Programming and Reprogramming Cellular Identity

By

Alexander Marson

A.B. Biology
Harvard College, 2001

M.Phil.
University of Cambridge, 2003

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

DOCTOR OF PHILOSOPHY
at the
MASSACHUSSETTS INSTITUTE OF TECHNOLOGY

June 2008]
MAY 2008

Signature of Author: _____
Department of Biology
April 11, 2008

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

Certified by: _____
Rudolf Jaenisch
Professor of Biology
Thesis Supervisor

Accepted by: _____
Stephen P. Bell
Professor of Biology
Chairman, Committee for Graduate Students

Acknowledgments

Let me start by acknowledging my loving parents, Ellen and Bernie, as well as my sister and confidant, Eve. Without their support, none of this would have been possible. I also owe my sanity, such as it is, to my wonderful friends, especially the crowd that has spent time in Cambridge over the years.

This work could not have been accomplished without close collaborations that I was lucky to establish, notably with Karsten Kretschmer, Harald von Boehmer and Garrett Frampton on Foxp3 studies, with Stuart Levine on miRNA gene regulation studies and with Ruth Foreman and Brett Chevalier on studies of Wnt signaling and reprogramming. Brett Chevalier has also been a tremendous help and great sounding board as I prepared this thesis. Many thanks to all of these collaborators.

I am grateful to Nancy Hannett, who was tolerant of my occasional oversights during my years in the lab, and who did the bulk of the work to clone a lentiviral library of murine microRNAs.

Thanks to everyone who I got a chance to work with and learn from at MIT and the Whitehead, especially all members of the Young and Jaenisch labs. Special thanks to current and former inhabitants of Room 513. Also, I must acknowledge Tony Lee, who has been an amazing teacher and friend over the past four years.

In addition, my graduate school experience has been enriched by scientific interactions with other labs including those of Harald von Boehmer, Peter Sicinski, Tom Look and David Hafler.

I am appreciative of the members of my thesis committee, David Bartel and David Housman, who have encouraged me and pushed me over the past few years. Thanks also to Cliff Tabin, who has generously agreed to serve as an outside member of this committee.

And, of course, profound thanks to Rick and Rudolf, who were enthusiastic cheerleaders, critics and teachers at all the right moments-- exactly what one hopes for in mentors.

Programming and Reprogramming Cellular Identity

by

Alexander Marson

Submitted to the Department of Biology
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy in Biology

Abstract

Every cell in the human body contains the same genetic information, with few exceptions, yet each cell type enacts a distinct gene expression program to allow for highly specialized functions. These tightly controlled programs are the results of transcriptional regulation, by transcription factors and chromatin regulators, as well as post-transcriptional regulation, mediated in part by microRNAs (miRNAs). Additionally, cells must respond to external cues, and signal transduction pathways converge on gene regulatory machinery to shape cellular identity. The work presented here focuses on the mechanisms by which transcription factors, chromatin regulators, miRNAs and signal transduction pathways coordinately regulate two particular medically important gene expression programs: (1) the program that controls pluripotency in embryonic stem (ES) cells, giving these cells the capacity to differentiate into every adult cell type, and (2) the program that allows regulatory T (T_{reg}) cells to prevent autoimmunity by suppressing the response of self-reactive conventional T cells. Genomic investigations of the core regulatory circuitry of each of these cells types presented here provide new insight into the genetics of pluripotency and autoimmunity, and suggest a strategy for reprogramming based on chemical manipulation of the cellular programs that control cell identity.

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

Table of Contents

Chapter 1

Introduction

A single fertilized egg can give rise to all of the diverse, specialized cell types that exist within an adult mammalian body. Because each adult cell type, with few exceptions, contains the same genetic information as the fertilized egg from which it derives, regulatory factors are crucial to determine the subset of genes that is actively expressed in any particular cell type. Examinations of the regulatory mechanisms that establish and maintain distinct cell types reveal the molecular "programs" that underlie specialized cellular functions. Aberrations in normal cellular programming likely lead to a subset of common human diseases; understanding cellular circuitry could help elucidate molecular underpinnings of disease. Additionally, "reprogramming" of cellular identity can now be accomplished by the introduction of defined regulatory factors into cells, converting them from one cell type to another.

The work presented here focuses on the gene expression programs of two cell types: (1) the program that controls pluripotency in embryonic stem (ES) cells, establishing and maintaining their capacity to differentiate into every adult cell type, and (2) the program that allows regulatory T (T_{reg}) cells to prevent autoimmunity by suppressing the response of self-reactive conventional T cells. For both of these cell types, master regulatory factors have been identified, allowing for the reprogramming of somatic cells to pluripotency, and the reprogramming of conventional T cells to suppressive T_{reg} cells. Genomic investigations of the core circuitry of each of these cells types presented here provide new insight into the genetics of pluripotency and

autoimmunity, and suggest a strategy for reprogramming based on chemical manipulation of the cellular programs that control cell identity.

Molecular control of cellular identity

A key challenge in studies of cellular identity is defining the specific factors that are responsible for giving a cell its specialized attributes. The landmark work of Davis et al. (1987) established that MyoD1, which is selectively expressed in myogenic cells, is sufficient to convert mouse fibroblasts to muscle-like cells when ectopically expressed. Consistent with its role in shaping cellular identity, MyoD1 is a transcription factor that binds to DNA and directly regulates gene expression. The role of MyoD1 in the developmental programming of muscle cells was clarified by genetic studies that demonstrated that, although MyoD1 is redundant with another transcription factor, Myf-5, the function of these two factors is required for the development of skeletal muscle in the mouse (Rudnicki et al., 1993). *In vivo* and *in vitro* genetic studies of these factors revealed key features of how a molecular program is able to control mammalian cell specification. The factors can autoregulate their own expression levels, regulate the expression levels of each other, and contribute to the regulation of common target genes (Weintraub, 1993). More recent examinations of the occupancy of genome by MyoD1 have revealed the specific DNA binding events that underlie this program (Blais et al., 2005; Cao et al., 2006).

Work in the hematopoietic system has further demonstrated that transcription factors play crucial roles in cell-fate determination (Pevny et al., 1991; Orkin and Zon, 2008). Genetic studies have identified a core set of transcription factors that are required for maintenance of multipotent hematopoietic stem cells. Distinct transcription factors, which are selectively expressed in cells at various stages of commitment in the myeloid and lymphoid lineages, are required for lineage commitment and proper development of mature hematopoietic cells (Orkin and Zon, 2008). Remarkably, normal cell-fate determinations can be reprogrammed by the ectopic expression of transcription factors from other hematopoietic lineages, underscoring the power of transcriptional regulation in shaping cellular identity (Xie et al., 2004; Orkin and Zon, 2008).

While transcription factors are crucial regulators of cellular identity, additional layers of regulation also contribute. Three other important modes of regulation, which are discussed below and in subsequent chapters, are microRNAs (miRNAs), extra-cellular signals and epigenetic modifications to histones and DNA. Transcription factors, miRNAs, signal transduction pathways and epigenetic chromatin marks coordinately control cellular identity (**Figure 1**).

miRNAs were first identified in *Caenorhabditis elegans* as genes that play key roles in the timing of cell-fate determinations (Chalfie et al., 1981; Lee et al., 1993; Wightman et al., 1993; Reinhart et al., 2000). These short, non-coding transcripts are an abundant class of transcripts in multi-cellular organisms and play important roles in development and cellular specification by specifying post-transcriptional repression of target mRNAs (Bartel, 2004; Ambros, 2004; Plasterk, 2006). In plants, the majority of

miRNA targets are developmental transcription factors (Rhoades et al., 2002), and, although the targets are more diverse in mammalian cells, miRNAs appear to have a pronounced regulatory effect on mRNAs involved in transcriptional control and development (Stark et al., 2005).

Extra-cellular cues and intra-cellular signal transduction pathways enable the local environment of a cell to contribute to its identity (Shaywitz and Melton, 2005). The terminal components of these signaling pathways often act as transcriptional regulators and help to reshape the gene expression program of a cell in response to the environment.

Epigenetic mechanisms, notably covalent modifications of histones and DNA bases, contribute to the stability of gene transcription required for cells to maintain their identities over time and even through cell-division. While cellular identity is shaped with inputs from extra-cellular signals, once cellular identity is established it tends to be robust to perturbations from the environment (Hadorn, 1968). Chromatin regulators and the epigenetic marks deposited on chromatin help to maintain this cellular memory. Modifications at key loci alter accessibility of chromatin to transcriptional machinery, provide novel binding sites for transcriptional regulators, or occlude existing binding sites for transcriptional regulators (Bird, 2002; Goll and Bestor, 2005; Margueron et al., 2005; Bernstein et al., 2007). Epigenetic marks on chromatin, which are mitotically heritable, are able to exert long-term effects on the transcriptional program of a cell and its progeny.

The current model of the molecular circuitry for pluripotency serves well to illustrate how a regulatory network involving signal transduction pathways, transcription

factors, post-transcriptional regulation by miRNAs and epigenetic modifications to chromatin and can control cellular identity. Over the past several years, much progress has been made on elucidating the molecular circuitry of pluripotent ES cells and, now, somatic cells can now be reprogrammed to pluripotency with defined factors. Known features of the regulatory circuitry of ES cells are highlighted below. This discussion is an introduction to studies presented in Chapter 2 and Chapter 3, which build on what was previously known about the molecular circuitry of pluripotency. In contrast, prior to the studies described in Chapter 4, little was known about the regulatory circuitry of immuno-suppressive T_{reg} cells. A discussion of the role of Foxp3, a key transcription factor in T_{reg} cells, as well as the roles of miRNAs and T-cell receptor signaling in shaping T_{reg} cell identity, introduces the global studies of Foxp3 transcriptional regulation described Chapter 4.

ES cells and pluripotency

ES cells are distinctive for their potential to both self-renew and differentiate into any adult cell type. These cells are derived *in vitro* from the inner cell mass (ICM) of blastocysts (Evans and Kaufman, 1981; Martin, 1981; Thomson et al., 1998). Under proper conditions, these cells can be maintained over prolonged periods in culture, retaining their pluripotency, as evidenced by the ability of murine ES cells to develop into to all embryonic and adult cell types in chimaeric mice, including germline cells.

Genetic studies have identified core transcriptional regulators of the ES pluripotent state, including Oct4, Sox2 and Nanog, which are required for the maintenance of pluripotency (Schöler, 1990; Nichols, 1998; Avilion, 2003; Chambers, 2003; Mitsui, 2003). These transcription factors are all expressed in pluripotent ES cells, as well as the inner cell mass of the blastocyst from which ES cells are derived, and *Oct4* and *Nanog* are rapidly silenced during early cellular differentiation. More recently, a family of Klf transcription factors, including Klf2, Klf4 and Klf5, was also shown to be required for pluripotency (Jiang et al., 2008).

The gene expression program underlying pluripotency can be restored to the genome of somatic cells either by nuclear transfer into recipient oocyte cytoplasm (Campbell et al., 1996; Wakayama et al., 1998; Wakayama et al., 2001;Hochedlinger and Jaenisch, 2002; Eggan et al., 2004; Li et al., 2004; Inoue et al., 2005) or cell-fusion with ES-cells (Tada et al., 2001; Cowan et al., 2005; Yu et al., 2006), especially ES cells over-expressing the transcription factor Nanog (Silva et al., 2006). Additionally, in what appears to be a special case, spermatogonial germ cells can acquire pluripotency under certain culture conditions (Kanatsu-Shinohara et al., 2004). These findings indicate that cellular, and perhaps extra-cellular factors, are sufficient to induce pluripotency in differentiated cells. Much effort went into defining the master regulatory factors that are sufficient to establish the gene expression program for pluripotency. The work in myogenic and hematopoietic cells, discussed above, suggested that strong candidates include transcription factors that are selectively expressed in pluripotent cells and contribute to the maintenance pluripotency.

Indeed, a defined set of ES cell transcription factors is now known to induce pluripotency in somatic cells (Takahashi and Yamanaka, 2006). Retroviral transduction of three nuclear factors, Oct4, Sox2 and Klf4 is sufficient to reprogram mouse and human somatic cells into induced pluripotent stem (iPS) cells that are indistinguishable from embryo derived embryonic stem (ES) cells (Takahashi and Yamanaka, 2006; Okita et al., 2007; Wernig et al., 2007; Maherali et al., 2007; Takahashi et al., 2007; Park et al., 2007; Nakagawa et al., 2008; Wernig et al., 2008). c-Myc, a powerful oncogene, is not required for the induction of pluripotency, but significantly increases the kinetics and efficiency of developmental reprogramming (Takahashi and Yamanaka, 2006; Nakagawa et al., 2008; Wernig et al., 2008). Recent experiments have demonstrated that other combinations of factors also can induce iPS cells (Yu et al., 2007; Nakagawa et al., 2008). Identification of these factors, coupled with studies of their occupancy of and effects on the genome, reveals key features of the molecular circuitry that establishes and maintains pluripotency.

The molecular circuitry of pluripotency

Genome-wide location analysis has provided insight into the molecular mechanisms by which transcription factors maintain pluripotency in human and murine ES cells (Boyer et al., 2005; Loh et al., 2006; Jiang et al., 2008; Kim et al., 2008). These experiments have yielded three key findings: (1) Oct4, Sox2 and Nanog form an interconnected autoregulatory loop, (2) all three factors often co-occupy the same sites in the promoters of their target genes, and finally, (3) Oct4, Sox2 and Nanog collectively

target two sets of genes, one that is actively expressed in ES cells and another that is silent in ES cells, but remains poised for subsequent expression during cellular differentiation (Boyer et al., 2005) (**Figure 2**). In particular, these studies reveal key features of the genetic logic of pluripotency, and more broadly, provide clues as to how multiple transcription regulators can coordinately control cell identity.

Interconnected autoregulation

Oct4, Sox2 and Nanog all bind to their own promoters, as well the promoters of the genes encoding the two other factors (Boyer et al., 2005) (**Figure 2**). This molecular circuitry appears to allow the factors to maintain their own expression. Autoregulation could provide stability of gene expression, allowing for the maintenance of the metastable pluripotent state (Alon et al., 2007). Such autoregulation is a genetic feature of many eukaryotic master regulators (Lee et al., 2002; Odom et al., 2004; Odom, 2006). Functional studies confirm that Oct4 and Sox2 co-occupy and activate the *Oct4* and *Nanog* genes (Kuroda et al., 2005; Okumura-Nakanishi et al., 2005). Experiments with an inducible *Sox2*-null murine ES-cell line provide compelling evidence for the existence of this interconnected autoregulatory loop and its role in the maintenance of pluripotency (Masui et al., 2007). *Sox2* deletion leads to rapid differentiation of these ES cells, but enforced expression of Oct4 can rescue the phenotype and maintain pluripotency, suggesting that the major required function of Sox2 is maintenance of proper Oct4 expression. This regulatory circuitry suggests an explanation for the rare iPS colonies formed when Oct4, Klf4 and c-Myc are over-expressed in MEFs (Takahashi and Yamanaka, 2006). When Oct4 is exogenously over-expressed it can contribute directly

to the activation of endogenous *Oct4* and endogenous *Sox2*, the products of which in turn contribute to the maintenance of these genes.

Co-occupancy of target genes

Furthermore, Oct4, Sox2 and Nanog co-occupy a large set of promoters, often at apparently overlapping genomic sites (Boyer et al., 2005) (**Figure 2**). This suggests that these pluripotency factors generally do not control their target genes independently, but rather act coordinately to maintain the transcriptional program required for pluripotency. Biochemical studies support his model. Recently, a large multi-protein complex, including Nanog and Oct4, was identified by iterative immunoprecipitation experiments in ES cells (Wang et al., 2006). Several transcription factor subunits of this complex were shown to be required for the maintenance of pluripotency, providing further evidence that multiple interacting proteins coordinately control pluripotency. Multiple ES cell transcription factors identified in this complex occupy overlapping sites in the genome (Kim et al., 2008). This model provides an explanation for the fact that iPS cell generation requires the combinatorial over-expression of multiple transcription factors; no single factor is sufficient to induce pluripotency. However, exogenous Nanog is not required, suggesting that not all components of this putative complex are required to initiate the process of reprogramming.

Regulating developmental regulators

Master regulators of pluripotency occupy the promoters of active genes that promote ES cell self-renewal, including transcription factors, signal transduction

components and chromatin modifying enzymes (Boyer et al., 2005). Oct4, Sox2 and Nanog also co-occupy the promoters of a smaller set of developmental transcription factors that are silent in ES cells, but which are associated with lineage commitment and cellular differentiation (Boyer et al., 2005) (**Figure 2**). These ES cell transcription factors, expressed in the ICM and the epiblast *in vivo*, contribute to the repression of transcription factors associated with other lineages present in the early embryo, the trophoblast and the primitive endoderm (Niwa et al., 2000; Boyer et al., 2005; Niwa et al., 2007). Additionally, Oct4, Sox2 and Nanog are found at the promoters of developmental transcription factors that are expressed later in development and contribute to lineage specification (Boyer et al., 2005). These findings suggest that Oct4, Sox2 and Nanog may help to maintain the pluripotent state of ES cells by contributing to repression of key lineage specification factors.

Many of the known developmental regulators targeted by Oct4, Sox2 and Nanog are also occupied by the Polycomb Group (PcG) proteins, which contribute to chromatin modifications and gene silencing (Lee et al., 2006; Bernstein et al., 2006; Boyer et al. 2006) (**Figure 2**). In ES cells, developmental regulator genes are silenced in a fashion that allows for subsequent expression in appropriate lineages that can descend from these pluripotent cells. H3K27 methylation, catalyzed by the Polycomb Repressive complex 2 (PRC2), is associated with this dynamic form of silencing, especially of regulators of development (Lee et al., 2006; Bernstein et al., 2006; Boyer et al. 2006). H3K27 methylation is selectively removed from the promoters of genes that are active in a given cell type (Lee et al., 2006; Bernstein et al., 2006; Boyer et al. 2006; Mikkelson, 2007; Zhao et al., 2007; Pan et al., 2007). H3K27 demethylases, UTX and JMJD3, contribute

to the removal of these epigenetic marks during vertebrate development (Lee et al., 2007; Lan et al., 2007). Some evidence suggests that a subset of genes that will not be required in a given lineage is silenced in a more permanent fashion in differentiated cells. Polycomb Group protein EZH2 may recruit DNA methyltransferases that catalyze CpG methylation, a stable epigenetic mark associated with silencing (Viré et al., 2006). However, many developmental regulators retain H3K27 methylation at genes that are silenced in a given lineage (Guenther et al., 2007; Mikkelson et al., 2007). Furthermore, recent work indicates that these genes often remain "poised" for activation, even in lineages where they are not used. RNA polymerase II is found initiated, poised to transcribe developmental regulator genes, but paused at transcription starts in lineages where the genes remain silent (Guenther et al., 2007). These findings suggest that Polycomb complexes, or associated proteins, could serve to pause RNA polymerase machinery at key regulators of development in pluripotent cells and in lineages where they are not expressed (**Figure 2**). Furthermore, the continued presence of RNA polymerase II at the start sites of genes that are not expressed could suggest a fundamental plasticity of cell type, even among cells that have already differentiated.

miRNA regulation in ES cells

Pluripotency transcription factors control a complex gene regulatory network. Autoregulation and control of other transcription factors involved in pluripotency and developmental decisions have already been discussed. One notable missing piece of this network model is analysis of the transcriptional control of miRNAs, non-coding RNAs that participate in post-transcriptional gene regulation.

Mature miRNAs, which specify post-transcriptional repression, arise from larger transcripts that are then processed (**Figure 3**). Primary miRNA sequences are transcribed and then cleaved in the nucleus by a multi-unit microprocessor to yield a precursor miRNA (pre-miRNA) hairpin (Denli et al., 2004; Gregory et al., 2004). The hairpin subsequently exits the nucleus and is cleaved by the RNase type III enzyme, dicer, to yield a mature 18-24 nucleotide miRNA (Bernstein et al., 2001; Hutvagner et al., 2001; Grishok et al., 2001; Knight and Bass, 2001). Mature miRNAs interact with complementary sequences in messenger RNA molecules and can contribute to destabilization of target mRNAs and repression of translation (Bartel, 2004; Ambros, 2004; Plasterk, 2006).

Several lines of evidence indicate that miRNAs contribute to the control of early development. miRNAs appear to regulate the expression of a significant percentage of all genes in a wide array of mammalian cell types (Lewis et al., 2005; Lim et al., 2005; Krek et al., 2005; Farh et al., 2005). A subset of miRNAs is preferentially expressed in ES cells or embryonic tissue (Houbaviy et al., 2003; Suh et al., 2004; Houbaviy et al., 2005; Mineno et al., 2006). Dicer-deficient mice fail to develop (Bernstein et al., 2003) and ES cells deficient in miRNA processing enzymes show defects in differentiation, self-renewal and perhaps viability (Kanellopoulou et al., 2005; Murchison et al., 2005; Wang et al., 2007). Specific miRNAs have been shown to participate in mammalian cellular differentiation and embryonic development (Stefani and Slack, 2008). However, functional interactions between transcription factors and miRNAs in the regulatory circuitry that controls early development have not yet been examined.

Signaling pathways

Pluripotent embryonic stem cells can self-renew in an undifferentiated, pluripotent state in culture, but are dependent on extra-cellular cues for this maintenance of pluripotency. MEF feeder cells provide a culture condition that promotes ES cell self-renewal, as does MEF-conditioned media, suggesting that MEFs secrete soluble factors that contribute to the maintenance of pluripotency (Okita and Yamanaka, 2006). LIF was the first soluble factor identified that allows for propagation of undifferentiated ES cells without the use of MEFs (Williams et al., 1988; Smith et al., 1988). Notably, LIF does not promote the maintenance of pluripotency in human ES cells, which could be due to a species-specific effect of LIF signaling, or could be the result of a difference in the developmental stage at which mouse and human ES cells are isolated (Brons et al., 2007; Tesar et al., 2007; Rossant, 2008). More recently, other soluble factors, including Wnt, BMP, FGF have been shown to promote pluripotency depending on species and cell isolation method (Okita and Yamanaka, 2006; Rossant, 2008). Human ES cells and autologously derived fibroblasts, were shown to send reciprocal paracrine signals of FGF and IGF, respectively, and these signals were sufficient to maintain the pluripotency of the ES cells (Bendall, 2007). These findings suggest that signals help to establish a local microenvironment *in vitro*, and presumably *in vivo*, to maintain pluripotency.

The Wnt signaling pathway (illustrated in **Figure 4b**) contributes to the maintenance of pluripotency in mouse and human ES cells (Sato et al., 2004; Ogawa et al., 2006; Singla et al., 2006; Cai et al., 2007), as well as the self-renewal of undifferentiated adult stem cells in multiple tissues (Reya and Clevers, 2005). This intra-

cellular signal transduction pathway feeds into the core transcriptional circuitry of pluripotency (Cole et al., 2008). The terminal components of the canonical Wnt pathway are Tcf/Lef family transcription factors, the most abundant of which in ES cells is Tcf3 (shown in **Figures 2 and 4**). In ES cells, Tcf3 occupies the genome at sites that overlap extensively with the binding sites of Oct4, Sox2 and Nanog. Wnt stimulation of ES cells, or depletion of Tcf3 in these cells, modulates the expression of key ES cell transcription factors including Oct4 and Nanog (Cole et al, 2008). In yeast experiments, terminal kinases downstream of signal transduction pathways were recently shown to directly interact with the genome and contribute to transcriptional regulation (Pokholok, 2006). This is likely a general feature of eukaryotic signal transduction pathways. Efforts are underway to investigate the ways in which the terminal components of multiple signaling pathways impinge on the core transcriptional circuitry of ES cells to contribute to the maintenance of pluripotency or cellular differentiation.

Molecular control of differentiation

Signaling pathways also contribute to cellular differentiation. When cellular lineage commitment occurs, the appropriate regulators of development lose Polycomb-mediated repression and are activated. In addition, regulators of pluripotency, including *Oct4* and *Nanog* are rapidly silenced. Retinoic acid is one particularly well-characterized inducer of differentiation, and has been shown to contribute to silencing of the *Oct4* locus (Okamoto, 1990; Pikarsky, 1994).

Further studies have begun to characterize the carefully regulated silencing of *Oct4* and other regulators of pluripotency that occurs during early differentiation. A set

of nuclear repressors has been identified that are induced in differentiating cells and are required for proper silencing of *Oct4*, including ARP-1, COUP-TF1 and GCNF (also referred to as Nr6a1) (Ben-Shushan, 1995; Fuhrmann, 2001; Gu, 2005; Gu, 2006). Pluripotency regulators also undergo major changes in epigenetic state during differentiation. At *Oct4*, marks of activation including H3K4me3 and H3K7 and H3K9 acetylation are lost. Furthermore, H3K9me2 and me3, marks associated with heterochromatin, are gained in a G9a histone methyltransferase-dependent manner (Feldman, 2006). Finally, in a process dependent on *de novo* DNA methyltransferases DNMT3a/3b, which are recruited directly or indirectly by G9a, the *Oct4* promoter undergoes CpG methylation. These findings indicate that *Oct4*, and other ES cell-specific genes including *Rex1*, but not *Nanog* or *Sox2*, undergo a multi-step, tightly regulated form of silencing, during which they adopt an epigenetic state characteristic of heterochromatin (Feldman, 2006). These epigenetic changes appear to enforce a more stable form of silencing compared to the more labile epigenetic silencing associated with H3K27 methylation at genes that must be dynamically regulated during development. Oct4 is not required in somatic cells in the adult mouse (Lengner, 2007). In fact, ectopic expression of *Oct4* has been shown to rapidly lead to massive hyperplasia of poorly differentiated cells, especially in the intestine, and rapid fatality, highlighting the strong evolutionary pressure to ensure complete silencing of pluripotency regulators in somatic cells (Hochedlinger, 2005). These multilayered marks of epigenetic silencing, including H3K9 methylation and CpG methylation, must be progressively removed in the process of generating iPS cells from somatic cells (**Figure 5**).

Induction of pluripotency in somatic cells with defined factors

The core regulatory circuitry of ES cell pluripotency outlined above immediately suggests key features of the molecular mechanism by which iPS cells are generated from somatic cells. Over-expression of Oct4 and Sox2 presumably leads to the activation of the endogenous loci encoding these factors due to their ability to contribute to interconnected autoregulation. Furthermore, they almost certainly directly activate the transcription of genes encoding other regulators, including *Nanog*, which are also required for maintenance of pluripotency. Immunostaining confirms that cells that ultimately adopt a true ES cell-like state reactivate expression of Nanog and other markers of pluripotency (Wernig, 2007).

Ultimately, iPS cells adopt a universal transcriptional program largely indistinguishable from that of ES cells (Okita, 2007; Wernig, 2007; Maherali, 2007). In order to accomplish this reactivation, reprogramming factors must reverse the multi-layered heterochromatin-like silencing that occurs at least at a subset of key pluripotency regulators (**Figure 5**). To date, H3K9 methylation status has not been examined at the promoters of pluripotency regulators in iPS cells, but bisulfite sequencing has revealed a loss of DNA methylation at these loci (Okita, 2007; Wernig, 2007; Maherali, 2007). The exogenous reprogramming factors and re-activated, endogenous pluripotency factors must directly or indirectly reestablish the global epigenetic state associated with pluripotency. In particular, as described above, repression of developmental regulators mediated by Polycomb Group proteins must be re-established, thus silencing the transcriptional program of the somatic cell of origin. Indeed, H3K27me3 chromatin modifications that had been lost in differentiation to MEFs are re-established in iPS cells (Wernig, 2007; Maherali, 2007; Takahashi et al., 2007).

Chromatin changes associated with developmental reprogramming appear to happen progressively over an extended period of time, and do not occur homogenously in all cells transduced with reprogramming factors. This was highlighted in initial experiments that selected for reprogrammed cells with drug selection linked to *Fbx15* expression (Takahashi and Yamanaka, 2006). In these experiments, re-activation of the *Fbx15* locus correlates with some degree of pluripotency induction, however the global expression pattern of these cells still differs greatly from true, germline-competent ES cells. Consistent with these differences in expression pattern, the cells have not fully adopted an ES cell-like epigenetic state, and partial CpG methylation is still found at the promoters of key regulators of pluripotency (Takahashi and Yamanaka, 2006) (**Figure 5**). However, with addition of the same four factors, other cells do in fact adopt a transcriptional and epigenetic program that is much closer to ES cells and allows for true pluripotency (Okita et al., 2007; Wernig et al., 2007; Maherali et al., 2007). These rare cells with more complete epigenetic reprogramming can either be selected with drug selection linked to *Oct4* or *Nanog* (Okita et al., 2007; Wernig et al., 2007; Maherali et al., 2007), or can be identified based on morphology (Meissner et al., 2007; Blelloch et al., 2007). Furthermore, over time, following introduction of the reprogramming factors, more cells gradually appear adopt a true pluripotent epigenetic state, consistent with models where these epigenetic changes occur progressively or stochastically (Brambrink et al., 2008; Stadtfeld et al., 2008) (**Figure 5**).

The observation that Klf4, and other Klf family members, co-occupy many of the promoters bound by Oct4, Sox2 and Nanog, although not necessarily at the same sites (Jiang et al., 2008, Kim et al., 2008), provides clues to how Klf4 contributes to

developmental reprogramming. The functional significance of this co-occupancy is highlighted by studies of transcriptional regulation of *Lefty1,* an ES-cell specific gene (Nakatake et al., 2006). Oct4 and Sox2 bind to overlapping sites in the *Lefty1* promoter, but are not sufficient to transcriptionally activate the gene in somatic cells. Remarkably, addition of one more factor, Klf4, allows for expression of *Lefty1* in somatic cells. Klf4 binds to a separate DNA in the sequence, more proximal to the transcriptional start site than the Oct4/Sox2 site (Nakatake et al., 2006). This suggests an attractive model for the role of Klf4 in reprogramming. Klf4 occupies a subset of Oct4/Sox2/Nanog bound promoters, and the combined actions of Oct4, Sox2 and Klf4 are required for proper expression of key ES cell genes.

Although, c-Myc does not appear to be strictly required for developmental reprogramming, this oncogene has a significant effect on the timing and efficiency of reprogramming (Nakagawa et al., 2008; Wernig et al., 2008). c-Myc, perhaps in concert with Klf4, could allow for a cancer-like transformation of somatic cells, conferring on MEFs the resistance to senescence and rapid proliferative phenotype associated with ES cells (Rodolfa and Eggan, 2006; Yamanaka, 2007; Knoepfler, 2008). A slightly different model posits that c-Myc plays an important role in priming the chromatin state of MEFs to allow the reprogramming factors to more efficiently switch the cell state (Rodolfa and Eggan, 2006; Yamanaka, 2007; Knoepfler, 2008). This model is supported by the extensive association of Myc protein with genome and its ability to modify chromatin (Cawley et al., 2004; Fernandez et al., 2003; Li et al., 2003; Knoepfler, 2006). Furthermore, genomic occupancy of c-Myc is distinct from that of core reprogramming factors, and overlaps highly with nucleosomes modified with epigenetic marks of active

transcription (Kim et al., 2008). In addition, a non-transcriptional role has been proposed for c-Myc (Knoepfler, 2008). c-Myc associates with pre-replication complexes and promotes DNA synthesis independent of transcriptional regulation (Dominguez-Sola et al., 2007). Rapid DNA replication, promoted by Myc over-expression, could provide an enhanced opportunity for the somatic genome to reset its epigenetic state in the presence of exogenous reprogramming factors.

Further studies of the programming of ES cells and the reprogramming of somatic cells

Presented here are two studies, in Chapter 2 and Chapter 3, that build on the existing model of the regulatory circuitry controlling pluripotency. Chapter 2 examines how miRNA genes are regulated by core transcription factors and Polycomb group proteins in ES cells. This study integrates miRNAs into a model of the multi-level network, involving both transcriptional and post-transcriptional regulation, that controls ES cell identity. Chapter 3 describes efforts to exploit knowledge of the regulatory network controlling ES cell pluripotency, to increase the efficiency of reprogramming somatic cells to pluripotency. In particular, the study uses transient stimulation of the Wnt signaling pathway, which connects directly to the core transcriptional circuitry of ES cells (Cole et al., 2008), to promote the induction of pluripotency in concert with nuclear reprogramming factors.

Foxp3 and control of CD4+CD25+ Regulatory T cells

$CD4^+CD25^+$ regulatory T (T_{reg}) cells are a specialized sub-population of T cells that are essential for the prevention of autoimmunity (Sakaguchi et al., 1995). T_{reg} cells have an attenuated cytokine response to T-cell receptor stimulation, and can suppress the proliferation and effector function of neighboring T-cells *in vivo* and *in vitro* (Shevach, 2002; von Boehmer, 2005). The molecular mechanisms of this suppression remain largely unknown (Shevach, 2002; von Boehmer, 2005). Failure to develop CD4+CD25+ T_{reg} cells is known to cause multi-organ autoimmunity both in mice and humans. Patients with common autoimmune diseases, such as multiple sclerosis, have normal numbers of T_{reg} cells, but these T_{reg} cells are deficient in suppressing conventional T cells as measured by *in vitro* suppression assays (Viglietta et al., 2004; Baecher-Allan et al., 2006).

The forkhead transcription factor Foxp3 is a key developmental determinant of T_{reg} cells. Foxp3 expression is restricted to T_{reg} cells, and *Foxp3* knockout causes a specific defect in T_{reg} cell development (Hori et al., 2003; Fontenot, 2003; Khattri et al., 2003; Fontenot, 2005). *Foxp3* knockout mice have a phenotype characterized by massive lymphoproliferation and multi-organ autoimmunity (Brunkow et al., 2001). In humans, *FOXP3* mutations have been linked to the autoimmune condition, Immune dysregulation, Polyendocrinopathy, Enteropathy, X-linked syndrome (IPEX) (Wildin et al., 2001; Bennett et al., 2001). Foxp3 is not only required for the establishment of functional T_{reg} cells, but also for the maintenance of mature T_{reg} cells in the periphery (Williams and Rudensky, 2007). Enforced expression of Foxp3 is sufficient to cause conventional T

cells to assume a suppressive phenotype, demonstrating its role as a master regulator of the T_{reg} cell identity (Hori et al., 2003). An understanding of the mechanism by which Foxp3 drives the differentiation of T_{reg} cells would help elucidate the genetic program that allows for the suppressive phenotype of T_{reg} cells and would also identify candidates for genes in which mutations contribute to aberrant T_{reg} cell function in human autoimmune disease.

miRNAs in T_{reg} cells

There is accumulating evidence of the role of miRNA-mediated gene regulation in the immune system. Targeted deletion of murine miR-155, a miRNA preferentially expressed in hematopoietic cells, revealed key roles for this transcript in regulation the adaptive immune response, including helper T cell differentiation (Rodriguez et al., 2007; Thai et al., 2007). miR-155-deficent T cells displayed aberrant expression of cytokines and transcription factors (Rodriguez et al., 2007).

miRNAs are also important specifically in T_{reg} function. *Dicer* depletion results in deficent T_{reg} cells and abnormal immune response (Cobb et al., 2006). T_{reg} cells have a miRNA expression profile distinct from that of conventional T cells, including higher levels of miR-155. Ectoptic expression of Foxp3 is able to shift the miRNA expression profile of conventional T cells towards the profile of T_{reg} cells, including activation of miR-155, suggesting that key miRNAs are under the control of this master regulator of the T_{reg} identity (Cobb et al., 2006).

Regulation of T_{reg} cells by T-cell stimulation

One of the defining functional characteristics of T_{reg} cells is a diminished cytokine response to stimulation of the T-cell receptor (TCR) relative to conventional T cells. Whereas conventional T cells secrete high doses of interleukin 2 (Il2) in response to T-cell stimulation, T_{reg} cells are refractory to secreting *Il2* (Schubert et al., 2001) (**Figure 6**). In fact, Foxp3 binds directly to the *Il2* promoter and represses transcription of this gene (Schubert et al., 2001; Chen et al., 2006; Wu et al., 2006). The consensus forkhead DNA sequence motif bound by Foxp3 in this promoter is adjacent to the binding motif for another transcription factor, Nfat (nuclear factor of activated T-cells) (Schubert et al, 2001.). Foxp3 interacts with and attenuates that transcriptional activity of Nfat and NF-kappaB, two transcription factors that translocate to the nucleus of T cells in response to T-cell stimulation (Bettelli et al., 2005). These findings suggest that a key function of Foxp3 might be to control a distinctive T_{reg} response to TCR stimulation (**Figure 6**). In fact, FOXP3 physically interacts with NFAT in a DNA-binding complex (Wu et al., 2006). The two transcription factors cooperate, both positively and negatively, to regulate the transcription of key known target genes, including *Il2*, which is repressed in T_{reg} cells, and *Il2ra (CD25),* a cell surface molecule preferentially expressed in T_{reg} cells (Wu et al., 2006). Nfat is a canonical effector of T-cell stimulation, suggesting that T_{reg} cell identity is cooperatively shaped by the lineage specification factor, Foxp3, and the transcriptional program induced by T-cell stimulation. However, the extent of Foxp3 cooperation with Nfat in shaping the gene expression program of in T_{reg} cells cannot be assessed fully without knowledge of the set of promoters occupied by Foxp3 throughout the genome.

Global control of the T_{reg} gene expression program by Foxp3

Chapter 4 describes studies to identify the global set of Foxp3 target genes to begin to elucidate how this transcription factor establishes T_{reg} cellular identity. Experiments presented were designed to identify genes whose promoters are occupied by Foxp3 and whose expression is dependent on this transcription factor. To investigate the interaction of Foxp3 with T-cell stimulation in controlling the gene expression program of T_{reg} cells, Foxp3 binding and Foxp3-dependent gene expression patterns were examined in the presence and absence of T-cell stimulation. This work was a step towards understanding how Foxp3 establishes the gene expression program required for T_{reg} cell function and how defects in this program might contribute to autoimmunity.

References

Alon, U. (2007). Network motifs: theory and experimental approaches. Nat Rev Genet *8*, 450-461.

Ambros, V. (2004). The functions of animal microRNAs. Nature *431*, 350-355.

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

Baecher-Allan, C., and Hafler, D. A. (2006). Human regulatory T cells and their role in autoimmune disease. Immunol Rev *212*, 203-216.

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

Ben-Shushan, E., Sharir, H., Pikarsky, E., and Bergman, Y. (1995). A dynamic balance between ARP-1/COUP-TFII, EAR-3/COUP-TFI, and retinoic acid receptor:retinoid X receptor heterodimers regulates Oct-3/4 expression in embryonal carcinoma cells. Mol Cell Biol *15*, 1034-1048.

Bendall, S. C., Stewart, M. H., Menendez, P., George, D., Vijayaragavan, K., Werbowetski-Ogilvie, T., Ramos-Mejia, V., Rouleau, A., Yang, J., Bosse, M., *et al.* (2007). IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. Nature *448*, 1015-1021.

Bennett, C. L., Christie, J., Ramsdell, F., Brunkow, M. E., Ferguson, P. J., Whitesell, L., Kelly, T. E., Saulsbury, F. T., Chance, P. F., and Ochs, H. D. (2001). The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nat Genet *27*, 20-21.

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

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Bernstein, E., Kim, S. Y., Carmell, M. A., Murchison, E. P., Alcorn, H., Li, M. Z., Mills, A. A., Elledge, S. J., Anderson, K. V., and Hannon, G. J. (2003). Dicer is essential for mouse development. Nat Genet *35*, 215-217.

Bettelli, E., Dastrange, M., and Oukka, M. (2005). Foxp3 interacts with nuclear factor of activated T cells and NF-kappa B to repress cytokine gene expression and effector functions of T helper cells. Proc Natl Acad Sci U S A *102*, 5138-5143.

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

Blais, A., Tsikitis, M., Acosta-Alvear, D., Sharan, R., Kluger, Y., and Dynlacht, B. D. (2005). An initial blueprint for myogenic differentiation. Genes Dev *19*, 553-569.

Blelloch, R., Venere, M., Yen, J., and Ramalho-Santos, M. (2007). Generation of induced pluripotent stem cells in the absence of drug selection. Cell Stem Cell *1*, 245-247.

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

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell *2*, 151-159.

Brons, I. G., Smithers, L. E., Trotter, M. W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A., and Vallier, L. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191-195.

Brunkow, M. E., Jeffery, E. W., Hjerrild, K. A., Paeper, B., Clark, L. B., Yasayko, S. A., Wilkinson, J. E., Galas, D., Ziegler, S. F., and Ramsdell, F. (2001). Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. Nat Genet *27*, 68-73.

Cai, L., Ye, Z., Zhou, B. Y., Mali, P., Zhou, C., and Cheng, L. (2007). Promoting human embryonic stem cell renewal or differentiation by modulating Wnt signal and culture conditions. Cell Res *17*, 62-72.

Campbell, D. J., and Ziegler, S. F. (2007). FOXP3 modifies the phenotypic and functional properties of regulatory T cells. Nat Rev Immunol *7*, 305-310.

Campbell, K. H., McWhir, J., Ritchie, W. A., and Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. Nature *380*, 64-66.

Cao, Y., Kumar, R. M., Penn, B. H., Berkes, C. A., Kooperberg, C., Boyer, L. A., Young, R. A., and Tapscott, S. J. (2006). Global and gene-specific analyses show distinct roles for Myod and Myog at a common set of promoters. Embo J *25*, 502-511.

Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J., *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*, 499-509.

Chalfie, M., Horvitz, H. R., and Sulston, J. E. (1981). Mutations that lead to reiterations in the cell lineages of C. elegans. Cell *24*, 59-69.

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

Chen, C., Rowell, E. A., Thomas, R. M., Hancock, W. W., and Wells, A. D. (2006). Transcriptional regulation by Foxp3 is associated with direct promoter occupancy and modulation of histone acetylation. J Biol Chem *281*, 36828-36834.

Cobb, B. S., Hertweck, A., Smith, J., O'Connor, E., Graf, D., Cook, T., Smale, S. T., Sakaguchi, S., Livesey, F. J., Fisher, A. G., and Merkenschlager, M. (2006). A role for Dicer in immune regulation. J Exp Med *203*, 2519-2527.

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

Cowan, C. A., Atienza, J., Melton, D. A., and Eggan, K. (2005). Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. Science *309*, 1369-1373.

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

Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., and Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231-235.

Dominguez-Sola, D., Ying, C. Y., Grandori, C., Ruggiero, L., Chen, B., Li, M., Galloway, D. A., Gu, W., Gautier, J., and Dalla-Favera, R. (2007). Non-transcriptional control of DNA replication by c-Myc. Nature *448*, 445-451.

Eggan, K., Baldwin, K., Tackett, M., Osborne, J., Gogos, J., Chess, A., Axel, R., and Jaenisch, R. (2004). Mice cloned from olfactory sensory neurons. Nature *428*, 44-49.

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

Farh, K. K., Grimson, A., Jan, C., Lewis, B. P., Johnston, W. K., Lim, L. P., Burge, C. B., and Bartel, D. P. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science *310*, 1817-1821.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol *8*, 188-194.

Fernandez, P. C., Frank, S. R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A., and Amati, B. (2003). Genomic targets of the human c-Myc protein. Genes Dev *17*, 1115-1129.

Fontenot, J. D., Gavin, M. A., and Rudensky, A. Y. (2003). Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. Nat Immunol *4*, 330-336.

Fontenot, J. D., Rasmussen, J. P., Williams, L. M., Dooley, J. L., Farr, A. G., and Rudensky, A. Y. (2005). Regulatory T cell lineage specification by the forkhead transcription factor foxp3. Immunity *22*, 329-341.

Fuhrmann, G., Chung, A. C., Jackson, K. J., Hummelke, G., Baniahmad, A., Sutter, J., Sylvester, I., Scholer, H. R., and Cooney, A. J. (2001). Mouse germline restriction of Oct4 expression by germ cell nuclear factor. Dev Cell *1*, 377-387.

Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem *74*, 481-514.

Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235-240.

Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., and Mello, C. C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell *106*, 23-34.

Gu, P., Le Menuet, D., Chung, A. C., and Cooney, A. J. (2006). Differential recruitment of methylated CpG binding domains by the orphan receptor GCNF initiates the repression and silencing of Oct4 expression. Mol Cell Biol *26*, 9471-9483.

Gu, P., LeMenuet, D., Chung, A. C., Mancini, M., Wheeler, D. A., and Cooney, A. J. (2005). Orphan nuclear receptor GCNF is required for the repression of pluripotency genes during retinoic acid-induced embryonic stem cell differentiation. Mol Cell Biol *25*, 8507-8519.

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

Hadorn, E. (1968). Transdetermination in cells. Sci Am *219*, 110-114 passim.

Harfe, B. D., McManus, M. T., Mansfield, J. H., Hornstein, E., and Tabin, C. J. (2005). The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. Proc Natl Acad Sci U S A *102*, 10898-10903.

He, L., Thomson, J. M., Hemann, M. T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S. W., Hannon, G. J., and Hammond, S. M. (2005). A microRNA polycistron as a potential human oncogene. Nature *435*, 828-833.

Hochedlinger, K., and Jaenisch, R. (2002). Monoclonal mice generated by nuclear transfer from mature B and T donor cells. Nature *415*, 1035-1038.

Hochedlinger, K., Yamada, Y., Beard, C., and Jaenisch, R. (2005). Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. Cell *121*, 465-477.

Hori, S., Nomura, T., and Sakaguchi, S. (2003). Control of regulatory T cell development by the transcription factor Foxp3. Science *299*, 1057-1061.

Houbaviy, H. B., Dennis, L., Jaenisch, R., and Sharp, P. A. (2005). Characterization of a highly variable eutherian microRNA gene. Rna *11*, 1245-1257.

Houbaviy, H. B., Murray, M. F., and Sharp, P. A. (2003). Embryonic stem cell-specific MicroRNAs. Dev Cell *5*, 351-358.

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., and Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Inoue, K., Wakao, H., Ogonuki, N., Miki, H., Seino, K., Nambu-Wakao, R., Noda, S., Miyoshi, H., Koseki, H., Taniguchi, M., and Ogura, A. (2005). Generation of cloned mice by direct nuclear transfer from natural killer T cells. Curr Biol *15*, 1114-1118.

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

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

Kanatsu-Shinohara, M., Inoue, K., Lee, J., Yoshimoto, M., Ogonuki, N., Miki, H., Baba, S., Kato, T., Kazuki, Y., Toyokuni, S., *et al.* (2004). Generation of pluripotent stem cells from neonatal mouse testis. Cell *119*, 1001-1012.

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

Khattri, R., Cox, T., Yasayko, S. A., and Ramsdell, F. (2003). An essential role for Scurfin in CD4+CD25+ T regulatory cells. Nat Immunol *4*, 337-342.

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

Knight, S. W., and Bass, B. L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science *293*, 2269-2271.

Knoepfler, P. S. (2008). Why Myc? An Unexpected Ingredient in the Stem Cell Cocktail. Cell Stem Cell *2*, 18-21.

Knoepfler, P. S., Zhang, X. Y., Cheng, P. F., Gafken, P. R., McMahon, S. B., and Eisenman, R. N. (2006). Myc influences global chromatin structure. Embo J *25*, 2723-2734.

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

Lan, F., Bayliss, P. E., Rinn, J. L., Whetstine, J. R., Wang, J. K., Chen, S., Iwase, S., Alpatov, R., Issaeva, I., Canaani, E., *et al.* (2007). A histone H3 lysine 27 demethylase regulates animal posterior development. Nature *449*, 689-694.

Lee, M. G., Villa, R., Trojer, P., Norman, J., Yan, K. P., Reinberg, D., Di Croce, L., and Shiekhattar, R. (2007). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science *318*, 447-450.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., *et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., *et al.* (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799-804.

Lengner, C. J., Camargo, F. D., Hochedlinger, K., Welstead, G. G., Zaidi, S., Gokhale, S., Scholer, H. R., Tomilin, A., and Jaenisch, R. (2007). Oct4 expression is not required for mouse somatic stem cell self-renewal. Cell Stem Cell *1*, 403-415.

Li, J., Ishii, T., Feinstein, P., and Mombaerts, P. (2004). Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons. Nature *428*, 393-399.

Li, Z., Van Calcar, S., Qu, C., Cavenee, W. K., Zhang, M. Q., and Ren, B. (2003). A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proc Natl Acad Sci U S A *100*, 8164-8169.

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

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

Margueron, R., Trojer, P., and Reinberg, D. (2005). The key to development: interpreting the histone code? Curr Opin Genet Dev *15*, 163-176.

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*, 7634-7638.

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

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Mineno, J., Okamoto, S., Ando, T., Sato, M., Chono, H., Izu, H., Takayama, M., Asada, K., Mirochnitchenko, O., Inouye, M., and Kato, I. (2006). The expression profile of microRNAs in mouse embryos. Nucleic Acids Res *34*, 1765-1771.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is

required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

Murchison, E. P., Partridge, J. F., Tam, O. H., Cheloufi, S., and Hannon, G. J. (2005). Characterization of Dicer-deficient murine embryonic stem cells. Proc Natl Acad Sci U S A *102*, 12135-12140.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol *26*, 101-106.

Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., Yagi, K., Miyazaki, J., Matoba, R., Ko, M. S., and Niwa, H. (2006). Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. Mol Cell Biol *26*, 7772-7782.

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

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

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

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

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

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

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

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313-317.

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

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

Orkin, S. H., and Zon, L. I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell *132*, 631-644.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G. A., Stewart, R., and Thomson, J. A. (2007). Whole-genome analysis of histone h3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell *1*, 299-312.

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

Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S. F., D'Agati, V., Orkin, S. H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature *349*, 257-260.

Pikarsky, E., Sharir, H., Ben-Shushan, E., and Bergman, Y. (1994). Retinoic acid represses Oct-3/4 gene expression through several retinoic acid-responsive elements located in the promoter-enhancer region. Mol Cell Biol *14*, 1026-1038.

Plasterk, R. H. (2006). Micro RNAs in animal development. Cell *124*, 877-881.

Pokholok, D. K., Zeitlinger, J., Hannett, N. M., Reynolds, D. B., and Young, R. A. (2006). Activated signal transduction kinases frequently occupy target genes. Science *313*, 533-536.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.

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

Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., and Bartel, D. P. (2002). Prediction of plant microRNA targets. Cell *110*, 513-520.

Rodolfa, K. T., and Eggan, K. (2006). A transcriptional logic for nuclear reprogramming. Cell *126*, 652-655.

