cells that are obtained from late drug selection, the *Oct4* locus is completely unmethylated, similar to what has been reported for ES cells and Nanog iPS cells (Okita, Ichisaka et al. 2007). However, in Fbx15-iPS cells that had early drug selection, the *Oct4* locus was largely methylated, indicating that epigenetic reprogramming had not completely occurred in these cells. The experiments described here suggest that similar to Oct4 or Nanog, activation of the Fbx15 locus can also be used as a criterion for reprogramming, and the partially reprogrammed cells are observed due to early drug selection that leads to incomplete epigenetic reprogramming.

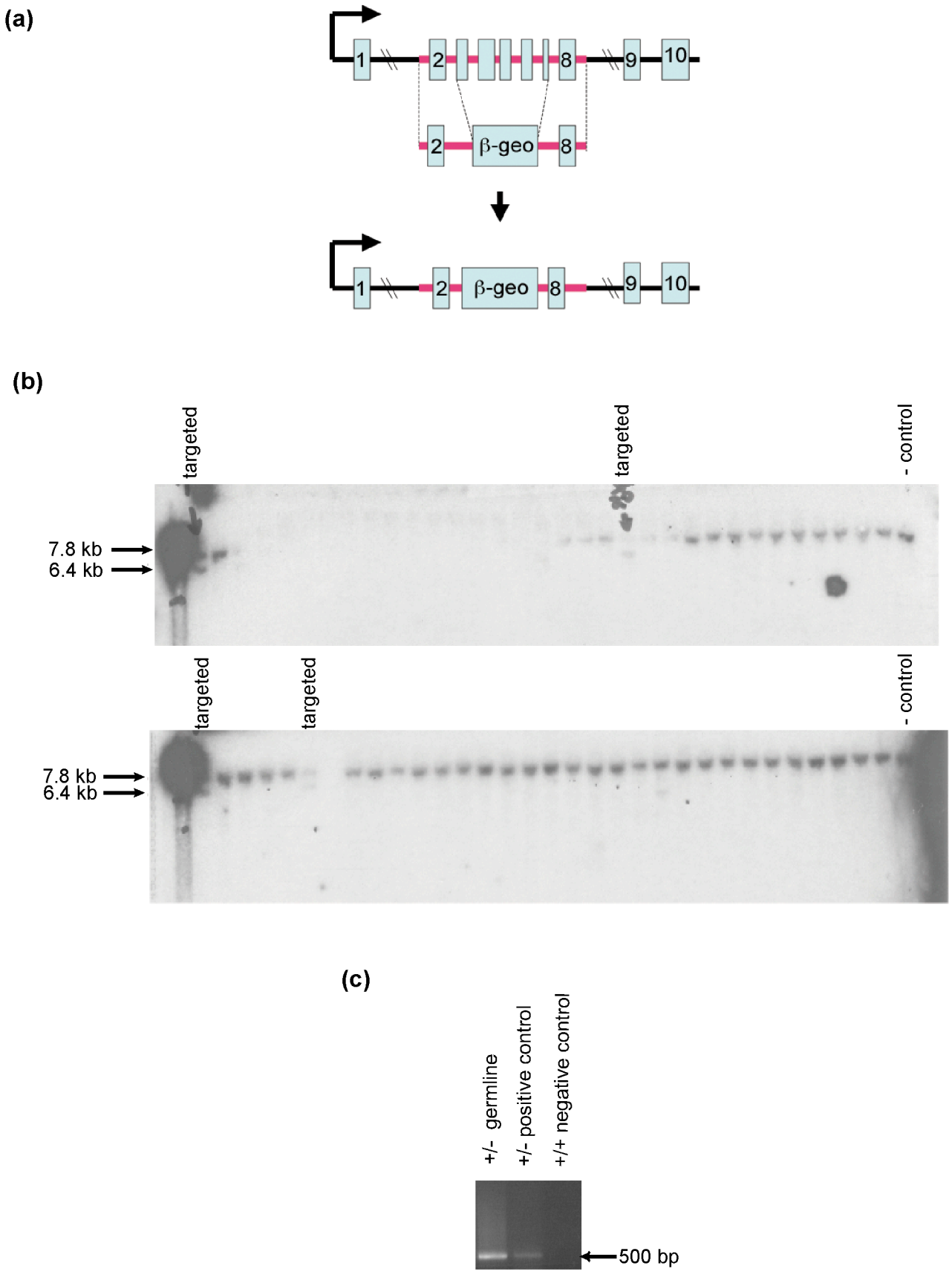
## RESULTS

### Generation of Fbx15-βgeo mice and fibroblasts

In order to develop a selection scheme for activation of the *Fbx15* locus, a β-geo cassette (a fusion of the *β-galactosidase* and neomycin resistance genes) was used to replace exons 3 to 7 of the *fbx15* gene (Figure 1a). This vector was targeted to the *Fbx15* locus by homologous recombination and positive clones were screened by Southern blot analysis (Figure 1b). Four out of forty six resistant colonies were correctly targeted and one of these clones was injected into C57/BL6 blastocysts to obtain chimeric mice containing the targeted allele.

**Figure 1. Generation of Fbx15-βgeo Mice.** (a) Targeting strategy to replace exons 3-7 of the *Fbx15* gene with a β-geo cassette. (b) Southern blot indicating ES cells clones correctly targeted with the Fbx15-β-geo construct. Untargeted ES cells were used as controls. A 7.4 kb band represents the wild type *Fbx15* allele, and a 6.8 kb band is of the targeted allele. (c) Mice carrying the Fbx15-β-geo allele in their germline were genotyped by a PCR strategy that amplified a 500 bp region of the *β-galactosidase* gene. Mice heterozygous for the targeted allele are represented by +/-. Fbx15-β-geo ES cells that were injected into blastocysts to make chimeras were used as positive controls. Untargeted ES cells served as negative controls.

**FIGURE 1.**


(a)
1
2
8
9
10
2
β-geo
8
1
2
β-geo
8
9
10
(b)
targeted
targeted
- control
7.8 kb
6.4 kb
targeted
targeted
- control
7.8 kb
6.4 kb
(c)
+/- germline
+/- positive control
+/+ negative control
500 bp

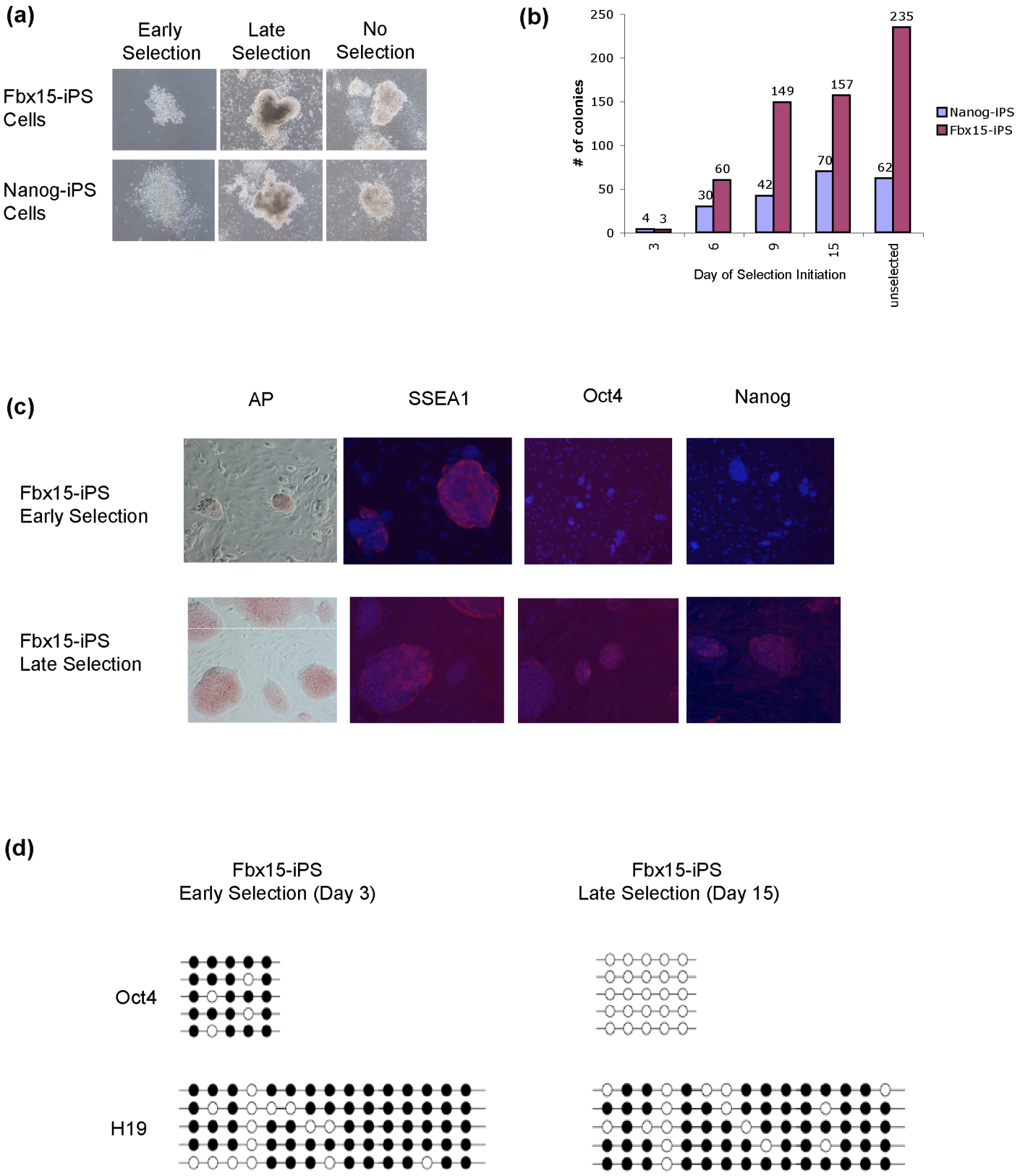
These chimeras were mated with B6D2F1 mice and offspring containing the Fbx15-βgeo allele were obtained. These mice were genotyped using a PCR strategy that screened for the presence of the *β-galactosidase* gene (Figure 1c). Since the Fbx15-βgeo construct had been targeted to ES cells containing a puromycin resistance marker, MEFs were generated from chimeric embryos made by injecting clones of the correctly targeted ES cell line into C57/BL6 blastocysts. MEFs containing the Fbx15-βgeo allele were selected with puromycin, and used for the experiments described below.

**Reprogramming of Fbx15-βgeo fibroblasts with early and late drug selection**

Fbx15-iPS cells were obtained by infecting Fbx15-βgeo fibroblasts with lentiviruses containing transgenes for the transcription factors, Oct4, Sox2, Klf4 and C-myc. As a control, Nanog-neo fibroblasts were also infected with these factors to obtain Nanog-iPS cells. Colonies were observed in both cases after initiating neomycin drug selection at days 3, 6, 9 and 15 post-infection. Additional controls without any drug selection were also done for both Fbx15-βgeo and Nanog-neo fibroblasts. Colonies were counted 21 days post-infection, based on morphological resemblance to ES cells (Figure 2a). As shown in Figure 2b, early drug selection in both cases resulted in very few neomycin-resistant colonies. In accordance with previous observations, the number of resistant colonies obtained increased when drug selection was started at later time points. It was interesting to note that at later time points of drug initiation,

**Figure 2. Analysis of Fbx15-iPS cells after Early and Late Drug Selection.** (a) Morphology of Fbx15- and Nanog-iPS cells obtained after early (day 3) and late (day 9) drug selection. (b) Colony counts of Fbx15-iPS cells and Nanog-iPS cells obtained after drug initiation at different time points. (c) Immunostaining of early (day 3) and late (day 15) drug selected Fbx-15 iPS colonies for pluripotency markers, Alkaline phosphatase (AP), SSEA1, Oct4 and Nanog. (d) Bisulfite sequencing of Oct4 and H19 DMR (control) loci in early (day3) and late (day 15) drug selected Fbx15-iPS cells.

