Sox2 Co-Occupies Distal Enhancer Elements with Cell-Type-Specific POU Factors to Specify Cell Identity in Embryonic Stem Cells and Neural Precursor Cells.

By
Michael A. Lodato

B.S. Biology
Hofstra University, 2005

Submitted to the Department of Biology in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biology
at the
Massachusetts Institute of Technology

June 2012

Signature of Author: _____
Michael A. Lodato, Department of Biology

Certified by:_____
Rudolf Jaenisch, Department of Biology, Thesis supervisor

Accepted by: _____
Stephen P. Bell, Co-Chair, Department Committee on Graduate Students

Sox2 Co-Occupies Distal Enhancer Elements with Cell-Type-Specific POU Factors to Specify Cell Identity in Embryonic Stem Cells and Neural Precursor Cells

By
Michael A. Lodato

Submitted to the Department of Biology on May 25th, 2012 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology at the Massachusetts Institute of Technology

ABSTRACT

Sox2 is a master regulator of two distinct cellular states, that of pluripotent embryonic stem cells (ESCs) and multipotent neural progenitor cells (NPCs), but what common or distinct roles Sox2 may play in these cell types not fully understood. Further, the molecular mechanisms by which Sox2 can specify two distinct cell identities are as of yet unclear. This thesis is aimed at answering these fundamental questions. In ESCs, Sox2 was associated with a subset of poised regulators of nervous system development, and upon differentiation into NPCs Sox2 selectively activates those which are important for progenitor cell state, while keeping others poised to become activated in later neural development. These data suggested that Sox2 might act as a pioneer factor for neural development throughout embryogenesis. While Sox2 is known to co-occupy target loci in ESCs with the POU factor Oct4, in NPCs Sox2 interacts with the central-nervous-system-expressed POU factors Brn1 and Brn2. By utilizing distinct composite Sox:Octamer motifs in each cell type, Sox2:POU modules control the expression of thousands of genes involved in the development of the neural lineage in a cell-type-specific manner. These data advance our understanding of the mechanism by which transcription factors control cell fate transitions, and indicate that combinatorial interactions between transcription factors may be a pervasive mechanism of transcriptional control in development.

Thesis Supervisor: Rudolf Jaenisch
Title: Professor of Biology

Acknowledgments

First, I would like to thank Rudolf Jaenisch, for allowing me to join his remarkable lab and giving me the freedom to follow my project wherever it led. This project could not have been possible without the help of Laurie Boyer, who contributed to this project at every level, from its conception, to teaching me the ChIP and other techniques, and finally to advising me on how to bring the project to a successful completion. I also have to thank Richard Young, the Chairman of my Thesis Committee. His generosity with the expertise and reagents generated in his lab was integral to the success of this project. Finally, David Page and Martha Bulyk, the two remaining members of my Thesis Committee, put much time and effort into giving me critical feedback on this written thesis which improved it immensely.

I would also like to thank all of the intelligent and creative graduate students and post-doctoral fellows I had the honor of working with during my time in the Jaenisch Lab. It was an interesting time in my life that I won't soon forget. Specifically, Tobias Brambrink, Menno Creyghton, Chris Lengner, Frank Soldner, and Grant Welstead all played major roles in my scientific development and have become good friends. I also have to acknowledge the critical reading of this thesis by Yarden Katz, Frank Soldner, and Grant Welstead, which was very helpful and insightful.

I must also thank my family. I come from a large extended family of loving grandparents, aunts, uncles, and cousins, and picturing them with me during the tough times of this process allowed me to keep pushing through. Most importantly, I have to thank my loving parents, Michael and Nancy Lodato. They are truly great people and I strive to be like them every day. My father taught me many valuable lessons; most importantly that anything worth doing is worth doing the right way. This idea has guided me in both my personal and scientific life. My mother is one of the most loving and caring people I know. She helped me navigate my childhood and adolescence, allowing an at times overly active boy to become the man I am today.

Finally, I have to thank my wife, Inna Lodato. Her constant love and support were instrumental in the success of this thesis. She smiled and encouraged me through the late nights, worked weekends, and missed holidays that accompany a task like this, and I couldn't have done it without her help. She always had faith that I would bring this body of work to a successful completion, and I was truly lucky to have her as a partner during this endeavor. She has also given me the greatest gift of all, my son Michael III and a daughter who will be coming into the world very soon. I optimistically look forward to beginning the next chapter of my life with my wife and children.

Thank you to you all.

Table of Contents

Preface

Sox2 is a master regulator of two distinct cellular states, that of pluripotent embryonic stem cells (ESCs) and that of multipotent neural progenitor cells (NPCs), but what common or distinct roles Sox2 may play in these cell types not fully understood. Further, the molecular mechanisms by which Sox2 can specify two distinct cell identities are as of yet unclear. This thesis is aimed at answering these fundamental questions. In Chapter 1, I will introduce important concepts for understanding this body of work. Next, in Chapter 2 I will outline the functional role of Sox2 in ESCs and NPCs, namely that Sox2 occupied both proximal promoter regions and a large group of distal enhancer regions in both cell types. I also will present data that suggests how Sox2 can function in a context-specific manner in ESCs and NPCs. While Sox2 is known to co-occupy target loci in ESCs with the POU factor Oct4, in NPCs Sox2 interacts with the central-nervous-system-expressed POU factors Brn1 and Brn2. By utilizing distinct composite Sox:Octamer motifs in each cell type, Sox2:POU modules control the expression of thousands of genes in a cell-type-specific manner. In the following Supplemental Chapter, I will present additional data I generated which delves deeper into the detailed genome-wide binding pattern of Sox2 in ESCs and NPCs. Specifically, while Sox2-bound regions in ESCs and NPCs were distinct, they were often linked to enhancers of the same genes. These genes tended to be involved in neural development, and while Sox2 was associated with poised enhancers of these loci in ESCs, upon differentiation into NPCs Sox2 associated with active enhancers of those genes which are important for NPC state, while associating with new poised enhancers of genes with the potential become activated in later neural development. These data suggested that Sox2 might act as a pioneer factor for neural development throughout embryogenesis. These data advance our understanding of the mechanism by which transcription factors control cell fate transitions, and indicate that combinatorial interactions between transcription factors may be a pervasive mechanism of transcriptional control in development.

Chapter I: Introduction

Abstract

Sox2 is a master regulator of several tissue specific stem/progenitor cells, yet how Sox2 controls distinct stem cell identities is not fully understood. This thesis will explore the genome-wide targets of Sox2 in two such populations, pluripotent embryonic stem cells (ESCs) and multipotent neuronal progenitor cells (NPCs). It is important to define the role of master regulatory transcription factors such as Sox2 in controlling cell identity because these factors have been shown to be central to defining the molecular characteristics of a cell. In order to facilitate a full appreciation of the data presented in later chapters, this introductory chapter will first discuss transcription factors and the mechanisms by which they control cell state, then introduce key concepts in ESC and NPC biology. Next, I will review the known roles of Sox2 in early development, specifically in ESCs and the partnership between Sox2 and the POU factor Oct4. I will then discuss the known roles of Sox2 in the developing and adult nervous system. I found that Sox2 co-occupied a large subset of enhancers in NPCs with the central-nervous-system-expressed POU factors Brn1 and Brn2, similarly to its interaction with Oct4 in ESCs. Therefore, I will end this introduction with a review of what is known about the function of these POU factors in the developing nervous system.

Transcription Factors Control Gene Expression Programs

Master Regulators are a special class of transcription factors which control cell identity

Transcription factors bind to DNA in a sequence specific manner and regulate which genes in the genome will be expressed in a given cell type, and thus are thought to sit at the top of a hierarchy controlling cellular identity. Master regulators comprise a special class of transcription factors which are especially potent regulators of cell state. In their associated cell type, master regulators are typically highly expressed and bind thousands of genomic targets involved in cellular identity (Young, 2011). For example, in embryonic stem cells (ESCs), master regulatory transcription factors such as the HMG-box containing factor Sox2 control both the ESC cell identity and set the stage for later developmental transitions (Boyer et al., 2005; Loh et al., 2006, below). Master regulators are typically necessary for the establishment and maintenance of cell state *in vivo*, and the potency of master regulatory transcription factors is perhaps best exemplified by the ability of a master regulator of one cell type to reprogram cellular identity when expressed ectopically in another cell type. For example, the transcription factor MyoD is a master regulator of muscle development, and when MyoD is ectopically expressed in fibroblast cells it can completely reprogram them into myotubes (Lassar et al., 1986). Similarly, ectopic expression of master regulatory transcription factors of other cell types have the ability to induce reprogramming by forcibly establishing their own cell-type-specific gene expression programs (Xie et al., 2004; Takahashi and Yamanaka, 2006; Feng et al., 2008; Seale et al., 2008; Zhou et al., 2008; Kim et al., 2011). Thus, master regulatory transcription factors are strong regulators of cell state.

Molecular mechanisms of transcription factor action

Transcription factors modulate transcriptional activity by recruiting the transcriptional apparatus to specific loci. Generally, a transcription factor is comprised of a DNA-binding domain and one or more transactivation domains, which mediate interactions with other proteins. Mechanistically, a stable interaction between the DNA-binding domain of a transcription factor with a genomic DNA target provides an interface at that locus for interactions between its transactivation domains and other factors,

which then affect transcriptional activity. These other factors include protein complexes which modify chromatin structure and covalently mark histone tails, including histone methyltransferases (and demethylases), acetyltransferases (and deacetylases), and enzymes which catalyze a host of other histone modifications (Bannister and Kouzarides, 2011; Tan et al., 2011; Meurs, 2011). Modified histone tails in turn serve as docking sites for effector proteins containing domains which recognize these specific modifications, such as chromodomains which recognize methylated lysines, and bromodomains which recognize acetylated lysines. Once properly localized, these proteins then act to regulate transcriptional activity (Kouzarides, 2007). Transcription factors also interact with co-activator protein complexes such as the mediator, which interact directly with the basal transcriptional apparatus to recruit the RNA polymerase II transcription initiation complex (Lee and Young, 2000). It has recently become clear that transcriptional initiation is only one step at which transcription factors act. A significant amount of regulation takes place at the transition from initiation to transcriptional elongation, and transcription factors such as c-Myc play a major role in releasing paused polymerase by recruiting release factors such as p-TEFb to promoters (Rahl et al., 2010). Thus, transcription factors play a central role in regulating the activity of promoters by recruiting co-activators and the transcriptional apparatus to these regions.

While a large part of transcriptional regulation occurs at proximal promoters, regions far from transcription start sites, known as distal enhancers, also play a critical role in modulating gene expression levels. In addition to regulating promoters directly, transcription factors also play a role in regulating these distal elements to control gene expression at a distance.

Distal enhancers control gene expression at a distance

Transcriptional control in Eukaryotes is a complex biological process involving proximal promoters, *trans*-acting protein and RNA factors, and importantly distal *cis*-regulatory elements called enhancers (Müller et al., 1988). Enhancers regulate transcriptional initiation at promoters from distances of several hundred base pairs to 1 Mb away from target transcriptional start sites (TSSs) (Lewin, 2008). Two general models of the role of enhancers have been proposed and validated *in vivo*: firstly, enhancers

can increase the expression of an already active promoter, or secondly an enhancer may activate expression of a primed but inactive promoter (Blackwood and Kadonaga, 1998). One of the first described enhancers was discovered in the genome of the polyomavirus SV40; when this sequence was linked to a test gene in a heterologous reporter assay it increased transcription by over 100-fold, even from distances as great as 3 kb (Banerji et al., 1981). Soon after, many other enhancers were discovered using heterologous reporter assays and enhancer trap techniques (O'Kane and Gehring, 1987; Brand and Perrimon, 1993; Kaiser, 1993). Enhancer discovery progressed rapidly with the advent of genome-wide analysis methods (Sakabe et al., 2012). The sequencing of the human and mouse genomes led to the realization that the vast majority of the DNA is non-protein coding, and while first hypothesized to be "junk DNA", it is now becoming clear that much of the non-coding mammalian genome is regulatory and contains many functional enhancer elements (Visel et al., 2009). For example, non-coding elements which are highly conserved between species have been shown to be enriched for functional, tissue-specific enhancers (Pennacchio et al., 2006). Further, genome-wide maps of chromatin state generate using ChIP-Seq (Figure 1) have shown that specific histone modifications are enriched far from promoters and associated with distal enhancers (Heintzman et al., 2007, 2009; Creyghton et al., 2010; Rada-Iglesias et al., 2011). Specifically, monomethylation of Lysine 4 of Histone H3 (H3K4me1) and acetylation of Lysine 27 of Histone H3 (H3K27Ac) have been shown to mark poised (H3K4me1+, H3K27Ac-) and active (H3K4me1+/H3K27Ac+ or H3K4me1-/H3K27Ac+) enhancers. Genes linked to active enhancers were shown to be preferentially expressed in the cell type in which the enhancers were marked, while those near poised enhancers were shown to be expressed in stimulus-dependent ways, such as upon differentiation. Thus, transcriptional regulation by enhancers is a conserved and widespread genomic phenomenon.

Figure 1

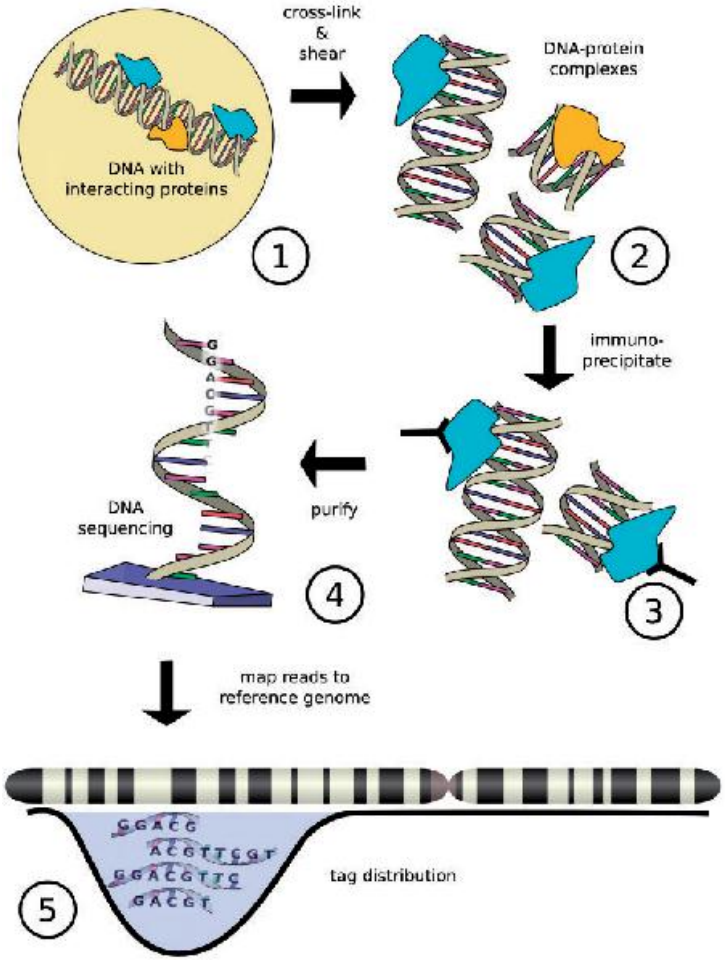


Figure 1-Schematic of a chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-Seq) experiment. Cross-linked chromatin is isolated (1) and sheared into fragments (2), then specific antibodies are used to immunoprecipitate proteins of interest and associated genomic DNA (3). Next, genomic DNA is purified away from bound protein and subjected to high-throughput short-read sequencing (4). When these reads are mapped back to a reference genome, regions enriched by ChIP can be identified and thus genomic targets of a given factors can be defined. ChIP-Seq has become a crucial tool for studying genome-wide binding profiles of many sequence-specific transcription factors, chromatin regulators, and modified histones. Modified from Szalkowski and Schmidt, 2011.

Transcription Factors Regulate Enhancer Activity

Transcription factors are key regulators of distal enhancers. Enhancers are typically comprised of multiple transcription factor binding sites, each of which are essential for function of the element (Lewin, 2008). The classic example of this phenomenon is the IFN-β enhancesome, which is a protein complex comprised of the structural protein HMG1 and sequence-specific transcription factors including NF-κB, IRF, ATF-Jun, and others (Maniatis et al., 1998). The recognition sites for all of these factors are essential to the activity of this element. These factors cooperatively bind to the IFN- β enhancer and then activate this element which positively regulates IFN- β promoter activity. Interestingly, in ESCs the vast majority of TSS-distal regions co-bound by Oct4, Sox2 and Nanog tested performed as enhancers in heterologous reporter assays. Further, proteomic data suggested that these factors interact physically with each other. These data suggest that the core-regulatory circuitry of ESCs may also function as an enhancesome (Kim et al., 2008; Chen et al., 2008). Generally, when transcription factors bind enhancers, they act as activators and recruit co-activators to enhancer elements (Figure 2). Mechanistically, enhancers physically interact with promoters to achieve regulatory function though looping of DNA sequences separating them (Mueller-Strum et al., 1989; Zhao and Dean, 2005; Fullwood et al., 2009). These loops deliver transcriptional co-activators to promoters, which in turn recruit the transcriptional machinery, including RNA polymerase (Figure 2). Thus, enhancers function by interacting with sequence-specific transcription factors, which then recruit co-activators or the transcriptional machinery itself, then finally physically loop to promoters to deliver these molecules, thus activating transcription.

Figure 2

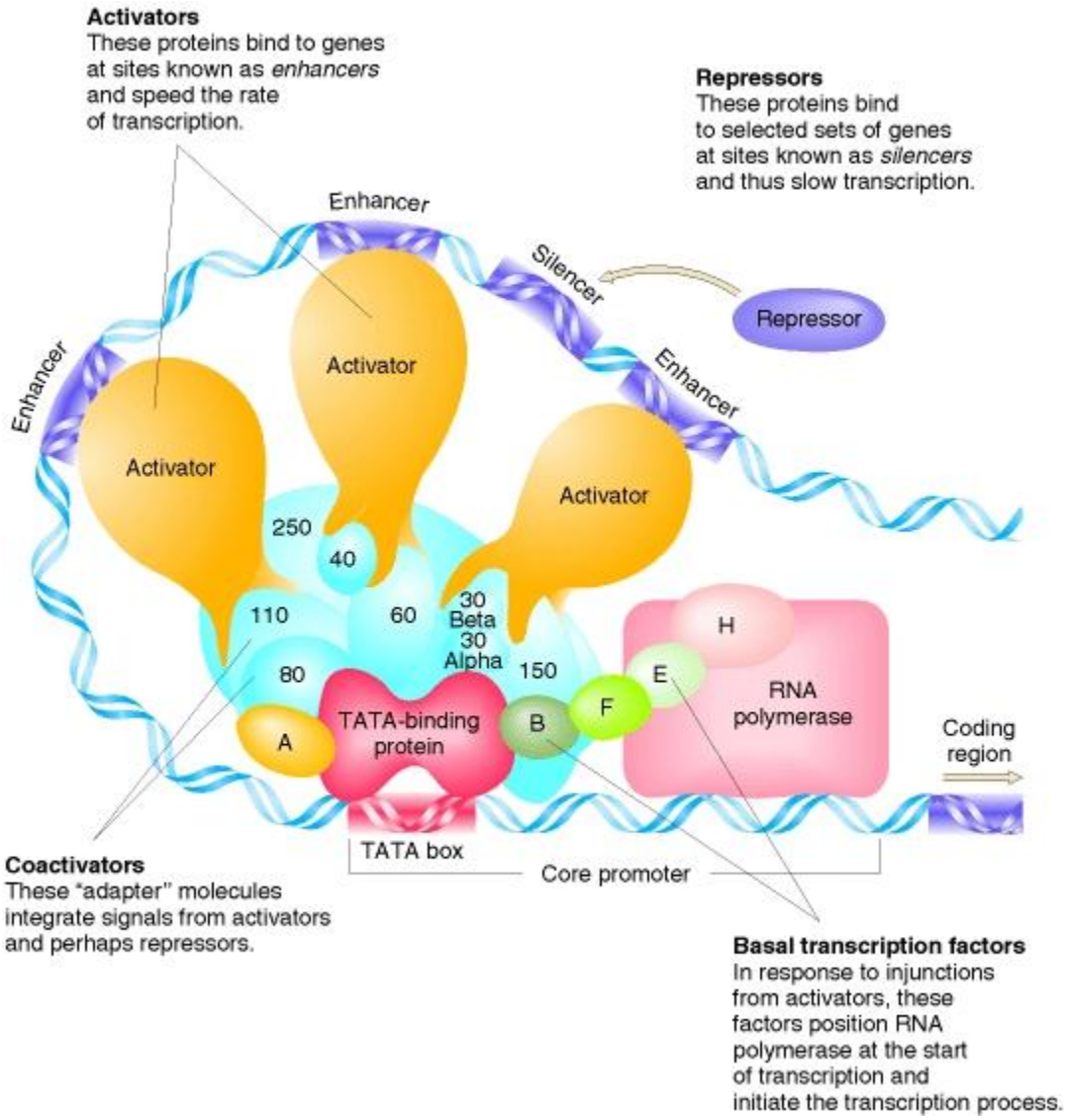


Figure 2-Enhancers loop DNA to deliver co-activators to promoters. The core transcriptional machinery, including basal transcription factors (labeled A, B, F, E, and H) and TATA-binding protein, is essential for transcription but cannot by itself increase or decrease its rate. That task falls to regulatory molecules such as activators, including sequence-specific transcription factors, which bind to regulatory elements such as enhancers and recruit co-activators, such as the Mediator complex. Then, DNA looping allows the co-activators to be delivered to core promoters, where they communicate with basal factors and activate transcription. Other elements such as silencers interact with repressors to negatively regulate transcription. Adapted from Griffiths et al., 2000.

Pioneer Factors Regulate Enhancers in Development

Transcription factors also influence enhancer activity by other mechanisms. The binding of enhancer regions by transcription factors is thought to depend in some part on the occupancy of the DNA sequence by nucleosomes, which can occlude transcription factor binding sites. Transcription factor binding at enhancers may be facilitated by the relatively nucleosome-depleted status of these elements, as evidenced by their hypersensitivity to DNAse treatment (Shones 2008; He et al., 2010). Intriguingly, this nucleosome depletion is thought to be first set-up by a special class of transcription factors called pioneer factors (Smale, 2010). These factors, such as the those belonging to the FoxA subgroup of the forkhead transcription factor family, are thought to bind DNA motifs in a non-histone-sensitive manner and evict nucleosomes from enhancers, which can then be bound by other effector transcription factors to activate transcription (Kaestner, 2010). In both ESCs and NPCs, Sox2 is known to bind a large number of sites in the genome linked to non-expressed genes, which later in development become bound by other factors and expressed (Boyer et al., 2005; Loh et al., 2005; Chen et al., 2008; Marson et al., 2008; Liber et al., 2010; Bergsland et al., 2011), suggesting that Sox2 may be a pioneer factor. Thus, sequence specific transcription factors may prime enhancer elements for later activation.

Sox2 is a master regulator of two distinct cellular states, that of pluripotent ESCs and multipotent NPCs. In order to allow for a comprehensive understanding of the role of Sox2 in these cells, I will first introduce key concepts in ESC and NPCs biology, and then return to the role the transcription factor Sox2 may play in each of these cell types

Embryonic Stem Cells and Neural Precursor Cells:

Two Stem Cell Types with Distinct Characteristics

Embryonic Stem Cells

Embryonic stem cells have been intensely studied for over three decades and since their discovery (Evans and Kaufman, 1981; Martin, 1981) and have become vital tools for studies in development, mouse genetics, and disease modeling. ESCs are the *in vitro* derivative of the inner cell mass (ICM) of the

mammalian blastocyst, and like the cells of the ICM ESCs are pluripotent, meaning they have the potential to give rise to all of the cell types of the embryo proper. When ESCs are injected back into host blastocysts, they can participate in development and have the potential to give rise to an entire adult mouse, including the germ cells (Bradley et al., 1984; Nagy et al., 1990). Due to this unique property, ESCs have been instrumental in the field of mouse genetics. Gene targeting of ESCs by homologous recombination and subsequent injection of targeted cells into blastocysts has made it possible for hundreds of knock-out and reporter mice to be studied (Doetschman et al., 1987; Thomas and Capecchi, 1987). Further, ESCs can be induced to differentiate *in vitro*, allowing for defined and accessible analysis of cell fate transitions in early development. The isolation of human ESCs focused much attention on the ability of these cells to play a role in regenerative medicine and human disease modeling; given that mouse ESCs have the potential to develop into any adult cell type, hESCs are thought to be a viable source for both tissue replacement therapy and disease modeling (Thompson et al., 1998; Mountford, 2008). Due to these distinctive properties, ESCs have been subjected to virtually every level of analysis from studies of individual molecular interactions to large scale genomic and proteomic studies (Young, 2011). Thus, ESCs are a well-characterized and important model of development.

Multiple extracellular signaling pathways have been shown to affect the ESC state. ESCs were originally derived in serum-containing medium in the presence of feeder cells, mouse embryonic fibroblasts (MEFs). The specific signals provided by serum and MEFs were soon identified and provided insights into the molecular characteristics of the ESC state. The primary effector though which serum supports ESCs is the bone morphogenetic protein BMP4, which signals to inhibitor of differentiation (Id) genes through the BMP effectors, SMADs (Ying et al., 2003). Other important factors in serum include Wnt proteins, and the activity of the Wnt pathway has been shown to inhibit differentiation of mouse ESCs (Kielman et al., 2002; Ogawa et al., 2006). The primary function of MEFs is to secret leukemia inhibitory factor (LIF) (Williams et al., 1988; Smith et al., 1988). LIF signaling is essential for ESC maintenance, and is thought to function by binding to LIF receptor causing it to dimerize with gp130,

leading to activation of the transcription factor Stat3 (Niwa et al., 1998). Part of the function of Stat3 is to counteract the function of FGF4, which is produced by ESCs in an autocrine manner. By signaling through FGFR, FGF4 activates the MAPK-pathway mediators ERK1 and ERK2, which promote differentiation of ESCs (Kunath et al., 2007). LIF also signals to the core-regulatory transcription factor network in ESCs which consists of Oct4, Sox2, and Nanog (see below) through Stat3 and the kinase Akt (Niwa et al., 2009). These data eventually allowed Austin Smith and colleagues to derive and grow mouse ESCs in completely defined chemical medium containing LIF and inhibitors of ERK signaling and GSK3β, a negative regulator of the Wnt pathway, a condition called "2i + LIF" (Ying et al., 2008). LIF could be omitted from the growth medium if a pharmacological inhibitor of FGF signaling was added in its place; this growth condition was called "3i" because it relied on these three inhibitors. Cells grown in 2i + LIF or 3i are fully pluripotent and self-renew indefinitely. Thus, the inputs of multiple signaling pathways are integrated by ESCs to make the developmental decision whether to remain in the ESC state or differentiate.

Neural Precursor Cells

During development, neural stem/progenitor cells (abbreviated NPCs herein) give rise to cells of the central nervous system (CNS). In the mouse, neural development begins after gastrulation when the ectoderm further develops into three tissues: the epidermis, which will form the skin; the neural crest, which will migrate to form the peripheral nervous system; and the neural tube, which is destined to form the brain and spinal cord, collectively known as the central nervous system (CNS) (Gilbert, 2000). As development proceeds, cells which remain closest to the lumen at the interior of the neural tube remain mitotic and give rise to neurons, astrocytes, oligodendrocyes, and in addition can self-renew. As the brain develops, these cells occupy specific regions of the developing brain, such as the subventricular zone of the cortex and the subgranular layer of the hippocampus. These cells are therefore known as neural stem cells or neural progenitor cells (NPCs) (Reynolds and Weiss, 1992; Davis and Temple, 1994; Gritti et al., 1996; Reynolds and Weiss 1996). As development of the CNS proceeds, NPCs exhibit a progressive

restriction of developmental potency; early on, NPCs become regionally specified along the anterior/posterior axis, and as development progresses NPCs give rise to specific cell types destined for specific subregions of the CNS in a temporally specific manner (Molyneaux et al., 2007). This program is thought to be intrinsic to the NPCs themselves (Qian et al., 2000; Shen et al., 2006). It was traditionally believed that the number of cells in the brain is fixed at a certain developmental stage, and that no mitotic cells exist in the adult brain. Work by Altman suggested the existence of mitotic cells in specific regions of the rat and feline brain, but it was not until later work in songbirds and mice that the existence of actively dividing cells in the adult brain was fully accepted (Altman, 1962, 1963, 1969, Goldman and Nottelbohm, 1983; Paton and Nottelbohm, 1984; Lois and Alvarez-Buylla, 1993; Johansson et al., 1999). In the adult brain, NPCs are also found in the subventricular zone and the subgranular layer, and adopt a radial glial morphology (Doetsch et al., 1999). Adult NPCs have mitotic and self-renewal potential, but remain quiescent for long periods of time, thus explaining the limited regenerative potential in the adult brain. Thus, NPCs are involved in the embryonic development of and adult maintenance of the CNS.

Multiple extracellular signaling pathways play a role in the development and function of NPCs. In fact, changes in the phenotype and developmental potential of mouse NPCs through ontogeny and into adulthood are accompanied by, and possibly driven by, changes in the responsiveness to extracellular and intracellular signals (Temple, 2001). For example, while mouse NPCs in the early embryo do not respond to EGF signaling, those isolated from late mouse embryos and adults rely on EGF signaling to proliferate (Gritti et al., 1999; Tropepe et al., 1999; Qian et al., 2000). Also, while early NPCs differentiate into neurons upon signaling BMPs, late embryonic NPCs commit to the astrocytic lineage in response to BMP signaling (Mehler et al., 2000). In the adult, the extracellular inhibitor of BMP signaling, Noggin, blocks this pathway to allow adult NPCs to again generate new neurons (Lim et al. 2000). In general, adult NPCs are thought to exist in a perivascular niche, where they are supported by endothelial cells which secrete factors which support the NPC state such as bFGF (Palmer et al., 2000; Shen et al 2004). Other

external signaling pathways, such as Wnt, Notch, and Sonic Hedgehog have been shown to be necessary for the function of NPCs in context specific ways and therefore the proper execution of neural development (Lai et al., 2003; Zhou et al., 2004; Yoon et al., 2005; Lie et al., 2005). Thus, NPCs are regulated at multiple levels to ensure proper execution of nervous system development.

The differential responses to and requirements for extracellular signals are thought to be the result of epigenetic changes in NPCs mediated by transcription factors over developmental time (Qian et al., 2000; Shen et al., 2006; Hirabayashi and Gotoh, 2010). Transcriptional regulators such those belonging to the Sox, Fox, POU and Hes families play critical roles in the maintenance of NPC identity at specific times by actively regulating the stem cell program and, in the case of Hes1 and Hes5, directly counteracting the function of pro-differentiation bHLH transcription factors such as Mash1, Math1, and Neurogenin (Ohtsuka et al., 1999, 2001; Ferri et al., 2004; Kageyama et al., 2005; Castro et al., 2006). These sequence-specific transcription factors activate gene expression programs by recruiting co-activator and co-repressor complexes which modify DNA and histones at specific loci, such as the bithorax group protein MLL1 and the Polycomb group protein Bmi1, both of which have been shown to affect NPC self-renewal (Molofsky et al., 2003; Lim et al., 2009). For example, in early neocortical development, NPCs do not respond to JAK-Stat signaling, but do respond to Wnt signaling by activating proneural transcriptional factors such as Ngn1, which then induce their differentiation to neurons. Later, Wnt signals are not able to induce neural differentiation, while JAK-Stat signaling induces astrogliogenesis. This switch was found to be mediated by DNA and histone methyltransferases which reversibly silenced proastroglial and proneural programs at specific developmental stages and rendered the cells responsive to certain signals but not others (Sun et al., 2001; Guillemot, 2005). Thus, a combination of extracellular signals and epigenetic changes control the progression of NPC competence though development.

The Role of the Transcription Factor Sox2 in Development

Sox2 is a member of a large family of developmental regulators

Sox2 (SRY-box 2) is part of the Sox family, a large group of sequence-specific, developmental regulatory transcription factors. The group comprises over 30 members identified across metazoans, 20 of which are conserved in mammals (Bowles et al., 2000; Keifer et al., 2007). Sox2 is part of the SoxB1 subfamily, along with Sox1 and Sox3, and is conserved from flies to humans (Soriano and Russell, 1998; Bowles et al., 2000). The Sox family was discovered by searching cDNA libraries for sequences homologous to the HMG domain of the male sex-determining factor, Sry (Sinclair et al., 1990; Gubbay et al., 1990; Denny et al., 1992; Wright et al., 1993). The HMG domain is the sequence-specific DNA-binding domain of Sox proteins; the name HMG is derived from the homology this domain has to the DNA binding domains of the high mobility group proteins HMG1 and HMG2, which bind DNA in a non-sequence specific manner and are thought to play architectural roles in chromatin (Thomas, 2001). All Sox family members are defined as being at least 50% identical in their HMG domain to Sry. Sox transcription factors play diverse roles in development, and as a group have been shown to be crucial for the genesis of many major organ systems in mammals (Lefebrve et al., 2007). Thus, defining the function of specific Sox family members such as Sox2 is key to understanding metazoan development in general.

Sox2 was one of the earliest identified Sox family members (Denny et al., 1992 (1); Wright et al., 1993). The *Sox2* locus exists on Chromosome 3 in the mouse genome, and contains a single exon. It encodes a 2.5kb mRNA, the expression of which in the embryo generally marks cell types with the broad developmental potential. Sox2 is expressed in the unfertilized oocyte, and after fertilization it is activated in the embryonic genome during the maternal-to-zygotic transition (Avillion et al., 2003, Zeng and Shultz, 2005; Pan and Shultz, 2011). It is also highly expressed in the inner cell mass (ICM), epiblast and extraembryonic endoderm in the peri-implantation embryo (Wood and Episkopou, 1999; Avillion et al., 2003). After gastrulation, Sox2 expression becomes restricted to the prospective ectoderm, neural plate and the chorion (Wood and Episkopou, 1999; Zappone et al., 2000; Avillion et al., 2003). During

somitogenesis, Sox2 is expressed highly in the neurectoderm, including the primitive streak, and also marks the genital ridge, gut endoderm, prospective sensory placodes, and the brachial arches (Collignon et al., 1996; Wood and Episkopou, 1999; Avillion et al., 2003). Sox2 mRNA is translated into a 39kDa protein containing an HMG DNA binding domain in its N-terminal region. The Sox2-HMG domain binds to DNA sequence specifically in the minor groove, and like other HMG domains the Sox2-HMG domain causes a dramatic bend in the double helix at the point of contact, which is essential for function (Kamachi et al., 1999; Scaffidi et al., 2001; Remenyi et al., 2003; Williams et al., 2004). The C-terminal region of Sox2 contains modular transactivation domains which mediate interactions with transcription factor partners such as Oct4 to stabilize DNA binding and with histone modifiers such as the histone acetyltransferase p300 to affect transcriptional activation (Nowling et al., 2000; Ambrosetti et al., 2000)

Sox2 interacts with partner factors to control gene expression programs

Transcription factors bind to DNA to recruit the transcriptional machinery to specific loci, and often more than one transcription factor binds to a given locus at the same time. By binding in different combinations with one another, transcription factors provide the cell with a diverse toolbox of stereochemical interfaces which may be integrated differentially by the transcriptional apparatus. In this way, transcription factors can to specify distinct gene expression programs utilizing the small number of transcription factors (about 2000) encoded in the genome to regulate all genes (20,000) across 200 cell types during development and in the face of a host of external stimuli (Remenyi et al., 2004).