Rossant, J. (2008). Stem cells and early lineage development. Cell *132*, 527-531.
Stefani, G., and Slack, F. J. (2008). Small non-coding RNAs in animal development. Nat Rev Mol Cell Biol *9*, 219-230.

Rudnicki, M. A., Schnegelsberg, P. N., Stead, R. H., Braun, T., Arnold, H. H., and Jaenisch, R. (1993). MyoD or Myf-5 is required for the formation of skeletal muscle. Cell *75*, 1351-1359.

Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M., and Toda, M. (1995). Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. J Immunol *155*, 1151-1164.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A. H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through

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

Schubert, L. A., Jeffery, E., Zhang, Y., Ramsdell, F., and Ziegler, S. F. (2001). Scurfin (FOXP3) acts as a repressor of transcription and regulates T cell activation. J Biol Chem *276*, 37672-37679.

Shaywitz, D. A., and Melton, D. A. (2005). The molecular biography of the cell. Cell *120*, 729-731.

Shevach, E. M. (2002). CD4+ CD25+ suppressor T cells: more questions than answers. Nat Rev Immunol *2*, 389-400.

Silva, J., Chambers, I., Pollard, S., and Smith, A. (2006). Nanog promotes transfer of pluripotency after cell fusion. Nature *441*, 997-1001.

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

Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C. G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. Nat Struct Mol Biol *15*, 259-267.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688-690.

Staal, F. J., and Clevers, H. C. (2005). WNT signalling and haematopoiesis: a WNT-WNT situation. Nat Rev Immunol *5*, 21-30.

Stadtfeld, M., Maherali, N., Breault, D. T., and Hochedlinger, K. (2008). Defining Molecular Cornerstones during Fibroblast to iPS Cell Reprogramming in Mouse. Cell Stem Cell *2*, 230-240.

Stark, A., Brennecke, J., Bushati, N., Russell, R. B., and Cohen, S. M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell *123*, 1133-1146.

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Suh, M. R., Lee, Y., Kim, J. Y., Kim, S. K., Moon, S. H., Lee, J. Y., Cha, K. Y., Chung, H. M., Yoon, H. S., Moon, S. Y., *et al.* (2004). Human embryonic stem cells express a unique set of microRNAs. Dev Biol *270*, 488-498.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr Biol *11*, 1553-1558.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

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

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L., and McKay, R. D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199

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

Turner, B. M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836-845.

Viglietta, V., Baecher-Allan, C., Weiner, H. L., and Hafler, D. A. (2004). Loss of functional suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis. J Exp Med *199*, 971-979.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J. M., *et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.

von Boehmer, H. (2005). Mechanisms of suppression by suppressor T cells. Nat Immunol *6*, 338-344.

Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R., and Yanagimachi, R. (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. Nature *394*, 369-374.

Wakayama, T., Tabar, V., Rodriguez, I., Perry, A. C., Studer, L., and Mombaerts, P. (2001). Differentiation of embryonic stem cell lines generated from adult somatic cells by nuclear transfer. Science *292*, 740-743.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T. W., and Orkin, S. H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.

Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat Genet *39*, 380-385.

Weintraub, H. (1993). The MyoD family and myogenesis: redundancy, networks, and thresholds. Cell *75*, 1241-1244.

Wencker, M., Sausse, C., Derse, D., Gazzolo, L., and Duc Dodon, M. (2007). Human T-cell leukemia virus type 1 Tax protein down-regulates pre-T-cell receptor alpha gene transcription in human immature thymocytes. J Virol *81*, 301-308.

Wernig, M., Meissner, A., Cassady, J. P., and Jaenisch, R. (2008). c-Myc Is Dispensable for Direct Reprogramming of Mouse Fibroblasts. Cell Stem Cell *2*, 10-12.

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

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.

Wildin, R. S., Ramsdell, F., Peake, J., Faravelli, F., Casanova, J. L., Buist, N., Levy-Lahad, E., Mazzella, M., Goulet, O., Perroni, L., *et al.* (2001). X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. Nat Genet *27*, 18-20.

Williams, L. M., and Rudensky, A. Y. (2007). Maintenance of the Foxp3-dependent developmental program in mature regulatory T cells requires continued expression of Foxp3. Nat Immunol *8*, 277-284.

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

Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A. D., Stroud, J. C., Bates, D. L., Guo, L., Han, A., Ziegler, S. F., *et al.* (2006). FOXP3 controls regulatory T cell function through cooperation with NFAT. Cell *126*, 375-387.

Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663-676.

Yamanaka, S. (2007). Strategies and new developments in the generation of patient-specific pluripotent stem cells. Cell Stem Cell *1*, 39-49.

Yates, A., and Chambers, I. (2005). The homeodomain protein Nanog and pluripotency in mouse embryonic stem cells. Biochem Soc Trans *33*, 1518-1521.

Yu, J., Vodyanik, M. A., He, P., Slukvin, II, and Thomson, J. A. (2006). Human embryonic stem cells reprogram myeloid precursors following cell-cell fusion. Stem Cells *24*, 168-176.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Zhao, X. D., Han, X., Chew, J. L., Liu, J., Chiu, K. P., Choo, A., Orlov, Y. L., Sung, W. K., Shahab, A., Kuznetsov, V. A., *et al.* (2007). Whole-genome mapping of histone h3 lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell *1*, 286-298.

Acknowledgments

Brett Chevalier, Rudolf Jaenisch and Rick Young provided useful comments during the preparation of this introduction. I thank Stuart Levine, Tom Dicesare, Megan Cole and Julia Zeitlinger for providing images to accompany the text. Figure 1 was adapted from J. Zeilinger and Jaenisch and Young, 2008. Figure 2 is adapted from Jaenisch and Young, 2008 and Cole et al., 2008. Figure 3 was provided by Tom Dicesare. Figure 4 is adapted from M. Cole and Staal and Clevers, 2004. Figure 5 is adapted from Jaenisch and Young, 2008. Figure 6 is adapted from Campbell and Ziegler, 2007.

Figure Legends

Figure 1. Model of the molecular circuitry controlling cellular identity. The expression of both protein-coding genes (pink rectangles) and miRNA genes (purple hexagons) are dependent on transcriptional regulation by transcription factors (blue circles) and chromatin regulators (green circles). The activity of these regulators is often dependent on extra-cellular cues that influence intra-cellular signal transduction pathways. The products of genes include transcriptional regulators and miRNAs that participate in the gene regulatory network.

Figure 2. Core transcriptional circuitry of ES cells. Multiple core ES cell transcription factors (blue circles) (1) participate in an interconnected autoregulatory loop (bottom left), (2) generally co-occupy the same sites in the promoters of their target genes, and, (3) collectively target two sets of genes, one that is actively expressed in ES cells and another that is silent in ES cells, but remains poised for subsequent expression during cellular differentiation (Boyer et al., 2005) Polycomb Group proteins (PcG, green circle) co-occupy a set of silent developmental regulators in ES cells that are expressed in differentiated cells in a lineage-specific fashion. These genes tend to show evidence of transcriptional initiation by RNA Polymerase II (Pol2, grey oval), but are not productively transcribed in ES cells.

Figure 3. miRNA biogenesis. Primary miRNAs are transcribed, generally by RNA Polymerase II, and then cleaved in the nucleus by Drosha and Dgcr8 to generate an RNA hairpin. The hairpin is exported from the nucleus and cleaved by Dicer, to yield a mature 18-24 nucleotide miRNA, which is unwound and incorporated into the RISC complex.

Mature miRNAs interact with complementary sequences in mRNAs and the RISC complex contributes to destabilization of target mRNAs and repression of translation (Bartel, 2004).

Figure 4. Core components of the Wnt/ß-catenin pathway. **a**. In the absence of extra-cellular Wnt ligand, Frizzled and LRP receptors (low-density-lipoprotein-receptor-related proteins 5 or 6) are not associated in the plasma membrane. A multi-protein complex, including APC (adenomatous polyposis coli), Axin (axis inhibitor) and Gsk3ß (glycogen-synthase kinase 3ß), phosphorylates cytoplasmic ß-catenin (B-cat) and tags it for degradation by the proteosome. This prevents the accumulation of nuclear B-cat. In the absence of nuclear B-cat, TCF3 associates transcriptional co-repressors of the GRG (groucho-related gene) family at the promoters of target genes. **b**. Wnt ligand promotes the association of Frizzled and LRP receptors. This leads to the phosphorylation and inactivation of Gsk3ß, mediated by Dvl (Dishevelled) (not shown). B-cat accumulates and migrates to the nucleus where it associates with Tcf/Lef proteins, including Tcf3 shown here, at target genes to regulate transcription.

Figure 5. Model of progressive epigenetic reprogramming in iPS cell generation. a. Fibroblasts transduced with reprogramming factors undergo gradual changes over weeks to adopt the true pluripotent epigenetic state of fully reprogrammed iPS cells (Brambrink et al., 2008; Stadtfeld et al., 2008). Fully reprogrammed iPS cells have a global gene expression program comparable to ES cells (Okita et al., 2007; Wernig et al., 2007; Maherali et al., 2007) **b.** In fibroblasts, the promoters of genes encoding core ES cell transcription factors, including *Oct4* and *Nanog*, are CpG methylated (indicated with

closed circles) and silenced. Initial experiments generating iPS cells (Takahashi and Yamanaka, 2006) identified cells that are partially reprogrammed based on their global expression pattern and partially demethylated (mixed closed circles and open circles) at the promoters of genes encoding key regulators of pluripotency. In fully reprogrammed iPS cells (Okita et al., 2007; Wernig et al., 2007; Maherali et al., 2007), promoters of genes encoding core ES cell transcription factors, including *Oct4* and *Nanog*, are CpG demethylated (indicated with open circles). **c.** Upon transduction, viral encoded reprogramming factors are expressed and presumably bind to the promoters of endogenous genes encoding core ES cell regulators, although the timing of this is not yet known. The endogenous genes encoding core ES regulators are gradually activated. In fully reprogrammed iPS cells, the viral genes are silenced, suggesting that the pluripotent state is maintained by core ES cell transcription factor autoregulation.

Figure 6. Foxp3 transcriptional control and T-cell signaling. a. Ligation of the T-cell receptor (TCR) and co-stimulatory molecules including CD28 leads to a productive cytokine response, including synthesis and secretion of Interleukin 2 (Il2), in conventional T cells. Transcription factors including Nfat, Ap1 and NF-kappaB (NFKB) contribute to this response to T-cell stimulation. **b.** One of the defining functional characteristics of T_{reg} cells is a diminished cytokine response to stimulation of the T-cell receptor (TCR) relative to conventional T cells. One potential function of Foxp3 in T_{reg} cells is direct inhibition of transcription at the promoters of genes, such as *Il2*, that would otherwise be activated by T-cell stimulation. **c.** Foxp3 could also regulate the expression of proteins that modulate the intra-cellular signaling pathway that mediates the response to T-cell stimulation. Shown here is a hypothetical gene X that is activated by Foxp3 and

inhibits T-cell stimulation, but similarly Foxp3 could repress a gene that would otherwise promote T-cell stimulation. Models of Foxp3 function presented in **b** and **c** are both consistent with the data presented in Chapter 4.

Figure 1

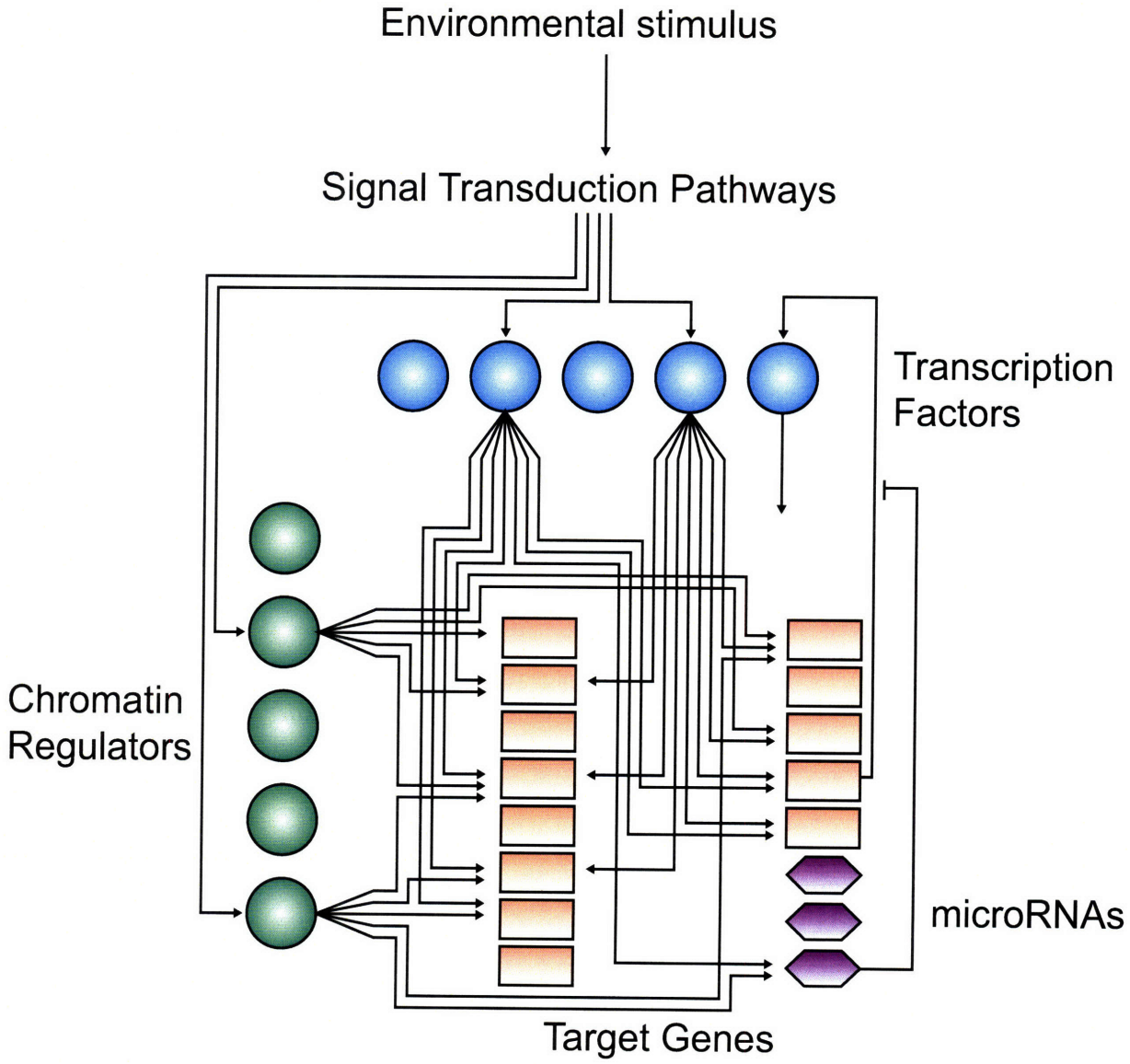

Environmental stimulus
Signal Transduction Pathways
Transcription Factors
Chromatin Regulators
microRNAs
Target Genes

Figure 2

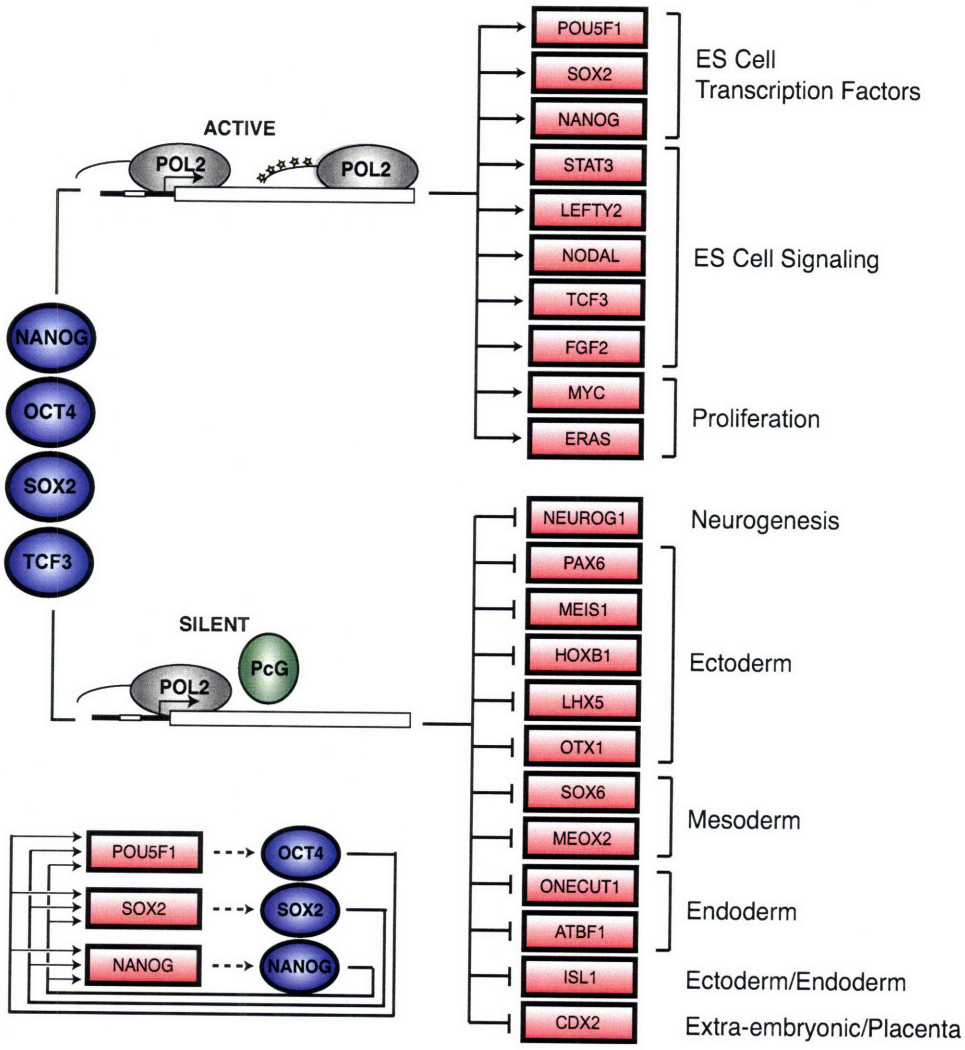

ACTIVE
POL2
POL2
NANOG
OCT4
SOX2
TCF3
SILENT
PcG
POL2
POU5F1
SOX2
NANOG
OCT4
SOX2
NANOG
POU5F1
SOX2
NANOG
STAT3
LEFTY2
NODAL
TCF3
FGF2
MYC
ERAS
ES Cell
Transcription Factors
ES Cell Signaling
Proliferation
NEUROG1
PAX6
MEIS1
HOXB1
LHX5
OTX1
SOX6
MEOX2
ONECUT1
ATBF1
ISL1
CDX2
Neurogenesis
Ectoderm
Mesoderm
Endoderm
Ectoderm/Endoderm
Extra-embryonic/Placenta

Figure 3

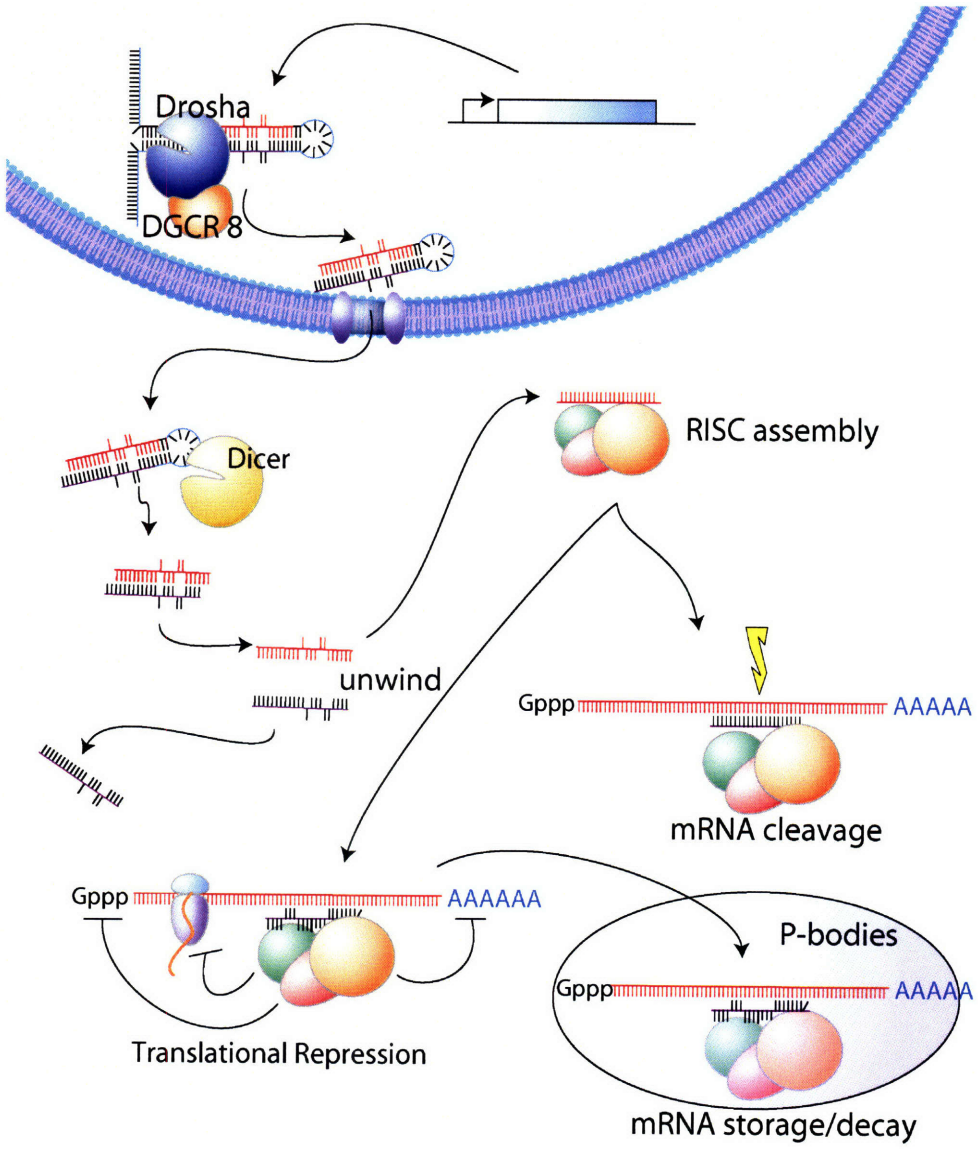

Drosha
DGCR 8
Dicer
RISC assembly
unwind
Gppp
AAAAA
mRNA cleavage
Gppp
AAAAAA
Translational Repression
P-bodies
Gppp
AAAAA
mRNA storage/decay

Figure 4

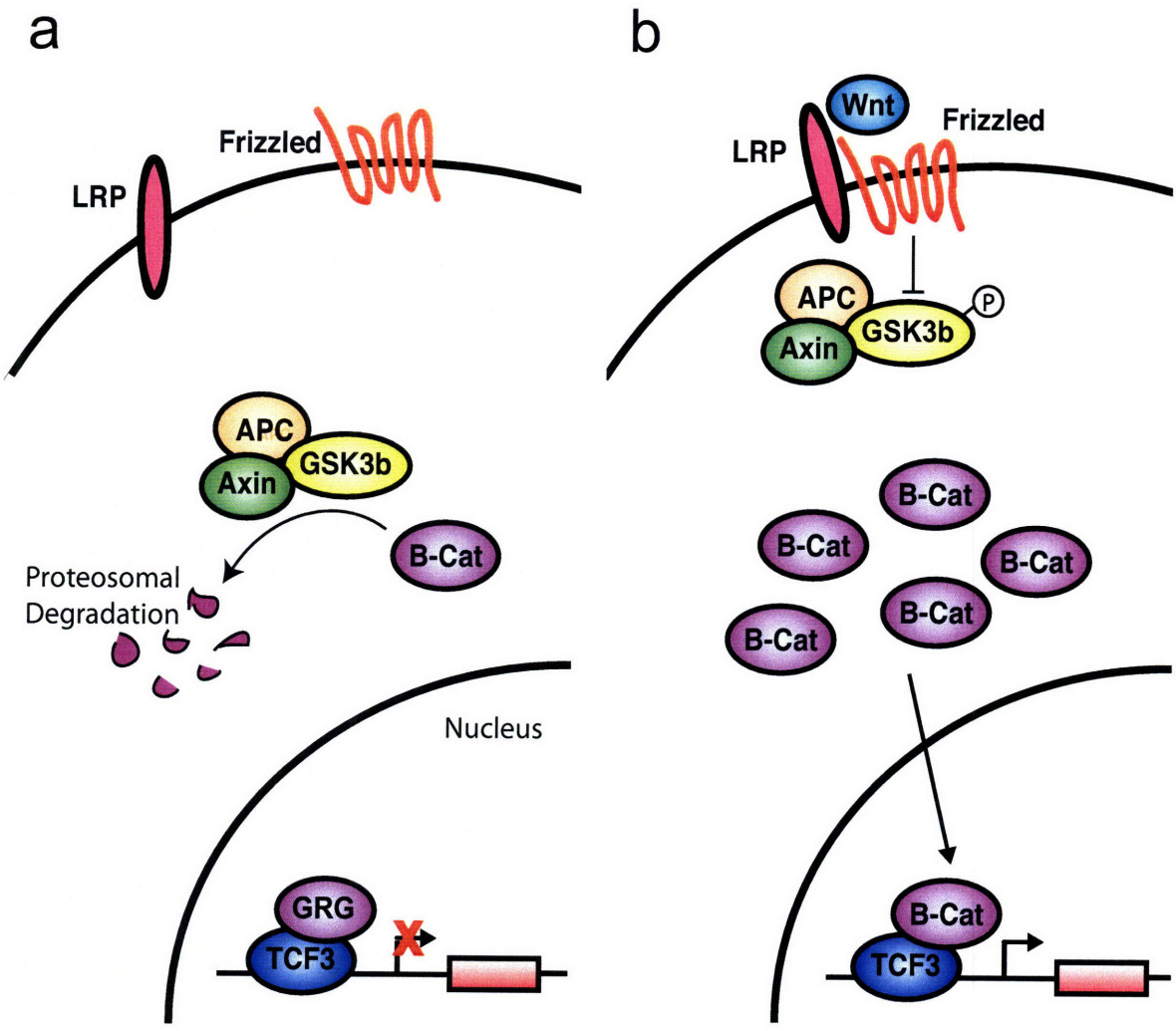

a
Frizzled
LRP
APC
Axin
GSK3b
B-Cat
Proteosomal
Degradation
Nucleus
GRG
TCF3
b
Wnt
Frizzled
LRP
APC
Axin
GSK3b
P
B-Cat
B-Cat
B-Cat
B-Cat
B-Cat
B-Cat
TCF3

Figure 5

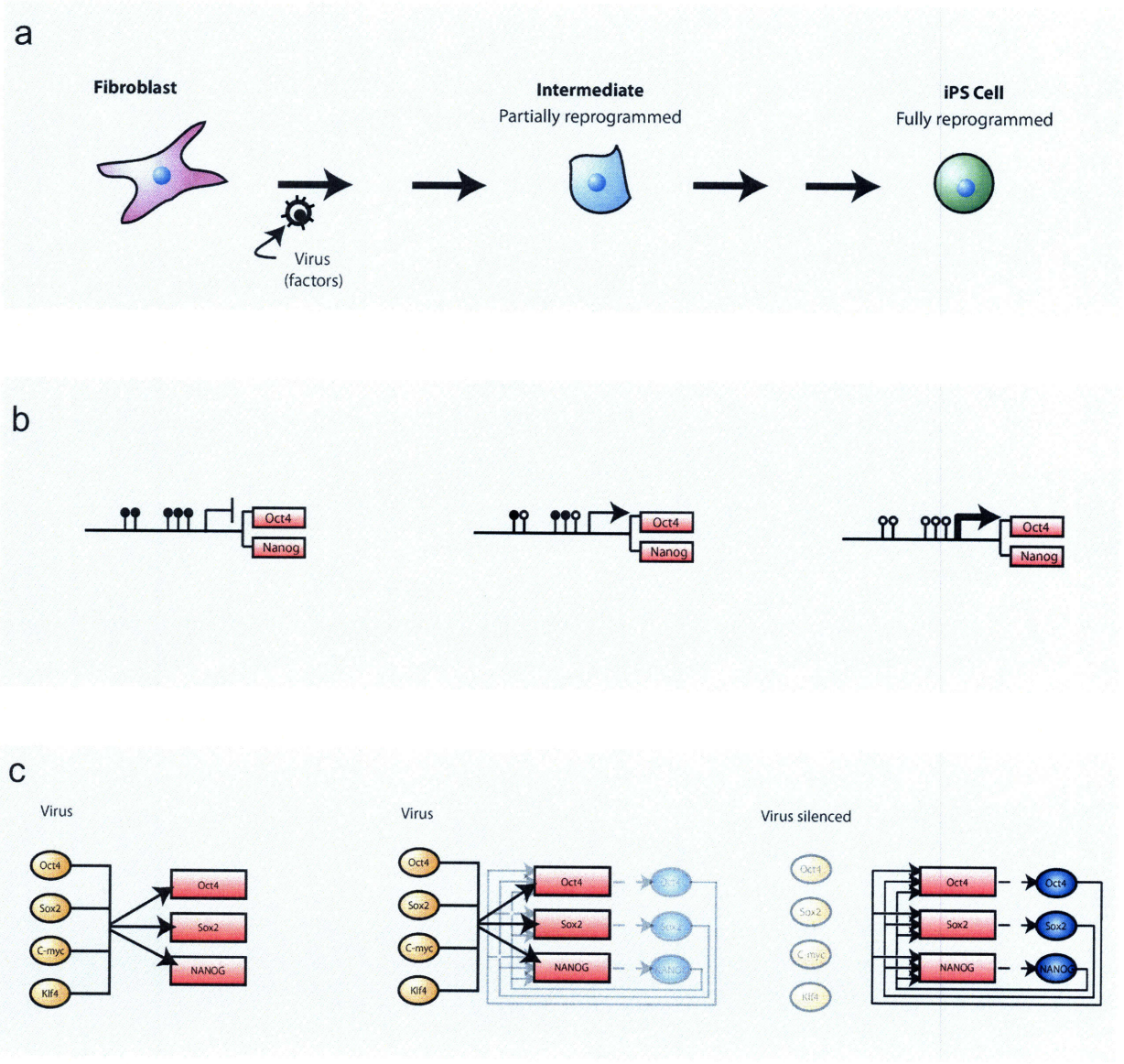
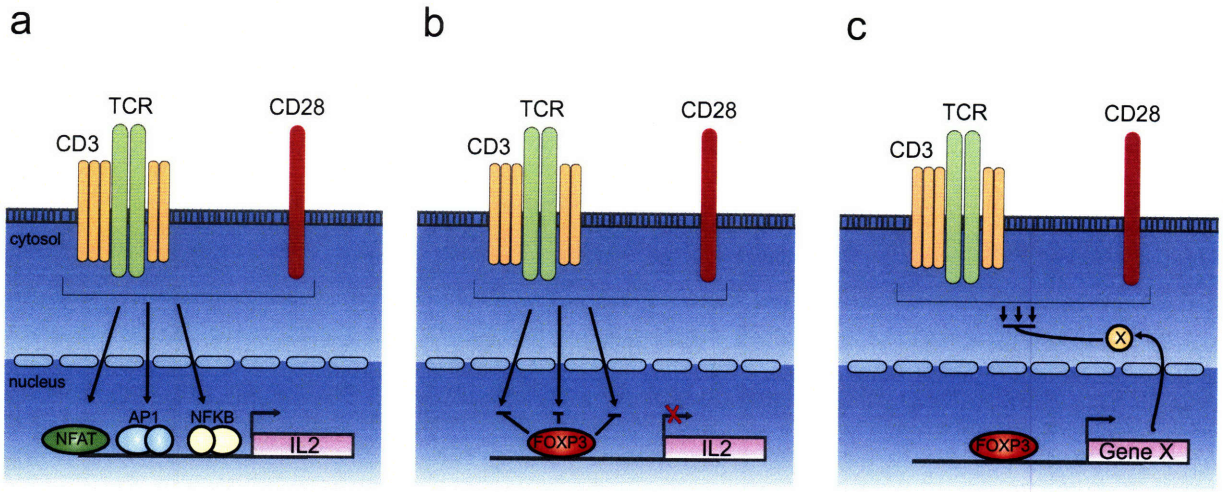

a
Fibroblast
Intermediate
Partially reprogrammed
iPS Cell
Fully reprogrammed
Virus
(factors)
b
Oct4
Nanog
Oct4
Nanog
Oct4
Nanog
c
Virus
Oct4
Sox2
C-myc
Klf4
Oct4
Sox2
NANOG
Virus
Oct4
Sox2
C-myc
Klf4
Oct4
Sox2
NANOG
Virus silenced
Oct4
Sox2
C-myc
Klf4
Oct4
Sox2
NANOG
Oct4
Sox2
NANOG

Figure 6


a
TCR
CD28
CD3
cytosol
nucleus
NFAT
AP1
NFKB
IL2
b
TCR
CD28
CD3
T
FOXP3
IL2
c
TCR
CD28
CD3
X
FOXP3
Gene X

Chapter 2

Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells

Alexander Marson*, Stuart S. Levine*, Megan F. Cole, Garrett M. Frampton, Tobias Brambrink, Matthew G. Guenther, Wendy K. Johnston, Marius Wernig, Jamie Newman, Thomas L. Volkert, David P. Bartel, Rudolf Jaenisch, Richard A. Young

**These authors contributed equally*

Personal Contribution to the Project

This chapter is the result of a close collaboration with Stuart Levine. We conceived of this study together. He was responsible for the computational analysis and I was responsible for overseeing the experiments and for writing the main text of the manuscript.

Summary

MicroRNAs (miRNAs) are crucial for normal embryonic stem (ES) cell self-renewal and cellular differentiation, but how miRNA gene expression is controlled by the key transcriptional regulators of ES cells has not been established. We describe here a new map of the transcriptional regulatory circuitry of ES cells that incorporates both protein-coding and miRNA genes, and which is based on high-resolution ChIP-seq data, systematic identification of miRNA promoters, and quantitative sequencing of short transcripts in multiple cell types. We find that the key ES cell transcription factors are associated with promoters for most miRNAs that are preferentially expressed in ES cells and with promoters for a set of silent miRNA genes. This silent set of miRNA genes is co-occupied by Polycomb Group proteins in ES cells and expressed in a tissue-specific fashion in differentiated cells. These data reveal how key ES cell transcription factors promote the miRNA expression program that contributes to self-renewal and cellular differentiation, and integrate miRNAs and their targets into an expanded model of the regulatory circuitry controlling ES cell identity.

Introduction

Embryonic stem (ES) cells hold significant potential for clinical therapies because of their distinctive capacity to both self-renew and differentiate into a wide range of specialized cell types. Understanding the transcriptional regulatory circuitry of ES cells and early cellular differentiation is fundamental to understanding human development and realizing the therapeutic potential of these cells. Transcription factors that control ES cell pluripotency and self-renewal have been identified (Chambers and Smith, 2004; Niwa, 2007; Silva and Smith, 2008) and a draft of the core regulatory circuitry by which these factors exert their regulatory effects on protein-coding genes has been described (Boyer et al., 2005; Loh et al., 2006; Lee et al., 2006; Boyer et al. 2006; Jiang et al., 2008; Cole et al., 2008; Kim et al., 2008). MicroRNAs (miRNAs) are also likely to play key roles in ES cell gene regulation (Kanellopoulou et al., 2005; Murchison et al., 2005; Wang et al., 2007), but little is known about how miRNAs participate in the core regulatory circuitry controlling self-renewal and pluripotency in ES cells.

Several lines of evidence indicate that miRNAs contribute to the control of early development. miRNAs appear to regulate the expression of a significant percentage of all genes in a wide array of mammalian cell types (Lewis et al., 2005; Lim et al., 2005; Krek et al., 2005; Farh et al., 2005). A subset of miRNAs is preferentially expressed in ES cells or embryonic tissue (Houbaviy et al., 2003; Suh et al., 2004; Houbaviy et al., 2005; Mineno et al., 2006). Dicer-deficient mice fail to develop (Bernstein et al., 2003) and ES cells deficient in miRNA processing enzymes show defects in differentiation, self-renewal and perhaps viability (Kanellopoulou et al., 2005; Murchison et al., 2005; Wang

et al., 2007; Calabrese et al., 2008). Specific miRNAs have been shown to participate in mammalian cellular differentiation and embryonic development (Stefani and Slack, 2008). However, how transcription factors and miRNAs function together in the regulatory circuitry that controls early development has not yet been examined.

The major limitation in connecting miRNA genes to the core transcriptional circuitry of ES cells has been sparse annotation of miRNA gene transcriptional start sites and promoter regions. Mature miRNAs, which specify post-transcriptional gene repression, arise from larger transcripts that are then processed (Bartel, 2004). Over 400 mature miRNAs have been confidently identified in the human genome (Landgraf et al., 2007), but only a minority of the primary transcripts have been identified and annotated. Prior attempts to connect ES cell transcriptional regulators to miRNA genes have searched for transcription factor binding sites only close to the annotated mature miRNA sequences (Boyer et al., 2005; Loh et al., 2006; Lee et al., 2006). Additionally, studies of the core transcriptional circuitry of ES cells have compared transcription factor occupancy to mRNA expression data, but have not systemically examined miRNA expression in ES cells and differentiated cell types, limiting our knowledge of transcriptional regulation of miRNA genes in these cells (Boyer et al., 2005; Loh et al., 2006; Lee et al., 2006; Cole et al. 2008).

To incorporate miRNA gene regulation into the model of transcriptional regulatory circuitry of ES cells, we began by generating new, high-resolution, genome-wide maps of binding sites for key ES cell transcription factors using massive parallel sequencing of chromatin immunoprecipitation (ChIP-seq). These data reveal highly

overlapping occupancy of Oct4, Sox2, Nanog and Tcf3 at the transcriptional start sites of miRNA transcripts, which we systematically mapped based on a method that uses chromatin landmarks and transcript data. We then carried out quantitative sequencing of short transcripts in ES cells, neural precursor cells (NPCs) and mouse embryonic fibroblasts (MEFs), which revealed that Oct4, Sox2, Nanog and Tcf3 occupy the promoters of most miRNAs that are preferentially or uniquely expressed in ES cells. Our data also revealed that a subset of the Oct4/Sox2/Nanog/Tcf3 occupied miRNA genes are silenced in ES cells by Polycomb Group proteins, but are expressed later in development in specific lineages. High-resolution transcription factor location analysis, systematic mapping of the primary miRNA transcriptional start sites in mouse and human, and quantitative sequencing of miRNAs in three different cell types provide a valuable data resource for studies of the gene expression program in ES and other cells and the regulatory mechanisms that control cell fate. The data also produce an expanded model of ES cell core transcriptional regulatory circuitry that now incorporates transcriptional regulation of miRNAs, and post-transcriptional regulation mediated by miRNAs, into the molecular understanding of pluripotency and early cellular differentiation.

Results

High-resolution genome-wide location analysis in ES cells with ChIP-seq

To connect miRNA genes to the core transcriptional circuitry of ES cells, we first generated high-resolution genome-wide maps of Oct4, Sox2, Nanog, and Tcf3 occupancy (**Figure 1**). ChIP-seq allowed us to map transcription factor binding sites and histone modifications across the entire genome at high resolution (Barski et al., 2007; Johnson et al., 2007; Mikkelsen et al., 2007; Robertson et al., 2007), and we optimized the protocol to allow for robust analysis of transcription factor binding in murine ES cells (Supplemental Material). Oct4, Sox2, Nanog and Tcf3 were found to co-occupy 14,230 sites in the genome (**Figure 1A, Supplementary Figures S1 and S2, Supplementary Tables S1-S3**). Approximately one quarter of these occurred within 8kb of the transcription start site of 3,289 annotated genes, another one quarter occurred within genes but more than 8kb from the start site, and almost half occurred in intergenic regions distal from start sites (**Supplementary Text**). Binding of the four factors at sites surrounding the *Sox2* gene (**Figure 1B**) exemplifies two key features of the data: all four transcription factors co-occupied the identified binding sites and the resolution was sufficient to determine the DNA sequence associated with these binding events to a resolution of <25bp. Composite analysis of all bound regions provided higher resolution and suggested how these factors occupy their common DNA-sequence motif (**Supplementary Figure S3, Supplementary Table S4**). Knowledge of these binding sites provided data necessary to map these key transcription factors to the promoters of miRNA genes.

Identification of miRNA promoters

Imperfect knowledge of the start sites of primary miRNA transcripts has limited our ability to identify the transcription factor binding events that control miRNA gene expression in vertebrates. Previous strategies to identify the 5' ends of primary miRNAs have been hampered because they relied on isolation of transient primary miRNA transcript, required knowledge of the specific cell type in which each given miRNA is transcribed, or focused only on potential start sites proximal to mature miRNAs (Fukao et al., 2007; Mikkelsen et al., 2007; Zhou et al., 2007; Barrera et al., 2008). To systematically identify transcriptional start sites for miRNA genes in the mouse and human genomes, we took advantage of the recent observation that histone H3 is tri-methylated at its lysine 4 residue (H3K4me3) at the transcriptional start sites of most genes in the genome, even when genes are not productively transcribed, and knowledge that this covalent modification is restricted to sites of transcription initiation (Barski et al., 2007; Guenther et al., 2007). We used the genomic coordinates of the H3K4me3 enriched loci derived from multiple cell types (**Supplementary Table S5,** Barski et al., 2007; Guenther et al., 2007; Mikkelsen et al., 2007) to create a library of candidate transcription start sites in both human and mouse (**Figure 2 and Supplementary Figure S4**).

High-confidence promoters were identified for over 80% of miRNAs in both mouse and human (**Figure 2, Supplementary Figure S4 and Supplementary Tables S6 and S7**). These promoters were associated with 185 murine primary microRNA

transcripts (pri-miRNAs) (specifying 336 mature miRNAs), and 294 human pri-miRNAs (specifying 441 mature miRNAs) **(Supplementary Table S6 and S7).** To identify promoters for miRNA genes, the association of candidate transcriptional start sites with regions encoding mature miRNAs was scored based on proximity to annotated mature miRNA sequences (Landgraf et al., 2007), available EST data, and conservation between species (**Figure 2A and Supplementary Figure S5 and Supplementary Text**). Four lines of evidence indicate that this approach identified genuine transcriptional start sites for miRNA genes. Existing EST data provided evidence that the predicted transcripts do in fact originate at the identified start sites and continue through the annotated loci of mature miRNAs (**Figure 2B and Supplementary Figures S5**). In addition to the chromatin signature of promoters, a high fraction of these regions contained CpG islands, a DNA sequence element often associated with promoters (**Figure 2B and Supplementary Table S6 and S7).** Third, in some instances where evidence of primary miRNA transcripts, which may be present only transiently before processing, were not available in published databases at the identified transcriptional start sites, chromatin marks associated with transcriptional elongation including nucleosomes methylated at H3 lysine 36 (H3K36me3) and H3 lysine 79 (H3K79me2), provided evidence that such transcripts are actively produced (**Figure 2C** and Mikkelsen et al., 2007). Finally, most miRNA promoters showed evidence of H3K4me3 enrichment in multiple tissues, as observed at the promoters of most protein-coding genes (Barski et al., 2007; Guenther et al., 2007; Heintzman et al., 2007) (**Figure 2D**).

Occupancy of miRNA promoters by core ES cell transcription factors

The binding sites of the ES cell transcription factors Oct4, Sox2, Nanog and Tcf3 were next mapped to these high-confidence miRNA promoters (**Figure 3**). In murine ES cells, Oct4, Sox2, Nanog, and Tcf3 co-occupied the promoters for 55 distinct miRNA transcription units, which included three clusters of miRNAs that are expressed as large polycistrons, thus suggesting that these regulators have the potential to directly control the transcription of 81 distinct mature miRNAs (**Figure 3A and Supplementary Tables S6**). This set of miRNAs occupied by Oct4/Sox2/Nanog/Tcf3 represents roughly 20 percent of annotated mammalian miRNAs, similar to the ~20 percent of protein-coding genes that are bound at their promoters by these key transcription factors (**Supplementary Table S2**).

To determine if transcription factor occupancy of miRNA promoters is conserved across species, we performed genome-wide location analysis for Oct4 in human ES cells using microarray-based analysis. We found extensive conservation of the set of miRNA genes that were occupied at their promoters by Oct4, as exemplified by the mir-302 cluster (**Figure 3A and 3B and Supplementary Tables S7 and S8**). Transcription factor occupancy does not necessarily mean that the adjacent gene is regulated by that factor; conserved transcription factor occupancy of a promoter, however, suggests gene regulation by that factor. Thus, our data identify a set of miRNA genes that are bound at their promoters by key ES cell transcription factors in mouse and human cells (**Figure 3C**), suggesting that core ES cell transcription factor regulation of these particular miRNA transcripts has functional significance.

Regulation of Oct4 bound miRNA transcripts during differentiation

Oct4 and Nanog are rapidly silenced as ES cells begin to differentiate (Chambers and Smith, 2004; Niwa, 2007). If the Oct4/Sox2/Nanog/Tcf3 complex is required for activation or repression of its target miRNAs, the targets should be differentially expressed when ES cells are compared to a differentiated cell-type. To test this hypothesis, Solexa sequencing of 18-30 nucleotide transcripts in ES cells, mouse embryonic fibroblasts (MEFs), and neural precursors (NPCs), was performed to obtain quantitative information on the abundance of miRNAs in pluripotent cells relative to two differentiated cell types (**Figure 4**).

In each cell type examined, a small subset of mature miRNA transcripts predominated (**Figure 4A**). Members of the mir-290-295 cluster, which encodes multiple miRNAs with the same seed sequence, constituted approximately two thirds of all mature miRNA transcripts in murine ES cells. Let-7 family members constituted roughly one quarter and one half of miRNAs in MEFs and NPCs, respectively. The mir-290-295 cluster, which dominated the expression profile of ES cells, but was scarce in both MEFs and NPCs, is occupied at its promoters by Oct4, Sox2, Nanog and Tcf3 (**Figure 3A**), consistent with the hypothesis that these factors are important for maintaining the expression of the mir-290-295 miRNA cluster in ES cells.

To determine if the behavior of the mir-290-295 cluster is typical of the Oct4/Sox2/Nanog/Tcf3-occupied miRNAs, we further examined the expression of this

set of miRNAs in the three cell types. **Figure 4B** shows how the abundance of this group of miRNAs changed in MEFs and NPCs relative to ES cells. Approximately half of the miRNAs dropped more than an order of magnitude in abundance in MEFs and NPCs relative to ES cells. A small subset of the Oct4/Sox2/Nanog/Tcf3-occupied miRNAs, which will be further discussed below, were expressed only at low levels in ES cells and showed increased abundance in MEFs and NPCs.

Oct4/Sox2/Nanog/Tcf3-occupied miRNAs are, in general, preferentially expressed in embryonic stem cells, as demonstrated by the analysis shown in **Figure 4C**. Whereas most miRNAs are unchanged in expression in ES cells relative to MEFs or NPCs, a significant portion of Oct4/Sox2/Nanog/Tcf3 occupied miRNAs are 100 fold more abundant in ES cells than in MEFs ($p<5 \times 10^{-15}$), and 1,000 fold more abundant in ES cells than in NPCs ($p<5 \times 10^{-9}$). This group of Oct4/Sox2/Nanog/Tcf3 bound miRNAs that is significantly more abundant in ES cells than in NPCs and MEFs, was also found to be actively expressed in induced pluripotent stem (iPS) cells (generated as described in Wernig et al., 2007), at levels comparable to that in ES cells, consistent with the hypothesis that core ES cell transcription factors maintain the expression of these miRNAs in pluripotent cells (**Supplementary Figure S6**).

Polycomb Group Proteins co-occupy tissue-specific miRNAs that are silenced in ES cells

We noted that the Oct4/Sox2/Nanog/Tcf3-bound miRNAs include the majority of miRNAs that were preferentially expressed in ES cells, but the data also revealed a second, smaller group of Oct4/Sox2/Nanog/Tcf3-bound miRNA genes that appeared to be transcriptionally inactive in ES cells (**Figure 4B**). This is reminiscent of previous observations with protein-coding genes in ES cells: Oct4 occupied a set of transcriptionally active genes, but also occupied, with Polycomb Group proteins, a set of transcriptionally repressed genes that are poised for expression upon cellular differentiation (Lee et al., 2006; Bernstein et al., 2006; Boyer et al., 2006). We reasoned that Polycomb complexes might also co-occupy Oct4 bound promoters for miRNA genes that showed little or no evidence for expression, and thus contribute to their silencing. Indeed, new ChIP-seq data for the Polycomb Goup protein Suz12 in murine ES cells supported this hypothesis (**Figure 5A and Supplementary Tables S6, S7, S10**). As expected, these promoters were also enriched for nucleosomes with histone H3K27me3, a chromatin modification catalyzed by Polycomb Group proteins (**Figure 5A and Supplementary Table S6** and Mikkelsen et al., 2007). In keeping with the repressive function of the Polycomb Group proteins reported at protein coding genes, miRNAs occupied at their promoters by Suz12 in ES cells were significantly less abundant in ES cells compared to all other miRNAs (**Figure 5B**). Approximately one quarter of the Oct4/Sox2/Nanog/Tcf3-occupied miRNAs belonged to the repressed set of miRNA genes bound by Suz12 in murine ES cells **(Supplementary Tables S6 and S7**).

To further examine the behavior of this set of miRNAs during embryonic cell-fate commitment, we returned to our quantitative sequencing data of short transcripts in ES cells, MEFs and NPCs (**Figure 5C**). Notably, miRNAs that were bound by Polycomb

Group proteins in ES cells are among the transcripts that are specifically induced in each of these cell types. For example, transcript levels of miR-9, a miRNA previously identified in neural cells and which promotes neural differentiation (Lagos-Quintana et al., 2002; Krichevsky et al., 2006), are significantly elevated in NPCs relative to ES cells, but this miRNA remains repressed in MEFs. Similarly, miR-218 and miR-34b/34c expression is induced in MEFs, but remains at low levels in NPCs (**Figure 5C**). Consistent with Polycomb-mediated repression of these lineage-specific miRNAs, the repressive chromatin mark deposited by Polycomb Group proteins, H3K27me3, is selectively lost at the promoters of the miRNAs in the cells in which they are induced (**Figure 5C** and Mikkelsen et al., 2007).