**FIGURE 2.**


(a)
Early Selection
Late Selection
No Selection
Fbx15-iPS Cells
Nanog-iPS Cells
(b)
# of colonies
250
200
150
100
50
0
4 3
30
60
42
149
70
157
62
235
3
6
9
15
unselected
Day of Selection Initiation
Nanog-iPS
Fbx15-iPS
(c)
AP
SSEA1
Oct4
Nanog
Fbx15-iPS Early Selection
Fbx15-iPS Late Selection
(d)
Fbx15-iPS Early Selection (Day 3)
Fbx15-iPS Late Selection (Day 15)
Oct4
H19

substantially more iPS cells were obtained by selection for Fbx15 than for Nanog. This observation has also been noted before (Okita, Ichisaka et al. 2007), although the reason for this difference in numbers remains unclear.

We performed immunostaining for the presence of the pluripotency markers, AP, SSEA1, Oct4 and Nanog in Fbx15-iPS cells obtained from drug selection initiated at different time points. Figure 2c shows that whereas AP and SSEA1 staining was seen in most if not all iPS cells, staining for Oct4 and Nanog was seen at greater intensity only in iPS cells obtained with late drug selection (days 9 and 15) or no selection. Moreover, bisulfite sequencing of the endogenous Oct4 locus revealed that in Fbx15-iPS cells obtained from early drug selection, the locus was largely methylated. On the other hand, in Fbx15-iPS cells attained from late drug selection, the Oct4 locus was completely unmethylated, as reported for Nanog-iPS cells and ES cells (Okita, Ichisaka et al. 2007). The control H19 DMR region was methylated in both cases (Figure 2d). These observations seem to suggest that early drug selection prevents complete epigenetic reprogramming and reactivation of pluripotency markers, thereby leading to partially reprogrammed intermediates. The use of Fbx15 activation as a selection criterion has little or no bearing on the process of complete reprogramming.

**DISCUSSION**

The generation of iPS cells by *in vitro* reprogramming of somatic cells has been a remarkable achievement for the scientific and medical communities. These cells

have been shown to hold the same potential that ES cells have in the treatment of different diseases, and are not plagued by the moral dilemmas surrounding ES cells. In proof-of-principle experiments, iPS cells have been used to treat a mouse model of sickle-cell anemia (Hanna, Wernig et al. 2007), as well as alleviate symptoms of Parkinson's disease in a rat model of the disorder (Wernig, Zhao et al. 2008). This process of reprogramming by the introduction of fhe transcription factors, Oct4, Sox2, Klf4 and c-Myc has also been accomplished in human cells (Takahashi, Tanabe et al. 2007; Yu, Vodyanik et al. 2007; Park, Zhao et al. 2008).

Despite this significant progress in reprogramming, not much is known about the mechanism by which this process occurs. Before iPS cells are used in medical applications, it will be important to gain more insight into the process by which they are obtained. As mentioned earlier, recent studies have begun to show that different markers of pluripotency, such as AP, SSEA1, Oct4 and Nanog are sequentially activated during the course of reprogramming (Brambrink, Foreman et al. 2008; Stadtfeld, Maherali et al. 2008). The presence of all these markers is indicative of complete reprogramming, since cells obtained at earlier time points are not pluripotent. Moreover, expression of the transgenes encoding the four reprogramming genes is required for at least 12-16 days in order to get stably reprogrammed cells. In addition to this, it is also known that early drug selection for the activation of ES cell markers leads to fewer reprogrammed cells, and starting such selection at later time points greatly increases the number of

reprogrammed cells obtained (Meissner, Wernig et al. 2007; Wernig, Meissner et al. 2007).

Although these studies had shed some light on the mechanism of reprogramming, it was not clear whether the choice of selection marker used for obtaining iPS cells had an effect on the reprogramming process. This question arose from the observation that iPS cells selected for the activation of Oct4 or Nanog were truly pluripotent, but those selected for Fbx15 were only partially reprogrammed, and represented an "intermediate" stage in reprogramming. Since iPS cells in the former case had been obtained after initiating drug selection at a later time point than in the latter case, another possibility was that these intermediates were seen due to early drug initiation.

In order to test both these possibilities, we performed reprogramming experiments with Fbx15-βgeo fibroblasts while initiating drug selection at different time points, and analyzed the iPS cells from each experiment by counting the number of drug-resistant colonies obtained at each time point, staining for the presence of pluripotency markers and examining the methylation status of the Oct4 locus. In accordance with previous observations, our results indicated that later initiation of drug selection led to more resistant colonies. Although all colonies showed expression of the pluripotency markers AP and SSEA1, only those at later time points stained positive for Oct4 and Nanog. It was not clear why all cells did not express Oct4 since they had been infected with a virus that

constitutively expressed the gene. It is possible that at the time of immunostaining the viral transgene was not expressed at a level that was high enough to be detected by this method. Regardless, the bisulfite sequencing results showed that the Oct4 locus in colonies obtained with early drug selection was largely methylated, whereas it was completely unmethylated in cells from late drug selection. These results suggested that early initiation of drug selection led to partially reprogrammed or "intermediate" cells since late selection for Fbx15 activation led to fully reprogrammed cells, as indicated by complete epigenetic reprogramming of the Oct4 locus. Such epigenetic reprogramming is possible at later stages of the process since *de novo* methyltransferases Dnmt3a and Dnmt3b are reactivated later.

An interesting observation in this study came from a comparison of colony counts between Fbx15- and Nanog-iPS cells. Even though the same number of fibroblasts had been infected with the reprogramming factors in both cases, the number of colonies obtained from Fbx15 selection was greater than from Nanog selection. This observation has also been made earlier by Yamanaka and colleagues, where the induction efficiency of Nanog-iPS cells was one-tenth that of Fbx15-iPS cells (Okita, Ichisaka et al. 2007). Although the reason for this difference in induction efficiency is not clear, it is possible that this discrepancy arose since the fibroblasts were of different passage numbers. In our experiments, the Fbx15-βgeo fibroblasts were of a much lower passage number (Passage 2) than the Nanog-neo fibroblasts, which may have provided them with

some growth advantage. If this is the case, then an experiment comparing induction efficiencies of Fbx15-βgeo and Nanog-neo fibroblasts of the same passage number will address this issue. The alternative possibility is that selection for Fbx15 activation leads to greater enrichment for drug resistant cells than selection for Nanog does. The proportion of completely reprogrammed cells in both these cases can be compared by including an Oct4-GFP marker in the fibroblasts, and screening for GFP-positive clones. If the proportion of GFP-positive cells is higher for Fbx15-iPS cells, then Fbx15 may be a more useful marker to enrich for fully reprogrammed cells. However, if this number is lower than that for Nanog-iPS cells, then Fbx15 selection may lead to enrichment for intermediates in the reprogramming process.

The results of this study emphasize the need for appropriate timing of drug selection for complete reprogramming, and indicate that Fbx15 could be added to the repertoire of selection markers that can be used in reprogramming assays. A number of technical issues still need to be addressed before iPS cells or their derivatives can be used in a patient. These include elimination of virus-mediated transduction of transgenes, since the integration of viruses in the host genome could be harmful. Moreover, even though iPS cells can now be derived without the oncogene c-Myc (Nakagawa, Koyanagi et al. 2008; Wernig, Meissner et al. 2008), it is still unclear whether the introduction of the other three factors will lead to tumors at later stages. In order to address many of these issues it will be necessary to screen for safer substitute factors or chemicals that can be used to

obtain iPS cells. In such screens, it may be advantageous to use ES cell markers, such as Fbx15 to select for reprogramming, while the expression of genes like Oct4 and Nanog, which are relevant to pluripotency, can be maintained at endogenous levels. The identification of other such markers may be useful in the future as different screening methods are devised to reprogram differentiated cells back to an embryonic state.

## METHODS

### Generation of Fbx15-βgeo ES cells and Mice

The strategy for generating the Fbx15-βgeo targeting construct was similar to the one described earlier (Tokuzawa, Kaiho et al. 2003), which replaced exons 3 to 7 of the *Fbx15* gene with an IRES (internal ribosome entry site)-βgeo (fusion of genes encoding for β-galactosidase and neomycin-resistance) cassette (Mountford, Zevnik et al. 1994). The construct was linearlized with *Spe*I and introduced into ES cells by electroporation. These ES cells had been derived from matings between mice that had a doxycycline-inducible Oct4 construct in the Collagen 1a1 locus (Hochedlinger, Yamada et al. 2005), and those that had a doxycycline-inducible Nanog construct in the same locus. Therefore, the ES cells used for targeting had inducible Oct4 and Nanog constructs in the collagen locus and a copy of the M2rtTA reverse transactivator in the Rosa26 locus (Beard, Hochedlinger et al. 2006).

Targeted ES cells were selected with neomycin (350 μg/ml), and antibiotic-resistant colonies were screened for homologous recombination by Southern blot analysis. DNA was isolated from 46 neomycin-resistant colonies, digested with *Hind*III and run on a 0.8% agarose gel. The DNA was nicked by soaking the gel in 0.25M HCl for 15 minutes. The gel was then soaked for another 15 minutes in 0.4M NaOH and the DNA was transferred to a nitrocellulose membrane in the same solution. After the transfer was complete, the membrane was washed in 0.2M Tris, pH 7.0 and 2X SSC for 30 seconds each, and air-dried. The membrane was prehybridized in Church buffer for 15 minutes at 65°C. A Strategene kit for labeling DNA was used to label 30 ng of probe with P-32, and purified probe was added to the membrane and Church buffer. After an overnight hybridization with the probe, the membrane was washed with 2X SSC, 0.2% SDS for 15 minutes at room temperature, then twice for 30 minutes each with 0.2% SSC, 0.2% SDS at 65°C. The membrane was exposed to an autoradiography film overnight at -80°C. The 500 bp external 3' probe from intron 8 (+31400 to +31999 bp) and produced A 7.4 kb band from the wild-type locus and a 6.8 kb band from the targeted one. PCR primers for amplifying this probe were: Forward (ATGTCTTGCTCTTTGGA GGGAGGCAG) and Reverse (TCCTCTGTACCTCCTCATGAGCATTC).

An ES cell clone that was correctly targeted with the Fbx15-βgeo construct was injected into diploid C57/BL6 blastocysts according to previously described procedures (Meissner, Wernig et al. 2007). These blastocysts were implanted in

pseudopregnant B6D2F1 females in order to obtain chimeric pups. Two embryos were extracted early at d13.5 in order to generate mouse embryonic fibroblasts (MEFs). MEFs from each embryo were selected with Puromycin (2μg/ml) after 24 hrs for 1 week. This selection eliminated all wild-type MEFs and only those carrying the Fbx15-βgeo remained on the dishes. These cells were expanded, frozen down, and used in the reprogramming experiments described here. The chimeras obtained were mated with B6D2F1 females to obtain germline transmission of the Fbx15-βgeo construct. Male and female mice carrying the targeted allele in their germline were obtained and screened by a PCR screening strategy. PCR primers were used to amplify approximately 500 bp of the region encoding for β-galactosidase in the Fbx15-βgeo construct: Forward (CGGTGATGGTGCTGCGTTGG) and Reverse (GAATCAGCAACGGC TTGCCG). The reaction was set up as described below:

*LacZ PCR Reaction for Genotyping Fbx15-βgeo Mice:*

Template DNA: 1 μl
10X Buffer (USB Fidelitaq): 2 μl
50 mM $MgCl_2$: 0.8 μl
10 mM dNTP: 0.4 μl
20 μM primer mix: 0.2 μl
Taq (USB Fidelitaq): 0.2 μl
$dH_2O$: 15.4 μl
Total Volume = 20 μl

*PCR Cycle*

1. 95°C: 1 min
2. 95°C: 45 s
3. 55°C: 45 s
4. 72°C: 1 min

5. Repeat steps 2-4 35X more
6. 72°C: 10 min

**Viral infections and Drug Selection Experiments**

Lentiviral constructs that constitutively expressed the *Oct4, Sox2, Klf4* and *c-Myc* transgenes, and virus preparation methods have been described previously (Brambrink, Foreman et al. 2008). Viruses carrying the four transgenes were used to infect $2.5x10^4$ Fbx15-βgeo and Nanog-neo MEFs/10 $cm^2$ plate coated with 0.2% gelatin. Neomycin selection was started at days 3, 6, 9 and 15 for both types of MEFs, and the media was changed every other day. Additional plates with no drug selection were also included in the experiment. The selection was carried out for 21 days, after which drug resistant colonies that morphologically resembled ES cells were scored, picked and passaged.