The Sox protein family interacts with several other transcription factor families to control gene expression. This is partially necessary for Sox proteins because partner factor interactions have been shown to stabilize them on DNA targets, for which they have relatively low binding affinity relative to other transcription factors (K_D = approximately 10^{-9} M for Sox-HMG compared to 10^{-11} M to 10^{-12} M for other transcription factors) (Kamachi, 2000). Sox2 has been shown to interact with a number of other factors in this way. In the developing lens, Sox2 has been shown to interact with the paired box transcription factor Pax6 and the bZIP transcription factor Maf to cooperatively bind to and regulate

crystalline gene expression (Kamachi et al., 2000; Shimada et al., 2003). Further, Sox2 and Pax6 have been shown to regulate Sox2 expression by binding to a distal enhancer of the Sox2 locus in the lens (Inoue et al., 2007). Perhaps one of the most important classes of partner factors for Sox proteins is the POU-homeobox family of transcription factors; the interaction between Sox and POU family members is conserved across metazoans from *Drosophila melanogaster* and *Caenorhabditis elegans* to mammals (Dailey and Basilico, 2001). Many Sox:POU modules have been described which cooperatively control gene expression in the mouse (Ambrosetti et al., 1997; Botquin et al., 1998; Kuhlbrodt et al., 1998; Tomioka et al., 2002; Remenyi et al., 2003; Tanaka et al., 2003; Donner et al., 2007; Reiprich et al., 2010), including the well characterized interplay between Sox2 and Oct4 in ESCs (below). In general, Sox:POU interactions are thought to stabilize binding to DNA targets and result in synergistic activation of target loci. Thus Sox proteins, and Sox2 in particular, utilize interactions with partner factors to achieve regulatory functions in many systems.

The role of Sox2 in early development

Effects of Sox2 loss-of function *in vivo*

Given the expression pattern of Sox2 in the early embryo and *in vitro* derivatives thereof, much interest was focused on defining the functional role of Sox2 during development *in vivo*. Therefore, a Sox2 loss-of-function allele was generated (Avillion et al., 2003). Animals heterozygous for this allele were grossly normal, save for a minor, background dependent decrease in male fertility. Sox2 homozygous null embryos failed around implantation; while Sox2-null blastocysts appeared normal, defects which manifested shortly after implantation resulted in a failure to properly form an epiblast, an embryonic tissue which is destined to give rise to all the tissues of the adult. Moreover, the inner cell mass (ICM) of these blastocysts failed to outgrow in culture into stable embryonic stem cell (ESC) lines. The ability of these embryos to form a phenotypically normal but ultimately nonfunctional ICM is attributed to the persistence of maternal stores of Sox2 mRNA and protein. Inspection of failed implantation sites derived from heterozygous intercrosses (presumed to be the remains of failed Sox2-null

embryos) revealed that while Oct4-positive epiblast cells were absent, disorganized H19-positive and Evx1-positive tissue were present, indicating that Sox2 is not necessary for early extraembryonic lineages, and that early non-epiblast ICM derivatives such as visceral endoderm were formed. Presumably, lack of signaling from a functional epiblast caused these tissues to fail to develop further. To test this, wild type ESCs were injected into Sox2-null blastocysts. While these cells were able to rescue the early embryonic lethality of these embryos, they still did not yield viable full-term fetuses. Thus, while Sox2 is dispensable for early extraembryonic lineages, it may function in later extraembryonic tissues, such as extraembryonic ectoderm, the chorion, trophoblast cells, and the placenta, where expression of Sox2 is indeed detected. Therefore, Sox2 is essential for mouse embryogenesis, both though the regulation of early, pluripotent, embryonic lineages and possibly by regulating later, extraembryonic tissues.

Role of Sox2 in ESCs

Since ESCs could not be derived from Sox2-null embryos, the loss-of-function phenotype in Sox2-null ESCs was addressed using a conditional knockout/inducible transgene strategy (Masui et al., 2007). In this system, both alleles of Sox2 were targeted with constructs which flanked the coding region with LoxP sites. Then, an inducible Sox2 transgene was introduced into these cells. Finally, Cre expression was induced to cause recombination of the Sox2 flanking LoxP sites resulting in loop-out of the coding sequence; simultaneously Sox2 transgene expression was induced, yielding ESCs dependent on this inducible Sox2 expression. Once expression of this transgene was turned off, the phenotype of acute loss of Sox2 in ESCs could be assessed. Loss of Sox2 expression caused ESCs to transdifferentiate into trophoblast-like cells, a fate which wild-type ESCs, ICM, and epiblast do not adopt (Beddington et al., 1989). This is the same phenotype as loss-of-function of Oct4 (Niwa et al., 2000), suggesting these factors control similar pathways. A similar result was obtained by another group, using a Sox2 dominant negative strategy (Li et al., 2007). While these Sox2-null ESCs could not be maintained and rapidly down-regulated many Sox2 targets such as Oct4, not all Sox2 target genes were immediately and fully

down regulated. This was due to compensation by other Sox factors expressed in ESCs, including Sox4, Sox11, and Sox15. Forced expression of Oct4 could rescue many of the defects of these ESCs by partnering with these other Sox factors, allowing Sox2-null ESCs to contribute to late stage chimeric embryos, although the viability of these embryos to term was not addressed. Thus, ESCs depend on Sox2 function to maintain pluripotency.

Sox2 and Oct4 co-operate to globally regulate ESC state

The first Sox2 target identified was an enhancer of the *Fgf4* locus in embryonic carcinoma cells (ECCs), an element which it co-regulates with Oct4. Studies of Sox family members identified general features of Sox binding to DNA targets (Nasrin et al., 1991; Denny et al., 1992 (2)), and using these data it was possible to scan regulatory regions of genes for candidate binding sites for Sox2. In ECCs, the *Fgf4* locus was known to be under the control of an intronic enhancer element containing an Octamer motif, the recognition site for the POU family of transcription factors (Curatola et al., 1990). This element was bound by the POU factors Oct1 and Oct3 (later renamed Oct4), in complex with another nuclear factor, provisionally named Fx (Dailey et al., 1994). The assembly of this ternary complex was dependent on a sequence nearby the Octamer motif, 5'-CTTTGTT-3' which was noted to be similar to the previously identified Sox motif. Using a degenerate PCR approach to amplify cDNAs homologous to the Sry HMG domain, Sox2 was found to be expressed in ECCs and it was identified as factor Fx (Yuan et al., 1995). The interaction between Sox2 and Oct4 at the *Ffg4* intronic enhancer was shown to result in cooperative binding to DNA and synergistic activation of this enhancer (Figure 3), a paradigm which would later manifest itself at many Sox2:Oct4 targets and further at many Sox:POU targets in other tissues.

Figure 3

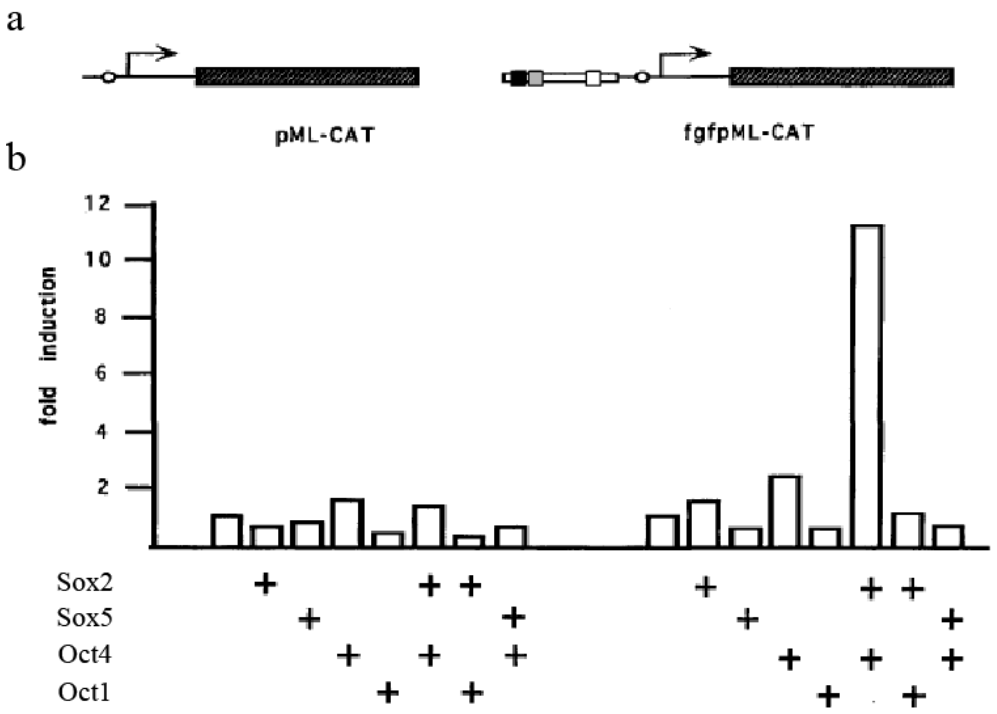


Figure 3: Synergistic activity of Sox2 and Oct4 at an *FGF4* enhancer. a. Schematic of control (left) and heterologous, FGF4-enhancer-driven (right, white rectangle) chloramphenicol acetyltransferase (CAT) reporter constructs. Black box, Sox2 binding site; gray box, Oct4 binding site; white box, binding site for another transcription factor, Sp1; white circle, TATA box; hashed box, CAT CDS. b. Ectopic Sox2, Sox5, Oct4, and Oct1 show limited ability to induce expression of this reporter alone, but Sox2 and Oct4 synergistically activate CAT expression when transfected together. Adapted from Yuan et al., 1995.

The partnership observed between Sox2 and Oct4 in ECCs was confirmed in ESCs at many target loci, and is mediated through defined domains in the DNA binding and transactivation regions of the Sox2 and Oct4 proteins. In addition to *Fgf4*, they have been shown to cooperatively bind and synergistically activate numerous targets in ESCs, including *UTF1*, *ZFP206*, and *Lefty1* (Nishimoto et al., 2005; Nakatake et al., 2006; Wang et al., 2007). Crystal structures of the HMG domain of Sox2 in complex with the POU domain of the Oct4 family member Oct1 at two genomic targets (enhancers of

Fgf4 and *HoxB1*) suggested that specific residues within the DNA binding domains of these proteins were critical for complex stabilization (Remenyi et al., 2003; Williams et al., 2004). Other studies of regions outside the HMG and POU domains of Sox2 and Oct4 demonstrated that non-DNA binding domains in the C-terminus of Sox2 (region R3) and in the C-terminus of Oct4 (region AD2) were modular transactivation domains which functioned to stabilize Sox2:Oct4:DNA complex formation, while other domains were defined (R1, R2, AD2) which are transactivating but not involved in this interaction directly (Ambrosetti et al., 2000). Thus, Sox2:Oct4 co-regulate many targets in ESCs through extensive protein: protein interactions which stabilize them on genomic targets.

Sox2:Oct4 co-regulation is a critical genome-wide phenomenon in ESCs. Genomic studies showed that Sox2 and Oct4 (along with the divergent homeodomain protein Nanog) co-occupy thousands of genomic target sites near promoter regions in mouse and human ESCs (Boyer et al., 2005; Loh et al., 2006) (Figure 4). Genes bound by Sox2:Oct4 included many regulators of pluripotency and the ESC state, including Oct4, Sox2, and Nanog themselves, indicating that these factors form an autoregulatory loop (Boyer et al., 2005; Chew et al., 2005; Okumura-Nakanishi et al., 2005; Kuroda et al., 2005; Rodda et al., 2005; Loh et al., 2006). Surprisingly, Oct4, Sox2, and Nanog also occupied lineage commitment factors which were repressed in ESCs (Figure 4). These repressed genes were demonstrated to be marked by the Polycomb repressive complex with trimethylation of Lysine 27 on Histone H3 (H3K27me3), a well characterized mechanism in gene silencing (Lee et al., 2006; Boyer et al., 2006). Later, whole genome studies allowed Sox2- and Oct4-occupied sites distal to TSSs to be mapped (Marson et al., 2008; Chen et al., 2008). These studies demonstrated that the vast majority of Sox2 targets in ESCs are co-occupied by Oct4, and that a large fraction of genomic loci bound by these factors are distal to transcriptional start sites. Further, when tested in heterologous reporter assays these distal regions behave as enhancers. Taken together with proteomic interaction data (Kim et al., 2008), this suggests that Sox2 and Oct4 may be at the center of an ESC-specific enhancesome which controls ESC identity. The potency of Sox2 and Oct4 to enforce the ESC state was definitively demonstrated in 2006, when

Takahashi and Yamanaka showed that the ectopic expression of four transcription factors, Oct4, Sox2, Klf4 and c-Myc could completely reprogram a somatic epigenome back to a pluripotent ESC-like state (Takahashi and Yamanaka, 2006, see below). Thus, the Sox2:Oct4 partnership sits at the center of a core-regulatory network controlling ESC identity.

Figure 4

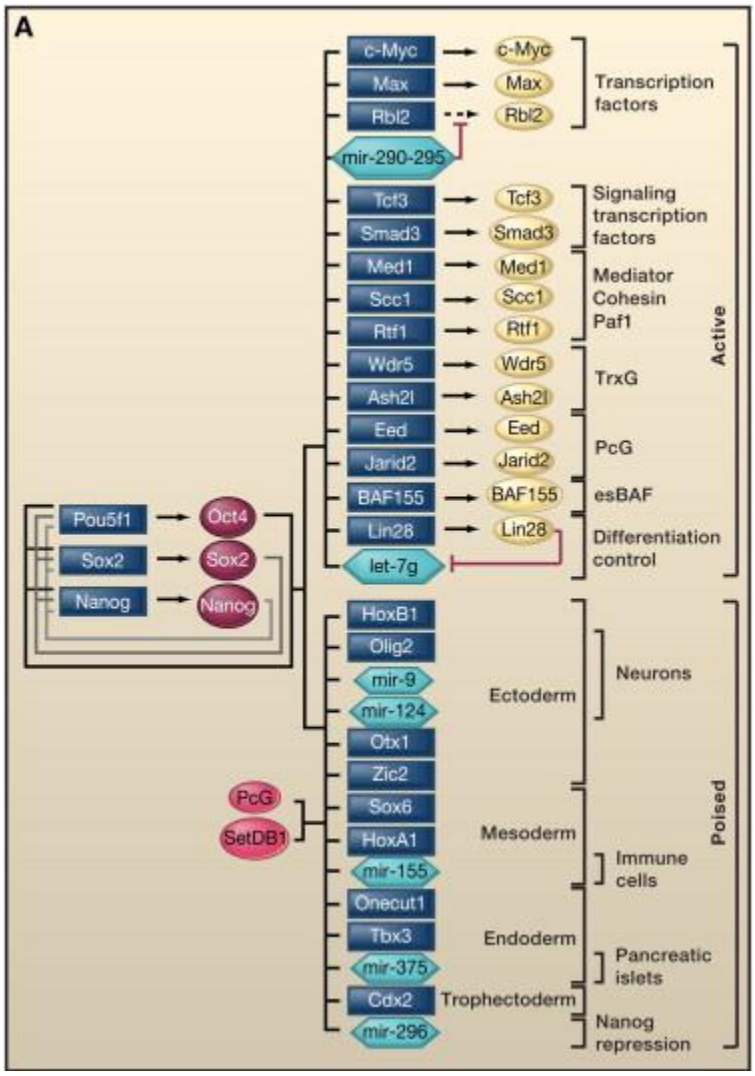


Figure 4: Sox2, Oct4, and Nanog form an autoregulatory loop which controls ESC pluripotency. Sox2, Oct4, and Nanog occupy regulatory regions of genes involved in both ESCs identity (top), including *Pou5f1* (the locus which encodes Oct4), *Nanog*, and *Sox2* themselves (left). These transcription factors also occupy loci encoding lineage commitment factors (bottom), which are poised to become

activated upon differentiation but kept repressed in the ESCs through the action of repressive epigenetic regulators such as Polycomb group proteins. Adapted from Young, 2011.

The role of Sox2 in neural development

Expression pattern of Sox2 in neural development

The central role Sox2 plays in the pluripotency of ESCs and the fact that Sox2 is expressed in the brain raised the possibility that Sox2 may regulate NPCs, the stem cells of the developing and adult nervous system. Detailed analysis of the expression pattern of Sox2 further validated this hypothesis. A series of β-Geo transgenes (encoding a fusion protein conferring neomycin resistance and β – galactosidase [β-Gal] activity) driven by Sox2 regulatory elements allowed for precise identification of Sox2-expressing regions of the developing and adult nervous system (Zappone et al., 2000). Using these tools, Sox2 was shown to mark neurogenic regions throughout development. Sox2 expression in the epiblast was maintained in the ectoderm specifically and lost in the mesoderm and endoderm. Sox2 then marks the neural plate at e7, and later the early neuroepithelium from e8.5 to e10.5. By e12.5 expression marks to the dorsal side of the telencephalic vesicles, which are destined to become the cortex. As this tissue develops, Sox2 remains high in the mitotic, undifferentiated regions near the ventricular zone, and is low in the higher, less mitotic and more differentiated layers. By e17.5 Sox2 expression persists in the neurogenic ventricular zone, and in the adult is maintained there, in the hippocampus, and in the rostral migratory stream, which consists of migrating mitotic neuroblasts destined to become fully differentiated olfactory neurons when they reach the olfactory bulb. Thus, Sox2 marks neurogenic regions of the developing and adult nervous system.

This pattern suggested that Sox2 may mark NPCs in the brain. To test this directly, telencephalic regions of e14.5 Sox2-β-Geo transgenic embryos were explanted in grown *in vitro* as neurospheres. The neurosphere assay was one of the first experimental tests developed to measure stem cell activity in the developing brain (Reynolds and Weiss, 1992). In these assays, cells are explanted and grown in suspension at clonal density. Cells with stem cell properties can proliferate to give rise to spherical

floating colonies of cells, containing a mix of differentiated cells and stem and progenitor cells. In this experiment, stem cells (as defined by the ability to give rise to neurospheres) showed both neomycin resistance and β-Gal activity, indicating that they expressed Sox2. Another study using mice harboring a Sox2-GFP reporter transgene demonstrated that the GFP-positive fraction of cells in the embryonic brain contained almost all of the cells with the potential to create neurospheres (D'amour and Gage, 2003). In an *in vivo* setting, lineage tracing analysis demonstrated that Sox2-expressing cells in the adult hippocampus give rise to neurons, astrocytes, oligodendrocyes, and importantly new Sox2-positive NPCs, demonstrating that they are indeed stem cells (Figure 3) (Suh et al., 2007). Thus, Sox2 marks NPCs, the stem cells of the developing and adult central nervous system.

Figure 3

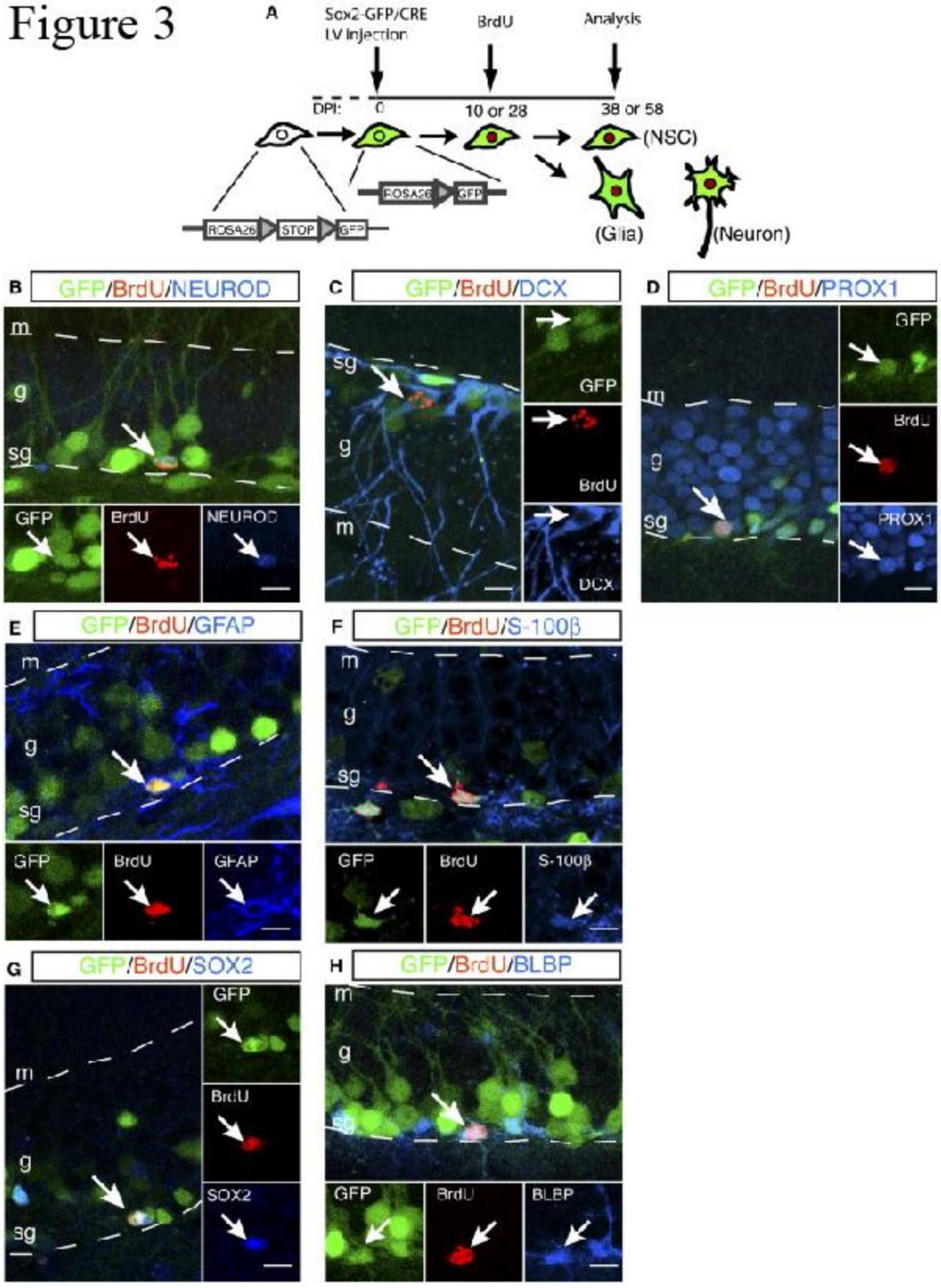

A
Sox2-GFP/CRE
I.V injection
BrdU
Analysis
DPI:
0
10 or 28
38 or 58
(NSC)
ROSA26
GFP
ROSA26
STOP
GFP
(Glia)
(Neuron)
B
GFP/BrdU/NEUROD
m
g
sg
GFP
BrdU
NEUROD
C
GFP/BrdU/DCX
sg
g
m
GFP
BrdU
DCX
D
GFP/BrdU/PROX1
m
g
sg
GFP
BrdU
PROX1
E
GFP/BrdU/GFAP
m
g
sg
GFP
BrdU
GFAP
F
GFP/BrdU/S-100β
m
g
sg
GFP
BrdU
S-100β
G
GFP/BrdU/SOX2
m
g
sg
GFP
BrdU
SOX2
H
GFP/BrdU/BLBP
m
g
sg
GFP
BrdU
BLBP

Figure 3: Sox2+ cells proliferate to give rise to differentiated neurons, glia, and new Sox2+ cells. A. Experimental outline. A lentiviral construct encoding GFP and Cre recombinase transgenes under the control of Sox2 regulatory elements (Sox2-GFP/CRE) was injected into the dentate gyrus of the hippocampus of ROSA26-lox-Stop-lox-GFP reporter mice. Since lentiviral Cre and GFP were under the control of Sox2 regulatory elements, Cre-mediated recombination of the floxed Stop cassette in the ROSA locus occurred in Sox2 expressing cells. Importantly, lentiviral GFP and Cre were expressed in Sox2+ cells only, and shown to be efficiently repressed as NPCs turned off endogenous Sox2 and differentiated. Thus, GFP expression in these experiments in non-NPC cell types likely originates from the ROSA26 locus after Cre-mediated excision of the floxed stop cassette. The recombination at the genomic level allowed lineage tracing of cells expressing Sox2 at the time of infection. Ten or 28 days after virus injection, BrdU (red) was administered to label proliferating cells. One month after BrdU injection, the fate of progeny was examined by staining for cell-type-specific markers and GFP. Sox2 expressing cells underwent cell proliferation as evidenced by BrdU incorporation and gave rise to neuronal precursors positive for NEUROD (B) or DCX (doublecortin) (C) and granular neurons expressing PROX1 (D). $GFAP^+$ (E) and $S\text{-}100\beta^+$ (F) glial cells were also generated from $Sox2^+$ cells by mitosis. $Sox2^+$ cells also gave rise to new undifferentiated cells positive for SOX2 (G) or BLBP (H) by mitosis. Abbreviations are the following: s, subgranular zone; g, granular layer; m, molecular layer. Scale bar, 10 μm. Adapted from Suh et al., 2007.

Functional role of Sox2 in NPCs

Sox2 is a marker of NPCs, but what functional role does it play in these cells? To answer this question, genetic studies were carried out to examine the gain-of-function and loss-of-function phenotypes of Sox2 in the nervous system specifically. Overexpression of Sox2 in NPCs in the developing chick prevented these cells from migrating into the outer layers of the neural tube and differentiating, indicating that exogenous Sox2 blocks exit from the progenitor state and adoption of differentiated identity (Graham et al., 2003; Bylund et al., 2003). This phenomenon was later confirmed in mouse, where overexpression of Sox2 prevented immature Schwann cell precursors from differentiating into myelinating oligodendrocyes in the spinal cord, and overexpression of Sox2 in the developing neocortex prevented the differentiation of precursor cells into neurons (Le et al., 2005; Bani-Yaghoub et al., 2006). Further, expression of a Sox2 dominant negative cDNA in the developing chick neural tube caused precocious differentiation of NPCs in neurons (Graham et al., 2003; Bylund et al., 2003). This indicated that loss of Sox2 function in NPCs forced them to exit the NPC state, but was complicated by the existence of other Sox family members expressed in NPCs, the function of which could also be affected by a Sox2 dominant negative. Taken together, these studies demonstrated that

Sox2 expression enforces the NPC state and suggests that Sox family function in NPCs is important for the maintenance of NPC identity.

Genetic studies in the mouse further refined our understanding of the role of Sox2 in NPCs. In one such system, compound heterozygous mice harboring one Sox2 knock-out (KO) allele and one hypomorphic allele were generated. These mice exhibited a wild-type Sox2 tissue expression pattern, but at a level 25%-30% of that of wild-type (Ferri et al., 2004). Sox2 KO/hypomorph mice exhibited reduced viability beginning in late embryogenesis and pleiotropic brain defects, including cerebral malformations with loss of mature neural subtypes and ventricle enlargement, loss of neurones, L-Dopa-rescuable circling behavior, and epilepsy. Importantly, in these mice NPC proliferation in the dentate gyrus, the ventricular zone, and the rostral migratory stream was dramatically decreased, and adult GFAP-positive/Nestin-positive hippocampal neural precursor cells are strikingly diminished. These data suggest that reduced expression of Sox2 results in brain malformation and compromised NPC function.

Later, Sox2 was conditionally ablated in NPCs Cre-Lox strategies, in which mice homozygous for a Sox2 allele in which the coding sequence is flanked by LoxP sites are bred into a strain which expresses a Cre recombinase transgene under the control of regulatory element of the *Nestin* gene (Miyagi et al., 2008; Favaro et al., 2009). Both studies demonstrated that the genesis and function of NPCs during neural development was affected in neurogenic regions of the developing CNS by the loss of Sox2, and that the maintenance of NPCs after birth was severely compromised in mutant animals. This was complicated by the expression of other Sox factors in NPCs which may compensate for the loss of Sox2, some of which were shown to be upregulated in Sox2-null NPCs (Miyagi et al., 2008). In a 129Ola background, Sox2 conditional mice did not survive longer than 12 hour post-partum (Miyagi et al., 2008). In a 129S1/SvI background, Sox2 conditional mice also showed high mortality and do not survive past four weeks (Favaro et al., 2009). The cause of this background specific difference is not clear. In 129S1/SvI Sox2-null mice, at P0 there was a reduction in neurogenesis, and by P7 there was a total loss of GFAP-positive/Nestin-positive NPCs in the brain, no proliferation, and an increase in apoptosis. Both

studies show that when cultured *in vitro* as neurospheres, Sox2-null NPCs have reduced proliferative and developmental potential. The loss-of-function phenotype of Sox2 in NPCs is similar to the loss of components of the Wnt and Sonic Hedgehog (Shh) signaling pathways in these cells (Machold et al., 2003). Interestingly, exogenous Shh or Shh chemical agonists can rescue some of the loss of Sox2 phenotype, indicating that part of the function of Sox2 in NPCs is to regulate this pathway. Thus, Sox2 is not only expressed in NPCs but is crucial for NPC maintenance and cell identity.

The genome-wide role of Sox2 in NPCs

Given the critical role Sox2 plays in NPCs, attempts were made to define the genome-wide targets of Sox2 in these cells. In one such study, Sox proteins were profiled from ESCs, where Sox2 is expressed, through NPCs which express Sox2 and Sox3, up to immature neurons which express Sox11 (Bergsland et al., 2011). These data confirmed that in ESCs, Sox2 occupied a set of repressed regulators of neural development and suggested that many of these loci became bound by Sox3 in NPCs and activated. While data presented indicating a high degree of overlap between Sox2 and Sox3 bound regions in NPCs, only Sox3 targets are characterized fully in this study. In NPCs, Sox3 was shown to occupy the regulatory regions of a subset of Sox2-ESC target genes involved in neural development, some of which were expressed in NPCs. This indicated that in NPCs Sox3 could activate a subset of the loci which were poised by Sox2 in ESCs. Finally, upon terminal differentiation some non-expressed Sox2-ESC/Sox3-NPC targets were passed off to Sox11 in immature neurons and activated. Thus, Sox proteins seem to act as pioneer factors throughout neural development. Left undefined by this study is the complete nature of Sox2 targets in NPCs, and importantly how such similar factors can have varying effects on targets depending on the cellular context. Another study compared Sox2 targets genome-wide in NPCs to the ATP-dependent chromatin remodeler Chd7 (Engelen et al., 2011). These factors overlapped extensively, and co-occupied many targets which are disregulated in CHARGE syndrome, which is known to be caused by haplo-insufficiency of Chd7. Interestingly, CHARGE syndrome shares many clinical features with a disease associated with loss of Sox2 function, Anophthalmia (see below

"Sox2 in Diseased States"). It was not clear from this study whether the interaction with a chromatin remodeler such as Chd7 could stabilize Sox2 on DNA targets in the same way a sequence specific transcription factor like Oct4 does in ESCs, so how Sox2 can bind to NPC-specific targets without complexing with a transcription factor such as Oct4 was not addressed. These studies have added to our understanding of the function of Sox2 in NPCs, but left many questions unanswered.

Sox2 in other Tissues

Sox2 marks stem/progenitor cells in non-neural lineages and functions to regulate stem/progenitor identity

Sox2 is expressed in tissues other than ESCs and NPCs, and many of these cells also display stem/progenitor cell characteristics. During development and in the adult, Sox2 marks cells in various tissues, including the lens, testes, lung, trachea, stomach, embryonic foregut, inner ear, taste bud, the stomach and many regions of the gastrointestinal tract, cervix, and thymus (Kamachi 1995; Cimpean et al., 2001; Mansukhani et al., 2005; Kiernan et al., 2005; Okubo et al., 2006; Gontan et al., 2008; Domyan et al., 2011; Arnold et al., 2011). In stem/progenitor cells of the testes, lung, trachea, taste bud, lens, and in osteoblasts, loss-of-function of Sox2 leads to a decrease in proliferation, loss of stem/progenitor cell markers, and disruption tissue homeostasis, demonstrating that Sox2 is important for stem cell identity in these contexts, similar to its role in ESCs and NPCs (Okubo et al., 2006; Que et al., 2009; Tompkins et al., 2009; Basu-Roy et al., 2011; Holmes et al., 2011; Arnold et al., 2011). Lineage tracing has confirmed that in the testes, glandular stomach, and in the lens, Sox2-positive cells give rise to differentiated cells (Arnold et al., 2011). For example, in the glandular stomach, Sox2-expressing cells give rise to enteroendocrine cells, mucus cells, and parietal cells, indicating that they are multipotent, and in the testes Sox2-positive spermatogonial stem cells can entirely reconstitute spermatogenesis when transplanted into infertile male mouse testes (Arnold et al., 2011). Thus, Sox2 is a marker of stem cells in multiple tissues.

Sox2 in diseased states

Given the potent effect of Sox2 on development, it is perhaps not surprising that dysfunction of Sox2 can lead to disease in humans. Mutations in Sox2 have been linked to the ocular malformation

disease Anophthalmia, and more broadly to Anophthalmia-Esophageal-Genital (AEG) syndrome, in which patients exhibit developmental defects in the eye, esophagus, and genital region (Fantes et al., 2003; Guichet et al., 2004; Hagstrom et al., 2005; Ragge et al., 2005; Williamson et al., 2006). Some patients with mutations in Sox2 also exhibit deafness (Hagstrom et al., 2005). Interestingly, Sox2 has been shown to play a role in the development of the eye, esophagus, testes, and inner ear tissues in the mouse (above). Sox2 has also been shown to mark many cancers in tissues in which Sox2 normally plays a role in stemness, including glioblastoma in the brain, stomach adenocarcinoma, lung and esophageal squamous cell carcinoma, thymoma, osteosarcoma and others (Vural et al., 2008; Bass et al., 2009; Cimpean et al., 2001; Basu-Roy et al., 2011; Fang et al., 2011), suggesting that these cancers might co-opt endogenous stem cell-like programs during oncogenesis in these tissues. Thus, the high potency of Sox2 in developmental fate transitions is reflected in the human diseases manifested when Sox2 expression is abnormal.

Brn1 and Brn2 are CNS Specific POU Factors Which May Partner with Sox2

Brn1 and Brn2 are part of the POU-homeobox family

The POU-homeobox family was first defined in 1988, when many labs pointed out the high degree of homology in the DNA binding domains between the pituitary expressed gene mapping to the *dwarf* locus in mouse, Pit-1, the mouse Octamer binding transcription factors Oct1 and Oct2, and the CNS expressed *C. elegans* gene Unc-86 (Herr et al., 1988). POU factors all contain a bi-partite DNA binding domain consisting of a POU homeodomain (POU_H) and a POU specific domain (POU_S), separated by an flexible linker (Ryan and Rosenfeld, 1997). These linkers are highly variable between POU factors, but highly conserved within sub-classes, even across species (Treacy and Rosenfeld, 1992). Outside the POU domain, there is little conservation between POU factors, but some amino-acid rich motifs are shared within subclasses, such as the presence of Glutamine-rich repeats in class III POU factors (which includes Brn1 and Brn2) (Treacy and Rosenfeld, 1992). Like Oct1 and Oct2, they all bind a variant of the canonical Octamer motif 5'-ATGCAAAT-3', originally discovered to drive expression of

immunoglobulin genes in the lymphoid lineage and be bound by a ubiquitously expressed factor (Oct1), and a lymphoid specific factor (Oct2) (Mason et al., 1985; Wirth et al., 1987; Wang et al., 1987; Staudt et al., 1988; LeBowitz et al., 1988). Using a degenerate PCR approach searching for cDNAs with homology to the POU domain, Brn1 and Brn2 were cloned, and probes from these cDNAs were used to show that Brn1 and Brn2 were widely expressed in the developing and adult rat nervous system in an almost completely overlapping way (He et al., 1989). Notably, Brn1 was detected in the rat kidney while Brn2 was absent in this organ. The overlapping expression pattern on the cellular level of Brn1 and Brn2 and the high homology in their DNA binding domain indicates that they may be able to regulate the same genomic targets *in vivo* (Treacy and Rosenfeld, 1992). Soon the mouse Brn1 and Brn2 homologs were cloned from the developing mouse brain, indicating their expression pattern is conserved (Hara et al., 1992). Many other POU factors were discovered in this way until the family was found to consist of seven subfamilies, Groups I-VII, and encompass more than 150 POU-domain containing proteins across species including at least 14 in mouse (Phillips and Luisi, 2000). Brn1 and Brn2 fall into POU group III, and are designated Pou3f3 (POU group III, factor three) and Pou3f2 (POU group III, factor two), respectively. Thus, like Sox2, Brn1 and Brn2 belong to a large and highly conserved family of transcription factors.