The tissue-specific expression pattern of miRNAs repressed by Polycomb in ES cells is consistent with these miRNAs serving as determinants of cell-fate decisions in a manner analogous to the developmental regulators whose genes are repressed by Polycomb in ES cells (Lee et al., 2006; Bernstein et al., 2006; Boyer et al., 2006). Such a function in cell-fate determination would require that these miRNAs remain silenced in pluripotent ES cells. Indeed, the miRNAs that are repressed in ES cells by Polycomb Group proteins appear to be induced, later in development, in a highly restricted subset of differentiated tissues specific to each miRNA (**Supplementary Figure S7**), unlike the majority of miRNAs identified in mouse (Landgraf et al., 2007). The miRNAs with promoters bound by Polycomb Group proteins in ES cells are significantly enriched ($p<0.005$) among the set of the most tissue-specific mammalian miRNAs (**Supplementary Fig. S7** and Landgraf et al., 2007). This suggests a model whereby

Polycomb Group proteins repress a set of tissue-specific miRNA genes in ES cells, a subset of which are co-occupied by Oct4, Sox2, Nanog and Tcf3 (**Figure 5D**).

Discussion

Here we provide new high-resolution, genome-wide maps of core ES cell transcription factors, identify promoter regions for most miRNA genes, and deduce the association of the ES cell transcription factors with these miRNA genes. We also provide quantitative sequence data of short RNAs in ES cells, NPCs and MEFs to examine changes in miRNA transcription. The key transcriptional regulators in ES cells collectively occupied the promoters of many of the miRNAs that were most abundant in ES cells, including those that were down-regulated as ES cells differentiate. In addition, these factors also occupied the promoters of a second, smaller set of miRNAs that were repressed in ES cells and were selectively expressed in specific differentiated cell types. This second group of miRNAs constitutes a subset of the miRNAs that were silenced by the Polycomb group proteins in ES cells, which is also known to silence key lineage-specific, protein-coding developmental regulators. Together these data reveal two key groups of miRNAs that are direct targets of Oct4/Sox2/Nanog/Tcf3, one group of miRNAs that is preferentially expressed in pluripotent cells and a second group that is silenced in ES cells by the Polycomb group proteins, and is poised to contribute to cell fate-decisions during mammalian development.

miRNA contribution to ES cell identity

Several miRNA ploycistrons, which encode the most abundant miRNAs in ES cells and which are silenced during early cellular differentiation (Houbaviy et al., 2003;

Suh et al., 2004; Houbaviy et al., 2005), were occupied at their promoters by Oct4, Sox2, Nanog and Tcf3. These include the mir-290-295 cluster, which contains multiple mature miRNAs that share seed sequences with members of the murine mir-302 cluster, as well as the human mir-371-373 and mir-302 clusters. miRNAs in the 17-92 cluster also share a highly similar seed sequence. miRNAs in this family have been implicated in cell proliferation (O'Donnell et al., 2005; He et al., 2005; Voorhoeve et al., 2006), consistent with the impaired self-renewal phenotype observed in miRNA-deficient ES cells (Kanellopoulou et al., 2005; Murchison et al., 2005; Wang et al., 2007). The zebrafish homologue of this miRNA family, mir-430, contributes to the rapid degradation of maternal transcripts in early zygotic development (Giraldez et al., 2006), and mRNA expression data suggests that this miRNA family also promotes the clearance of transcripts in early mammalian development (Farh et al., 2005).

In addition to promoting the rapid clearance of transcripts as cells transition from one state to another during development, miRNAs also likely contribute to the control of cell identity by fine-tuning the expression of genes. mir-430, the zebrafish homologue of the mammalian mir-302 family, serves to precisely tune the levels of Nodal antagonists Lefty1 and Lefty 2 relative to Nodal, a subtle modulation of protein levels that has pronounced effects on embryonic development (Choi et al., 2007). Recently, a list of ~250 murine ES cell mRNAs that appear to be under the control of miRNAs in the mir-290-295 cluster was reported (Sinkkonen et al., 2008). This study reports that *Lefty1* and *Lefty2* are evolutionarily conserved targets of the mir-290-295 miRNA family. These miRNAs also maintain the expression of *de novo* DNA methyltransferases 3a and 3b (Dnmt3a and Dnmt3b), perhaps by dampening the expression of the transcriptional

repressor Rbl2, helping to poise ES cells for efficient methylation of *Oct4* and other pluripotency genes during differentiation.

Knowledge of how the core transcriptional circuitry of ES cells connects to both miRNAs and protein-coding genes, reveals recognizable network motifs downstream of Oct4/Sox2/Nanog/Tcf3, involving both transcriptional and post-transcriptional regulation, that further reveal how this circuitry controls ES cell identity (**Figure 6**). *Lefty1* and *Lefty2*, both actively expressed in ES cells, are directly occupied at their promoters by Oct4/Sox2/Nanog/Tcf3. Therefore, the core ES cell transcription factors appear to promote the active expression of *Lefty1* and *Lefty2*, but also fine-tune the expression of these important signaling proteins by activating a family miRNAs that target the *Lefty1* and *Lefty2* 3'UTRs. This network motif whereby a regulator exerts both positive and negative effects on its target, termed "incoherent feed-forward" regulation (Alon, 2007), provides a mechanism to fine-tune the steady-state level or kinetics of a target's activation (**Figure 6A**). Over a quarter of the proposed targets of the mir-290-295 miRNAs also are likely under the direct transcriptional control of Oct4/Sox2/Nanog/Tcf3 based on our binding maps, suggesting that these miRNAs could participate broadly in tuning the effects of ES cell transcription factors (**Figure 6A**).

The miRNA expression program directly downstream of Oct4/Sox2/Nanog/Tcf3 could help poise ES cells for rapid and efficient differentiation, consistent with the phenotype of miRNA-deficient cells (Kanellopoulou et al., 2005; Murchison et al., 2005; Wang et al., 2007; Calabrese et al., 2008). Oct4/Sox2/Nanog/Tcf3 also likely contributes to this poising by their occupancy of the Let-7g promoter. Mature Let-7 transcripts are

scarce in ES cells, but were among the most abundant miRNAs in both MEFs and NPCs (**Figure 3**). Primary Let-7g transcript is abundant in ES cells, but its maturation is blocked by Lin28 (Viswanathan et al., 2008 and data not shown). We now report that the promoters of both Let-7g and *Lin28* are occupied by Oct4/Sox2/Nanog/Tcf3, suggesting that the core ES cell transcription factors promote the transcription of both primary Let-7g and *Lin28*, which blocks the maturation of Let-7g. In this way Let-7 and Lin-28 appear to participate in an incoherent feed-forward circuit downstream of Oct4/Sox2/Nanog/Tcf3 to contribute to rapid cellular differentiation (**Figure 6B**). Notably, ectopic expression of Lin28 in human fibroblasts promotes the induction of pluripotency (Yu et al., 2007), suggesting blocked maturation of pri-Let-7 transcripts plays an important role in the pluripotent state. Additionally, *Dnmt3a and Dnmt3b*, which are indirectly up-regulated by the mir-290-25 miRNAs (Sinkkonen et al., 2008), are also occupied at their promoters by Oct4/Sox2/Nanog/Tcf3, providing examples of "coherent" regulation of important target genes by ES cell transcription factors and the ES cell miRNAs maintained by those transcription factors **(Figure 6C)**.

Multi-layer Regulatory Circuitry of ES cell identity

The regulatory circuitry we present for miRNAs in ES cells can now be integrated into the model of core regulatory circuitry of pluripotency we have proposed previously (Boyer et al., 2005; Lee et al., 2006; Cole et al., 2008), as illustrated in **Figure 7**. Our data reveal that Oct4, Sox2, Nanog and Tcf3 occupy the promoters of two key sets of miRNAs, similar to the two sets of protein-coding genes regulated by these factors: one

set that is actively expressed in pluripotent ES cells and another that is silenced in these cells by Polycomb Group proteins and whose later expression might serve to facilitate establishment or maintenance of differentiated cell states.

The expanded circuit diagram presented here integrates transcription factor occupancy of miRNA genes and existing data on miRNA targets into our model of the molecular control of the pluripotent state. These data suggest that miRNAs that are activated in ES cells by Oct4/Sox2/Nanog/Tcf3, serve to modulate the direct effects of these transcription factors, participating in incoherent feed-forward regulation to tune levels of key genes, and modifying the gene expression program to help poise ES cells for efficient differentiation. Core ES cell transcription factors and the miRNAs under their control coordinately contribute transcriptional and post-transcriptional gene regulation to the network that maintains ES cell identity.

Concluding Remarks

The regulatory circuitry controlled by ES cell transcription factors, Oct4, Sox2, Nanog and Tcf3, and the Polycomb Group proteins, which is required for the normal ES cell state, has offered insights into the molecular control of ES cell pluripotency and self-renewal and cellular reprogramming (Jaenisch and Young, 2008). We now provide high-resolution genome-wide location analysis of these factors provided by ChIP-seq data, and quantitative sequencing of short transcripts in multiple cell types, to connect miRNA genes to the core circuitry of ES cells. This information should prove useful as

investigators continue to probe the role of miRNAs in pluripotency, cell-fate decisions, and perhaps regenerative medicine.

Experimental Procedures

A detailed description of all materials and methods used can be found in Supplementary Information.

Cell Culture

V6.5 (C57BL/6-129) murine ES cells were grown under typical ES conditions (see Supplementary Information) on irradiated mouse embryonic fibroblasts (MEFs). For location analysis, cells were grown for one passage off of MEFs, on gelatinized tissue-culture plates. To generate neural precursor cells, ES cells were differentiated along the neural lineage using standard protocols (see Supplementary Information). V6.5 ES cells were differentiated into neural progenitor cells (NPCs) through embryoid body formation for 4 days and selection in ITSFn media for 5–7 days, and maintained in FGF2 and EGF2 (R&D Systems) (See Supplementary Information). Mouse embryonic fibroblasts were prepared and cultured from DR-4 strain mice as previously described (See Supplementary Information).

Antibodies and ChIP assays

Detailed descriptions of antibodies, antibody specificity and ChIP methods used in this study are provided in Supplementary Information.

Crosslinked cells (~1 x 10^7 per IP) were lysed and sonicated using a Misonix 3000 sonicator to solubilize and shear crosslinked DNA to a 200bp-1000bp fragment size. Batch sonicated whole cell extract was incubated 12-18 hours at 4°C with 100 µl of Dynal Protein G magnetic beads (Dynal) that has been pre-bound to 10 µg of the appropriate antibody. Immunoprecipitates were washed with RIPA buffer and the DNA eluted in 1% SDS at 65°C for 1 hour. Chemical cross-links were reversed for 10 hours to allow isolation of immunoenriched DNA fragments. Immunoprecipitated DNA and control whole cell extract DNA were purified by treatment with RNAse A, proteinase K and two consecutive phenol:chloroform:isoamyl alcohol extractions.

ChIP-seq

Crosslinked cells (1 x 10^7 per IP) were lysed and sonicated using a Misonix 3000 sonicator to solubilize and shear crosslinked DNA to a 200bp-1000bp fragment size. Batch sonicated whole cell extract was incubated 12-18 hours at 4°C with 100 µl of Dynal Protein G magnetic beads (Dynal) that has been pre-bound to 10 µg of the appropriate antibody. Immunoprecipitates were washed with RIPA buffer and the DNA eluted in 1% SDS at 65°C for 1 hour. Chemical cross-links were reversed for 10 hours to allow isolation of immunoenriched DNA fragments. Immunoprecipitated DNA and control whole cell extract DNA were purified by treatment with RNAse A, proteinase K and two consecutive phenol:chloroform:isoamyl alcohol extractions.

Purified immunoprecipitated DNA were prepared for sequencing according to a modified version of the Solexa Genomic DNA protocol. Fragmented DNA was end repaired and subjected to 18 cycles of LM-PCR using oligos provided by Illumina. Amplified fragments between 150 and 300bp (representing shear fragments between 50 and 200nt in length and ~100bp of primer sequence) were isolated by agarose gel electrophoresis and purified. High quality samples were confirmed by the appearance of a smooth smear of fragments from 100-1000bp with a peak distribution between 150 and 300bp. 3ng of Linker-ligated DNA was applied to the flow-cell using the Solexa Cluster Station fluidics device. Following bridge amplification the cluster density and morphology were confirmed by microscopic analysis of flow-cells stained with a 1:5000 dilution of SYBR Green I (Invitrogen). Samples were then subjected to 26 bases of sequencing according to Illumina's standard protocols.

Images acquired from the Solexa sequencer were processed through the bundled Solexa image extraction pipeline and aligned to both mouse NCBI build 36 and 37 using ELAND. Only sequences uniquely matching the reference genome without mismatches were used. Mapped reads were extended to 200bp and allocated into 25bp bins. Groups of bins containing statistically significant enrichment for the epigenetic modification were identified by comparison to a Poissonian background model as well as comparison to an empirical distribution of reads obtained from whole cell extract DNA.

Quanititative short RNA sequencing

A method of cloning the 18-30nt transcripts previously described (Lau et al., 2001) was modified to allow for Solexa (Illumina) sequencing (manuscript submitted). Single-stranded cDNA libraries of short transcripts were generated using size selected RNA from mouse embryonic stem cells, mouse neural precursors, and mouse embryonic fibroblasts. RNA extraction was performed using Trizol, followed by RNeasy purification (Qiagen).

5μg of RNA was size selected and gel purified. 3' Adaptor (pTCGTATGCCGTCTTCTGTTG [idT]) was ligated to RNA with T4 RNA ligase and also, separately with RNA Ligase (Rnl2(1-249)k->Q). Ligation products were gel purified and mixed. 5' adaptor (GUUCAGAGUUCUACAGUCCGACGAUC) was ligated with 4 RNA Ligase.

RT-PCR (Superscript II, Invitrogen) was performed with 5' primer (CAAGCAGAAGACGGCATA). Splicing of overlapping ends PCR (SOEPCR) was performed (Phusion, NEB) with 5' primer and 3' PCR primer (AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA), generating cDNA with extended 3' adaptor sequence. PCR product (40 μl) was denatured (85°C, 10 min, formamide loading dye), and the differently sized strands were purified on a 90% formamide, 8% acrylamide gel, yielding single-stranded DNA suitable Solexa sequencing.

The single-stranded DNA samples were resuspended in 10mM Tris (EB buffer)/0.1% Tween and then used as indicated in the standard Solexa sequencing protocol (Illumina). Each library was run on one lane of the Solexa sequencer.

Promoter array design and data extraction

The design of the oligonucleotide-based whole genome array set and data extraction methods are described in Lee et al., 2006. The microarrays used for location analysis in this study were manufactured by Agilent Technologies (http://www.agilent.com).

References

Alon, U. (2007). Network motifs: theory and experimental approaches. Nat Rev Genet *8*, 450-461.

Barrera, L. O., Li, Z., Smith, A. D., Arden, K. C., Cavenee, W. K., Zhang, M. Q., Green, R. D., and Ren, B. (2008). Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. Genome Res *18*, 46-59.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

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

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bernstein, E., Kim, S. Y., Carmell, M. A., Murchison, E. P., Alcorn, H., Li, M. Z., Mills, A. A., Elledge, S. J., Anderson, K. V., and Hannon, G. J. (2003). Dicer is essential for mouse development. Nat Genet *35*, 215-217.

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

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

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

Choi, W. Y., Giraldez, A. J., and Schier, A. F. (2007). Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. Science *318*, 271-274.

Farh, K. K., Grimson, A., Jan, C., Lewis, B. P., Johnston, W. K., Lim, L. P., Burge, C. B., and Bartel, D. P. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science *310*, 1817-1821.

Fukao, T., Fukuda, Y., Kiga, K., Sharif, J., Hino, K., Enomoto, Y., Kawamura, A., Nakamura, K., Takeuchi, T., and Tanabe, M. (2007). An evolutionarily conserved

mechanism for microRNA-223 expression revealed by microRNA gene profiling. Cell *129*, 617-631.

Giraldez, A. J., Mishima, Y., Rihel, J., Grocock, R. J., Van Dongen, S., Inoue, K., Enright, A. J., and Schier, A. F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75-79.

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

He, L., Thomson, J. M., Hemann, M. T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S. W., Hannon, G. J., and Hammond, S. M. (2005). A microRNA polycistron as a potential human oncogene. Nature *435*, 828-833.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Houbaviy, H. B., Dennis, L., Jaenisch, R., and Sharp, P. A. (2005). Characterization of a highly variable eutherian microRNA gene. Rna *11*, 1245-1257.

Houbaviy, H. B., Murray, M. F., and Sharp, P. A. (2003). Embryonic stem cell-specific MicroRNAs. Dev Cell *5*, 351-358.

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

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

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

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

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

Krek, A., Grun, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., MacMenamin, P., da Piedade, I., Gunsalus, K. C., Stoffel, M., and Rajewsky, N. (2005). Combinatorial microRNA target predictions. Nat Genet *37*, 495-500.

Krichevsky, A. M., Sonntag, K. C., Isacson, O., and Kosik, K. S. (2006). Specific microRNAs modulate embryonic stem cell-derived neurogenesis. Stem Cells *24*, 857-864.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Curr Biol *12*, 735-739.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A. O., Landthaler, M., *et al.* (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. Cell *129*, 1401-1414.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., *et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*,
15-20.

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S., and Johnson, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769-773.

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

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Mineno, J., Okamoto, S., Ando, T., Sato, M., Chono, H., Izu, H., Takayama, M., Asada, K., Mirochnitchenko, O., Inouye, M., and Kato, I. (2006). The expression profile of microRNAs in mouse embryos. Nucleic Acids Res *34*, 1765-1771.

Murchison, E. P., Partridge, J. F., Tam, O. H., Cheloufi, S., and Hannon, G. J. (2005). Characterization of Dicer-deficient murine embryonic stem cells. Proc Natl Acad Sci U S A *102*, 12135-12140.

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

O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V., and Mendell, J. T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. Nature *435*, 839-843.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., *et al.* (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods *4*, 651-657.

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

Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C. G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. Nat Struct Mol Biol *15*, 259-267.

Stefani, G., and Slack, F. J. (2008). Small non-coding RNAs in animal development. Nat Rev Mol Cell Biol *9*, 219-230.

Suh, M. R., Lee, Y., Kim, J. Y., Kim, S. K., Moon, S. H., Lee, J. Y., Cha, K. Y., Chung, H. M., Yoon, H. S., Moon, S. Y., *et al.* (2004). Human embryonic stem cells express a unique set of microRNAs. Dev Biol *270*, 488-498.

Viswanathan, S. R., Daley, G. Q., and Gregory, R. I. (2008). Selective Blockade of MicroRNA Processing by Lin-28. Science.

Voorhoeve, P. M., le Sage, C., Schrier, M., Gillis, A. J., Stoop, H., Nagel, R., Liu, Y. P., van Duijse, J., Drost, J., Griekspoor, A., *et al.* (2006). A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Cell *124*, 1169-1181.

Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat Genet *39*, 380-385.

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

Yi, R., O'Carroll, D., Pasolli, H. A., Zhang, Z., Dietrich, F. S., Tarakhovsky, A., and Fuchs, E. (2006). Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. Nat Genet *38*, 356-362.

Yi, R., Poy, M. N., Stoffel, M., and Fuchs, E. (2008). A skin microRNA promotes differentiation by repressing 'stemness'. Nature *452*, 225-229.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Zhou, X., Ruan, J., Wang, G., and Zhang, W. (2007). Characterization and identification of microRNA core promoters in four model species. PLoS Comput Biol *3*, e37.

Acknowledgements

We thank members of the Young, Jaenisch and Bartel laboratories, especially T. Lee, for discussions and critical review of the manuscript. We also thank M. Calabrese and A. Ravi for helpful discussions. We are grateful to S. Gupta and J. Love at The Whitehead Institute Center for Microarray Technology (WICMT) who helped optimize and perform ChIP-seq, and L.A. Boyer, B. Chevalier, R. Kumar, and T. Lee who were instrumental in performing location analysis in hES cells. We also thank Biology and Research Computing (BaRC), as well as E. Herbolsheimer for computational and technical support and the Whitehead Institute Center for Microarray Technology (WICMT) for assistance with microarray expression analysis. This work was supported in part by NIH grants 5-RO1-HDO45022, 5-R37-CA084198, and 5-RO1- CA087869 to R.J. and by NIH grant HG002668 and a grant from the Whitehead Institute to R.A.Y.

Figure 1 High-resolution genome-wide mapping of core ES cell transcription factors with ChIP-seq. A. Summary of binding data for Oct4, Sox2, Nanog and Tcf3. 14,230 sites are co-bound genome wide and mapped to either promoter proximal (TSS +/- 8kb, dark green) (27% of binding sites), genic (>8kb from TSS, middle green) (30% of binding sites), or intergenic (light green) (43% of binding sites). The promoter proximal binding sites are associated with 3,289 genes. **B.** (upper) Binding of Oct4 (blue), Sox2 (purple), Nanog (orange) and Tcf3 (red) across 37.5kb of mouse chromosome 3 surrounding the *Sox2* gene (black below the graph, arrow indicates transcription start site). Short sequences uniquely and perfectly mapping to the genome were extended to 200bp (maximum fragment length) and scored in 25bp bins. The score of the bins were then normalized to the total number of reads mapped. Highly enriched regions are highlighted by a dotted box. Oct4/Sox2 DNA binding motifs (Loh et al., 2006) were mapped across the genome and are shown as grey boxes below the graph. Height of the box reflects the quality of the motif. (lower) Detailed analysis of three enriched regions (Chromosome 3: 4,837,600-34,838,300, 34,845,300-34,846,000, and 34,859,900-34,860,500) at the *Sox2* gene indicated with boxes above. The 5' most base from ChIP-seq were separated by strand and binned into 25bp regions. Sense (darker tone) and anti-sense (light tone) of each of the four factors tested are directed towards the binding site, which in each case occurs at a high-confidence Oct/Sox2 DNA binding motif indicated below.

Figure 2 Identification of miRNA promoters. A. Description of algorithm for miRNA promoter identification. A library of candidate transcriptional start sites was generated with histone H3 lysine 4 tri-methyl (H3K4me3) location analysis data from multiple tissues (Barski et al., 2007; Guenther et al., 2007; Mikkelsen et al., 2007). Candidates were scored to assess likelihood that they represent true miRNA promoters. Based on scores, a list of mouse and human miRNA promoters was assembled. Additional details can be found in Supplemental Text. **B.** Examples of identified miRNA promoter regions are shown. A map of H3K4me3 enrichment is displayed in regions neighbouring selected human and mouse miRNAs for multiple cell types: human ES cells (hES), REH

human pro-B cell line (B cell), primary human hepatocytes (Liver), primary human T cells (T cell), mouse ES cells (mES), neural precursor cells (NPCs) and mouse embryonic fibroblasts (MEFs). miRNA promoter coordinates were confirmed by distance to mature miRNA genomic sequence, conservation and EST data (shown as solid line where available). Predicted transcriptional start site and direction of transcription are noted by an arrow, with mature miRNA sequences indicated (red). CpG islands, commonly found at promoters, are indicated (green). Dotted lines denote presumed transcripts. **C.** Confirmation of predicted transcription start sites for active miRNAs using chromatin modifications. Normalized ChIP-seq counts for H3K4me3 (red), H3K79me2 (blue) and H3K36me3 (green) are shown for two miRNA genes where EST data was unavailable. Predicted start site (arrow), CpG islands (green bar), presumed transcript (dotted lines) and miRNA positions (red bar) are shown. **D.** Most human and mouse miRNA promoters show evidence of H3K4me3 enrichment in multiple tissues.

Figure 3 Oct4, Sox2, Nanog and Tcf3 occupancy of miRNA promoters. A. Oct4 (blue), Sox2 (purple), Nanog (orange) and TCF3 (red) binding is shown at four murine miRNA genes as in Figure 1A. H3K4me3 enrichment in ES cells is indicated by shading across genomic region. Presumed transcripts are shown as dotted lines. Coordinates for the mmu-mir-290-295 cluster are derived from NCBI build 37. **B.** Oct4 ChIP enrichment ratios (ChIP-enriched versus total genomic DNA) are shown across human miRNA promoter region for the hsa-mir-302 cluster. H3K4me3 enrichment in ES cells is indicated by shading across genomic region. **C.** Schematic of miRNAs with conserved binding by the core transcription factors in ES cells. Transcription factors are represented by dark blue circles and miRNAs are represented by purple hexagons. miRNAs from the miR-302 cluster and miR290-295 (mouse)/371-372(human) cluster are selectively expressed in ES cells (Houbaviy et al., 2003).

Figure 4 Regulation of Oct4/Sox2/Nanog/TCF3-bound miRNAs during differentiation. A. Pie charts showing relative contributions of miRNAs to the complete population of miRNAs in mES cells (red) , MEFs (blue) and neural precursors (NPCs,

green) based on quantification of miRNAs from by small RNA sequencing. A full list of the miRNAs identified can be found in Supplementary Table S6. **B.** Normalized frequency of detection of individual mature miRNAs whose primary transcripts are occupied by Oct4, Sox2, Nanog and Tcf3 in mouse. Red line in center and right panel show the level of detection in ES cells. **C.** Histogram of changes in frequency of detection. Changes for miRNAs whose primary transcripts are occupied by Oct4, Sox2, Nanog and Tcf3 in mouse are shown as bars (red for ES enriched, blue for MEF enriched and green for NPC enriched). The background frequency for non-occupied miRNAs is shown as a grey line.

Figure 5. Polycomb represses lineage-specific miRNAs in ES cells. A. Suz12 (light green) and H3K27me3 (dark green, Mikkelsen et al., 2007) binding are shown for two miRNA genes in murine ES cells. Predicted start sites (arrow), CpG islands (green bar), presumed miRNA primary transcript (dotted line) and mature miRNA (red bar) are shown. B. Expression analysis of miRNAs from mES cells based on quantitative small RNA sequencing. Cumulative distributions for PcG bound miRNAs (green line) and all miRNAs (grey line) are shown. C. Expression analysis of miRNAs bound by Suz12 in mES cells. Relative counts are shown for mES (red), NPCs (orange) and MEFs (yellow). miR-9 transcript levels were selectively induced in NPCs, while miR-218 and miR-34c were induced in MEFs. H3K27me3 (green line) was lost from the miR-9-1 and the miR-9-2 promoters in NPCs, while the promoters retained H3K4me3 (blue line) (Mikkelsen et al., 2007). H3K27me3 was lost at the miR-218 and miR-34c promoters in MEFs. D. Schematic of a subset of miRNAs bound by Suz12 in both mES and hES cells. Cells known to selectively express these miRNAs based on computation predictions (Farh et al., 2005) or experimental confirmation (Yi et al., 2006; Yi et al., 2008; Landgraf et al.,2007) are indicated. Transcription factors are represented by dark blue circles, and Suz12 by a green circle. miRNA gene promoters are represented by purple hexagons.

Figure 6. miRNA modulation of the gene regulatory network in ES cells. A. An incoherent feed-forward motif (Alon 2007) involving a miRNA repression of a

transcription factor target gene is illustrated (left). Transcription factors are represented by dark blue circles, miRNAs in purple hexagons, protein-coding gene in pink rectangles and proteins in orange ovals. Particular instances of this netowrks motif identified in ES cells, where signaling molecules or transcriptional regulators directly downstream of Oct4/Sox2/Nanog/Tcf3 are tuned or silenced by miRNAs maintained in ES cells by Oct4/Sox2/Nanog/Tcf3, are illustrated (right). B. Another incoherent feed-forward motif (Alon 2007) where a protein, encoded by a gene under the control of Oct4/Sox2/Nanog/Tcf3, inhibits the maturation of a primary miRNA transcript maintained in ES cells by Oct4/Sox2/Nanog/Tcf3, is illustrated (left). In ES cells, Lin28 blocks the maturation of primary Let-7g (Visiwanthan et al., 2008). *Lin28* and the Let-7g gene are occupied by Oct4/Sox2/Nanog/Tcf3. Also, noted by the purple dashed line is the Targetscan prediction (Grimson et al., 2007), that mature Let-7g would target Lin28 (right). c, A coherent feed-forward motif (Alon 2007) where a miRNA represses the expression of transcriptional repressor, which indirectly activates the expression of a gene maintained in ES cells by Oct4/Sox2/Nanog/Tcf3, is illustrated (left). This motif is found in ES cells, where mir-290-295 miRNAs repress Rbl2 indirectly maintaining the expression of *Dnmt3a* and *Dnmt3a*, which are also occupied at their promoters by Oct4/Sox2/Nanog/Tcf3 (right).

Figure 7. Multi-level regulatory network controlling ES cell identity. Updated map of ES cell regulatory circuitry is shown. Interconnected auto-regulatory loop is shown to the left. Active transcripts are shown at the top right, and PcG silenced transcripts are shown at the bottom. Transcription factors are represented by dark blue circles, and Suz12 by a green circle. Gene promoters are represented by red rectangles, gene products by orange circles, and miRNA promoters are represented by purple hexagons.

Figure 1

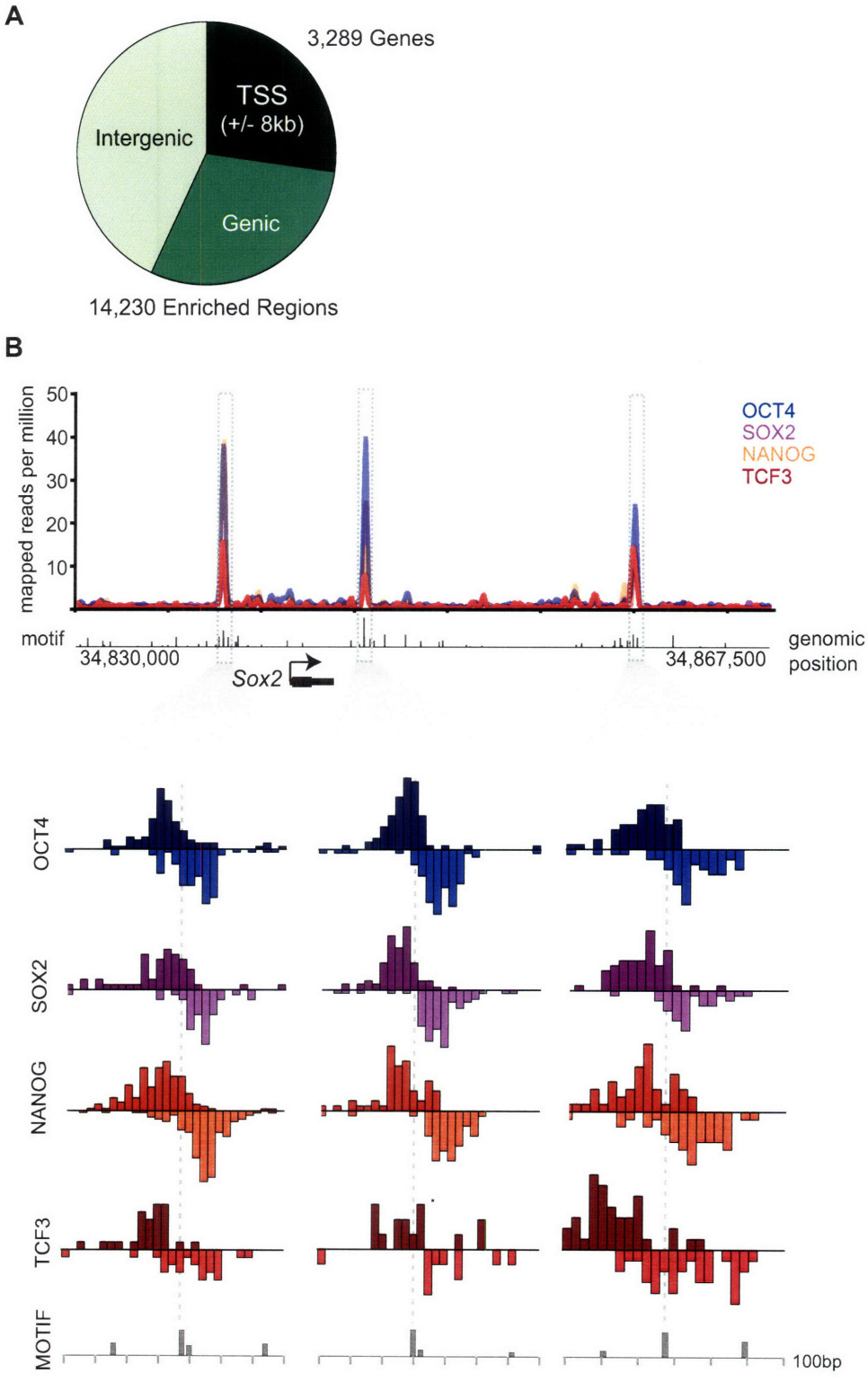

A
3,289 Genes
TSS
(+/- 8kb)
Intergenic
Genic
14,230 Enriched Regions
B
mapped reads per million
50
40
30
20
10
OCT4
SOX2
NANOG
TCF3
motif
34,830,000
34,867,500
genomic
position
Sox2
OCT4
SOX2
NANOG
TCF3
MOTIF
100bp

Figure 2

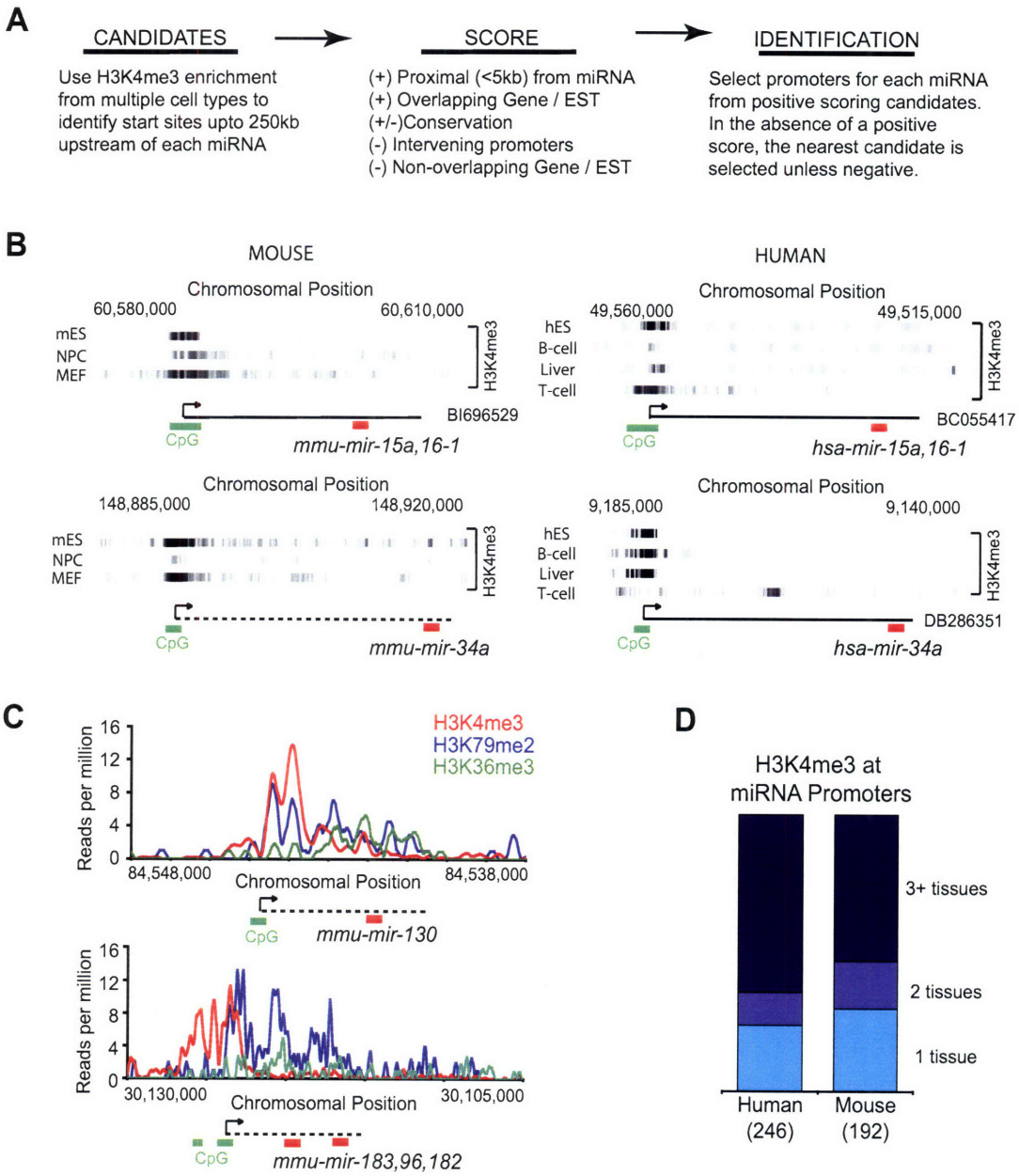

A
CANDIDATES
Use H3K4me3 enrichment from multiple cell types to identify start sites upto 250kb upstream of each miRNA
SCORE
(+) Proximal (<5kb) from miRNA
(+) Overlapping Gene / EST
(+/-)Conservation
(-) Intervening promoters
(-) Non-overlapping Gene / EST
IDENTIFICATION
Select promoters for each miRNA from positive scoring candidates. In the absence of a positive score, the nearest candidate is selected unless negative.
B
MOUSE
HUMAN
Chromosomal Position
60,580,000
60,610,000
mES
NPC
MEF
H3K4me3
BI696529
CpG
mmu-mir-15a,16-1
49,560,000
49,515,000
hES
B-cell
Liver
T-cell
BC055417
hsa-mir-15a,16-1
148,885,000
148,920,000
mmu-mir-34a
9,185,000
9,140,000
DB286351
hsa-mir-34a
C
H3K4me3
H3K79me2
H3K36me3
Reads per million
84,548,000
84,538,000
mmu-mir-130
30,130,000
30,105,000
mmu-mir-183,96,182
D
H3K4me3 at miRNA Promoters
3+ tissues
2 tissues
1 tissue
Human (246)
Mouse (192)

Figure 3

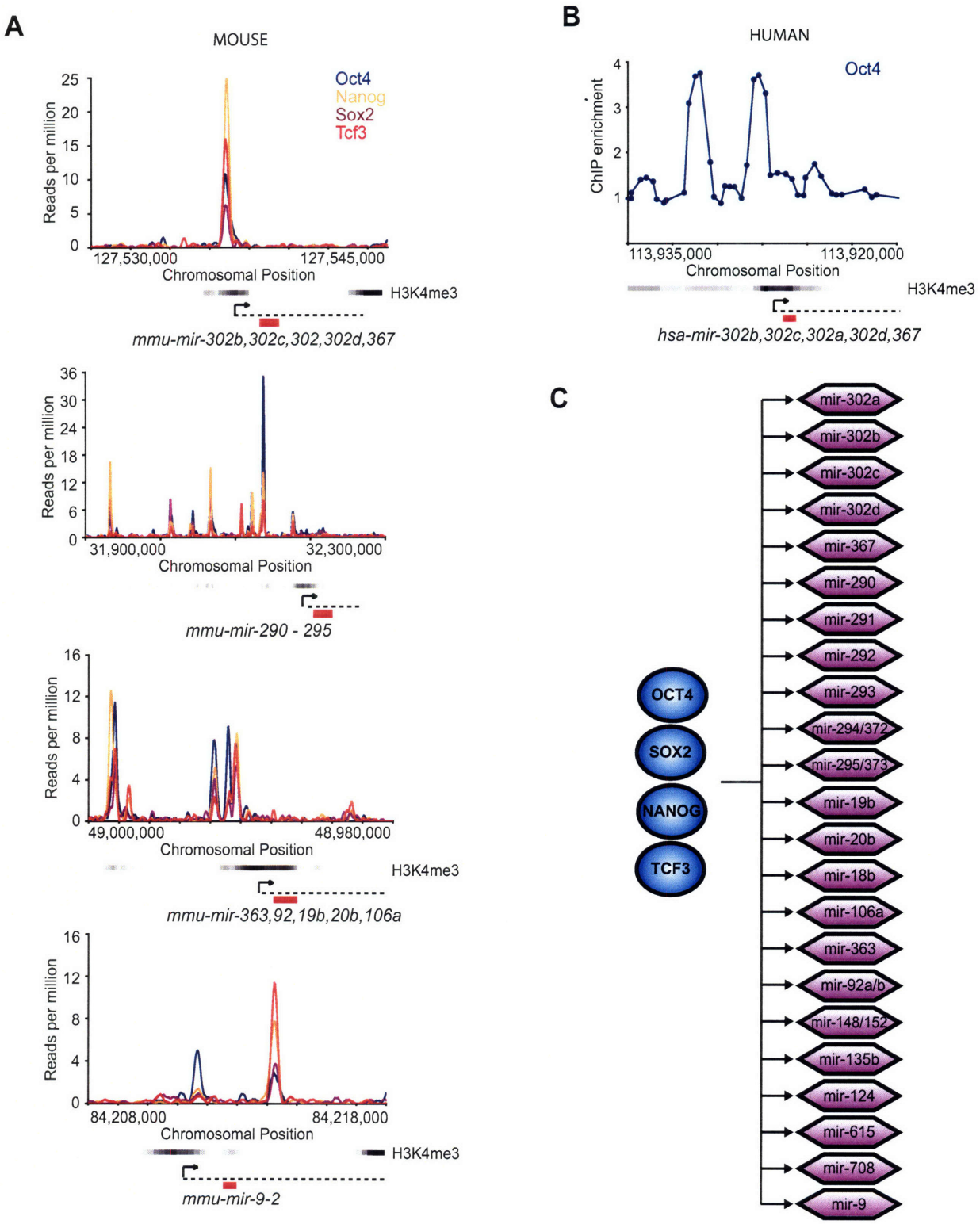

A
MOUSE
Oct4
Nanog
Sox2
Tcf3
Reads per million
Chromosomal Position
127,530,000
127,545,000
H3K4me3
mmu-mir-302b,302c,302,302d,367
31,900,000
32,300,000
mmu-mir-290 - 295
49,000,000
48,980,000
mmu-mir-363,92,19b,20b,106a
84,208,000
84,218,000
mmu-mir-9-2
B
HUMAN
Oct4
ChIP enrichment
113,935,000
113,920,000
hsa-mir-302b,302c,302a,302d,367
C
OCT4
SOX2
NANOG
TCF3
mir-302a
mir-302b
mir-302c
mir-302d
mir-367
mir-290
mir-291
mir-292
mir-293
mir-294/372
mir-295/373
mir-19b
mir-20b
mir-18b
mir-106a
mir-363
mir-92a/b
mir-148/152
mir-135b
mir-124
mir-615
mir-708
mir-9

Figure 4

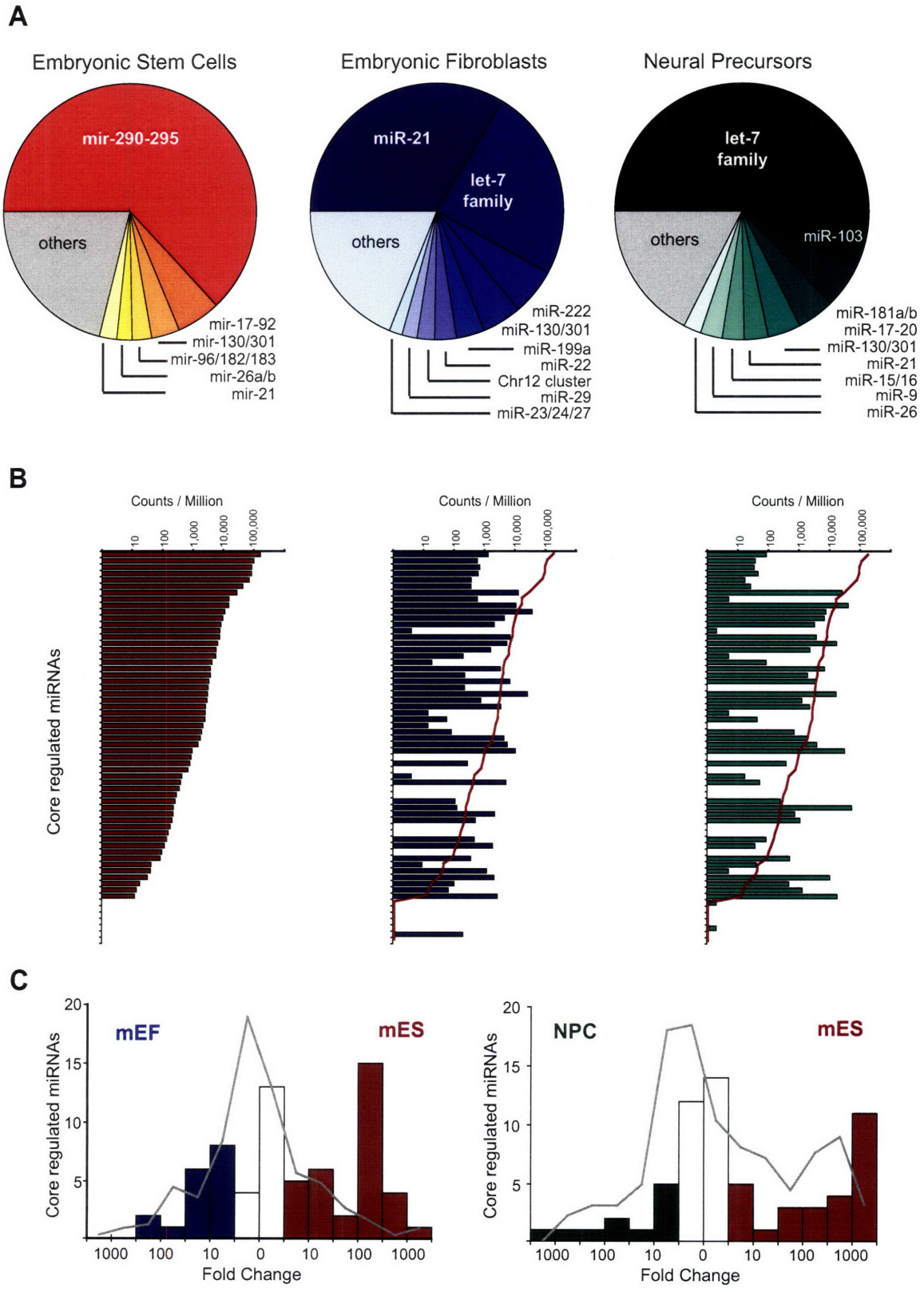

A
Embryonic Stem Cells
mir-290-295
others
mir-17-92
mir-130/301
mir-96/182/183
mir-26a/b
mir-21
Embryonic Fibroblasts
miR-21
let-7 family
others
miR-222
miR-130/301
miR-199a
miR-22
Chr12 cluster
miR-29
miR-23/24/27
Neural Precursors
let-7 family
others
miR-103
miR-181a/b
miR-17-20
miR-130/301
miR-21
miR-15/16
miR-9
miR-26
B
Counts / Million
10
100
1,000
10,000
100,000
Core regulated miRNAs
C
mEF
mES
NPC
mES
Core regulated miRNAs
20
15
10
5
1000
100
10
0
10
100
1000
Fold Change

Figure 5

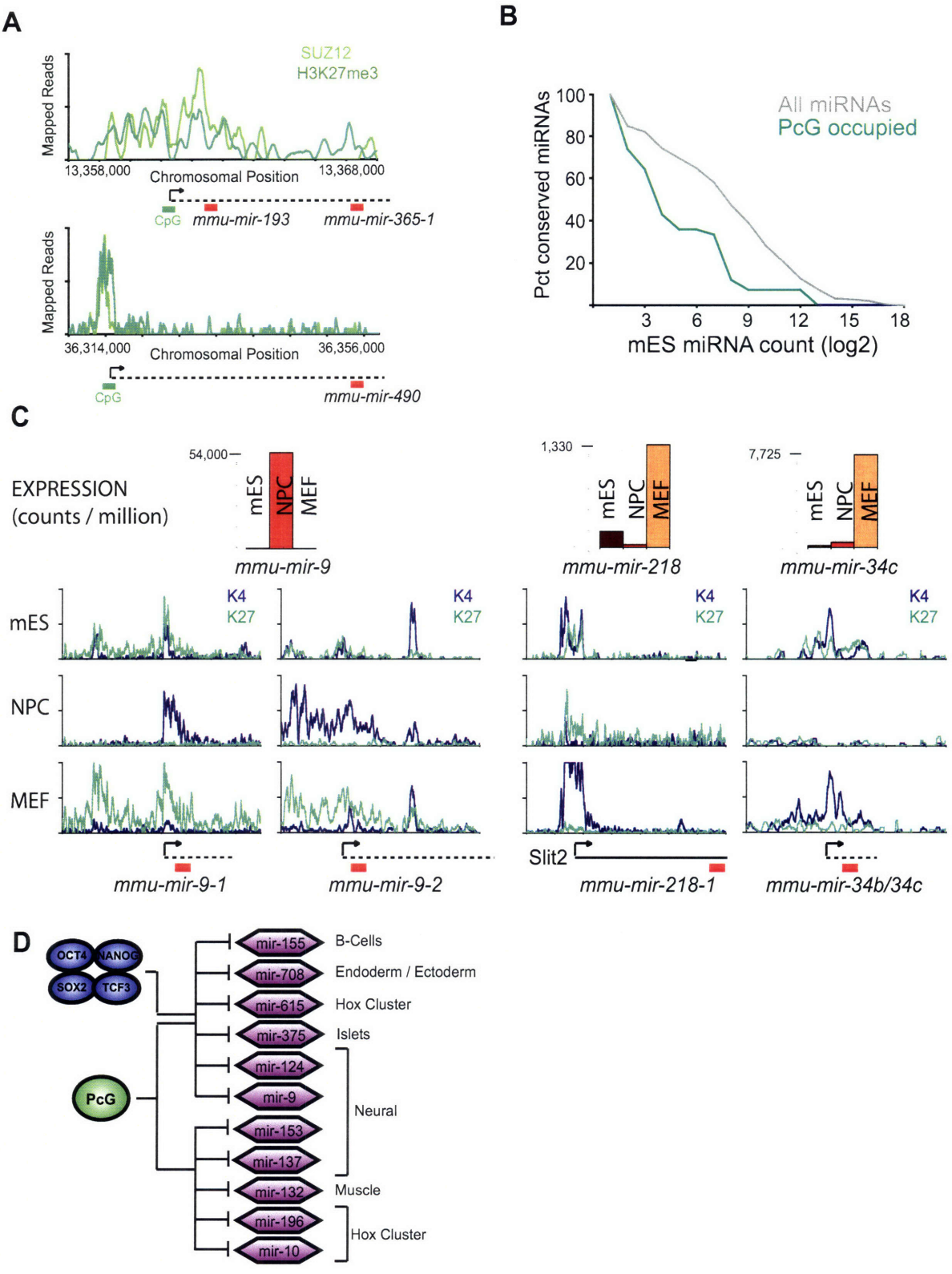

A
SUZ12
H3K27me3
Mapped Reads
13,358,000
Chromosomal Position
13,368,000
CpG
mmu-mir-193
mmu-mir-365-1
Mapped Reads
36,314,000
Chromosomal Position
36,356,000
CpG
mmu-mir-490
B
All miRNAs
PcG occupied
Pct conserved miRNAs
100
80
60
40
20
3
6
9
12
15
18
mES miRNA count (log2)
C
EXPRESSION
(counts / million)
54,000
mES
NPC
MEF
mmu-mir-9
1,330
mmu-mir-218
7,725
mmu-mir-34c
mES
NPC
MEF
K4
K27
mmu-mir-9-1
mmu-mir-9-2
Slit2
mmu-mir-218-1
mmu-mir-34b/34c
D
OCT4
NANOG
SOX2
TCF3
PcG
mir-155
mir-708
mir-615
mir-375
mir-124
mir-9
mir-153
mir-137
mir-132
mir-196
mir-10
B-Cells
Endoderm / Ectoderm
Hox Cluster
Islets
Neural
Muscle
Hox Cluster