**Immunostaining protocol**

For immunostaining, iPS cells were grown in 6-well dishes on feeders. After removing the ES cell media, they were washed once with Hepes buffer. Using a cotton tip applicator, each well was marked along its edge and divided into four quadrants. These edges were then sealed with a pap-pen. The cells were fixed in 4% paraformaldehyde for 30 minutes at room temperature, and washed once with 1X PBS. For the quadrants in which AP staining was done, an AP staining kit for Vector was used. Briefly, in 5 ml of AP buffer (100 mM Tris, pH 8.3) 2 drops of reagents 1, 2 and 3 were added. This solution was added to the cells that were stained for AP, and was removed after 10 minutes when positive

controls stained red for the presence of AP. In the other quadrants, cells were stained for the presence of SSEA, Oct4 and Nanog. These wells were blocked with 5% CCS (Cosmic Calf Serum). For cells that were to be stained for Oct4 and Nanog, 0.1% TritonX-100 was also added to the blocking solution to permeablize the cells. After removing the block, primary antibodies against Oct4 (Santa Cruz, sc-9081 (H-134)), Nanog (Bethyl, BL1662) and SSEA1 (Developmental Studies Hybridoma Bank) were added at a dilution of 1:100 in 0.1% FBS, 0.1% TritonX-100 (no TritonX-100 was used for SSEA1 staining), and incubated overnight at 4°C. The cells were washed 3x with 1X PBS and incubated with secondary antibodies (in the same solutions used for diluting the primary antibodies) for 1.5 hours at room temperature. The cells were then counterstained with DAPI for 5 minutes and washed 3 times with 1X PBS. Immmunofluorescence was detected using an Olympus Fluorescence microscope and pictures were taken with a Zeiss Axiocam camera.

**Bisulfite sequencing protocol**

For bisulfite sequencing, iPS cells were grown in 6-well plates. For DNA preparation, the cells were washed once with Hepes buffer and incubated in 500 μl of lysis buffer containing 200 μg/ml of Proteinase K for 2 hours at 37°C. The DNA was then extracted with an equal volume of Phenol/choloroform using phaselock tubes (Eppendorf), followed by an isopropanol and ethanol precipitation. DNA was suspended in 150 μl of TE buffer and incubated overnight at 55°C. 2μg of DNA was used for bisulfite treatment according to the

protocol described in the Epitect Bisulfite Kit (Qiagen). The treated DNA was purified and eluted in 20 μl of elution buffer, and used in the following nested PCR reactions to examine the methylation status of the Oct4 locus and H19 DMR region:

*Oct4 PCR Reaction 1*

Template DNA: 1.5 μl
10X Buffer: 2.5 μl
50 mM $MgCl_2$: 1 μl
100 μM Primer F1: 0.25 μl
100 μM Primer R: 0.25 μl
10mM dNTPs: 0.5 μl
Taq: 1 μl
$dH_2O$: 18 μl
(Total Volume = 25 μl)

*Oct4 PCR Reaction 2*

Template DNA (from reaction 1): 1 μl
10X Buffer: 2.5 μl
50 mM $MgCl_2$: 1 μl
100 μM Primer F2: 0.25 μl
100 μM Primer R: 0.25 μl
10mM dNTPs: 0.5 μl
Taq: 1 μl
$dH_2O$: 18.5 μl
(Total Volume = 25 μl)

*Oct4 PCR Cycle (Reactions 1 and 2)*

1. 94°C: 4 min
2. 94°C: 30 s
3. 56°C: 1 min (-1°C for each cycle)
4. 72°C: 1 min
5. Repeat steps 2-4 4X more
6. 94°C: 30 s
7. 51°C: 45 s
8. 72°C: 1 min 20 s
9. Repeat steps 6-8 29X more
10. 72°C: 10 min

*H19 PCR Reaction 1*

Template DNA: 4 μl
10X Buffer: 5 μl
50 mM $MgCl_2$: 2 μl
100 μM Primer F1: 1 μl
100 μM Primer R1: 1 μl
10mM dNTPs: 1 μl
Taq: 1 μl
$dH_2O$: 35 μl
(Total Volume = 50 μl)

*H19 PCR Reaction 2*

Template DNA (from reaction 1): 2 μl
10X Buffer: 5 μl
50 mM $MgCl_2$: 2 μl
100 μM Primer F2: 1 μl
100 μM Primer R2: 1 μl
10mM dNTPs: 1 μl
Taq: 1 μl
$dH_2O$: 37 μl
(Total Volume = 50 μl)

*H19 PCR Cycle (Reaction1):*

1. 94°C: 4 min
2. 55°C: 2 min
3. 72°C: 2 min
4. Repeat steps 1-3 1X more
5. 94°C: 1 min
6. 55°C: 2 min
7. 72°C: 2 min
8. Repeat steps 5-7 34X more
9. 72°C: 10 min

*H19 PCR Cycle (Reaction 2):*

1. 94°C: 1 min
2. 55°C: 2 min
3. 72°C: 2 min
4. Repeat steps 1-3 34X more
5. 72°C: 10 min

A region of the Oct4 promoter was amplified using two forward primers (F1 and F2) and one reverse primer (R): F1 (GTTGTTTTGTTTTGGTTTTGGATAT), F2 (ATGGGTTGAAATATTGGGTTTATTTA) and R (CCACCCTCTAACCTTAACCTC TAAC) (Blelloch, Wang et al. 2006). The H19 DMR regions was amplified using two forward primers (F1 and F2) and two reverse primers (R1 and R2): F1 (GAG TAT TTA GGA GGT ATA AGA ATT), F2 (GTA AGG AGA TTA TGT TTA TTT TTG G), R1 (ATC AAA AAC TAA CAT AAA CCC CT) and R2 (CCTCATTAATCC CATAACTAT) (Lucifero, Mertineit et al. 2002). All products from the second PCR reactions were run on a 1% gel, purified using a gel purification kit (Qiagen) and Topo-cloned into pCR2.1 vector. The colonies were grown on LB plates containing X-Gal to screen for white colonies that would contain an insert of the purified PCR product. These colonies were grown in liquid cultures and DNA was extracted from them and sequenced using an M13 Reverse primer. The DNA sequences were aligned to the corresponding *Oct4* or *H19* regions and analyzed for their methylation status using the program Sequencher 4.7.

**ACKNOWLEDGEMENTS**

I would like to thank Jessie Daussman for performing blastocyst injections, and Alex Meissner and Caroline Beard for helpful discussions with this project.

## REFERENCES


Beard, C., K. Hochedlinger, et al. (2006). "Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells." Genesis **44**(1): 23-8.

Blelloch, R., Z. Wang, et al. (2006). "Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus." Stem Cells **24**(9): 2007-13.

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

Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-55.

Hanna, J., M. Wernig, et al. (2007). "Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin." Science **318**(5858): 1920-3.

Hochedlinger, K., Y. Yamada, et al. (2005). "Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues." Cell **121**(3): 465-77.

Lucifero, D., C. Mertineit, et al. (2002). "Methylation dynamics of imprinted genes in mouse germ cells." Genomics **79**(4): 530-8.

Maherali, N., R. Sridharan, et al. (2007). "Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution." Cell Stem Cell **1**(1): 55-70.

Meissner, A., M. Wernig, et al. (2007). "Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells." Nat Biotechnol **25**(10): 1177-81.

Mitsui, K., Y. Tokuzawa, et al. (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-42.

Mountford, P., B. Zevnik, et al. (1994). "Dicistronic targeting constructs: reporters and modifiers of mammalian gene expression." Proc Natl Acad Sci U S A **91**(10): 4303-7.

Nakagawa, M., M. Koyanagi, et al. (2008). "Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts." Nat Biotechnol **26**(1): 101-6.

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

Okita, K., T. Ichisaka, et al. (2007). "Generation of germline-competent induced pluripotent stem cells." Nature.

Park, I. H., R. Zhao, et al. (2008). "Reprogramming of human somatic cells to pluripotency with defined factors." Nature **451**(7175): 141-6.

Stadtfeld, M., N. Maherali, et al. (2008). "Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse." Cell Stem Cell **2**(3): 230-40.

Takahashi, K., K. Tanabe, et al. (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-72.

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

Tokuzawa, Y., E. Kaiho, et al. (2003). "Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development." Mol Cell Biol **23**(8): 2699-708.

Wernig, M., A. Meissner, et al. (2008). "c-Myc is dispensable for direct reprogramming of mouse fibroblasts." Cell Stem Cell **2**(1): 10-2.

Wernig, M., A. Meissner, et al. (2007). "In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state." Nature.

Wernig, M., J. P. Zhao, et al. (2008). "Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease." Proc Natl Acad Sci U S A **105**(15): 5856-61.

Winston, J. T., D. M. Koepp, et al. (1999). "A family of mammalian F-box proteins." Curr Biol **9**(20): 1180-2.

Yu, J., M. A. Vodyanik, et al. (2007). "Induced pluripotent stem cell lines derived from human somatic cells." Science **318**(5858): 1917-20.

# Chapter 4

# Perspectives

## Genome-wide Approaches to Identify Protein-DNA Interactions

The last decade has seen the insurgence of a number of chromatin immunoprecipitation (ChIP)-based technologies to facilitate the identification of genomic binding targets of proteins, such as transcription factors. These regions can be identified by hybridizing ChIP DNA to array platforms (ChIP-chip) (Horak and Snyder 2002), or through sequencing based technologies, such as ChIP-PET (Loh, Wu et al. 2006), ChIP-SACO (Impey, McCorkle et al. 2004) and ChIP-STAGE (Bhinge, Kim et al. 2007). Most recently, ChIP-Seq was introduced as a sequencing based method to examine protein-DNA binding events throughout the genome (Johnson, Mortazavi et al. 2007). This technology combines the next-generation Solexa sequencing method (Bentley 2006) with traditional ChIP, in order to obtain short sequence reads that can be mapped to the reference genome to identify ChIP-enriched fragments. This method offers several advantages over ChIP-chip, including more rapid and extensive genome coverage, as well as higher resolution of binding sites. The sequencing method used in ChIP-seq also provides greater sequencing depth than ChIP-PET, and does not involve the cloning of ChIP-enriched DNA. Therefore, it is likely that in the near future, newer technologies like ChIP-seq are going to be used instead of more traditional array and sequencing based methods to identify protein-DNA interactions in a more rapid and comprehensive manner.

As newer technologies emerge, there is a need to assess the advantages and limitations of each platform, and compare and integrate data obtained from them

with previously known results. Chapter 2 discusses ways in which such analyses can be done for data obtained for the embryonic stem (ES) cells transcription factors, Oct4 and Nanog from two platforms, ChIP-chip and ChIP-PET. The results of this study indicated that the data set from each platform was only a partial representation of the overall network, and these data should be used in a complementary fashion to obtain a more thorough overview of the transcriptional circuitry. These results also stress the need for performing such comparative analyses to assess protein-DNA interaction data from other emerging platform technologies, such as ChIP-seq. Since the methods for performing such assessments are discussed in greater detail in Chapter 2, the remainder of this section will focus on the uses and limitations of ChIP-based genome-wide approaches for identifying protein-DNA interactions.