Loss-of-function of Brn1 and Brn2 causes pleiotropic brain defects

Knock-out (KO) of Brn1 and Brn2 resulted in dysfunction of NPCs and subsequent malformation of specific brain regions. Brn2 KO animals were not viable, exhibiting lethality between P0 and P10 (Nakai et al., 1995). They exhibited a hypoplastic pituitary gland which led to a small body size and a decrease in body fat. In the brain, specific neuronal subtypes (mangocellular neurons of the paraventricular nuclei and the supraoptic nuclei of the hypothalamus) are totally absent, due to death of NPCs at e12.5 which are destined to give rise to these cell types. Brn1 KO animals also die shortly after birth, probably due to renal failure (Nakai et al., 2003), but also exhibit multiple brain defects including a malformed hippocampus due to a failure of NPCs to properly migrate to the outer layers of this tissue

(McEvilly et al., 2002). Interestingly, expression of Brn1 was increased 30% at the protein level in homozygous Brn2 knockout animals, thus it is possible that the upregulation of Brn1 in Brn2 KO animals was compensating for the loss of Brn2 and masking the true role of this factor. Further, detailed analysis of Brn1 and Brn2 expression revealed that only Brn1 was detected in the CA1 region of the hippocampus. Brn2 was not upregulated in hippocampus of Brn1 KO animals, possibly explaining why this region was specifically affected by loss of Brn1. To address the question of compensation Brn1/Brn2 double knock-out (DKO) mice were generated. In Brn1/Brn2 DKO animals, regions affected in each single KO were similarly malformed, but new brain regions not known to be affected in either single KO were also affected, such as the cerebral cortex and specifically the neocortex (McEvilly et al., 2002; Sugitani et al., 2002). In this region, NPCs in the subventricular zone at e13.5 and beyond were non-proliferative and failed to migrate to the outer layers of the cortex. This defect was shown to be cell autonomous, as other cells in the region which guide NPC migration neither expressed Brn1/Brn2 nor were affected by their loss. Thus, Brn1 and Brn2 have both distinct and overlapping roles in the developing brain, specifically in hypothalamic, hippocampal, and cortical NPCs.

Brn1 and Brn2 function with Sox2 at NPC specific target loci

Brn1 and Brn2 therefore seemed to play roles in the development of the brain similar to that of Sox2. Interestingly, known target loci of Brn1 and Brn2 were shown to be shared with Sox2, indicating that they may constitute a new Sox:POU module regulating development. The Nestin *Nes30* enhancer lies within the first intron of the *Nestin* locus and is conserved in mouse, rat, and human. This enhancer contains Sox and POU binding sites in close proximity to one another, and these sites must maintain correct orientation in respect to one another for proper enhancer activity (Tanaka et al., 2004). *In vitro*, Sox2 and Brn2 were able to synergistically drive expression from a reporter construct regulated by this sequence, but not if the spacing or orientation of the Sox and Octamer site is altered. Another important enhancer in the developing nervous system is *SRR1*, a 5' regulatory element of *Sox2* which drives Sox2 expression in the neural lineage. Electrophoretic mobility shift experiments demonstrated that two

protein:DNA complexes formed on this element when it was incubated with mouse brain lysates; one Brn1 dependent, another Brn2 dependent (Catena et al., 2004), indicating that this element may be bound and regulated by Brn1 and Brn2 in the mouse brain. Direct evidence of Brn1 and Brn2 regulation of Sox2 *in vivo* was provided when chromatin immunoprecipitation assays from the forebrain of adult mice showed that Brn1, Brn2 and Sox2 occupied a 3' regulatory region of *Sox2*, *SRR2* (Miyagi et al., 2006). Recently, in an analogous fashion to the role of Sox2 and Oct4 in factor mediated reprogramming, ectopic expression of Sox2 and Brn2 with the forkhead factor FoxG1 was shown to transdifferentiate fibroblast cells into neural precursor-like cells, which could differentiate and self-renew (Lujan et al., 2012). Therefore, part of the function of Sox2 and the Brn factors may be to regulate gene expression programs in NPCs in a combinatorial way to ensure proper execution of CNS development.

Summary

Sox2 is a master regulator of both pluripotent ESCs and multipotent NPCs. These cell types share certain characteristics, such as high developmental potency and self-renewal, but are molecularly and phenotypically distinct. The major question of this thesis will be how Sox2 can regulate these two distinct identities. Part of the answer lies with the partnership between Sox2 and cell-type-specific POU factors; the partnership between Sox2 and Oct4 in ESCs is well characterized, and the interaction between Sox2 and the CNS specific POU factors Brn1 and Brn2 is an emerging Sox:POU module which may prove important for NPC biology. Advances in genome-wide analytic technologies have elucidated the pervasiveness of distal enhancers in transcriptional control, and while Sox2 and Oct4 have been shown to bind a multitude of enhancers in ESCs, the global regulation of these elements in NPCs by Sox2, Brn1 and Brn2 is only recently becoming clear. Data presented in the following chapters will resolve key questions regarding the role of Sox2 and its POU factor partners in controlling transcription and developmental transitions.

References

Altman, J. (1962). "Are new neurons formed in the brains of adult mammals?" Science **135**(3509): 1127.

Altman, J. (1963). "Autoradiographic investigation of cell proliferation in the brains of rats and cats." The Anatomical Record **145**(4): 573-591.

Altman, J. (1969). "Autoradiographic and histological studies of postnatal neurogenesis. IV. Cell proliferation and migration in the anterior forebrain, with special reference to persisting neurogenesis in the olfactory bulb." The Journal of Comparative Neurology **137**(4): 433-457.

Ambrosetti, D. C., C. Basilico, et al. (1997). "Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites." Molecular and Cellular Biology **17**(11): 6321.

Ambrosetti, D. C., H. R. Schöler, et al. (2000). "Modulation of the Activity of Multiple Transcriptional Activation Domains by the DNA Binding Domains Mediates the Synergistic Action of Sox2 and Oct-3 on the Fibroblast Growth Factor-4 Enhancer." Journal of Biological Chemistry **275**(30): 23387.

Arnold, K., A. Sarkar, et al. (2011). "Sox2+ Adult Stem and Progenitor Cells Are Important for Tissue Regeneration and Survival of Mice." Cell Stem Cell **9**(4): 317-329.

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

Banerji, J., S. Rusconi, et al. (1981). "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences." Cell **27**(2): 299-308.

Bani-Yaghoub, M., R. G. Tremblay, et al. (2006). "Role of Sox2 in the development of the mouse neocortex." Developmental Biology **295**(1): 52-66.

Bannister, A. J. and T. Kouzarides (2011). "Regulation of chromatin by histone modifications." Cell Research **21**(3): 381-395.

Bass, A. J., H. Watanabe, et al. (2009). "SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas." Nature Genetics **41**(11): 1238-1242.

Basu-Roy, U., D. Ambrosetti, et al. (2010). "The transcription factor Sox2 is required for osteoblast self-renewal." Cell Death & Differentiation **17**(8): 1345-1353.

Beddington, R. and E. Robertson (1989). "An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo." Development **105**(4): 733-737.

Bergsland, M., D. Ramsköld, et al. (2011). "Sequentially acting Sox transcription factors in neural lineage development." Genes & Development **25**(23): 2453-2464.

Blackwood, E. M. and J. T. Kadonaga (1998). "Going the distance: a current view of enhancer action." Science **281**(5373): 60.

Botquin, V., H. Hess, et al. (1998). "New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2." Genes & Development **12**(13): 2073-2090.

Bowles, J., G. Schepers, et al. (2000). "Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators." Developmental Biology **227**(2): 239-255.

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

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

Bradley, A., M. Evans, et al. (1984). "Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines." Nature **309**(5965): 255-256.

Brand, A. H. and N. Perrimon (1993). "Targeted gene expression as a means of altering cell fates and generating dominant phenotypes." Development **118**(2): 401-415.

Bylund, M., E. Andersson, et al. (2003). "Vertebrate neurogenesis is counteracted by Sox1-3 activity." Nature Neuroscience **6**(11): 1162-1168.

Castro DS, Skowronska-Krawczyk D et al. (2006). "Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif." Developmental Cell **11** (6): 834-844.

Catena, R., C. Tiveron, et al. (2004). "Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells." Journal of Biological Chemistry **279**(40): 41846-41857.

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

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

Cimpean, A., S. Encica, et al. (2010). "Expression and potential role of SOX2 gene in human thymus and thymomas." EJC Supplements **8**(5): 174-174.

Collignon, J., S. Sockanathan, et al. (1996). "A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2." Development **122**(2): 509.

Creyghton, M. P., A. W. Cheng, et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proceedings of the National Academy of Sciences **107**(50): 21931.

Curatola, A. and C. Basilico (1990). "Expression of the K-fgf proto-oncogene is controlled by 3'regulatory elements which are specific for embryonal carcinoma cells." Molecular and Cellular Biology **10**(6): 2475-2484.

Dailey, L. and C. Basilico (2001). "Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes." Journal of Cellular Physiology **186**(3): 315-328.

Dailey, L., H. Yuan, et al. (1994). "Interaction between a novel F9-specific factor and octamer-binding proteins is required for cell-type-restricted activity of the fibroblast growth factor 4 enhancer." Molecular and Cellular Biology **14**(12): 7758.

D'Amour, K. A. and F. H. Gage (2003). "Genetic and functional differences between multipotent neural and pluripotent embryonic stem cells." Proceedings of the National Academy of Sciences **100**(Suppl 1): 11866.

Davis, A. A. and S. Temple (1994). "A self-renewing multipotential stem cell in embryonic rat cerebral cortex." Nature **372**(6503): 263-266.

De Santa, F., I. Barozzi, et al. (2010). "A large fraction of extragenic RNA pol II transcription sites overlap enhancers." PLoS Biol **8**(5): e1000384.

Denny, P., S. Swift, et al. (1992). "A conserved family of genes related to the testis determining gene, SRY." Nucleic Acids Research **20**(11): 2887-2887.

Denny, P., S. Swift, et al. (1992). "An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein." The EMBO journal **11**(10): 3705.

Desai, A. R. and S. K. McConnell (2000). "Progressive restriction in fate potential by neural progenitors during cerebral cortical development." Development **127**(13): 2863-2872.

Detrait, E. R., T. M. George, et al. (2005). "Human neural tube defects: developmental biology, epidemiology, and genetics." Neurotoxicol Teratol **27**(3): 515-524.

Doetsch, F., I. Caille, et al. (1999). "Subventricular zone astrocytes are neural stem cells in the adult mammalian brain." Cell **97**(6): 703-716.

Doetschman, T., R. G. Gregg, et al. (1987). "Targetted correction of a mutant HPRT gene in mouse embryonic stem cells." Nature **330**, 576 - 578

Doetschman, T., N. Maeda, et al. (1988). "Targeted mutation of the Hprt gene in mouse embryonic stem cells." Proceedings of the National Academy of Sciences **85**(22): 8583-8587.

Domyan, E. T., E. Ferretti, et al. (2011). "Signaling through BMP receptors promotes respiratory identity in the foregut via repression of Sox2." Development **138**(5): 971-981.

Donner, A. L., V. Episkopou, et al. (2007). "Sox2 and Pou2f1 interact to control lens and olfactory placode development." Developmental Biology **303**(2): 784-799.

Engelen, E., U. Akinci, et al. (2011). "Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes." Nature Genetics **43**(6): 607-611.
Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-156.
Fang, X., J. G. Yoon, et al. (2011). "The SOX2 response program in glioblastoma multiforme: an integrated ChIP-seq, expression microarray, and microRNA analysis." BMC genomics **12**(1): 11.
Fantes, J., N. K. Ragge, et al. (2003). "Mutations in SOX2 cause anophthalmia." Nature Genetics **33**(4): 462-463.
Fasano, C. A., T. N. Phoenix, et al. (2009). "Bmi-1 cooperates with Foxg1 to maintain neural stem cell self-renewal in the forebrain." Genes & Development **23**(5): 561-574.
Favaro, R., M. Valotta, et al. (2009). "Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh." Nature Neuroscience **12**(10): 1248-1256.
Feng, R., S. C. Desbordes, et al. (2008). "PU. 1 and C/EBPα/β convert fibroblasts into macrophage-like cells." Proceedings of the National Academy of Sciences **105**(16): 6057.
Ferri, A. L. M., M. Cavallaro, et al. (2004). "Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain." Development **131**(15): 3805-3819.
Fullwood, M. J., M. H. Liu, et al. (2009). "An oestrogen-receptor-alpha-bound human chromatin interactome." Nature **462**(7269): 58-64.
Gilbert, SF (2000), "Developmental Biology, 6th Edition." Sinauer Associates, Sunderland, MA.
Goldman, S. A. and F. Nottebohm (1983). "Neuronal production, migration, and differentiation in a vocal control nucleus of the adult female canary brain." Proceedings of the National Academy of Sciences **80**(8): 2390.
Gontan, C., A. de Munck, et al. (2008). "Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation." Developmental Biology **317**(1): 296-309.
Grabowski, P. (2011). "Alternative splicing takes shape during neuronal development." Curr Opin Genet Dev. (4): 388-94
Graham, V., J. Khudyakov, et al. (2003). "SOX2 functions to maintain neural progenitor identity." Neuron **39**(5): 749-765.
Griffiths, A.J.F., Miller J.H. et al. (2000). "Introduction to Genetic Analysis." W.H. Freeman. New York.
Gritti, A., E. Parati, et al. (1996). "Multipotential stem cells from the adult mouse brain proliferate and self-renew in response to basic fibroblast growth factor." The Journal of Neuroscience **16**(3): 1091-1100.
Gubbay, J., J. Collignon, et al. (1990). "A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes." Nature **346**(6281): 245-250.
Guichet, A., S. Triau, et al. (2004). "Prenatal diagnosis of primary anophthalmia with a 3q27 interstitial deletion involving SOX2." Prenatal Diagnosis **24**(10): 828-832.
Guillemot, F. (2005). "Cellular and molecular control of neurogenesis in the mammalian telencephalon." Curr Opin Cell Biol **17**(6): 639-647.
Hagstrom, S. A., G. J. T. Pauer, et al. (2005). "SOX2 mutation causes anophthalmia, hearing loss, and brain anomalies." American Journal of Medical Genetics Part A **138**(2): 95-98.
Hara, Y., A. C. Rovescalli, et al. (1992). "Structure and evolution of four POU domain genes expressed in mouse brain." Proceedings of the National Academy of Sciences **89**(8): 3280.
He, H. H., C. A. Meyer, et al. (2010). "Nucleosome dynamics define transcriptional enhancers." Nature Genetics **42**(4): 343-347.
He, X., M. N. Treacy, et al. (1989). "Expression of a large family of POU-domain regulatory genes in mammalian brain development." Nature **340**(6228): 35-41.
Heintzman, N. D., G. C. Hon, et al. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.
Heintzman, N. D., R. K. Stuart, et al. (2007). "Distinct and predictive chromatin signatures of

transcriptional promoters and enhancers in the human genome." Nature Genetics **39**(3): 311-318.
Holmes, G., T. G. Bromage, et al. (2011). "The Sox2 high mobility group transcription factor inhibits mature osteoblast function in transgenic mice." Bone.
Inoue, M., Y. Kamachi, et al. (2007). "PAX6 and SOX2-dependent regulation of the Sox2 enhancer N-3 involved in embryonic visual system development." Genes to Cells **12**(9): 1049-1061.
Johansson, C. B., S. Momma, et al. (1999). "Identification of a neural stem cell in the adult mammalian central nervous system." Cell **96**(1): 25-34.
Kaestner, K. H. (2010). "The FoxA factors in organogenesis and differentiation." Curr Opin Genet Dev **20**(5): 527-532.
Kageyama, R., T. Ohtsuka, et al. (2005). "Roles of bHLH genes in neural stem cell differentiation." Experimental Cell Research **306**(2): 343-348.
Kaiser, K. (1993). "Second generation enhancer traps." Current Biology: CB **3**(8): 560.
Kamachi, Y., K. S. E. Cheah, et al. (1999). "Mechanism of regulatory target selection by the SOX high-mobility-group domain proteins as revealed by comparison of SOX1/2/3 and SOX9." Molecular and Cellular Biology **19**(1): 107.
Kamachi, Y., S. Sockanathan, et al. (1995). "Involvement of SOX proteins in lens-specific activation of crystallin genes." The EMBO journal **14**(14): 3510.
Kamachi, Y., M. Uchikawa, et al. (2000). "Pairing SOX off: with partners in the regulation of embryonic development." Trends in Genetics **16**(4): 182-187.
Kamachi, Y., M. Uchikawa, et al. (2001). "Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development." Genes & Development **15**(10): 1272.
Kiefer, J. C. (2007). "Back to basics: Sox genes." Developmental Dynamics **236**(8): 2356-2366.
Kielman, M. F., M. Rindapää, et al. (2002). "Apc modulates embryonic stem-cell differentiation by controlling the dosage of bold beta-catenin signaling." Nature Genetics **32**: 594-605.
Kiernan, A. E., A. L. Pelling, et al. (2005). "Sox2 is required for sensory organ development in the mammalian inner ear." Nature **434**(7036): 1031-1035.
Kim, J., J. Chu, et al. (2008). "An extended transcriptional network for pluripotency of embryonic stem cells." Cell **132**(6): 1049-1061.
Kim, T. K., M. Hemberg, et al. (2010). "Widespread transcription at neuronal activity-regulated enhancers." Nature **465**(7295): 182-187.
Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.
Kuhlbrodt, K., B. Herbarth, et al. (1998). "Cooperative function of POU proteins and SOX proteins in glial cells." Journal of Biological Chemistry **273**(26): 16050.
Kunath, T., M. K. Saba-El-Leil, et al. (2007). "FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment." Development **134**(16): 2895-2902.
Kuroda, T., M. Tada, et al. (2005). "Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression." Molecular and Cellular Biology **25**(6): 2475.
Lai, E. C. (2004). "Notch signaling: control of cell communication and cell fate." Development **131**(5): 965-973.
Lassar, A. B., B. M. Paterson, et al. (1986). "Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts." Cell **47**(5): 649-656.
Le, N., R. Nagarajan, et al. (2005). "Analysis of congenital hypomyelinating Egr2Lo/Lo nerves identifies Sox2 as an inhibitor of Schwann cell differentiation and myelination." Proceedings of the National Academy of Sciences **102**(7): 2596.
LeBowitz, J., T. Kobayashi, et al. (1988). "Octamer-binding proteins from B or HeLa cells stimulate transcription of the immunoglobulin heavy-chain promoter in vitro." Genes & Development **2**(10): 1227.
Lee, T. I., R. G. Jenner, et al. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-313.
Lee, T. I. and R. A. Young (2000). "Transcription of eukaryotic protein-coding genes." Annual Review of

Genetics **34**(1): 77-137.
Lefebvre, V., B. Dumitriu, et al. (2007). "Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors." The International Journal of Biochemistry & Cell Biology **39**(12): 2195-2214.
Lewin, B. and G. Dover (2008). Genes IX, Oxford University Press New York.
Li, J., G. Pan, et al. (2007). "A dominant-negative form of mouse SOX2 induces trophectoderm differentiation and progressive polyploidy in mouse embryonic stem cells." Journal of Biological Chemistry **282**(27): 19481.
Liber, D., R. Domaschenz, et al. (2010). "Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage." Cell Stem Cell **7**(1): 114-126.
Lie, D. C., S. A. Colamarino, et al. (2005). "Wnt signalling regulates adult hippocampal neurogenesis." Nature **437**(7063): 1370-1375.
Lim, D. A., Y. C. Huang, et al. (2009). "Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells." Nature **458**(7237): 529-533.
Lim, D. A., A. D. Tramontin, et al. (2000). "Noggin antagonizes BMP signaling to create a niche for adult neurogenesis." Neuron **28**(3): 713-726.
Loh, Y. H., Q. Wu, et al. (2006). "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells." Nature Genetics **38**(4): 431-440.
Lois, C. and A. Alvarez-Buylla (1993). "Proliferating subventricular zone cells in the adult mammalian forebrain can differentiate into neurons and glia." Proceedings of the National Academy of Sciences **90**(5): 2074.
Lujan, E., S. Chanda, et al. (2012). "Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells." Proceedings of the National Academy of Sciences **109**(7): 2527-2532.
Machold, R., S. Hayashi, et al. (2003). "Sonic hedgehog is required for progenitor cell maintenance in telencephalic stem cell niches." Neuron **39**(6): 937-950.
Mansukhani, A., D. Ambrosetti, et al. (2005). "Sox2 induction by FGF and FGFR2 activating mutations inhibits Wnt signaling and osteoblast differentiation." The Journal of Cell Biology **168**(7): 1065-1076.
Marson, A., S. S. Levine, et al. (2008). "Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells." Cell **134**(3): 521-533.
Martin, G. R. (1981). "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." Proceedings of the National Academy of Sciences **78**(12): 7634.
Mason, J. O., G. T. Williams, et al. (1985). "Transcription cell type specificity is conferred by an immunoglobulin VH gene promoter that includes a functional consensus sequence." Cell **41**(2): 479-487.
Masui, S., Y. Nakatake, et al. (2007). "Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells." Nature Cell Biology **9**(6): 625-635.
McEvilly, R. J., M. O. de Diaz, et al. (2002). "Transcriptional regulation of cortical neuron migration by POU domain factors." Science **295**(5559): 1528.
Mehler, M. F., P. C. Mabie, et al. (2000). "Developmental changes in progenitor cell responsiveness to bone morphogenetic proteins differentially modulate progressive CNS lineage fate." Developmental Neuroscience **22**(1-2): 74-85.
Miyagi, S., S. Masui, et al. (2008). "Consequence of the loss of Sox2 in the developing brain of the mouse." FEBS letters **582**(18): 2811-2815.
Miyagi, S., M. Nishimoto, et al. (2006). "The Sox2 regulatory region 2 functions as a neural stem cell-specific enhancer in the telencephalon." Journal of Biological Chemistry **281**(19): 13374.
Molofsky, A. V., R. Pardal, et al. (2003). "Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation." Nature **425**(6961): 962.
Mountford, J. (2008). "Human embryonic stem cells: origins, characteristics and potential for regenerative therapy." Transfusion Medicine **18**(1): 1-12.

Mueller-Storm, H. P., J. M. Sogo, et al. (1989). "An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge." Cell **58**(4): 767-777.
Muers, M. (2011). "Chromatin: A haul of new histone modifications." Nature Reviews Genetics **12**(11): 744-744.
MÜLLER, M. M., T. GERSTER, et al. (1988). "Enhancer sequences and the regulation of gene transcription." European Journal of Biochemistry **176**(3): 485-495.
Nagy, A., E. Gocza, et al. (1990). "Embryonic stem cells alone are able to support fetal development in the mouse." Development **110**(3): 815-821.
Nakai, S., H. Kawano, et al. (1995). "The POU domain transcription factor Brn-2 is required for the determination of specific neuronal lineages in the hypothalamus of the mouse." Genes & Development **9**(24): 3109-3121.
Nakai, S., Y. Sugitani, et al. (2003). "Crucial roles of Brn1 in distal tubule formation and function in mouse kidney." Development **130**(19): 4751-4759.
Nakatake, Y., N. Fukui, et al. (2006). "Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells." Molecular and Cellular Biology **26**(20): 7772.
Nasrin, N., C. Buggs, et al. (1991). "DNA-binding properties of the product of the testis-determining gene and a related protein." Nature **354**(6351): 317-620.
Nishimoto, M., S. Miyagi, et al. (2005). "Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus." Molecular and Cellular Biology **25**(12): 5084.
Niwa, H., T. Burdon, et al. (1998). "Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3." Genes & Development **12**(13): 2048.
Niwa, H., J. Miyazaki, et al. (2000). "Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells." Nature Genetics **24**(4): 372-376.
Niwa, H., K. Ogawa, et al. (2009). "A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells." Nature **460**(7251): 118-122.
Nowling, T. K., L. R. Johnson, et al. (2000). "Identification of the transactivation domain of the transcription factor Sox-2 and an associated co-activator." Journal of Biological Chemistry **275**(6): 3810.
Ogawa, K., R. Nishinakamura, et al. (2006). "Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells." Biochemical and Biophysical Research Communications **343**(1): 159-166.
Ohtsuka, T., M. Ishibashi, et al. (1999). "Hes1 and Hes5 as notch effectors in mammalian neuronal differentiation." The EMBO journal **18**(8): 2196-2207.
Ohtsuka, T., M. Sakamoto, et al. (2001). "Roles of the basic helix-loop-helix genes Hes1 and Hes5 in expansion of neural stem cells of the developing brain." Journal of Biological Chemistry **276**(32): 30467-30474.
O'Kane, C. J. and W. J. Gehring (1987). "Detection in situ of genomic regulatory elements in Drosophila." Proceedings of the National Academy of Sciences **84**(24): 9123.
Okubo, T., L. H. Pevny, et al. (2006). "Sox2 is required for development of taste bud sensory cells." Genes & Development **20**(19): 2654.
Okumura-Nakanishi, S., M. Saito, et al. (2005). "Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells." Journal of Biological Chemistry **280**(7): 5307.
Palmer, T. D., A. R. Willhoite, et al. (2000). "Vascular niche for adult hippocampal neurogenesis." The Journal of Comparative Neurology **425**(4): 479-494.
Pan, H. and R. M. Schultz (2011). "SOX2 Modulates Reprogramming of Gene Expression in Two-Cell Mouse Embryos." Biology of Reproduction.
Paton, J. A. and F. N. Nottebohm (1984). "Neurons generated in the adult brain are recruited into functional circuits." Science **225**(4666): 1046.
Pennacchio, L. A., N. Ahituv, et al. (2006). "In vivo enhancer analysis of human conserved non-coding sequences." Nature **444**(7118): 499-502.

Phillips, K. and B. Luisi (2000). "The virtuoso of versatility: POU proteins that flex to fit1." Journal of Molecular Biology **302**(5): 1023-1039.
Qian, X., Q. Shen, et al. (2000). "Timing of CNS Cell Generation:: A Programmed Sequence of Neuron and Glial Cell Production from Isolated Murine Cortical Stem Cells." Neuron **28**(1): 69-80.
Que, J., X. Luo, et al. (2009). "Multiple roles for Sox2 in the developing and adult mouse trachea." Development **136**(11): 1899-1907.
Rada-Iglesias, A., R. Bajpai, et al. (2010). "A unique chromatin signature uncovers early developmental enhancers in humans." Nature **470**(7333): 279-283.
Ragge, N. K., B. Lorenz, et al. (2005). "SOX2 anophthalmia syndrome." American Journal of Medical Genetics Part A **135**(1): 1-7.
Rahl, P. B., C. Y. Lin, et al. (2010). "c-Myc regulates transcriptional pause release." Cell **141**(3): 432-445.
Reiprich, S., J. Kriesch, et al. (2010). "Activation of Krox20 gene expression by Sox10 in myelinating Schwann cells." Journal of Neurochemistry **112**(3): 744-754.
Reményi, A., K. Lins, et al. (2003). "Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers." Genes & Development **17**(16): 2048-2059.
Reményi, A., H. R. Scholer, et al. (2004). "Combinatorial control of gene expression." Nature Structural & Molecular Biology **11**(9): 812-815.
Reynolds, B. A. and S. Weiss (1992). "Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system." Science **255**(5052): 1707.
Reynolds, B. A. and S. Weiss (1996). "Clonal and population analyses demonstrate that an EGF-responsive mammalian embryonic CNS precursor is a stem cell." Developmental Biology **175**(1): 1-13.
Rodda, D. J., J. L. Chew, et al. (2005). "Transcriptional regulation of Nanog by OCT4 and SOX2." Journal of Biological Chemistry **280**(26): 24731-24737.
Ryan, A. K. and M. G. Rosenfeld (1997). "POU domain family values: flexibility, partnerships, and developmental codes." Genes & Development **11**(10): 1207.
Sakabe, N. J., D. Savic, et al. (2012). "Transcriptional enhancers in development and disease." Genome Biology **13**(1): 238.
Scaffidi, P. and M. E. Bianchi (2001). "Spatially precise DNA bending is an essential activity of the sox2 transcription factor." Journal of Biological Chemistry **276**(50): 47296.
Schones, D. E., K. Cui, et al. (2008). "Dynamic regulation of nucleosome positioning in the human genome." Cell **132**(5): 887-898.
Seale, P., B. Bjork, et al. (2008). "PRDM16 controls a brown fat/skeletal muscle switch." Nature **454**(7207): 961-967.
Shen, Q., S. K. Goderie, et al. (2004). "Endothelial cells stimulate self-renewal and expand neurogenesis of neural stem cells." Science **304**(5675): 1338.
Shen, Q., Y. Wang, et al. (2006). "The timing of cortical neurogenesis is encoded within lineages of individual progenitor cells." Nature Neuroscience **9**(6): 743-751.
Shimada, N., T. Aya-Murata, et al. (2003). "Cooperative action between L-Maf and Sox2 on delta-crystallin gene expression during chick lens development." Mechanisms of Development **120**(4): 455-465.
Sinclair, A., P. Berta, et al. (1990). "A gene from the human sex-determining region encodes a protein with." Nature **346**(6281): 240-244.
Smale, S. T. (2010). "Pioneer factors in embryonic stem cells and differentiation." Curr Opin Genet Dev **20**(5): 519-526.
Smith, A. G., J. K. Heath, et al. (1988). "Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides." Nature **336**(6200): 688-690.
Soriano, N. S. and S. Russell (1998). "The Drosophila SOX-domain protein Dichaete is required for the development of the central nervous system midline." Development **125**(20): 3989.

Stanfield, B. B. and W. M. Cowan (1979). "The development of the hippocampus and dentate gyrus in normal and reeler mice." J Comp Neurol **185**(3): 423-459.
Staudt, L. M., R. G. Clerc, et al. (1988). "Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif." Science **241**(4865): 577.
Staudt, L. M., H. Singh, et al. (1986). "A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes." Nature **323**(6089): 640-643.
Sturm, R. A. and W. Herr (1988). "The POU domain is a bipartite DNA-binding structure." Nature **336** (6199): 601-644.
Sugitani, Y., S. Nakai, et al. (2002). "Brn-1 and Brn-2 share crucial roles in the production and positioning of mouse neocortical neurons." Genes & Development **16**(14): 1760.
Suh, H., A. Consiglio, et al. (2007). "In vivo fate analysis reveals the multipotent and self-renewal capacities of Sox2+ neural stem cells in the adult hippocampus." Cell Stem Cell **1**(5): 515-528.
Sun, Y., M. Nadal-Vicens, et al. (2001). "Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms." Cell **104**(3): 365-376.
Szalkowski, A.M. and Schmid C.D. (2011) "Rapid Innovation in ChIP-Seq peak-calling algorithms is outdistancing benchmarking efforts." Briefings in Bioinformatics **12** (6) 626-633.
Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.
Tan, M., H. Luo, et al. (2011). "Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification." Cell **146**(6): 1016-1028.
Tanaka, S., Y. Kamachi, et al. (2004). "Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells." Molecular and Cellular Biology **24**(20): 8834.
Thomas, J. O. and A. A. Travers (2001). "HMG1 and 2, and related 'architectural' DNA-binding proteins." Trends in Biochemical Sciences **26**(3): 167-174.
Thomas, K. R. and M. R. Capecchi (1987). "Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells." Cell **51**(3): 503-512.
Thomson, J. A., J. Itskovitz-Eldor, et al. (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145.
Tomioka, M., M. Nishimoto, et al. (2002). "Identification of Sox-2 regulatory region which is under the control of Oct-3/4–Sox-2 complex." Nucleic Acids Research **30**(14): 3202-3213.
Tompkins, D. H., V. Besnard, et al. (2009). "Sox2 is required for maintenance and differentiation of bronchiolar Clara, ciliated, and goblet cells." PLoS One **4**(12): e8248.
Treacy, M. N. and M. G. Rosenfeld (1992). "Expression of a family of POU-domain protein regulatory genes during development of the central nervous system." Annual Review of Neuroscience **15**(1): 139-165.
Tropepe, V., M. Sibilia, et al. (1999). "Distinct neural stem cells proliferate in response to EGF and FGF in the developing mouse telencephalon." Developmental Biology **208**(1): 166-188.
Visel, A., M. J. Blow, et al. (2009). "ChIP-seq accurately predicts tissue-specific activity of enhancers." Nature **457**(7231): 854-858.
Vural, B., L. C. Chen, et al. (2005). "Frequency of SOX Group B (SOX1, 2, 3) and ZIC2 antibodies in Turkish patients with small cell lung carcinoma and their correlation with clinical parameters." Cancer **103**(12): 2575-2583.
Wang, J., K. Nishiyama, et al. (1987). "Purification of an octamer sequence (ATGCAAAT)-binding protein from human B cells." Nucleic Acids Research **15**(24): 10105.
Wang, Z. X., C. H. L. Teh, et al. (2007). "Oct4 and Sox2 directly regulate expression of another pluripotency transcription factor, Zfp206, in embryonic stem cells." Journal of Biological Chemistry **282**(17): 12822.
Wathelet, M. G., C. H. Lin, et al. (1998). "Virus infection induces the assembly of coordinately activated transcription factors on the IFN-beta enhancer in vivo." Molecular Cell **1**(4): 507-518.
Williams, D. C., M. Cai, et al. (2004). "Molecular Basis for Synergistic Transcriptional Activation by Oct1 and Sox2 Revealed from the Solution Structure of the 42-kDa Oct1· Sox2· Hoxb1-DNA

Ternary Transcription Factor Complex." Journal of Biological Chemistry **279**(2): 1449.
Williams, R. L., D. J. Hilton, et al. (1988). "Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells." Nature **336**(6200): 684-687.
Williamson, K. A., A. M. Hever, et al. (2006). "Mutations in SOX2 cause anophthalmia-esophageal-genital (AEG) syndrome." Human Molecular Genetics **15**(9): 1413.
Wirth, T., L. Staudt, et al. (1987). "An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid-specific promoter activity." Nature **329**(6135): 174-178.
Wood, H. B. and V. Episkopou (1999). "Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages." Mechanisms of Development **86**(1-2): 197-201.
Wright, E. M., B. Snopek, et al. (1993). "Seven new members of the Sox gene family expressed during mouse development." Nucleic Acids Research **21**(3): 744.
Xie, H., M. Ye, et al. (2004). "Stepwise reprogramming of B cells into macrophages." Cell **117**(5): 663-676.
Ying, Q. L., J. Nichols, et al. (2003). "BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3." Cell **115**(3): 281-292.
Ying, Q. L., J. Wray, et al. (2008). "The ground state of embryonic stem cell self-renewal." Nature **453**(7194): 519-523.
Yoon, K. and N. Gaiano (2005). "Notch signaling in the mammalian central nervous system: insights from mouse mutants." Nature Neuroscience **8**(6): 709-715.
Young, R. A. (2011). "Control of the embryonic stem cell state." Cell **144**(6): 940-954.
Yuan, H., N. Corbi, et al. (1995). "Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3." Genes & Development **9**(21): 2635.
Zappone, M. V., R. Galli, et al. (2000). "Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells." Development **127**(11): 2367.
Zeng, F. and R. M. Schultz (2005). "RNA transcript profiling during zygotic gene activation in the preimplantation mouse embryo." Developmental Biology **283**(1): 40-57.
Zhao, H. and A. Dean (2005). "Organizing the genome: enhancers and insulators." Biochem Cell Biol **83**(4): 516-524.
Zhou, C. J., C. Zhao, et al. (2004). "Wnt signaling mutants have decreased dentate granule cell production and radial glial scaffolding abnormalities." The Journal of Neuroscience **24**(1): 121-126.
Zhou, Q., J. Brown, et al. (2008). "In vivo reprogramming of adult pancreatic exocrine cells to beta-cells." Nature **455**(7213): 627-632.

Chapter 2: Sox2 co-occupies distal enhancer elements with distinct POU factors to specify cell identity

Michael A. Lodato[1,2], Chris Ng[3], Joseph A. Wamstad[1], Albert W. Cheng[1,2], Kevin K. Thai[1], Ernest Fraenkel[3], Laurie A. Boyer[1], Rudolf Jaenisch[1,2*]

[1]Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue Cambridge, MA 02139 USA; [2]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142 USA; [3]Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139 USA.