Figure 6

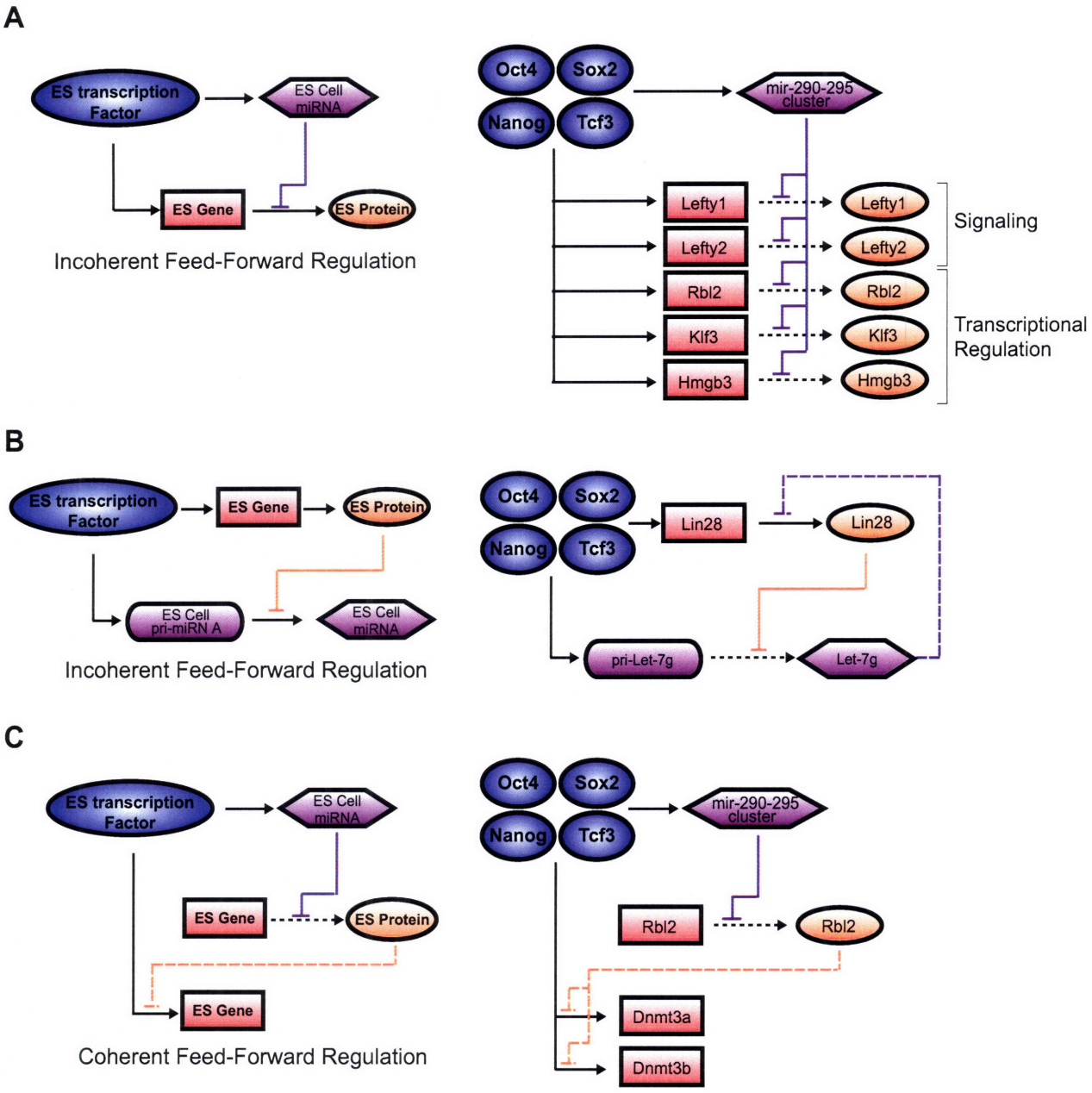
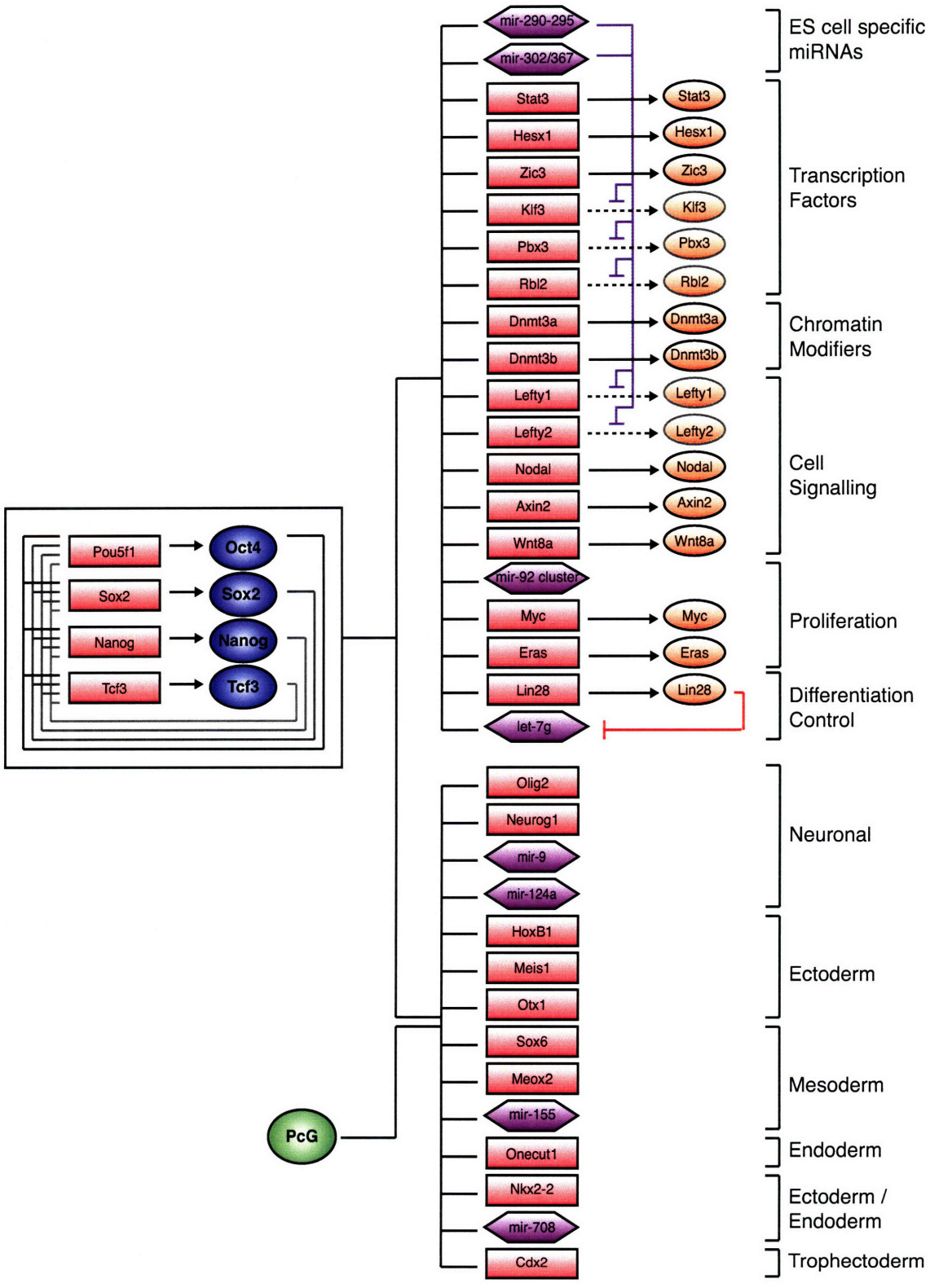

A
ES transcription Factor
ES Cell miRNA
ES Gene
ES Protein
Incoherent Feed-Forward Regulation
Oct4
Sox2
Nanog
Tcf3
mir-290-295 cluster
Lefty1
Lefty2
Rbl2
Klf3
Hmgb3
Lefty1
Lefty2
Rbl2
Klf3
Hmgb3
Signaling
Transcriptional Regulation
B
ES transcription Factor
ES Gene
ES Protein
ES Cell pri-miRN A
ES Cell miRNA
Incoherent Feed-Forward Regulation
Oct4
Sox2
Nanog
Tcf3
Lin28
Lin28
pri-Let-7g
Let-7g
C
ES transcription Factor
ES Cell miRNA
ES Gene
ES Protein
ES Gene
Coherent Feed-Forward Regulation
Oct4
Sox2
Nanog
Tcf3
mir-290-295 cluster
Rbl2
Rbl2
Dnmt3a
Dnmt3b

Figure 7


mir-290-295
mir-302/367
Stat3
Hesx1
Zic3
Klf3
Pbx3
Rbl2
Dnmt3a
Dnmt3b
Lefty1
Lefty2
Nodal
Axin2
Wnt8a
mir-92 cluster
Myc
Eras
Lin28
let-7g
Pou5f1
Sox2
Nanog
Tcf3
Oct4
Sox2
Nanog
Tcf3
PcG
Olig2
Neurog1
mir-9
mir-124a
HoxB1
Meis1
Otx1
Sox6
Meox2
mir-155
Onecut1
Nkx2-2
mir-708
Cdx2
ES cell specific miRNAs
Transcription Factors
Chromatin Modifiers
Cell Signalling
Proliferation
Differentiation Control
Neuronal
Ectoderm
Mesoderm
Endoderm
Ectoderm / Endoderm
Trophectoderm

Chapter 3

Wnt signaling promotes reprogramming of somatic cells to pluripotency

Alexander Marson*, Ruth Foreman*, Brett Chevalier*, Steve Bilodeau, Michael Kahn, Richard A. Young‡, Rudolf Jaenisch‡

*These authors contributed equally to this work

‡Corresponding authors

Personal Contribution to the Project

This chapter is the result of a close collaboration with Ruth Foreman and Brett Chevalier. This work was inspired by the work of Megan Cole, Sarah Johnstone, Michael Kagey and Rick Young (Cole et al., 2008). Rick suggested that I examine the role of Wnt signaling in iPS cell generation. I conceived of the initial experiments to test if Wnt signaling could promote reprogramming and conducted the cell culture experiments described here along with Ruth and Brett. I prepared the manuscript.

Summary

Somatic cells can be reprogrammed to induced pluripotent stem (iPS) cells by retroviral transduction of four transcription factors, Oct4, Sox2, Klf4 and c-Myc. While the reprogrammed pluripotent cells are thought to have great potential for regenerative medicine, genomic integrations of the retroviruses increase the risk of tumorigenesis. Thus, new approaches are required that circumvent the need for retroviruses to induce reprogramming. We report here that soluble Wnt3a promotes the generation of iPS cells in the absence of c-Myc retrovirus. Furthermore, chemicals that target the Wnt pathway are able to modulate the efficiency of iPS cell formation. These data demonstrate that signal transduction pathways and transcription factors can act coordinately to reprogram differentiated cells to a pluripotent state, and suggest the first synthetic small molecule that promotes the reprogramming process.

Introduction

Retroviral transduction of the four nuclear factors, Oct4, Sox2, Klf4 and c-Myc has been used to reprogram mouse and human somatic cells into induced pluripotent stem (iPS) cells that are indistinguishable from embryo derived embryonic stem (ES) cells (Takahashi and Yamanaka, 2006; Okita et al., 2007; Wernig et al., 2007; Maherali et al., 2007; Takahashi et al., 2007). More recent experiments have demonstrated that other combinations of factors also can induce iPS cells (Yu et al., 2007; Park et al., 2007; Nakagawa et al., 2008). While retroviruses afford an efficient means of over-expressing these transcription factors for an extended period of time, they preclude the clinical utility of the presently derived iPS cells because of the danger of insertional mutagenesis and the high risks of tumorigenesis associated with the transduction of oncogenes such as c-Myc (Okita et al., 2007). Recently, iPS cells have been generated without c-Myc retrovirus ($Myc^{[-]}$) and chimeric mice arising from these $Myc^{[-]}$ cells have a significantly lower incidence of tumors than mice generated from Myc infected cells (Nakagawa et al., 2008; Wernig et al., 2008). While they carry a reduced risk of causing cancer, $Myc^{[-]}$ iPS cells can only be generated at extremely low efficiencies and with delayed kinetics relative to Myc transduced cells (Nakagawa et al., 2008; Wernig et al., 2008). To realize the therapeutic potential of iPS cells (Hanna et al., 2007), there is great interest in identifying methods that can replace genetic manipulations of the donor cells and substitute for the retroviruses and oncogenes currently used for the induction of the reprogramming process. Here we identify a defined small molecule that could promote the generation of $Myc^{[-]}$ iPS cells.

Knowledge of the transcriptional regulatory circuitry controlled by Oct4, Sox2, as well as Nanog, which promotes reprogramming of human cells (Yu et al., 2007), offers important clues to the mechanism of reprogramming (Boyer et al., 2005; Loh et al., 2006; Jaenisch and Young, 2008; Kim et al., 2008). These ES transcription factors control a large number of genes, both promoting the transcription of genes required for self-renewal, and repressing the transcription of genes that would cause cellular differentiation. Additionally, Oct4, Sox2 and Nanog all bind to their own promoters, as well as the promoters of the genes encoding the two other factors. This interconnected auto-regulatory loop likely contributes to the stability of the pluripotent state and suggests that exogenous transcription factors directly promote the transcription of the endogenous genes to induce reprogramming. Klf4 was more recently shown to occupy the promoters of an overlapping set of genes, including *Oct4*, *Sox2* and *Nanog* (Jiang et al., 2008; Kim et al., 2008). During iPS cell formation, once key genes such as *Oct4* and *Nanog*, whose promoters are methylated and silenced in somatic cells, are induced by the reprogramming factors (Okita et al., 2007; Wernig et al. 2007; Maherali et al., 2007), the transcriptional program of pluripotency is reactivated. At this point, expression of exogenous reprogramming factors becomes dispensable because the endogenous transcriptional regulatory circuitry is able to maintain itself in iPS cells under normal ES cell culture conditions (Wernig et al., 2007; Brambrink et al., 2008; Stadtfeld et al., 2008).

Naturally occurring signaling molecules that modulate the expression of ES cell transcription factors are promising candidates for soluble factors that enhance reprogramming. The Wnt signaling pathway contributes to the maintenance of

pluripotency in mouse and human ES cells (Sato et al., 2004; Ogawa et al., 2006; Singla et al., 2006; Cai et al., 2007), as well as the self-renewal of undifferentiated adult stem cells in multiple tissues (Reya and Clevers, 2005). In addition, initial studies of iPS generation suggested that constitutively active ß-catenin, another downstream component of the Wnt pathway, might promote reprogramming of fibroblasts to pluripotency (Takahashi and Yamanaka, 2006).

Recent genome-wide location analysis of Tcf3, one of the key transcriptional regulators downstream of the Wnt pathway in embryonic stem cells, has revealed that this factor co-occupies almost all promoter regions occupied by ES cell specific transcription factors, including Oct4 and Nanog, and is bound to the promoters of *Oct4, Sox2, Nanog* and *c-Myc* in ES cells (Cole et al, 2008). This demonstrates that the Wnt signaling pathway directly converges on the core regulatory circuitry of ES cell transcription, consistent with the observed effects of this pathway on pluripotency. Since the Wnt pathway is so intimately connected to the core circuitry of pluripotency, we hypothesized that the stimulation of the pathway using soluble factors could modulate the efficiency of inducing pluripotency in somatic cells.

Here we examine the influence of Wnt stimulation on the reprogramming of murine fibroblasts. Consistent with the observation that downstream components of the Wnt signaling pathway converge on the core regulatory circuitry of pluripotency, we find that Wnt stimulation is able to increase the efficiency of reprogramming by nuclear factors in the absence of *c-Myc* retrovirus. Furthermore, we find chemicals that are reported to modulate the ß-catenin pathway can contribute both positively and negatively

to reprogramming. Notably, ICG-427 appears to enhance the efficiency of generating iPS cells without exogenous c-Myc under standard ES cell culture conditions.

Results

Wnt3a conditioned medium promotes reprogramming of MEFs to an ES-like state

To determine the effect of Wnt3a on reprogramming, we infected mouse embryonic fibroblasts (MEFs) that harbor a doxycyline (DOX)-inducible *Oct4* cDNA (Hochedlinger et al., 2005) with inducible lentiviral vectors encoding Sox2, Klf4, and c-Myc (Brambrink et al., 2008). These cells also contain a G418 resistance cassette in the endogenous *Oct4* locus allowing for drug selection of iPS cells (Wernig et al., 2007).

Four factor expression was induced by addition of DOX in cells cultured in the presence or absence of Wnt3a conditioned medium (Wnt3a-CM) (Willert et al., 2003), G418 selection was initiated after 5 days and the number of drug resistant colonies was determined 24 days after induction (**Figure 1a**). **Figure 1b** shows that the total number of drug resistant colonies was increased more than 7 fold when the cells were cultured in Wnt3a-CM. We also noted that the drug resistant colonies were larger and more ES-cell like by morphology when cultured in Wnt3a-CM than in ES cell medium (**Figure 1c**). Furthermore, the colonies that appeared in Wnt3a-CM with G418 selection initiated on Day 5 could be further propagated, in contrast to the small colonies derived in standard ES cell medium.

Given that Wnt3a-CM had a positive effect on reprogramming in concert with the four transcription factors, we next examined if Wnt3a-CM could promote iPS cell generation in MEFs lacking transduction of any of the factors. In parallel experiments, fibroblasts were transduced with subsets of the transcription factors and observed in the presence and absence of Wnt3a-CM (**Figures 1d and 1e**). No resistant colonies formed

in the absence of Oct4 or Klf4 infection. One colony was observed in the absence of Sox2 retrovirus, but this colony could not be not be further propagated in ES cell culture conditions. In contrast, in the presence of Wnt3a-CM multiple robust G418-resistant colonies formed in the absence of c-Myc in cells transduced only with Oct4, Sox2 and Klf4 (**Figures 1d and 1e**). Similar to colonies from MEFs transduced with all four factors, these iPS lines could be propagated in standard ES cell media without further selection and retained ES cell morphology. In replicate experiments, G418-resistant colonies were formed occasionally with no c-Myc transduction in the absence Wnt3a-CM. However, consistent with published reports (Nakagawa et al., 2008; Wernig et al., 2008), these colonies were rare. In the following, iPS cells generated with only three factors and without c-Myc will be designated as $Myc^{[-]}$ iPS cells.

To more closely examine the effects of Wnt3a-CM treatment on the reprogramming process, Oct4/Sox2/Klf4 and Oct4/Sox2/Klf4/c-Myc over-expressing MEFs were cultured with and without Wnt3a-CM, and G418 selection was initiated at different times after DOX addition. **Figure 1f** shows that when three factor over-expressing cells were cultured in Wnt3a-CM medium about 3 fold more $Myc^{[-]}$ iPS colonies appeared when G418 was added at day 5 and about 20 fold more colonies when G418 was added at 10 days after induction as compared to cultivation in ES cell medium **(Figure 1f, left panel).** Wnt3a-CM medium also increased the number of drug resistant colonies after induction of all four factors, though the fold increase was less pronounced than in three factor transduced cells (**Figure 1f, right panel**). These results indicate that Wnt3a-CM increased the number of drug resistant colonies in both three factor and four

factor induced cells, with the most pronounced effect on three factor over-expressing cells with selection applied at the later time point.

Rare Myc[-] iPS clones have been shown previously to arise without any genetic selection (Nakagawa et al., 2008; Wernig et al., 2008). We tested whether Wnt3-CM would also aid in the generation of Myc[-] iPS cells in the absence of selection for *Oct4* reactivation. For this, cells with GFP driven by the endogenous *Oct4* promoter were utilized (Meissner et al., 2007). Oct4/Sox2/Klf4 infected cells with and without Wnt3a-CM treatment were analyzed for GFP expression by flow cytometry at days 10, 15 and 20 after DOX induction. No GFP positive cells were present with or without Wnt3a-CM treatment on day 10 or day 15. By day 20 a small population of GFP expressing cells was detected in cells cultured in Wnt3a-CM, but not in standard ES cell medium (**Figure 1g**). The Wnt3a-CM exposed cultures formed GFP expressing colonies with morphology typical for ES or iPS cells (**Figure 1h**). However, unlike four factor transduced cells, which usually form a highly heterogeneous population of cells when propagated without selection, the Oct4/Sox2/Klf4/Wnt3a-CM colonies appeared homogenously ES-like, similar to previously reported Myc[-] iPS clones (Nakagawa et al., 2008).

Developmental potential of Wnt3a-CM-derived Myc[-] iPS cells

Several assays were performed to characterize the developmental potential of Myc[-] iPS cells derived with Wnt3a-CM treatment. Immunocytochemistry confirmed the expression of markers of pluripotency, including the nuclear factor Nanog (**Figures 2a and 2b**), and the surface glycoprotein SSEA1 (**Figures 2c and 2d**). Functional assays confirmed that, like ES cells, these iPS cells were pluripotent. When injected into SCID

mice subcutaneously, the Myc[-] iPS cells gave rise to teratomas with histological evidence of cells differentiating into all three germ layers (**Figures 2e, 2f and 2g**). More importantly, Myc[-] iPS cells derived with Wnt3a-CM treatment contributed to the formation of somatic tissues as well as the germline in chimeric mice (**Figure 2h**). These results indicate that Wnt3a-CM treated Myc[-] clones are pluripotent cells that are morphologically and functionally indistinguishable from ES cells.

Small molecule modulators of Wnt pathway affect reprogramming

Screens for chemicals that modulate the effects of Wnt signaling have identified compounds that selectively inhibit either the differentiation promoting effects of the Wnt pathway (ICG-427) or the self-renewal promoting effects of the pathway (ICG-001) (Teo et al., 2005; McMillan and Kahn, 2005 and unpublished data). These small molecules were attractive candidates to test for their effects on reprogramming because of the effect we observed with Wnt3a-CM.

To quantify the effects of Wnt3a-CM, triplicate experiments were performed on Oct4/Sox2/Klf4-inducible, G418 selectable MEFs (**Figure 3b**). Myc[-] G418 resistant colonies were extremely scarce in standard ES cell culture conditions, with 0-3 colonies forming on each ten centimeter plate. However, when G418 selection was initiated in Wnt3a-CM at day 15 after factor induction, about 20 fold more drug resistant colonies formed, confirming the efficacy of this medium in enhancing reprogramming. It should be noted that conditioned medium from control fibroblasts lacking Wnt3a over-expression also caused a marked increase in the number of G418-resistant colonies,

which was roughly half of the Wnt3a-CM effect. Thus, normal fibroblasts may secrete factors, perhaps including Wnt3a, that promote reprogramming.

We then examined if the compounds described above modulated the Wnt3a-CM dependent effects on iPS cell formation. **Figure 3b** (right columns) shows that ICG-001 completely inhibited the effect of Wnt3a-CM on $Myc^{[-]}$ iPS formation. In contrast, ICG-427, which has been shown to contribute to maintenance of pluripotency in ES cells (unpublished data), did not have a significant effect on the number of G418 resistant colonies growing in Wnt3a-CM, although there upward trend in colony number was observed.

The effects of Wnt3a-CM and the same Wnt pathway inhibitors were also examined in MEFs over-expressing all four reprogramming factors, including c-Myc (**Figure 3c**). High numbers of G418 resistant colonies were observed in standard ES cell medium with four factor reprogrammed cells and in Wnt3a-CM, with only a subtle increase in the number of colonies in Wnt3a-CM. In contrast to the dramatic effect of ICG-001 on $Myc^{[-]}$ cells, the compound had only a subtle effect on the number of G418 colonies in c-Myc transduced cells, and a relatively high number of resistant colonies were observed under these conditions. These results are consistent with the notion that Wnt3a can, at least in part, replace the role of c-Myc in reprogramming.

Because the Wnt3a-CM medium may contain, in addition to Wnt3a, unknown factors that could enhance reprogramming, we used the selective Wnt inhibitors to investigate the role of the Wnt pathway in defined ES cell medium. **Figure 3d** shows that the addition of ICG-427 significantly increased efficiency of reprogramming in 3

factor over-expressing cells (formation of Myc$^{[-]}$ G418 resistant colonies). In contrast, ICG-001 treatment slightly lowered the number of Myc$^{[-]}$ G418 resistant colonies as compared to that observed in standard ES medium. Thus, even without conditioned medium, the chemical modulation of the Wnt pathway had a reproducible positive effect on the induction of pluripotency in somatic cells. We thus identify ICG-427 as a defined small molecule that appears to promote the reprogramming process.

Discussion

Our data establish that the Wnt signaling pathway can contribute to the generation of iPS cells. We observed that reprogramming by four nuclear factors is accelerated in the presence of Wnt3a-CM and that at early time-points after induction robust iPS colonies were observed only in the presence of Wnt3a-CM. More importantly, Wnt3a-CM significantly accelerated and enhanced the efficiency of reprogramming without transduction of c-Myc. While in standard ES cell conditions rare $Myc^{[-]}$ iPS appeared only late after infection (Nakagawa et al., 2008; Wernig et al., 2008), drug resistant or GFP positive iPS colonies were detected already at 5 days or 20 days, respectively, after factor induction in Wnt3a-CM. When selection was initiated at later time points, Wnt3a had an even more pronounced effect on the formation of $Myc^{[-]}$ iPS cells, increasing the number of G418 resistant colonies ~20 fold. While the details of the kinetics of the effects of Wnt on reprogramming need to be explored further in future experiments, these data demonstrate that Wnt stimulation is an effective substitute for Myc-encoding retrovirus in aiding the action of Oct4, Sox2 and Klf4 (**Figure 4**). Although Wnt stimulation of three factor over-expressing cells produced fewer G418 resistant colonies than four factor over-expression, the $Myc^{[-]}$ iPS colonies that were formed were more homogeneously ES-like in appearance and growth characteristics than cells over-expressing four factors. It is possible that more optimal concentrations of Wnt3a or small molecule modulators of this pathway may yield as many colonies as the addition of Myc retrovirus.

Recent reports have demonstrated that Myc is dispensable for inducing pluripotency in MEFs (Nakagawa et al., 2008; Wernig et al., 2008), establishing Oct4, Sox2 and Klf4 as a core set of reprogramming factors, although a slightly different combination of factors also functions in human cells (Yu et al., 2007). Epigenetic reprogramming of MEFs with defined factors is a gradual process, where transcription factors progressively re-establish the core circuitry of pluripotency over the course of weeks (Brambrink et al, 2008). Myc accelerates this process (Nakagawa et al., 2008; Wernig et al., 2008), conceivably because it enhances cell proliferation allowing the changes to unfold more rapidly. Additionally, Myc is suspected to have widespread effects on chromatin state (Knoepfler, 2008; Kim et al., 2008) and could facilitate the productive binding of the core reprogramming factors to their appropriate genomic targets (Jaenisch and Young, 2008). Wnt stimulation seems to at least partially be able to replace the requirement for c-Myc in reprogramming consistent with the notion that Wnt stimulation and c-Myc act in the same signaling pathway.

The Wnt signaling pathway has been shown to connect directly to the core transcriptional regulatory circuitry of ES cells, suggesting a mechanism by which this pathway could directly promote the induction of the pluripotency in the absence of c-Myc transduction (**Figure 4**). In ES cells, Tcf3 occupies and regulates the promoters of *Oct4*, *Sox2* and *Nanog* (Cole et al., 2008). In MEFs, these endogenous pluripotency transcription factors are silenced. During reprogramming, as exogenous Oct4, Sox2 and Klf4 contribute to the reactivation of the endogenous pluripotency factors, Wnt signaling could directly potentiate the effect of these transcription factors, as it does in ES cells (Cole et al., 2008). Additionally, Wnt could serve to activate endogenous *c-Myc* directly,

thereby substituting for exogenous c-Myc. Indeed, *c-Myc* is a well-established target of the Wnt pathway in colorectal cancer cells (He et al., 1998). In ES cells, Tcf3 occupies the *c-Myc* promoter, and Wnt3a positively contributes to expression of the gene (Cole et al., 2008). The fact that enforced over-expression of c-Myc counteracts the negative effect of the Wnt inhibitor ICG-001 on the reprogramming process suggests that Wnt stimulation could be acting upstream endogenous Myc. Wnt-induced effects on cell proliferation, mediated by c-Myc or other endogenous proliferation factors, could help to accelerate the sequential process of forming $Myc^{[-]}$ iPS colonies. In these models, which are not mutually exclusive, the small molecule ICG-427 modulates the Wnt/b-catenin signaling pathway to promote the reprogramming process, while ICG-001 impedes the process (**Figure 4**).

Although the finding still needs to be reproduced, the use of ICG-427 could be a critical first step towards generating iPS with chemical cues instead of retroviral transduction of transcription factors, which harbor significant risks of oncogenic transformation. Wnt modulators, including ICG-427, could provide a means of boosting the efficiency of generating iPS cells without Myc over-expression, which otherwise only form at extremely low efficiency with delayed kinetics. These compounds will likely prove useful as chemicals, or other transient cues, are identified that substitute for the remaining three reprogramming retroviruses. Finally, this provides an important proof of principal that chemical manipulation of the core circuitry controlling cellular identity can significantly contribute to the directed alteration of cell state. Here, based on knowledge that the Wnt signals connect directly to this core circuitry in ES cells, we provide functional evidence that small molecule modulation of this pathway regulates the

induction of pluripotency in somatic cells. Future efforts to manipulate other signaling pathways and other core regulators of cell identity that control the pluripotent state will likely further help to alleviate the current reliance on retroviruses in the generation of iPS cells.

Experimental Procedures

Cell Culture

V6.5 (C57BL/6-129) murine ES cells and iPS cells were grown under typical ES conditions on irradiated mouse embryonic fibroblasts (MEFs). Transgenic MEFs used in the infections with DOX-inducible lentiviruses (Brambrink et al., 2008) were harvested at 13.5dpc and selected on 2ug/ml puromycin from embryos after blastocyst injection of Oct4-IRES-GFPneo/Oct4-inducible ES cells (Wernig et al., 2007) or harvested from F1 matings between R26-M2rtTA mice (Beard et al., 2006) and Oct4-GFP mice (Meissner et al., 2007). Wnt3a conditioned media and control conditioned media was generated according to standard protocols (ATCC) (Willert et al., 2003). Wnt modulators ICG-427 and ICG-001 were dissolved in DMSO to a stock concentration of 0.1M. The final, working concentration of all 3 inhibitors was 4uM.

Viral Transduction

Tetracycline inducible lentiviral constructs expressing the cDNAs for Oct4, Klf-4, Sox2 and c-Myc were used as previously described (Brambrink et al., 2008). Virus was prepared by transfecting HEK293T cells with a mixture of viral plasmid and packaging constructs expressing the viral packaging functions and the VSV-G protein. Medium was replaced 24 hours after transfection and viral supernatants were collected at 48 hours and 72 hours. After filtration, supernatants were pooled and MEFs were incubated with viral supernatants and fresh media at a ratio of 1:1 for 24 hours and subsequently cultured in

ES-cell medium, conditioned medium or in medium supplemented with chemical inhibitors.

Immunostaining and Antibodies

Cells were stained as described previously (Wernig et al., 2007). Antibodies against Nanog (Bethyl) and SSEA1 (R&D systems, Minneapolis, MN) were used according to supplier recommendations.

Teratoma Formation

Teratoma formation was assayed as previously described (Hochedlinger et al., 2005). Briefly, cells were trypsinized and $5x10^5$ cells were injected subcutaneously into SCID mice. After 14-21 days, teratomas were dissected, fixed in 10% phosphate-buffered formalin overnight and subsequently embedded in paraffin wax using a Tissue-Tek VIP embedding machine (Miles Scientific, Naperville, IL) and a Thermo Shandon Histocenter 2 (Thermo Fisher Scientific, Waltham, MA). Sections were cut at a thickness of 2 μm using a Leica RM2065 (Leica, Wetzlar, Germany) and stained with hematoxylin and eosin (Hochedlinger et al., 2005).

Blastocyst Injection

Injections of iPS cells into Balb/c host blastocysts were carried out as previously described (Beard et al., 2006).

References

Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis *44*, 23-28.

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

Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell *2*, 151-159.

Cai, L., Ye, Z., Zhou, B. Y., Mali, P., Zhou, C., and Cheng, L. (2007). Promoting human embryonic stem cell renewal or differentiation by modulating Wnt signal and culture conditions. Cell Res *17*, 62-72.

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

Hanna, J., Wernig, M., Markoulaki, S., Sun, C. W., Meissner, A., Cassady, J. P., Beard, C., Brambrink, T., Wu, L. C., Townes, T. M., and Jaenisch, R. (2007). Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. Science *318*, 1920-1923.

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

Hochedlinger, K., Yamada, Y., Beard, C., and Jaenisch, R. (2005). Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. Cell *121*, 465-477.

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

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

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

Knoepfler, P. S. (2008). Why Myc? An unexpected ingredient in the stem cell cocktail. Cell Stem Cell *2*, 18-21.

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

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

McMillan, M., and Kahn, M. (2005). Investigating Wnt signaling: a chemogenomic safari. Drug Discov Today *10*, 1467-1474.

Meissner, A., Wernig, M., and Jaenisch, R. (2007). Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. Nat Biotechnol *25*, 1177-1181.

Muroyama, Y., Kondoh, H., and Takada, S. (2004). Wnt proteins promote neuronal differentiation in neural stem cell culture. Biochem Biophys Res Commun *313*, 915-921.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol *26*, 101-106.

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

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313-317.

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

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

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

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A. H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through

activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med *10*, 55-63.

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

Stadtfeld, M., Maherali, N., Breault, D. T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *in press*.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

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

Teo, J. L., Ma, H., Nguyen, C., Lam, C., and Kahn, M. (2005). Specific inhibition of CBP/beta-catenin interaction rescues defects in neuronal differentiation caused by a presenilin-1 mutation. Proc Natl Acad Sci U S A *102*, 12171-12176.

Wernig, M., Meissner, A., Cassady, J. P., and Jaenisch, R. (2008). c-Myc is dispensable for direct reprogramming of mouse fibroblasts. Cell Stem Cell *2*, 10-12.

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

Willert, K., Brown, J. D., Danenberg, E., Duncan, A. W., Weissman, I. L., Reya, T., Yates, J. R., 3rd, and Nusse, R. (2003). Wnt proteins are lipid-modified and can act as stem cell growth factors. Nature *423*, 448-452.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Acknowledgements

We thank members of the Jaenisch, Young and Kahn laboratories, especially Jamie Newman, Megan Cole and Tony Lee for discussions and critical review of the manuscript. We are grateful to Jamie Newman, Megan Cole, and Sarah Johnstone for sharing data prior to publication and for Wnt3a conditioned media, and to Seshamma Reddy, Jessica Dausman, Ruth Flannery and Betty Zhou for technical assistance. We also thank Tom DiCesare for graphic assistance. This work was supported by NIH grants 5-RO1-HDO45022, 5-R37-CA084198, and 5-RO1- CA087869 to RJ and by NIH grant HG002668 and a grant from the Whitehead Institute to RY.

Figure Legends

Figure 1 Wnt3a promotes epigenetic reprogramming. **a.** Schematic representation of the experimental time-line. MEFs were infected with DOX-inducible lentivirus, split into cultures with and without Wnt3-CM treatment (Willert et al., 2003), and then induced with DOX (day 0). G418 selection was initiated at fixed time points after induction amd Wnt3a-CM treatment was maintained for 7 days of selection. DOX and G418 were maintained until resistant colonies were assessed. **b.** G418-resistant colony counts from MEFs overexpressing Oct4/Sox2/Klf4/c-Myc in standard ES cell media or with Wnt3a-CM treatment. **c.** Phase images of G418 resistant colonies formed with and without Wnt3a-CM treatment. **d.** G418-resistant colony counts from MEFs infected with different combinations of reprogramming factors in the presence and absence of Wnt3a-CM. G418 resistant colonies emerged without c-Myc transduction in the presence of Wnt3a-CM. **e.** Phase image of $Myc^{[-]}$ G418 resistant colony formed with Wnt3a-CM treatment. In this experiment, no colonies were observed for $Myc^{[-]}$ cells in the absence of Wnt3a-CM. **f.** G418-resistant colony counts from MEFs over-expressing Oct4/Sox2/Klf4 ($Myc^{[-]}$) or Oct4/Sox2/Klf4/c-Myc ($Myc^{[+]}$) in the presence (red bars) and absence (gray bars) of Wnt3a-CM. G418 selection was initiated on day 5 or day 10 post-induction as indicated and colonies (in a $32cm^2$ area) were assessed on day 20. **g.** Scatter plots comparing GFP intensity to autofluoresence, using flow cytommetry, in *Oct4*-GFP cells on day 20 post-induction of Oct4/Sox2/Klf4, reveal a GFP expressing population of cells (indicated with an arrow) only with Wnt3a-CM treatment. **h.** Phase image of GFP expressing $Myc^{[-]}$ cells derived with Wnt3a-CM treatment and without any genetic selection.

Figure 2 Induction of Pluripotency in Wnt Stimulated cells. a-d. Immunostaining reveals induction of pluripotency markers, Nanog (**a**,**b**) and SSEA-1(**c**,**d**) in Wnt3a-CM treated $Myc^{[-]}$ cells. **e-g.** Wnt3a-CM treated $Myc^{[-]}$ lines formed teratomas when injected into SCID mice subcutaneously. Teratomas from Oct4/Sox2/Klf4/Wnt3aCM iPS lines showed evidence of differentiated cells of three germ layers similar to teratomas formed from V6.5 mES injections. Arrows indicated neural tissue in (**e**), cartilage in (**f**), and

endodermal cells in (**g**), **h**. Oct4/Sox2/Klf4/Wnt3aCM iPS lines derived without selection gave rise to chimeric mice (as shown on the left) with agouti coat color and pigmented eyes (in contrast to wild type Balb/c mouse, right) providing evidence of contribution to somatic cells. Coat color of offspring confirmed that the Oct4/Sox2/Klf4/Wnt3aCM iPS line generated here is germline-competent (data not shown).

Figure 3 Small molecule modulators of Wnt pathway affect reprogramming a. Counts are shown for G418 resistant colonies in Oct4/Sox2/Klf4 over-expressing MEFs cultured in ES cell media, MEF conditioned media, Wnt3a over-expressing conditioned media, and Wnt3a over-expressing conditioned media with ICG-001 or ICG-427. Selection was initiated on day 15 post-induction, and colonies were assessed on day 28. Mean number of counts from triplicate experiments is displayed with error bars indicating S.D. **b**. Counts are shown for G418 resistant colonies (in a $32cm^2$ area) in Oct4/Sox2/Klf4/c-Myc over-expressing MEFs cultured in ES cell media, Wnt3a over-expressing conditioned media, and Wnt3a over-expressing conditioned media with ICG-001 or ICG-427. Selection was initiated on day 10 post-induction, and colonies were assessed on day 20. **c**. Counts are shown for G418 resistant colonies in Oct4/Sox2/Klf4 infected MEFs cultured in standard ES cell media alone or with ICG-001 or ICG-427. These MEFs were co-infected with dsRed as a control for viral titer. Selection was initiated on day 15 post-induction, and colonies were assessed on day 28. Mean number of counts from triplicate experiments is displayed with error bars indicating S.D.

Figure 4 Model of Wnt stimulation and induction of pluripotency. Wnt stimulation promotes the formaton of iPS cells in the absence of c-Myc transduction. This could be due to: i) direct regulation by the Wnt pathway of key endogenous pluripotency factors, such as *Oct4*, *Sox2* and *Nanog* as suggested by genomic studies in ES cells (Cole et al., 2008), ii) Wnt pathway-induced activation of endogenous *Myc* (He et al., 1998; Cole et al., 2008), or other cell proliferation genes, accelerating the sequential process of forming

iPS colonies. The small molecule ICG-427 modulates the Wnt/b-catenin signaling pathway to promote the reprogramming process, while ICG-001 impedes the process.

Figure 1

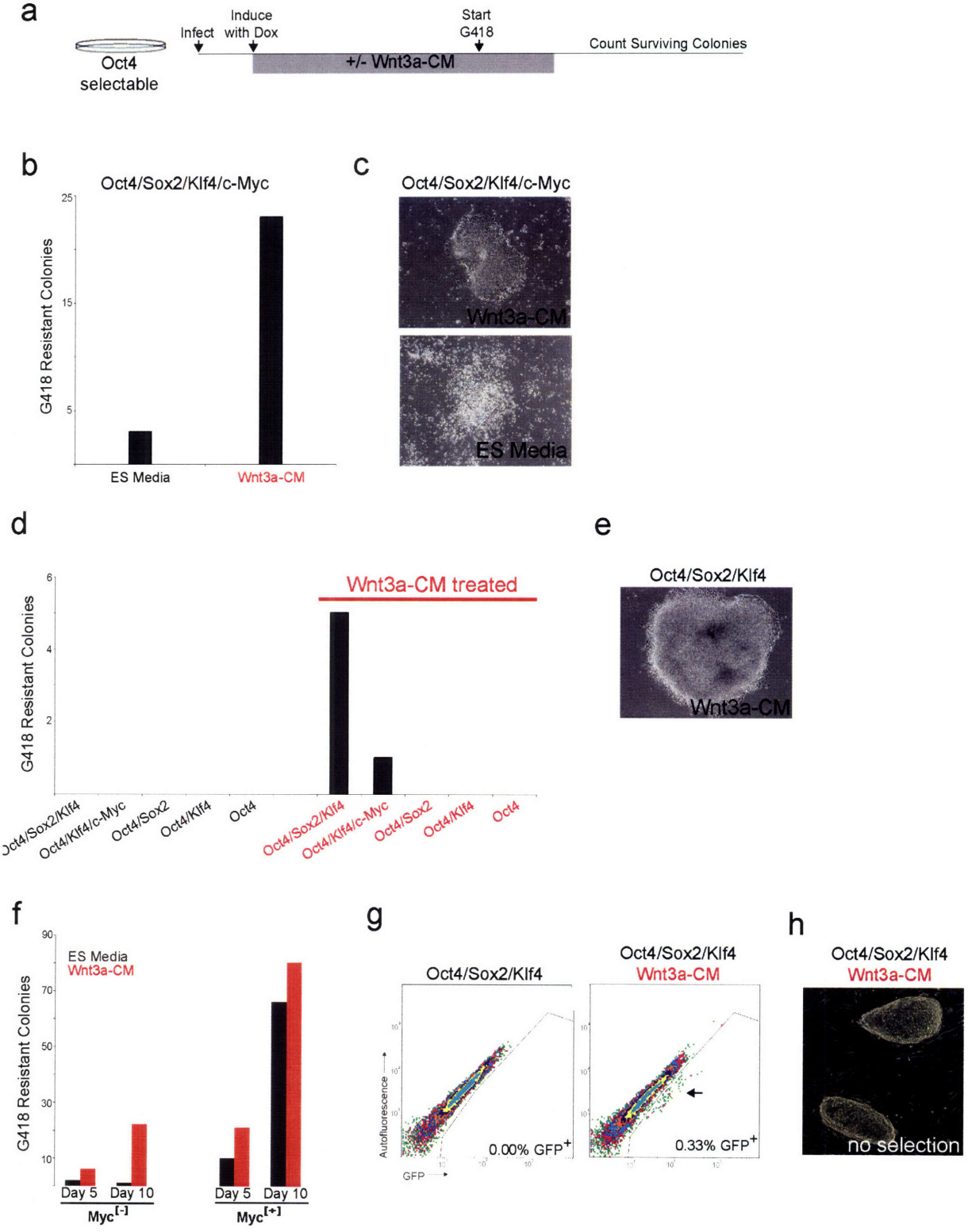

a
Induce
Infect with Dox
Start
G418
Count Surviving Colonies
Oct4
selectable
+/- Wnt3a-CM
b
Oct4/Sox2/Klf4/c-Myc
G418 Resistant Colonies
25
15
5
ES Media
Wnt3a-CM
c
Oct4/Sox2/Klf4/c-Myc
Wnt3a-CM
ES Media
d
Wnt3a-CM treated
G418 Resistant Colonies
6
4
2
Oct4/Sox2/Klf4
Oct4/Klf4/c-Myc
Oct4/Sox2
Oct4/Klf4
Oct4
Oct4/Sox2/Klf4
Oct4/Klf4/c-Myc
Oct4/Sox2
Oct4/Klf4
Oct4
e
Oct4/Sox2/Klf4
Wnt3a-CM
f
ES Media
Wnt3a-CM
G418 Resistant Colonies
90
70
50
30
10
Day 5
Day 10
Myc[-]
Day 5
Day 10
Myc[+]
g
Oct4/Sox2/Klf4
Oct4/Sox2/Klf4
Wnt3a-CM
Autofluorescence
GFP
0.00% GFP+
0.33% GFP+
h
Oct4/Sox2/Klf4
Wnt3a-CM
no selection