*Uses of ChIP-based Approaches in Examining Protein-DNA Interactions*

The availability of whole genome sequences for different organisms has paved the way for genome-wide analyses of DNA-binding factors, largely based on the ChIP-based approaches mentioned earlier. These studies have been instrumental in the identification of DNA-binding patterns of a number of proteins, such as transcription factors and chromatin-remodeling and modification complexes, in simple organisms like yeast (Lee, Rinaldi et al. 2002), as well as in mammalian systems (Horak, Mahajan et al. 2002; Odom, Zizlsperger et al. 2004). For example, ChIP-chip studies on determining transcriptional regulation by the Polymerase III machinery (RNA Polymerase III, TFIIIB and TFIIIC)

identified a number of target genes that encode untranslated RNAs. Furthermore, under starvation conditions, the expression of these RNAs was regulated by inhibiting the recruitment of RNA Polymerase III and TFIIIB to these sites. However, TFIIIC remained bound to this set of target genes (Harismendy, Gendrel et al. 2003; Roberts, Stewart et al. 2003; Moqtaderi and Struhl 2004). Therefore, these studies demonstrate the manner in which ChIP-chip can be used to examine the patterns of gene regulation by different components of the transcriptional machinery under varying cell states and conditions. Such ChIP-based, genome-wide investigations can be expanded to other model organisms, as well as *in vitro* systems such as ES cells, to explore the manner in which different DNA-binding factors modulate gene expression patterns under a wide variety of conditions.

*Validation of Data Obtained from Genome-Wide Protein-DNA Interaction Studies*

It is worth appreciating that even though such genome-wide approaches yield vast amounts of information on protein-DNA interactions, the functional relevance of these binding events must still be validated. Approaches, such as RNAi can be used as starting points for performing such analyses, by knocking down the expression of genes that are bound by the protein of interest. Certain genomic targets of the transcription factors Oct4 and Nanog, such as *rif1* and *esrrb* have been examined in this manner to investigate their roles in maintaining ES cell pluripotency (Loh, Wu et al. 2006). Such loss-of-function analyses can also be carried out in a high throughput manner using genome-wide RNAi libraries.

However, in mammalian systems, these large-scale studies are currently feasible only in cells that can be cultured *in vitro*, e.g. ES cells. Therefore, in order to characterize these genomic targets at the organismal level, gene knockouts and overexpression analyses must be done using conventional transgenic methods. Since the design and characterization of transgenic animals is a time- and labor-intensive process, it is often not feasible to validate all targets obtained from these genome-wide ChIP-based methods. This limitation poses a major bottleneck in the extraction of biologically relevant information from these binding data. In the case of Oct4 and Nanog binding events in ES cells, one possible way to prioritize genetic targets for validation would be to first examine those that are conserved between human and mouse ES cells. Given the similarity in properties of ES cells derived from the two species, it is likely that these key transcription factors act in similar ways in both types of ES cells. For this purpose, it will be advantageous to determine the overlap between the binding data for both species, which was obtained using comparable platforms (e.g. ChIP-chip). However, this technique of target prioritization based on conservation across species has the disadvantage of eliminating targets that may account for the differences in the properties of human and mouse ES cells, and have important roles in pluripotency. Therefore, in the future, shifting the focus of high-throughput technology development, from target identification to target validation will be essential in order to use these data to reach meaningful conclusions. Nevertheless, ChIP-based analyses have provided significant

advancements in putting together protein-DNA interaction maps, which can be used as starting frameworks for deciphering mechanisms of gene regulation.

*Limitations of ChIP-Based Approaches*

Despite the wealth of protein-DNA interaction information obtained from ChIP-based studies, these analyses pose several limitations that remain to be addressed. One of these challenges stems from the use of formaldehyde as an agent to cross-link proteins to DNA. Since formaldehyde can create protein-protein and protein-DNA crosslinks, the interaction data obtained from studies can represent both direct as well as indirect associations between proteins and DNA. Moreover, proteins that are otherwise unbound to chromatin can also get cross-linked to it, depending on their proximity to the surrounding DNA and proteins. Therefore, the development of protocols, where proteins can be efficiently immunoprecipitated along with the interacting DNA fragments would be extremely useful in addressing some of these issues. Additionally, since laser-induced UV rays only crosslink proteins to DNA, if methods to reverse such crosslinks can be developed, this technique can be used to distinguish between direct and indirect protein-DNA interactions (Hockensmith, Kubasek et al. 1991).

Aside from the technical issues of the ChIP procedure, one of the other hurdles with obtaining protein-DNA interaction data is that the binding of several proteins, such as transcription factors varies during different stages of the cell cycle. Therefore, synchronization of cells to the same stage of the cell cycle is often

necessary to address this issue. However, it is not feasible to grow cells, such as ES cells in this manner, thereby making the interpretation of protein-DNA interaction data from these unsynchronized cells more tedious. Moreover, these interactions can also vary with changes in the cell's chromatin state. For instance, the association of certain chromatin proteins with DNA is hyperdynamic in ES cells, and can be difficult to capture using ChIP (Meshorer, Yellajoshula et al. 2006).

One of the most important limitations with interpreting binding data is that it often difficult to determine the regulatory role for a binding event. This is particularly true for a number of transcription factors, which not only bind promiscuously throughout the genome, but can also regulate transcription from sites far away from the target gene promoters. Such gene regulation has been observed for the transcription factors, c-Myc and p53, whose DNA-binding sites are located in regions that are distant from the promoters of protein-coding genes (Cawley, Bekiranov et al. 2004). Similar to transcriptional mechanisms observed in prokaryotes (Mossing and Record 1986), eukaryotic transcription factors may also employ methods, such as forming DNA loops to regulate genes at distant sites. In order to address this issue, it will be important to determine the proteins interacting with a transcription factor that impart specificity to its gene-regulatory functions. An attempt at identifying proteins interacting with the ES cell transcription factor, Nanog has been described in the Appendix. Combining protein-protein interaction information with protein-DNA binding data will provide

better insight into the mechanisms by which a transcription factor can modulate the downstream components of its regulatory network.

## The Future of iPS Cell Technology: Designing Reprogramming and Differentiation Screens

Recently, the use of four transcription factors to induce somatic cells into a pluripotent state, has been used to address the issue of reprogramming, one of the biggest challenges in ES cell biology (Takahashi and Yamanaka 2006; Maherali, Sridharan et al. 2007; Okita, Ichisaka et al. 2007; Wernig, Meissner et al. 2007). This technique involves virus-mediated transduction of transgenes encoding the transcription factors, Oct4, Sox2, c-Myc and Klf4 into somatic cells, followed by selection for the activation of pluripotency markers in order to obtain induced pluripotent (iPS) cells. Even though this technique of reprogramming is a major technological feat, as discussed in Chapters 1 and 3, the use of viral-infection methods poses a significant challenge in the application of iPS cells in medical therapies. Therefore, safer non-viral substitutes must be devised for reprogramming somatic cells into a pluripotent state. To this end, it would be desirable to gain further insight into the mechanism of reprogramming, by defining intermediate steps in the process. Furthermore, in order to use iPS cells for therapeutic purposes, differentiation screens need to be designed, in order to obtain a variety of patient-specific cell types that can be used in the treatment of different diseases.

The classification of steps in the reprogramming process, as well as the design of differentiation screens will need an appropriate choice of markers. For instance, the activation of pluripotency factors, such as Nanog is used to select for cells that have been completely reprogrammed to a pluripotent state. Moreover, as shown in Chapter 3, the activation of *fbx15*, a target of Oct4 can also be used to select for reprogramming. However, the reasons for why more drug-resistant colonies are obtained with Fbx15 selection than with Nanog selection are still unclear. It will be interesting to examine this issue by introducing an Oct4-GFP marker in both types of cells, and analyzing the number of completely reprogrammed cells in each case (as indicated by activation of GFP). Such an experiment will tell us whether Fbx15 selection leads to an enrichment for completely reprogrammed iPS cells, or partially reprogrammed intermediates. Additionally, the ChIP-based studies described in Chapter 2 identify a number of downstream targets of the factors, Oct4 and Nanog in ES cells. Some of these genes, which are transcriptionally activated by Oct4 and Nanog, may be useful markers for classification of intermediates in reprogramming. Moreover, Oct4 and Nanog also bind to a number of genes that are silent in pluripotent cells, but are expressed upon differentiation into specific cells lineages. Therefore, such genes could be used in differentiation screens in order to select for cells that have differentiated into specific lineages, and express these markers.

Reprogramming and differentiation screens can be designed using both chemical and biological approaches. Chemical screens can be used to identify small

molecule compounds that can successfully reprogram somatic cells into iPS cells, and differentiate iPS cells into other specific cell lineages. Additionally, a number of developmentally important signaling pathways, such as Wnt (Cole, Johnstone et al. 2008), JAK/STAT (Niwa, Burdon et al. 1998; Matsuda, Nakamura et al. 1999)and NF-κB (Torres and Watt 2008), are known to regulate pluripotency, as well as differentiation. Therefore, another approach to designing reprogramming and differentiation screens would be to identify chemical and biological molecules (e.g. antibodies) that specifically target components of these pathways. The use of such multiple, parallel approaches will allow us to determine safe and efficient ways in which pluripotent cells can be obtained and employed in the treatment of a number of medical disorders.

**REFERENCES:**


Bentley, D. R. (2006). "Whole-genome re-sequencing." Curr Opin Genet Dev **16**(6): 545-52.

Bhinge, A. A., J. Kim, et al. (2007). "Mapping the chromosomal targets of STAT1 by Sequence Tag Analysis of Genomic Enrichment (STAGE)." Genome Res **17**(6): 910-6.

Cawley, S., S. Bekiranov, et al. (2004). "Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs." Cell **116**(4): 499-509.

Cole, M. F., S. E. Johnstone, et al. (2008). "Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells." Genes Dev **22**(6): 746-55.

Harismendy, O., C. G. Gendrel, et al. (2003). "Genome-wide location of yeast RNA polymerase III transcription machinery." Embo J **22**(18): 4738-47.

Hockensmith, J. W., W. L. Kubasek, et al. (1991). "Laser cross-linking of protein-nucleic acid complexes." Methods Enzymol **208**: 211-36.

Horak, C. E., M. C. Mahajan, et al. (2002). "GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis." Proc Natl Acad Sci U S A **99**(5): 2924-9.

Horak, C. E. and M. Snyder (2002). "ChIP-chip: a genomic approach for identifying transcription factor binding sites." Methods Enzymol **350**: 469-83.

Impey, S., S. R. McCorkle, et al. (2004). "Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions." Cell **119**(7): 1041-54.

Johnson, D. S., A. Mortazavi, et al. (2007). "Genome-wide mapping of in vivo protein-DNA interactions." Science **316**(5830): 1497-502.

Lee, T. I., N. J. Rinaldi, et al. (2002). "Transcriptional regulatory networks in Saccharomyces cerevisiae." Science **298**(5594): 799-804.

Loh, Y. H., Q. Wu, et al. (2006). "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells." Nat Genet **38**(4): 431-40.

Maherali, N., R. Sridharan, et al. (2007). "Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution." Cell Stem Cell **1**(1): 55-70.

Matsuda, T., T. Nakamura, et al. (1999). "STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells." Embo J **18**(15): 4261-9.

Meshorer, E., D. Yellajoshula, et al. (2006). "Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells." Dev Cell **10**(1): 105-16.

Moqtaderi, Z. and K. Struhl (2004). "Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes." Mol Cell Biol **24**(10): 4118-27.

Mossing, M. C. and M. T. Record, Jr. (1986). "Upstream operators enhance repression of the lac promoter." Science **233**(4766): 889-92.

Niwa, H., T. Burdon, et al. (1998). "Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3." Genes Dev **12**(13): 2048-60.

Odom, D. T., N. Zizlsperger, et al. (2004). "Control of pancreas and liver gene expression by HNF transcription factors." Science **303**(5662): 1378-81.