Chapter Synopsis

Sox2 is a master regulator of several stem cell compartments, including pluripotent embryonic stem cells (ESCs) and multipotent neuronal progenitor cells (NPCs), yet how Sox2 controls distinct stem cell identities is not fully understood. Here we show by genome-wide analysis that while Sox2 occupied promoter regions in ESCs and NPCs with cell-type-specific functions, the majority of Sox2-bound regions mapped to a distinct set of distal enhancer elements linked to loci encoding developmental regulators. Notably, we found that Sox2 co-occupied a large subset of enhancers with the central-nervous-system expressed POU factors Brn1 and Brn2 in NPCs, similarly to its interaction with the POU family member Oct4 in ESCs. Consistent with a role in neural specification, forced expression of Brn2 in ESCs leads to up-regulation of neural markers and premature differentiation. Further analysis revealed that Sox2:Brn1 and Sox2:Brn2 sites were distinguished by a SOX-POU composite motif in a distinct configuration compared to Sox2:Oct4 regions, suggesting that particular Sox2-POU combinations bind DNA in different conformations. As Sox2 is a key factor in lineage specification, somatic cell reprogramming, and direct transdifferentiation, our findings have important implications for understanding these processes.

Author Contributions

M.A.L., L.A.B., and R.J. conceived experiments and interpreted data. M.A.L. and A.W.C. analyzed microarray data. M.A.L and C.N. performed motif spacing and peak distribution analyses. M.A.L and J.A.W. edited figures M.A.L. and K.T. performed Sox2 ESC ChIP. M.A.L performed all other experiments and analyzed data. M.A.L. and L.A.B. wrote the chapter.

Introduction

Transcription factors bind to DNA in a sequence-specific manner and regulate which genes in the genome will be expressed in a given cell type, and thus are thought to sit at the top of a hierarchy controlling cellular identity. In embryonic stem cells (ESCs), master regulatory transcription factors, such as the HMG-box-containing factor Sox2, control both the ESC cell state and set the stage for later developmental transitions. Given this crucial role, it is not surprising that Sox2 is essential for the development of a functional inner cell mass (ICM) in the mammalian embryo (Avillion et al., 2003) and for maintenance of ESCs *in vitro* (Miyagi et al., 2006). Sox2 is also necessary for the proper function of neural precursor cells (NPCs) in the developing nervous system and functions in multiple adult stem cell populations in the gastrointestinal tract, the respiratory tract, the testes, and many other tissues (Collignon et al., 1996; Kamachi 1995; Cimpean et al., 2001; Mansukhani et al., 2005; Kiernan et al., 2005; Okubo et al., 2006; Gontan et al., 2008; Domyan et al., 2011; Arnold et al., 2011). Thus, Sox2 is a critical regulator of distinct stem cell states, but how it can serve this multi-functional role is not fully understood.

In ESCs, Sox2 is part of the core transcriptional regulatory network that controls pluripotency and self-renewal. Together with Oct4 and Nanog, Sox2 binds to the proximal promoter regions of many genes with known roles in pluripotency (including the promoters of Oct4, Sox2, and Nanog), as well as developmental regulators that are not expressed in ESCs but maintain the potential for activation upon differentiation (Boyer et al., 2005; Loh et al., 2006; Marson et al., 2008; Chen et al., 2008). These data suggested that Sox2 regulates pluripotency in ESCs by actively promoting pluripotency and by marking the regulatory regions of developmental genes for future activation. A powerful demonstration of the importance of Sox2 in specifying the ESC identity is the role it plays in factor mediated reprogramming, by which differentiated adult cells are converted into an pluripotent state by the exogenous expression of three to four transcription factors, critically including Oct4 and Sox2 (Takahashi et al., 2006, Wernig et al., 2007; Takahashi et al., 2007).

In the central nervous system, Sox2 regulates development by controlling cell identity in NPCs. Sox2 expression marks NPCs, proliferating cells which have the ability to self-renew and differentiate into multiple neural lineages, in the developing and adult brain (D'Amour and Gage, 2003; Ferri et al., 200; Suh et al., 2007). Consistent with this, Sox2 loss-of-function in the developing central nervous system leads to multiple brain defects, including precocious progenitor differentiation, premature lethality and a reduced proliferating cell population in the brain (Graham et al., 2003; Bylund et al., 2003; Miyagi et al., 2008, Favaro et al., 2009). Furthermore, forced expression of Sox2 has been shown to block terminal differentiation of NPCs (Graham et al., 2003; Bylund et al., 2003; Le et al., 2005; Bani-Yaghoub, 2006). Thus, Sox2 is necessary and sufficient for maintenance of the stem cell state of NPCs in the brain.

While the genetic role of Sox2 in the pregastrulation embryo and in the development of the CNS at midgestation is well understood, the molecular mechanisms by which Sox2 can regulate global gene expression networks in a context-specific manner is only recently becoming clear, and much remains to be learned. In ESCs, Sox2 regulates active genes but also functions as a pioneer factor at repressed developmental regulators, facilitating a transcriptionally competent at these loci. For example, in ESCs Sox2 and FoxD3 were shown to co-occupy a B-cell specific enhancer of the λ5-VpreB1 locus and maintain this element as inactive until commitment to the B-cell lineage, where Sox2 and FoxD3 were replaced by Sox4 and expression of the locus was activated (Liber et al., 2010). Further, many enhancers linked to genes involved in neural development and bound by Sox2 in ESCs were recently shown to be bound by other Sox family members, namely Sox3 and Sox11, as development progressed from pluripotent ESCs through multipotent NPCs (Sox3) and finally to immature neurons (Sox11) (Bergsland et al., 2011). Notably, a large overlap was observed between Sox2 and Sox3 bound regions in this dataset, but a direct comparison of Sox2 target loci in ESCs and NPCs was not reported. Thus, one facet of Sox2 function genome-wide seems to be the priming of distal enhancers for activation later in development by other Sox factors. This is a very critical role, as recently it has become apparent that

distal enhancers are regulated in a more dynamic way than proximal promoters over developmental time, and thus may play a larger role than promoter regions in dictating cellular identity. For example, monomethylation of Lysine 4 of Histone H3 (H3K4me1) and acetylation of Lysine 27 of Histone H3 (H3K27Ac) are two chromatin marks associated with distal enhancers, and the genome-wide deposition of these marks varies more between cell types than that of trimethylation of Lysine 4 of Histone H3 (H3K4me3), which is known to mark transcriptional start sites (Heintzman et al., 2007, 2009; Creyghton et al., 2010; Rada-Iglesias et al., 2011). Thus, the regulation of these elements by pioneer transcription factors such as Sox2 may set the stage for all subsequent developmental decisions. While previous genome-wide studies regarding Sox2 target regions have contributed to our understanding of Sox2 function by revealing a subset of binding sites in the two cell types, it is likely that transcription factors that cooperate with Sox2 in a context dependent manner will also influence its function during lineage commitment. These factors may mediate target selection by Sox2 in a given cellular context, or they may modulate the readout of Sox2 binding by the transcriptional apparatus. Thus, functional interactions with cell-type specific partner factors may be a critical determinant of Sox2 function across development.

These functional interactions between transcription factors are critical for developmental decisions, yet we lack a clear understanding of the underlying molecular mechanisms controlling these interactions. Sox family members are thought to weakly bind DNA and to cooperate with partner factors to stabilize interactions on genomic targets (Kamachi et al., 2000; Wegner et al., 2010). Interactions between members of the POU and Sox transcription factor families have been highly conserved during evolution and are important for specifying gene expression programs during early embryonic development in metazoans (Daily and Basilico 2001). In ESCs, Sox2 and Oct4 have been shown to cooperatively bind DNA targets in distinct conformations to form stable ternary complexes (Yuan et al., 1995; Nishimoto et al., 1999; Remenyi et al, 2003). These distinct binding conformations have been suggested to confer functional specificity to Sox:POU binding by providing allosterically diverse interfaces for interactions with downstream effectors. While the Sox2:Oct4 partnership in ESCs is well

characterized, the transcription factors which partner with Sox2 at the genomic level in NPCs are not known. Recently, Sox2 has been shown to co-occupy target regions with the ATP-dependent chromatin remodeler Chd7 in NPCs (Engelen et al., 2011), but it is not known whether an interaction with a chromatin modifier such as Chd7 could stabilize ternary complex formation between it, Sox2, and a DNA target in the same way transcription factors such as Oct4 have been shown to do. Thus, elucidating transcription factors that specifically co-operate with Sox2 in ESCs and NPCs will be key to understanding how Sox2 can function to specify these distinct stem cell states.

Here we show that Sox2 occupied a distinct set of promoter regions in ESCs and NPCs encoding both lineage regulators and more widely expressed metabolic genes. However, the majority of enriched regions map to distal enhancers. We noted a very limited overlap in Sox2 occupied regions; the vast majority of Sox2 target regions were cell-type specific. We confirmed that Oct4 occupied most Sox2 regions in ESCs and found that Sox2 and the POU transcription factors Brn1 and Brn2 co-occupied many distal enhancers in NPCs. These POU factors are co-expressed with Sox2 in neurogenic regions of the brain, and loss-of-function of Brn1 or Brn2 causes pleiotropic brain defects, including defects in NPC function (He et al., 1989; Hara et al., 1992; Nakai et al., 1995; McEvilly et al., 2002; Sugitani et al., 2002). Our data provide evidence that Sox2 binding at enhancers can dictate lineage-specific gene expression programs, in part, by partnering with tissue-specific POU factors. Consistent with this, forced expression of Brn2 in differentiating ESCs promotes their commitment towards the neural fate. Further analysis revealed that Sox2:Brn1 and Sox2:Brn2 targets harbor a distinct set of Sox:POU composite motifs compared to Sox2:Oct4 targets. Together, these data reveal new insights into how Sox2 can function in a context-dependent manner to specify distinct stem cell populations during development.

Results

Sox2 occupies distinct regions in ESCs and NPCs

Sox2 is a master regulator of pluripotent ESCs and multipotent NPCs, yet how the same transcription factor can specify these distinct stem cell states in a context specific fashion has not been

clear. We reasoned that detailed analysis of Sox2 genomic binding patterns and sequence context might reveal how Sox2 can regulate distinct gene expression programs in ESCs and NPCs. To this end, we differentiated ESCs toward NPCs using established protocols (Okabe et al., 1996), and then interrogated Sox2 binding sites in each cell type by chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-Seq). ChIP-Seq analysis on Sox2 in ESCs and NPCs identified 13,717 and 16,685 enriched regions, respectively (Methods). Biological replicates displayed high concordance and our results are consistent with prior data sets (Marson et al., 2008) (Figure S1) indicating that we have generated high quality data sets.

We assigned Sox2 ChIP-enriched regions to genomic features to identify bound gene promoters and regulatory elements in ESCs and NPCs. We identified over 29,000 Sox2-enriched regions at high confidence between both cell types and found that while a subset of these regions mapped to TSSs, the majority mapped to intragenic and intergenic regions (Figure 1a). Notably, we find that Sox2 occupied a largely non-overlapping set of targets in ESCs and NPCs; using a stringent threshold >95% of Sox2 targets are bound uniquely in one cell type or another (1,274 of the 29,129 union regions common to both datasets) (Figure 1b). For example, Sox2 was associated with the Nanog promoter in ESCs but absent from this locus in NPCs (Figure 1c), while Sox2 was bound to the locus encoding the Notch ligand Dll1 in NPCs but not in ESCs (Figure 1d). Importantly, Nanog and Dll1 are critical regulators of the ESC and NPC state, respectively (Mitsui et al., 2003; Chambers et al., 2003; Rocha et al., 2009; Ramos et al., 2010). Like Nanog and Dll1, as a group genes bound by Sox2 tend to be highly expressed (Figures 1e, 1g). In ESCs Sox2 promoter targets code for epigenetic and other transcriptional regulators, including core histone genes (Figure 1f). In NPCs Sox2 also occupied the promoters epigenetic and transcriptional regulators, but also genes involved in other cellular metabolic functions such as the regulation of the Ubiquitin cycle and regulation of translation (Figure 1h, not shown). A large group of these genes encode factors involved in RNA processing, specifically regulators of RNA splicing (Figure 1h). While splicing is also a general cellular function, alternative splicing is known to play a key role in the development of

the brain (Grabowski, 2011), and many factors involved in alternative splicing are Sox2 target in NPCs. Thus, Sox2 is bound to distinct sets of active promoters involved in cell identity in both cell types.

The fact that the majority of Sox2 binding in the genome occurred distal to promoters indicated that Sox2 may be controlling larger network of genes than those it bound near the TSS. Recently, much attention has been on distal enhancers and the role these elements play in controlling gene expression. Work from our lab and others has shown that distal enhancers are enriched for nucleosomes marked by monomethylation on Lysine 4 of Histone H3 (H3K4me1) and/or acetylation of Lysine 27 on Histone H3 (H3K27Ac) (Heintzman et al., 2007, 2009; Creyghton et al., 2010; Rada-Iglesias et al., 2011). These marks are distributed in the genome is a highly cell type specific manner, and genes closest these marked regions are highly expressed. In order to ascertain whether Sox2 binds near these elements, we surveyed 1 kb up and downstream of peaks of H3K4me1 and H3K27Ac enrichment for Sox2-bound regions in ESCs and NPCs. Indeed, we find that many Sox2-enriched regions that could not be assigned to promoters were highly correlated with enhancer chromatin marks in both ESCs and NPCs. Further, we find that Sox2 binds to more regions modified with H3K4me1 and/or H3K27Ac than annotated start sites (Figure 1i). Thus Sox2 occupied two distinct sets of active promoters and distal enhancers in ESCs and NPCs.

Figure 1

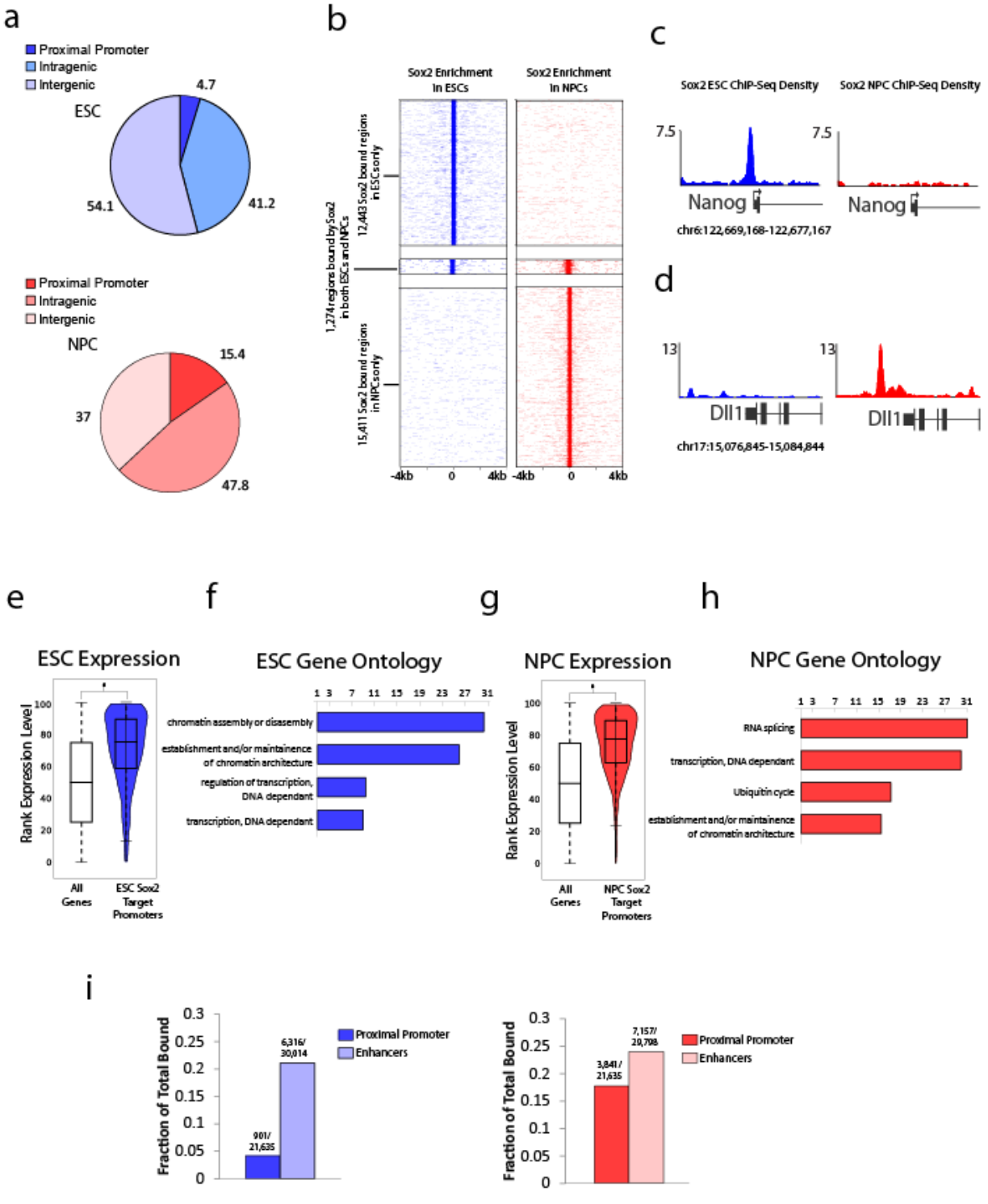

a
Proximal Promoter
Intragenic
Intergenic
ESC
4.7
54.1
41.2
Proximal Promoter
Intragenic
Intergenic
NPC
15.4
37
47.8
b
Sox2 Enrichment in ESCs
Sox2 Enrichment in NPCs
12,443 Sox2 bound regions in ESCs only
1,274 regions bound by Sox2 in both ESCs and NPCs
15,411 Sox2 bound regions in NPCs only
-4kb
0
4kb
c
Sox2 ESC ChIP-Seq Density
Sox2 NPC ChIP-Seq Density
7.5
Nanog
chr6:122,669,168-122,677,167
d
13
Dll1
chr17:15,076,845-15,084,844
e
ESC Expression
Rank Expression Level
All Genes
ESC Sox2 Target Promoters
f
ESC Gene Ontology
chromatin assembly or disassembly
establishment and/or maintainence of chromatin architecture
regulation of transcription, DNA dependant
transcription, DNA dependant
g
NPC Expression
Rank Expression Level
All Genes
NPC Sox2 Target Promoters
h
NPC Gene Ontology
RNA splicing
transcription, DNA dependant
Ubiquitin cycle
establishment and/or maintainence of chromatin architecture
i
Fraction of Total Bound
Proximal Promoter
Enhancers
901/ 21,635
6,316/ 30,014
3,841/ 21,635
7,157/ 29,798

Figure 1 - Sox2 targets distinct sets of promoters and enhancers in ESCs and NPCs. (a) Location analysis breakdown of Sox2 peaks relative to annotated transcriptional start sites. Numbers on pie chart indicate fraction of bound regions which fall into each category. **(b)** Heat maps of Sox2 enrichment in ESCs and NPCs centered on peaks of enrichment and extended 4kb in each direction. **(c,d)** Gene plots showing Sox2 density at *Nanog* and *Dll1* in ESCs (left, blue) and in NPCs (right, red). y-axis corresponds to reads per million. Scale bar corresponds to 1 kb. Genomic positions reflect NCBI Mouse Genome Build 36 (mm8). **(e, g)** Box and Violin plots representing expression data from Affymetrix arrays of genes linked to Sox2 target TSSs. y-axis corresponds to percentile expression rank, * = p-value < 0.01, Student's T-test. **(f,h)** Gene ontology analysis of genes linked to Sox2 target TSSs. **(i)** Fraction of total start sites or total marked enhancers associated with Sox2 enriched region within 1 kb in ESCs and NPCs. Numbers above bars reflect absolute numbers of bound regions and genomic features.

Brn1 and Brn2 co-occupy enhancer regions with Sox2 in NPCs

Our observation that Sox2 bound to distinct sets of target loci in ESCs and NPCs suggests that Sox2 may utilize a distinct mechanism to bind to genomic targets in these cells. We performed an unbiased motif analysis of Sox2 targets and identified sequences highly similar to the canonical motif for Sox2 binding, 5'-CTTTGTT-3' (Yuan et al., 1995; Collignon et al., 1996, Kamachi et al., 1999; Salmon-Divon M. et al., 2010), enriched in ESC and NPC Sox2 targets, indicating that Sox2 uses the same DNA recognition sequence to bind targets in both cells (Figure 2a). This suggested that Sox2 may collaborate with other cellular factors to recognize tissue-specific genomic targets.

While Sox2 functions with the POU-domain-containing transcription factor Oct4 in ESCs, Oct4 is not expressed in NPCs. In order to define a set of possible partner factors for Sox2 in NPCs, we analyzed transcription factor binding motifs near peaks of Sox2 enrichment in in these cells. We determined enrichment of all annotated transcription factor binding motifs in the TRANSFAC database and found recognition motifs for many transcription factors to be enriched. Notably, we identified a motif recognized by the POU factor Brn2 as highly enriched (Figure 2b). The presence of this motif caught our attention for several reasons. First, the interaction between Sox and POU factors is a conserved partnership in all metazoans (Dailey and Basilico, 2001). Second, our transcriptome data analysis identified Brn2 (encoded by Pou3f2) and its paralog Brn1 (encoded by Pou3f3) as POU factors that were highly expressed in NPCs but not in ESCs (Figure S2). Third, Brn2 and Brn1 are co-expressed with Sox2

in NPCs and loss-of-function of these factors is detrimental to NPC function (He et al., 1989; Hara et al., 1992; McEvilly et al., 2002; Sugitani et al., 2002). Additionally, the combination of Sox2, Brn2, and the forkhead transcription factor FoxG1 has been shown to enforce a NPC-like state on fibroblast cells, indicating that Sox2 and Brn2 may combinatorially control NPC identity in an analogous fashion to Sox2 and Oct4 in ESCs (Lujan et al., 2012). Therefore, we hypothesized that Brn1 and Brn2 could function as Sox2 partner factors in NPCs.

We performed ChIP-Seq to test whether Brn1 and Brn2 bound to a similar set of genomic regions as Sox2 in NPCs. Our analysis identified 8,867 and 6,574 enriched regions for Brn1 and Brn2, respectively. We observed that Oct4 and the Brn factors target distinct regions of the genome in ESCs and NPCs and that those bound by Brn1 and Brn2 in NPCs are highly correlated (Figure 2c). Brn1 and Brn2 sites, like Oct4 bound regions, were enriched for a sequence highly similar to a canonical octamer motif 5'-ATGCA/T A/T A/T A/T -3' (Staudt et al., 1996; Phillips and Luisi, 2000) (Figure S3). The high correlation between Brn1- and Brn2-bound genomic regions is consistent with the homology in their DNA binding domains (Treacy and Rosenfeld, 1992), motif preference, ability to compensate for each other in knockout experiments (McEvilly et al., 2002; Sugitani et al., 2002), and known individual targets known to be bound by both Brn1 and Brn2 *in vivo* (Catena et al., 2004; Tanaka et al., 2004). Interestingly, like Sox2, Brn1 and Brn2 bind to more enhancer regions than promoters (Figure S3). Oct4 is known to regulate genes involved in pluripotency and lineage commitment, whereas Brn1 and Brn2 occupied enhancers appear to reside near genes with roles in NPC identity and differentiation into neurons and glia (Figure S3). Thus, Brn1 and Brn2 occupy many genomic regions by utilizing a canonical Octamer motif in NPCs, and these regions are distinct from those bound by Oct4 in ESCs.

We then examined the overlap between the POU factors and Sox2 in ESCs and NPCs. We confirmed that Sox2 and Oct4 showed significant overlap at promoters and enhancer regions in ESCs (Figure S4) (Boyer et al., 2005; Loh et al., 2006; Marson et al., 2008; Chen et al., 2008). In contrast, binding of Sox2 and Brn1 and Brn2 overlapped predominantly at distal enhancer regions (Figure 2d;

Figure S4) but was not present at many Sox2 target promoters. For example, Sox2, Brn1, and Brn2 were bound at putative enhancers 5', 3', and within and intron of the Notch1 locus, while only Sox2 was present the Notch1 TSS. Also, Sox2, Brn1, and Brn2 co-localized 3' of the Ascl1 locus, while Sox2 was bound to the promoter of this gene without Brn1 or Brn2 (Figure 2e). Thus, Sox2, Brn1, and Brn2 predominantly co-occupied enhancer regions in NPCs, unlike Sox2 and Oct4 in ESCs which showed extensive co-localization at promoters as well.

We next sought to define the functional role Sox2, Brn1, and Brn2 may have in binding enhancer regions in NPCs. These factors bound more H3K4me1+/H3K27Ac-, or poised, enhancers than H3K4me1+ /H3K27Ac+ or H3K4me1-/H3K27Ac+, active elements (Figure 2f). Genes linked to Sox2:Brn1 and Sox2:Brn2 bound active enhancers were expressed at a higher level than those genes linked to poised elements (Figure 2g). Gene ontology analysis revealed that genes associated with Sox2, Brn1, and Brn2 bound enhancers in NPCs are involved in early nervous system development, particularly in neurogenesis (Figures 2h, 2i). Genes linked to active enhancers included transcription factors known to play a role in NPC identity such as Sox2 and Notch1, components of cell signaling cascades which support NPC function such as the Trk and EGF pathways (Xian and Zhou, 2004; Bartkowska et al., 2007), and genes involved in cell migration and cell adhesion, the balance of which plays a pivotal role in shaping the developing CNS (Morest and Silver, 2003). Interestingly, genes linked to poised enhancers are specifically enriched for regulators of later neuronal developmental stages, such as the proneural transcription factor Math1, genes involved in axon guidance, and genes involved in signaling pathways which are crucial for neuronal development, such as the L1CAM, Nectin, and Wnt (Haney et al., 1999; Fradkin et al., 2005; Park et al., 2008), supporting the notion that Sox2 may act as a pioneer factor to set the stage for further development of NPCs and suggesting Brn1 and Brn2 may also be in this class of genes. Taken together, these data demonstrate an enhancer-specific, genome-wide Sox:POU interaction in NPCs may be critical for both stem cell maintenance and fate commitment.

Figure 2

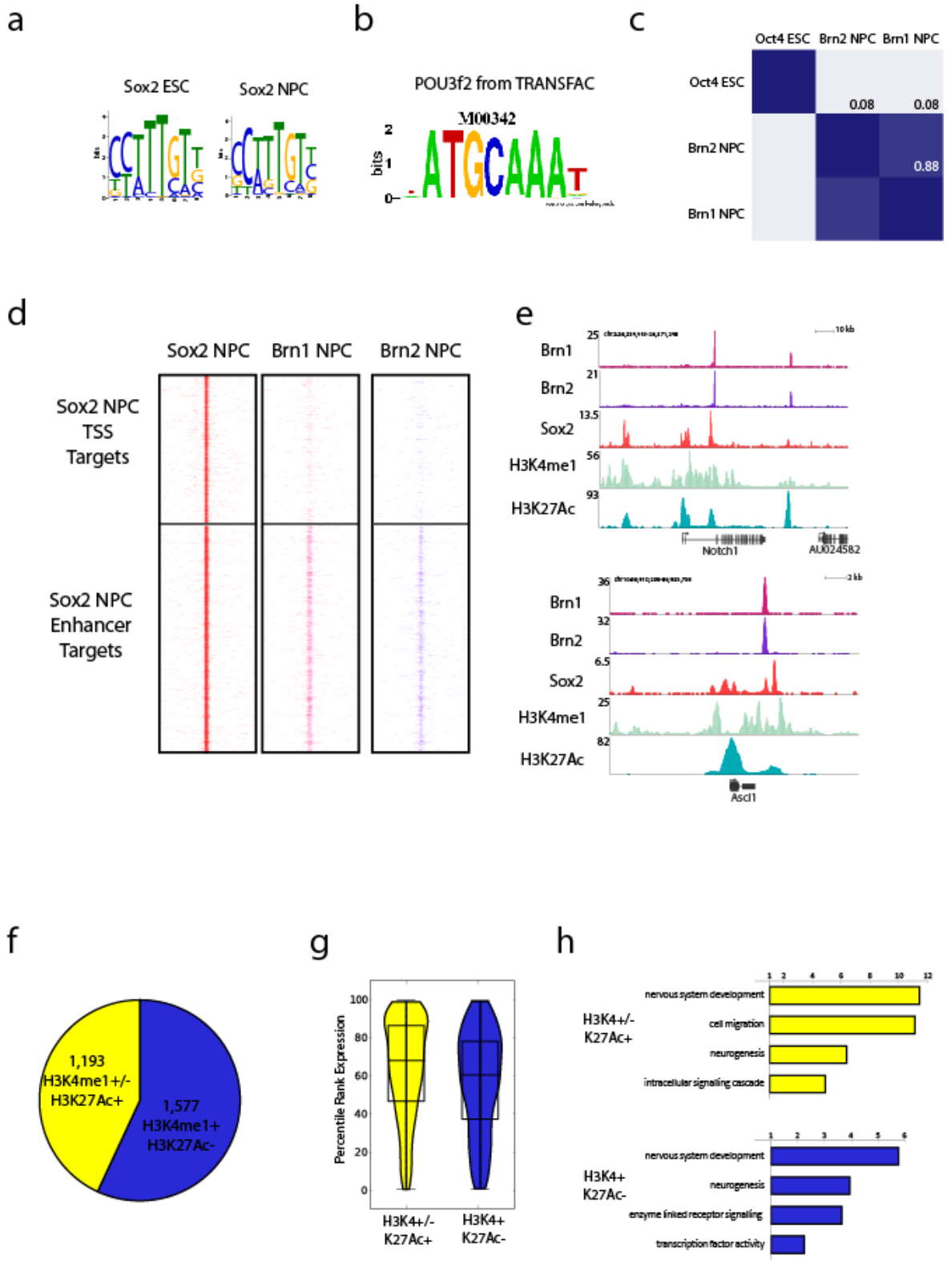

a
Sox2 ESC
Sox2 NPC
b
POU3f2 from TRANSFAC
M00342
bits
c
Oct4 ESC
Brn2 NPC
Brn1 NPC
0.08
0.08
0.88
d
Sox2 NPC
Brn1 NPC
Brn2 NPC
Sox2 NPC TSS Targets
Sox2 NPC Enhancer Targets
e
Brn1
Brn2
Sox2
H3K4me1
H3K27Ac
Notch1
AU024582
Ascl1
f
1,193 H3K4me1+/- H3K27Ac+
1,577 H3K4me1+ H3K27Ac-
g
Percentile Rank Expression
H3K4+/- K27Ac+
H3K4+ K27Ac-
h
H3K4+/- K27Ac+
nervous system development
cell migration
neurogenesis
intracellular signaling cascade
H3K4+ K27Ac-
nervous system development
neurogenesis
enzyme linked receptor signaling
transcription factor activity

Figure 2 - Sox2 co-localizes with Brn1 and Brn2 at enhancer regions of genes involved in neural development in NPCs. (a) Motifs similar to the canonical Sox2 motif enriched by MEME motif analysis of Sox2 bound regions in ESCs and NPCs **(b)** Brn2 motif enriched in Sox2 target regions in NPCs. **(c)** Correlation matrix comparing Oct4, Brn1, and Brn2 enriched regions in ESCs and NPCs. **(d)** Heat maps of Sox2, Brn1, and Brn2 enrichment in NPCs centered on peaks of enrichment and extended 4kb in each direction at TSS regions and marked enhancer regions. **(e)** Gene plots showing density of indicated ChIP-Seq datasets at indicated loci. y-axis corresponds to reads per million. **(f)** Breakdown of number of Sox2:Brn1 and Sox2:Brn2 target enhancers are H3K4me1+, H3K27Ac- (poised) or H3K4me1+/-, H3K27Ac+ (active). **(g)** Expression level of genes linked to poised and active Sox:POU target enhancers in NPCs. **(h)** Gene ontology analysis of genes linked to poised and active Sox:POU target enhancers in NPCs.

Different Sox2:POU combinations display a distinct binding site configurations

While Sox2 and the POU factors seem to use highly similar Sox and Octamer motifs to bind genomic targets in ESCs and NPCs, the loci these factors actually bind are almost mutually exclusive. We therefore sought to refine our understanding of how Sox2, Oct4, Brn1 and Brn2 find genomic targets in ESCs and NPCs. First, we searched Sox2:Oct4, Sox2:Brn1, and Sox2:Brn2 bound regions for other conserved transcription factor binding sites. Indeed, we found that Sox2:Oct4 co-occupied regions also contained a set of binding sites for other transcription factors, and this set was distinct from those which associate with Sox2:Brn1 and Sox2:Brn2 (Figure S5). Interestingly, the recognition sites for other known regulators of neurogenesis which are expressed in NPCs, such as Rfx factors and NF-I, were enriched in Sox2:Brn1 and Sox2:Brn2 bound regions (Figure S5). These data indicate that Sox2, Brn1, and Brn2 might partner with these other factors to regulate transcription in NPCs, and raises the possibility that the interaction with distinct sets of partner factors in ESCs and NPCs may in part guide Sox2:POU target selection.

Previous studies have indicated that the relative position of Sox and Octamer motifs to each other could affect the ability of Sox and POU factors to utilize these composite motifs. Sox2 and Oct4 are known to bind in very close proximity to each other to a composite Sox:POU motif consisting of a 5' Sox site, a 1 bp spacer, then a 3' Octamer site at many targets in ESCs. Some variants of this motif orientation can also be bound by Sox2 and Oct4, while others cannot (Ambrosetti et al., 1997; Remenyi et

al., 2003). Further, different motif conformations have different effects on the ability of binding to induce transcription (Botquin et al., 1998). This sensitivity to orientation extends to other pairings of Sox and POU family members (Kuhlbrodt et al., 1998; Tomoika et al., 2002; Tanaka et al., 2004; Reiprich et al., 2010). Thus, we sought to define the motif preferences for Sox2:Brn1 and Sox2:Brn2 in NPCs to possibly explain how these modules can regulate global gene expression networks in a distinct way from Sox2:Oct4.

In some cases, it seems Sox2:Oct4, Sox2:Brn1, and Sox2:Brn2 can bind to a given site in the genome, but have different effect on transcription. For example, an intronic enhancer of the *Nestin* locus contains a Sox:Pou composite motif comprised or a 5' Octamer site followed by a 7-bp spacer and then an Sox site (Tanaka et al., 2004). Reversing the orientation of these motifs to one similar to the known 5' Sox, 3'Octamer pattern in Sox2:Oct4 targets rendered it unable to be activated by Sox2:Brn2, which drives activity of this element in the brain. This site can be bound by Sox2:Oct4, Sox2:Brn1, and Sox2:Brn2 *in vitro*, and this binding can drive expression of a reporter construct regulated by this element, while other Sox:POU combinations such as Sox2:Brn5 or Sox9:Brn2 were not able to drive this expression. We find that Sox2 and Oct4 can also bind to this site *in vivo*, as can Sox2 and Brn1 (Figure S6a), and confirm that Sox2:Oct4 cannot support expression of Nestin in ESCs but binding of that Sox2 and Brn1 is compatible with expression in NPCs (Figure S6b). This suggests that while Sox2:Oct4 can drive expression utilizing this element *in vitro*, *in vivo* this pairing may be acting as a pioneer factor module for this enhancer, to be later replaced by Sox2:Brn1 in NPCs in order to active this element. Thus, binding of Sox2 with different partner POU factors may influence the functional consequence of this interaction with a genomic locus.

The majority of Sox2:Brn targets are not shared by Sox2:Oct4, so we therefore sought to understand what mechanism could underlie this selection. Given that we identified similar Sox and Octamer consensus motifs bound in both cell types, we hypothesized that analysis of the motif configuration at the global level between Sox2 and the cell-type-specific POU factors, Oct4, Brn1, and

Brn2 may reveal mechanistic insights into how these factors recognize distinct genomic sites. Consistent with previous data, we determined that Sox2 binding occurred within close proximity of Oct4 binding in ESCs at many genomic targets (Figure 3a). In contrast, regions of Sox2 and Brn1 or Brn2 co-occupancy were often more widely distributed (Figure 3b, c). For example, the *Jam2* locus was bound by Sox2 and POU factors in both ESCs and NPCs. Sox2 and Oct4 occupied a region upstream of the TSS in close proximity to each other, while the peaks of Sox2 binding in NPCs was approximately 200 bp away from peaks of Brn1 and Brn2 binding (Figure 3d). Moreover, many known Sox2:Oct4 target loci harbor a composite Sox-Octamer motif, consisting of a 5' Sox site followed by a 1 bp spacer and then a 3' Octamer site, and we observed that this 5' Sox-Oct composite motif is the most frequent configuration at Sox2:Oct4 ESC targets (Figure 3e). This configuration was also the most frequently observed in Sox2:Brn1 and Sox2:Brn2 co-bound regions, but in addition a significant fraction of Sox2:Brn1 and Sox2:Brn2 target regions displayed the reverse orientation (3' Sox, 5' Octamer) with distances of 5 bp and 24 bp between the motifs (Figure 3f, g). For example, a Sox2:Brn1 target region within an intron of the Dhrs2 locus, encoding a widely expressed metabolic gene, contained a Sox:Octamer composite motif in the -24bp configuration, and was not bound by Sox2:Oct4 in ESCs (Figure 3h). These data suggest that differences in target selection by Sox2 and its cell-type-specific partner factors may be in part dictated by differential Sox:Oct motif preferences by Sox2:Oct4 vs. Sox2:Brn1 or Sox2:Brn2.