Figure 2

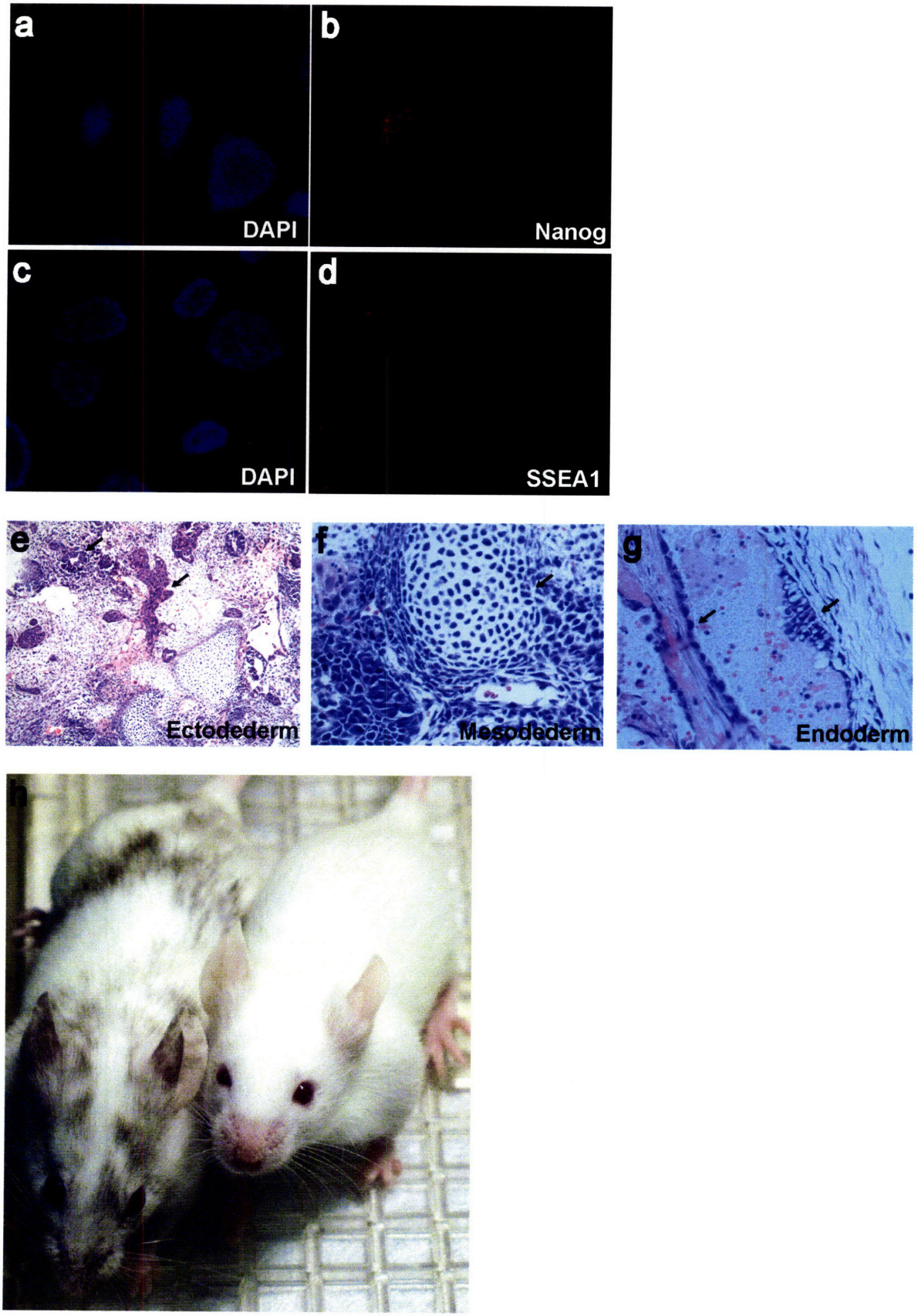

a
DAPI
b
Nanog
c
DAPI
d
SSEA1
e
Ectodederm
f
Mesodederm
g
Endoderm
h

Figure 3

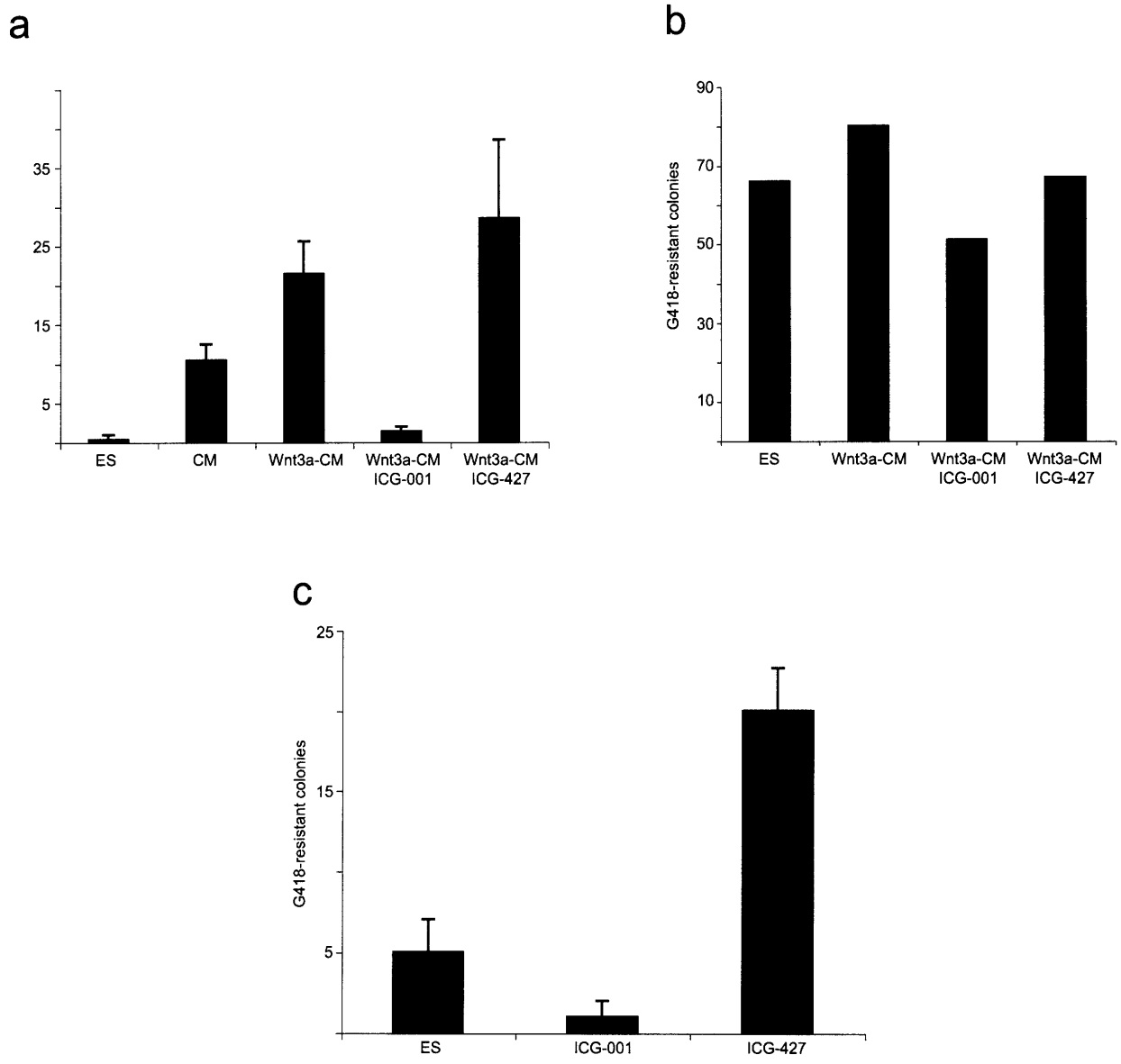
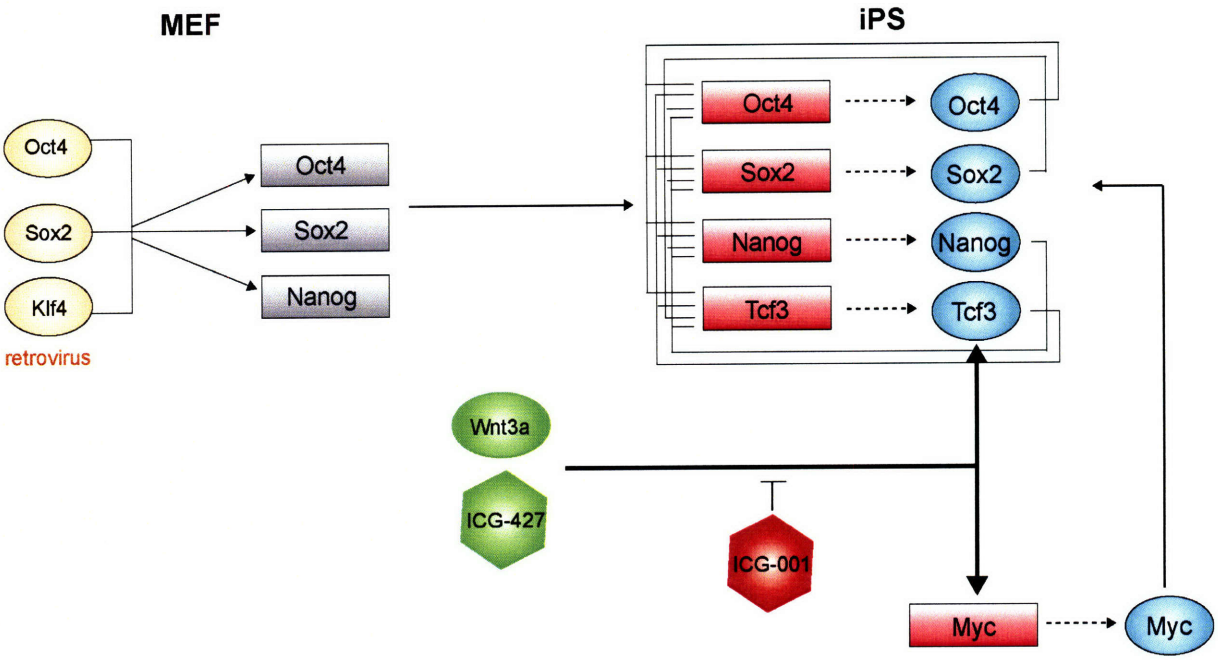

a
35
25
15
5
ES
CM
Wnt3a-CM
Wnt3a-CM ICG-001
Wnt3a-CM ICG-427
b
90
70
50
30
10
G418-resistant colonies
ES
Wnt3a-CM
Wnt3a-CM ICG-001
Wnt3a-CM ICG-427
c
25
15
5
G418-resistant colonies
ES
ICG-001
ICG-427

Figure 4


MEF
iPS
Oct4
Sox2
Klf4
retrovirus
Oct4
Sox2
Nanog
Oct4
Sox2
Nanog
Tcf3
Oct4
Sox2
Nanog
Tcf3
Wnt3a
ICG-427
ICG-001
Myc
Myc

Chapter 4

Foxp3 occupancy and regulation of key target genes during T-cell stimulation

Alexander Marson*, Karsten Kretschmer*, Garrett M. Frampton, Elizabeth S. Jacobsen, Julia K. Polansky, Kenzie D. MacIsaac, Stuart S. Levine, Ernest Fraenkel, Harald von Boehmer†, Richard A. Young†

**These authors contributed equally to this work*

† *Corresponding authors*

Personal Contribution to the Project

This work is the result of an equal collaboration with Karsten Kretschmer, who was at the time a post-doctoral fellow in the von Boehmer lab. Garrett Frampton was also instrumental to this project in providing computational expertise. I conceived of these experiments and established the collaboration with the von Boehmer lab. I oversaw Foxp3 ChIP-chip experiments, with technical assistance from Beth Jacobsen. In addition, I coordinated all analyses and prepared the main text of the manuscript.

Summary

Foxp3$^+$CD4$^+$CD25$^+$ regulatory T (T_{reg}) cells are essential for the prevention of autoimmunity (Sakaguchi et al., 1995; Baecher-Allan et al., 2006). T_{reg} cells have an attenuated cytokine response to T-cell receptor stimulation, and can suppress the proliferation and effector function of neighbouring T-cells (Shevach, 2002; von Boehmer, 2005). The forkhead transcription factor Foxp3 (forkhead box P3) is selectively expressed in T_{reg} cells, is required for T_{reg} development and function, and is sufficient to induce a T_{reg} phenotype in conventional CD4$^+$CD25$^-$ T-cells (Hori et al., 2003; Fontenot et al., 2003; Khattri et al., 2003; Fontenot et al., 2005). Mutations in *Foxp3* cause severe, multi-organ autoimmunity in both human and mouse (Brunkow et al., 2001; Bennett et al., 2001; Wildin et al., 2001). FOXP3 can cooperate in a DNA-binding complex with NFAT (nuclear factor of activated T-cells) to regulate the transcription of several known target genes (Wu et al., 2006b). However, the global set of genes regulated directly by Foxp3 is not known and consequently, how this transcription factor controls the gene expression program for T_{reg} function is not understood. Here we identify Foxp3 target genes and report that many of these are key modulators of T-cell activation and function. Remarkably, the predominant, although not exclusive, effect of Foxp3 occupancy is to suppress the activation of target genes on T-cell stimulation. Foxp3 suppression of its targets appears to be crucial for the normal function of T_{reg} cells, because overactive variants of some target genes are known to be associated with autoimmune disease.

We developed a strategy to identify genes whose promoters are bound by Foxp3 and whose expression is dependent on that transcription factor (**Figure 1**). To generate two cell lines that are genetically matched except for Foxp3, we transduced a Foxp3$^-$CD4$^+$ murine T-cell hybridoma with FLAG-tagged Foxp3. This approach was favoured over comparison of *ex vivo* cells, which are heterogeneous with regard to activation status. The lines provided sufficient numbers of homogeneous cells with appropriate controls to facilitate both location analysis and expression analysis. FACS (fluorescence-activated cell sorting) analysis confirmed that Foxp3 is expressed in the hybridoma at levels comparable to those in *ex vivo* CD4$^+$CD25$^+$ T_{reg} cells (**Supplementary Figure S1**). Previous work has shown that conventional CD4$^+$ T-cells ectopically expressing Foxp3 do not upregulate interleukin 2 (Il2) secretion following T-cell receptor (TCR) dependent stimulation (Schubert et al., 2001). To confirm that the Foxp3$^+$ hybridomas contain functional Foxp3, we assayed Foxp3$^-$ and Foxp3$^+$ cells for Il2 secretion. Indeed, FACS analysis revealed that Il2 secretion is strongly inhibited in phorbol myristate acetate (PMA)/ionomycin stimulated Foxp3$^+$ hybridomas compared to stimulated Foxp3$^-$ hybridomas (**Supplementary Figure S2**).

To identify direct targets of Foxp3, DNA sequences occupied by the transcription factor were identified in a replicate set of experiments using chromatin-immunoprecipitation (ChIP) combined with DNA microarrays. For this purpose, DNA microarrays were used that contain 60-mer oligonucleotide probes covering the region from –8 kilobases (kb) to +2 kb relative to the transcript start sites for approximately

16,000 annotated mouse genes (Boyer et al., 2006). The sites occupied by Foxp3 were identified as peaks of ChIP-enriched DNA that span closely neighbouring probes (**Figure 2**). Foxp3 was found to occupy the promoters of 1,119 genes in PMA/ionomycin stimulated hybridomas (Supplementary Tables S1 and S2). The well-characterized Foxp3 target gene (Wu et al., 2006b; Chen et al., 2006), *Il2*, was among the genes occupied by Foxp3 (**Figure 2A**). Most of the promoters occupied by Foxp3 in stimulated T-cells were also occupied in unstimulated cells (**Supplementary Tables S1** and **S2**, and **Supplementary Figure S3**). However, at some promoters Foxp3 binding increased considerably in cells stimulated with PMA/ionomycin (**Figure 2A** and **Supplementary Figure S3**). Control immunoprecipitation experiments in Foxp3$^-$ cells, which produced few positive signals, confirmed the specificity of these results (**Supplementary Figure S4**). Our confidence in the binding data was further strengthened by the discovery of a DNA sequence motif, which matches the consensus forkhead motif, at the genomic loci that were bound by Foxp3 (**Figure 2B** and **Supplementary Table S5**). This motif distinguishes Foxp3 bound regions from unbound regions tiled on the promoter arrays with a high level of confidence ($P < 10^{-41}$). Instances of this motif are significantly more likely to be conserved in Foxp3 bound regions than in promoter regions that are not bound by Foxp3 ($P < 10^{-23}$), suggesting that these sites serve a functional role (**Supplementary Table S5**).

To gain insights into the cellular functions that are directly regulated by Foxp3 transcriptional control, we compared the list of genes occupied by Foxp3 in stimulated hybridomas to the biological pathways annotated by the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2000). Among these pathways, Foxp3 target

genes are most strongly associated with the TCR signalling pathway ($P = 1.4 \times 10^{-5}$) (**Figure 2C**). Foxp3 target genes encode proteins that participate at multiple levels of this pathway, including cell surface molecules, signalling components and transcriptional regulators. Many genes with known roles in T-cells are missed with existing automated surveys, so we also manually inspected the list of Foxp3 target genes. This revealed that many additional Foxp3 targets are probably important for T-cell function, including microRNAs that are differentially expressed between T_{reg} cells and conventional T-cells (**Figure 2A** and **Supplementary Table S3** and Cobb et al., 2006). Surprisingly, *Ctla4* was not among the Foxp3 targets, but analysis by RT–PCR (polymerase chain reaction with reverse transcription) revealed no *Ctla4* expression in Foxp3$^-$ and Foxp3$^+$ hybridomas (data not shown). However, the Foxp3 targets included genes previously reported to be upregulated in T_{reg} cells, such as *Il2ra* (*CD25*) (Sakaguchi et al., 1995), *Tnfrsf18* (*GITR*) (McHugh et al, 2002), *Nrp1* (Bruder et al., 2004) and *Ccr4* (Iellem et al., 2001), consistent with predictions that these are directly regulated by Foxp3.

Previous reports have shown that only a portion of transcription factor binding events is associated with transcriptional regulation (Harbison et al., 2004). To identify the set of genes whose expression is dependent on Foxp3, we performed gene expression profiling on Foxp3$^-$ and Foxp3$^+$ T-cell hybridoma cells before and after PMA/ionomycin stimulation. Comparison of data from unstimulated Foxp3$^-$ and Foxp3$^+$ cells revealed few differentially expressed genes, suggesting that the transcription factor had little influence on global gene expression in unstimulated hybridomas (**Supplementary Table S8**). In contrast, PMA/ionomycin-stimulated Foxp3$^-$ and Foxp3$^+$ cells showed significant differences in expression of almost 1% of mouse genes. Many of these genes were

directly occupied by Foxp3, and Foxp3 binding was predominantly, but not exclusively, associated with genes whose expression is downregulated in stimulated Foxp3$^+$ hybridomas. Foxp3 occupied the promoters of approximately half of the genes in this cluster (**Figure 3A**). This set of downregulated target genes is enriched for genes that are implicated in TCR signalling ($P = 6.1 \times 10^{-3}$). Underscoring the significance of this finding, E2f4, an unrelated control transcription factor that occupies ~800 genes in these hybridomas, was found to occupy only one of these promoters (**Supplementary Table S4**).

Only a subset of all Foxp3 occupied genes was found to be differentially expressed. One reason for this could be that Foxp3 requires cofactors to modulate transcription. Recently, FOXP3 was shown to cooperate with NFAT in a DNA-binding complex to activate or repress target gene expression (Wu et al., 2006b). Consistent with this report, Foxp3 exerts a more pronounced transcriptional effect in stimulated hybridoma cells than in non-stimulated cells. Importantly, in our experiments, the set of Foxp3 bound genes are enriched ($P < 10^{-19}$) for the presence of an Nfat DNA sequence motif neighbouring the sites of Foxp3 occupancy (**Supplementary Table S5**).

We next examined whether the genes regulated by Foxp3 in the T-cell hybridomas show similar regulatory behaviour in *ex vivo* T-cells. Hybridoma cells were used initially because they afforded cells that differ only by Foxp3 status, they provided a homogenous population of non-stimulated T-cells, and we could have confidence in genome-wide location data using FLAG-tagged Foxp3. Nonetheless, microarray expression profiling was performed on *ex vivo* T-cells from different mice that express a

transgenic TCR alone or together with the TCR agonist ligand (Klein et al., 2003), from which relatively pure populations of naïve $CD4^+CD25^-$ T helper and $Foxp3^+CD4^+CD25^+$ T_{reg} cells, respectively, could be isolated. This analysis revealed that many, although not all, of the regulated targets that were identified in the hybridomas show consistent expression patterns in stimulated *ex vivo* cells (**Figure 3B** and **Supplementary Fig. S5**). Some differences in gene expression, especially genes that are activated by Foxp3 only in T_{reg} cells, may be due to the fact that *ex vivo* T_{reg} cells are generated through antigenic stimulation (Jordan et al., 2001; Kretschmer et al., 2005) and hence could contain transcriptional cofactors that differ from those in hybridoma cells.

Our findings were further validated for a panel of nine Foxp3 targets. Site-specific primers were used to confirm the binding of Foxp3 to the promoters of these genes (**Figure 3C**) and quantitative RT–PCR was used to assay messenger RNA levels in non-stimulated and stimulated hybridomas and *ex vivo* cells, in the presence and absence of 2 μM cyclosporin A (**Figure 3D**). These experiments confirmed the direct effects of Foxp3 at targets that were identified in the genome scale experiments. Notably, in these experiments (**Figure 3D**) all genes that were activated following stimulation in $Foxp3^-$ cells and repressed in $Foxp3^+$ cells were activated in a calcineurin dependent manner, consistent with the notion that Nfat is involved in their activation. In addition, just as Foxp3 regulates the protein levels of secreted Il2 (**Supplementary Fig. S2**), cell surface staining and FACS analysis show that Foxp3 regulates the level of Ly6a protein, demonstrating that Foxp3 transcriptional regulation of its targets modulates protein levels (**Supplementary Fig. S6**).

Taken together, the results from *ex vivo* T-cells and the hybridoma system identify a core set of Foxp3 regulatory targets (**Figure 4A**), most of which showed suppressed activation in stimulated $Foxp3^+$ cells. A smaller number of Foxp3 target genes was upregulated in stimulated $Foxp3^+$ cells, including some encoding cell surface molecules with known roles in immunoregulation, such as Ly6a (Stanford et al., 1997) and Tnfrsf9 (4-1BB) (Myers et al., 2005). The results from the hybridoma system indicate that Foxp3 occupies regions of most of its target promoters in both unstimulated and stimulated conditions, but increased binding at some promoters and regulation of most targets is observed after stimulation. Furthermore, in hybridomas the major function of Foxp3 at these genes is to suppress the level of gene activation that would occur if this transcription factor were not expressed (**Figure 4B**). Conceivably, the Foxp3 dependent downregulation of T-cell activation and cytokine genes, and upregulation of immunosuppressive cell surface molecules, contribute to both the hyporesponsive and suppressive T_{reg} phenotype.

Mutations in some Foxp3 target genes are already known to be associated with autoimmune disease. The protein tyrosine phosphatase Ptpn22 is a notable example. In our experiments, *Ptpn22* is one of the highest confidence direct targets of Foxp3; it is upregulated on stimulation in $Foxp3^-$ cells, and this upregulation is inhibited in $Foxp3^+$ hybridomas (**Figure 3A**) and *ex vivo* T_{reg} cells (**Figures 3B** and **3D**). Ptpn22 modulates the signal cascade downstream of the TCR, and mutations in the human *PTPN22* have been associated with type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus and Graves' disease, as well as other autoimmune diseases (Bottini et al., 2004; Wu et al., 2006a; Bottini et al., 2006). A recent report suggests that one *PTPN22* single-nucleotide

polymorphism associated with autoimmunity is a gain-of-function mutation (Vang et al., 2005). Our findings are compatible with the hypothesis that the gain-of-function mutation might be pathogenic if mutant PTPN22 is overactive in T_{reg} cells (Vang et al., 2005).

In summary, our data indicate that Foxp3 binds to the promoters of well-characterized regulators of T-cell activation and function. In the T-cell hybridomas studied here, the major role of this transcription factor is to dampen the induction of key genes when T_{reg} cells are stimulated. In *ex vivo* T_{reg} cells, Foxp3 could also activate the expression of a greater number of genes, perhaps owing to the greater abundance of certain transcriptional cofactors in these cells. Some of the identified Foxp3 target genes have been previously implicated in autoimmune diseases, implying that a therapeutic strategy to recapitulate the function of this transcription factor may have clinical utility for these diseases.

Experimental Procedures

A detailed description of all materials and methods used can be found in Supplementary Information.

Growth of murine $CD4^+$ T-cell hybridomas and *ex vivo* T-cells

$CD4^+$ 5B6-2 hybridoma cells expressing a $PLP_{139-151}$-specific TCR, which was kindly provided by V. Kuchroo, were cultured in Dulbecco's modified Eagle medium (Invitrogen). Primary murine $CD4^+$ T-cells were cultured in RPMI-1640 medium (Invitrogen). For gene expression profiling, real-time RT–PCR, and location analysis, cells were cultured in the absence or presence of 50 ng ml^{-1} phorbol 12-myristate 13-acetate (PMA) and 200 ng ml^{-1} ionomycin at 37 °C and harvested after 6 h. Where indicated, cells were preincubated for 1 h with 2μM cyclosporin A. Details of cell generation and isolation are provided in Supplementary Information.

Antibodies and ChIP assays

Detailed descriptions of antibodies, antibody specificity and ChIP methods used in this study are provided in Supplementary Information.

Promoter array design and data extraction

The design of the oligonucleotide-based promoter array set and data extraction methods are described in Supplementary Information. The microarrays used for location analysis in this study were manufactured by Agilent Technologies (http://www.agilent.com).

Motif analysis

Discovery of the Foxp3 sequence motif from the ChIP-chip binding data was performed using the THEME algorithm. The Foxp3 motif learned by THEME and the Nfat motif from the TRANSFAC database (version 8.3) were used to scan all arrayed sequences to identify matches to the motifs.

Functional classification of bound genes

Comparison of Foxp3 target genes to annotated KEGG biological pathways was performed using the online DAVID tool (http://niaid.abcc.ncifcrf.gov/).

Gene expression profiling

For each hybridoma culture condition, total RNA was prepared from 1×10^7 cells using Trizol (Gibco) followed by additional purification using the RNeasy Mini Kit (Qiagen). Biotinylated antisense cRNA was then prepared according to the Affymetrix standard labelling protocol (one amplification round). For each primary T-cell culture condition, total RNA was isolated from 5×10^5 cells with RNeasy. Biotinylated antisense cRNA was prepared by two rounds of *in vitro* amplification using the BioArray RNA Amplification and Labelling System (Enzo Life Sciences) according to the protocol for 10–1,000 ng of input RNA provided by the manufacturer. Biotinylated cRNAs of hybridomas and primary T-cells were fragmented and hybridized to Affymetrix GeneChip Mouse Expression Set 430 2.0 arrays at the Microarray Core Facility (Dana-Farber Cancer Institute).

Quantitative RT-PCR

To determine transcript levels in T-cell hybridomas and *ex vivo* T-cells, RNA was isolated, reverse-transcribed and subjected to real-time PCR performed on an ABI PRISM thermal cycler using SYBR Green PCR core reagents (Applied Biosystems). Detailed information is provided in Supplementary Information.

Reference List

Baecher-Allan, C., and Hafler, D. A. (2006). Human regulatory T cells and their role in autoimmune disease. Immunol Rev *212*, 203-216.

Bennett, C. L., Christie, J., Ramsdell, F., Brunkow, M. E., Ferguson, P. J., Whitesell, L., Kelly, T. E., Saulsbury, F. T., Chance, P. F., and Ochs, H. D. (2001). The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nat Genet *27*, 20-21.

Bottini, N., Musumeci, L., Alonso, A., Rahmouni, S., Nika, K., Rostamkhani, M., MacMurray, J., Meloni, G. F., Lucarelli, P., Pellecchia, M., *et al.* (2004). A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet *36*, 337-338.

Bottini, N., Vang, T., Cucca, F., and Mustelin, T. (2006). Role of PTPN22 in type 1 diabetes and other autoimmune diseases. Semin Immunol *18*, 207-213.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Bruder, D., Probst-Kepper, M., Westendorf, A. M., Geffers, R., Beissert, S., Loser, K., von Boehmer, H., Buer, J., and Hansen, W. (2004). Neuropilin-1: a surface marker of regulatory T cells. Eur J Immunol *34*, 623-630.

Brunkow, M. E., Jeffery, E. W., Hjerrild, K. A., Paeper, B., Clark, L. B., Yasayko, S. A., Wilkinson, J. E., Galas, D., Ziegler, S. F., and Ramsdell, F. (2001). Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. Nat Genet *27*, 68-73.

Chen, C., Rowell, E. A., Thomas, R. M., Hancock, W. W., and Wells, A. D. (2006). Transcriptional regulation by Foxp3 is associated with direct promoter occupancy and modulation of histone acetylation. J Biol Chem *281*, 36828-36834.

Cobb, B. S., Hertweck, A., Smith, J., O'Connor, E., Graf, D., Cook, T., Smale, S. T., Sakaguchi, S., Livesey, F. J., Fisher, A. G., and Merkenschlager, M. (2006). A role for Dicer in immune regulation. J Exp Med *203*, 2519-2527.

Fontenot, J. D., Gavin, M. A., and Rudensky, A. Y. (2003). Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. Nat Immunol *4*, 330-336.

Fontenot, J. D., Rasmussen, J. P., Williams, L. M., Dooley, J. L., Farr, A. G., and Rudensky, A. Y. (2005). Regulatory T cell lineage specification by the forkhead transcription factor foxp3. Immunity *22*, 329-341.

Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J. B., Reynolds, D. B., Yoo, J., *et al.* (2004). Transcriptional regulatory code of a eukaryotic genome. Nature *431*, 99-104.

Hori, S., Nomura, T., and Sakaguchi, S. (2003). Control of regulatory T cell development by the transcription factor Foxp3. Science *299*, 1057-1061.

Iellem, A., Mariani, M., Lang, R., Recalde, H., Panina-Bordignon, P., Sinigaglia, F., and D'Ambrosio, D. (2001). Unique chemotactic response profile and specific expression of chemokine receptors CCR4 and CCR8 by CD4(+)CD25(+) regulatory T cells. J Exp Med *194*, 847-853.

Jordan, M. S., Boesteanu, A., Reed, A. J., Petrone, A. L., Holenbeck, A. E., Lerman, M. A., Naji, A., and Caton, A. J. (2001). Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. Nat Immunol *2*, 301-306.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

Khattri, R., Cox, T., Yasayko, S. A., and Ramsdell, F. (2003). An essential role for Scurfin in CD4+CD25+ T regulatory cells. Nat Immunol *4*, 337-342.

Klein, L., Khazaie, K., and von Boehmer, H. (2003). In vivo dynamics of antigen-specific regulatory T cells not predicted from behavior in vitro. Proc Natl Acad Sci U S A *100*, 8886-8891.

Kretschmer, K., Apostolou, I., Hawiger, D., Khazaie, K., Nussenzweig, M. C., and von Boehmer, H. (2005). Inducing and expanding regulatory T cell populations by foreign antigen. Nat Immunol *6*, 1219-1227.

McHugh, R. S., Whitters, M. J., Piccirillo, C. A., Young, D. A., Shevach, E. M., Collins, M., and Byrne, M. C. (2002). CD4(+)CD25(+) immunoregulatory T cells: gene expression analysis reveals a functional role for the glucocorticoid-induced TNF receptor. Immunity *16*, 311-323.

Myers, L. M., and Vella, A. T. (2005). Interfacing T-cell effector and regulatory function through CD137 (4-1BB) co-stimulation. Trends Immunol *26*, 440-446.

Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M., and Toda, M. (1995). Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. J Immunol *155*, 1151-1164.

Schubert, L. A., Jeffery, E., Zhang, Y., Ramsdell, F., and Ziegler, S. F. (2001). Scurfin (FOXP3) acts as a repressor of transcription and regulates T cell activation. J Biol Chem *276*, 37672-37679.

Shevach, E. M. (2002). CD4+ CD25+ suppressor T cells: more questions than answers. Nat Rev Immunol *2*, 389-400.

Stanford, W. L., Haque, S., Alexander, R., Liu, X., Latour, A. M., Snodgrass, H. R., Koller, B. H., and Flood, P. M. (1997). Altered proliferative response by T lymphocytes of Ly-6A (Sca-1) null mice. J Exp Med *186*, 705-717.

von Boehmer, H. (2005). Mechanisms of suppression by suppressor T cells. Nat Immunol *6*, 338-344.

Wildin, R. S., Ramsdell, F., Peake, J., Faravelli, F., Casanova, J. L., Buist, N., Levy-Lahad, E., Mazzella, M., Goulet, O., Perroni, L., *et al.* (2001). X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. Nat Genet *27*, 18-20.

Wu, J., Katrekar, A., Honigberg, L. A., Smith, A. M., Conn, M. T., Tang, J., Jeffery, D., Mortara, K., Sampang, J., Williams, S. R., *et al.* (2006a). Identification of substrates of human protein-tyrosine phosphatase PTPN22. J Biol Chem *281*, 11002-11010.

Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A. D., Stroud, J. C., Bates, D. L., Guo, L., Han, A., Ziegler, S. F., *et al.* (2006b). FOXP3 controls regulatory T cell function through cooperation with NFAT. Cell *126*, 375-387.

Acknowledgements

We thank members of the Young, von Boehmer and Fraenkel laboratories, as well as R. Jaenisch and D.K. Gifford, for discussions and critical review of the manuscript, especially T.I. Lee, J. Zeitlinger and D.T. Odom. We also thank Biology and Research Computing (BaRC), especially T. Dicesare for graphic assistance, as well as E. Herbolsheimer for computational and technical support. K.K. was supported in part by the fellowship grant KR2316/1-1 from the German Research Foundation. This work was supported in part by a donation from E. Radutzky, a grant from the Whitaker Foundation to E.F., NIH Grant R37 AI53102 to H.v.B. and NIH grant AI055021 to R.A.Y.

Author Information

All microarray data from this study are available from ArrayExpress at the EBI (http://www.ebi.ac.uk/arrayexpress) under accession code E-TABM-154. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.A.Y. (young@wi.mit.edu) or H. v. B. (Harald_von_Boehmer@dfci.harvard.edu).

Figure Legends

Figure 1 Strategy to identify direct Foxp3 transcriptional effects. Genetically matched $Foxp3^+$ and $Foxp3^-$ cell populations were generated by transduction of FLAG-tagged Foxp3 into a $Foxp3^-$ murine T-cell hybridoma. Foxp3 binding sites at promoters across the genome were identified by ChIP experiments with an anti-FLAG antibody. Foxp3 dependent transcriptional regulation was identified by gene expression profiling performed on each of these cell types.

Figure 2 Direct Foxp3 targets include key modulators of T-cell function. a, Foxp3 ChIP enrichment ratios (ChIP-enriched versus total genomic DNA) across indicated promoters are shown for stimulated (pink) and unstimulated (blue) cells. Exons (blocks) and introns (lines) of genes and the mir-155 precursor (grey) are drawn to scale below the plots, with direction of transcription noted by an arrow. **b**, Foxp3 bound genomic regions are enriched for the presence of a forkhead DNA motif, represented here in WebLogo (http://weblogo.berkeley.edu). **c**, The KEGG (Kanehisa et al, 2000) TCR signalling pathway, enriched ($P = 1.4 \times 10^{-5}$) for proteins encoded by direct targets of Foxp3 (blue outline), is displayed.

Figure 3 Foxp3 directly suppresses the activation of target genes. a, Replicate expression data for the 125 genes with Foxp3 dependent differential expression in stimulated hybridomas (False Disovery Rate [FDR] < 0.05) - were hierarchically clustered and displayed. - The Z-score normalized induction (red) or repression (green) is shown for each gene. Direct targets of Foxp3 in stimulated hybridomas are indicated (dark blue for FDR < 0.05, light blue for FDR < 0.10). **b**, For repressed (green) and induced (red) Foxp3-bound targets in **a**, $\log_2$(fold difference) in expression between stimulated *ex vivo* effector (T_{eff}) T-cells and T_{reg} cells is displayed. *Slc17a6* and *Adam10* are not expressed in the *ex vivo* samples. **c**, Site-specific PCR on 10 ng of ChIP DNA confirmed selected targets. Immunoprecipitated (IP) DNA was compared to serial

dilutions (90, 30 and 10 ng of DNA) of unenriched whole cell extract (WCE) DNA. Enrichment ratios, shown at top right of each sub-panel, are normalized relative to the unenriched *b actin* control. DNA fragment size (bp) is indicated on the left of each row. **d**, The transcript levels of the panel of Foxp3 targets presented in **c** and of *Foxp3* were analysed by real-time RT–PCR in stimulated and unstimulated cells, with (grey) and without (black) cyclosporin A. Mean values ± s.d. of relative expression, determined in triplicate, are shown for indicated genes.

Figure 4 Core direct regulatory effects of Foxp3. a, Shown here are a subset of direct Foxp3 targets that exhibit consistent transcriptional behaviour in hybridomas and in *ex vivo* T-cells (Supplementary Fig. S5). **b**, Foxp3 binds to a large set of promoters both in unstimulated and stimulated T-cells, but Foxp3 transcriptional regulation is more extensive in stimulated T-cells. The genomic regions where Foxp3 binds are enriched for an Nfat binding site DNA motif. In the hybridomas, Foxp3 predominantly acts to directly suppress the activation of its target genes.

Figure 1

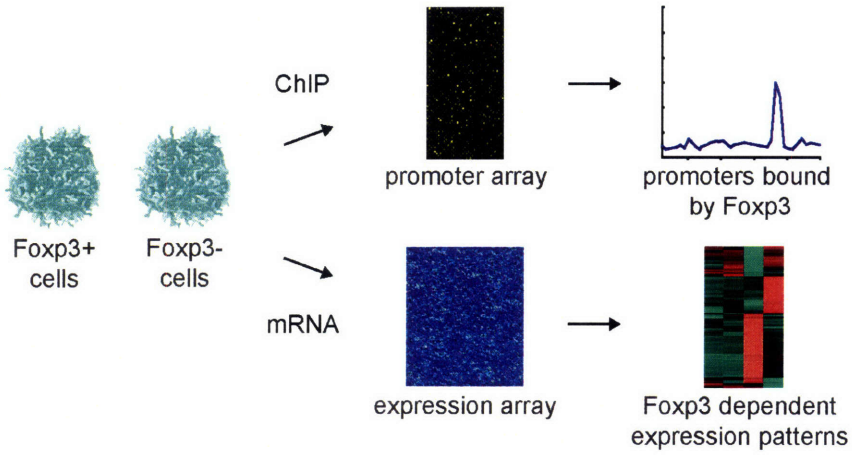
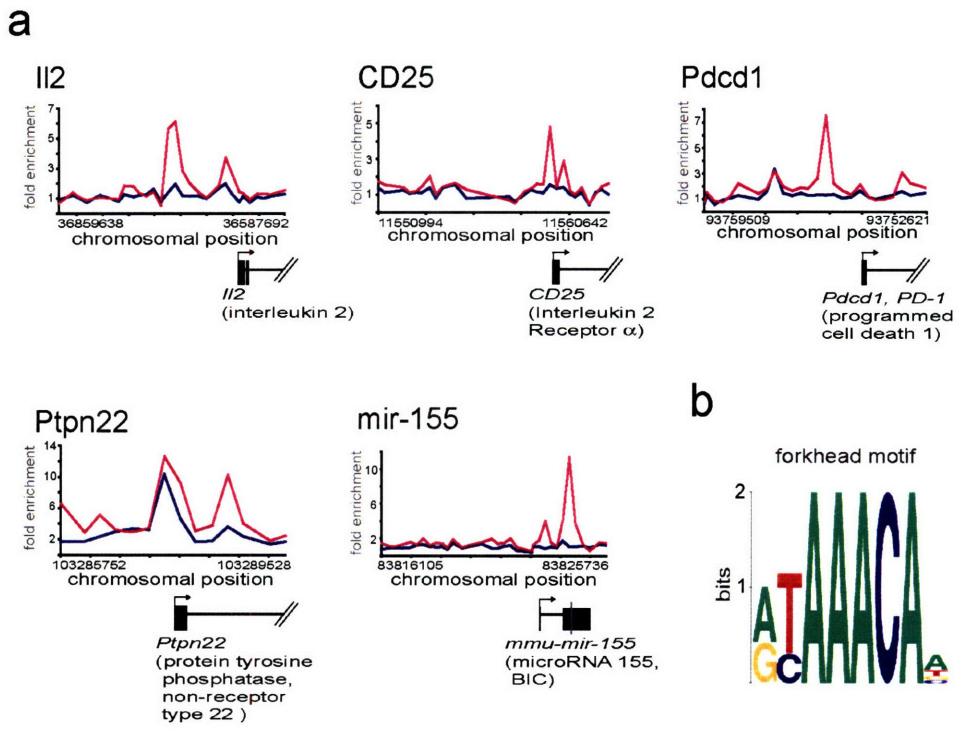

ChIP
promoter array
promoters bound by Foxp3
Foxp3+ cells
Foxp3- cells
mRNA
expression array
Foxp3 dependent expression patterns

Figure 2


a
Il2
fold enrichment
36859638
36587692
chromosomal position
Il2
(interleukin 2)
CD25
11550994
11560642
chromosomal position
CD25
(Interleukin 2
Receptor α)
Pdcd1
93759509
93752621
chromosomal position
Pdcd1, PD-1
(programmed
cell death 1)
Ptpn22
103285752
103289528
chromosomal position
Ptpn22
(protein tyrosine
phosphatase,
non-receptor
type 22)
mir-155
83816105
83825736
chromosomal position
mmu-mir-155
(microRNA 155,
BIC)
b
forkhead motif
bits

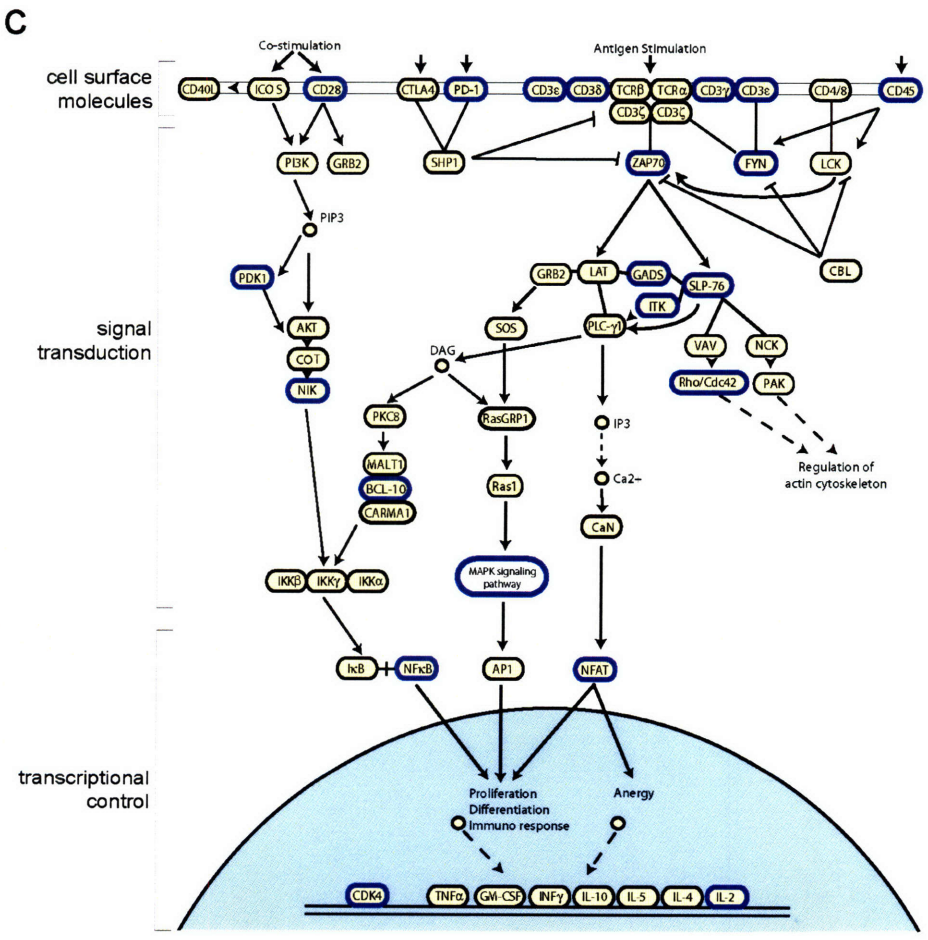

c
cell surface
molecules
signal
transduction
transcriptional
control
Co-stimulation
Antigen Stimulation
CD40l
ICOS
CD28
CTLA4
PD-1
CD3ε
CD3δ
TCRβ
TCRα
CD3γ
CD3ε
CD4/8
CD45
CD3ζ
CD3ζ
PI3K
GRB2
SHP1
ZAP70
FYN
LCK
PIP3
PDK1
GRB2
LAT
GADS
SLP-76
CBL
ITK
AKT
SOS
PLC-γ1
VAV
NCK
COT
DAG
Rho/Cdc42
PAK
NIK
PKCθ
RasGRP1
IP3
Regulation of
actin cytoskeleton
MALT1
BCL-10
CARMA1
Ras1
Ca2+
CaN
IKKβ
IKKγ
IKKα
MAPK signaling
pathway
IκB
NFκB
AP1
NFAT
Proliferation
Differentiation
Immuno response
Anergy
CDK4
TNFα
GM-CSF
INFγ
IL-10
IL-5
IL-4
IL-2

Figure 3

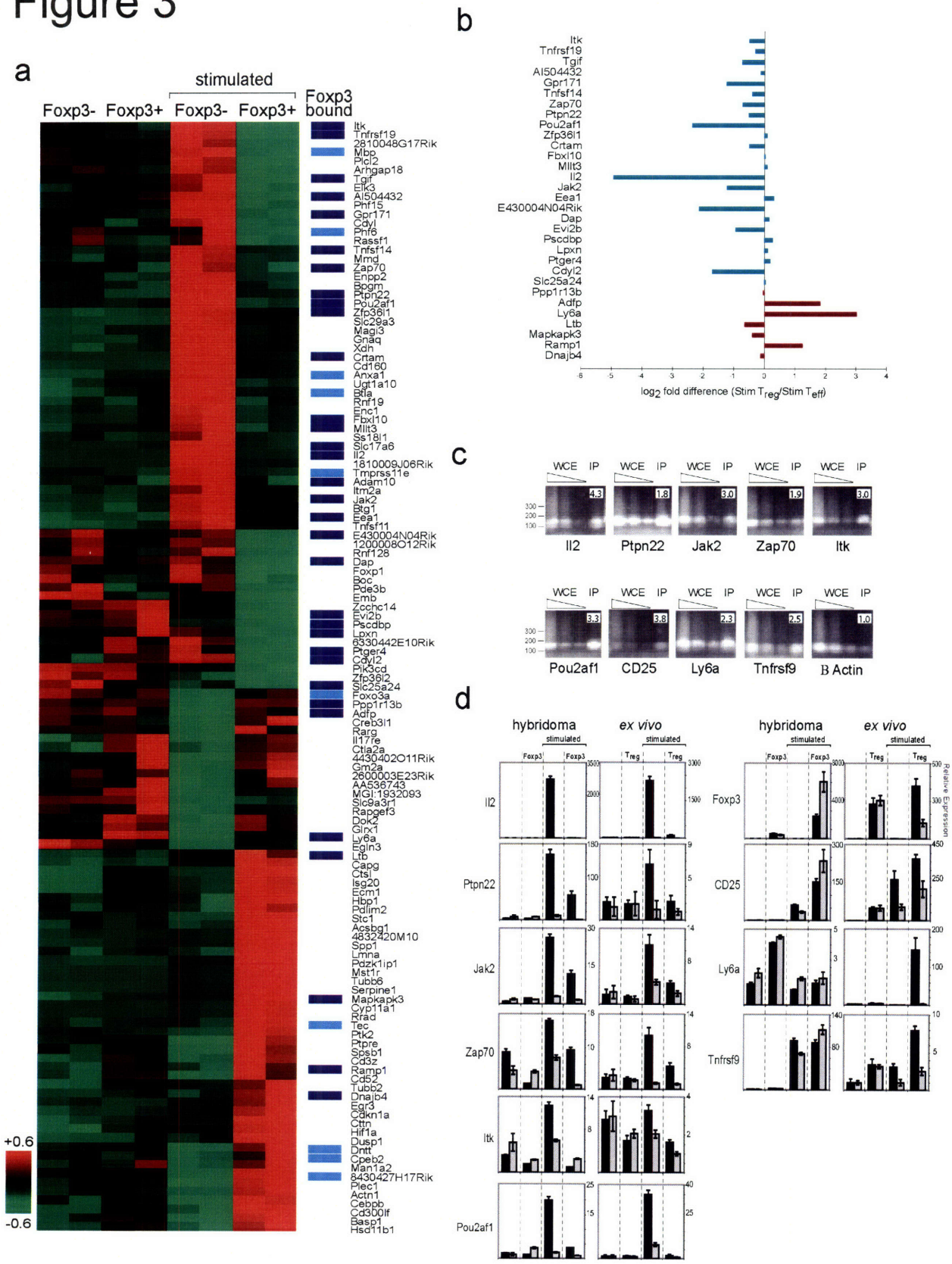
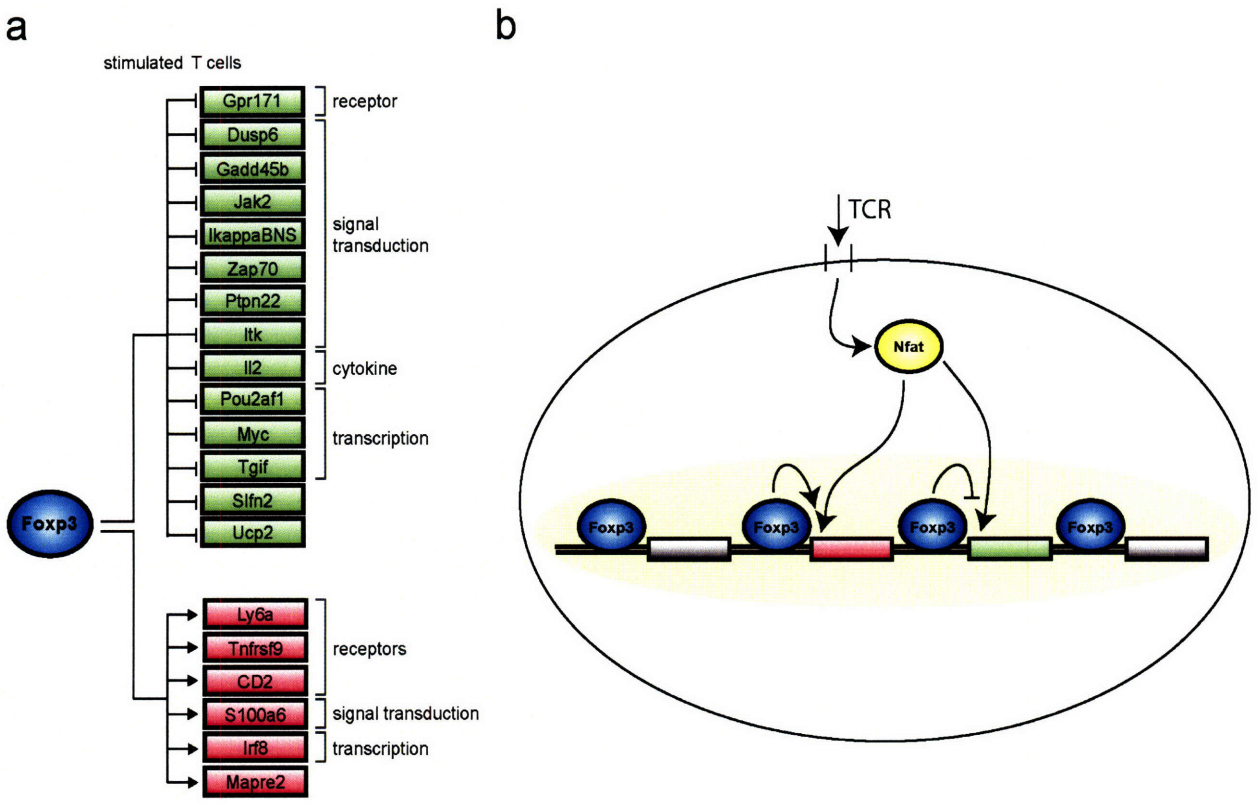

a
stimulated
Foxp3-
Foxp3+
Foxp3-
Foxp3+
Foxp3 bound
+0.6
-0.6
Itk
Tnfrsf19
2810048G17Rik
Mbp
Ptcl2
Arhgap18
Tgif
Elk3
AI504432
Phf15
Gpr171
Cdyl
Phf6
Rassf1
Tnfsf14
Mmd
Zap70
Enpp2
Bcgrn
Ptpn22
Pou2af1
Zfp36l1
Slc29a3
Magi3
Gnaq
Xdh
Crtam
Cd160
Anxa1
Ugt1a10
Btla
Rnf19
Enc1
Fbxl10
Mllt3
Ss18l1
Slc17a6
Il2
1810009J06Rik
Tmprss11e
Adam10
Itm2a
Jak2
Btg1
Eea1
Tnfsf11
E430004N04Rik
1200008O12Rik
Rnf128
Dap
Foxp1
Boc
Pde3b
Emb
Zcchc14
Evi2b
Pscdbp
Lpxn
6330442E10Rik
Ptger4
Cdyl2
Pik3cd
Zfp36l2
Slc25a24
Foxo3a
Ppp1r13b
Adfp
Creb3l1
Rarg
Il17re
Ctla2a
4430402O11Rik
Gm2a
2600003E23Rik
AA536743
MGI:1932093
Slc9a3r1
Rapgef3
Dok2
Glrx1
Ly6a
Egln3
Ltb
Capg
Ctsl
Isg20
Ecm1
Htbp1
Pdlim2
Stc1
Acsbg1
4832420M10
Spp1
Lmna
Pdzk1ip1
Mst1r
Tubb6
Serpine1
Mapkapk3
Cyp11a1
Rrad
Tec
Ptk2
Ptpre
Spsb1
Cd3z
Ramp1
Cd52
Tubb2
Dnajb4
Egr3
Cdkn1a
Cttn
Hif1a
Dusp1
Dntt
Cpeb2
Man1a2
8430427H17Rik
Plec1
Actn1
Cebpb
Cd300lf
Basp1
Hsd11b1
b
Itk
Tnfrsf19
Tgif
AI504432
Gpr171
Tnfsf14
Zap70
Ptpn22
Pou2af1
Zfp36l1
Crtam
Fbxl10
Mllt3
Il2
Jak2
Eea1
E430004N04Rik
Dap
Evi2b
Pscdbp
Lpxn
Ptger4
Cdyl2
Slc25a24
Ppp1r13b
Adfp
Ly6a
Ltb
Mapkapk3
Ramp1
Dnajb4
-6 -5 -4 -3 -2 -1 0 1 2 3 4
log2 fold difference (Stim Treg/Stim Teff)
c
WCE IP
4.3
1.8
3.0
1.9
3.0
300
200
100
Il2
Ptpn22
Jak2
Zap70
Itk
3.3
3.3
2.3
2.5
1.0
Pou2af1
CD25
Ly6a
Tnfrsf9
B Actin
d
hybridoma
ex vivo
stimulated
Foxp3
Treg
Il2
Ptpn22
Jak2
Zap70
Itk
Pou2af1
Foxp3
CD25
Ly6a
Tnfrsf9
Relative Expression

Figure 4


a
stimulated T cells
Foxp3
Gpr171
Dusp6
Gadd45b
Jak2
IkappaBNS
Zap70
Ptpn22
Itk
Il2
Pou2af1
Myc
Tgif
Slfn2
Ucp2
receptor
signal transduction
cytokine
transcription
Ly6a
Tnfrsf9
CD2
S100a6
Irf8
Mapre2
receptors
signal transduction
transcription
b
TCR
Nfat
Foxp3
Foxp3
Foxp3
Foxp3

Chapter 5

Future Directions and Conclusion

The work presented here focuses on molecular programs controlling two cell types, embryonic stem (ES) cells and regulatory T (T_{reg}) cells. Chapter 2 extends the model of regulatory circuitry downstream of core ES cell transcription factors Oct4, Sox2, Nanog and Tcf3 to incorporate miRNA genes as well as protein-coding genes. Oct4, Sox2, Nanog and Tcf3 co-occupy the promoters of miRNAs that are preferentially expressed in ES cells, as well as a set of miRNAs that are repressed in ES cells by Polycomb Group proteins and expressed in differentiated cells in a tissue-specific pattern. Therefore in ES cells, Oct4, Sox2, Nanog and Tcf3 appear to regulate not only proteins that are preferentially expressed in ES cells and contribute to self-renewal and normal pluripotency, but also miRNAs. Additionally, miRNAs that are repressed in ES by Polycomb Group proteins could serve as lineage determinants, similar to the developmental transcription factors that are similarly silenced in ES cells.