Okita, K., T. Ichisaka, et al. (2007). "Generation of germline-competent induced pluripotent stem cells." Nature.

Roberts, D. N., A. J. Stewart, et al. (2003). "The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships." Proc Natl Acad Sci U S A **100**(25): 14695-700.

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

Torres, J. and F. M. Watt (2008). "Nanog maintains pluripotency of mouse embryonic stem cells by inhibiting NFkappaB and cooperating with Stat3." Nat Cell Biol **10**(2): 194-201.

Wernig, M., A. Meissner, et al. (2007). "In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state." Nature.

# Appendix

## IDENTIFICATION OF PROTEINS INTERACTING WITH THE PLURIPOTENCY REGULATOR NANOG

Divya Mathur[1,2], Betty Chang[3], Steven A. Carr[3] and Rudolf Jaenisch[1,2].

[1]Department of Biology, Massachusetts Institute of Technology, 32 Ames Street, Cambridge, MA 02139, USA

[2]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

[3]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA

### RESPECTIVE CONTRIBUTIONS

DM made the tagged-Nanog constructs and performed the protein purifications. BC did the mass spectrometry analyses.

## INTRODUCTION

This study aims to gain insight into the regulation of pluripotency by identifying the proteins interacting with the pluripotency factor Nanog in murine embryonic stem (ES) cells. Nanog is a novel homeodomain transcription factor, and has been shown to play a key role in maintaining ES cell identity as well as in early cell fate decisions in the developing embryo (Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003). Along with Oct4, Nanog is one of the earliest expressed transcription factors known to be critical for developmental fate decisions, since embryos and ES cells deficient for Nanog differentiate into primitive endoderm (Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003).

Expression analyses of Nanog have suggested a dual role for the protein in maintaining not only an undifferentiated state but stem cell self-renewal as well. In the preimplantation embryo, *nanog's* expression begins in the inner cells of the morula and is ultimately downregulated by early post-implantation stages when the epiblast differentiates into different germ lineages. *Nanog* is expressed predominantly in pluripotent cells, including cells of the inner cell mass (ICM), ES cells, embryonic germ cells, proliferating germ cells and embryonic carcinoma cells (Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003; Wang, Tsai et al. 2003; Hart, Hartley et al. 2004; Yamaguchi, Kimura et al. 2005). Interestingly, ES cells that overexpress Nanog can bypass the otherwise essential stimulation by the cytokine Leukemia Inhibitory Factor (LIF), which activates the transcription factor Stat3 to maintain self-renewal (Niwa, Burdon et al. 1998; Matsuda,

Nakamura et al. 1999; Chambers, Colby et al. 2003). Hyperactivation of Stat3 has no effect on the endogenous levels of Nanog, indicating that Nanog's role in regulating self-renewal is independent of Stat3.

The ability to sustain ES cell self-renewal in the absence of LIF is a unique property of Nanog, additional to its known function of preventing differentiation. Therefore, Nanog may play a bifunctional role – one as an activator of self-renewal genes and the other as a repressor of differentiation genes (Pan and Pei 2003; Pan and Pei 2005). Currently, most evidence suggests that Nanog acts as a repressor of differentiation. The DNA recognition motif of Nanog is present in the promoter of *gata6*, an inducer of primitive endoderm differentiation (Mitsui, Tokuzawa et al. 2003). Activation of the GATA factors, GATA4 and GATA6, is required for differentiation into primitive endoderm (Fujikura, Yamato et al. 2002). Forced expression of these factors leads to spontaneous differentiation into primitive endoderm. In fact, *nanog* deficient cells, which have a similar phenotype show upregulation of these GATA factors. It remains to be seen if indeed Nanog blocks the effects of GATA6 and GATA4 in order to maintain pluripotency. Although there has been no *in vivo* evidence of Nanog acting as a transcriptional activator, luciferase reporter assays indicate that Nanog can activate transcription via its C-terminal domain (Pan and Pei 2005). Further investigations of how Nanog can activate or repress the transcription of its targets are required to establish its suggested bifunctional role in sustaining pluripotency.

This project aims to gain better insight into the mechanism by which Nanog regulates the transcription of its targets by identifying physiologically relevant binding partners of Nanog in ES cells. In order to accomplish this goal, a functional FLAG-HA epitope-tagged construct of Nanog was generated and targeted to the endogenous Nanog locus. We purified proteins associated with Nanog from ES cells expressing this tagged construct, and identified these using mass spectrometry. Our preliminary data include the proteins BAF60b, γ-catenin, histone variant H2A.Z, Myb binding protein and eIF4a3. Further validation of these protein interactions, and an examination of their roles in the maintenance of pluripotency will lead to a better understanding of the mechanism by which the transcription factor Nanog regulates cell-fate determination.

## RESULTS

### Generation of FLAG-HA Tagged Nanog constructs

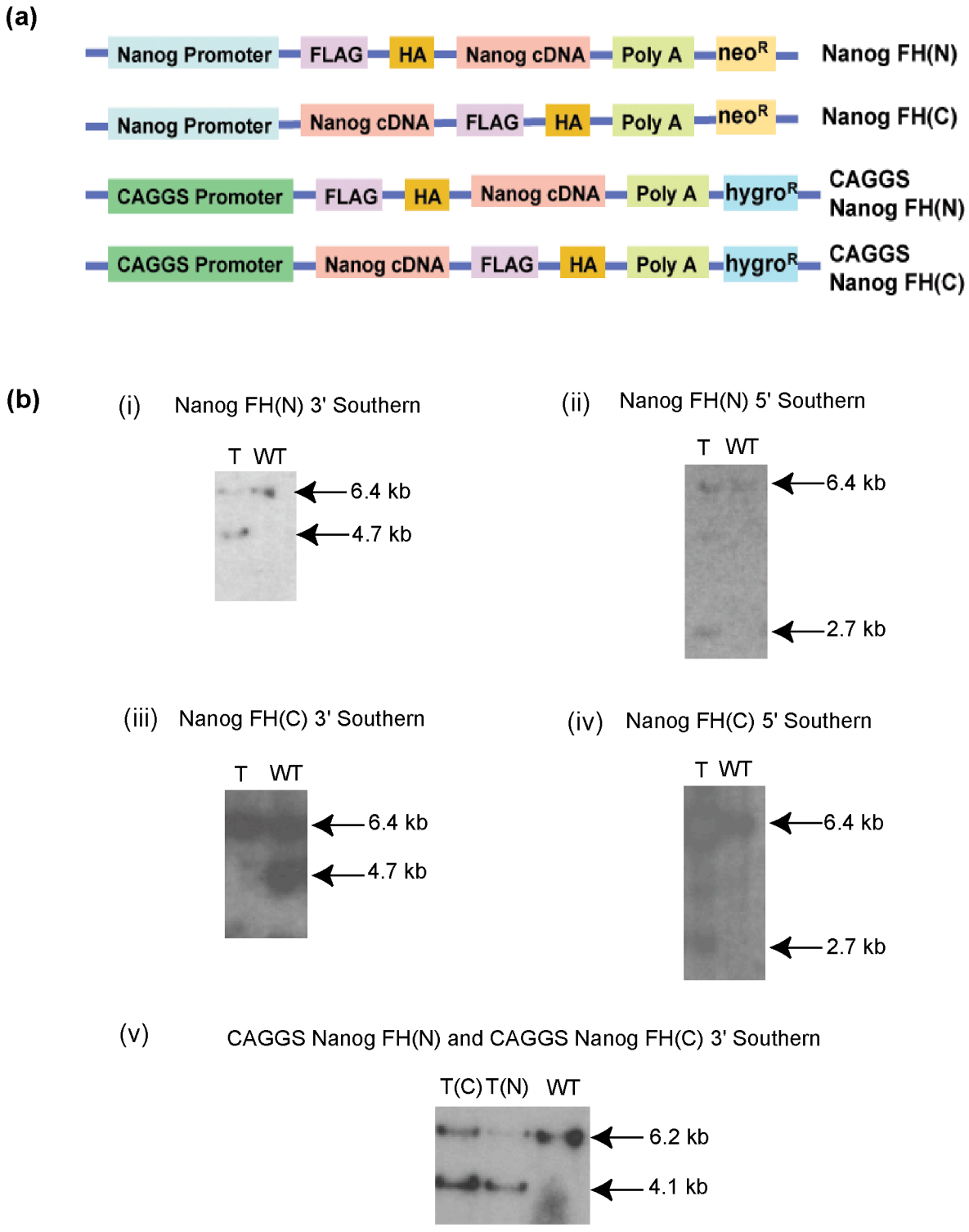
Four different constructs were generated with *nanog* cDNA tagged with FLAG and HA epitopes (Figure 1a). Two of these had the FLAG-HA sequences tagged to the N-terminus of the *nanog* cDNA, and the other two had these tags on the C-terminus end. One N-terminally tagged construct and another C- terminally tagged one were targeted downstream of the *nanog* promoter, These constructs were termed Nanog FH(N) and Nanog FH(C), respectively. The other two constructs were inserted into an overexpression cassette, under the control of the ubiquitously strong CAGGS promoter. These constructs, hereafter referred to as CAGGS Nanog FH(N) (for the N-terminally tagged construct) and CAGGS

Nanog FH(C) (for the C-terminally tagged construct), were targeted downstream of the *collagen1a1* locus (used for high targeting efficiency) using FLP/frt mediated site-specific integration.

The constructs Nanog FH(N) and Nanog FH(C) were targeted to V6.5 ES cells, and CAGGS Nanog FH(N) and CAGGS Nanog FH(C) were targeted to C10 ES cells, which contained an frt site at the *collagen1a1* locus. After drug selection, all clones were screened by Southern blot analyses to check for correctly targeted ES cells (Figure 1b). An external probe in the 3' end of *nanog* intron 1 was used to screen targeted with the Nanog FH(N) and Nanog FH (C) constructs, where a 6.4 kb band represented the wild-type *nanog* allele, and a 4.7 kb one represented the targeted one. A 5' end internal probe from the Nanog promoter was also used to screen these ES cells, and in this case, the wild-type *nanog* allele gave a 6.4 kb band, and the targeted allele gave a 2.7 kb one. Using this strategy, 4 out of 48 clones for Nanog FH (N) and 6 out of 48 for Nanog FH(C) were determined as correctly targeted. ES cells targeted with the overexpression constructs were screened with an internal 3' probe, where the wild-type allele gave a 6.2 kb band, the frt allele gave a 6.7 kb band and a correctly targeted flp-in allele gave a 4.1 kb band. All resistant ES cell colonies for CAGGS Nanog FH(N) (I out of 1) and CAGGS Nanog FH(C) (20 out of 20) were correctly targeted.

**Figure 1. FLAG-HA Epitope Tagged Nanog Constructs.** (a) The four targeting constructs of Nanog tagged with FLAG and HA epitopes at the N- or C-termini and targeted to the *nanog* or *collagen1a1* locus. (b) Southern blot analyses to screen for ES cells clones correctly targeted with the tagged Nanog constructs. Lanes with DNA from correctly targeted ES cell clones are labeled 'T' and untargeted wild type controls are labeled 'WT'. In Southerns (i) and (iii), the wild type locus produced a 6.4 kb band and the targeted one produced a 4.7 kb band. In Southerns (ii) and (iv), the wild type locus produced a 6.4 kb band and the targeted one produced a 2.7 kb one. In Southern (v) the wild type locus is represented by a 6.2 kb band and the targeted one produced a 4.1 kb band. In this blot, 'T(C)' represents correctly targeted CAGGS Nanog FH(C), and 'T(N)' represents correctly targeted CAGGS Nanog FH(N).