Figure 3

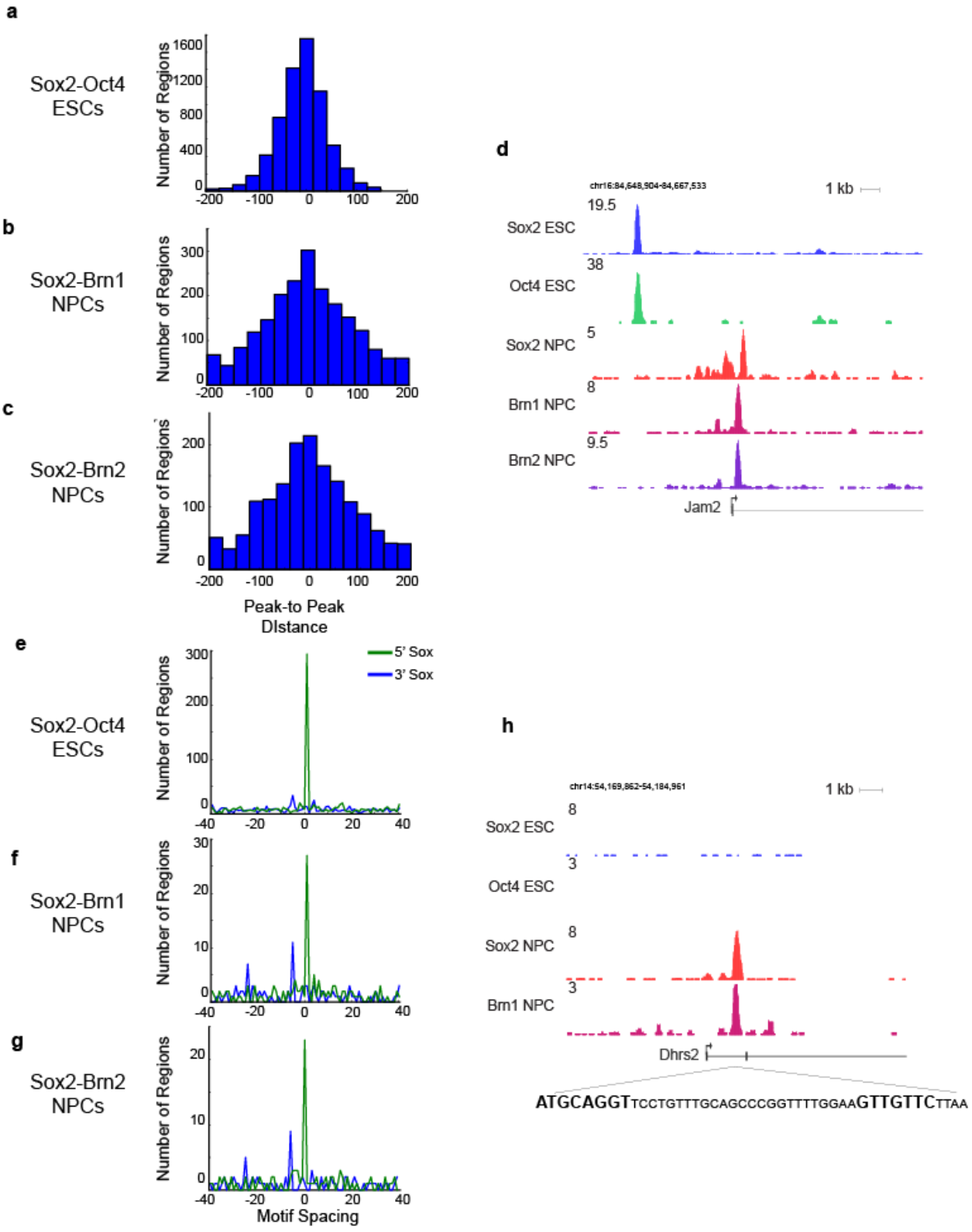

a
Sox2-Oct4 ESCs
Number of Regions
b
Sox2-Brn1 NPCs
Number of Regions
c
Sox2-Brn2 NPCs
Number of Regions
Peak-to Peak DIstance
d
chr16:84,648,904-84,667,533
1 kb
Sox2 ESC
Oct4 ESC
Sox2 NPC
Brn1 NPC
Brn2 NPC
Jam2
e
Sox2-Oct4 ESCs
5' Sox
3' Sox
Number of Regions
f
Sox2-Brn1 NPCs
Number of Regions
g
Sox2-Brn2 NPCs
Number of Regions
Motif Spacing
h
chr14:54,169,862-54,184,961
1 kb
Sox2 ESC
Oct4 ESC
Sox2 NPC
Brn1 NPC
Dhrs2
ATGCAGGTTCCTGTTTGCAGCCCGGTTTTGGAAGTTGTTCTTAA

Figure 3. Motif configuration affects both the consequence and the occurrence of binding by Sox2 and cell-type specific POUs. (a-c) Distribution of distances in base pairs between peaks of Oct4 (left) Brn1 (middle) and Brn2 (right) from Sox2 bound peaks. **(d)** Gene plots at the *Jam2* locus. y-axis corresponds to reads per million. **(e-g)** Distribution of distance in base pairs of Sox and POU motifs within Sox2:Oct4, Sox2:Brn1, and Sox2:Brn2 bound regions. Green plot represents 3'Sox-Octamer-5' configuration, while the blue plot represents the reverse orientation. **(h)** Gene plots at the *Dhrs2* locus. Inset is genomic sequence containing Sox and Octamer motifs under peaks of Sox2 and Brn1. y-axis corresponds to reads per million.

Brn2 expression in ESCs leads to accelerated neural differentiation

We next sought to understand what the functional consequences of Sox:Brn binding to enhancers could be. Previous studies from our lab and others have shown that epigenetic marks associated with distal enhancers vary more widely than those associated with promoters (Heintzman et al., 2007, 2009; Creyghton et al., 2010), indicating that they play a crucial role in regulating cellular identity. Thus, we reasoned that Sox2, Brn1, and Brn2, through their preference for binding distal enhancers, would be strong regulators of cell state. To test this idea, we generated ESC lines that harbored a tetracycline inducible Brn2 transgene and assayed their potential to differentiate into NPCs by culturing then in serum-free medium without LIF (Ying et al., 2003) (Figure S7a, b). ESCs showed distinct morphological changes from round cells that grow in colonies in control cultures to polarized, Nestin positive cells at Day 1 of differentiation when exposed to ectopic Brn2, but took several days to accomplish this transition in control cells (Figure 4A). Consistent with this, neural lineage genes such as Nestin and Sox1 were up-regulated in Brn2-induced, differentiating ESCs (Figure 4B). Notably, Brn2 expression did not affect the normal transition in the levels of Oct4, Sox2, and Nanog compared to control cells indicating that Brn2 cannot completely overcome the Oct4:Sox2 autoregulatory circuitry present in ESCs (Figure 4B; Figure S7C). Thus, Brn2 can induce ESCs to differentiate towards the neural lineage, suggesting that the genomic binding pattern of Brn2 at distal enhancers functions to enforce neural identity.

In order to determine the mechanism by which Brn2 affects this change, we performed ChIP-Seq on Brn2 at Day 2 of this time course to define a set of genomic targets which may be occupied ectopic

Brn2. We also performed ChIP-Seq on Sox2 in these cells to test the hypothesis that ectopic Brn2 may collaborate with the endogenous Sox2 in these cells to co-regulate genomic targets, analogous to the scenario in NPCs. We defined 8,401 genomic regions enriched by Brn2 ChIP and 12,362 genomic regions enriched by Sox2 ChIP at high confidence in these cells. As in ESCs and NPCs, the majority of Sox2 and Brn2 occupied regions were distal to annotated TSS (Figure S7D). Globally, ectopic Brn2 targets correlated with those bound by Sox2 and Brn2 in NPCs (Figure 4C), while Sox2 targets in these cells correlated with Sox2:Oct4 ESC targets (Figure 4C). However, we were able to define approximately 1,500 regions which were co-occupied by ectopic Brn2 and Sox2 in these cells. We then compared these regions to Sox2-bound regions in ESCs, NPCs, and control day 2 differentiating ESCs without ectopic Brn2 to test whether ectopic Brn2 was utilizing endogenous Sox2:POU circuitry with Sox2 to affect cell fate (Figure 4D). Indeed, we found that ectopic Brn2:Sox2 occupied regions bound by Sox2 in ESCs (4D, green, purple). Many of these regions were also bound by Sox2 in control differentiating ESCs without Brn2, suggesting that Sox2 occupancy of these regions was Brn2 independent (data not shown). Interestingly, ectopic Brn2 and Sox2 also occupied regions not bound by Sox2 in ESCs. A subset of these regions were Sox2 targets in NPCs (4D, dark blue), whereas others were specific to these cells and bound in neither ESCs nor in NPCs (4D, orange). A fraction of these regions were also bound by Sox2 in control differentiating ESCs which did not have ectopic Brn2 (4D, red, cyan), but the majority were bound only in cells with ectopic Brn2. This suggests that ectopic Brn2 was able to bind to these regions and recruit Sox2, suggesting it may act as a pioneer factor.

We then tested whether the binding of ectopic Brn2 and Sox2 influenced the expression of the closest gene to the co-bound region. We measured the expression of the nearest genes to three classes of ectopic Brn2:Sox2 bound regions: those which were occupied by Sox2:Oct4 in ESCs and Sox2:Brn2 in NPCs (Class I), those which were occupied only in NPCs by Sox2:Brn2 (Class II), and regions which were only bound by Brn2 in NPCs (Class III) (Figure 4E). In all three contexts, ectopic Brn2 was sufficient to induce expression of some of these genes (Figure 4F, S7E), indicating that ectopic Brn2 can

utilize diverse Sox2-dependant and independent mechanisms to activate gene expression and influence cell identity.

Figure 4

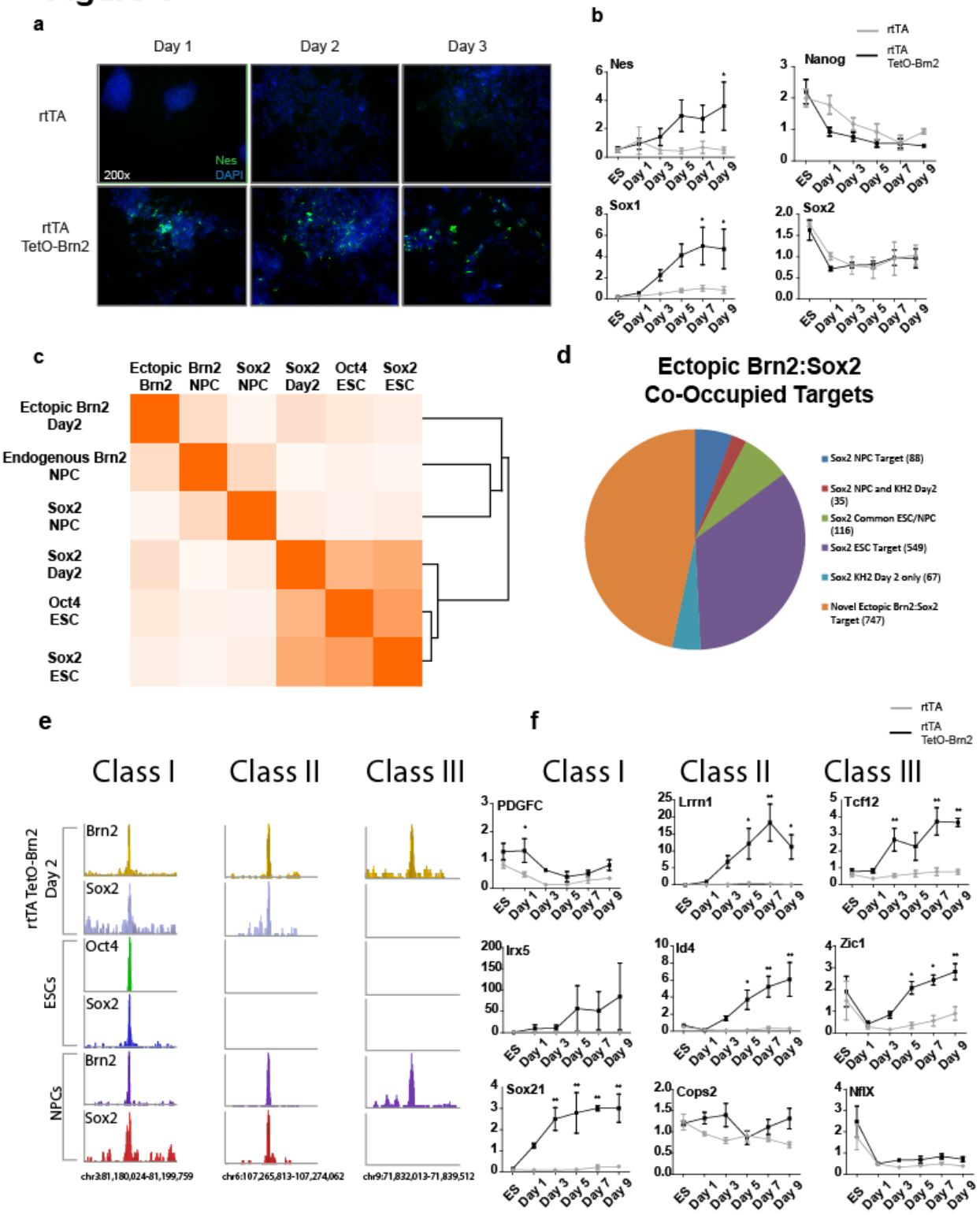


Figure 4. Brn2 biases ES cells towards neural differentiation. **(a)** Immunocytochemistry of Nestin (green) in ES cells induced to differentiate in adherent cultures with or without ectopic Brn2. **(b)** Quantitative Real-Time PCR of the indicating genes in ESCs with (black lines) and without (gray lines) ectopic Brn2 expression through differentiation. y-axis represents relative expression normalized to

GapDH in 3 biological replicates, measured in triplicate. ESC time point is ESCs without doxycycline, and d1-d9 time points represent time in differentiation medium. **(c)** Correlation matrix comparing ectopic Brn2 binding at Day 2 of differentiation and Sox2 at Day2 in Brn2 induced cells with the binding of Sox2 and Brn2 in NPCs and Sox2 and Oct4 in ESCs. **(d)** Breakdown of the fractions of ectopic Brn2:Sox2 co-occupied regions which are also bound by endogenous Sox2 in ESCs, NPCs, or Day 2 of differentiation in control cells. **(e)** Gene plots depicting representative loci distal to the start sites of PDGFC, Lrrn1, and Tcf12 constituting three classes (Class I, II, and III, respectively) of ectopic Brn2:Sox2 co-occupied regions. **(f)** Quantitative Real-Time PCR time course of expression of genes in each class in control and Brn2 expressing differentiating ESCs. y-axis represents relative expression normalized to GapDH in 3 biological replicates, measured in triplicate. * = p-value < 0.05, * = p-value < 0.01 ANOVA with Bonneferoni correction

Discussion

In this study, we investigated the genome-wide binding profiles of the HMG-box transcription factor Sox2 in two cell types in which it is thought be a master regulator of cell identity: pluripotent embryonic stem cells and multipotent neural precursor cells. How this master regulator can control cellular identity in two distinct stem cell types has been an unsolved question for some time. In ESCs, Sox2 was known to partner with the POU factor Oct4 to stabilize its interaction with many genomic targets. These targets included both genes involved in ESC identity and lineage commitment factors which would become active only later in development, indicating that it could act as a pioneer factor. How Sox2 could function in the cellular environment of an NPC which lacks Oct4, and whether Sox2 acted both to activate gene expression in some contexts and poise expression in others was not clear. Recently, analysis of Sox3 in NPCs revealed that a subset of Sox2 target regions in ESCs linked to repressed genes involved in NPC identity were bound by Sox3 in NPCs (Bergsland et al., 2011). This study also suggested that Sox2 may co-occupy many Sox3 bound sites in NPCs. These genes in turn fell into two classes: those which were expressed in NPCs, and those which were repressed in NPCs and were later bound by Sox11 as these cells completed their differentiation towards mature neurons. Our data confirms that indeed Sox2 as well as Sox3 seems to also act as a pioneer factor in NPCs. This study left open one question: How can Sox2 play such a key role in two very different cell types yet bind many of the same genomic targets in these cells? Another genome-wide analysis of Sox2 in NPCs revealed its co-localization with the ATP-dependent histone remodeler Chd7, indicating that Sox2 interfaces with this

epigenetic regulator in modulate transcription (Engelen et al., 2011). While this interaction revealed a possible mechanism for Sox2 function, it is not known if an interaction with a chromatin remodeler like Chd7 could stabilize ternary complex formation between itself, Sox2, and a DNA target in the same way transcription factors such as Oct4 have been shown to do. Thus the question remains: How can Sox2 bind tightly to genomic targets in NPCs in the absence of Oct4?

To begin to answer these open questions, we performed ChIP-Seq on Sox2 in ESCs and NPCs. We confirmed that in ESCs, Sox2 bound the promoters of genes encoding transcription factors and epigenetic regulators. In NPCs Sox2 also occupied the promoters of genes in these classes, but in addition genes involved in RNA splicing, which is known to play a major role in the development of the CNS (Boutz et al., 2007; Grabowski, 2011). We also confirmed that the majority of Sox2 bound regions were distal to known TSSs in both cell types, and in support of the notion that these regions might represent distal enhancers, showed that a significant fraction of these non-TSS associated Sox2 bound regions correlate with known marks of enhancers, H3K4me1 and H3K27Ac. Consistent with the idea that Sox2 may act as a pioneer factor to prime enhancers for later activation (Smale, 2010; Wegner et al., 2011), we showed that a large number of the enhancers are H3K4me1+ but H3K27Ac-, a state characteristic of poised enhancers. Thus, Sox2 occupied two classes of targets: regulatory regions highly expressed genes involved in NPC state, and poised genes involved in later lineage commitment.

Sox2 occupied almost mutually exclusive sites in ESCs and NPCs, despite using the same DNA motif to recognize these genomic targets. How could this selection be mediated? Given that Sox proteins frequently bind target loci with partner factors, exemplified by the partnership of Sox2 in ESCs with the POU factor Oct4, we tested Sox2-bound regions in NPCs for enrichment of the DNA binding motifs of known transcription factors from the TRANSFAC database. Interestingly, we found the recognition site for another POU factor, Brn2, enriched in these regions. Given the evolutionary conservation of the Sox:POU interaction, the co-expression of Sox2, Brn2 and the closely related Brn1 in neurogenic regions

of the brain, and the neurodevelopmental defects associated with loss of function of Brn1 and Brn2, we hypothesized that these neural POU factors could partner with Sox2 in NPCs.

To test this, we performed ChIP-Seq on Brn1 and Brn2 in NPCs. Interestingly, while Sox2 co-occupied most binding sites with Oct4 in ESCs, we found that the interaction between Sox2 and Brn1 and Brn2 was enhancer specific. Many of these enhancers were active (H3K27Ac+), and genes linked to them included regulates of the NPC state such as Notch1. These data validate two described enhancers bound by Sox2, Brn1, and Brn2 in neurogenic regions of the brain, the *Nes30* enhancer of the *Nestin* locus (Tanaka et al., 2004) and the 3' enhancer of the *Sox2* locus, *SRR2* (Miyagi et al., 2006), and extend them to thousands of neural specific enhancers. Interestingly, both of these enhancers are also bound by Sox2:Oct4 in ESCs; *SRR2* is maintained as an active enhancer, while at the *Nestin* locus, Sox2:Oct4 may be acting as pioneer factors to facilitate Sox2;Brn1 binding in NPCs. Thus, cell-type-specific POU factors collaborate with Sox2 to hand off targets during developmental transitions. Another subset of elements were poised, H3K4me1+/H3K27Ac- enhancers which were linked to genes which encode regulators of development passed the NPC state. This suggests that Brn1 and Brn2, like Sox2, may be acting as pioneer factors in NPCs. Interestingly, the POU factor Brn5, like Sox factors Sox11 and Sox21, is expressed in differentiated neurons and some glia in the adult CNS and thought to play a role in regulating cell state in these tissues (Okamoto et al., 1993; Andersen et al., 2003; Cui et al., 1997, 1998; Wu et al., 2001; Sandberg 2005; Ferletta et al., 2011). The comparison of the genome-wide binding profiles of Brn5 and Sox21 in neurons to Sox11 in these cells and Sox and POU factors in earlier developmental stages will be instrumental in clarifying the precise roles of these factors in priming enhancers for later developmental decisions in the neural lineage.

Many recent studies have suggested that enhancers may play a crucial role in regulating cellular state, so we hypothesize that Sox2, Brn1, and Brn2, through their preferential association with these elements, would be potent regulators of cellular identity. To test this, we generated ESCs harboring a tetracycline inducible Brn2 transgene and assayed the potential of these cells to differentiate into NPCs,

and showed that Brn2 accelerated this process. This observation is supported by other experiments in the literature which suggest that Brn2 is an early marker of neural commitment. In retinoic acid induced P19 embryonic carcinoma cells, Brn2 is induced early as these cells adopt the neural fate *in vitro* (Jin et al., 2009). Interestingly, another study showed that by switching the suite of nuclear importins, which regulate traffic between the cytoplasm and the nucleus, expressed in ESCs, these cells could be biased towards differentiating towards the neural fate due to differential import of Oct4 and Brn2 by these factors (Yasuhara et al., 2007). Finally, a recent study *in vivo* demonstrated that Sox2 expression is maintained upon differentiation of epiblast to neural plate by a hand off of a Sox2 enhancer from Sox2:Oct4 to Sox2:Brn2 (Iwafuchi-Doi et al, 2011). Mechanistically, we showed that ectopic Brn2 can recruit Sox2 to new enhancer targets in the cells, and this co-binding is sufficient to induce expression of nearby genes. Thus, Brn2 may act as a specificity factor of Sox2, allowing it to bind target loci to which it cannot bind alone.

The fact that POU factors directed Sox2 to cell-type-specific targets in both ESCs and NPCs raised a question: If Sox2 uses the same recognition motif in both cell types, and Oct4, Brn1, and Brn2 all bind highly similar Octamer motifs, how is this context specific binding pattern achieved? To address this paradox, we sought to refine our understanding of the precise rules of Sox2:Oct4, Sox2:Brn1, and Sox2:Brn2 binding. We found that while the motifs bound by the factors were highly similar, the configuration of the Sox and Octamer motifs in Sox2:Brn1 and Sox2:Barn2 target regions differed significantly. At some loci, as in the case of *Nestin*, this distinction manifested itself by the locus being bound by Sox2:Oct4 in ESCs but not activated, and only becoming activate upon differentiation to and NPCs and the binding of the element by Sox2:Brn1. In other cases, the distinct orientation of Sox and Octamer motifs in Sox2:Brn1 and Sox2:Brn2 regions precluded binding by Sox2:Oct4 in ESCs. This sensitivity to the orientation of Sox and Octamer motifs has been noted at several individual targets of Sox and POU factors (Ambrosetti et al., 1997; Botquin et al., 1998; Kuhlbrodt et al., 1998; Tomoika et al., 2002; Remenyi et al., 2003 Tanaka et al., 2003; Reiprich et al., 2010), but this is the first time this

observation has been extended genome wide. This suggests that allosteric interactions between transactivation domains of Sox and POU factors may be key in stabilizing ternary complexes between them and DNA targets globally, examples of which have been observed at individual targets of Sox2:Oct4 in ESCs (Ambrosetti et al., 2000).

This set of experiments demonstrates that collaboration between Sox and POU factors controls a subset of distal enhancers from the pluripotent ESC state to the multipotent NPC state. Structural differences in the DNA sequence of these enhancers as well as the intrinsic characteristics of the specific Sox and POU factors expressed in a given cell type along this pathway likely contribute to the diversity of gene expression programs observed during this transition. Taken together with other recently published work, these data suggest that this Sox:POU interaction may regulate this pathway from the earliest stages of development all the way through terminal, neuronal differentiation. Going forward, understanding the combinatorial code regulation development in this and other lineages will be key to understanding how enhancers are regulated in metazoan development.

Supplemental Figures

Figure S1

Comparison	Correlation
Sox2 NP Replicates	0.7
Sox2 ES (Lodato et al. vs Marson et al.)	0.77
Sox2 ES vs Sox2 NP	0.1

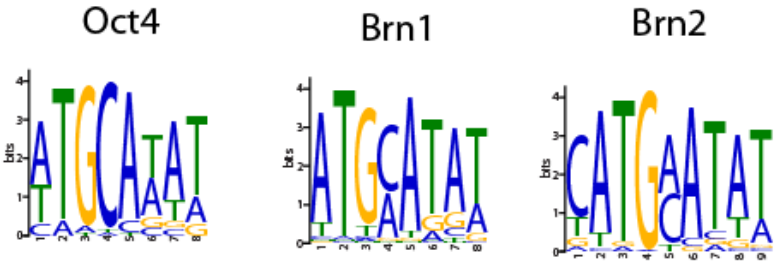
Figure S2

ESC Expression	Normalized Expression
Pou5f1 (Oct4)	4178.83
Pou2f1	419.37
Pou5f2	46.10
Pou6f1	39.27
Pou4f1	38.70
Pou2f3	27.17
Pou4f2	21.90
Pou3f3 (Brn1)	17.87
Pou2f2	17.60
Pou6f2	11.53
Pou3f2 (Brn2)	9.80
Pou3f4	9.53
Pou4f3	4.77
Pou1f1	2.77

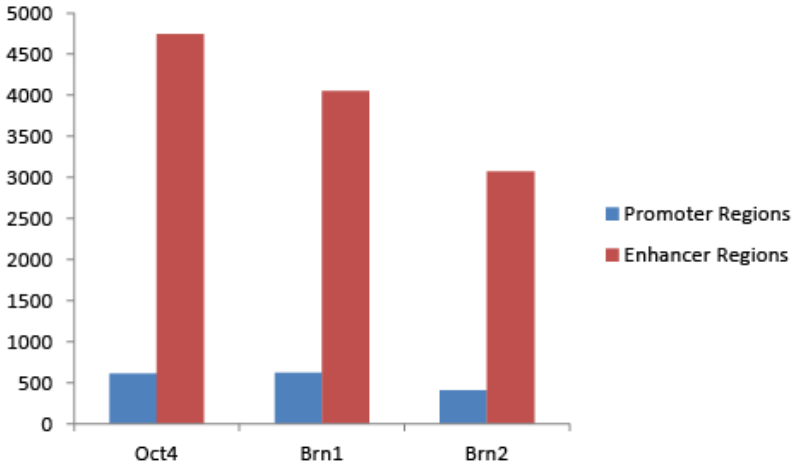
NPC Expression	Normalized Expression
Pou3f3 (Brn1)	1269.97
Pou6f1	271.40
Pou2f1	182.63
Pou3f4	81.40
Pou5f2	67.77
Pou4f1	54.60
Pou3f2 (Brn2)	53.37
Pou2f3	20.40
Pou6f2	11.83
Pou5f1 (Oct4)	9.60
Pou2f2	9.53
Pou4f2	2.73
Pou4f3	2.37
Pou1f1	0.93

Figure S3

a



b



c

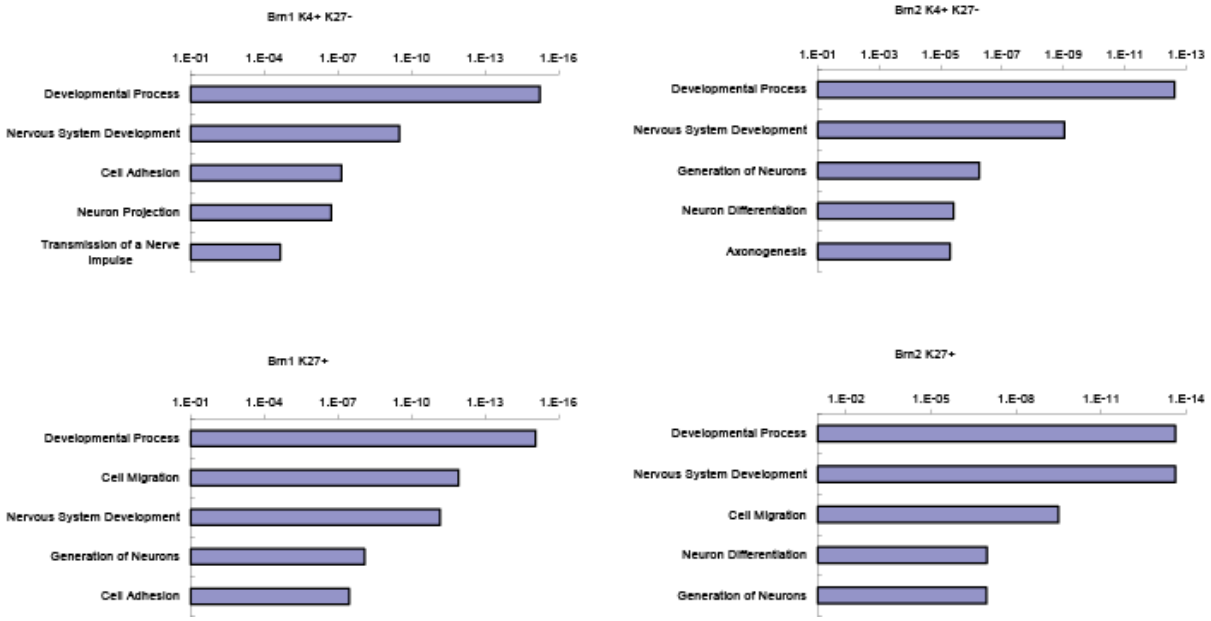


Figure S4

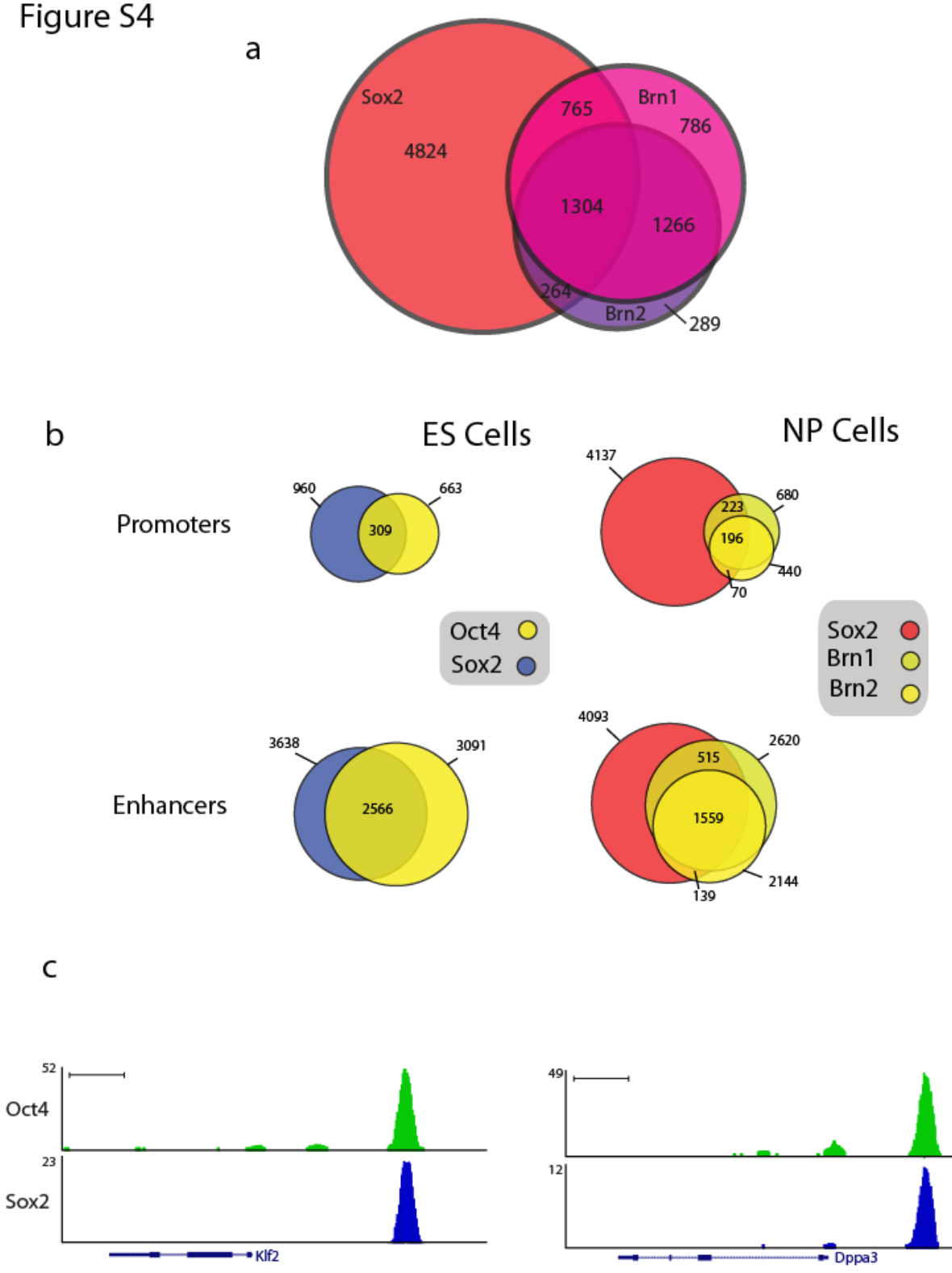
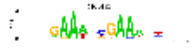

a
Sox2
4824
765
Brn1
786
1304
1266
264
Brn2
289
b
ES Cells
NP Cells
Promoters
960
663
309
4137
680
223
196
440
70
Oct4
Sox2
Sox2
Brn1
Brn2
Enhancers
3638
3091
2566
4093
2620
515
1559
2144
139
c
52
Oct4
23
Sox2
Klf2
49
12
Dppa3

Figure S5

Sox2-Oct4 ESCs

Transcription Factor	TRANSFAC Motif Logo	P-Value
Zeb1, TBX5		3.08e-82
MEF-2		7.55e-79
SATB1, HNF1, Nkx3.1, HP1, Pit-1, Cdx Family		2.17e-76
PLZF, HOXA5, Pax-4, STAT5A, AIRE		2.14e-69
IRF Family, Blimp-1, ISGF-3		3.74e-46
TRF1		5.51e-44
GATA-5, GATA-1, Lmo2		2.98e-43
FOXO3, FOXO1, FOXO4, Foxd1		1.26e-42
Homez, Pbx Family, Onecut1		4.65e-40
Obox Family, Pitx Family, Otx Family, Gsc, Crx		8.36e-38

Sox2-Brn2 NPCs

Transcription Factor	TRANSFAC Motif Logo	P-Value
NF-1, CTF1		8.20e-134
c-Myc:Max, USF1, Srebf1, Arnt, Stra13, CLOCK:BMAL		5.41e-70
HIC1, RFX Family, MTF-1		3.56e-67
LBP-1, BEN, CBF1, HEB		1.09e-63
E2F Family		6.40e-49
Zfp206, NRF-1		8.05e-47
NeuroD, MYF, AP-4, Zeb1		4.01e-46
Ahr, Foxn1		1.44e-42
HES1, SREBP1		7.44e-39
Trim28, AP-2, AP-2gamma		1.13e-34