Chapter 3 describes efforts to reprogram somatic cells to pluripotency with a signaling pathway that connects directly to the core regulatory circuitry of ES cells. Tcf3, a transcriptional regulator downstream of the canonical Wnt pathway in ES cells, occupies genomic regions that are highly overlapping with Oct4, Sox2 and Nanog binding sites, and Wnt signaling contributes to the maintenance of pluripotency in ES cells. Data presented in Chapter 3 demonstrates that stimulation of the Wnt pathway promotes the generation of induced pluripotent stem (iPS) cells in concert with nuclear reprogramming factors.

The work presented in Chapter 4 focuses on the gene expression program under the control of Foxp3, a crucial transcription factor in T_{reg} cells. In contrast to ES cells, where several key studies had already examined the core transcriptional circuitry, no global analysis of Foxp3 transcriptional regulation had been conducted previously. We were able to identify binding sites for this master regulator across the murine genome and demonstrate that this factor suppresses the expression of a set of target genes that would otherwise be induced by T-cell receptor (TCR) stimulation. The over-representation of consensus Nfat binding DNA motifs that we observed neighboring Foxp3 binding sites is consistent with a model where Foxp3, Nfat and perhaps other key factors, cooperate to shape T_{reg} cell identity (Wu et al., 2006; Ono et al., 2007). Among the genes that are targeted by Foxp3 are some of the genes most commonly mutated in a range of autoimmune diseases, including the phosphatase *Ptpn22* (Bottini et al. 2004; Wu et al., 2006; Bottini et al., 2006; Vang et al., 2006), consistent with a model where defects in the T_{reg} cellular circuitry are a common cause of autoimmunity.

Knowledge of the molecular programs that control cellular identity provides clues as to how cellar identity can be manipulated. Furthermore, this knowledge has potential to shed new light on human diseases that result from failures in the programs that normally control cellular identity. Two major directions for future research, which follow from the work presented here, are discussed below. Briefly, we have constructed a library of murine miRNAs to test their ability to shape cellular identity. miRNAs that are now connected to the core ES cell regulatory circuitry are particularly intriguing candidates to examine for their roles in promoting pluripotency or in cell-fate determination. Second, because Foxp3 was observed to target genes that are commonly

mutated in autoimmune disease, we are eager to study further the relationship between Foxp3 occupancy of the genome and polymorphisms associated with autoimmunity. Complementary investigations of the core circuitry of T_{reg} cells and genetic studies of patients have potential to elucidate the etiology of autoimmune diseases.

Manipulating miRNA expression to reprogram cellular identity

We have designed a library of annotated murine miRNA sequences cloned into a lentiviral vector. Known mature miRNA sequences and genomic flanking sequence were inserted into Gateway entry vectors and recombined into lentiviral expression vectors. This system drives miRNA transcription using an RNA polymerase II *Ubiquitin* promoter. The vector also includes a blasticidin resistance gene, allowing for the selection of infected cells. Northern blots on infected neural precursor cells have confirmed that the system offers enforced expression of individual mature miRNA transcripts.

The connections observed between miRNAs and the core transcriptional circuitry transcriptional circuitry of ES cells, described in Chapter 2, are a strong motivation to examine the contributions of individual miRNAs to pluripotency and cellular differentiation. The library that we have created, which allows for enforced expression of miRNAs, will serve as a tool to further characterize the functions of miRNAs. We hope to screen for miRNAs that help to maintain pluripotency upon LIF withdrawal, or increase the efficiency of generating iPS cells in combination with transcription factors.

Additionally, we hope to characterize the effects of miRNAs on the developmental potential of ES cells and identify miRNAs that are useful to direct ES cell-fate decisions.

The miRNA library will allow for multiple high-throughput approaches to examine miRNA contribution to cellular identity. Experiments can be conducted in an arrayed fashion, testing in parallel the effect of each cloned miRNA. However, for efficiency and also to identify potential combinatorial effects of miRNAs on cellular identity, we can also infect cells with pools of virus. For example, to identify miRNAs that promote lineage specification, we can infect ES cells with pools of virus, allow the cells to differentiate into heterogeneous embryoid bodies, and then sort for cells of different lineage markers. PCR amplification of the miRNA provirus, followed by either microarray analysis (as described in Voorhoeve et al., 2006) or quantitative sequencing, should allow us to determine which exogenous miRNAs are preferentially associated with cells of a specific lineage suggesting that they could have directed the differentiation of infected ES cells. Notably, these studies are not limited to ES cells; similar experimental designs could be applied to examine miRNA contribution to other cell types or cell functions, including screens to identify miRNAs that contribute to the suppressive phenotype of T_{reg} cells.

Molecular circuitry of T_{reg} cell and insight into autoimmune disease

The transcription factor Foxp3 is required for the establishment and maintenance of functional T_{reg} cells (Hori et al., 2003; Fontenot et al., 2003; Khattri et al., 2003;

Fontenot et al., 2005). Foxp3 is sufficient to induce a suppressive phenotype in conventional T cells (Hori et al., 2003). *Foxp3* mutations are found in a rare human autoimmune syndrome, IPEX, characterized by multi-organ autoimmunity including enteropathy, thyroiditis and diabetes mellitus (Bennett et al, 2001; Wildin et al., 2001). Patients with this disease, as well as mice with targeted *Foxp3* deletion, lack T_{reg} cells (Fontenot et al., 2005; Baecher-Allan and Hafler, 2006). In contrast, in more common forms of autoimmune disease, such as multiple sclerosis, patients have normal numbers of T_{reg} cells, but these cells appear defective based on *in vitro* suppression assays (Viglietta et al., 2004; Baecher-Allan and Hafler, 2006).

One compelling hypothesis is that Foxp3 itself is not mutated in common autoimmune diseases, but targets of the transcription factor are mutated or aberrantly regulated leading to T_{reg} cell dysfunction. The core circuitry of T_{reg} cells could be disturbed at different nodes to cause different dysfunctions associated with distinct autoimmune diseases. These speculations are supported by our observation that Foxp3 occupies the promoters of some of the genes most commonly associated with autoimmune diseases. *PTPN22* is a Foxp3 target gene and polymorphisms in this gene are associated with type I diabetes, rheumatoid arthritis, systemic lupus erythematosus and other autoimmune diseases, but no significant risk association has been identified with multiple sclerosis (Bottini et al. 2004; Bottini et al., 2006; Vang et al., 2006; De Jager et al., 2006 and D.A.H., personal communication). Meanwhile, *IL2RA* (CD25), which is also a direct target of Foxp3 and which is preferentially expressed in T_{reg} cells, is strongly implicated in multiple sclerosis ($p \leq 1 \times 10^{-27}$), as well as other common autoimmune diseases (Hafler et al., 2007; D.A.H., personal communication).

To examine the relationship between the core regulatory circuitry of T_{reg} cells and common human autoimmune diseases, we are taking a three-pronged approach. First, we are planning ChIP-seq experiments to establish high-resolution maps of genomic loci in human T_{reg} cells associated with Foxp3 and other key transcription factors, terminal components of signaling pathways known to effect T_{reg} cells, and chromatin modifications. Second, these maps will be compared to data that is currently being generated by genome-wide association studies of common human autoimmune diseases, including type I diabetes and multiple sclerosis. New sequencing technologies will allow for dense re-sequencing of genomic regions of interest in patient cells and healthy controls. Third, we will compare gene expression profiles of T_{reg} cells from patients to those of healthy controls. The literature already contains some examples suggesting that approach will be fruitful. Runx1/Aml1 is a transcription factor that physically interacts with Foxp3 and contributes to T_{reg} cell function in cooperation with Foxp3 (Ono et al., 2007). Recently, several polymorphisms associated with distinct autoimmune diseases have been identified that ablate Runx1 binding sites in the genome (Prokunina et al., 2002; Helms et al., 2003; Tokuhiro et al., 2003). The experimental approach outlined above has potential to identify systematically mutations that disrupt the core circuitry of T_{reg} cells and contribute to common autoimmune diseases.

Concluding Remarks

The work presented here focuses on the molecular circuitry of two cell types, ES cells and T_{reg} cells. These studies highlight ways that transcription factors, miRNAs,

signal transduction pathways and chromatin modifications coordinately regulate cellular identity. Furthermore, these studies argue that knowledge of these complex programs can offer clues to reprogramming cellular identity. While these two cell types are of considerable scientific and clinical interest, the studies described here might also be considered part of a general approach to studies of cellular identity. There is now significant potential to examine systematically the programs that shape the diverse set of cell types constituting the human body, to examine how these programs fail in human disease, and to examine how cells can be reprogrammed for therapeutic purposes.

Acknowledgements

Nancy Hannett provided major technical assistance in cloning the genomic regions for the miRNA library. Marius Wernig and Andria Balogh helped perform initial lentiviral infections of neural precursor cells, and Calvin Jan assisted with Northern blot assay to confirm miRNA expression.

Genomic studies of T_{reg} cells in human autoimmunity are currently being planned in close collaboration with David Hafler and Lisa Maier.

References

Bennett, C. L., Christie, J., Ramsdell, F., Brunkow, M. E., Ferguson, P. J., Whitesell, L., Kelly, T. E., Saulsbury, F. T., Chance, P. F., and Ochs, H. D. (2001). The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nat Genet *27*, 20-21.

Bottini, N., Musumeci, L., Alonso, A., Rahmouni, S., Nika, K., Rostamkhani, M., MacMurray, J., Meloni, G. F., Lucarelli, P., Pellecchia, M., *et al.* (2004). A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet *36*, 337-338.

Bottini, N., Vang, T., Cucca, F., and Mustelin, T. (2006). Role of PTPN22 in type 1 diabetes and other autoimmune diseases. Semin Immunol *18*, 207-213.

Brunkow, M. E., Jeffery, E. W., Hjerrild, K. A., Paeper, B., Clark, L. B., Yasayko, S. A., Wilkinson, J. E., Galas, D., Ziegler, S. F., and Ramsdell, F. (2001). Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. Nat Genet *27*, 68-73.

De Jager, P. L., Sawcer, S., Waliszewska, A., Farwell, L., Wild, G., Cohen, A., Langelier, D., Bitton, A., Compston, A., Hafler, D. A., and Rioux, J. D. (2006). Evaluating the role of the 620W allele of protein tyrosine phosphatase PTPN22 in Crohn's disease and multiple sclerosis. Eur J Hum Genet *14*, 317-321.

Fontenot, J. D., Gavin, M. A., and Rudensky, A. Y. (2003). Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. Nat Immunol *4*, 330-336.
Fontenot, J. D., Rasmussen, J. P., Williams, L. M., Dooley, J. L., Farr, A. G., and Rudensky, A. Y. (2005). Regulatory T cell lineage specification by the forkhead transcription factor foxp3. Immunity *22*, 329-341.

Hafler, D. A., Compston, A., Sawcer, S., Lander, E. S., Daly, M. J., De Jager, P. L., de Bakker, P. I., Gabriel, S. B., Mirel, D. B., Ivinson, A. J., *et al.* (2007). Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med *357*, 851-862.

Helms, C., Cao, L., Krueger, J. G., Wijsman, E. M., Chamian, F., Gordon, D., Heffernan, M., Daw, J. A., Robarge, J., Ott, J., *et al.* (2003). A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis. Nat Genet *35*, 349-356.

Hori, S., Nomura, T., and Sakaguchi, S. (2003). Control of regulatory T cell development by the transcription factor Foxp3. Science *299*, 1057-1061.

Khattri, R., Cox, T., Yasayko, S. A., and Ramsdell, F. (2003). An essential role for Scurfin in CD4+CD25+ T regulatory cells. Nat Immunol *4*, 337-342.

Prokunina, L., Castillejo-Lopez, C., Oberg, F., Gunnarsson, I., Berg, L., Magnusson, V., Brookes, A. J., Tentler, D., Kristjansdottir, H., Grondal, G., *et al.* (2002). A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. Nat Genet *32*, 666-669.

Tokuhiro, S., Yamada, R., Chang, X., Suzuki, A., Kochi, Y., Sawada, T., Suzuki, M., Nagasaki, M., Ohtsuki, M., Ono, M., *et al.* (2003). An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. Nat Genet *35*, 341-348.

Voorhoeve, P. M., le Sage, C., Schrier, M., Gillis, A. J., Stoop, H., Nagel, R., Liu, Y. P., van Duijse, J., Drost, J., Griekspoor, A., *et al.* (2006). A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Cell *124*, 1169-1181.

Wildin, R. S., Ramsdell, F., Peake, J., Faravelli, F., Casanova, J. L., Buist, N., Levy-Lahad, E., Mazzella, M., Goulet, O., Perroni, L., *et al.* (2001). X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. Nat Genet *27*, 18-20.

Wu, J., Katrekar, A., Honigberg, L. A., Smith, A. M., Conn, M. T., Tang, J., Jeffery, D., Mortara, K., Sampang, J., Williams, S. R., *et al.* (2006a). Identification of substrates of human protein-tyrosine phosphatase PTPN22. J Biol Chem *281*, 11002-11010.

Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A. D., Stroud, J. C., Bates, D. L., Guo, L., Han, A., Ziegler, S. F., *et al.* (2006b). FOXP3 controls regulatory T cell function through cooperation with NFAT. Cell *126*, 375-387.

Appendix A: Supplementary Material

Connecting microRNA genes to the core transcriptional circuitry of embryonic stem cells

Supplementary Methods and Discussion

Index of Supplementary Tables

Index of Supplementary Figures

Figure S1 Promoters for known genes occupied by Oct4/Sox2/Nanog/Tcf3 in mES cells.
Figure S2 Comparison of ChIP-seq and ChIP-chip genome wide data for Oct4, Nanog, and Tcf3.
Figure S3 High resolution analysis of Oct4/Sox2/Nanog/Tcf3 binding based on Meta-analysis
Figure S4 Algorithm for Identification of miRNA promoters.
Figure S5 Summary of miRNA promoter classification.
Figure S6 miRNA genes occupied by the core master regulators in ES cells are expressed in induced Pluripotent Stem cells (iPS).
Figure S7 Tissue specific expression of PcG bound miRNAs

Index of Supplementary Files

The following files contain data formatted for upload into the UCSC genome browser (Kent et al., 2002). To upload the files, first copy the files onto a computer with internet access. Then use a web browser to go to http://genome.ucsc.edu/cgi-bin/hgCustom?hgsid=105256378 for mouse and http://genome.ucsc.edu/cgi-bin/hgCustom?hgsid=104842340 for human. In the "Paste URLs or Data" section, select "Browse..." on the right of the screen. Use the pop-up window to select the copied files, then select "Submit". The upload process may take some time.

mouse_miRNA_track.mm8.bed – Map of predicted miRNA genes in mouse. Transcripts with EST or gene evidence are shown as black lines. Presumed transcripts are shown as grey lines. Positions of the mature miRNAs are annotated as thicker lines.
human_miRNA_track.hg17.bed – Map of predicted miRNA genes in human. Transcripts with EST or gene evidence are shown as black lines. Presumed transcripts are shown as grey lines. Positions of the mature miRNAs are annotated as thicker lines.
mES_regulator_ChIPseq.mm8.WIG.gz – ChIP-seq data for Oct4, Sox2, Nanog and Tcf3 in mES cells. Top track for each data set illustrates the normalized number of reads assigned to each 25bp bin. Bars in the second track identify regions of the genome enriched at $p < 10^{-9}$.
mES_chromatin_ChIPseq.mm8.WIG.gz – ChIP-seq data for H3K4me3, H3K79me2, H3K36me3 and Suz12 in mES cells. Top track for each data set illustrates the normalized number of reads assigned to each 25bp bin. Bars in the second track identify regions of the genome enriched at $p < 10^{-9}$. Enriched regions were not identified for H3K79me2 and H3K36me3

Supplemental References

Supplementary Tables and Files are available from the Young Lab

Growth Conditions and Quality Control for Human Embryonic Stem Cells

Human embryonic stem (ES) cells were obtained from WiCell (Madison, WI; NIH Code WA09) and grown as described. Cell culture conditions and harvesting have been described previously (Boyer et al., 2005; Lee et al., 2006; Guenther et al., 2007). Quality control for the H9 cells included immunohistochemical analysis of pluripotency markers, alkaline phosphatase activity, teratoma formation, and formation of embryoid bodies and has been previously published as supplemental material (Boyer et al., 2005; Lee et al., 2006).

Growth Conditions for Murine Embryonic Stem Cells

V6.5 (C57BL/6-129) murine ES cells were grown under typical ES cell culture conditions on irradiated mouse embryonic fibroblasts (MEFs) as previously described (Boyer et al., 2006). Briefly, cells were grown on gelatinized tissue culture plates in DMEM-KO (Gibco/Invitrogen) supplemented with 15% fetal bovine serum (characterized from Hyclone), 1000 U/ml leukemia inhibitory factor (LIF) (Chemicon; ESGRO ESG1106), non-essential amino acids, L-glutamine, Penicillin/Streptomycin andß-mercaptoethanol. Immunostaining was used to confirm expression of pluripotency markers, SSEA 1 (Developmental Studies Hybridoma Bank) and Oct4 (Santa Cruz, SC-5279). For location analysis, cells were grown for one passage off of MEFs, on gelatinized tissue-culture plates.

Antibodies

Oct4-bound genomic DNA was enriched from whole cell lysate using an epitope specific goat polyclonal antibody purchased from Santa Cruz (sc-8628) and compared to a reference whole cell extract (Boyer et al., 2005). A summary of regions occupied with high confidence for this antibody identified by ChIP-seq in mES cells is provided in **Table S3** and by ChIP-chip on genome-wide tiling arrays in hES cells are on **Table S8**. Oct4 ChIP-seq data can be visualized on the UCSC browser by uploading supplemental file:
mES_regulator_ChIPseq.mm8.WIG.gz

Sox2-bound genomic DNA was enriched from whole cell lysate using an affinity purified goat polyclonal antibody purchased from R&D Systems (AF2018) and compared to a reference whole cell extract (Boyer et al., 2005). A summary of regions occupied with high confidence for this antibody identified by ChIP-seq in mES cells is provided in **Table S3**. Sox2 ChIP-seq data can be visualized on the UCSC browser by uploading supplemental file:
mES_regulator_ChIPseq.mm8.WIG.gz

Nanog-bound genomic DNA was enriched from whole cell lysate using an affinity purified rabbit polyclonal antibody purchased from Bethyl Labs (bl1662) and compared to a reference whole cell extract (Boyer et al., 2005).. A summary of regions bound with high confidence for this antibody is provided in **Table S3**.

Nanog ChIP-seq data can be visualized on the UCSC browser by uploading supplemental file:

mES_regulator_ChIPseq.mm8.WIG.gz

Tcf3-bound genomic DNA was enriched from whole cell lysate using an epitope specific goat polyclonal antibody purchased from Santa Cruz (sc-8635) and compared to a reference whole cell extract (Cole et al., 2008). A summary of regions occupied with high confidence for this antibody identified by ChIP-seq in mES cells is provided in **Table S3**. Tcf3 ChIP-seq data can be visualized on the UCSC browser by uploading supplemental file:

mES_regulator_ChIPseq.mm8.WIG.gz

Suz12-bound genomic DNA was enriched from whole cell lysate using an affinity purified rabbit polyclonal antibody purchased from Abcam (AB12073) and compared to a reference whole cell extract (Lee et al., 2006). A summary of regions bound with high confidence for this antibody is provided in **Table S10**. Suz12 ChIP-seq data can be visualized on the UCSC browser by uploading:

supplemental file mES_chomatin_ChIPseq.mm8.WIG.gz

H3K4me3-modified nucleosomes were enriched from whole cell lysate using an epitope-specific rabbit polyclonal antibody purchased from Abcam (AB8580) (Santos-Rosa et al., 2002; Guenther et al., 2007). Samples were analyzed using ChIP-seq. Comparison of this data with ChIP-seq published previously

(Mikkelsen et al., 2007) showed near identify in profile and bound regions (**Table S5**). H3K4me3 ChIP-seq data can be visualized on the UCSC browser by uploading supplemental file:
mES_chomatin_ChIPseq.mm8.WIG.gz

H3K79me2-modified nucleosomes were isolated from mES whole cell lysate using Abcam antibody AB3594 (Guenther et al., 2007). Chromatin immunoprecipitations against H3K36me3 were compared to reference WCE DNA obtained from mES cells. Samples were analyzed using ChIP-seq and were used for visual validation of predicted miRNA promoter association with mature miRNA sequences only **(Figure 2)**. H3K79me2 ChIP-seq data can be visualized on the UCSC browser by uploading supplemental file:
mES_chomatin_ChIPseq.mm8.WIG.gz

H3K36me3-modified nucleosomes were isolated from mES whole cell lysate using rabbit polyclonal antibody purchased from Abcam (AB9050) (Guenther et al., 2007). Chromatin immunoprecipitations against H3K36me3 were compared to reference WCE DNA obtained from mES cells. Samples were analyzed using ChIP-seq and were used for visual validation of predicted miRNA promoter association with mature miRNA sequences only **(Figure 2)**. H3K36me3 ChIP-seq data can be visualized on the UCSC browser by uploading supplemental file: mES_chomatin_ChIPseq.mm8.WIG.gz

Chromatin Immunoprecipitation

Protocols describing all materials and methods have been previously described (Lee et al. 2007) and can be downloaded from:

http://web.wi.mit.edu/young/hES_PRC/

Briefly, we performed independent immunoprecipitations for each analysis. Embryonic stem cells were grown to a final count of $5x10^7$ – $1x10^8$ cells for each location analysis experiment. Cells were chemically crosslinked by the addition of one-tenth volume of fresh 11% formaldehyde solution for 15 minutes at room temperature. Cells were rinsed twice with 1xPBS and harvested using a silicon scraper and flash frozen in liquid nitrogen. Cells were stored at –80°C prior to use.

Cells were resuspended, lysed in lysis buffers and sonicated to solubilize and shear crosslinked DNA. Sonication conditions vary depending on cells, culture conditions, crosslinking and equipment. We used a Misonix Sonicator 3000 and sonicated at approximately 28 watts for 10 x 30 second pulses (90 second pause between pulses). For ChIP of Oct4, Nanog, Tcf3 and Suz12 in murine ES cells, SDS was added to lysate after sonication to a final concentration of 0.1%. Samples were kept on ice at all times.

The resulting whole cell extract was incubated overnight at 4°C with 100 µl of Dynal Protein G magnetic beads that had been preincubated with approximately

10 µg of the appropriate antibody. Beads were washed 4-5 times with RIPA buffer and 1 time with TE containing 50 mM NaCl. For ChIP of Oct4, Nanog, Tcf3 and Suz12 in murine ES cells, the following 4 washes for 4 minutes each were used instead of RIPA buffer: 1X low salt (20mM Tris pH 8.1, 150mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS), 1X high salt (20mM Tris pH 8.1, 500mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS), 1X LiCl (10mM Tris pH 8.1, 250mM LiCl, 1mM EDTA, 1% deoxycholate, 1% NP-40), and 1X TE+ 50mM NaCl. Bound complexes were eluted from the beads by heating at 65°C with occasional vortexing and crosslinking was reversed by overnight incubation at 65°C. Whole cell extract DNA (reserved from the sonication step) was also treated for crosslink reversal.

ChIP-Seq Sample Preperation and Analysis

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

Sample Preparation

Purified immunoprecipitated (ChIP) DNA were prepared for sequencing according to a modified version of the Illumina/Solexa Genomic DNA protocol. Fragmented DNA was prepared for ligation of Solexa linkers by repairing the ends and adding a single adenine nucleotide overhang to allow for directional

ligation. A 1:100 dilution of the Adaptor Oligo Mix (Illumina) was used in the ligation step. A subsequent PCR step with limited (18) amplification cycles added additional linker sequence to the fragments to prepare them for annealing to the Genome Analyzer flow-cell. After amplification, a narrow range of fragment sizes was selected by separation on a 2% agarose gel and excision of a band between 150-300 bp (representing shear fragments between 50 and 200nt in length and ~100bp of primer sequence). The DNA was purified from the agarose and diluted to 10 nM for loading on the flow cell.

Polony generation on Solexa Flow-Cells

The DNA library (2-4 pM) was applied to the flow-cell (8 samples per flow-cell) using the Cluster Station device from Illumina. The concentration of library applied to the flow-cell was calibrated such that polonies generated in the bridge amplification step originate from single strands of DNA. Multiple rounds of amplification reagents were flowed across the cell in the bridge amplification step to generate polonies of approximately 1,000 strands in 1μm diameter spots. Double stranded polonies were visually checked for density and morphology by staining with a 1:5000 dilution of SYBR Green I (Invitrogen) and visualizing with a microscope under fluorescent illumination. Validated flow-cells were stored at 4°C until sequencing.

Sequencing

Flow-cells were removed from storage and subjected to linearization and annealing of sequencing primer on the Cluster Station. Primed flow-cells were loaded into the Illumina Genome Analyzer 1G. After the first base was incorporated in the Sequencing-by-Synthesis reaction the process was paused for a key quality control checkpoint. A small section of each lane was imaged and the average intensity value for all four bases was compared to minimum thresholds. Flow-cells with low first base intensities were re-primed and if signal was not recovered the flow-cell was aborted. Flow-cells with signal intensities meeting the minimum thresholds were resumed and sequenced for 26 cycles.

Solexa Data Analysis

Images acquired from the Illumina/Solexa sequencer were processed through the bundled Solexa image extraction pipeline, which identified polony positions, performed base-calling and generated QC statistics. Sequences were aligned using the bundled ELAND software using murine genome NCBI Build 36 and 37 (UCSC mm8, mm9) as the reference genome. Alignments to build 37 were used for analysis of the mmu-mir-290 cluster only as that cluster is not represented on build 36. Only sequences perfectly and uniquely mapping to the genome were used. A summary of the number of reads used is shown in **Table S1**.

The analysis methods used were derived from previously published methods (Johnson et al., 2007, Mikkelsen et al., 2007). Sequences from all lanes for each chromatin IP were combined, extended 200bp (maximum fragment length

accounting for ~100bp of primer sequence), and allocated into 25 bp bins. Genomic bins containing statistically significant ChIP-seq enrichment were identified by comparison to a Poissonian background model, using a p-value threshold of 10^{-9}. A list of the numbers of counts in a genomic bins required for each sample to meet this threshold are provided in **Table S1**. Addtionally, we used and empirical background model obtained from identical Solexa sequencing of DNA from whole cell extract (WCE) from matched cell samples (> 5x normalized enrichment across the entire region, see below). A summary of the bound regions and their relation to gene targets can be found in **Tables S2, S3, S5 and S10**.

The p-value threshold was selected to minimize the expected false-positive rate. Assuming background reads are spread randomly throughout the genome, the probability of observing a given number of counts can be modelled as a Poisson process where the expectation can be calculated as the number of mapped reads times the number of bins per read (8) divided by the total number of bins available (we assumed 50% as a very conservative estimate). With the genome divided into ~10^8 bins of 25 bp, a probability of $p<10^{-9}$ represents the likelihood that ~1 experiment in 10 will randomly enrich one bin in the genome.

The Poisson background model assumes a random distribution of binding events, however we have observed significant deviations from this expectation in ChIP-seq datasets. These non-random events can be detected as sites of

enrichment using control IPs and create a significant number of false positive events for actual ChIP-seq experiments. To remove these regions, we compared genomic bins and regions that meet the statistical threshold for enrichment to an empirical distribution of reads obtained from Solexa sequencing of DNA from whole cell extract (WCE) from matched cell samples. We required that enriched regions have five-fold greater ChIP-seq density in the specific IP sample as compared with the non-specific WCE sample, normalized for the total number of reads. This served to filter out genomic regions that are biased to having a greater than expected background density of ChIP-seq reads. We observed that ~200-500 regions in the genome showed non-specific enrichment in these experiments.

Identifications of regions enriched for Oct4/Sox2/Nanog/Tcf3

The identification of enriched regions in ChIP-chip and ChIP-seq experiments is typically done using threshold for making a binary determination of enriched or not enriched. Unfortunately, there is not actually a clear delineation between truly bound and and unbound regions. Instead, enrichment is a continuum and the threshold is set to minimize false positives (high-confidence sites). This typically requires that thresholds be set at a level that allows a high false-negative rate (~30% for ChIP-chip, Lee et al). When multiple factors are compared, focusing only on the intersection of the different data sets compounds

this effect, leading to higher false negative rates and the loss of many critical target genes.

Oct4, Sox2, Nanog and Tcf3 co-occupy promoters throughout the genome (Cole, **Figure 1)** and cluster analysis of enriched sites reveals apparent co-enrichment for all 4 factors at >90% of sites (Frampton & Young, unpublished data). However, the overlap for any two factors at the cut-off for high-confidence enrichment is only about two thirds **(Figure S1, Tables S2 and S3)**. Therefore many of these sites must have enrichment that is below the high-confidence threshold for at least some of the participating factors. Variability in the enrichment observed for each factor at different binding sites is common in the data **(Figures 1b, 3, and S2)**.

To determine a threshold of binding for multiple factors, we used two complementary methods to examine high-confidence targets of the four regulators. First, the classes of genes enriched by different numbers of factors at high-confidence were compared to the known classes of targets based on gene ontology **(Figure S1b,** http://gostat.wehi.edu.au/cgi-bin/goStat.pl, Beissbarth and Speed, 2004**)**. The highest confidence targets (those with high levels of immuno-enrichment observed for all for factors) preferentially encoded factors involved in DNA binding, regulation of transcription and development as has been previously shown (Boyer et al., 2005). These gene ontology categories continued to be overrepresented among high-confidence targets of either 3 of the 4 factors or 2

of 4 the factors, albeit at lower levels, but were barely enriched among high confidence targets of only one factor.

As a second test, we examined how different numbers of overlapping high-confidence targets affected the overlap with our previous genome-wide studies using ChIP-chip. Because not all regions of the genome are tiled with equal density on the microarrays used for ChIP-chip, we first determined the minimum probe density required to confirm binding detected by ChIP-seq (**Figure S2**). At most genes with high probe density, the ChIP-seq and ChIP-chip data were very highly correlated. However, regions of the genome with microarray coverage of less than three probes per kilobase were generally unreliable in detecting these enrichment. These regions, which had low probe coverage on the microarrays, represent approximately 1/3 of all sites co-enriched for the four factors by ChIP-seq. In regions where probe density was greater than three probe per kilobase the fraction of ChIP-seq sites confirmed by ChIP-chip experiments increased with additional factors co-binding with a large fall off below 2 factors (data not shown). Based on these two analyses, we elected to choose targets occupied at high-confidence by 2 or more of the 4 factors tested for further analysis in this manuscript. **(Figures 1a and S1a [red line]).**

While a majority of the miRNA promoters identified as occupied by Oct4/Sox2/Nanog/Tcf3 are not occupied by all four factors at high-confidence, it is interesting to note that all of the miRNA genes that share highly similar seeds

to miR-302 are occupied at high confidence by all four factors (miR-302 cluster, miR-290 cluster and miR-106a cluster), similar to the promoters of core transcriptional regulators of ES cells. By comparison, promoters also occupied by Suz12 almost never showed high-confidence binding for all four factors (**Table S4**, see mmu-miR-9-2 in **Figure 3**). Similar effects were observed for protein-coding genes in mES cells (Lee et al., 2006). Whether this is caused by reduced epitope availability in PcG bound regions or reflects reduced protein binding is unclear.

DNA Motif Discovery and High-resolution Binding-Site Analysis

DNA motif discovery was performed on the genomic regions that were enriched for Oct4 at high-confidence. In order to obtain maximum resolution, a modified version of the ChIP-seq read mapping algorithm was used. Genomic bins were reduced in size from 25 bp to 10 bp. Furthermore, a read extension that placed greater weight towards the middle of the 200 bp extension was used. This model placed 1/3 count in the 8 bins from 0-40 and 160-200 bp, 2/3 counts in the 8 bins from 40-80 and 120-160 bp and 1 count in the 4 bins from 80-120 bp. This allowed increased precision for determination of the peak of ChIP-seq density in each Oct4 bound region. 100bp surrounding the 500 Oct4 bound regions with the greatest peak ChIP-seq density were submitted to the motif disocvery tool MEME (Bailey and Elkan, 1995; Bailey, 2006) to search for over-represented DNA motifs. A single sixteen basepair motif was discovered by the MEME

algorithm (**Table S4, Figure S2i**). This motif was significantly ($p<10^{-100}$) over-represented in the Oct4 bound input sequences and occurred in 445 of the 500 hundred basepair sequences.

As a default, MEME uses the individual nucleotide frequencies within input sequences to model expected motif frequencies. This simple model might result discovery of motifs which are enriched because of non-random di-, tri-, etc. nucleotide frequencies. Consequently, three different sets of control sequences of identical length were used to ensure the specificity of the motif discovery results. First, the sequences immediately flanking each input sequence were used as control sequences. Second, randomly selected sequences having the same distribution of distances from transcription start sites as the Oct4 input sequences were used as control sequences. Third, sequences from completely random genomic regions were used as control sequences. Each of these sets of control seqeunces were also examined using MEME. For each of these controls, the motif discovered from actual Oct4 bound sequences was not identified in the control sequences.

The motif discovery process was repeated using different numbers and lengths of sequences, but the same motif was discovered for a wide array of input sequences. Furthermore, when motif discovery was repeated with the top 500 Sox2, Nanog, and Tcf3 occupied regions, the same motif was identified. Overall, the motif occurs within 100 bp of the peak of ChIP-seq density at more than 90%

of the top regions enriched in each experiment, while occuring in the same span at 24-28% of control regions and within 25 bp of the ChIP-seq peak at more than 80% of regions versus 9-11% of control regions.

We next attempted to determine the precise sites on the genome bound by Oct4, Sox2, Nanog, and Tcf3 at basepair resolution using composite analysis of the bound regions for each factor. In particular, we examined if the different factors tended to associate with specific sequences within the assymetric DNA motif identified at a high fraction of the sites occupied by Oct4, Sox2, Nanog, and Tcf3. A set of ~2,000 of the highest confidence bound regions was determined for each factor based on a count threshold approximately two fold higher than the threshold for high-confidence regions shown in **Table S1** (Poissoin: $p < 10^{-9}$). Regions without a motif within 50bp of the peak of ChIP-seq enrichment, typically ~10% of regions, were removed from this analysis. The distance from the first base of the central motif in each bound region to the 5' end of all reads within 250bp was tabulated, keeping reads mapping to the same strand as the motif separate from reads mapping to the oppositie strand. The difference in ChIP-seq read frequency between reads mapping to the same strand as the motif and the reads mapping to the oppositite strand was calculated at every basepair within the 500 bp window **Figure S3**. We made the assumption that the precise peak of the ChIP-seq distribution was the point at which this strand bias was equal to zero.

To determine the precise position where the strand bias was equal to zero, we created a simplified model of the strand bias for each transcription factor. We chose a function with 4 parameters (A,B, C, and M), one of which (M) was the point at which the curve crosses the x-axis.

$$f(x) = A \times \arctan\left(\frac{x-M}{B}\right) \times e^{-\left|\frac{x-M}{B}\right|^{C}}$$

Least squares curve fitting was performed GNUplot (http://www.gnuplot.info/) using an approximated set of initial conditions (A = -1000, B = 100, C = 2, M = 10). The variablity in M was detemined by monte carlo simulation (n=25) using a random set of half of the ChIP-seq reads in each dataset and is shown in **Figure S3.**

Identification of miRNA promoters in human and mouse

To better understand the regulation of miRNAs, we sought to identify the sites of transcription initiation for all miRNAs in both human and mouse, at least to low resolution (~1kb). Most methods used to identify promoters require active transcription of the miRNA and isolation of rare primary miRNA transcripts. We decided to use an approach based on *in vivo* chromatin signature of promoters. This approach has two principle advantages. First, the required data has been published by a variety of laboratories and is readily accessible and second, it does not require the active transcription of the miRNA primary transcript.

Recent results using genome-wide location analysis of H3K4me3 indicate that between 60 and 80% of all protein-coding genes in any cell population have promoters enriched in methylated nucleosomes, even where the gene is not detected by typical transcription profiling (Guenther et al., 2007) Importantly, over 90% of the H3K4me3 enriched regions in these cells map to known or predicted promoters, suggesting that H3K4me3 can be used as a proxy for sites of active initiation. Our strategy to identify miRNA promoters, therefore, uses H3K4me3 enriched sites from as many sources as possible as a collection of promoters. In human, H3K4me3 sites were identified in ES cells (H9), hepatocytes, a pro-B cell line (REH cells) (Guenther et al., 2007) and T cells (Barski et al., 2007). Mouse H3K4me3 sites were identified from ES cells (V6.5), neural precursors, and embryonic fibroblasts (Mikkelsen et al., 2007). In total, we identified 34,793 high-confidence H3K4me3 enriched regions in human and 34,096 high-confidence regions enriched in mouse, collectively present at ~75% of all protein-coding genes.

The list of miRNAs identified in the miRNA atlas (Landgraf et al., 2007) was used as the basis for our identification. The total list consists of 496 miRNAs in human, 382 miRNAs in mouse. ~65% of the murine miRNAs can be found in both species.

For each of these miRNAs, possible start sites were derived from both all H3K4me3 enriched regions within 250kb upstream of the miRNA as well as all known start sites for any miRNAs that were identified as being within known transcripts from RefSeq (Pruitt et al., 2005) Mammalian Gene Collection (MGC) (Gerhard et al., 2004) Ensembl (Hubbard et al., 2005), or University of California Santa Cruz (UCSC) Known Genes (genome.ucsc.edu) (Kent et al., 2002) for which EntrezGene (http://www.ncbi.nlm.nih.gov/entrez/) gene IDs had been generated. Where an annotated start site was found to overlap an H3K4me3 enriched region, the known start was used in place of the enriched region.

A scoring system was derived empirically to select the most likely start sites for each miRNA. Each possible site was given a bonus if it was either the start of a known transcript that spanned the miRNA or of an EST that spanned the miRNA. Scores were reduced if the H3K4me3 enriched region was assignable instead to a transcript or EST that did not overlap the miRNA. Additional positive scores were given to enriched sites within 5kb of the miRNA, while additional negative scores were given based on the number of intervening H3K4me3 sites between the test region and the miRNA. Finally, each enriched region was tested for conservation between human and mouse using the UCSC liftover program (Hinrichs et al., 2006). If two test regions overlapped, they were considered to be conserved (21%). In the cases where human and mouse disagreed on the quality of a site, if the site had an EST or gene overlapping the miRNA, that site was given a high score in both species. Alternatively, if one species had a non-

overlapping site, that site was considered to be an unlikely promoter in both species. Finally, for miRNAs where a likely promoter was identified in only one species, we manually checked the homologous region of the other genome to search for regions enriched for H3K4me3-modified nucleosomes that may have fallen below the high-confidence threshold. Start sites were considered to be likely if the total score was ≥ 0 **(Figure S4 and S5)**. In total, we identified likely start sites for ~85% of all miRNAs in both species **(Tables S6 and S7)**. Predicted miRNA genes can be visualised on the UCSC browser by uploading the supplemental files:

mouse_miRNA_track.mm8.bed and **human_miRNA_track.hg17.bed**

Several lines of evidence suggest the high quality of these predictions. First, previous studies have found that miRNAs within 50kb of each other are likely to be co-regulated (Lagos-Quintana et al., 2001; Lau et al., 2001). While the nature of these clusters was not included in our analysis, nearly all miRNAs within a cluster end up identifying the same promoter region (see **Figures 2, 3, 5 and S3)**. The only exceptions to this are found in the large clusters of repeat derived miRNAs found in chromosome 12 of mouse and chromosome 14 in human where a single H3K4me3 enriched region splits the clusters. Second, consistent with the frequent association of CpG islands with the transcriptional start sites for protein-coding genes, ~50% of the miRNA promoters identified here overlap CpG islands **(Tables S6 and S7)**. Finally, for miRNAs that were active in ES cells,

histone modifications associated with elongation were able to "connect" the mature miRNAs to the predicted transcription start site **(Figure 2)**.

To further ascertain the accuracy of our promoter predictions, we compared our predicted start sites to those identified in recent studies. Predictions were tested against mmu-mir-34b / mmu-mir-34c (Corney et al., 2007), hsa-mir-34a (Chang et al., 2007) mmu-mir-101a, mmu-mir-202, mmu-mir-22, mmu-mir-124a-1, mmu-mir-433 (Fukao et al., 2007), and hsa-mir-17/18a/19a/20a/19b-1/92a-1 (O'Donnell et al., 2005). Additional miRNA promoters in these manuscripts were not predicted strongly by the above algorithm. For these 14 miRNAs, H3K4me3 sites were identified within 1kb of all but two of the sites. mmu-mir-202 was predicted about 20kb upstream of the annotated start site, but may reflect an H3K4me3 site absent from the tissues sampled. mmu-mir-433 is in the middle of a large cluster of miRNAs on mouse chromosome 12. The annotated TSS lies within the cluster between mir-433 and mir-431 suggesting the promoter may be incorrect. Overall, the accuracy of the promoter predictions is believed to be ~75% (6/8). Additional H3K4me3 data sets and EST data should allow for improved accuracy in predicting and validating these initiation sites.

ChIP-chip Sample Preparation and Analysis

Immunoprecipitated DNA and whole cell extract DNA were purified by treatment with RNAse A, proteinase K and multiple phenol:chloroform:isoamyl alcohol extractions. Purified DNA was blunted and ligated to linker and amplified using a

two-stage PCR protocol. Amplified DNA was labeled and purified using Bioprime random primer labeling kits (Invitrogen): immunoenriched DNA was labeled with Cy5 fluorophore, whole cell extract DNA was labeled with Cy3 fluorophore.

Labeled DNA was mixed (~5 µg each of immunoenriched and whole cell extract DNA) and hybridized to arrays in Agilent hybridization chambers for up to 40 hours at 40°C. Arrays were then washed and scanned.

Slides were scanned using an Agilent DNA microarray scanner BA. PMT settings were set manually to normalize bulk signal in the Cy3 and Cy5 channel. For efficient batch processing of scans, we used Genepix (version 6.0) software. Scans were automatically aligned and then manually examined for abnormal features. Intensity data were then extracted in batch.

44k Human Whole Genome Array

The human promoter array was purchased from Agilent Technology (www.agilent.com). The array consists of 115 slides each containing ~44,000 60mer oligos designed to cover the non-repeat portion of the human genome. The design of these arrays are discussed in detail elsewhere (Lee et al., 2006).

Data Normalization and Analysis

We used GenePix software (Axon) to obtain background-subtracted intensity values for each fluorophore for every feature on the whole genome arrays. Among the Agilent controls is a set of negative control spots that contain 60-mer

sequences that do not cross-hybridize to human genomic DNA. We calculated the median intensity of these negative control spots in each channel and then subtracted this number from the intensities of all other features.

To correct for different amounts of each sample of DNA hybridized to the chip, the negative control-subtracted median intensity value of control oligonucleotides from the Cy3-enriched DNA channel was then divided by the median of the control oligonucleotides from the Cy5-enriched DNA channel. This yielded a normalization factor that was applied to each intensity in the Cy5 DNA channel.

Next, we calculated the log of the ratio of intensity in the Cy3-enriched channel to intensity in the Cy5 channel for each probe and used a whole chip error model to calculate confidence values for each spot on each array (single probe p-value). This error model functions by converting the intensity information in both channels to an X score which is dependent on both the absolute value of intensities and background noise in each channel using an f-score calculated as described for promoter regions or using a score of 0.3 for tiled arrays. When available, replicate data were combined, using the X scores and ratios of individual replicates to weight each replicate's contribution to a combined X score and ratio. The X scores for the combined replicate are assumed to be normally distributed which allows for calculation of a p-value for the enrichment ratio seen at each feature. P-values were also calculated based on a second model assuming that, for any range of signal intensities, IP:control ratios below 1

represent noise (as the immunoprecipitation should only result in enrichment of specific signals) and the distribution of noise among ratios above 1 is the reflection of the distribution of noise among ratios below 1.

High Confidence Enrichment

To automatically determine bound regions in the datasets, we developed an algorithm to incorporate information from neighboring probes. For each 60-mer, we calculated the average X score of the 60-mer and its two immediate neighbors. If a feature was flagged as abnormal during scanning, we assumed it gave a neutral contribution to the average X score. Similarly, if an adjacent feature was beyond a reasonable distance from the probe (1000 bp), we assumed it gave a neutral contribution to the average X score. The distance threshold of 1000 bp was determined based on the maximum size of labeled DNA fragments put into the hybridization. Since the maximum fragment size was approximately 550 bp, we reasoned that probes separated by 1000 or more bp would not be able to contribute reliable information about a binding event halfway between them.

This set of averaged values gave us a new distribution that was subsequently used to calculate p-values of average X (probe set p-values). If the probe set p-value was less than 0.001, the three probes were marked as potentially bound.

As most probes were spaced within the resolution limit of chromatin immunoprecipitation, we next required that multiple probes in the probe set provide evidence of a binding event. Candidate bound probe sets were required to pass one of two additional filters: two of the three probes in a probe set must each have single probe p-values < 0.005 or the centre probe in the probe set has a single probe p-value < 0.001 and one of the flanking probes has a single point p-value < 0.1. These two filters cover situations where a binding event occurs midway between two probes and each weakly detects the event or where a binding event occurs very close to one probe and is very weakly detected by a neighboring probe. Individual probe sets that passed these criteria and were spaced closely together were collapsed into bound regions if the centre probes of the probe sets were within 1000 bp of each other.

Comparing Enriched Regions to Known Genes and miRNAs

Enriched regions were compared relative to transcript start and stop coordinates of known genes compiled from four different databases: RefSeq (Pruitt et al., 2005), Mammalian Gene Collection (MGC) (Gerhard et al., 2004), Ensembl (Hubbard et al., 2005), and University of California Santa Cruz (UCSC) Known Genes (genome.ucsc.edu) (Kent et al., 2002). All human coordinate information was downloaded in January 2005 from the UCSC Genome Browser (hg17, NCBI build 35). Mouse data was downloaded in June of 2007 (mm8, NCBI build 36).

To convert bound transcription start sites to more useful gene names, we used conversion tables downloaded from UCSC and Ensembl to automatically assign

EntrezGene (http://www.ncbi.nlm.nih.gov/entrez/) gene IDs and symbols to the RefSeq, MGC, Ensembl, UCSC Known Gene. Comparisons of Oct4, Sox2, Nanog, Tcf3, H3K4me3 and Suz12 to annotated regions of the genomes can be found in Tables S3, S5, S8 and S10

For miRNAs start sites, two separate windows were used to evaluate overlaps. For chromatin marks and non-sequence specific proteins, miRNA promoters were considered bound if they were within 1kb of an enriched sequence. For sequence specific factors such as Oct4, we used a more relaxed region of 8kb surrounding the promoter, consistent with previous work we have published (Boyer et al., 2005). A full list of the high confidence start sites bound to promoters can be found in Tables S6 and S7.

Growth Conditions for Neural Precursors, mouse embryonic fibroblasts, and induced pluripotent stem cells.

To generate neural precursor cells, ES cells were differentiated along the neural lineage using standard protocols. V6.5 ES cells were differentiated into neural progenitor cells (NPCs) through embryoid body formation for 4 days and selection in ITSFn media for 5–7 days, and maintained in FGF2 and EGF2 (R&D Systems) (Okabe et al., 1996).

Mouse embryonic fibroblasts were prepared from DR-4 strain mice as previously described (Tucker et al., 1997). Cells were cultured in Dulbecco's modified

Eagle medium supplemented with 10% cosmic calf serum, β-mercaptoethanol, non-essential amino acids, L-glutamine and penicillin/streptomycin.