**FIGURE 1.**


(a)
Nanog Promoter
FLAG
HA
Nanog cDNA
Poly A
neoR
Nanog FH(N)
Nanog Promoter
Nanog cDNA
FLAG
HA
Poly A
neoR
Nanog FH(C)
CAGGS Promoter
FLAG
HA
Nanog cDNA
Poly A
hygroR
CAGGS Nanog FH(N)
CAGGS Promoter
Nanog cDNA
FLAG
HA
Poly A
hygroR
CAGGS Nanog FH(C)
(b)
(i) Nanog FH(N) 3' Southern
T WT
6.4 kb
4.7 kb
(ii) Nanog FH(N) 5' Southern
T WT
6.4 kb
2.7 kb
(iii) Nanog FH(C) 3' Southern
T WT
6.4 kb
4.7 kb
(iv) Nanog FH(C) 5' Southern
T WT
6.4 kb
2.7 kb
(v) CAGGS Nanog FH(N) and CAGGS Nanog FH(C) 3' Southern
T(C) T(N) WT
6.2 kb
4.1 kb

## LIF-Independence Functionality Tests for Tagged Nanog Constructs

In order the check for the functionality of the N- and C-terminally tagged Nanog constructs, we utilized the ability of ES cells that overexpress Nanog to grow in the absence of LIF. In this assay, the ES cells targeted with the overexpression constructs provided a functional assay for the tagged proteins. All four types of targeted ES cell lines, Nanog FH(N), Nanog FH(C), CAGGS Nanog FH(N) and CAGGS Nanog FH(C), as well as untargeted V6.5 ES cells were grown in the presence and absence of LIF (Figure 2a). All ES cell lines grew stably in the presence of LIF. When LIF was withdrawn from the growth media, the cells expressing Nanog at endogenous levels differentiated, but the CAGGS Nanog FH(C) ES cell line remained undifferentiated. Surprisingly, the CAGGS Nanog FH(N) ES cell line that overexpressed Nanog tagged at the N-terminus failed to self-renew in the absence of LIF. Therefore, C-terminally tagged Nanog was determined to be functional, whereas the N-terminally tagged Nanog construct was not. All subsequent experiments were carried out with the C-terminally tagged Nanog ES cells. These cells produced Nanog protein that was detectable by western blots using anti-Nanog and anti-HA antibodies (Figure 2b)

## Immunoprecipitation of Nanog and Associated Proteins from ES cell Nuclear Extracts

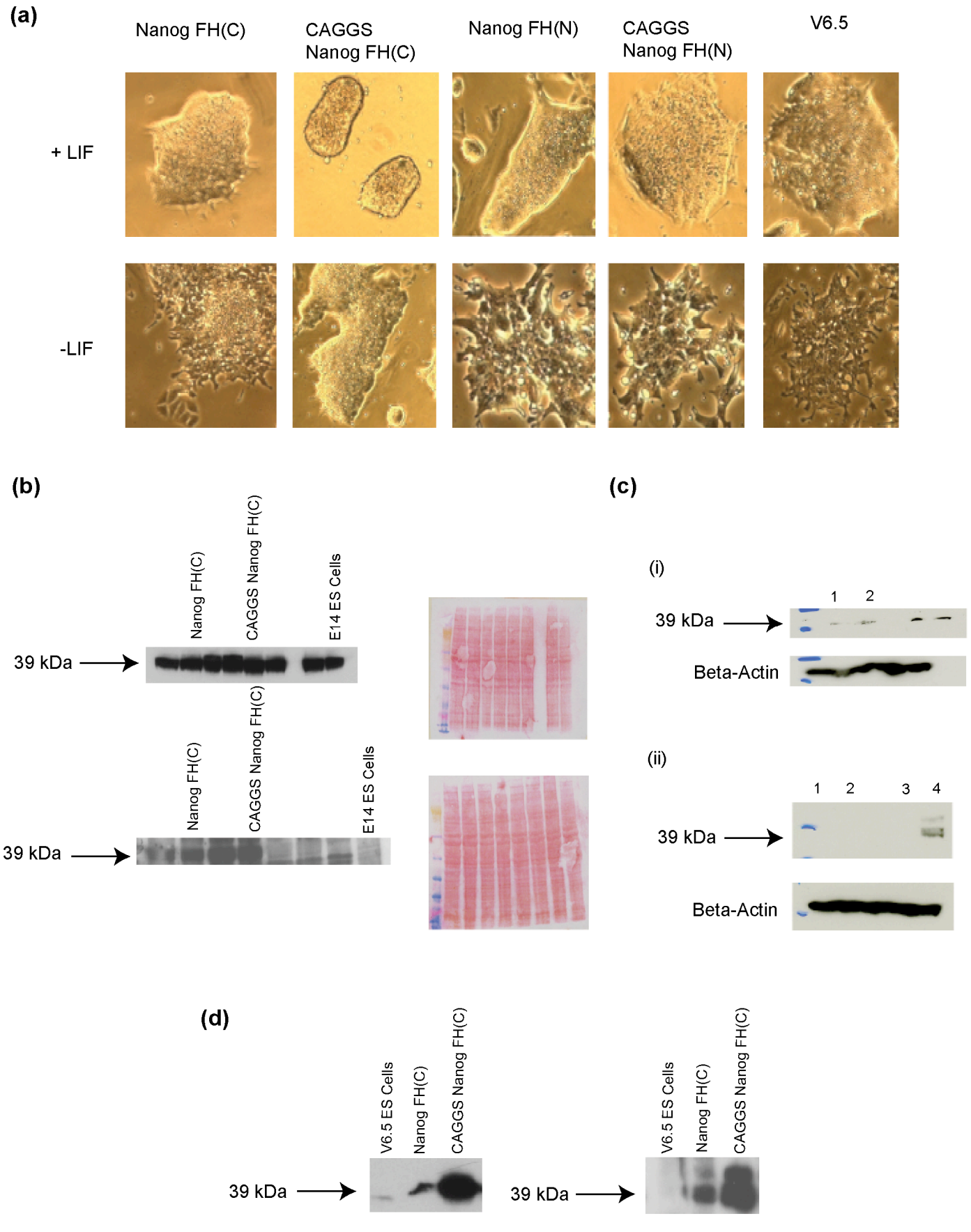
Nuclear and cytoplasmic extracts were made from the CAGGS Nanog FH(C) and Nanog FH(C) ES cell lines. Western blots of these fractions performed with an anti-HA antibody revealed that the tagged Nanog transcription factor, which is a

nuclear protein was being sequestered into the cytoplasm in the overexpression CAGGS Nanog FH(C) ES cell line (Figure 2c). Since such sequestering can lead to excessive background, we conducted the large-scale purifications with nuclear extracts from the Nanog FH(C) cell line that expressed Nanog at endogenous levels.

Before proceeding with large-scale purifications of Nanog, preliminary immunoprecipitations were performed with an anti-FLAG antibody on whole cell-extracts from the Nanog FH(C) and CAGGS Nanog FH(C) ES cell lines. Western blot analyses on the immunoprecipitated material with anti-HA and anti-Nanog antibodies confirmed that the tagged protein of approximately 39 kDa could be pulled down from these cells. No protein was detectable in similar immunoprecipitations performed with untagged V6.5 ES cells (Figure 2d). Large-scale anti-FLAG immunoprecipations were performed on approximately$10^9$ Nanog FH(C) ES cells and V6.5 control ES cells, and the eluates in each case were analyzed by mass spectrometry. Background proteins identified in the V6.5 ES cells were eliminated from the proteins identified in the Nanog FH(C) immunoprecipitated material. These analyses identified the proteins H2A.Z, Baf60b, Myb-binding protein, $\gamma$-catenin and eIF4a3 as ones that interacted with Nanog.

**Figure 2. Characterization of Tagged Nanog ES cells.** (a) LIF-independence tests to examine functionality of tagged Nanog constructs. Nanog FH(C), CAGGS Nanog FH(C), Nanog FH(N), CAGGS Nanog FH(N) and untagged V6.5 ES cells were grown in the presence (+) or absence (-) of LIF. (b) Western blots with anti-Nanog (top gel) and anti-HA (bottom gel) antibodies to show Nanog expression (39 kDa band) in Nanog FH(C), CAGGS Nanog FH(C) and untagged E14 ES cells. Ponceau stained membranes to check for equal protein loading are shown to the right of each gel. (c) Western blots with anti-HA antibody showing Nanog expression (39 kDa band) in (i) CAGGS Nanog FH(C) nuclear extract (lane 1) and cytoplasmic extract (lane 2), as well as in (ii) V6.5 ES cell nuclear extract (lane 1), cytoplasmic extract (lane 2) and Nanog FH(C) nuclear extract (lane 3) and cytoplasmic extract (lane 4). Beta-Actin loading controls are shown for each gel. (d) Western blots with anti-Nanog antibody (left) and anti-HA antibody (right) following anti-FLAG immunoprecipitations in untagged V6.5 ES cells, Nanog FH(C) and CAGGS Nanog FH(C) ES cells. Nanog protein is represented by a 39 kDa band.

**FiIGURE 2.**


(a)
Nanog FH(C)
CAGGS Nanog FH(C)
Nanog FH(N)
CAGGS Nanog FH(N)
V6.5
+ LIF
-LIF
(b)
Nanog FH(C)
CAGGS Nanog FH(C)
E14 ES Cells
39 kDa
Nanog FH(C)
CAGGS Nanog FH(C)
E14 ES Cells
39 kDa
(c)
(i)
1 2
39 kDa
Beta-Actin
(ii)
1 2 3 4
39 kDa
Beta-Actin
(d)
V6.5 ES Cells
Nanog FH(C)
CAGGS Nanog FH(C)
39 kDa
V6.5 ES Cells
Nanog FH(C)
CAGGS Nanog FH(C)
39 kDa

## DISCUSSION

This project aimed at identifying proteins interacting with Nanog in mouse ES cells, in order to gain further insight into the mechanism by which this transcription factor regulates the expression of its targets to maintain pluripotency. In order to address this issue, we employed a strategy that involved immunoprecipitation of Nanog and associated proteins from ES cells expressing FLAG-HA epitope tagged Nanog from the endogenous *nanog* locus. The proteins interacting with Nanog were identified by mass spectrometry. Using this approach, we identified five potential candidates that interact with Nanog—Histone variant H2A.Z, Myb-binding protein 1a (Mybb1a), γ-catenin, BAF60b and eIF4a3.

Although the results of these experiments must be reproduced and validated, some of the candidate proteins identified seem to have functions that may be interesting and relevant to pluripotency and early cell fate decisions. For instance, the histone H2A.Z, which is a variant of histone H2A, has been shown to be critical for early mammalian development and ES cell pluripotency (Faast, Thonglairoam et al. 2001). Interestingly, the *H2A.Z* knockout embryo phenocopies a *nanog* knockout embryo, since both die between days 3.5 and 4.5 postcoitum and differentiate into primitive endoderm. Similarly, ES cells deficient for H2A.Z or Nanog also differentiate into primitive endoderm. A role for H2A.Z has been suggested in transcriptional regulation, since it is essential in maintaining an active chromatin state and preventing the spread of

heterochromatin (Fan, Gordon et al. 2002). Therefore, it will be interesting to further validate the association between H2A.Z and Nanog in ES cells, and examine the relevance of this interaction with respect to pluripotency.

Another protein that was identified as a factor interacting with Nanog was Mybbp1a. Although this protein has been previously shown to regulate transcription by binding to the transcription factor, c-Myb (Tavner, Simpson et al. 1998), it was also identified as a novel repressor of NF-κB activity (Owen, Elser et al. 2007). Interestingly, in recent studies, Nanog was also shown to maintain ES cell pluripotency by inhibiting NF-κB activity (Torres and Watt 2008). Therefore, if the interaction between Mybbp1a and Nanog can be validated, it could lend further insight into the mechanism of transcriptional regulation by Nanog.