Figure S6

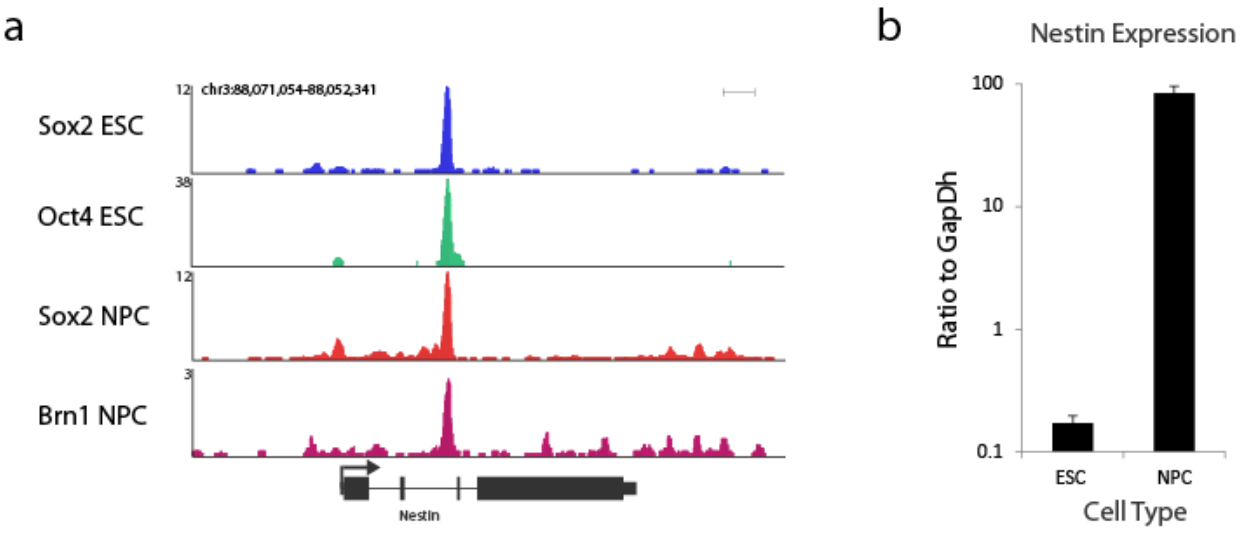
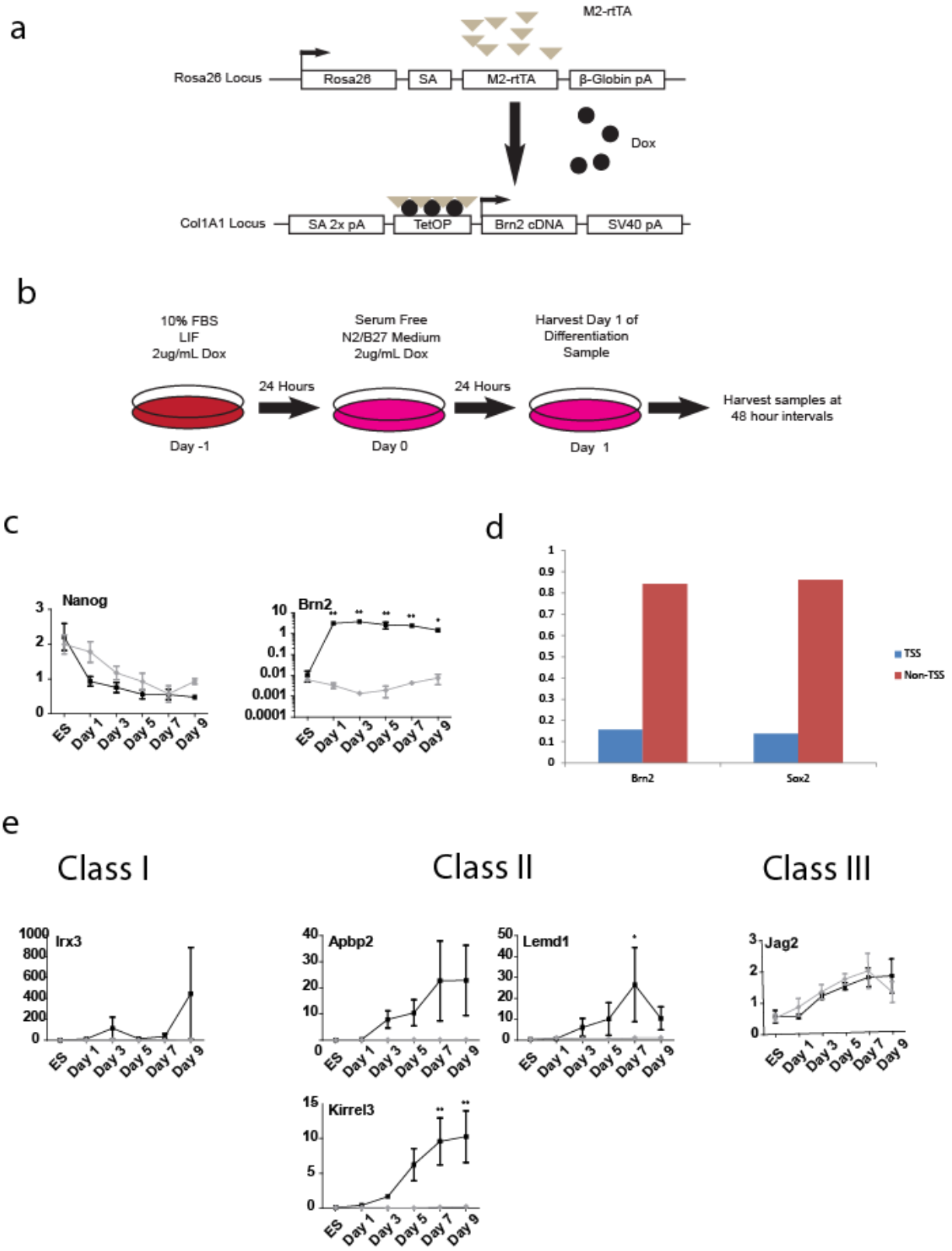


Figure S7


a
M2-rtTA
Rosa26 Locus
Rosa26
SA
M2-rtTA
β-Globin pA
Dox
Col1A1 Locus
SA 2x pA
TetOP
Brn2 cDNA
SV40 pA
b
10% FBS
LIF
2ug/mL Dox
24 Hours
Serum Free
N2/B27 Medium
2ug/mL Dox
24 Hours
Harvest Day 1 of
Differentiation
Sample
Harvest samples at
48 hour intervals
Day -1
Day 0
Day 1
c
Nanog
Brn2
ES
Day 1
Day 3
Day 5
Day 7
Day 9
d
TSS
Non-TSS
Brn2
Sox2
e
Class I
Class II
Class III
Irx3
Apbp2
Lemd1
Jag2
Kirrel3

Supplemental Figure Legends

Figure S1. Correlation of ChIP-Seq datasets

Figure S2. Expression of POU factors in ESCs and NPCs. Mean, RMA normalized values from highest detected probe set for each gene across three biological replicates interrogated with Affymetrix Gene Expression arrays are presented.

Figure S3. (a) Sequences similar to the canonical octamer motif enriched in Oct4, Brn1, and Brn2 targets by MEME analysis. (**b)** Breakdown of POU factor binding to promoters and enhancers. **(c)** GOSTAT gene ontology analysis of Brn1 and Brn2 bound poised and active enhancers.

Figure S4. (a) Venn diagram depicting overlap between Sox2, Brn1, and Brn2 enhancer associated bound regions in NPCs. **(b)** Venn diagrams depicting overlap between Sox2 and Oct4 in ESCs and Sox2, Brn1, and Brn2 at promoters and enhancer at the level of genes. **(c)** Gene plots of Sox2 and Oct4 binding to indicated loci. y-axis corresponds to reads per million. Scale bar corresponds to 1 kb.

Figure S5. Mann-Whitney U analysis of transcription factor motifs enriched in Sox2 and Brn1 and Brn2 bound regions in NPCs (left) and Sox2 and Oct4 bound regions in ESCs. Sox2 and POU motifs have been excluded. P-values reflect enrichment in Sox2:Brn2 overlapping regions. Gene whose names are in black are detected expressed in their given cell type while those in green are not detected.

Figure S6. (**a)** Gene plots showing density of Oct4 and Sox2 in ESCs and Sox2 and Brn1 in NPCs at a validated enhancer of the *Nestin* locus. y-axis corresponds to reads per million. Scale bar corresponds to 1 kb. **(b)** Q-PCR analysis of Nestin expression in ESCs and NPCs. Error bars represent standard deviation across three technical replicates.

Figure S7. (a) Schematic of transgenic system to inducibly express Brn2 from the Col1A1 locus (adapted from Hochedlinger et al., 2005) **(b)** Experimental design. Adherent Brn2 targeted and control ESCs were exposed to dox, then 24 hours later the medium was switched to differentiation medium. The next day cells were harvested as the Day 1 time point. **(c, e)** Quantitative Real-Time PCR time course of Nanog and Brn2 (c) and genes in Class I, II, and III (e) in control (gray) and Brn2 induced (black) differentiating ESCs. y-axis represents relative expression normalized to GapDH in 3 biological replicates, measured in triplicate. * = p-value < 0.05, * = p-value < 0.01 ANOVA with Bonneferoni correction. (**d**) Fraction of bound regions mapping within 1 kb of TSSs vs. distal to TSSs for indicated factors in d2 cells.

Methods

Cell growth and culture conditions

C57/BL6-129JAE (V6.5) mouse embryonic stem cells were cultured in DMEM supplemented with 10% Fetal Bovine Serum (Hyclone), 1,000 units/ml leukemia inhibitory factor (LIF) 0.001 % β-mercapto-ethanol (Sigma, M7522)., 100 μM nonessential amino acids (Invitrogen, 11140-050), 2 mM L-glutamine (Invitrogen, 25030-081), 100 U/mL penicillin, 100 μg/mL streptomycin (Invitrogen, 15140-122), Neural progenitors were derived via *in vitro* differentiation from V6.5 ES cells as described (Okabe et al., 1996). Briefly, embryoid bodies (EBs) were formed from V6.5 ES cells by culturing them in suspension without

LIF, and on day 4 these EBs were plated back onto adherent dishes. After 24 hours selection for neural precursors began by switching the medium to ITSFn, which contains 1:1 DMEM:F12 (Ham) supplemented with 5 μg/ml insulin (Invitrogen I6634), 50ug/ml transferrin (Sigma T2036), 0.5uM Sodium Selenite (Sigma S1382), 5ug/ml fibronectin (Invitrogen 33016015) and 100 U/mL penicillin, 100 μg/mL streptomycin (Invitrogen, 15140-122). After 8 days cultures were trypsinized to single cells, and trypsin was quenched with serum-containing medium. They were then passaged on to 15 μg /ml polyornithine (Sigma P3655) and 1 μg/ml laminin (Invitrogen 23017) coated dishes and cultured from that point onwards in N3 medium, which contains 1:1 DMEM:F12 (Ham), 25 μg /ml insulin, 50ug/ml transferrin, 0.5uM Sodium Selenite, 100 U/mL penicillin, 100 μg/mL streptomycin, 20nM progesterone (Sigma 8783), 100nM putricine (Sigma 5780), 5ng/ml bFGF (R and D Biosystems 233-FB-025) and 20ng/ml EGF (R and D Biosystems 236-EG-200). In the presence of growth factors these the vast majority of these cells can be labeled homogenously with antibodies against Nestin, Sox2, and Pax6. Upon growth factor withdrawal, the cells differentiate into TUJ1-positive neurons.

Chromatin Immunoprecipitation

Sox2 NPC ChIP, Brn1 NPC ChIP, Brn2 NPC ChIP

ChIP was performed as described (Lee et al., 2006). Briefly, approximately $5x10^8$ cells were cross-linked and chromatin fractions were isolated. Chromatin was sheared by sonication, whole cell extract (WCE) was removed, and ChIP was performed using the following antibodies:

Sox2 R and D Systems AF2018 goat polyclonal

Brn1 Santa Cruz Biotechnology sc-6028 goat polyclonal

Brn2 Santa Cruz Biotechnology sc-6029 goat polyclonal

ChIP and WCE DNA was then purified and genomic libraries were prepared using the ChIP-Seq Sample Prep Kit (Illumina 1003473) according to the manufacturers protocol (Illumina 11257047) for selecting library fragments between 200 and 350 bp. Samples were run using the GA2X genome sequencer (SCS v2.6, pipeline 1.5).

Brn2 and Sox2 ChIP in differentiating ESCs

ESCs at day 2 of differentiation were cross-linked and harvested as above. The same Sox2 and Brn2 antibodies were used. Approximately $5x10^7$ formaldhyde-crosslinked cells were lysed and as above on an IP-Star (Diagenode). Chromatin was sonicated on the Bioruptor (Diagenode) to an average size of 0.2-1

kb. Sox2 ChIP was performed with 3 µg of antibody (above) using the IP-Star Automated System (Diagenode) and 2.5% of chromatin was used for each whole cell extract (WCE). Following reversal of crosslinks, sample and WCE DNA was purified. ChIP DNA was dissolved in water and placed on the SPRI-TE (Beckman Coulter) for Illumina sample preparation. ChIP-Seq and WCE libraries were barcoded and multiplexed on the HiSeq 2000 (Illumina).

Sox2 ESC

Approximately $5x10^7$ formaldhyde-crosslinked ESCs were lysed and as above on an IP-Star (Diagenode). Chromatin was sonicated on the Bioruptor (Diagenode) to an average size of 0.2-1 kb. Sox2 ChIP was performed with 3 µg of antibody (above) using the IP-Star Automated System (Diagenode) and 2.5% of chromatin was used for each whole cell extract (WCE). Following reversal of crosslinks, sample and WCE DNA was purified. ChIP DNA was dissolved in water and placed on the SPRI-TE (Beckman Coulter) for Illumina sample preparation. ChIP-Seq and WCE libraries were barcoded and multiplexed on the HiSeq 2000 (Illumina).

ChIP-Seq Data Analysis

Images acquired from the Illumina/Solexa sequencer were processed using the bundled Solexa image extraction pipeline. Sequences were aligned using Bowtie software (http://bowtie-bio.sourceforge.net/index.shtml) using murine genome NCBI Build 36 and 37 (UCSC mm8) as the reference genome with default settings for mismatch tolerance, non-unique mapping events, etc. Analysis of our sequence data was done based on previous models (4, 5). Sequences were extended +200 bp for transcription factors and allocated in 25 bp bins (1.05×10^8 bins total). Statistically significant enriched bins were identified using a Poissonian background model, generally with a p-value threshold of 10^{-8} to minimize false positives. We used an empirical background model (whole cell extracts (WCE)) that require genomic bins to be enriched at least 5 fold above background to correct for non-random enrichment observed previously. Genes with enriched regions within 1kb of their start sites ware called bound at their promoters, while enriched regions which were within 1 kb of peaks of H4K4me1 or H3K27Ac (Creyghton et al., 2010) called bound to enhancers and were assigned the closest gene up or downstream.

Gene ontology

Gene ontology analysis was performed using GOstat (http://gostat.wehi.edu.au/cgi-bin/goStat.pl) or the GREAT algorithm (http://great.stanford.edu/) as indicated. GOSTAT was performed using the mgi

(mouse) GO annotation database. For enhancer locations from each cell type we selected the closest gene either upstream or downstream. GREAT analysis was preferred using mm9 bound regions obtained by lift-over of mm8 called regions, using the Galaxy web tool (*main.g2.bx.psu.edu/).*

Gene Plots

Single gene ChIP-Seq density plots were generated using the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway). Wiggle files were generated from ChIP-Seq reads and density was normalized to reads-per-million. Published datasets were also used to correlate Sox2 to epigenetic marks (Mikkelsen et al., 2007; Creyghton et al., 2010). These wiggle files were uploaded to the genome browser and tracks of ChIP-Seq density at selected loci were downloaded and formatted for inclusion in the manuscript.

Unbiased Motif Search

MEME *(*meme.sdsc.edu/) was used to find DNA sequences enriched in Sox2 bound regions in ESCs and NPCs. Plus/minus 75 base pairs surrounding peaks of minimum peak height of 100 of Sox2 enrichment above were input into MEME and motif logos were generated from position weight matrices obtained.

RNA isolation and microarray analysis

Three biological replicates for V6.5 ESCs and V6.5-derived NPCs were prepared, hybridized to arrays, and analyzed independently, then normalized and compared to yield final expression values.

RNA was isolated using Trizol reagent (Invitrogen) according to the manufacturer's protocol and DNAse treated using the DNA-Free RNA kit (Zymo Research R1028). Samples were then prepared for Affymetrix GeneChip Expression Array analysis. 5 µg total RNA was used to prepare biotinylated cRNA according to the manufacturer's protocol (Affymetrix One Cycle cDNA Synthesis Kit). Briefly, this method involves SuperScript II-directed reverse transcription using a T7-Oligo(dT) Promoter Primer to create first strand cDNA. RNase H-mediated second strand cDNA synthesis is followed by T7 RNA Polymerase directed *in vitro* transcription, which incorporates a biotinylated nucleotide analog during cRNA amplification.

Samples were prepared for hybridization using 15 µg biotinylated cRNA in a 1X hybridization cocktail according the Affymetrix hybridization manual. Additional hybridization cocktail components were provided in the Affymetrix GeneChip Hybridization, Wash and Stain Kit. GeneChip arrays (Mouse 430) were hybridized in a GeneChip Hybridization Oven at 45°C for 16 hours at 60 RPM. Washing was done using a GeneChip Fluidics Station 450 according to the manufacturer's instructions, using the buffers provided in the Affymetrix GeneChip Hybridization, Wash and Stain Kit. Arrays were scanned

on a GeneChip Scanner 3000 and images were extracted and analyzed using GeneChip Operating Software v1.4.

Biological replicates were RMA normalized using updated annotation from the BrainCDF the site remapped from Ensembl Gene ID to Gene Name from Biomart, and the mean intensity for each probe across three arrays was calculated. Maximum probe mean values for each gene were taken as gene expression levels. Box and Violin plots were constructed depicting median values as center line, and bottom and top of the box representing the 25th and 75th percentiles, respectively. Whiskers depict + 1.5*IQR (interquartile range) for top, -1.5*IQR for the bottom.

Correlation of ChIP-Seq Datasets

Comparison of ChIP-Seq datasets was performed using a similarity metric based on a correlation coefficient (Bilodeau et al., 2009). This analysis generates a correlation coefficient between zero and one reflecting the similarity of genomic regions occupied in two datasets. A value of one would reflect total overlap between two datasets while a value of zero would reflect overlap expected by random association. Values in Figure S1 were generated in this way. The matrix in Figure 2C was generated by clustering these data using hierarchical clustering along the horizontal and vertical axis with an average linkage similarity metric using the software Cluster3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/). This matrix was visualized using Java TreeView (http://jtreeview.sourceforge.net/).

Transgenic ES cell generation

ES cells which inducibly overexpress mouse Brn2 were generated using the "flp-in" system described previously (Beard et al., 2004). Briefly, Brn2, cDNA were cloned into pBS31, which contains the PGK promoter followed by an ATG start codon and an frt site, a splice acceptor-double polyA cassette, the tetracycline operator followed by a CMV minimal promoter, a unique EcoRI site for subcloning, and an SV40 polyA signal. These constructs were electroporated into KH2 ES cells, which harbor an M2-rtTA gene into the ROSA26 locus a construct containing a frt-flanked PGK-neomycin-resistance gene followed by a promoterless, ATG-less hygromycin-resistance gene downstream of the Col1a1 locus. To flip-in the tetracyclin-responsive Brn2 constructs into these cells, pBS31 plasmids was co-transfected with a plasmid expressing the FlpE recombinase, and ES cells were selected for hygromycin resistance. Cells were treated with the tetracycline analog doxycycline to achieve induced expression.

Brn induced neural differentiation of ES cells to NP cells

Dox inducible Brn2 expressing ES cells and rtTA only ES cells were passaged off feeders in ES medium plus 2 μg/ml Dox. Twenty four hours later the medium was switched to N2B27 plus Dox. Cells were

fixed with 4% paraformaldehyde in PBS and stained with anti-Nestin (Developmental Hybridoma Bank) and DAPI, and assayed by Quantitative Real-Time PCR (below) at 24 hour intervals.

Quantitative Real-Time PCR

Trizol-isolated RNA from three biologically independent samples for each sample was purified, DNAse treated (DNA free RNA Kit, Zymo Research) then reverse transcribed using a First Strand Synthesis Kit (Invtirogen). cDNA was analyzed by quantitative PCR analysis performed in technical triplicate using an ABI Prism 7000 (Applied Biosystems) with Platinum SYBR green qPCR SuperMix-UDG with ROX (Invitrogen). The following primers were used:

	Forward (5'-3')	Reverse (5'-3')
Oct4	ACATCGCCAATCAGCTTGG	AGAACCATACTCGAACCACATCC
Sox2	ACAGATGCAACCGATGCACC	TGGAGTTGTACTGCAGGGCG
Nestin	TCTACAGGCAGCGCTAACAGTC	TCCCCTAACTCATCTGCCTCAC
Sox1	AGTGGAAGGTCATGTCCGAGG	GCCAGCGAGTACTTGTCCTTCTT
Brn2	AGCAGTTCGCCAAGCAATTC	CGAGAACACGTTGCCGTACA
GapDH	TTCACCACCATGGAGAAGGC	CCCTTTTGGCTCCACCCT
Apba2	ACGGACAGAGTGTGGTAGCC	CGTGAGGAGCCTAAACATGG
Cops2	CTGATGTGGAGAGCTTGCTG	CACCCCTCTTCTGATGATCC
Id4	ACTCACCCTGCTTTGCTGAG	AGAATGCTGTCACCCTGCTT
Irx3	AGTGCCTTGGAAGTGGAGAA	CGTCCAGATGGTTCTGTGG
Irx5	ACAGAAGCCCGAGGACAAG	TAAAATCCGAGTCGCTGAGG
Jag2	CGTGGCTGCTATCACTCAGA	AGCCACAGCACACTGAACAC
Kirrel3	TGTGCCATCCCTGAATATGA	TGCTCTCCTGAGAGGTGGTT
Lemd1	ATTCACACCTGGCCCAATAC	TCACTGTCATCGCTGTCCTC
Lrrn1	GTCCTCATCCTCCGGCTAGT	CACAGGTCCTTGTGGGAGTT
NfIX	CTTTGTGACGTCTGGGGTCT	TGTAGTAGCTGGGGCTCTCC
PDGFC	GTGGAGGAAATTGTGCCTGT	CCCTTGACTCCAGTTTTTGG
Sox21	TGGTGTTTGCTTTGCACTTC	GGAGGGAGGAAGGATGAGAC
Tcf12	CCGTGGCAGTCATCCTTAGT	GCTGACGCAGCAGAGACTTT
Zic1	CCTTTGCAAGATGTGCGATA	CTGTGAGCCCTGAGAAGAGG

Data were extracted from the lineage range, and the standard curve method was used to obtain relative expression values. Technical replicates were averaged, then biological replicates were averaged. Statistical significance was determined using Graphpad Prism to perform an ANOVA with Bonnferroni Correction for multiple testing.

Overrepresented motifs within Sox2 bound sites.

Sox2-bound regions (in ES and NP cells) regions were examined for motif overrepresentation. Sox2-bound regions were defined as 100 base pair windows around the peak. We used a hypothesis-based approach to identify known protein-DNA recognition elements enriched in each dataset. The set of hypotheses are derived from all vertebrate position specific scoring matrices (PSSMs) from TRANSFAC (Wingender et al., 1996) filtered for sufficient information content (IC>8 total bits). The final set of motifs was preclustered based on pairwise distance by KL-divergence of the PSSMs using Affinity Propagation. The TAMO programming environment (Gordon et al., 2005) were used to store the PSSMs and calculate the max motif score for each sequence across all *k*-mers in the sequence for a motif of width *k*). Overrepresentation of motifs in a foreground set of sequences was assessed against a background set of randomly generated sequences which matched the size and sequence complexity of the test set using the Mann-Whitney Wilcoxon ranked sum test. For each independent motif test, the U statistic and sample sizes were used to calculate a p-value for motif enrichment in the foreground set of sequences relative to background and the area under the curve for the receiver operating characteristic (AUC ROC). For each foreground set of sequences, a background was selected to maximally capture undesirable biases in the foreground set of sequences such as GC content, CpG content and TSS proximity. Out of each cluster, the statistics and logo for the most highly enriched motif by minimum p-value was presented.

Genome-wide distances between Sox2 and cofactors

Distances between Sox2 bound sites and cofactor bound sites in ESCs and NPCs were calculated as follows. Overlapping regions of Sox2 and POU factors were defined as regions with at least 1-bp of overlap. Peaks from these overlapping regions were then used to define distances between factor binding. In particular, we calculated distances between Sox2:Brn1 site pairs (NPCs), Sox2: Brn2 site pairs (NPCs), and Sox2: Oct4 site pairs (ESCs). Site pairs were defined by matching each Sox2 bound site to the closest cofactor bound site within 200 base. Distance was calculated as the cofactor chromosomal coordinate subtracted from the Sox2 chromosomal coordinate.

Spacing between motif matches.

Spacing between Sox and Oct family binding was also determined using a motif-based approach to determine whether spatial arrangement of the motifs are exclusive to particular Sox2:POU pairs. In particular, the Sox2:Oct4 (ESCs), Sox2:Brn1 (NPCs), and Sox2:Brn2 (NPCs) co-bound regions as defined as aforementioned were examined. Max motif scores were calculated as aforementioned and normalized as in Equation 1. Motif matches to Sox were defined as normalized scores greater than 0.85 to

a general Sox TRANSFAC matrix, M01308. Similarly, Oct family motif matches were defined as normalized scores greater than 0.85 to a general Oct TRANSFAC matrix, M00342.

For each sequence *i* and motif *j*, a motif score s_{ij} : $s_{ij} = \frac{s_i - \min(s_j)}{\max(s_j) - \min(s_j)}$ (1)

Spacing was defined as the number of base positions between the Oct4 and Sox2 motif matches relative to the Sox2 motif match (positive if Oct4 was 3' to Sox2, negative if Oct4 was 5' to Sox2). For each sequence, the pair of motif matches could lie in a forward or reverse orientation. Each pair of motif matches was designated as forward orientation if both matches were either located on the forward or reverse strand. Each pair of motif matches was designated as reverse orientation if one of the matches was located on the forward strand and the other on the reverse strand. For example

Motif	Trace	Spacing	Explanation
CTTTGTT-N-ATGCAAAT	Green	+1	Sox is 5' of Oct, both in same 5' to 3' orientation
ATGCAAAT-N-CTTTGTT	Green	-1	Sox is 3' of Oct, both in same 5' to 3' orientation
AACAAAG-N-ATGCAAAT	Blue	+1	Sox is 5' of Oct, Sox is reverse compliment

References

Ambasudhan, R., M. Talantova, et al. (2011). "Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions." Cell Stem Cell.

Ambrosetti, D. C., C. Basilico, et al. (1997). "Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites." Molecular and Cellular Biology **17**(11): 6321.

Ambrosetti, D. C., H. R. Schöler, et al. (2000). "Modulation of the Activity of Multiple Transcriptional Activation Domains by the DNA Binding Domains Mediates the Synergistic Action of Sox2 and Oct-3 on the Fibroblast Growth Factor-4Enhancer." Journal of Biological Chemistry **275**(30): 23387.

Andersen, B., M. Schonemann, et al. (1993). "Brn-5 is a divergent POU domain factor highly expressed in layer IV of the neocortex." Journal of Biological Chemistry **268**(31): 23390.

Arnold, K., A. Sarkar, et al. (2011). "Sox2+ Adult Stem and Progenitor Cells Are Important for Tissue Regeneration and Survival of Mice." Cell Stem Cell **9**(4): 317-329.

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

Bani-Yaghoub, M., R. G. Tremblay, et al. (2006). "Role of Sox2 in the development of the mouse neocortex." Developmental Biology **295**(1): 52-66.

Bartkowska, K., A. Paquin, et al. (2007). "Trk signaling regulates neural precursor cell proliferation and differentiation during cortical development." Development **134**(24): 4369-4380.

Bergsland, M., D. Ramsköld, et al. (2011). "Sequentially acting Sox transcription factors in neural lineage development." Genes & Development **25**(23): 2453-2464.

Botquin, V., H. Hess, et al. (1998). "New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2." Genes & Development **12**(13): 2073-2090.

Boutz, P. L., P. Stoilov, et al. (2007). "A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons." Genes & Development **21**(13): 1636.

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

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

Bulger, M. and M. Groudine (2011). "Functional and mechanistic diversity of distal transcription enhancers." Cell **144**(3): 327-339.

Bylund, M., E. Andersson, et al. (2003). "Vertebrate neurogenesis is counteracted by Sox1-3 activity." Nature Neuroscience **6**(11): 1162-1168.

Catena, R., C. Tiveron, et al. (2004). "Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells." Journal of Biological Chemistry **279**(40): 41846-41857.

Cau, E., G. Gradwohl, et al. (2000). "Hes genes regulate sequential stages of neurogenesis in the olfactory epithelium." Development **127**(11): 2323.

Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-655.

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

Cimpean, A., S. Encica, et al. (2010). "692 Expression and potential role of SOX2 gene in human thymus and thymomas." EJC Supplements **8**(5): 174-174.

Collignon, J., S. Sockanathan, et al. (1996). "A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2." Development **122**(2): 509.

Creyghton, M. P., A. W. Cheng, et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proceedings of the National Academy of Sciences **107**(50): 21931.

Cui, H. and R. F. Bulleit (1997). "Expression of the POU transcription factor Brn-5 inhibits proliferation of NG108-15 cells." Biochemical and biophysical research communications **236**(3): 693-696.

Cui, H. and R. F. Bulleit (1998). "Expression of the POU transcription factor Brn-5 is an early event in the terminal differentiation of CNS neurons." J Neurosci Res **52**(6): 625-632.

Dailey, L. and C. Basilico (2001). "Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes." Journal of cellular physiology **186**(3): 315-328.

D'Amour, K. A. and F. H. Gage (2003). "Genetic and functional differences between multipotent neural and pluripotent embryonic stem cells." Proceedings of the National Academy of Sciences **100**(Suppl 1): 11866.

Domyan, E. T., E. Ferretti, et al. (2011). "Signaling through BMP receptors promotes respiratory identity in the foregut via repression of Sox2." Development **138**(5): 971-981.

Eminli, S., J. Utikal, et al. (2008). "Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression." Stem Cells **26**(10): 2467-2474.

Favaro, R., M. Valotta, et al. (2009). "Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh." Nature Neuroscience **12**(10): 1248-1256.

Ferletta, M., D. Caglayan, et al. (2011). "Forced expression of Sox21 inhibits Sox2 and induces apoptosis in human glioma cells." International Journal of Cancer.

Ferri, A. L. M., M. Cavallaro, et al. (2004). "Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain." Development **131**(15): 3805-3819.

Fradkin, L. G., G. Garriga, et al. (2005). "Wnt signaling in neural circuit development." The Journal of

Neuroscience **25**(45): 10376.
Gontan, C., A. de Munck, et al. (2008). "Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation." Developmental Biology **317**(1): 296-309.
Gordon, D. B., L. Nekludova, et al. (2005). "TAMO: a flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs." Bioinformatics **21**(14): 3164-3165.
Graham, V., J. Khudyakov, et al. (2003). "SOX2 functions to maintain neural progenitor identity." Neuron **39**(5): 749-765.
Haney, C. A., Z. Sahenk, et al. (1999). "Heterophilic binding of L1 on unmyelinated sensory axons mediates Schwann cell adhesion and is required for axonal survival." The Journal of Cell Biology **146**(5): 1173.
Hara, Y., A. C. Rovescalli, et al. (1992). "Structure and evolution of four POU domain genes expressed in mouse brain." Proceedings of the National Academy of Sciences **89**(8): 3280.
He, X., M. N. Treacy, et al. (1989). "Expression of a large family of POU-domain regulatory genes in mammalian brain development."
Heintzman, N. D., G. C. Hon, et al. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.
Heintzman, N. D., R. K. Stuart, et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nature Genetics **39**(3): 311-318.
Hochedlinger, K., Y. Yamada, et al. (2005). "Ectopic Expression of Oct-4 Blocks Progenitor-Cell Differentiation and Causes Dysplasia in Epithelial Tissues." Cell **121**(3): 465-477.
Iwafuchi-Doi, M., Y. Yoshida, et al. (2010). "The Pou5f1/Pou3f-dependent but SoxB-independent regulation of conserved enhancer N2 initiates Sox2 expression during epiblast to neural plate stages in vertebrates." Developmental Biology.
Jin, Z., L. Liu, et al. (2009). "Different transcription factors regulate nestin gene expression during P19 cell neural differentiation and central nervous system development." Journal of Biological Chemistry **284**(12): 8160.
Kamachi, Y., K. S. E. Cheah, et al. (1999). "Mechanism of regulatory target selection by the SOX high-mobility-group domain proteins as revealed by comparison of SOX1/2/3 and SOX9." Molecular and Cellular Biology **19**(1): 107.
Kamachi, Y., S. Sockanathan, et al. (1995). "Involvement of SOX proteins in lens-specific activation of crystallin genes." The EMBO journal **14**(14): 3510.
Kamachi, Y., M. Uchikawa, et al. (2000). "Pairing SOX off: with partners in the regulation of embryonic development." Trends in Genetics **16**(4): 182-187.
Kiernan, A. E., A. L. Pelling, et al. (2005). "Sox2 is required for sensory organ development in the mammalian inner ear." Nature **434**(7036): 1031-1035.
Kim, J. B., B. Greber, et al. (2009). "Direct reprogramming of human neural stem cells by OCT4." Nature **461**(7264): 649-643.
Kim, J. B., V. Sebastiano, et al. (2009). "Oct4-induced pluripotency in adult neural stem cells." Cell **136**(3): 411-419.
Kuhlbrodt, K., B. Herbarth, et al. (1998). "Cooperative function of POU proteins and SOX proteins in glial cells." Journal of Biological Chemistry **273**(26): 16050.
Le, N., R. Nagarajan, et al. (2005). "Analysis of congenital hypomyelinating Egr2Lo/Lo nerves identifies Sox2 as an inhibitor of Schwann cell differentiation and myelination." Proceedings of the National Academy of Sciences **102**(7): 2596.
Lee, T. I., R. G. Jenner, et al. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-313.
Liber, D., R. Domaschenz, et al. (2010). "Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage." Cell Stem Cell **7**(1): 114-126.
Loh, Y. H., Q. Wu, et al. (2006). "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells." Nature Genetics **38**(4): 431-440.

Lujan, E., S. Chanda, et al. (2012). "Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells." Proceedings of the National Academy of Sciences.
Mansukhani, A., D. Ambrosetti, et al. (2005). "Sox2 induction by FGF and FGFR2 activating mutations inhibits Wnt signaling and osteoblast differentiation." The Journal of Cell Biology **168**(7): 1065-1076.
Marson, A., S. S. Levine, et al. (2008). "Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells." Cell **134**(3): 521-533.
McEvilly, R. J., M. O. de Diaz, et al. (2002). "Transcriptional regulation of cortical neuron migration by POU domain factors." Science **295**(5559): 1528.
Mitsui, K., Y. Tokuzawa, et al. (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-642.
Miyagi, S., S. Masui, et al. (2008). "Consequence of the loss of Sox2 in the developing brain of the mouse." FEBS letters **582**(18): 2811-2815.
Morest, D. K. and J. Silver (2003). "Precursors of neurons, neuroglia, and ependymal cells in the CNS: What are they? Where are they from? How do they get where they are going?" Glia **43**(1): 6-18.
Nakai, S., H. Kawano, et al. (1995). "The POU domain transcription factor Brn-2 is required for the determination of specific neuronal lineages in the hypothalamus of the mouse." Genes & Development **9**(24): 3109-3121.
Nishimoto, M., S. Miyagi, et al. (2005). "Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus." Molecular and Cellular Biology **25**(12): 5084.
Ohtsuka, T., M. Sakamoto, et al. (2001). "Roles of the basic helix-loop-helix genes Hes1 and Hes5 in expansion of neural stem cells of the developing brain." Journal of Biological Chemistry **276**(32): 30467-30474.
Okabe, S., K. Forsberg-Nilsson, et al. (1996). "Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro." Mechanisms of Development **59**(1): 89-102.
Okamoto, K., M. Wakamiya, et al. (1993). "A novel class of murine POU gene predominantly expressed in central nervous system." Journal of Biological Chemistry **268**(10): 7449-7457.
Okubo, T., L. H. Pevny, et al. (2006). "Sox2 is required for development of taste bud sensory cells." Genes & Development **20**(19): 2654.
Pang, Z. P., N. Yang, et al. (2011). "Induction of human neuronal cells by defined transcription factors." Nature **476**(7359): 220-223.
Park, J., B. Liu, et al. (2008). "Disruption of Nectin-like 1 cell adhesion molecule leads to delayed axonal myelination in the CNS." The Journal of Neuroscience **28**(48): 12815-12819.
Phillips, K. and B. Luisi (2000). "The virtuoso of versatility: POU proteins that flex to fit1." Journal of Molecular Biology **302**(5): 1023-1039.
Rada-Iglesias, A., R. Bajpai, et al. (2010). "A unique chromatin signature uncovers early developmental enhancers in humans." Nature **470**(7333): 279-283.
Ramos, C., S. Rocha, et al. (2010). "Two Notch ligands, Dll1 and Jag1, are differently restricted in their range of action to control neurogenesis in the mammalian spinal cord." PLoS One **5**(11): e15515.
Reiprich, S., J. Kriesch, et al. (2010). "Activation of Krox20 gene expression by Sox10 in myelinating Schwann cells." Journal of Neurochemistry **112**(3): 744-754.
Reményi, A., K. Lins, et al. (2003). "Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers." Genes & Development **17**(16): 2048-2059.
Reynolds, B. A. and S. Weiss (1992). "Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system." Science **255**(5052): 1707.
Rocha, S. F., S. S. Lopes, et al. (2009). "Dll1 and Dll4 function sequentially in the retina and pV2 domain of the spinal cord to regulate neurogenesis and create cell diversity." Developmental Biology **328**(1): 54-65.