Analysis of Mature miRNA Frequency by Solexa Sequencing

Polony generation on Solexa Flow-Cells

The DNA library (2-4 pM) was applied to the flow-cell (8 samples per flow-cell) using the Cluster Station device from Illumina. The concentration of library applied to the flow-cell was calibrated such that polonies generated in the bridge amplification step originate from single strands of DNA. Multiple rounds of amplification reagents were flowed across the cell in the bridge amplification step to generate polonies of approximately 1,000 strands in 1μm diameter spots. Double stranded polonies were visually checked for density and morphology by staining with a 1:5000 dilution of SYBR Green I (Invitrogen) and visualizing with a microscope under fluorescent illumination. Validated flow-cells were stored at 4°C until sequencing.

Sequencing and Analysis

Flow-cells were removed from storage and subjected to linearization and annealing of sequencing primer on the Cluster Station. Primed flow-cells were loaded into the Illumina Genome Analyzer 1G. After the first base was incorporated in the Sequencing-by-Synthesis reaction the process was paused for a key quality control checkpoint. A small section of each lane was imaged and the average intensity value for all four bases was compared to minimum

thresholds. Flow-cells with low first base intensities were re-primed and if signal was not recovered the flow-cell was aborted. Flow-cells with signal intensities meeting the minimum thresholds were resumed and sequenced for 36 cycles. Images acquired from the Illumina/Solexa sequencer were processed through the bundled Solexa image extraction pipeline which identified polony positions, performed base-calling and generated QC statistics. Sequences were then assigned to a miRNA if they perfectly matched at least the first 20bp of the mature miRNA sequences downloaded from targetScan (http://www.targetscan.org/). Mature miRNA frequencies were then normalized to each other by determining the expected frequency in mapped reads/million. A full list of the miRNAs detected can be found in **Table S9**.

miRNA Microarray Expression Analysis

Mouse embryonic fibroblasts and neural precursor cells were cultured as described above. Murine induced pluripotent (iPS) cells, derived as previously described (Wernig et al., 2007), were cultured under the same conditions as murine embryonic stem cells (described above). RNA was extracted with RNeasy (Qiagen) reagents. 5 µg total RNA from treated and control samples were labeled with Hy3™ and Hy5™ fluorescent label, using the miRCURY™ LNA Array labeling kit (Exiqon, Denmark) following the procedure described by the manufacturer. The labeled samples were mixed pair-wise and hybridized to the miRNA arrays printed using miRCURY™ LNA oligoset version 8.1 (Exiqon, Denmark). Each miRNA was printed in duplicate, on codelink slides (GE), using

GeneMachines Omnigrid 100. The hybridization was performed at 60C overnight using the Agilent Hybridization system - SurHyb, after which the slides were washed using the miRCURY™ LNA washing buffer kit (Exiqon, Denmark) following the procedure described by the manufacturer. The slides were then scanned using Axon 4000B scanner and the image analysis was performed using Genepix Pro 6.0.

Median minus background signal intensities for all microarray probes were tabulated and quantile normalized. Within each sample, each probe was given a signal value of the average signal of the probe of that rank, across the full dataset. Intensities were then floored at one unit and log normalized. Control probes were removed from further analysis. Statistically significant differential expression was calculated using the online NIA Array Analysis Tool (http://lgsun.grc.nia.nih.gov/ANOVA/).

Probes were tested for differential expression using the following settings:

Threshold z-value to remove outliers: 10000

Error Model: Max(Average,Bayesian)

Error variance averaging window: 100

Proportion of highest error variances to be removed: 0.01

Bayesian degrees of freedom: 5

FDR threshold: 0.10

Of 1008 probes, 230 were determined to be differentially expressed between 3 MEF and 2 ES samples. Expression data for the iPS samples were not used for identifying differentially expressed miRNAs.

For clustering and heat map display, expression data were Z-score normalized. Centroid linkage, Spearman rank correlation distance, hierarchical clustering of genes and arrays was performed using Gene Cluster 3.0 (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm#ctv). Heatmaps were generated using Java Treeview (http://jtreeview.sourceforge.net/) with color saturation at 0.6 standard deviations. Complete miRNA microarray expression data, differentially expression results, and clustergram data are provided (Supplementary Tables S99).

Tissue-Specificity of miRNAs

To determine the global tissue-specificity for miRNAs we used data from the recent publication of the miRNA atlas[4]. Specificity scores were taken from Table S34 Node 0 from Landgraf et al. (2007). Of the 45 distinct mature miRNAs with specificity scores >1 that are not bound only by Oct4/Sox2/Nanog/Tcf3, 16 were identified as Suz12 targets. These 16 represent over 40% of the distinct mature miRNAs whose promoters are occupied by Suz12 ($p < 5x10^{-4}$ for specificity scores > 1.3)

Identification of Oct4/Sox2/Nanog/Tcf3 occupied feed forward loops

To identify feed forward loops we examined the recent data set identifying functional targets of the miR-290 cluster (Sinkkonen et al., 2008). In their study, Sinkkonen et al. identified miR-290 targets by both looking at mRNAs that increase in level in a Dicer -/- cell line and overlap that data set with mRNAs that decrease in expression when miR-290 is added back to the cells. Because the promoter of the miR-290 gene is occupied by Oct4/Sox2/Nanog/Tcf3, any targets of the miRNA cluster that are also occupied by the 4 factors would represent feed forward targets. Of the 245 miR-290 cluster targets identified in the intersect of the two data sets, promoters for 64 are occupied by Oct4/Sox2/Nanog/Tcf3. This is approximately 50% more interactions then would be expected by random (binomial p-value < $1x10^{-4}$).

Interestingly, only a small minority of these genes is also occupied by significant quantities of the PRC2 subunit Suz12. Of the 64 targets whose promoters are occupied by Oct4/Sox2/Nanog/Tcf3, only 5 are occupied by domains of Suz12 binding >500bp (larger region sizes have been correlated with gene silencing, Lee et al., 2006). This may be because PcG bound proteins are not functional targets of mir-290 in mES cells or because these proteins are not expressed following Dicer deletion, they are excluded from the target list, but may be targets at other stages of development. In the later case, the miRNAs may serve as a redundant silencing mechanism for ES cells to help prevent even low levels of expression of the developmental regulators bound by PcG complexes.

SUPPLEMENTAL FIGURES

Figure S1. Promoters for known genes occupied by Oct4/Sox2/Nanog/Tcf3 in mES cells. a. Overlap of genes whose promoters are within 8kb of sites enriched for Oct4, Sox2, Nanog, or Tcf3. Not shown are the Nanog:Oct4 overlap (289) and Sox2:Tcf3 overlap (26). Red line deliniates genes considered occupied by Oct4/Sox2/Nanog/Tcf3. **b.** Enrichment for selected GO-terms previously reported to be associated with Oct4/Sox2/Nanog binding (Boyer et al., 2005) was tested on the sets of genes occupied at high-confidence for 1 to 4 of the tested DNA binding factors. Hypergeometric p-value is shown for genes annotated for DNA binding (blue), Regulation of Transcription (green) and Development (red).

Figure S2. Comparison of ChIP-seq and ChIP-chip genome wide data for Oct4, Nanog and Tcf3. a. Binding of Oct4 (blue), Nanog (orange) and Tcf3 (red) across 17kb surrounding the Wnt8a and Nodal genes (black below the graph, arrow indicates transcription start site) as in **Figure 1b**. (upper) Binding derived from ChIP-seq data plotted as reads per million. (lower) Binding derived from ChIP-chip enrichment ratios (Cole et. al., 2008) **b.** Poor probe density prevents detection of ~1/3 of ChIP-seq binding events on Agilent genome-wide tiling arrays. Top panel shows the fraction of regions that are occupied by Oct4/Sox2/Nanog/Tcf3 at high-confidence in mES cells as identified by ChIP-seq that are enriched for Oct4 (blue), Nanog (orange) and Tcf3 (red) on Agilent

genome-wide microarrays (Cole et al., 2008). Numbers on the x-axis define the boundaries used to classify probe densities for the histogram. Bottom panel illustrates a histogram of the microarray probe densities of the enriched regions identified.

Figure S3. High resolution analysis of Oct4/Sox2/Nanog/Tcf3 binding based on Meta-analysis. a-d. Short sequence reads for **a.** Oct4, **b**. Sox2, **c.** Nanog, **d.** Tcf3 mapping within 250bp of 2000 highly enriched regions where the peak of binding was found within 50bp of a high quality Oct4/Sox2 motif were collected. Composite profiles were created at base pair resolution for forward and reverse strand reads centered on the Oct4/Sox2 motif (aligned at +1). The difference between the number of positive and negative strand reads are shown for each base pair (circles). The best fit line is shown for each factor (see Supplemental Text). **e-h.** Zoomed in region of a-d showing 20bp surrounding the Oct4/Sox2 motif. Dashed line indicates the position where the best fit line crosses the X-axis. For reference, the motif is shown below each graph. **i.** Summary of meta-analysis for Oct4, Sox2, Nanog and Tcf3. Arrows indicate the nucleotide where each transcription factor switches from a positive strand bias to a negative strand bias. The octomer and HMG box motifs are indicated.

Figure S4. Algorithm for Identification of miRNA promoters. a, Flowchart describing the method used to identify the promoters for primary miRNA transcripts in human and mouse. For a full description, see supplemental text. **b,**

Two examples of identification of miRNA promoters. Top, Initial identification of possible start sites based on H3K4me3 enriched regions from four cell types. Enrichment of H3K4me3-modified nucleosomes is shown as shades of gray. Red bar represents the position of the mature miRNA. Black bars below the graph are regions enriched for H3K4me3. Initial scores are shown below the black bars. The region on the far right was excluded from the analysis (score = X) since it is downstream of the mature miRNA. Middle, Identification of candidate start sites <5kb upstream of the mature miRNA (yellow shaded area). Bottom, identification of candidate start sites that either initiate overlapping (left) or non-overlapping (right) transcripts. EST and transcript data is shown. Scores associated with identified genes are shown bold.

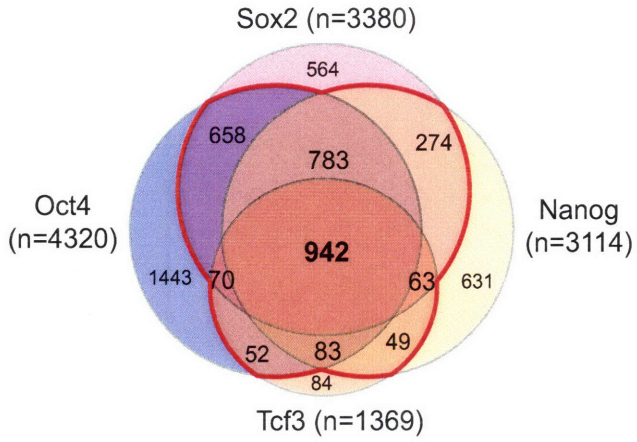
Figure S5. Summary of miRNA promoter classification. a. Promoters assigned to mature miRNAs were classified by the dominant feature of their scoring. Green: miRNAs that were found to have overlapping ESTs or genes confirming their promoters. Orange: miRNAs that were found to have a candidate start site within 5kb of the mature miRNA. Gray: miRNAs with either no candidates within 250kb of the mature miRNA or where all candidates had a score less then zero (**see Fig. S4b, right**). Yellow: miRNAs for which the closest candidate start site was selected solely on the basis of its proximity. **b**, The basis of miRNA promoter identification, including Gene or EST evidence (green), distance of <5 kilobases to mature miRNA (orange), nearest possible promoter to miRNA (yellow), tended to be conserved between human and mouse

Figure S6. miRNA genes occupied by the core transcriptional regulators in ES cells are expressed in induced Pluripotent Stem (iPS) cells. miRNAs were purified from MEFs (lanes 1-3), mES cells (lane 4,5) and iPS cells (lane6) and hybridized to LNA miRNA arrays. Differentially expressed probes enriched in either mEFs or mES cells are shown (FDR < _10%, see supplemental text, iPS cells were not used to determine differential expression). miRNA probes were Z-score normalized, and cell types were clustered hierarchically (top). Probes associated with active miRNAs occupied at their promoters by Oct4/Sox2/Nanog/Tcf3 are listed to the right.

Figure S7. PcG occupied miRNAs are generally expressed in a tissue specific manner. Mature miRNAs derived from genes occupied by Suz12 and H3K27me3-modified nucleosomes were compared to the list of tissue specific miRNAs derived from the miRNA expression atlas (Landgraf et al., 2007). Vertical axis represents tissue-specificity and miRNAs with specificity score ≥1 are shown. miRNAs bound by Oct4/Sox2/Nanog/Tcf3 and expressed in mES cells are not shown (largely ES cell specific miRNAs). Among the tissue specific miRNAs there is significant enrichment ($p < 0.005$ by hypergeometric distribution) for miRNAs occupied by Suz12 (green).

Figure S1

A


Sox2 (n=3380)
564
658
783
274
Oct4
(n=4320)
Nanog
(n=3114)
942
1443
70
63
631
52
83
49
84
Tcf3 (n=1369)

B

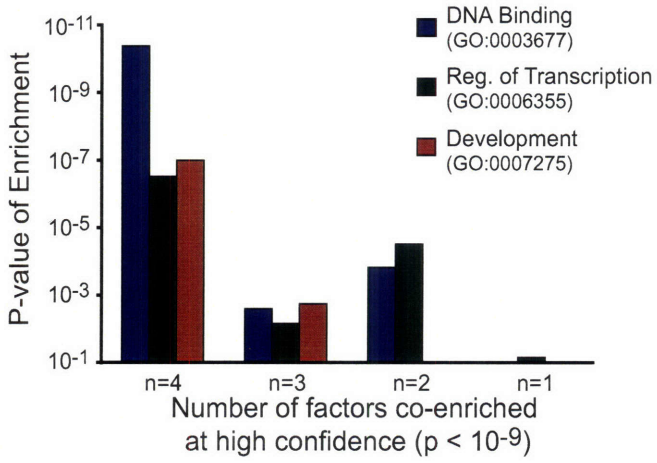

DNA Binding
(GO:0003677)
Reg. of Transcription
(GO:0006355)
Development
(GO:0007275)
P-value of Enrichment
10-11
10-9
10-7
10-5
10-3
10-1
n=4
n=3
n=2
n=1
Number of factors co-enriched
at high confidence (p < 10-9)

Figure S2

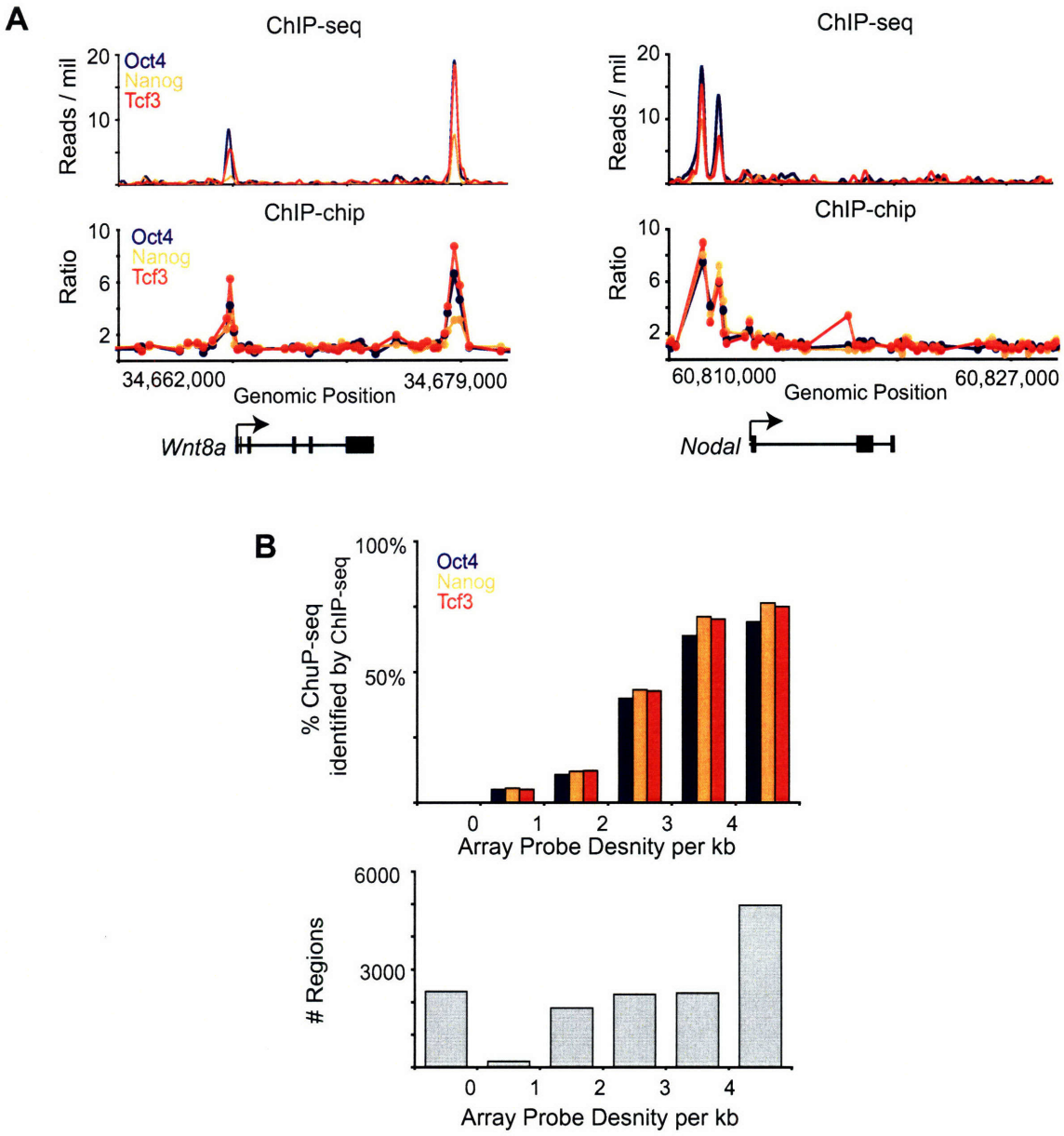

A
ChIP-seq
Reads / mil
20
10
Oct4
Nanog
Tcf3
ChIP-chip
Ratio
10
6
2
34,662,000
34,679,000
Genomic Position
Wnt8a
ChIP-seq
Reads / mil
20
10
ChIP-chip
Ratio
10
6
2
60,810,000
60,827,000
Genomic Position
Nodal
B
% ChuP-seq identified by ChIP-seq
100%
50%
Oct4
Nanog
Tcf3
0
1
2
3
4
Array Probe Desnity per kb
Regions
6000
3000
0
1
2
3
4
Array Probe Desnity per kb

Figure S3

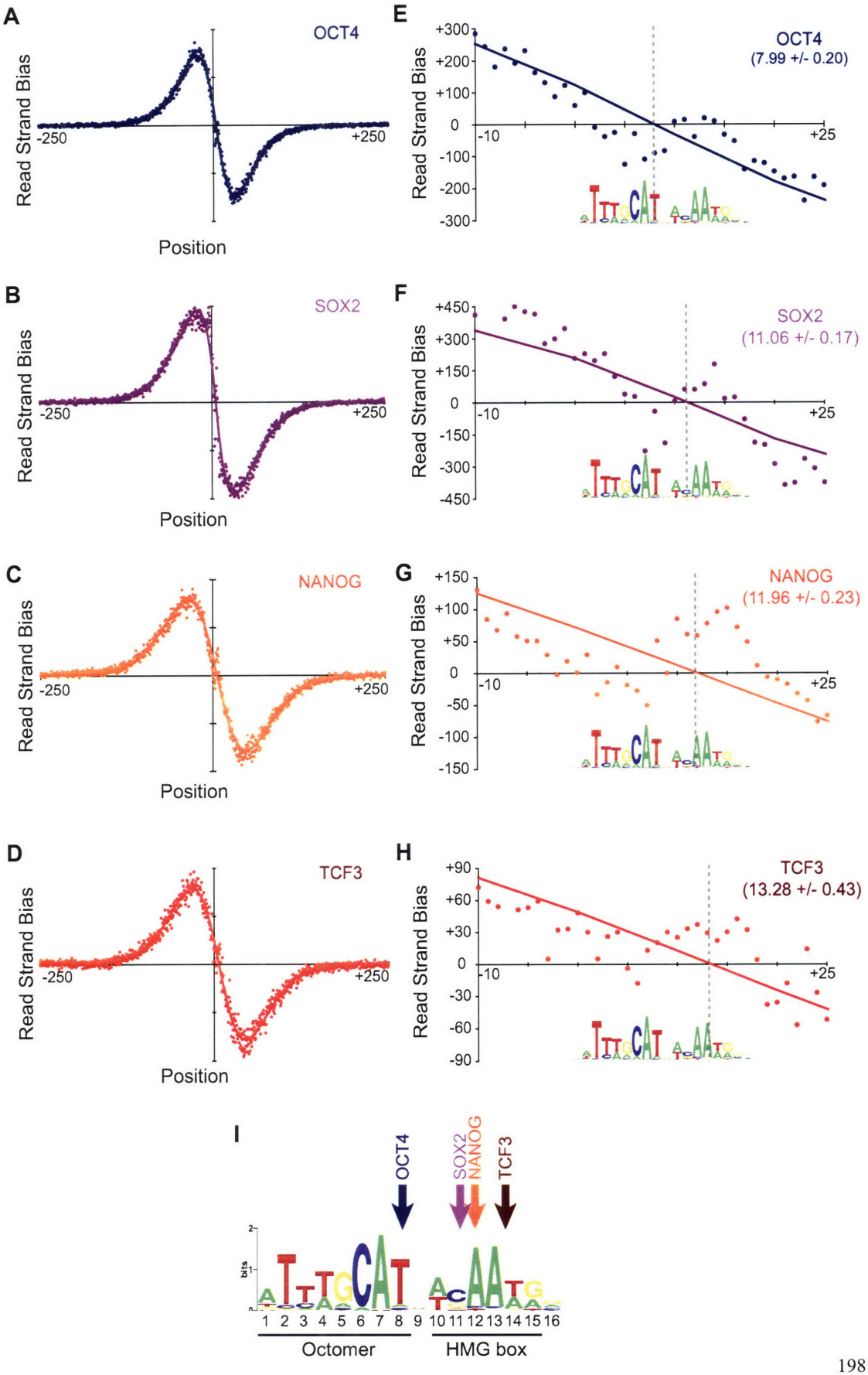

A
OCT4
Read Strand Bias
-250
+250
Position
E
OCT4
(7.99 +/- 0.20)
+300
+200
+100
0
-100
-200
-300
-10
+25
B
SOX2
F
SOX2
(11.06 +/- 0.17)
+450
+300
+150
-150
-300
-450
C
NANOG
G
NANOG
(11.96 +/- 0.23)
+150
+100
+50
-50
-100
-150
D
TCF3
H
TCF3
(13.28 +/- 0.43)
+90
+60
+30
-30
-60
-90
I
OCT4
SOX2
NANOG
TCF3
bits
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Octomer
HMG box

Figure S4

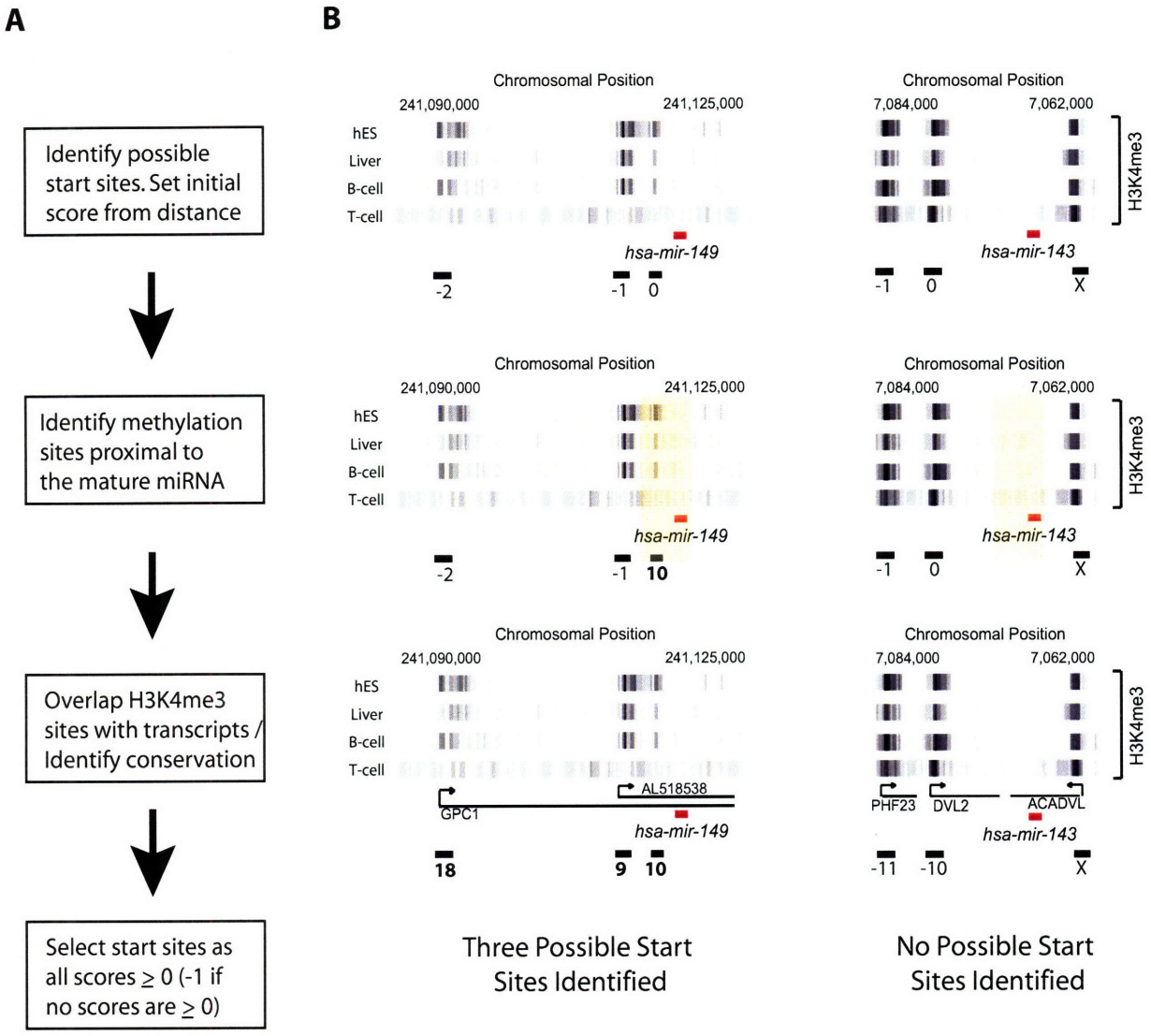

A
B
Identify possible start sites. Set initial score from distance
Identify methylation sites proximal to the mature miRNA
Overlap H3K4me3 sites with transcripts / Identify conservation
Select start sites as all scores ≥ 0 (-1 if no scores are ≥ 0)
Chromosomal Position
241,090,000
241,125,000
hES
Liver
B-cell
T-cell
hsa-mir-149
-2
-1
0
10
18
9
GPC1
AL518538
H3K4me3
7,084,000
7,062,000
hsa-mir-143
X
PHF23
DVL2
ACADVL
-11
-10
Three Possible Start Sites Identified
No Possible Start Sites Identified

Figure S5

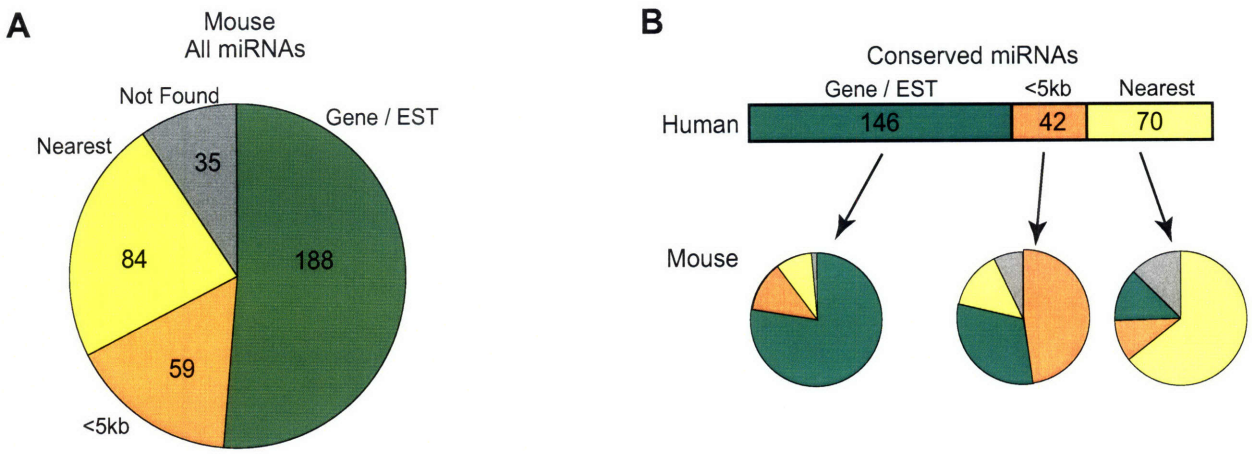
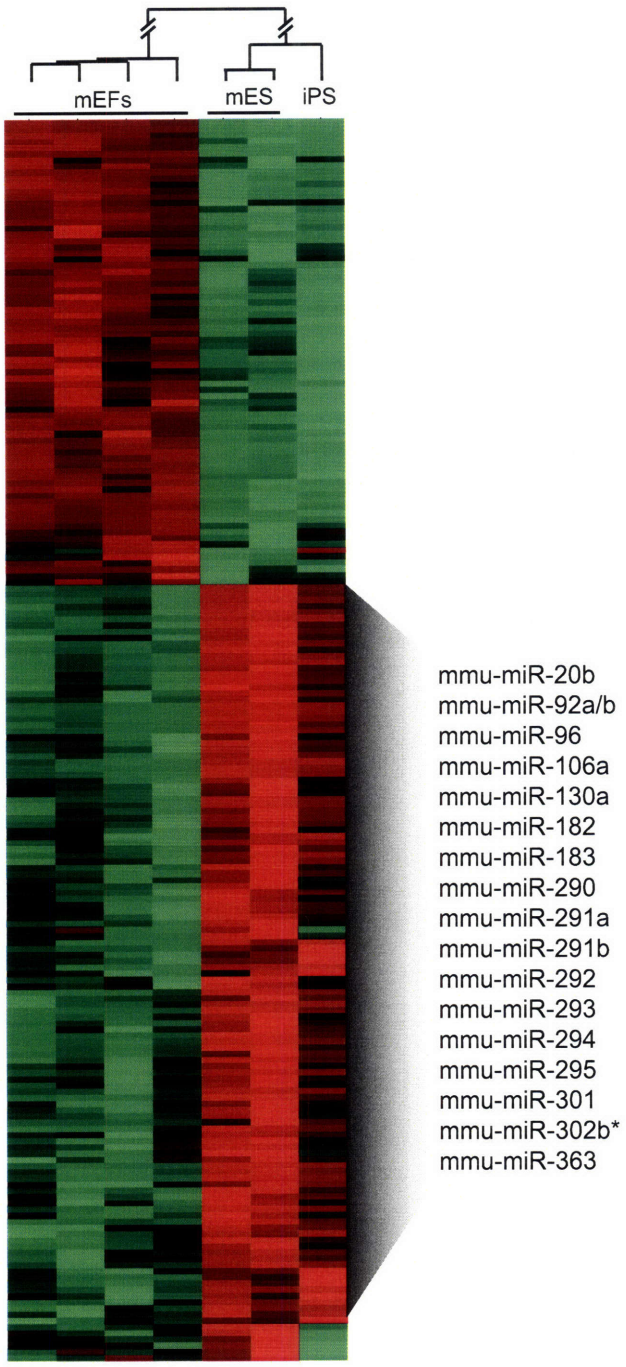
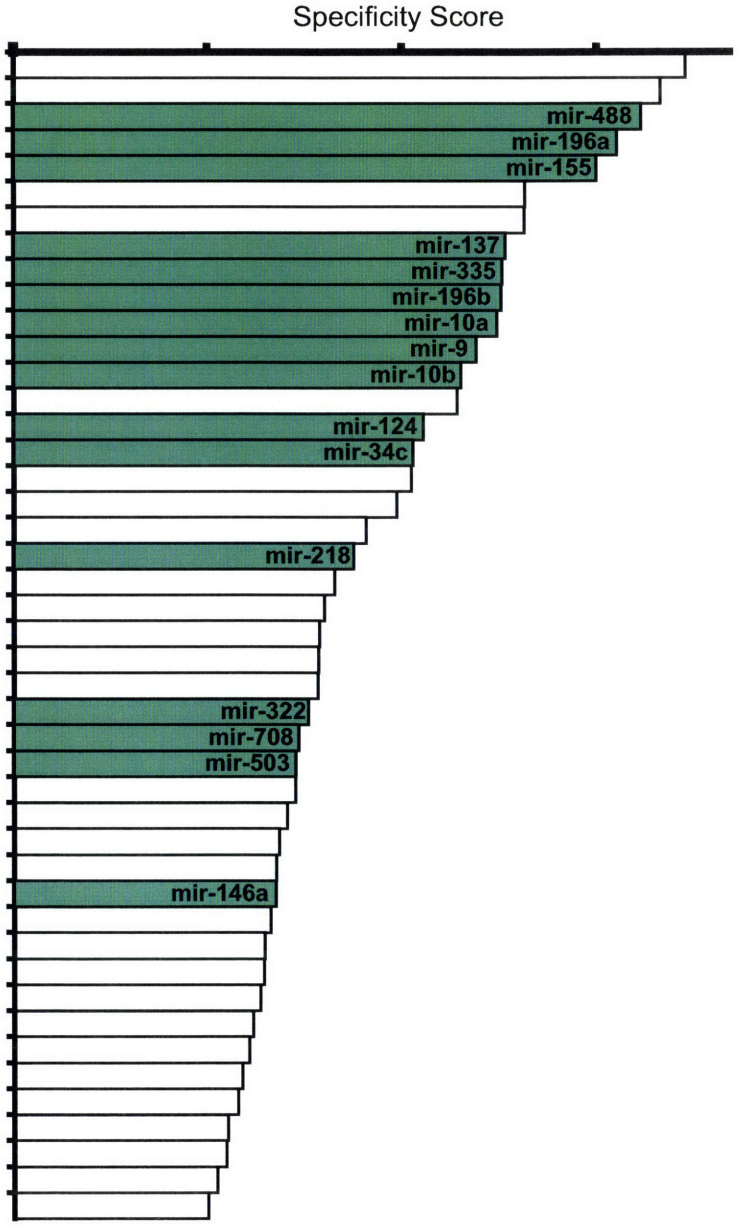

A
Mouse
All miRNAs
Not Found
Gene / EST
Nearest
35
84
188
59
<5kb
B
Conserved miRNAs
Gene / EST
<5kb
Nearest
Human
146
42
70
Mouse

Figure S6


mEFs
mES
iPS
mmu-miR-20b
mmu-miR-92a/b
mmu-miR-96
mmu-miR-106a
mmu-miR-130a
mmu-miR-182
mmu-miR-183
mmu-miR-290
mmu-miR-291a
mmu-miR-291b
mmu-miR-292
mmu-miR-293
mmu-miR-294
mmu-miR-295
mmu-miR-301
mmu-miR-302b*
mmu-miR-363


Specificity Score
mir-488
mir-196a
mir-155
mir-137
mir-335
mir-196b
mir-10a
mir-9
mir-10b
mir-124
mir-34c
mir-218
mir-322
mir-708
mir-503
mir-146a

References

Bailey, T. L., and Elkan, C. (1995). The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol *3*, 21-29.

Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res *34*, W369-373.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Beissbarth, T., and Speed, T. P. (2004). GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics *20*, 1464-1465.

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

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Chang, T. C., Wentzel, E. A., Kent, O. A., Ramachandran, K., Mullendore, M., Lee, K. H., Feldmann, G., Yamakuchi, M., Ferlito, M., Lowenstein, C. J., *et al.* (2007). Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. Mol Cell *26*, 745-752.

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

Corney, D. C., Flesken-Nikitin, A., Godwin, A. K., Wang, W., and Nikitin, A. Y. (2007). MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth. Cancer Res *67*, 8433-8438.

Fukao, T., Fukuda, Y., Kiga, K., Sharif, J., Hino, K., Enomoto, Y., Kawamura, A., Nakamura, K., Takeuchi, T., and Tanabe, M. (2007). An evolutionarily conserved mechanism for microRNA-223 expression revealed by microRNA gene profiling. Cell *129*, 617-631.

Gerhard, D. S., Wagner, L., Feingold, E. A., Shenmen, C. M., Grouse, L. H., Schuler, G., Klein, S. L., Old, S., Rasooly, R., Good, P., *et al.* (2004). The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). Genome Res *14*, 2121-2127.

Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91-105.

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

Hinrichs, A. S., Karolchik, D., Baertsch, R., Barber, G. P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T. S., Harte, R. A., Hsu, F., *et al.* (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Res *34*, D590-598.

Hubbard, T., Andrews, D., Caccamo, M., Cameron, G., Chen, Y., Clamp, M., Clarke, L., Coates, G., Cox, T., Cunningham, F., *et al.* (2005). Ensembl 2005. Nucleic Acids Res *33*, D447-453.

Johnson, D., Martazavai, A., Myers, R., Wold, B., (2007). Genome-wide mapping of inn vivo protein-DNA interactions. Science *316*, 1441-2.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res *12*, 996-1006.
Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853-858.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A. O., Landthaler, M., *et al.* (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. Cell *129*, 1401-1414.

Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858-862.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., *et al.* (2006a). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

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

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V., and Mendell, J. T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. Nature *435*, 839-843.

Okabe, S., Forsberg-Nilsson, K., Spiro, A. C., Segal, M., and McKay, R. D. (1996). Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. Mech Dev *59*, 89-102.

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res *33*, D501-504.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C. G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. Nat Struct Mol Biol *15*, 259-267.

Tucker, K. L., Wang, Y., Dausman, J., and Jaenisch, R. (1997). A transgenic mouse strain expressing four drug-selectable marker genes. Nucleic Acids Res *25*, 3745-3746.

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

Appendix B: Supplementary Material

Foxp3 occupancy and regulation of key target genes during T-cell stimulation

Supplementary Methods and Discussion

- Cell Generation and Culture Conditions
  - Cell Culture, Stimulation and Analysis of Cytokine Production
  - Generation of Foxp3-expressing $CD4^+$ T cell Hybridoma Clones
  - Transgenic Mice
  - Purification of Primary $CD4^+$ T cells
- Foxp3 Location Analysis
  - Chromatin Immunoprecipitation
  - Antibodies for ChIP
  - Ten Slide Promoter Array
  - Selection of Regions and Design of Subsequences
  - Compiled Probes and Controls
  - Single Slide Proximal Promoter Array
  - Array Scanning and Data Extraction
  - Data Normalization and Analysis
  - Identification of Bound Regions
  - Comparing Bound Regions to Known Genes
  - Control Location Analysis Experiments
  - Site-Specific PCR Analysis
- DNA Motif Analysis
- Gene Expression Analysis
  - Gene Expression Profiling
  - Expression Data Normalization
  - Identification of Differentially Expressed Genes
  - Hierarchical Clustering and Heatmap Display
  - Real Time RT-PCR
- Significance of List Overlap
- Functional Annotation of Gene Lists

Supplementary Notes

Supplemental Figure Legends and Figures S1-S6

Supplementary Tables are available online at:
http://www.nature.com/nature/journal/v445/n7130/suppinfo/nature05478.html

Cell Generation and Culture Conditions

Cell Culture, Stimulation and Analysis of Cytokine Production

$CD4^+$ 5B6-2 hybridoma cells expressing a $PLP_{139-151}$-specific TCR, which was kindly provided by Vijay Kuchroo, were cultured in Dulbecco's modified Eagle medium (Invitrogen). Primary murine $CD4^+$ T cells were cultured in RPMI-1640 medium (Invitrogen). Media were supplemented with 10% FCS, 2 mM Glutamax, 1 mM HEPES, 1 mM sodium pyruvate, 0.1 mM non-essential amino acids, 0.55 mM 2-mercaptoethanol, 100U/ml penicillin/streptomycin and 0.1 mg/ml gentamycin. For gene expression profiling, real-time RT-PCR, and location analysis, cells were cultured in the absence or presence of 50 ng/ml phorbol 12-myristate 13-acetate (PMA) and 200 ng/ml ionomycin at 37°C and harvested after 6h. Where indicated, cells were preincubated for 1h with 2 μM cyclosporin A. For the analysis of cytokine production of 5B6-2 hybridoma cells, 10 mg/ml brefeldin A was added for the last 4h of 6 and 36h cultures. Cells were harvested at various time points and intracellular cytokine staining was performed using the Cytofix/Cytoperm kit (Becton Dickinson) according to manufacturer's recommendations and phycoerythrin-conjugated anti-mouse IL-2 (JES6–5H4) and TNF-a (MP6-XT22) antibodies.

Generation of Foxp3-expressing $CD4^+$ T cell Hybridoma Clones

The murine full-length Foxp3 was amplified from cDNA of purified BALB/c $CD4^+CD25^+$ spleen cells using 5'-ATGCCCAACCCTAGGCCAGCCAA-3' as the sense and 5'-TCAAGGGCAG GGATTGGAGCAC-3' as the antisense primers. A

minimal Kozak consensus sequence (double-underlined) just upstream of the initiator codon and a FLAG-tag (single-underlined) and were added using 5'-GAATTCACCATGATGGACTACAAGGACGACGACGACAAGCCCAACCCTAGG CCAGCCAA-3' as the sense and 5'-GGATC CTCAAGGGCAGGGATTGGAGCAC-3' as the antisense primers. For further cloning sense and antisense primer sequences contained 5' EcoRI and BamHI restriction sites, respectively. Lentiviral vectors encoding the N-terminal FLAG-tagged Foxp3-IRES-GFP or Empty-IRES-GFP control were generated from the pLenti6/V5-D-TOPO vector (Invitrogen). The integrity of cloned cDNA was confirmed by sequencing and sequence comparison to GenBank accession no. NM_054039. Concentrated culture supernatants of 293FT transfected with lentiviral vectors were used to infect 5B6-2 hybridoma cells. Stably transduced 5B6-2-[Foxp3]-IRES-GFP or 5B6-2-[Empty]-IRES-GFP 5B6-2 hybridoma cell clones were established after flow cytometric single-cell sorting.

Transgenic Mice

TCR-hemagglutinin (TCR-HA) BALB/c mice express a transgenic TCR specific for the H2-IE^d $HA_{107-119}$ peptide. Double-transgenic TCR-HA x pgk-HA mice additionally express the HA protein under the control of the phosphoglycerate kinase (pgk) promoter and are characterized by high frequencies of TCR-HA-expressing T_{reg} cells (Klein et al., 2003). Mice were bred in the Dana–Farber Cancer Institute animal facility under specific pathogen-free conditions. Animal

care and all procedures were in accordance with the guidelines of the Animal Care and Use Committee of the Dana-Farber Cancer Institute.

Purification and FACS analysis of primary $CD4^+$ T cells

Single cell suspensions of pooled spleen and lymph node cells were prepared from TCR-HA and TCR-HA x pgk-HA mice for the purification of naïve and T_{reg} cells, respectively. Cells were stained with fluorochrome-conjugated anti-mouse mAbs CD4 (RM4-5), CD25 (PC61) and TCR-HA (6.5). $HA_{107-119}$–specific $CD4^+CD25^-$ naïve T cells and Foxp3-expressing $CD4^+CD25^{high}$ T_{reg} cells were then purified using a FACSAria cell sorter and FACSDiva software (Becton Dickinson). Cells were ≥98% pure upon re-analysis. Intracellular staining with the anti-mouse/rat mAb FJK-16s (eBioscience) revealed that ~95% of purified $TCR\text{-}HA^+CD4^+CD25^{high}$ cells express Foxp3. $CD4^+CD25^-$ naïve T cells showed negligible staining for Foxp3 (≤0.5%). For the analysis of Ly6a surface expression on freshly isolated or *in vitro* stimulated antigen-specific T_{eff} and T_{reg} cells, primary T cells were purified as described above and additionally stained using a phycoerythrin-conjugated anti-mouse mAb Ly6a/e (E13-161.7).

Foxp3 Location Analysis

Chromatin Immunoprecipitation

Protocols describing ChIP methods are available from: http://jura.wi.mit.edu/young_public/hESregulation/ChIP.html

Briefly, for each location analysis reaction ~10^8 N-terminal FLAG-tagged Foxp3-IRES-GFP or Empty-IRES-GFP 5B6-2 hybridomas were chemically crosslinked by the addition of 11% formaldehyde solution for 20 min at room temperature. Cells were washed twice with 1 x PBS and pellets were stored at -80ºC prior to use. Cells were resuspended, lysed, and sonicated to solubilize and shear crosslinked DNA. We used a Misonix Sonicator 3000 and sonicated at power 7 for 10 to 18, 20 second pulses (60 second pause between pulses) at 4°C while samples were immersed in an ice bath. The resulting whole cell extract was incubated overnight at 4°C with 100 ul of Dynal Protein G magnetic beads preincubated with 10 µg of the appropriate antibody for at least 4 hrs. Beads were then washed 4 times with RIPA buffer and 1 time with TBS. Bound complexes were eluted from the beads in elution buffer by heating at 65°C with occasional vortexing, and crosslinking was reversed by ~6 hour incubation at 65°C.Whole cell extract DNA (reserved from the sonication step) was also treated for crosslink reversal. Immunoprecipitated DNA and whole cell extract DNA were then purified by treatment with RNaseA, proteinase K and multiple phenol:chloroform:isoamyl alcohol extractions. Purified DNA was blunted, ligated to a universal linker and amplified using a two-stage PCR protocol. Amplified DNA was labeled using Invitrogen Bioprime random primer labelling kits (immunoenriched DNA was labeled with Cy5 fluorophore, whole cell extract DNA was labeled with Cy3 fluorophore). Labeled and purified DNA was combined (4 – 5 µg each of immunoenriched and whole cell extract DNA) and hybridized to

arrays in Agilent hybridization chambers for 40 hours at 40°C. Arrays were then washed and scanned.

Antibodies for ChIP

For ChIP experiments, we used Anti-FLAG (Sigma M2) and anti-E2F4 (Santa Cruz 1082) antibodies. E2F4 antibody has been shown to specifically recognize previously reported E2F4 target genes (Ren et al., 2002; Weinmann et al., 2002). Anti-FLAG antibody has also been demonstrated to work for chromatin immunoprecipitation (Henry et al., 2003).

Ten Slide Promoter Array

This study employed a 10-slide mouse promoter array set that has been used in previously published work (Boyer et al., 2006). These arrays were designed to contain oligonucleotides that cover approximately 10 kb around the transcription start sites of approximately 16,000 of the best annotated mouse transcription start sites. Arrays were manufactured by Agilent Technologies (www.agilent.com).

Selection of Regions and Design of Subsequences

To define transcription start sites, we first selected transcripts from three of the most commonly used databases for sequence information (Refseq, Ensembl, MGC). Transcription start sites within 500 bp of each other were considered single start sites. To restrict our array to the most likely transcription start sites,

we selected only those that were found in at least two of the three databases. We also included microRNAs from the RFAM database.

25 kb of sequence around each transcription start site (20 kb upstream to 5 kb downstream) was initially extracted for analysis from the repeat-masked sequence derived from the May 2004 build of the mouse genome. Because we were balancing feature number, exact number of transcription start sites, tiling density and extent of upstream genomic coverage for our array design, we chose to design oligos across a much larger region than we were likely to fit on the array. Each transcription start site was considered independent, even if the 25 kb region overlapped with the 25 kb region of another transcription start site. While we anticipated not being able to use all of these oligos, this allowed us flexibility in later design steps to add oligos for additional upstream genomic coverage if space became available. The subset of probes from –8 kb to +2 kb was selected for the actual array. We used the program ArrayOligoSelector 10 (AOS; http://arrayoligosel.sourceforge.net/) to score 60-mers for every unmasked subsequence greater than 62 bp across all promoter regions. The scores for each oligo were retained but not put through the built-in AOS selection process. Instead, the collection of scored 60-mers was divided by promoter and sorted by genomic position. Each set of 60-mers was then filtered based on the AOS oligo scoring criteria: GC content, self-binding, complexity and uniqueness. For our most stringent filter, we selected the following ranges for each parameter: GC content between 30 percent and 100 percent, self-binding score less than 100,

complexity score less than or equal to 24, uniqueness greater than or equal to –40.

From this subset of 60-mers, we selected oligos designed to cover the promoter region with an estimated density of one probe every 280 bp. At this point, we restricted oligo selection to those oligos found within the region 8 kb upstream to 2 kb downstream of the transcription start site. To achieve more uniform tiling, we instituted a simple method to find probes within a particular distance from each other. Starting at the upstream end of the region, we selected the first qualified probe and then selected the next qualified probe located 150 bp to 280 bp away. If there were multiple eligible probes, we chose the most distal probe within the 280 bp limit. If no probes were identified within this limit, we continued scanning until we found the next acceptable probe. The process was then repeated with the most recently selected probe until we reached the end of the promoter region.

For regions that were not covered by high quality probes, we returned to the full set of scored 60-mers and filtered using less stringent criteria. This gave us an additional set of 60-mers that we then used to fill gaps in our coverage. After this second pass, we identified gaps in our coverage and added oligos that were properly spaced and best fit our criteria regardless of whether they passed the filter cutoffs. This iterative process gave us a compromise between optimal probe quality and optimal probe spacing.

Compiled Probes and Controls

There are 407,355 features split over 10 arrays. The probes are arranged such that array 1 begins with the first selected transcript start site on the left arm of chromosome 1, array 2 picks up where array 1 ends, array 3 picks up where array 2 ends, and so on. Over 16,000 genes are represented on the arrays and each promoter region corresponding to a unique start site is covered by approximately 25-27 probes. A true average distance between probes is difficult to calculate due to the presence of large gaps in the probe tiling. Most of these gaps simply represent the distance between the first and last oligos of two different sets of probes designed against two different genes. Other gaps are caused by lack of available sequence information, repeat masking or sequences that are otherwise highly repetitive and not suitable for oligo design.