For the remaining factors that were pulled down with Nanog in our experiments, their roles in transcriptional regulation and ES cell pluripotency need to be further explored. γ-catenin has been shown to be share sequence similarity with the transcription factor β-catenin, but its role in activating the Wnt/TCF/Lef pathway is still unclear (Shimizu, Fukunaga et al. 2008). Recent studies indicate that Nanog shares a number of its downstream genomic targets in ES cells with those of Tcf3, a component of the Wnt signaling pathway (Cole, Johnstone et al. 2008). It would be of interest to examine the potential involvement of γ-catenin in mediating such transcriptional regulation. There is little known about the roles of

BAF-60b (Wang, Xue et al. 1996), a component of the Swi/Snf complex, and eIF4A3 (Chan, Dostie et al. 2004), a component of the exon-junction complex with respect to transcriptional regulation or maintenance of pluripotency.

In the future, the Nanog-interacting factors identified in these experiments should be reproduced in multiple immunoprecipitations, and validated by other methods such as co-immunoprecipitations or gel-shift assays. It would also be interesting to examine whether these proteins are bound to Nanog freely in the nucleus, or on chromatin, and whether they form one or more complexes with Nanog. The stoichiometry of the components of such a complex (or complexes) should be determined. Importantly, characterizing the relevance of these interactions to transcriptional regulation and the maintenance of pluripotency will provide a better understanding of the molecular mechanisms by which the transcription factor Nanog can regulate cell fate decisions.

## METHODS

### Generation of FLAG-HA Tagged Nanog constructs for Endogenous Expression and Overexpression

*(i) Cloning of Nanog FH(N) and Nanog FH(C) Targeting Constructs*

The initial cloning vector for both constructs was generated by subcloning a PGK-neomycin cassette flanked by 2 loxP sites into a pSP72 vector with *BamH*I and *Xho*I. For both targeting vectors, the 5' arm, the tagged *nanog* cDNA, and an RBGpA sequence were inserted between this *BamH*I site, and a *Cla*I site in

the pSP72 vector. The 5' arms of both vectors spanned 1188 bp of the *nanog* promoter and were flanked by a *Cla*I site at the 5' end and a *BamH*I site at the 3' end. These restriction sites were added to the PCR primers that were used for amplifying the 5' arm from a BAC containing this region of the *nanog* promoter. The primers for the 5'arm PCR were: Forward (ATCGATCTGGGTTAGAGTGCT TTCACTCAC) and Reverse (GGATCCGTCAGTGTGATGGCGAGGGAAGGG). The *nanog* cDNA tagged with FLAG and HA sequences at either the N- or C-terminus was cloned downstream of the 5'arm. These cDNAs were flanked by a *BamH*I site at the 5'end and an *Xba*I site at the 3' end. These restriction sites, as well as the FLAG and HA epitope tags were added to the PCR primers used for amplification of the cDNAs. For the Nanog FH(N) cDNA, the PCR primers were: Forward (GGATCCGCTCGATGAACCATGGACTACAAGGACGACGATGACAA GCTCGATGGAGGATACCCCTACGACGTGCCCGACTACGCCAGTGTGGGTC TTCCTGGTCCCCACAGTTTGCC) and Reverse (TCTAGATCATATTTCACCTG GTGGAGTCAC), and for the Nanog FH(C) cDNA, the primers used were: Forward (GGATCCATGAGTGTGGGTCTTCCTGGTCCC) and Reverse (TCTAG ACTAGGCGTAGTCGGGCACGTCGTAGGGGTATCCTCCAGCGGCCGACTTG TCATCGTCGTCCTTGTAGTCTATTTCACCTGGTGGAGTCACAGAGTAGTTCA GG). The FLAG and HA sequences used for cloning have been described earlier (Nakatani and Ogryzko 2003). For the N-terminal tag, a Kozak sequence was also included as suggested in this paper. For both Nanog FH(N) and Nanog FH(C), an RBGpA sequence was cloned downstream of the tagged cDNAs using an *Xba*I site at the 3'end of the cDNAs and a *BamH*I site at the 5'end of the

PGKneo2lox fragment. The 3' arm for both targeting vectors was amplified from a BAC, and spanned 1.4 kb of *nanog* intron 1. The PCR primers used for the 3' arm were: Forward (GCGGCCGCGTAAGGAATTCAGTCCCCGAA) and Reverse (GCGGCCGCCTCGAGGCCCTCTTCTGGAGTGTCTGAAGAC). For Nanog FH(N), the 3'arm was cloned into the targeting vector with a *Not*I site at the 5' end and *Xho*I site at the 3' end. In the case of Nanog FH(C), this arm was cloned into the vector using the *Xho*I restriction site.

*(ii) Cloning of CAGGS Nanog FH(N) and CAGGS Nanog FH(C) Flp-in Constructs*

In order to obtain the CAGGS Nanog FH(N) construct, the N-terminally tagged *nanog* cDNA that had been PCR-amplified and TOPO cloned into pCR2.1, was subloned into the *EcoR*1 site of the vector PGK(ATG)frt-CAGGS-RBGpA through blunt end ligation. The C-terminally tagged cDNA was isolated from an *Xba*I/*BamH*i/*Bgl*II digest of Nanog FH(C), and gel-purifying a 990 bp fragment that represented the cDNA. This cDNA was also cloned into the *EcoR*I site of PGK(ATG)frt-CAGGS-RBGpA through blunt end ligation, in order to get the CAGGS Nanog FH(C) construct.

**Targeting of FLAG-HA Tagged Nanog Constructs to ES Cells**

*(i) Targeting of Nanog FH(N) and Nanog FH(C)*

25 μg each of Nanog FH(N) and Nanog FH(C) targeting vectors were linearized with *Pvu*I in 500 μl reactions. The digested DNA was ethanol precipitated, resuspended in 50 μl of T.E.buffer (10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA) and

electroporated into $10^6$ v6.5 ES cells, by zapping twice at 400V and 25 μF. These electroporated ES cells were selected with neomycin (350 μg/ml) for 10 days, starting at 24 hours after electroporation. Neomycin resistant clones were picked and expanded for screening with Southern blot analysis.

*(ii) Targeting of CAGGS Nanog FH(N) and CAGGS Nanog FH(C)*

50 μg each of CAGGS Nanog FH(N) and CAGGS Nanog FH(C) were ethanol precipitated along with 25 μg of the pCAGGS flpE recombinase plasmid. The precipitated DNA in each case was resuspended in 400 μl of HEPES buffer and electroporated into $10^6$ C10 ES cells (Beard, Hochedlinger et al. 2006) by zapping twice at 500V and 25 μF. The electroporated ES cells were selected with hygromycin (140 μg/ml) for 13 days, starting at 24 hours after electroporation. Drug-resistant clones were picked and expanded for screening by Southern blots.

## Southern Blot Analyses to Screen for Correctly Targeted FLAG-HA Tagged Nanog Constructs

*(i) Southern Blot Analyses for Nanog FH(N) and Nanog FH(C) Targeted ES cells*

In order to screen for ES cells targeted with the Nanog FH(N) and Nanog FH(C) constructs, DNA from neomycin-resistant ES cell clones was digested with *Pvu*II and run on a 0.8% agarose gel. For examining homologous recombination at the 3' end, a 415 bp external probe was PCR-amplified from *nanog* intron 1 using the primers: Forward (GCTACCTGAGACCCTATCCCTTAG) and Reverse

(CATCTCACCAG CCCTACATACAGTG). This probe identified a 6.4 kb band representing the wild-type *nanog* allele, and a 4.7 kb band representing a targeted allele (Nanog FH(N) or Nanog FH(C)). In order to screen for 5'end homologous recombination, an internal probe was isolated by purifying an approximately 400 bp fragment from a *Pvu*II/*Xba*I digest of the Nanog FH(C) targeting vector. This probe identified a 6.4 kb band representing the wild type *nanog* allele and a 2.7 kb band representing the targeted allele (Nanog FH(N) or Nanog FH(C)).

*(ii) Southern Blot Analyses for CAGGS Nanog FH(N) and CAGGS Nanog FH(C) Targeted ES cells*

DNA from ES cell clones targeted with CAGGS Nanog FH(N) and CAGGS Nanog FH(C) were digested with *Spe*I. In order to generate a 3' internal probe, the Col1ABSKS plasmid was digested with *Pst*I and *Xba*I, and an 850 bp fragment was gel-purified. This probe recognized a 6.2 kb band from the wild type *collagen1a1* allele, a 6.7 kb band of the frtpgkneo allele, and a 4.1 kb band of the targeted allele. Southern blots were carried out as described in Chapter 3.

**LIF-Independence Functionality Tests for Tagged Nanog constructs**

ES cells that were correctly targeted with the four tagged Nanog constructs, as well as untagged V6.5 ES cells were cultured in duplicate 6-well dishes in ES media containing LIF. These cells had been passaged once without MEFs in order to avoid LIF produced from the feeder layer. After 48 hours, LIF was

withdrawn from the media in half of the wells and the morphology of ES cells was examined after another 48 hours.

**Preparation of ES Cell Whole-cell, Nuclear and Cytoplasmic Extracts**

*(i) Preparation of ES Cell Whole Cell Extracts for Small Scale Protein Purifications*

ES cells were grown on one gelatin coated, 15 $cm^2$ dish until they were approximately 80% confluent. The cells were grown without MEFs for at least one passage to avoid feeder contamination. The cells were washed twice with cold 1X PBS, pH 7.4 and 500 μl of 1X Cell Lysis Buffer (Cell Signaling Technology) was added to the ES cells. 1 mM PMSF and a protease inhibitor cocktail (Roche) had been added to the lysis buffer immediately before use. The plate was incubated on ice for 5 minutes and cells were scraped using a polyethylene cell lifter (Corning), and transferred into a 1.5 ml microcentrifuge tube. The extract was incubated on a rocker at 4°C, and spun for 10 minutes at 14000 RPM for another 10 minutes in a cold microcentrifuge. The supernatant was removed and either used immediately for further purifications, or flash frozen in liquid $N_2$ and stored at -80°C. Bradford assays were used to determine the protein concentration.

*Preparation of Nuclear and Cytoplasmic Extracts for Large-Scale Protein Purifications*

ES cells were grown on 25 gelatin coated, 15 $cm^2$ dish until they were approximately 80% confluent. The cells were grown without MEFs for at least one passage to avoid feeder contamination. After this step, all procedures were carried out in the cold room at 4°C. The ES cells were washed twice with cold 1X PBS, pH 7.4. 5 ml of PBS were added to each plate, the cells were scraped off with a polyethylene cell lifter (Corning), and transferred into a 250 ml centrifuge tube (Corning). The cells were spun at 3000 RPM for 5 minutes, and the residual PBS was discarded. The packed cell volume (PCV) was measured at this time and the extraction protocol described below was carried out. The cell pellets could also be flash frozen in liquid $N_2$ at this point and stored at -80°C.

Buffers for Nuclear and Cytoplasmic Extract Protocol

*High Salt Buffer*

20 mM HEPES, pH 7.9, stored at 4°C
25% glycerol
1.5 mM $MgCl_2$
1.2 M KCl
0.2 mM EDTA
0.5 mM PMSF (added dropwise before use)
1 mM DTT (added before use)

*Hypotonic Buffer*

10 mM HEPES, pH 7.9, stored at 4°C
1.5 mM $MgCl_2$
10mM KCl
0.5 mM PMSF (added dropwise before use)
1 mM DTT (added before use)

*Low Salt Buffer*

20 mM HEPES, pH 7.9, stored at 4°C
25% glycerol
1.5 mM $MgCl_2$
0.02 M KCl
0.2 mM EDTA
0,5 mM PMSF (added dropwise before use)
1 mM DTT (added before use)

Isolation of Nuclei

1. Resuspend the cell pellet in a 50 ml Falcon tube in 4X PCV of 1X PBS, pH 7.4

2. Centrifuge the cells for 10 minutes at 3000 RPM using a JS4.2 rotor, and remove the supernatant gently with a pipette. Note: A turbid supernatant indicates cell lysis.