Salmon-Divon, M., H. Dvinge, et al. (2010). "PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci." BMC Bioinformatics **11**: 415.
Sandberg, M., M. Kallstrom, et al. (2005). "Sox21 promotes the progression of vertebrate neurogenesis." Nature Neuroscience **8**(8): 995-1001.
Smale, S. T. (2010). "Pioneer factors in embryonic stem cells and differentiation." Curr Opin Genet Dev **20**(5): 519-526.
Sugitani, Y., S. Nakai, et al. (2002). "Brn-1 and Brn-2 share crucial roles in the production and positioning of mouse neocortical neurons." Genes & Development **16**(14): 1760.
Takahashi, K., K. Tanabe, et al. (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-872.
Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.
Tanaka, S., Y. Kamachi, et al. (2004). "Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells." Molecular and Cellular Biology **24**(20): 8834.
Tomioka, M., M. Nishimoto, et al. (2002). "Identification of Sox-2 regulatory region which is under the control of Oct-3/4–Sox-2 complex." Nucleic acids research **30**(14): 3202-3213.
Treacy, M. N. and M. G. Rosenfeld (1992). "Expression of a family of POU-domain protein regulatory genes during development of the central nervous system." Annual review of Neuroscience **15**(1): 139-165.
Wegner, M. (2011). "SOX after SOX: SOXession regulates neurogenesis." Genes & Development **25**(23): 2423-2428.
Wernig, M., A. Meissner, et al. (2007). "In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state." Nature **448**(7151): 318-324.
Wingender, E., P. Dietze, et al. (1996). "TRANSFAC: a database on transcription factors and their DNA binding sites." Nucleic Acids Research **24**(1): 238-241.
Wu, R., M. Jurek, et al. (2001). "The POU gene Brn-5 is induced by neuregulin and is restricted to myelinating Schwann cells." Molecular and Cellular Neuroscience **17**(4): 683-695.
Xian, C. J. and X. F. Zhou (2004). "EGF family of growth factors: essential roles and functional redundancy in the nerve system." Front Biosci **9**: 85-92.
Yasuhara, N., N. Shibazaki, et al. (2006). "Triggering neural differentiation of ES cells by subtype switching of importin-α." Nature Cell Biology **9**(1): 72-79.
Ying, Q. L., M. Stavridis, et al. (2003). "Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture." Nature Biotechnology **21**(2): 183-186.
Zhao, Y., K. M. Kwan, et al. (2007). "LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum." Proceedings of the National Academy of Sciences **104**(32): 13182-13186.

Supplemental Chapter: Sox2 controls neural development from the embryonic stem cell state to the neural precursor cell state

Michael A. Lodato[1,2], Kevin Thai[1], Albert W. Cheng[1,2], Laurie A. Boyer[1], Rudolf Jaenisch[1,2]

[1]Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue Cambridge, MA 02139 USA; [2]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142 USA

Chapter Synopsis

In this Supplemental Chapter, I will provide further analysis of Sox2-occupied regions in ESCs and NPCs. As mentioned in Chapter 2, Sox2 occupied promoter and enhancer regions in ESCs and NPCs, and these regions were largely distinct, exhibiting cell type specificity. I will show more detailed analysis of Sox2-bound promoters and enhancers in these cell types here. Further, while Sox2-bound regions were almost mutually exclusive at the global level in ESCs and NPCs, I will highlight key regions of overlap between these datasets. Specifically, Sox2-occupied regions proximal to promoters were more likely to be occupied in both ESCs and NPCs that the group of bound regions as a whole. Further I identified a subset of promoters which were bound by Sox2 at distinct sites in ESCs and NPCs, suggesting that Sox2 was using distinct sites in the same promoter to regulate common genes. Surprisingly, though Sox2-bound enhancer regions were highly divergent, a large fraction of the genes associated with Sox2 binding through enhancers in ESCs were linked to new, distinct, Sox2-occupied enhancers in NPCs. These genes tended to be involved in neural development. These data suggested that Sox2 was utilizing distinct enhancers to regulate a set of genes involved in neural development from pluripotent ESCs to multipotent NPCs, and was thus regulating neural development from very early in embryogenesis though neural commitment.

Author Contributions

M.A.L, L.A.B., and R.J conceived experiments and interpreted data. M.A.L and A.W.C. analyzed microarray data. M.A.L and K.T. performed Sox2 ESC ChIP-Seq. M.A.L. performed all other experiments, analyzed data, and wrote the chapter.

Results

Sox2 binding to promoter regions in ESCs and NPCs

Sox2 is a master regulator of two distinct cell states: the pluripotent ESC state and the multipotent NPC state. In order to ascertain how one transcription factor could control these two cellular programs, we mapped the genome-wide binding profile of Sox2 in ESCs and NPCs using chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-Seq) to define genes and networks of genes which may be regulated by this factor. We performed ChIP-Seq analysis on Sox2 in ESCs, and identified 13,717 Sox2 enriched regions throughout the genome using a Poissonian error model (Marson et al. 2008; Methods). This dataset was highly correlated to a previously published one (Marson et al., 2008), indicating that it is of high quality (Chapter 2). To probe Sox2 binding in NPCs, we derived these cells *in vitro* using established protocols (Okabe et al., 1996) from genetically identical ESCs as used in the above analysis. Sequencing data of the Sox2-enriched DNA fragments from two biological replicates were highly correlated (Chapter 2). This analysis identified 16,685 enriched regions in NPCs. Using these data we then sought to define genes regulated by Sox2 in ESCs and NPCs.

Given that proximal promoter regions are known to play a crucial role in gene regulation, we examined Sox2-bound regions within 1 kb of annotated transcriptional start sites in ESCs and NPCs to gain insight into genes which may be regulated by Sox2. We confirmed that Sox2 occupied known targets in ESCs such as the *Nanog* and *Pou5f1* loci, demonstrating that Sox2 regulates other regulators of pluripotency (Figure 1a). At loci encoding regulators of NPC identity, such as the *Hes5* and *Dll1* loci, (Cau et al., 2000; Ohtsuka et al., 2001; Rocha et al., 2009; Ramos et al., 2010) Sox2 binding was observed in NPCs. Generally, Sox2 targets were highly expressed in the cell type in which they were bound (Figure 1b). While many Sox2 target genes in ESCs were expressed, Sox2 is known to occupy poised developmental regulators in ESCs which are repressed and marked by trimethylation of Lysine 27 on Histone H3 (H3K27me3), a repressive chromatin mark catalyzed by Polycomb group proteins (Loh et al., 2005; Boyer et al., 2005, 2006; Lee et al., 2006). In NPCs, the overlap between Sox2 and H3K27me3

was significantly smaller than in ESCs (Figure 1c). This is in agreement with the fact that while ESCs must keep developmental programs for all somatic tissue types competent to be activated, NPCs have more restricted developmental potential and thus can maintain fewer poised developmental programs. Unbiased MEME motif analysis revealed that the Sox2-bound regions in NPCs frequently harbored an E-box motif (Figure 1d). This motif is recognized by the bHLH family or transcription factors, including the Myc subfamily and mediators of Notch signaling such as Hes1 and Hes5. c-Myc is a potent activator of transcription, specifically through the ability of c-Myc to release paused RNA polymerase II at transcriptional start sites and thereby allowing elongation to occur (Rahl et al., 2010). N-Myc, L-Myc, and c-Myc are expressed in NPCs and important for NPC identity and lineage commitment (Bernard et al., 1992; Kuwahara et al., 2010; Wey et al., 2010). There is a small but statistically significant difference in the expression level of Sox2 bound genes in ESCs and NPCs (Figure 1b), which may be related to the higher overlap between Sox2 and H3K27me3 in ESCs than in NPCs, and the prevalence of an E-Box motif in many Sox2 bound regions in NPCs, which may be bound by strong activators such as c-Myc. Taken together, these data indicate that Sox2 is associated with a set of active promoters in ESCs and NPCs with distinguishing sequence and epigenetic features.

Figure 1

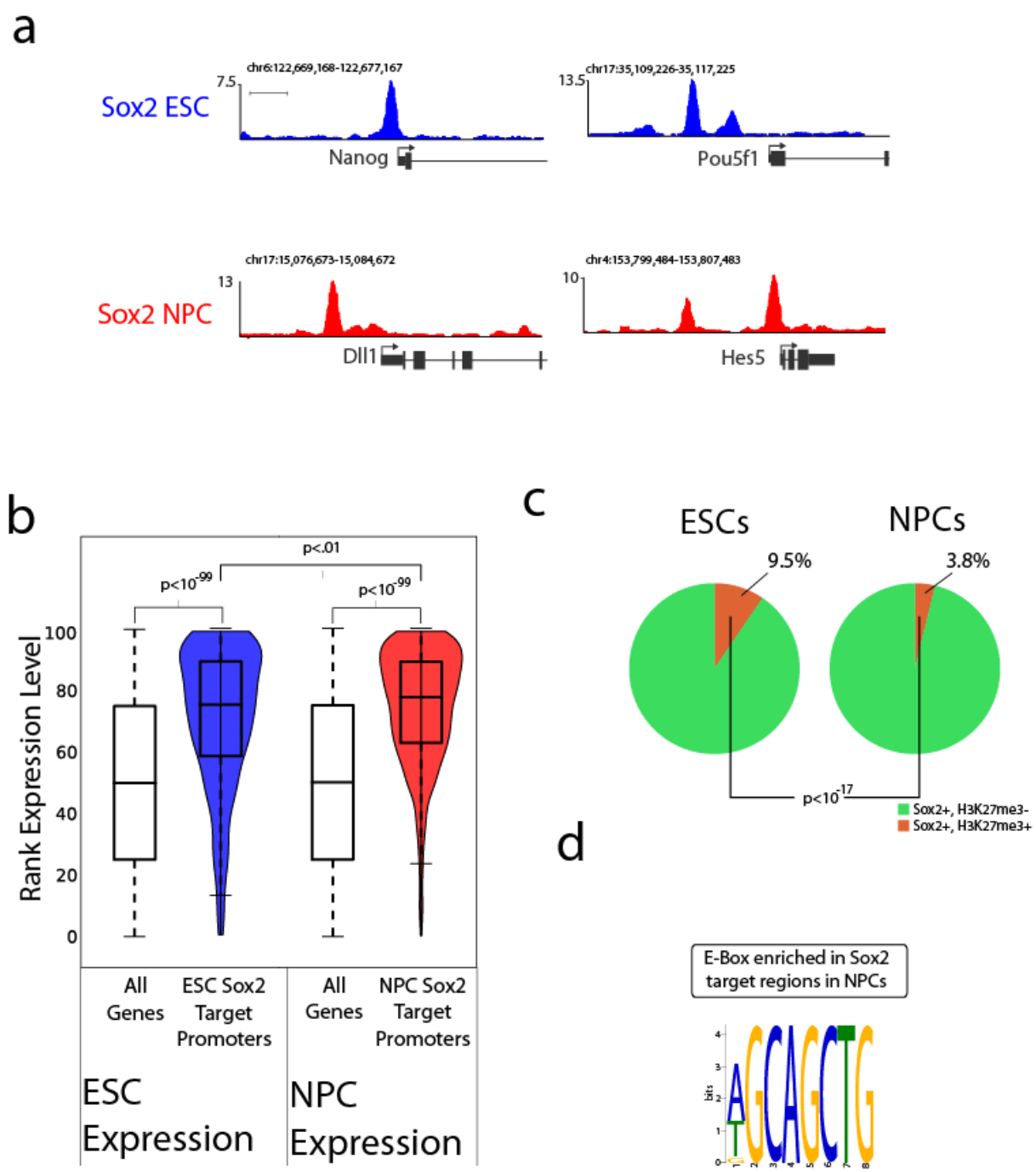


Figure 1 Sox2 binding at promoter regions. a. Gene plots depicting Sox2 ChIP-Seq density at indicated regions (mm8 genome). y-axis reflects reads per million. b. Box and violin plots representing gene expression values for genes linked to Sox2-bound promoters. p-values reflect non-parametric Mann-

Whitney U test values. c. Fraction of Sox2 bound promoters which are also marked by H3K27me3. p-value reflects hypergeometric distribution. d. E-box motif enriched in Sox2 bound regions in NPCs.

To define the functional classes to which Sox2-bound promoter target genes belonged, we used the GREAT algorithm (McLean et al., 2010), a gene ontology tool which tests ChIP-Seq bound regions for the function of the nearest genes using binomial test over the input genomic regions, incorporating annotations from twenty ontologies (Figure 2). In ESCs, we find Sox2 was associated with both developmental regulators and a large number of broadly expressed housekeeping genes, including the promoters of core histone genes. These genes included canonical regulators of pluripotency such as *Sox2* itself, *Oct4*, and *Nanog*, and epigenetic regulators such as *Mbd3*, *Set*, and *Smarc2*. In NPCs, Sox2 also occupied the promoters of lineage specific and housekeeping genes. Many were highly expressed and known regulators of the NPC state, such as Pax6, Rfx4 and again Sox2 itself. The largest group of genes regulated by Sox2 in NPCs included regulators of RNA processing. This group includes genes involved in mRNA capping, poly-adenylation, RNA editing, and tRNA modification, with the largest subgroup comprised of genes involved in mRNA splicing. Sox2 binds to promoters of over 200 genes encoding both general splicing factors, such as components of the splicesome, as well as developmental and tissue specific regulators of splicing such as Mbln1 and Mbln2, and PTB and nPTB, which are known to play a crucial role in the development of neurons from NPCs (Boutz et al., 2007). Interestingly, a large group of the housekeeping genes regulated by Sox2 in NPCs are confirmed targets of c-Myc in Burkitt's lymphoma **(**Ben-Porath et al., 2008). Thus, Sox2 in NPCs regulates lineage specific and housekeeping genes, possibly by partnering with c-Myc. Therefore, in both cell types, Sox2 plays the dual role of regulating cell-type specific genes and widely expressed housekeeping genes.

Figure 2

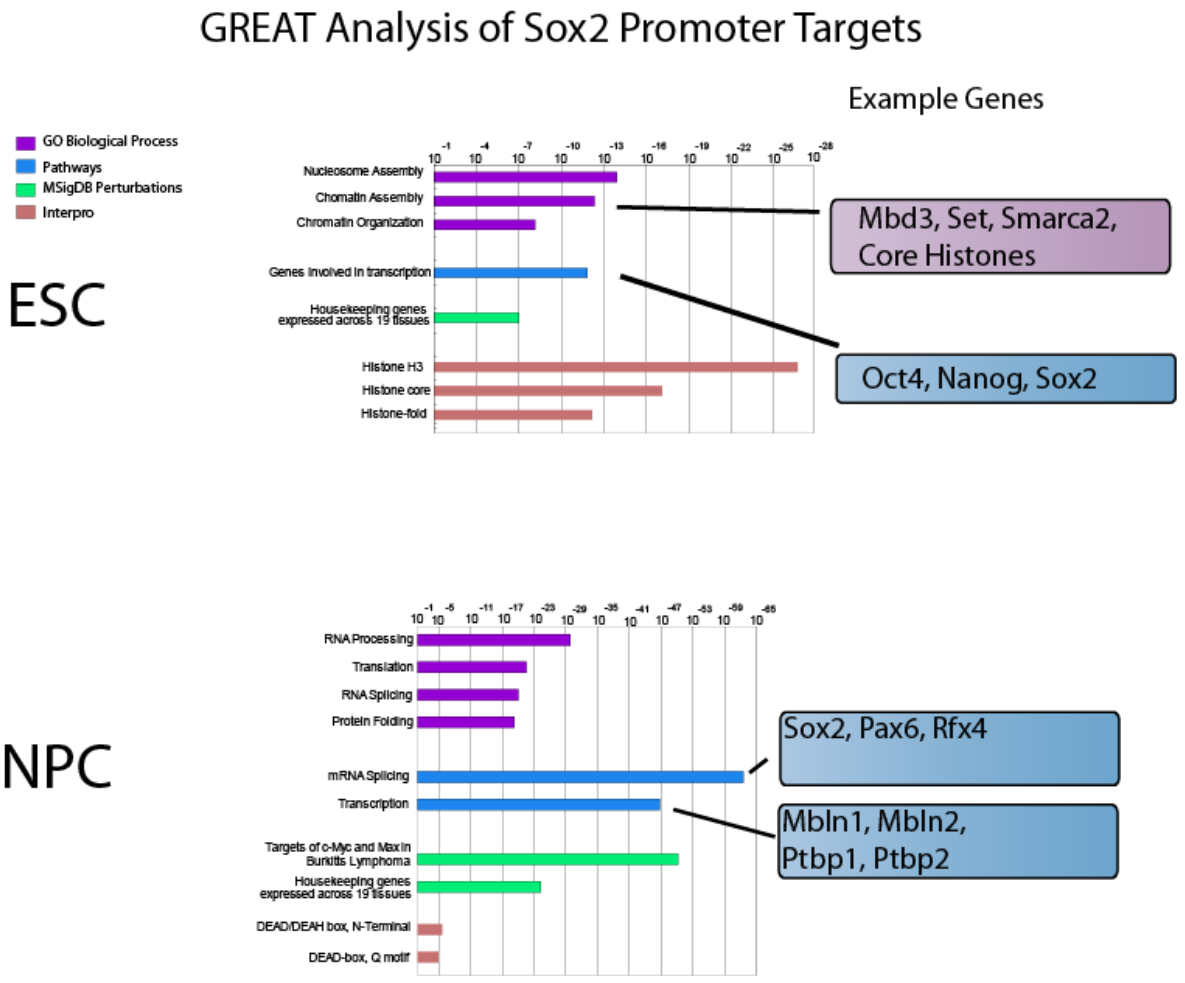


Figure 2 GREAT analysis of genes linked to Sox2 bound promoters. Classes of genes bound by Sox2 in ESCs (top) and NPCs (bottom). x-axis reflects binomial raw p-value of enrichment of given category against a whole genome background. Example genes highlighted on the right.

Sox2 Binding at Distal Enhancers

Interestingly, the majority of Sox2-bound regions in both ESCs and NPCs were not within 1 kb of known TSSs (Figure 3a). Recently it has been shown that H3K4me1 and H3K27Ac mark thousands of poised (H3K4me1+/H3K27Ac+) and active (H3K4me1+/H3K27Ac+, H3K4me1-/H3K27Ac+)

enhancers in the genome of many cell types, including ESCs and NPCs. In order to ascertain whether Sox2 binds near these elements, we surveyed 1 kb up and downstream of peaks of H3K4me1 and H3K27Ac enrichment for Sox2-bound regions in ESCs and NPCs. Indeed, we find that Sox2 binds to more regions modified with H3K4me1 and/or H3K27Ac than annotated start sites (Figure 3b). For example, in ESCs Sox2 marks H3K4me1+/H3K27Ac+ regions upstream of the Nanog promoter and downstream of its transcriptional stop, in addition to binding the Nanog promoter. Sox2 also is associated with H3K4me1+/H3K27Ac+ regions upstream and within an intron of Jam2, a cell adhesion molecule which is thought to be part of a stem cell gene expression signature shared by ESCs, NPCs, and hematopoietic stem cells (Ivanova et al., 2006) (Figure 3c). In NPCs, Sox2 overlaps with H3K4me1 and H3K27Ac downstream of Ascl1, a key regulator of neurogenesis, and upstream and within an intron of Notch1, a well-characterized regulator of NPCs identity (Figure 3d). Thus, our data are consistent with the notion that Sox2 regulates a larger set of target genes through distal enhancer elements.

Figure 3

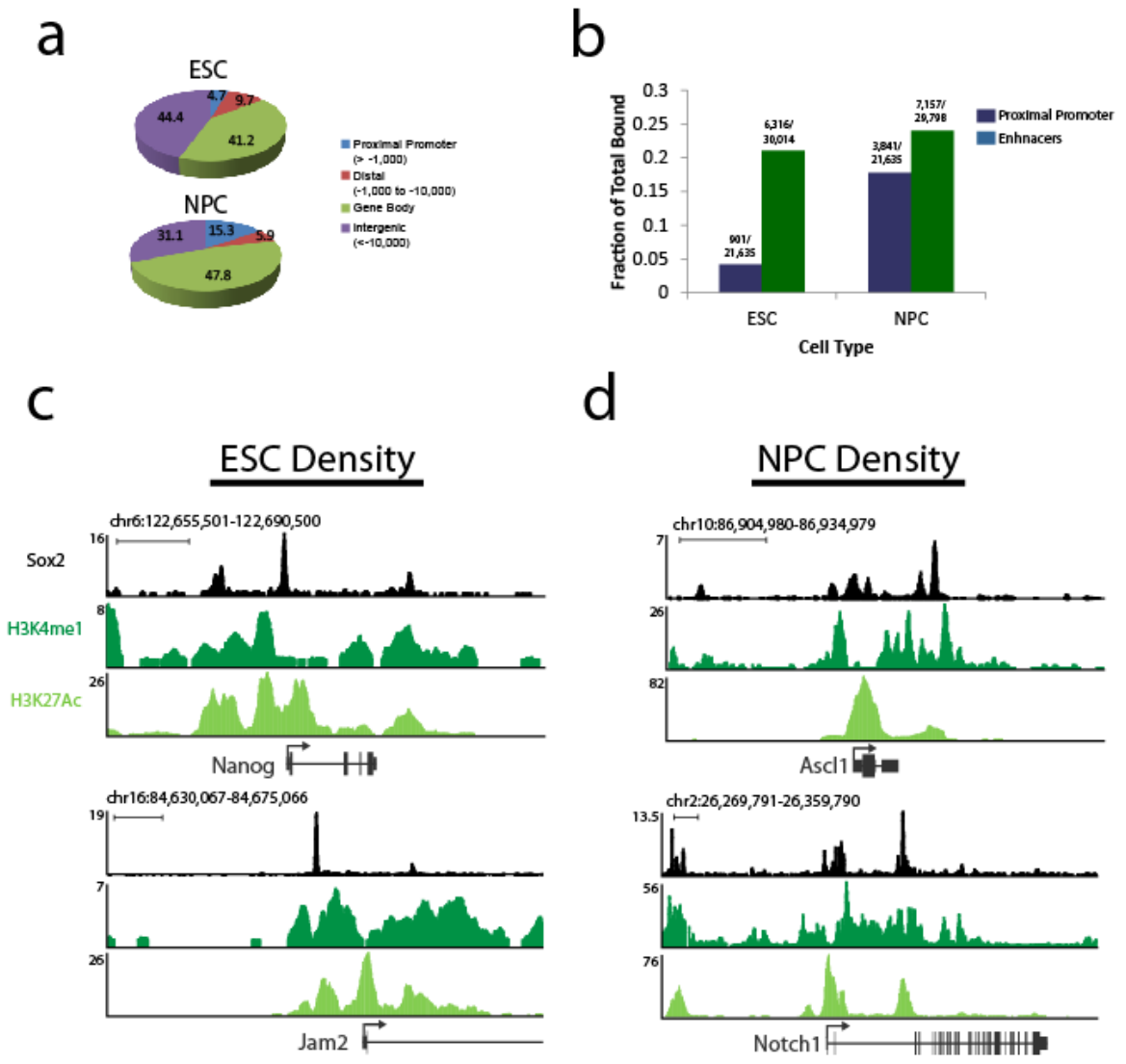


Figure 3 Sox2 binds more enhancer regions than promoters in ESCs and NPCs. a. Location analysis of Sox2 bound regions relative to known TSSs. b. Comparison of number of Sox2-bound regions which map to promoters versus marked (H3K4me1 and/or H3K27Ac) enhancers, plotted as fraction of total promoters and enhancers in the genome. Numbers above bars reflect raw numbers of bound regions. c,d. Gene plots reflecting Sox2 density at indicated loci (mm8 genome). y-axis reflects reads per million.

We used nearest neighboring gene analysis to predict the genes associated with putative enhancer regions bound by Sox2, as previously described (Heintzman et al., 2009, Visel et al., 2009, Creyghton et al., 2010). In both cell types, Sox2 was associated with more poised enhancers than active enhancers

(Figure 4a). As expected, genes linked to Sox2+, poised enhancers were expressed at a lower level on average than genes linked to Sox2+, active enhancers (Figure 4b,c). Since the genes in this class were lowly expressed in the cell type in which they were bound by Sox2, we sought to define in which tissues these genes would become expressed. We used the GREAT algorithm to compare the closest set of genes near poised Sox2 target enhancers in both cell types to gene expression signatures across mouse development from the MGI database, and we found that in ESCs, genes linked to poised, Sox2+ enhancers are expressed at later stages of early embryonic development, while in NPCs Sox2+ poised genes are expressed in later stages of neural development. (Figure 4d, e). Then, we used GREAT analysis to define the gene ontology categories and pathways to which Sox2-bound enhancers belonged. This revealed that genes bound by Sox2 through enhancers in ESCs are involved in early morphogenesis, axis specification, stem cell differentiation, and development of the neural lineage (Figure 5). In NPCs, Sox2 bound enhancers are linked to genes involved in neural development. Genes closest to H3K27Ac+ enhancers are involved in NPC identity, and include transcription factors, genes involved in stem cell differentiation and neurogenesis, and members of the Cdc42 pathway (Figure 5). Genes linked to poised enhancers tend to be involved in later neural development and the morphogenesis of brain regions such as the cerebellum and hippocampus. They also include regulators of many major signaling pathways in neural development, such as the Reelin pathway that is responsible for migration of immature neurons during cortical morphogenesis (Honda et al., 2011), and CamKII and NMDA signaling pathways, which are involved in the transmission of signals between mature neurons (Wang et al., 2004). Thus, Sox2 occupies enhancers linked to genes involved in early developmental steps from the ESC state through the NPC state, in which Sox2 supports NPC function through binding to active enhancers and may prime certain loci to be expressed upon terminal differentiation though binding poised enhancers.

Figure 4

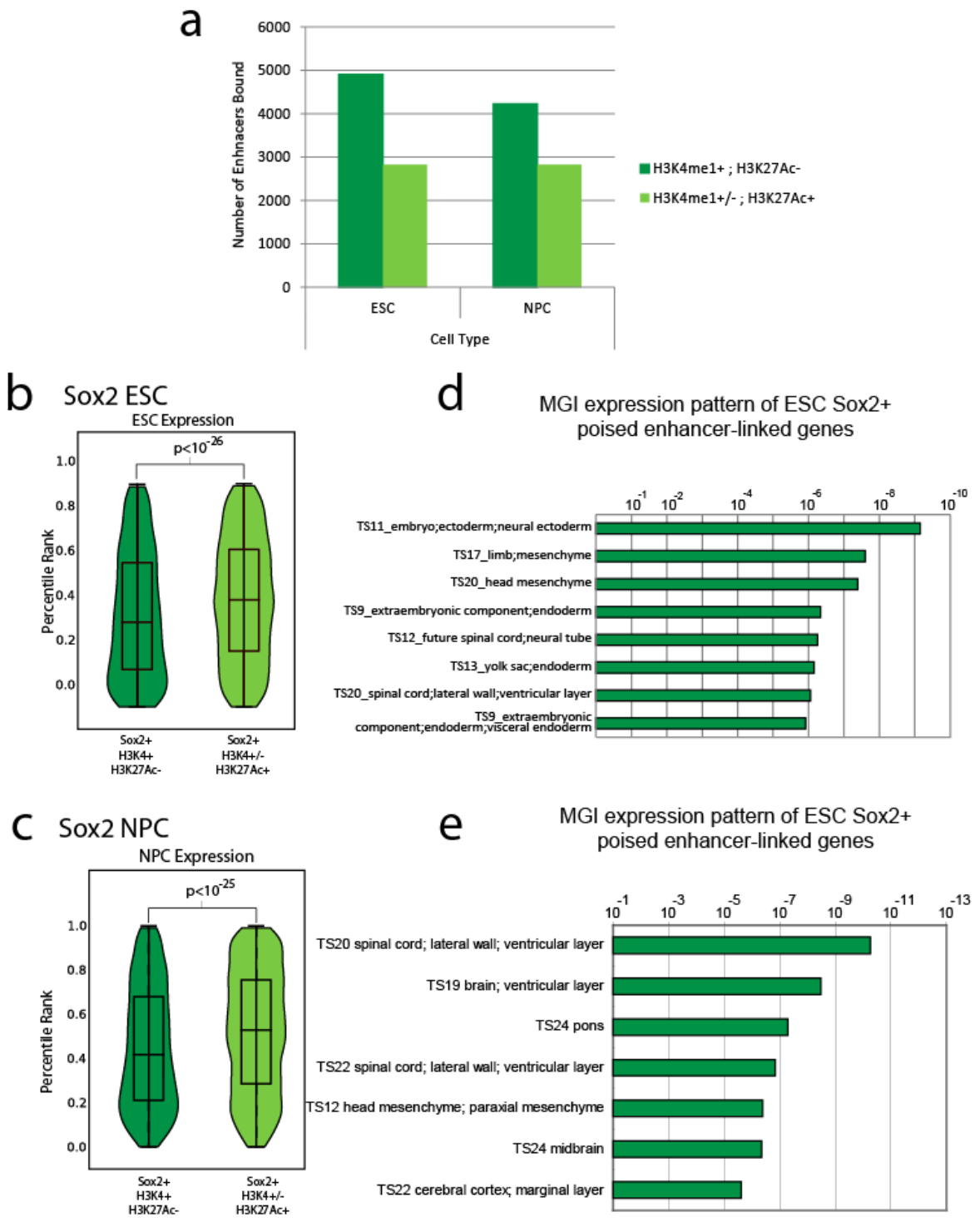


Figure 4 Expression analysis of Sox2 bound promoters. a. Comparison of number of Sox2-bound regions associated with poised (H3K4me1+/H3k27Ac-) enhancers versus active (H3K4me1+/H3K27Ac+

and H3K4me1-/H3K27Ac+) enhancers. b,c. Box and violin plots representing gene expression values for genes linked to Sox2-bound poised and active enhancers in ESCs (top) and NPCs (bottom). p-values reflect non-parametric Mann-Whitney U test values. d,e. MGI expression analysis of genes linked to poised enhancers in ESCs (top) and NPCs (bottom). x-axis reflects raw binomial p-value of enrichment vs. a whole genome background set.

Figure 5

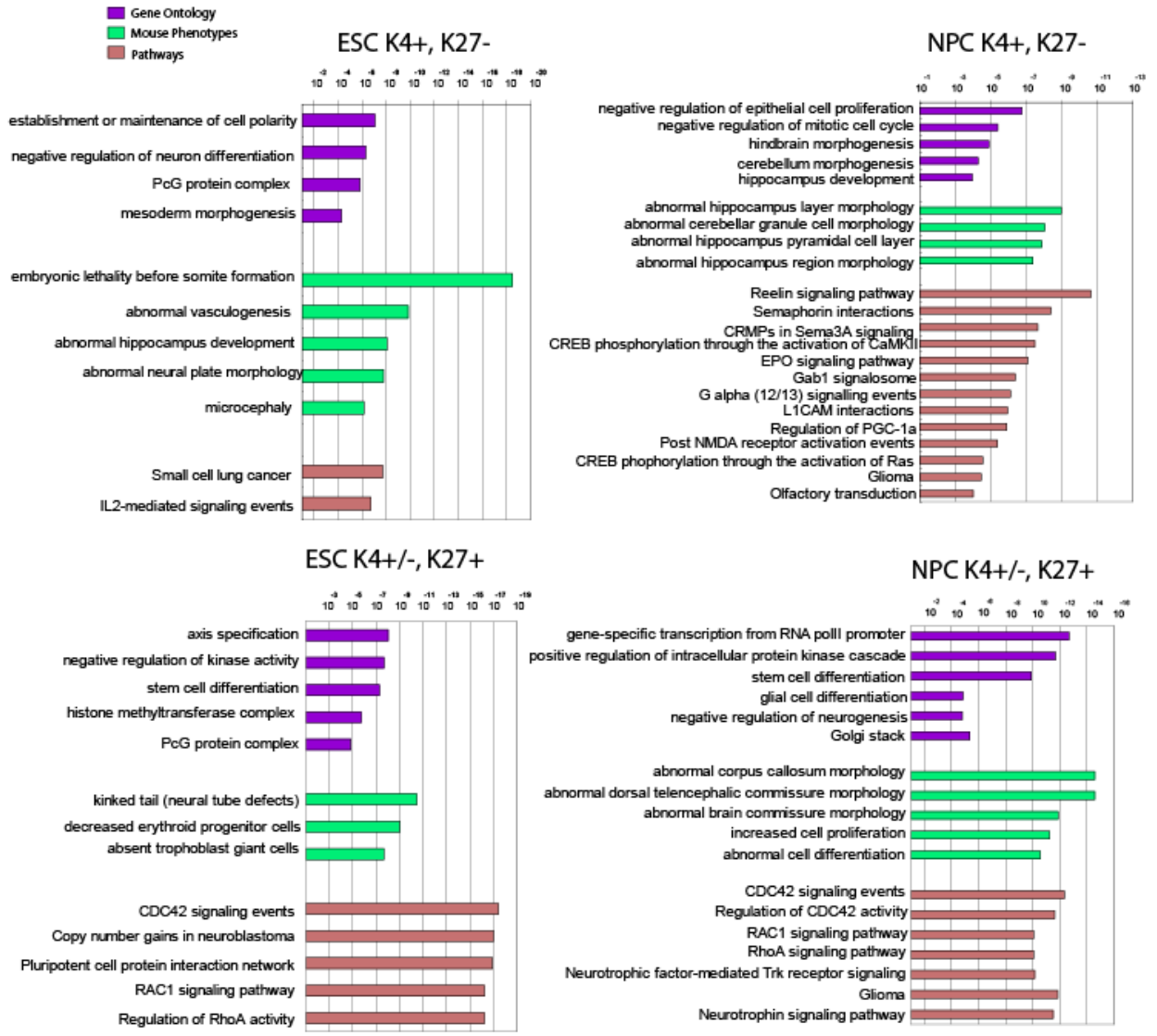


Figure 5 GREAT analysis of genes linked to Sox2 bound enhancers. Classes of genes linked to poised (top) and active (bottom) enhancers bound by Sox2 in ESCs (left) and NPCs (right). x-axis reflects raw binomial p-value of enrichment vs. a whole genome background set.