Several sets of controls were added. A total of 353 oligos representing Arabidopsis thaliana genomic sequence were included. These Arabidopsis oligos were BLASTed against the mouse genome and do not register any significant hits. These oligos were intended to check background signal. We added a total of 186 oligos representing five proximal promoter regions of genes that are known targets of the transcriptional regulator Oct4 (Pipox, Foxh1, Oct4/Pou5f1, Msh2 and Hoxb1). Each of the four promoters is represented by 21 - 32 different oligos that are evenly positioned across the regions. The oligos for the Hoxb1 region are printed an additional two times. These promoter regions

can be used as positive controls. There are 481 gene desert controls. To identify these probes, we identified intergenic regions of 1 Mb or greater and designed probes in the middle of these regions. These are intended to identify genomic regions that are most likely to be unbound by promoter-binding transcriptional regulators (by virtue of their extreme distance from any known gene). We have used these as normalization controls in situations where a factor binds to a large number of promoter regions. There are 224 features printed as intensity controls; 37 oligos are printed twice and 25 of these 37 are printed an additional six times. Based on a limited number of test hybridizations, this set of oligos gives signal intensities that cover the entire dynamic range of the array. Our intention was that this set could serve as a way to normalize intensities across multiple slides throughout the entire signal range. There are 2,256 controls added by Agilent (standard) and the remainder of each array consists of blank spots.

Single Slide Proximal Promoter Array

Anti-FLAG Foxp3 IPs were compared to control IPs (E2F4, empty vector) on a single slide array with ~95,000 probes (Supplemental Figure S4). Oligo probes were designed essentially as described above. Probes were designed to tile the entire genome with 1 probe placed every ~250bp. A subset of these probes, covering 800bp upstream and 200bp downstream of annotated transcriptional start sites, were then selected to cover the proximal promoters of approximately 18,000 genes.

Array Scanning and Data Extraction

Slides were scanned using an Agilent DNA microarray scanner BA. PMT settings were set manually to normalize bulk signal in the Cy3 and Cy5 channel. For efficient batch processing of scans, we used GenePix 6.0 software (Molecular Devices). Scans were automatically aligned and then manually examined for abnormal features. Intensity data were then extracted in batch. The complete ChIP-chip datasets have been submitted to the online data repository ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) and are associated with accession code E-TABM-154.

Data Normalization and Analysis

GenePix was used to obtain background-subtracted intensity values for each fluorophore for every feature on the array. To obtain set-normalized intensities we first calculated, for each slide, the median intensities in each channel for a set of control probes that are included on each array. We then calculated the average of these median intensities for the set of 10 slides. Intensities were then normalized such that the median intensity of each channel for an individual slide equalled the average of the median intensities of that channel across all slides.

Each slide contains a set of negative control spots that contain 60-mer sequences that do not cross-hybridize to murine genomic DNA. We calculated the median intensity of these negative control spots in each channel and then subtracted this number from the set-normalized intensities of all other features.

To correct for different amounts of genomic and immunoprecipitated DNA hybridized to the chip, negative control-subtracted median intensity value of the IP enriched DNA channel for the set of intensity control probes described above was then divided by the median of the genomic DNA channel for the same set of probes. This yielded a normalization factor that was applied to each intensity in the genomic DNA channel.

Because binding events are rare in the genome, DNA fragments can only be enriched using ChIP, and not anti-enriched. Therefore, the distribution of probes that are below the 1:1 axis (X-score = 0) can provide an empirical non-parametric noise distribution for the experiment. For each X score above 0, the probability of enrichment was calculated to be equal to the number of probes with an X score greater then (more enriched) the test score divided by the total number of probes (enriched + noise) with an absolute value of the X score greater then the test score. This calculation removes the assumption that the X scores on a given array are normally distributed.

Identification of Bound Regions

To automatically determine bound regions in the datasets, we developed an algorithm to incorporate information from neighbouring probes. For each 60-mer, we calculated the average X score of the 60-mer and its two immediate neighbours. If a feature was flagged as abnormal during scanning, we assumed it

gave a neutral contribution to the average X score. Similarly, if an adjacent feature was beyond a reasonable distance from the probe (1000 bp), we assumed it gave a neutral contribution to the average X score. The distance threshold of 1000 bp was determined based on the maximum size of labeled DNA fragments put into the hybridization. Since the maximum fragment size was approximately 550 bp, we reasoned that probes separated by 1000 or more bp would not be able to contribute reliable information about a binding event halfway between them. This set of averaged values gave us a new distribution that was subsequently used to calculate test statistics for each probe set.

As most probes were spaced within the resolution limit of chromatin immunoprecipitation, we next required that multiple probes in the probe set provide evidence of a binding event. Bound probes were required to have a single probe probability of enrichment and a probe set probability of enrichment greater than 0.95 (5% False Discovery Rate) for high stringency binding calls and .90 (10% False Discovery Rate) for low stringency binding calls. Individual probe sets that passed these criteria and were spaced closely together were collapsed into bound regions if the centre probes of the probe sets were within 1000 bp of each other.

Comparing Bound Regions to Known Genes

The location of all bound regions was compared to a composite set of transcripts compiled from three databases: RefSeq (Pruitt et al., 2005), Ensembl (Hubbard et

al., 2005), and UCSC annotated known genes (http://genome.ucsc.edu/cgi-bin/hgTrackUi?g=knownGene) that were associated with Entrez Gene identifiers and miRNAs downloaded from the the RFAM database (http://www.sanger.ac.uk/Software/Rfam/). By this method approximately 16,000 Entrez Genes have a least one probe within -8 kb to +2 kb of their transcription start site. All coordinate information was downloaded from the UCSC Genome Browser (NCBI build 6; March 2005). For genes, we assigned bound regions within the -8 kb to +2 kb of the transcriptional start site, which is consistent with our array design. MicroRNAs allowed a less stringent binding within 10kb upstream or downstream of the 5' end of the miRNA.

Control Location Analysis Experiments

Control location analysis experiments demonstrate the specificity of our antibodies for the appropriate targets. When Foxp3 location analysis was performed in hybridoma cells that were not transduced with FLAG-tagged Foxp3 almost no enrichment of DNA was detected. Location analysis experiments with anti-E2F4 antibodies identified expected E2F4 targets, which are dissimilar from the Foxp3 targets that were discovered. These control experiments were performed using the single slide proximal promoter arrays.

Site-Specific PCR Analysis

We used site-specific PCR to confirm binding of Foxp3 to a panel of Foxp3 target genes identified by ChIP-chip. A subset of the bound probe sets was selected

and primer pairs were designed to amplify a 150-200 bp region around the genomic location of probes that show peak levels of immunoenrichment. PCR was performed on ligation-mediated PCR amplified IP samples (Figure 3c). 10 ng of immunoprecipitated (IP) DNA was used in PCR reactions. For input whole cell extract (WCE) samples, a range of DNA amounts (90, 30 and 10 ng of DNA) was used. The PCR was performed for 23 cycles and products were visualized on an agarose gel stained with SYBR Gold (Amersham) and quantified using ImageQuant (Amersham). Enrichment ratios shown in Figure 3c were calculated as the ratio of the intensity of the PCR product from 10ng IP DNA to the product from 10ng WCE DNA. Ratios were normalized relative to the ratio observed for the unenriched *ß actin* control. For site-specific PCR in unamplified IP material from cyclosporin A treated cells (Figure S3), ChIP protocol was followed through the proteinase K digestion. For site-specific PCR, 2ul of IP material was used. Again, for whole cell extract (WCE) samples a range of DNA amounts (90, 30 and 10 ng of DNA) was used. 28 cycles of PCR were performed and products were visualized on agarose gel with ethidium bromide. For all site-specific ChIP analysis, a region of the *ß actin* promoter where no Foxp3 binding is expected is used as a negative control.

GENE	5' OLIGO	3' OLIGO	SIZE
Itk	GCTGTTCTTCCAGGAGGATG	AGGCTGGCTGATGCTGATAG	190 bp
Jak2	ACGGCAGGACTAATTGTTGC	GAAAGGGGGAGAAAGAGACG	180 bp
Zap70	TCTAGGACAGGAACACATTGG	GTGTCGGGAACACAAGAGGA	158 bp
Ptpn22	TTCTGCCTTTCTTCTGGGAAT	CTAGCGCCTTCCTTTCTCAA	162 bp
Il2	GTCCTCATGGGCTCAACATC	GGGAGGCCAACCTTTGTAAT	156 bp
Pou2af1	TTCATGAGACGGAAACCACA	CACATCTACAGGAGGGAACCA	156 bp
Ly6a	CCCAGCACAGTGGTAAGAGG	GGCAGGGTTTATCACTTGGA	182 bp

Tnfrsf9	TGTGTGTGTGAAGAGGGGTTT	TCCACAGACGTGACAAGGAG	151 bp
CD25	GGGTGAAAAGACAGCTTGGT	GGGTGTGGGATTCACAAATG	151 bp
ß Actin	AGGGTACCACCGGAAAAGTC	CCCCAAAGGCTGAGAAGTTA	150 bp

DNA Motif Analysis

Discovery of the Foxp3 sequence motif from the ChIP-chip binding data was performed using the THEME algorithm (MacIsaac et al., 2006). THEME tests specific, biologically informed hypotheses about a transcription factor's binding specificity and identifies a motif consistent with both the binding data and prior knowledge regarding the protein's DNA-binding domain structural family.

We extracted genomic sequence corresponding to the regions bound by Foxp3 at high confidence in stimulated cells for use as the "foreground" data set in THEME. We then extracted sequence regions at random from unbound regions on the array for use as the "background" data set. The length of these unbound sequences was matched to the average size of the bound set (700bp) to avoid biasing either set towards motif presence or absence. We then ran THEME, testing hypotheses consistent with the Forkhead DNA-binding domain family, and identified the motif that yielded the lowest bound vs. unbound classification error after 5-fold cross-validation. The statistical significance of the best motif's cross-validation error was assessed by running THEME on randomized data (using the starting hypothesis from which the best motif was derived) and calculating a Z-score for the observed error under the null hypothesis that the sequences were selected at random from the background set. We then tested hypotheses corresponding to 35 other distinct DNA-binding domains in an identical fashion,

and found that the best motif identified from the Forkhead family yielded a lower cross-validation error than any motif from all other families tested.

The Foxp3 motif learned by THEME, and the Nfat motif from the TRANSFAC (version 8.3) database (Wingender et al., 1996) were used to scan all arrayed sequences to identify matches to the motif. Each potential site was assessed by summing the position-specific scores from the motif log-odds matrix. Sites were identified as matches if their score was greater than or equal to the threshold, determined by the THEME algorithm, which classified bound and unbound sequences with the lowest error during motif discovery. For the Nfat motif from TRANSFAC a threshold of 60% of the maximum possible score was used. Foxp3 motif conservation in bound regions was determined using mm6/hg17 mouse-human pairwise alignments obtained from the UCSC Genome Browser (Karolchik et al., 2003). If the human sequence directly aligned to the motif match in mouse also met the score threshold we identified that site as a conserved match. We performed the same conservation calculations for randomly selected unbound microarray sequence regions. Statistical significance was determined by fitting a hypergeometric distribution to the data and testing the null hypothesis that the number of conserved Foxp3 motifs observed in bound regions arose from random selection without replacement from the background population. We then examined the distribution of spacings between Foxp3 and Nfat motifs (when they occurred together in the same Foxp3 bound region). Statistical significance was determined using the hypergeometric test

described above. Results are summarized in Supplemental Table S5. The online tool WebLogo (http://weblogo.berkley.edu) was used to generate DNA motif sequence logo in Figure 2b.

Gene Expression Analysis

Gene Expression Profiling

For each hybridoma culture condition, total RNA was prepared from 1 x 10^7 cells using Trizol (Gibco) followed by additional purification using the RNeasy Mini Kit (Qiagen). Biotinylated antisense cRNA was then prepared according to the Affymetrix standard labelling protocol (one amplification round). For each primary T cell culture condition, total RNA was isolated from 5 x 10^5 cells with RNeasy. Biotinylated antisense cRNA was prepared by two rounds of *in vitro* amplification using the BioArray RNA Amplification and Labeling System (Enzo Life Sciences) according to the protocol for 10-1000 ng of input RNA provided by the manufacturer. Biotinylated cRNAs of hybridomas and primary T cells were fragmented and hybridized to Affymetrix GeneChip Mouse Expression Set 430 2.0 arrays at the Microarray Core Facility (Dana-Farber Cancer Institute). Arrays were stained, scanned, and quantified according to standard Affymetrix protocols. Data were annotated according the NetAffx database (http://www.affymetrix.com/analysis/index.affx) as of March, 2006. The complete expression datasets have been submitted to the online data repository ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) and are associated with accession code E-TABM-154.

Expression Data Normalization

Quantile normalization was performed separately on the hybridoma and *ex vivo* expression datasets. Expression data were ranked within each sample. Each probeset was given the value of the average signal of the probesets of that rank, across the dataset.

Identification of Differentially Expressed Genes

The hybridoma dataset was analyzed for statistically significant differential expression using the online NIA Array Analysis Tool (Sharov et al., 2005): (http://lgsun.grc.nia.nih.gov/ANOVA/).

Probesets were tested for differential expression using the following settings:

Threshold z-value to remove outliers: 10000

Error Model: Max(Average,Bayesian)

Error variance averaging window: 200

Proportion of highest error variances to be removed: 0.01

Bayesian degrees of freedom: 5

FDR threshold: .05

Of 45,101 probesets on the Affymetrix Mouse 430 2.0 array, 256 were differentially expressed between Foxp3- and Foxp3+ stimulated hybridomas, while 23 probesets were differentially expressed in the unstimulated cells. The expression data for these probesets are provided (Supplemental Tables S7 and

S8). Probesets were excluded that had an average signal intensity that was not in the upper tercile on the arrays (54.4 units). Probesets that did not map to the genes for which we had Foxp3 binding data and probesets that mapped to multiple genes were also excluded. In the cases where multiple probesets mapping to one gene were differentially expressed, only the probeset showing the largest differential expression was displayed.

It is worth noting that a stringent cutoff to call differential expression is used. This identifies expression changes with high confidence but produces an underestimate because there are many genes that show small changes. Our goal is to gain insight into Foxp3 action, and by focusing on only the most pronounced transcriptional effects, we aim to minimize the effects of noise in the expression data. As a result of this approach, some genes that are likely to be regulated directly by Foxp3 exhibit small transcriptional effects that are not called differentially expressed. For example the genes *Bcl10*, *Cd53*, *Rbpsuh*, and *Rgs1* are all direct Foxp3 binding targets that are known to play a role in regulation of T cells, and have the characteristic expression pattern of suppressed activation, but do not meet the FDR < 5% statistical significance cutoff for differential expression.

Genes with consistent Foxp3 dependent differential expression in between stimulated *ex vivo* T helper and T_{reg} cells and between stimulated Foxp3- and Foxp3+ hybridoma cells were determined according to the following method.

Probesets were excluded that had an average signal intensity that was smaller than the median signal on the arrays (26.2 units). Probesets that did not map to the genes for which we had Foxp3 binding data and probesets that mapped to multiple genes were also excluded. In the cases where multiple probesets mapping to one gene were differentially expressed, only the probeset showing the largest differential expression was displayed.

A score for Foxp3 dependent differential expression was calculated in the hybridoma and *ex vivo* datasets separately. The product of these scores was used to sort the genes and identify those with the largest Foxp3 dependent differential expression, which was consistent in the two cell types. To calculate each score, the average signal intensity within the two groups being compared was calculated. The difference in signal between the groups being compared was divided by the median signal intensity of all probesets on the array (26.2 units) plus one eighth of the average signal intensity for that probeset.

$(A - B) / ((A + B) / 16 + median)$

This generated a differential expression score that is comparable to a signal to noise ratio, where noise is estimated be a linear function of signal intensity. In Figure S5, the 125 genes with the highest overall differential expression score were displayed, to match the number of genes that are shown in Figure 3A.

Hierarchical Clustering and Heatmap Display

For clustering and heat map display, expression data were Z-score normalized separately within the hybridoma and *ex vivo* datasets. For heatmap display in Figure S5 data were Z-score normalized within the full hybridoma and *ex vivo* datasets including the unstimulated samples, though only the stimulated samples are displayed. Average linkage, correlation distance, centered, hierarchical clustering was performed using Gene Cluster (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm#ctv). Heatmaps were generated using Java Treeview (http://jtreeview.sourceforge.net/). Cluster branches were flipped about tree nodes for optimal display. In Figure 3B, those genes from panel A that are bound by Foxp3 and are expressed in *ex vivo* cells are displayed in panel B. *Slc17a6* and *Adam10* are excluded because they are not expressed in the *ex vivo* samples.

Real-time RT-PCR

Total RNA was prepared from hybridoma cells or FACS purified primary T cell populations using the RNeasy kit (Qiagen) followed by DNase digestion (Qiagen). cDNA was synthesized from total RNA using Superscript II reverse transcriptase and oligo(dT) (Invitrogen Life Technologies) according to the manufacturer's recommendations. Real-time RT-PCR was performed on an ABI PRISM thermal cycler (Applied Biosystems) using SYBR® Green PCR core reagents (Applied Biosystems). Real-time RT-PCR primer sets were either obtained from SuperArray or are available upon request.

Significance of List Overlap

Statistical significance of overlap between differentially expressed genes and Foxp3 bound genes was calculated using a standard Chi-Square test.

Functional Annotation and Statistical Significance of Gene Lists

Functional annotation and statistical significance of gene lists was performed with the on-line tool, DAVID (http://niaid.abcc.ncifcrf.gov/) (Dennis et al., 2003). Genes were imported as EntrezGene IDs and, using the Functional Annotation tool, compared to KEGG pathways (Kanehisa et al., 2000).

The Foxp3 binding targets in PMA/ionomycin-stimulated $CD4^+$ hybridomas were enriched for genes associated with the following KEGG pathways:

KEGG Pathway	**P-Value**
T CELL RECEPTOR SIGNALING PATHWAY	1.4E-5
CELL CYCLE	4.3E-3
FATTY ACID ELONGATION IN MITOCHONDRIA	2.5E-2
CYTOKINE-CYTOKINE RECEPTOR INTERACTION	5.9E-2
PYRIMIDINE METABOLISM	9.2E-2

The Foxp3 target genes that are downregulated in Foxp3+ stimulated hybridomas relative to their levels in Foxp3- stimulated hybridomas were enriched for genes associated with the following KEGG pathways:

KEGG Pathway	P-Value
T CELL RECEPTOR SIGNALING PATHWAY	6.1E-3
CYTOKINE-CYTOKINE RECEPTOR INTERACTION	3.6E-2

Supplementary Notes

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Chen, C., Rowell, E. A., Thomas, R. M., Hancock, W. W., and Wells, A. D. (2006). Transcriptional regulation by Foxp3 is associated with direct promoter occupancy and modulation of histone acetylation. J Biol Chem *281*, 36828-36834.

Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., and Lempicki, R. A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Henry, K. W., Wyce, A., Lo, W. S., Duggan, L. J., Emre, N. C., Kao, C. F., Pillus, L., Shilatifard, A., Osley, M. A., and Berger, S. L. (2003). Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes Dev *17*, 2648-2663.

Hubbard, T., Andrews, D., Caccamo, M., Cameron, G., Chen, Y., Clamp, M., Clarke, L., Coates, G., Cox, T., Cunningham, F., *et al.* (2005). Ensembl 2005. Nucleic Acids Res *33*, D447-453.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., Roskin, K. M., Schwartz, M., Sugnet, C. W., Thomas, D. J., *et al.* (2003). The UCSC Genome Browser Database. Nucleic Acids Res *31*, 51-54.

Klein, L., Khazaie, K., and von Boehmer, H. (2003). In vivo dynamics of antigen-specific regulatory T cells not predicted from behavior in vitro. Proc Natl Acad Sci U S A *100*, 8886-8891.

Macisaac, K. D., Gordon, D. B., Nekludova, L., Odom, D. T., Schreiber, J., Gifford, D. K., Young, R. A., and Fraenkel, E. (2006). A hypothesis-based approach for identifying the binding specificity of regulatory proteins from chromatin immunoprecipitation data. Bioinformatics *22*, 423-429.

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res *33*, D501-504.

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

Sharov, A. A., Dudekula, D. B., and Ko, M. S. (2005). A web-based tool for principal component and significance analysis of microarray data. Bioinformatics *21*, 2548-2549.

Weinmann, A. S., Yan, P. S., Oberley, M. J., Huang, T. H., and Farnham, P. J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev *16*, 235-244.

Wingender, E., Dietze, P., Karas, H., and Knuppel, R. (1996). TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res *24*, 238-241.

Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A. D., Stroud, J. C., Bates, D. L., Guo, L., Han, A., Ziegler, S. F., *et al.* (2006). FOXP3 controls regulatory T cell function through cooperation with NFAT. Cell *126*, 375-387.

Supplemental Figure Legends

Figure S1. Flow cytometric analysis of Foxp3 expression. 5B6-2-[Foxp3]-IRES-GFP or 5B6-2-[Empty]-IRES-GFP 5B6-2 hybridoma cells (**a**) or FACS-purified *ex vivo* TCR-HA transgenic $CD4^+CD25^-$ naïve or $CD4^+CD25^+$ Treg cells (**b**) were cultured in the absence or presence of PMA/ionomycin. After 6h cells were harvested and intracellular staining was performed using the mAb FJK16s (anti-Foxp3). Histograms show relative levels Foxp3 protein expression, which demonstrates that Foxp3 is expressed at a similar level in the transduced hybridomas as in *ex vivo* T_{reg} cells.

Figure S2. Cytokine production of Foxp3-transduced 5B6-2 hybridomas. 5B6-2-[empty]-IRES-GFP (**a**) or 5B6-2-[Foxp3]-IRES-GFP (**b**) hybridoma cells were cultured in the presence of PMA/ionomycin for the indicated times. Brefeldin A was added for the last 4h of 6h and 36h cultures. After surface staining for CD4 expression and fixation, intracellular staining with the indicated cytokine antibodies or appropriate isotype controls was performed. Numbers in dot plots indicate the frequencies of cells in the respective quadrant. These data show that Foxp3 transduction suppresses production of Il2, but does not suppress production of TNF-a.

Figure S3. The promoters of most Foxp3 target genes are bound by Foxp3 before and after T cell stimulation. (a) The lists of genes whose promoters are

bound by Foxp3 in unstimulated (blue) and stimulated (pink) hybridoma cells are shown in a Venn diagram. The genes occupied by Foxp3 in stimulated T cells (FDR < .05) are represented by the pink circle. The genes occupied by Foxp3 in unstimulated T cells (FDR < .05) are represented by the blue circle. The dotted light blue circle represents the genes occupied by Foxp3 in unstimulated cells if the threshold is relaxed to FDR < .10. Most of the genes bound in stimulated cells are prebound in unstimulated cells. **(b)** Although most Foxp3 bound genes in stimulated cells were also bound in unstimulated cells, the binding profiles were not identical in the two conditions. *Ptpn22* and *Jak2* were representative of target genes where strong immunoenrichment was observed in both conditions. The binding profile for Foxp3 is shown at these promoters with binding in unstimulated cells displayed with a blue line and binding in stimulated cells displayed with a pink line. The profile across these promoter regions indicates that additional Foxp3 binding events are stabilized in response to PMA/ionomycin stimulation. This phenomenon is observed at several Foxp3 targets. In contrast to the strong immunoenrichment observed in both conditions at *Ptpn22* and *Jak2*, recent reports (Wu et al., 2006; Chen et al., 2006) indicate that Foxp3 is stabilized at the promoters of *Il2* and *CD25* in response to T cell stimulation. This finding is confirmed in our Foxp3 binding data as shown here. **(c)** Foxp3 binding at the *Ptpn22* and *Jak2* promoters in unstimulated cells was independently confirmed with site-specific ChIP in cyclosporin A treated cells. Primers flanking the binding peaks indicated with asterisks in **(b)** were used for ChIP PCR reactions shown here. Immunoenrichment at the *Ptpn22* and *Jak2* promoters

was observed. As expected, immunoenrichment was not observed at the *Il2, CD25*, and control *ß actin* promoters.

Figure S4. Control experiments confirm the specificity of Foxp3 ChIP-chip. When log_2 intensity values of IP (Cy5 label) material are plotted against log_2 intensity values of whole cell extract (Cy3 label), considerable enrichment of IP material is oberserved in ChIP experiments in FLAG-tagged Foxp3+ cells both before **(a)** and after **(b)** PMA/ionomycin stimulation. In contrast, very little IP enrichment is observed when the same IPs are performed in Foxp3- hybridomas **(c)**. A positive control IP with an anti-E2F4 antibody identifies expected IP enriched E2F4 targets **(d)**, which are largely distinct from the identified Foxp3 target genes (see Supplemental Table S4).

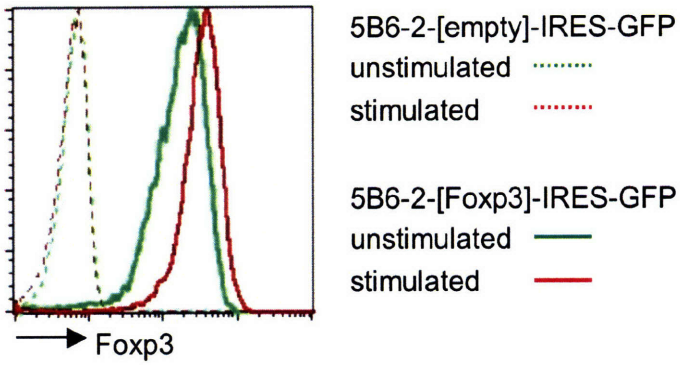
Figure S5. Many Foxp3 targets show consistent Foxp3 dependent differential expression in *ex vivo* and hybridoma cells. Genes were selected that showed consistent Foxp3 dependent differential expression in *ex vivo* and hybridoma cells according to methods described in the Supplementary section, Identification of Differentially Expressed Genes. 125 genes are displayed to match the number of genes in Figure 3a. For clustering and heatmap display data were Z-score normalized within the full hybridoma and *ex vivo* datasets including the unstimulated samples, though only the stimulated samples are displayed. Data were hierarchically clustered and are displayed in a heatmap. The Z-score normalized induction (red) or repression (green) is shown for each

gene. Direct targets of Foxp3 are signified with blue bars, with dark blue representing genes called bound with a false discovery rate of 5% and light blue representing a false discovery threshold of 10%. There is a significant enrichment of direct Foxp3 targets among the genes that are downregulated in stimulated Foxp3+ cells ($p < 10^{-19}$).

Figure S6. Analysis of Ly6a protein expression on primary T_{reg}. Surface expression of Ly6a on $CD4^{+}6.5^{+}$ naïve T cells from TCR-HA mice or Foxp3-expressing $CD4^{+}CD25^{high}6.5^{+}$ T_{reg} cells from double-transgenic TCR-HA x pgk-HA mice was analyzed on freshly FACS-purified cells **(a)** or after 18h of culture in the absence or presence of 50 ng/ml PMA and 200 ng/ml ionomycin with or without a 1h preincubation with 2 μM cyclosporin A **(b)**.

Figure S1

a


5B6-2-[empty]-IRES-GFP
unstimulated
stimulated
5B6-2-[Foxp3]-IRES-GFP
unstimulated
stimulated
Foxp3

b

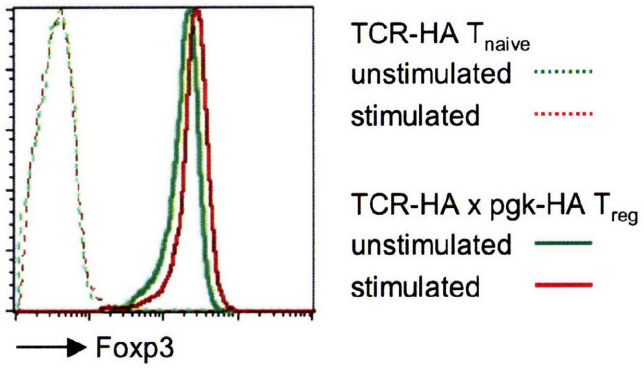
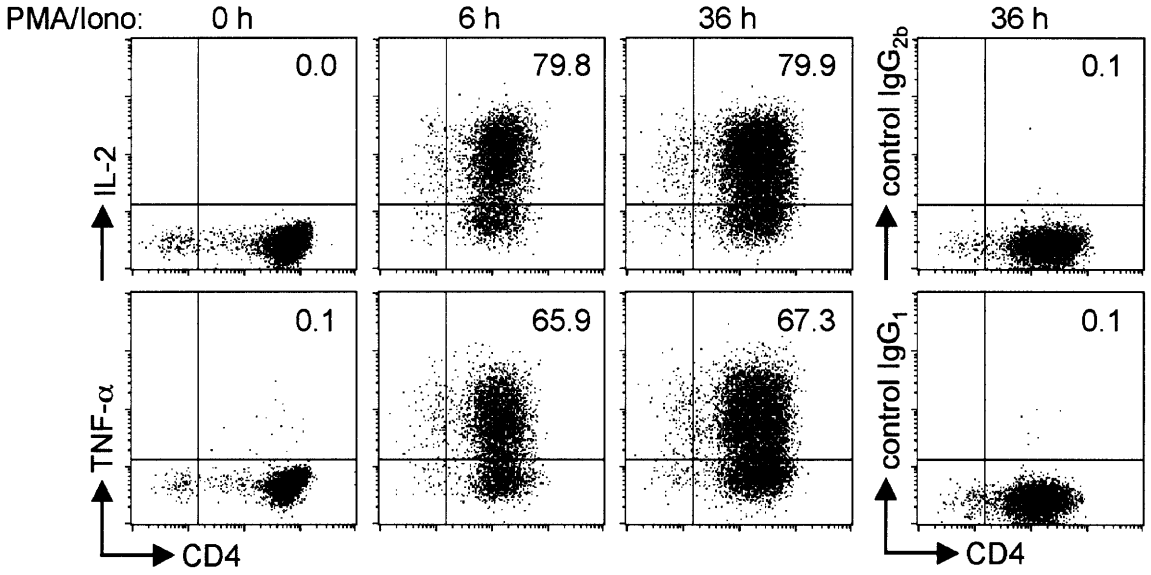

TCR-HA T_{naive}
unstimulated
stimulated
TCR-HA x pgk-HA T_{reg}
unstimulated
stimulated
Foxp3

Figure S2

a


5B6-2-[empty]-IRES-GFP:
PMA/Iono:
0 h
6 h
36 h
36 h
0.0
79.8
79.9
0.1
IL-2
control IgG2b
0.1
65.9
67.3
0.1
TNF-α
control IgG1
CD4
CD4

b

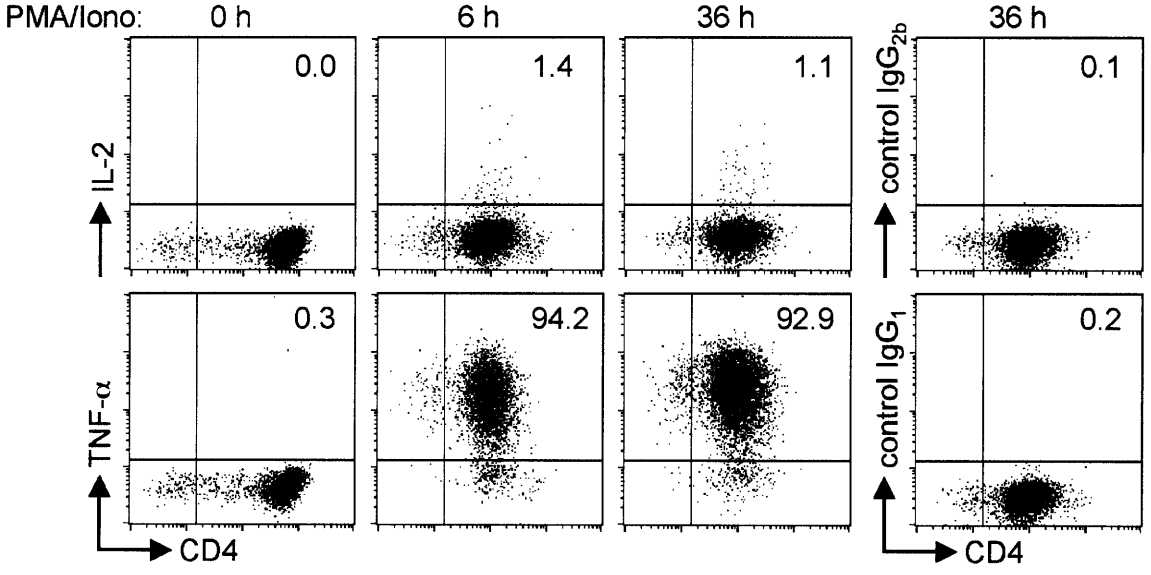

5B6-2-[Foxp3]-IRES-GFP:
PMA/Iono:
0 h
6 h
36 h
36 h
0.0
1.4
1.1
0.1
IL-2
control IgG2b
0.3
94.2
92.9
0.2
TNF-α
control IgG1
CD4
CD4

Figure S3

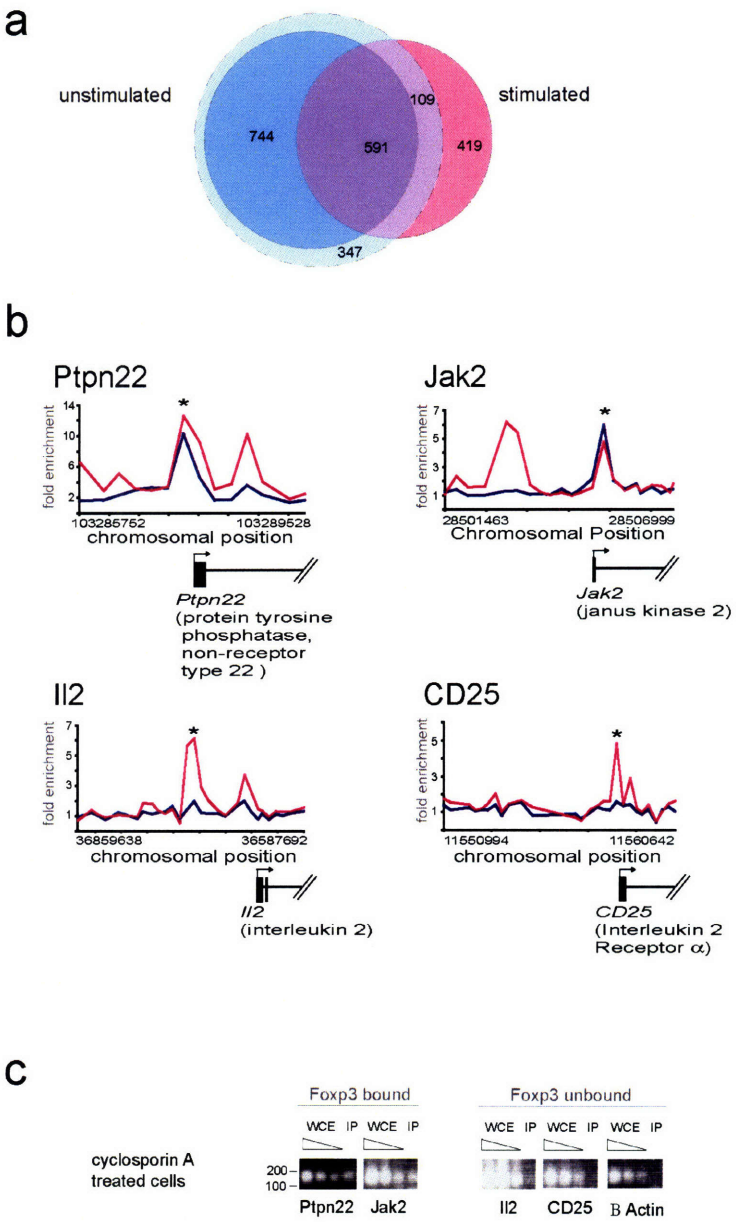

a
unstimulated
stimulated
744
591
109
419
347
b
Ptpn22
fold enrichment
chromosomal position
103285752
103289528
Ptpn22
(protein tyrosine phosphatase, non-receptor type 22)
Jak2
Chromosomal Position
28501463
28506999
Jak2
(janus kinase 2)
Il2
36859638
36587692
Il2
(interleukin 2)
CD25
11550994
11560642
CD25
(Interleukin 2 Receptor α)
c
cyclosporin A treated cells
Foxp3 bound
Foxp3 unbound
WCE IP WCE IP
WCE IP WCE IP WCE IP
200
100
Ptpn22 Jak2
Il2 CD25 B Actin

Figure S4

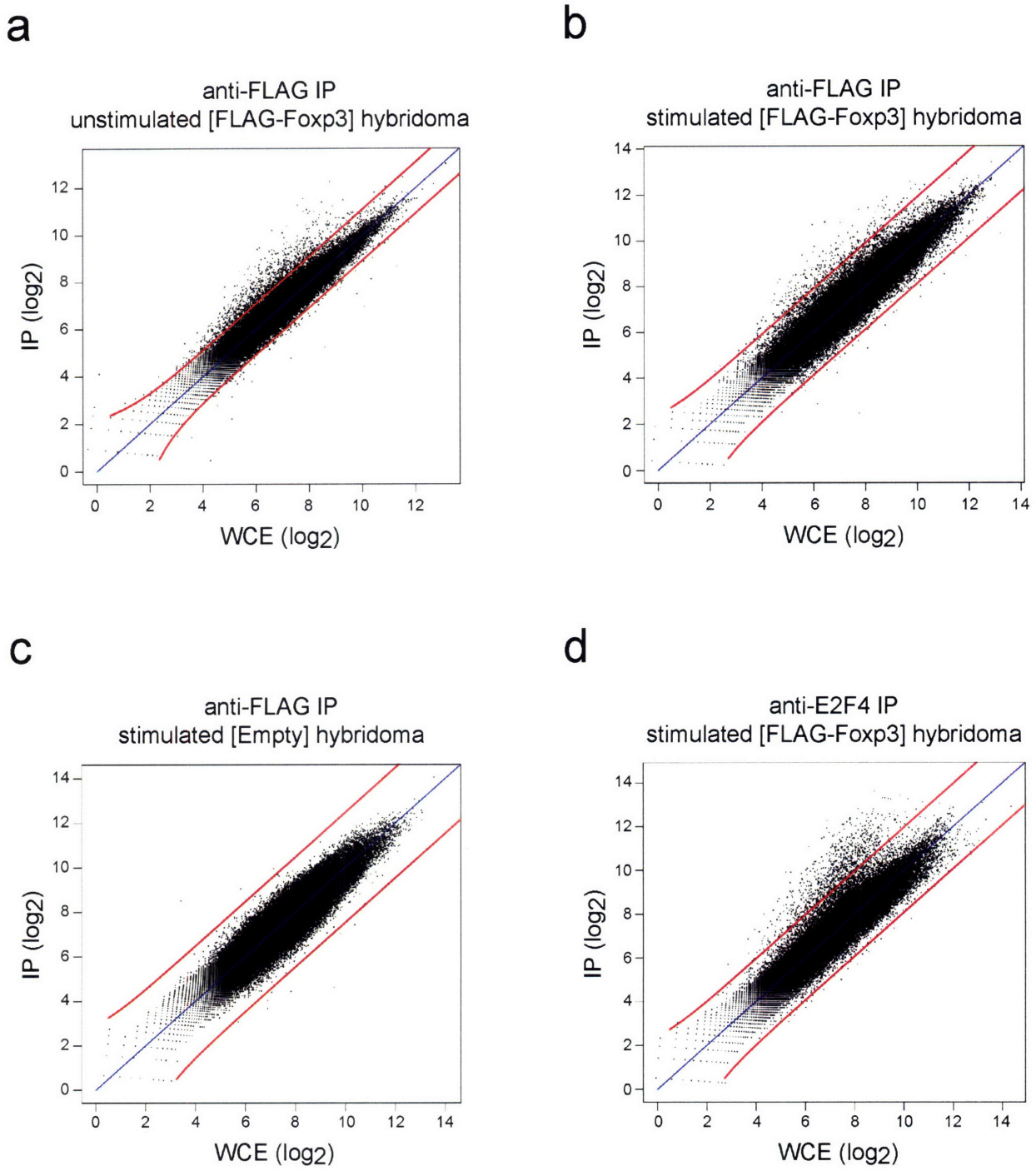

a
anti-FLAG IP
unstimulated [FLAG-Foxp3] hybridoma
IP (log2)
WCE (log2)
b
anti-FLAG IP
stimulated [FLAG-Foxp3] hybridoma
IP (log2)
WCE (log2)
c
anti-FLAG IP
stimulated [Empty] hybridoma
IP (log2)
WCE (log2)
d
anti-E2F4 IP
stimulated [FLAG-Foxp3] hybridoma
IP (log2)
WCE (log2)

Figure S5

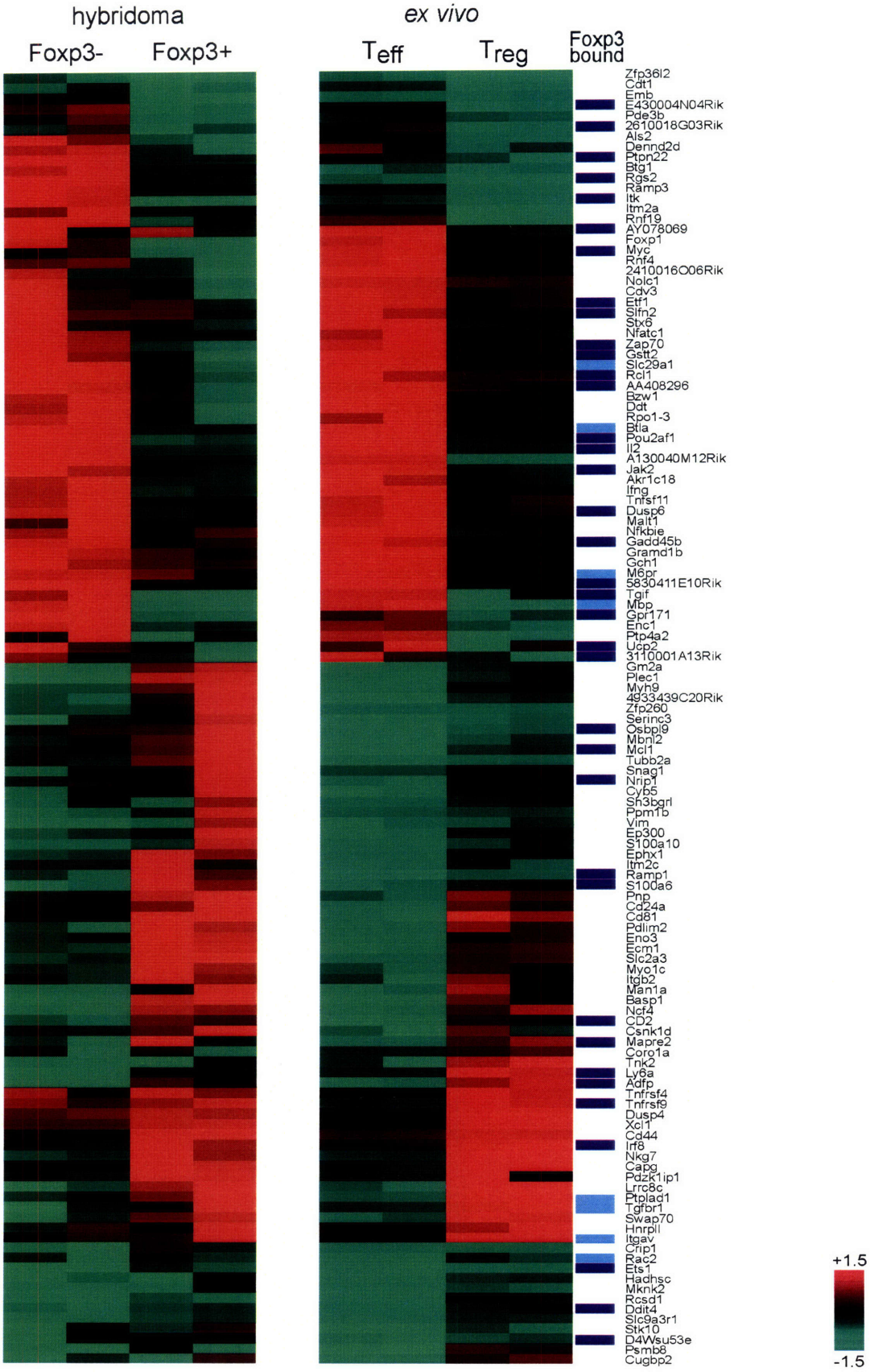
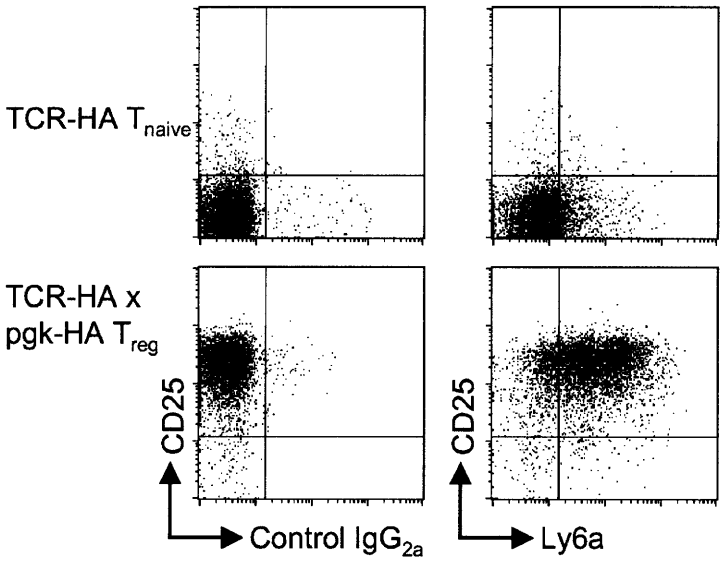
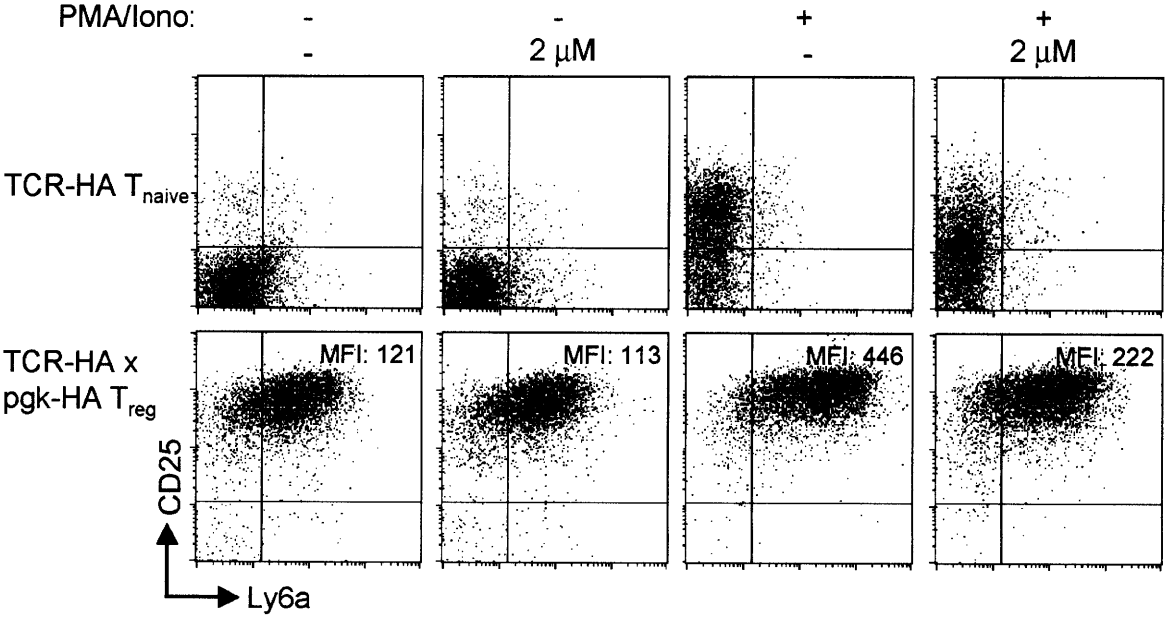

hybridoma
Foxp3-
Foxp3+
ex vivo
Teff
Treg
Foxp3 bound
Zfp36l2
Cdt1
Emb
E430004N04Rik
Pde3b
2610018G03Rik
Als2
Dennd2d
Ptpn22
Btg1
Rgs2
Ramp3
Itk
Itm2a
Rnf19
AY078069
Foxp1
Myc
Rnf4
2410016O06Rik
Nolc1
Cdv3
Etf1
Slfn2
Stx6
Nfatc1
Zap70
Gstt2
Slc29a1
Rcl1
AA408296
Bzw1
Ddt
Rpo1-3
Btla
Pou2af1
Il2
A130040M12Rik
Jak2
Akr1c18
Ifng
Tnfsf11
Dusp6
Malt1
Nfkbie
Gadd45b
Gramd1b
Gch1
M6pr
5830411E10Rik
Tgif
Mbp
Gpr171
Enc1
Ptp4a2
Ucp2
3110001A13Rik
Gm2a
Plec1
Myh9
4933439C20Rik
Zfp260
Serinc3
Osbpl9
Mbnl2
Mcl1
Tubb2a
Snag1
Nrip1
Cyb5
Sh3bgrl
Ppm1b
Vim
Ep300
S100a10
Ephx1
Itm2c
Ramp1
S100a6
Pnp
Cd24a
Cd81
Pdlim2
Eno3
Ecm1
Slc2a3
Myo1c
Itgb2
Man1a
Basp1
Ncf4
CD2
Csnk1d
Mapre2
Coro1a
Tnk2
Ly6a
Adfp
Tnfrsf4
Tnfrsf9
Dusp4
Xcl1
Cd44
Irf8
Nkg7
Capg
Pdzk1ip1
Lrrc8c
Ptplad1
Tgfbr1
Swap70
Hnrpll
Itgav
Crip1
Rac2
Ets1
Hadhsc
Mknk2
Rcsd1
Ddit4
Slc9a3r1
Stk10
D4Wsu53e
Psmb8
Cugbp2
+1.5
-1.5

Figure S6

a


TCR-HA T_{naive}
TCR-HA x
pgk-HA T_{reg}
CD25
Control IgG_{2a}
CD25
Ly6a

b


PMA/Iono:
-
-
-
2 μM
+
-
+
2 μM
TCR-HA T_{naive}
TCR-HA x
pgk-HA T_{reg}
MFI: 121
MFI: 113
MFI: 446
MFI: 222
CD25
Ly6a