3. Resuspend the cells rapidly in 2-3X PCV of hypotonic buffer and quickly vortex for less than 10 seconds.

4. Centrifuge the cells for 5 minutes at 3000 RPM, and discard the supernatant. Note: This step should also be carried out rapidly since proteins can leak out of the cells. This step removes the salt in the PBS in order for efficient swelling to occur in the next step. Some swelling is also seen at this stage.

5. Resuspend the cells in 3X original PCV of hypotonic buffer and allow them to swell on ice for 10 minutes.

6. Transfer the cells to a glass Dounce homogenizer (15 ml capacity), and homogenize the cells with 25 strokes of a type B pestle. Note: The pestle should be moved up and down slowly to avoid air bubbles from forming.

7. Check cells for lysis using Trypan Blue. Mix 2.5 μl of cells with 20 μl of Trypan Blue dye, and place them on a slide. Cover the sample with a coverslip

and observe it under a microscope. The nuclei of lysed cells stain blue or purple. Approximately 90% lysis should be seen after 25 strokes.

8. Transfer the cells to a 50 ml Falcon tube and centrifuge for 15 minutes at 4000 RPM in a JS4.2 rotor.

9. Remove the supernatant, which is the cytoplasmic extract. This extract can be flash frozen in liquid $N_2$ and stored at -80°C.

Extraction of Nuclei

10. Measure the packed nuclear volume (PNV) using the graduations on the Falcon tube. Resuspend the nuclei in 1/2 PNV of low salt buffer by vortexing briefly.

11. Hold the tube on a vortex with one hand, and while gently vortexing, slowly and dropwise, add 1/2 PNV of high salt buffer. Note: Continuous mixing and dropwise addition of high salt buffer is necessary to prevent nuclei from clumping together.

12. Use a rocking platform to gently mix the buffers and nuclei. Let the extraction continue for 30 minutes. Note: The extract should seem non-homogenously viscous (mucus-like).

13. Pellet the nuclei by centrifuging for 30 minutes in a Beckman JA-20 rotor at 14500 RPM. Remove the supernatant, which is the nuclear extract. This can be flash frozen in liquid $N_2$ and stored at -80°C, or used immediately for further purification. The nuclei in the pellet can also be similarly saved to obtain chromatin extracts.

**Anti-FLAG Immunoprecipitation of Nanog**

Note: All immunoprecipitation reactions were carried out at 4°C.

*Small Scale Anti-FLAG Purification of Nanog from Tagged ES cells*

1. Aliquot 40-50 µl of FLAG antibody beads slurry (Sigma) per sample.

2. Wash the slurry with twice the amount of 1X cold PBS and spin at 5000 RPM for 1 min.

3. Repeat the wash after removing most of the PBS and adding more cold 1X PBS.

4. Remove the PBS and add an equivalent amount of cold 1X PBS to the slurry (i.e., if 250 µl of slurry is being used, then add 250 µl of PBS).

5. In fresh microcentrifuge tubes, add 80 µl of slurry for each sample.

6. Measure the protein concentration with a Bradford assay and aliquot the desired amount of lysate to mix with the slurry.

7. Incubate the immunoprecipitation reactions at 4°C for 2 hours.

8. Prepare Elution Buffer: HEPES 50 mM (pH 7.4), 100 mM NaCl. For protein elution add 490 µl of elution buffer + 10 µl of 50X FLAG peptide (Sigma).

9. Spin immunoprecipitation reactions for 1 min at 5000 RPM. Wait for a minute for the beads to settle down and then remove most of the buffer,

10. Add 800 µl of 1% Triton buffer (wash buffer) and mix,

11. Spin at 5000 RPM for 1 min. Wait for beads to settle, remove the buffer and repeat the wash twice more with wash buffer.

12. Repeat the wash with 800 µl of elution buffer (without the FLAG peptide). Remove all the liquid with a 25 G needle.

13. Add 70 µl of elution buffer (with the FLAG peptide). Use a tip cut off at the end so that beads can be mixed well by pipetting.

14. Incubate at room temperature for 10 minutes.

15. Add 15 µl of 5X protein sample buffer (NuPage) to 60 µl of eluate, boil for 5 minutes and store at -80°C or use for further analyses.

16. Add 20 µl of sample buffer to the beads, boil them for 5 minutes and store in the freezer or use for further analyses.

*Large-Scale Anti-FLAG Immuno-affinity Purification of Tagged Nanog from ES Cells*

Reagents and Buffers for FLAG-IP Protocol

*2X BC0*

40% glycerol
40 mM HEPES, pH 7.9
0.4 mM EDTA
1 mM DTT
0.4 mM PMSF

*BC100*

50 ml 2X BC0
46 ml $dH_2O$
4 ml 2.5 M KCl
100 µl 200 mM PMSF
50 µl 1 M DTT

*BC150*

50 ml 2X BC0
44 ml $dH_2O$

6 ml 2.5 M KCl
100 µl 200 mM PMSF
50 µL 1 M DTT

*BC300*

50 ml 2X BC0
38 ml $dH_2O$
12 ml 2.5 M KCl
100 µl 200 mM PMSF
50 µl 1 M DTT

*BC500*

50 ml 2X BC0
30 ml $dH_2O$
20 ml 2.5 M KCl
100 µl 200 mM PMSF
50 µl 1 M DTT

*200 X TLCK*

10 mg/ml in 1 mM HCl

Add protease inhibitor cocktail (Roche) to all solutions immediately before use.

Immunoprecipitation Protocol

1. To immunoprecipitate 3 samples, wash 1.5 ml (i.e. 3 ml of slurry) of FLAG M2 resin (Sigma) on a BioRad Polyprep chromatography column 3 times with BC150 (add 7.5 ml of BC150 to the resin, spin at 1000 RPM for 1 min and remove supernatant).

2. Incubate with 7.5 ml of 5X Denhardt's Solution (USB) for 30 minutes.

3. Wash the resin twice more with 7.5 ml BC150.

4. Bring the total volume up to 4 ml and aliquot 1 ml in each of three 15 ml Falcon tubes (leave 1 ml spare in case of pipetting errors).

5. Add 5 ml of BC150 to each tube and wash the resin one more time.

6. Add fortified nuclear extracts (including 0.2 mM DTT, 1X TLCK and protease inhibitor cocktail (Roche)) to the washed resin and incubate on a rocker overnight.

7. Pour the resin back into a polyprep column gently, by pouring along the edge of the column (do not pipette, since this may damage the resin).

8. Collect the flow through, and towards the end add 1 ml of BC150 to the tube and pour it into the column. Note: Do not let the column dry out.

9. Add 250 µl of BC150 and let it soak in. Repeat again with another 250 µl of BC150. Add 5 ml of BC150 and let it flow through the column.

Repeat the previous step with 500µl + 5 ml of BC300, BC500, BC300 and BC100, in that order.

10. Elute the protein by using a 20:1 molar ratio of peptide to antibody (prepare a stock solution of 2.5 mg of 3X FLAG peptide (Sigma) in 4.7 ml of BC100). Add 150 µl of peptide (in BC100) to the column and incubate for 80 min by stopping the column. Layer on another 150 µl of peptide solution, and collect the flow through.

11. Repeat the previous step four more times.

12. Combine the eluates from the last two steps and concentrate the volume to 100 µl using protein concentration columns (Vivaspin2, 5000 MWCO, PES). The eluates can be frozen at -80°C or sent for Mass spectrometry analysis for identification of proteins present in the samples.

## Western Blot Analyses

100 µg of protein samples were run on 4-12% gradient NuPAGE (Novex) gels at 200V for 50 minutes in MOPS buffer. The samples had been prepared with NuPAGE sample buffer and 1/0$^{th}$ volume of DTT (according to specifications by NuPAGE). The proteins were transferred to a nitrocellulose membrane at 350 mA for 45 minutes at 4°C in transfer buffer (12 mM Tris. 96 mM glycine, 20% methanol, pH 8.3). The membrane was blocked in PBS+1% non-fat milk powder for 15 minutes at room temperature. It was incubated with primary antibody diluted in PBS + 1% non-fat milk powder + 0.1% Tween-20 for 90 minutes at room temperature. The membrane was washed 3 times for 10 minutes with PBS+ 0.1% Tween-20. It was incubated with the secondary antibody diluted in PBS+ 1% non-fat milk powder +0.1% Tween-20 for 1 hour at room temperature. The membrane was washed 3 times for 10 minutes with PBS + 0.1% Tween-20 and developed using ECL reagent.

## Antibodies

For western blots, the primary antibodies used were anti-Nanog antibody (BL1162, Bethyl) anti-HA-HRP (3F10, Roche), anti-HA (Covance), anti-β-actin (Abcam 8226). HRP-conjugated rabbit and mouse secondary antibodies were used from Santa Cruz. For FLAG immunoprecipitations, anti-FLAG M2 affinity gel (Sigma) was used.

**ACKNOWLEDGEMENTS**

I would like to thank Caroline Beard for help in designing the tagging strategy and advice on this project. I also appreciate the guidance from members in Bob Kingston's lab at Harvard Medical School on the nuclear extraction and protein purification protocols.

## REFERENCES


Beard, C., K. Hochedlinger, et al. (2006). "Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells." Genesis **44**(1): 23-8.

Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-55.

Chan, C. C., J. Dostie, et al. (2004). "eIF4A3 is a novel component of the exon junction complex." Rna **10**(2): 200-9.

Cole, M. F., S. E. Johnstone, et al. (2008). "Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells." Genes Dev **22**(6): 746-55.

Faast, R., V. Thonglairoam, et al. (2001). "Histone variant H2A.Z is required for early mammalian development." Curr Biol **11**(15): 1183-7.

Fan, J. Y., F. Gordon, et al. (2002). "The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states." Nat Struct Biol **9**(3): 172-6.

Fujikura, J., E. Yamato, et al. (2002). "Differentiation of embryonic stem cells is induced by GATA factors." Genes Dev **16**(7): 784-9.

Hart, A. H., L. Hartley, et al. (2004). "Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human." Dev Dyn **230**(1): 187-98.

Matsuda, T., T. Nakamura, et al. (1999). "STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells." Embo J **18**(15): 4261-9.

Mitsui, K., Y. Tokuzawa, et al. (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-42.

Nakatani, Y. and V. Ogryzko (2003). "Immunoaffinity purification of mammalian protein complexes." Methods Enzymol **370**: 430-44.

Niwa, H., T. Burdon, et al. (1998). "Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3." Genes Dev **12**(13): 2048-60.

Owen, H. R., M. Elser, et al. (2007). "MYBBP1a is a novel repressor of NF-kappaB." J Mol Biol **366**(3): 725-36.

Pan, G. and D. Pei (2005). "The stem cell pluripotency factor NANOG activates transcription with two unusually potent subdomains at its C terminus." J Biol Chem **280**(2): 1401-7.

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

Shimizu, M., Y. Fukunaga, et al. (2008). "Defining the roles of beta-catenin and plakoglobin in LEF/T-cell factor-dependent transcription using beta-catenin/plakoglobin-null F9 cells." Mol Cell Biol **28**(2): 825-35.

Tavner, F. J., R. Simpson, et al. (1998). "Molecular cloning reveals that the p160 Myb-binding protein is a novel, predominantly nucleolar protein which may play a role in transactivation by Myb." Mol Cell Biol **18**(2): 989-1002.

Torres, J. and F. M. Watt (2008). "Nanog maintains pluripotency of mouse embryonic stem cells by inhibiting NFkappaB and cooperating with Stat3." Nat Cell Biol **10**(2): 194-201.

Wang, S. H., M. S. Tsai, et al. (2003). "A novel NK-type homeobox gene, ENK (early embryo specific NK), preferentially expressed in embryonic stem cells." Gene Expr Patterns **3**(1): 99-103.

Wang, W., Y. Xue, et al. (1996). "Diversity and specialization of mammalian SWI/SNF complexes." Genes Dev **10**(17): 2117-30.

Yamaguchi, S., H. Kimura, et al. (2005). "Nanog expression in mouse germ cell development." Gene Expr Patterns **5**(5): 639-46.