Common targets of Sox2 across development

Sox2 was associated with genes involved in neural development and housekeeping functions in both cell types, but surprisingly very few regions bound by Sox2 were common between the datasets; approximately 9% of regions bound by Sox2 in ESCs were bound in NPCs, and less than 5% of Sox2

bound regions were common between ESCs and NPCs (Figure 6a, Table 1). In spite of this, we sought to define any commonalities between these targets to uncover mechanisms by which Sox2 might regulate ESCs and NPCs in a similar manner. In fact, of the few common Sox2 bound regions between ESCs and NPCs, many occur within promoter regions (Figure 6b). Approximately 38% of Sox2-bound regions near TSSs in ESCs were also occupied by Sox2 in NPCs. Further, at the gene level, we see that approximately 57% of the genes bound by Sox2 in ESCs are also bound by Sox2 in NPCs, including loci where Sox2 is bound the same site and loci in which Sox2 is associated with the same gene at different sites in the promoter (Figure 6c). These genes tend to be highly expressed in both ESCs and NPCs (Figure 6d), and encode epigenetic and transcriptional regulators, regulators of the cell cycle, and a lowly enriched group of regulators of nervous system development (Figure 6e). Thus, Sox2 regulates a subset of promoters in both ESCs and NPCs, sometimes utilizing different binding sites to do so, and these genes encode transcriptional, epigenetic, and cell cycle regulators which are involved in housekeeping and development of the neural lineage.

Figure 6

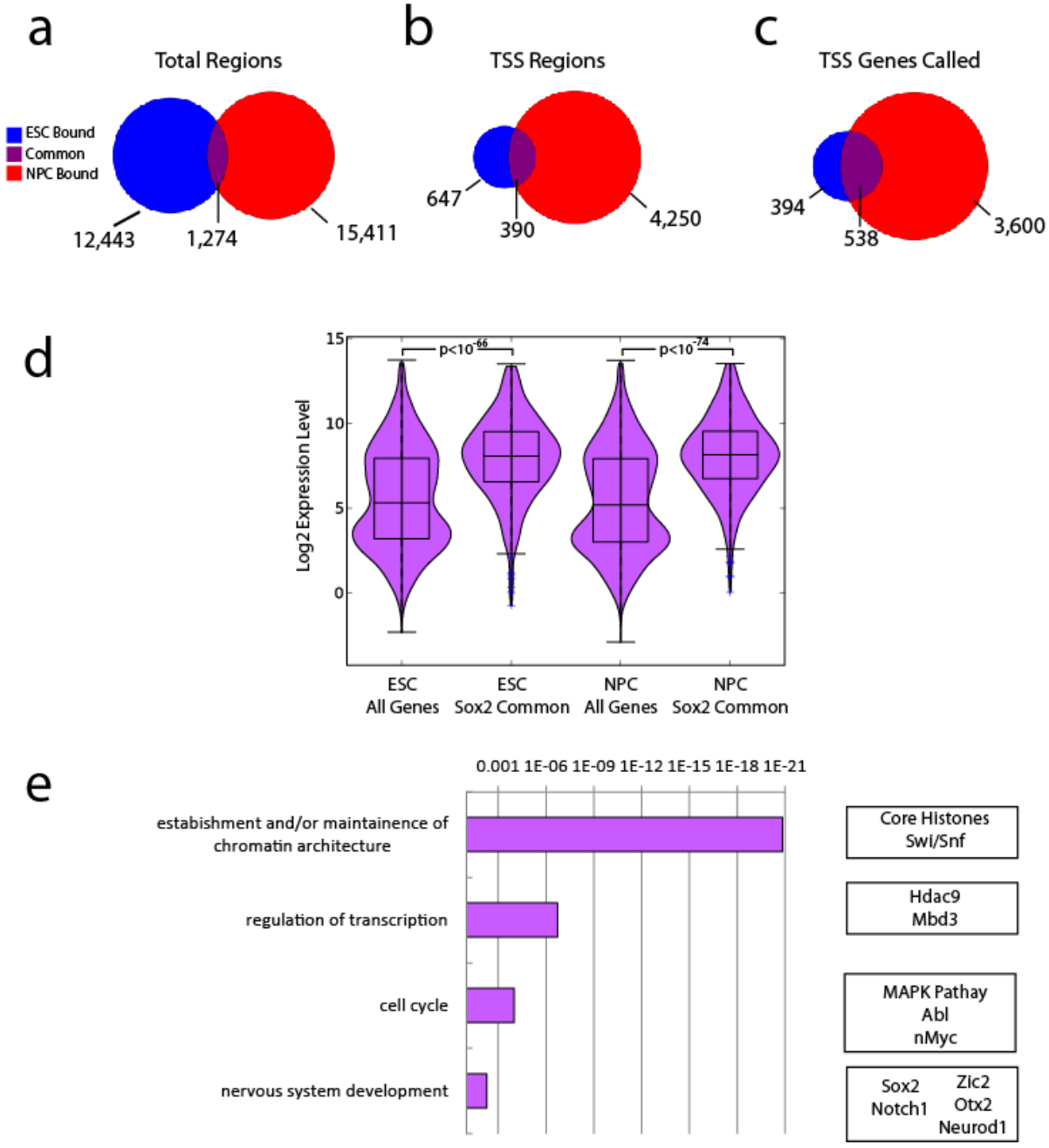


Figure 6 - Many promoters bound by Sox2 in ESCs remain bound in NPCs. a. Overlap of all Sox2 bound regions in ESCs (blue) and NPCs (red). b. Overlap of Sox2-bound regions within 1 kb of annotated start sites in ESCs and NPCs. c. Overlap of genes called bound by Sox2 at promoters, at overlapping and distinct sites within 1kb of known TSSs. d. Box and violin plots representing gene expression values for genes linked to Sox2-bound promoters common to both ESCs and NPCs. p-values

reflect non-parametric Mann-Whitney U test values. e. GOSTAT gene ontology analysis of common targets. Example genes from each category are shown on right.

At enhancers, genome-wide trend of low overlap between Sox2 in ESCs and Sox2 in NPCs hold true; less than 4% of Sox2 bound enhancers in ESCs are bound by Sox2 in NPCs (Figure 7a). Surprisingly, 43% of genes associated with a Sox2-bound enhancer in ESCs were also associated with a Sox2-bound enhancer in NPCs (Figure 7b). Thus, distinct, cell-type specific enhancers occupied by Sox2 in ESCs and NPCs were often associated with the same nearest gene. For example, at the Id2, Pax6, and Sox2 loci, Sox2 occupied multiple regions upstream and downstream of the TSS, but these regions differed depending on the cellular context (Figure 7c). When we separated these enhancers based on their chromatin state, we found that genes linked to Sox2-bound poised enhancers in ESCs acquired new Sox2 enhancers in either poised and active states in NPCs (Figure 7c). Conversely, genes linked to Sox2-bound active enhancers in ESCs were associated with new Sox2–bound poised or active enhancers in NPCs (Figure 7d). Thus, Sox2 occupied a distinct set of enhancer regions in ESCs and NPCs, but many of the genes linked to these enhancers were common between the two cell types.

Figure 7

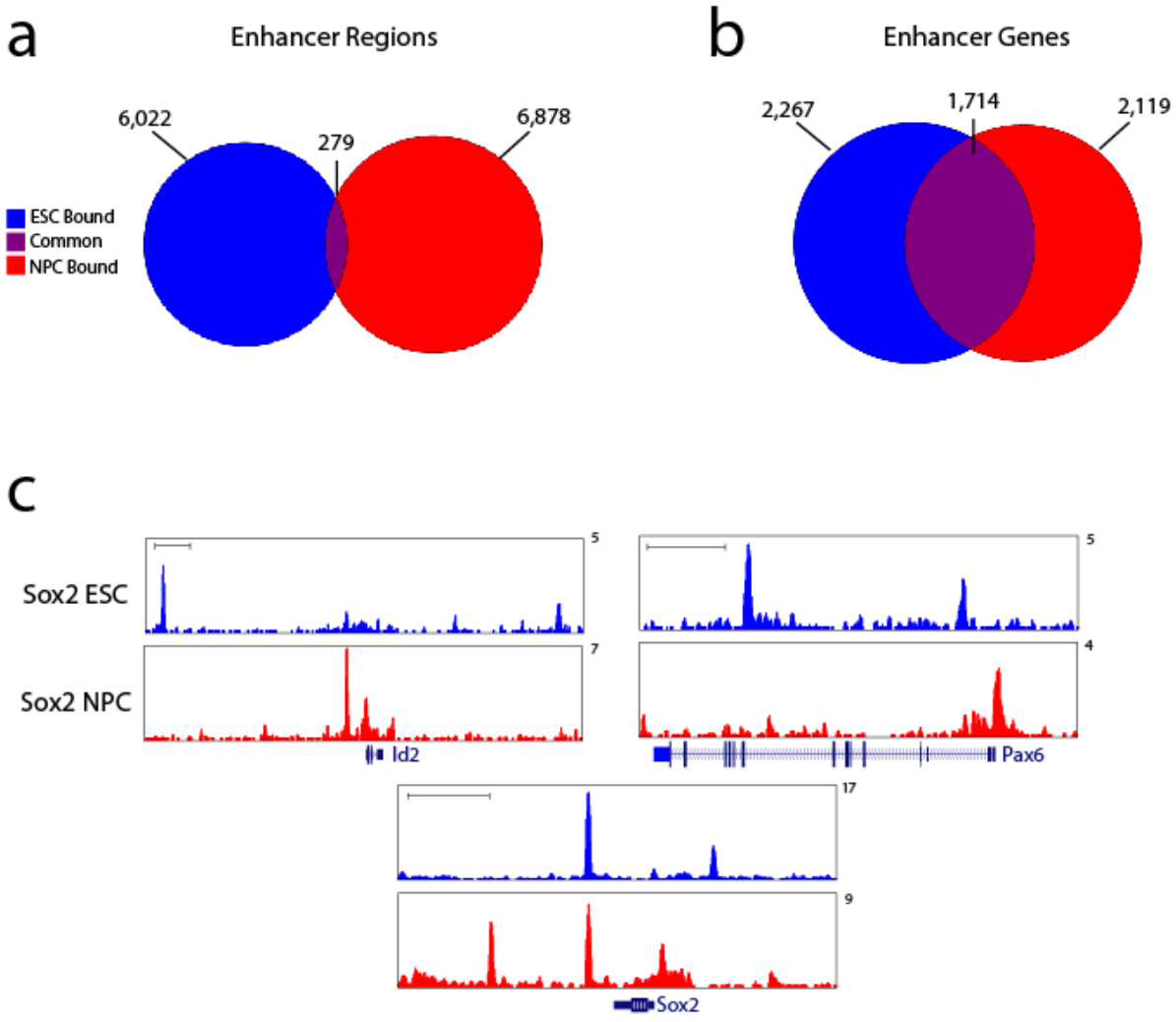

a
Enhancer Regions
6,022
279
6,878
ESC Bound
Common
NPC Bound
b
Enhancer Genes
2,267
1,714
2,119
c
Sox2 ESC
Sox2 NPC
Id2
Pax6
Sox2
d
ESC Poised → NPC
317
330
923
NPC Poised
NPC Active
ESC Only
e
ESC Active → NPC
613
454
1,344

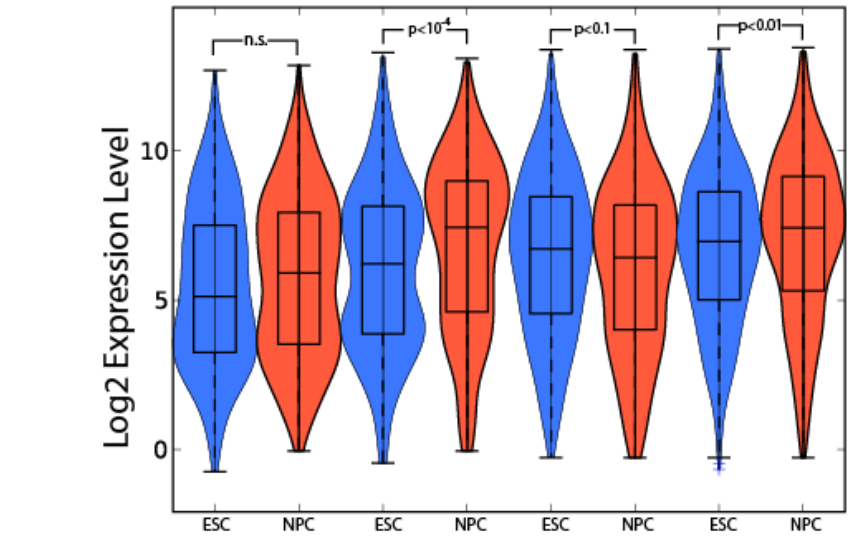
Figure 7 - Differential binding of Sox2 at enhancers in ESCs and NPCs. a. Overlap of Sox2-bound regions associated with enhancers. b. Overlap of nearest gene to Sox2-bound enhancers. c. Gene plots depicting Sox2 ChIP-Seq density at indicated loci (mm8 genome). Sox2 is associated with the same locus but at different sites in ESCs and NPCs in many cases. y-axis reflects reads per million. d. Pie charts depicting fraction of genes linked to poised (left) and active (right) Sox2-bound enhancers in ESCs which become associated with new Sox2-bound poised enhancers, Sox2-bound active enhancers, or no new Sox2-bound enhancers in NPCs

Next, we sought to understand the function of genes associated with distinct sets of Sox2-bound enhancers in both cell types commonly. Genes linked to Sox2-poised enhancers in ESCs and NPCs were lowly expressed in both cell types (Figure 8a) and tend to code for genes involved in differentiation past the NPC state, such as the proneural bHLH transcription factor Math1 and the Lim family homeobox transcription factors Lhx5 and Lhx8 (Figure 8b) (Zhao et a., 1999; Gowan et al., 2001; Zhao et a., 2007; Miquelajauregui et al., 2010). Genes which switch from poised, Sox2+ enhancers to active, Sox+ enhancers are more highly expressed in NPCs than ESCs (Figure 8a) and encode for regulators of NPC function such as EGF receptor and cell signaling molecules involved in the cytoskeleton such as Focal-Adhesion Kinase (Fak) (Figure 8c) (Reynolds and Weiss, 1992; Brown et al., 2005). Putative Sox2 targets through active ESC enhancers are also maintained in NPCs through both poised and active Sox2+ enhancers. Genes which switch from active to poised decrease in expression level from ESCs to NPCs (Figure 8a) and include signaling molecules and transcription factors such as Otx2, which is involved in both early specification of pluripotent cells to ectoderm and the differentiation of NPCs to specific neural subtypes (Figure 8d) (Vernay et al., 2005). Finally there is a subset of genes associated with Sox+ active enhancers in both ESCs and NPCs, which are expressed in both cell types but expressed at a higher level in NPCs (Figure 8a). These genes and encode transcription factors involved in both ESC and NPC identity such as Sox2 itself (Figure 8e). Interestingly, genes associated with active, Sox2+ enhancers in ESCs involved in cell migration split into two categories in NPCs, some being linked to active, Sox2+ enhancers and other linked to poised, Sox2+ enhancers. While NPCs must keep cell migration/differentiation and cell adhesion/stem cell maintenance in equilibrium to ensure proper

development of the CNS, the cells of the ICM do not maintain a pool of themselves as development proceeds, and in fact all migrate to new positions in the epiblast and finally in one of three germ layers soon after the ICM is formed, suggesting that cell migration pathways might be mostly active in ESCs while becoming more sensitively regulated in NPCs. Our data suggest that Sox2 might be central in achieving this balance. Thus, Sox2 is associated with poised and active enhancer elements in both ESCs and NPCs, and in general these enhancers are linked to genes which progressively regulate the development of the nervous system.

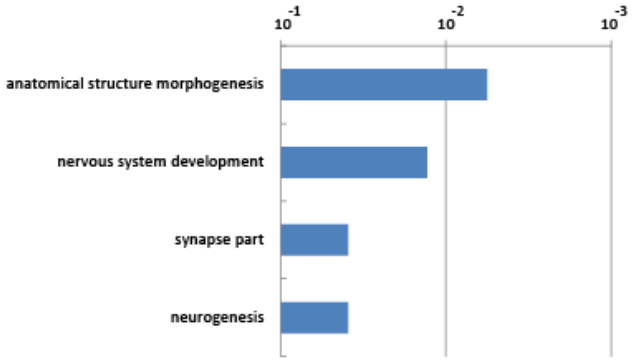
Figure 8

a

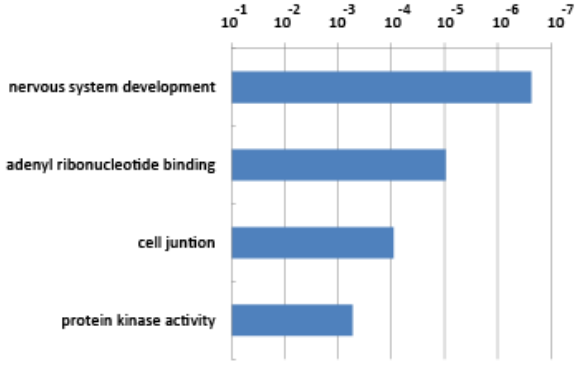


ESC Enhancer	Poised	Poised	Active	Active
NPC Enhancer	Poised	Active	Poised	Active

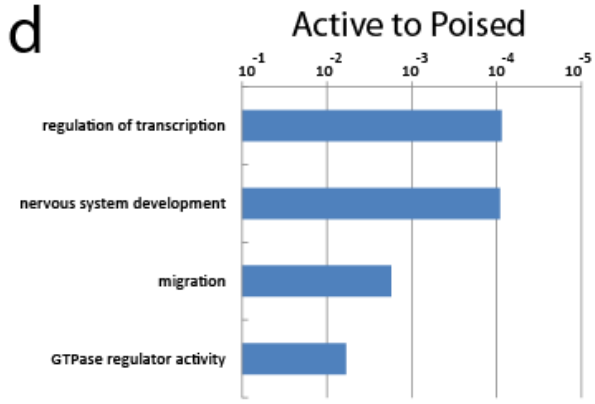
b



c



d



e

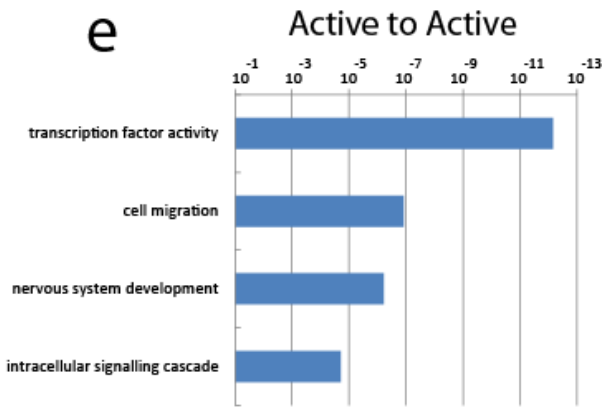


Figure 8 - Characterization of Sox2+ enhancer associated genes. a. Box and violin plots representing gene expression values for genes linked to Sox2-bound enhancers grouped by change or maintenance of enhancer state between cell types. p-values reflect non-parametric Mann-Whitney U test values. b-e. GOSTAT gene ontology analysis of genes in above classes.

Discussion

In this study, we investigated the genome-wide binding profiles of the HMG-box transcription factor Sox2 in two cell types in which it is thought be a master regulator of cell identity: pluripotent embryonic stem cells and multipotent neural precursor cells. How this master regulator can control cellular identity in two distinct stem cell types has been an unsolved question for some time. In ESCs, Sox2 was known to occupy genes involved in ESC identity and lineage commitment factors which would become active only later in development, indicating that it could act as a pioneer factor (Liber et al., 2010). Recently, analysis of Sox3 in NPCs revealed that a subset of Sox2 target regions in ESCs linked to repressed genes involved in neural precursor identity were bound by Sox3 in NPCs (Bergsland et al., 2011). This study also suggested that Sox2 may co-occupy many Sox3 bound sites in NPCs. These genes in turn fell into two classes: those which were expressed in NPCs, and those which were repressed but destined to be activated later in development. The latter group was shown to be bound by Sox11 as NPCs completed their differentiation towards immature neurons. Thus, a subset of regulators of nervous system development was "handed off" between Sox factors from pluripotent ESC though multipotent NPCs, and finally to immature neurons. This study advanced our understanding of Sox function in neural development, but left open many questions, including how Sox2 could function as both an activator of transcription at some loci and a pioneer factor at others. Along those lines, can distinguishing features of Sox2-bound regions be identified which differentiate between these two functions?

To begin to answer these open questions, we performed ChIP-Seq on Sox2 in ESCs and NPCs. We confirmed that in ESCs, Sox2 occupied the promoters of genes encoding transcription factors and epigenetic regulators. In NPCs Sox2 also occupied the promoters of genes in these classes, but in

addition genes involved in RNA splicing, which is known to play a major role in the development of the CNS (Grabowski, 2011). For example, in NPCs Sox2 occupied the promoters of two splicing factors, PTB and nPTB (neuronal PTB), which form a negative feedback switch between non-neuronal (including NPC) alternative splicing programs mediated by PTB and neuronal alternative spicing programs mediated by nPTB (Boutz et al., 2007). Our data suggests that the loss of Sox2 in differentiating neurons could be a key player in activating this switch, allowing PTB levels to fall which would alleviate the repression PTB exerts on nPTB. We also confirmed that the majority of Sox2 bound regions were distal to known TSSs in both cell types, and in support of the notion that these regions might represent distal enhancers, we showed that a significant fraction of these non-TSS associated Sox2-bound regions correlate with known marks of enhancers, H3K4me1 and H3K27Ac. Consistent with the idea that Sox2 may act as a pioneer factor to prime enhancers for later activation (Wegner et al., 2011), we show that a large number of the enhancers are H3K4me1 positive, a mark of poised enhancers, but not H3K27Ac positive, a mark of active enhancers.

Sox2 occupied almost mutually exclusive sites in ESCs and NPCs, yet seemed to regulate some similar processes. How could this dichotomy be achieved? Upon inspection we found that while Sox2-bound regions were divergent between ESCs and NPCs, many genes were regulated by Sox2 in both cell types. Indeed when focused on promoter-associated Sox2-bound regions, we found a higher overlap between Sox2-bound regions in ESCs and NPCs than that observed among all Sox2-bound regions, and further we found a large number of genes which displayed Sox2 binding at distinct sites within the same promoter. This phenomenon was even more pronounced when considering enhancers. Enhancer associated Sox2-bound regions were highly divergent between ESCs and NPCs, but many of the genes linked to these enhancers were common to both cell types. Of particular interest were genes linked to poised, Sox2+ enhancers in ESCs which became associated with active Sox2 enhancers in NPCs; these genes tended to be highly expressed in NPCs relative to ESCs and be involved in NPC identity, confirming the presumed role of Sox2 in ESCs as a pioneer factor and indicating that it indeed “hands

off" targets to itself upon transaction to the NPC state. Furthermore, a subset of genes linked to Sox2-bound, poised enhancers in ESCs were associated with new Sox2-bound, poised enhancers in NPCs, and tended to code for regulators of later neural development. Thus, Sox2 acts as a pioneer factor in NPCs as well, possibly handing off targets to Sox factors expressed in more differentiated neural cells, such as Sox11 or Sox21 (Sandberg et al., 2005; Ferletta et al., 2011). Therefore, a large number of genes crucial for the proper execution of neural development are regulated by Sox2 from early development in pluripotent ESCs though the multipotent NPC state, where they are poised to become activated upon the proper developmental cues (Figure 9).

Figure 9

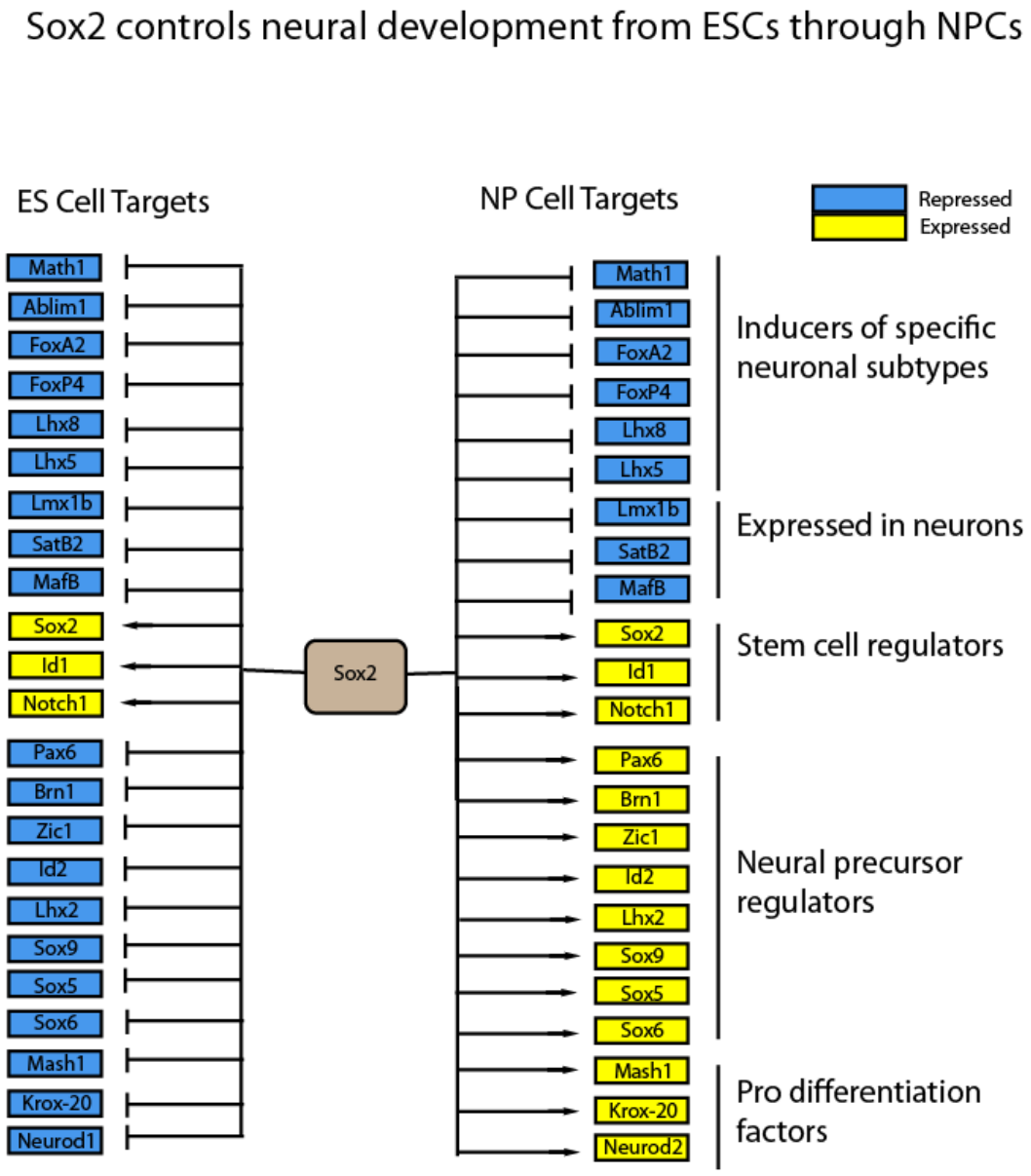


Figure 9 - Sox2 controls neural development from ESCs to NPCs. Network diagram of Sox2 bound genes in ESCs (left) which are involved in nervous system development and generally repressed in ESCs (blue) and remain bound by Sox2 in NPCs (right). In some cases these genes become active in NPCs (yellow), but in others they remain poised to become active in differentiated cells of the nervous system.

Methods

Chromatin Immunoprecipitation

H3K27me3 NPC ChIP

ChIP was performed as described (Lee et al., 2006). Briefly, approximately $5x10^8$ cells were cross-linked and chromatin fractions were isolated. Chromatin was sheared by sonication, whole cell extract (WCE) was removed, and ChIP was performed using H3K27me3 and Sox2 antibodies (Millipore 07-449 rabbit polyclonal). ChIP and WCE DNA was then purified and genomic libraries were prepared using the ChIP-Seq Sample Prep Kit (Illumina 1003473) according to the manufacturers protocol (Illumina 11257047) for selecting library fragments between 200 and 350 bp. Samples were run using the GA2X genome sequencer (SCS v2.6, pipeline 1.5).

For all other methods, see Chapter 2.

References

Ben-Porath, I., M. W. Thomson, et al. (2008). "An embryonic stem cell–like gene expression signature in poorly differentiated aggressive human tumors." Nature Genetics **40**(5): 499-507.

Bergsland, M., D. Ramsköld, et al. (2011). "Sequentially acting Sox transcription factors in neural lineage development." Genes & Development **25**(23): 2453-2464.

Bernard, O., J. Drago, et al. (1992). "L-myc and N-myc influence lineage determination in the central nervous system." Neuron **9**(6): 1217-1224.

Boutz, P. L., P. Stoilov, et al. (2007). "A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons." Genes & Development **21**(13): 1636.

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

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

Cau, E., G. Gradwohl, et al. (2000). "Hes genes regulate sequential stages of neurogenesis in the olfactory epithelium." Development **127**(11): 2323.

Creyghton, M. P., A. W. Cheng, et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proceedings of the National Academy of Sciences **107**(50): 21931.

Ferletta, M., D. Caglayan, et al. (2011). "Forced expression of Sox21 inhibits Sox2 and induces apoptosis in human glioma cells." International Journal of Cancer.

Gowan, K., A. W. Helms, et al. (2001). "Crossinhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons." Neuron **31**(2): 219-232.

Grabowski, P. (2011). "Alternative splicing takes shape during neuronal development." Curr Opin Genet Dev. (4): 388-94

Heintzman, N. D., G. C. Hon, et al. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.

Heintzman, N. D., R. K. Stuart, et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nature Genetics**39**(3): 311-318.

Honda, T., K. Kobayashi, et al. (2011). "Regulation of cortical neuron migration by the Reelin signaling

pathway." Neurochem Res **36**(7): 1270-1279.
Huang, Y.S., Cheng C.Y. et al. (2011) "Involvement of SHP2 in focal adhesion, migration and differentiating neural stem cells." Brain Dev Epub ahead of print 11/24/2011
Ivanova, N., R. Dobrin, et al. (2006). "Dissecting self-renewal in stem cells with RNA interference." Nature **442**(7102): 533-538.
Kuwahara, A., Y. Hirabayashi, et al. (2010). "Wnt signaling and its downstream target N-myc regulate basal progenitors in the developing neocortex." Development **137**(7): 1035.
Lee, T. I., R. G. Jenner, et al. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-313.
Liber, D., R. Domaschenz, et al. (2010). "Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage." Cell Stem Cell **7**(1): 114-126.
Loh, Y. H., Q. Wu, et al. (2006). "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells." Nature Genetics **38**(4): 431-440.
Marson, A., S. S. Levine, et al. (2008). "Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells." Cell **134**(3): 521-533.
McLean, C. Y., D. Bristor, et al. (2010). "GREAT improves functional interpretation of cis-regulatory regions." Nature Biotechnology **28**(5): 495-501.
Miquelajauregui, A., A. Varela-Echavarria, et al. (2010). "LIM-homeobox gene Lhx5 is required for normal development of Cajal-Retzius cells." J Neurosci **30**(31): 10551-10562.
Ohtsuka, T., M. Sakamoto, et al. (2001). "Roles of the basic helix-loop-helix genes Hes1 and Hes5 in expansion of neural stem cells of the developing brain." Journal of Biological Chemistry **276**(32): 30467-30474.
Okabe, S., K. Forsberg-Nilsson, et al. (1996). "Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro." Mechanisms of Development **59**(1): 89-102.
Rahl, P. B., C. Y. Lin, et al. (2010). "c-Myc regulates transcriptional pause release." Cell **141**(3): 432-445.
Ramos, C., S. Rocha, et al. (2010). "Two Notch ligands, Dll1 and Jag1, are differently restricted in their range of action to control neurogenesis in the mammalian spinal cord." PLoS One **5**(11): e15515.
Reynolds, B. A. and S. Weiss (1992). "Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system." Science **255**(5052): 1707.
Rocha, S. F., S. S. Lopes, et al. (2009). "Dll1 and Dll4 function sequentially in the retina and pV2 domain of the spinal cord to regulate neurogenesis and create cell diversity." Developmental Biology **328**(1): 54-65.
Sandberg, M., M. Kallstrom, et al. (2005). "Sox21 promotes the progression of vertebrate neurogenesis." Nature Neuroscience **8**(8): 995-1001.
Vernay, B., M. Koch, et al. (2005). "Otx2 regulates subtype specification and neurogenesis in the midbrain." J Neurosci **25**(19): 4856-4867.
Visel, A., M. J. Blow, et al. (2009). "ChIP-seq accurately predicts tissue-specific activity of enhancers." Nature **457**(7231): 854-858.
Wang, J. Q., Q. Tang, et al. (2004). "Glutamate signaling to Ras-MAPK in striatal neurons: mechanisms for inducible gene expression and plasticity." Molecular Neurobiology **29**(1): 1-14.
Wegner, M. (2011). "SOX after SOX: SOXession regulates neurogenesis." Genes & Development **25**(23): 2423-2428.
Wey, A. and P. S. Knoepfler (2010). "C-myc and N-myc promote active stem cell metabolism and cycling as architects of the developing brain." Oncotarget **1**(2): 120.
Zhao, Y., Y. J. Guo, et al. (1999). "Isolated cleft palate in mice with a targeted mutation of the LIM homeobox gene lhx8." Proceedings of the National Academy of Sciences **96**(26): 15002-15006.
Zhao, Y., K. M. Kwan, et al. (2007). "LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum." Proceedings of the National Academy of Sciences **104**(32): 13182-13186.

Chapter 3: Conclusions

The focus of this thesis was the role of the Sry-related, HMG-box containing transcription factor Sox2 in embryonic stem cells (ESCs) and neural precursor cells (NPCs). Sox2 occupied more enhancer regions than promoters in both cell types, and did so in a cell-type-specific manner; less than 5% of Sox2 bound regions are shared in ESCs and NPCs. While regions of Sox2 binding were distinct, the genes associated with these regions were shared. Interestingly, Sox2 associated with a large subset of H3K4me1+, H3K27Ac- negative, so-called "poised" enhancers in ESCs which were linked to genes involved in neural development. Many of these genes were also linked to Sox2-bound enhancers in NPCs and fell into two groups: those which were now associated with H3K4me1+/-, H3K27Ac+, so-called "active" enhancers and involved in NPC identity, and those which remained poised and were involved in differentiation past the NPC state. Thus, Sox2 may act as a pioneer factor, which sets the stage in ESCs for later development down the neural lineage by priming specific genes for later activation. In NPCs, some of these genes become active, while at other loci Sox2 continued to keep genes poised to become expressed in differentiated neurons and glia.

How can Sox2 bind to totally distinct regions in ESCs and NPCs? In Chapter 2 I outlined experiments attempting to answer this question. The answer did not lie in the motif used by Sox2 in the two cell types, as Sox2 utilized the canonical Sox2 recognition site to bind targets in both ESCs and NPCs. While Sox2 is known to partner with the POU factor Oct4 in ESCs, Oct4 is absent in NPCs. Interestingly, other POU factor family members are expressed in NPCs, including Brn1 and Brn2. These factors are co-expressed with Sox2 in neurogenic regions of the brain and were previously thought to play some role in NPC function. Genome-wide analysis of their binding revealed that they co-occupied distal enhancers with Sox2 in NPCs, but showed little overlap with Sox2 at promoter regions. While Brn1, Brn2, and Oct4 all utilized Octamer motifs to bind to genomic targets in ESCs and NPCs, Sox2:Oct4 and Sox2:Brn bound regions displayed distinct orientations of Sox and Octamer motifs, suggesting that the unique stereochemical interactions between Sox2 and POU factors may play a role in

target selection by these modules. Further, ectopic Brn2 in differentiating ESCs recruited Sox2 to new target loci and accelerated neural differentiation. Thus, the suite of partner factors expressed in a given cellular context may influence the function of Sox2 in terms of the genomic loci to which it binds.

While this thesis has addressed unanswered questions about the role of Sox2 in ESCs and NPCs, many questions remain. How can Sox2 act as a pioneer in some contexts and actively promoter transcription in others? Sox2 must interact with at least two sets of co-factors to achieve this dual function. Oct4, Brn1, and Brn2 co-localize with Sox2 at both active and poised genomic loci, so proteomic studies should be performed to determine which other factors interact with endogenous untagged Sox2 in genetically matched ESCs and NPCs and attempts should be made to define distinct complexes which localize to poised versus active regions. A good place to start would be the motif analysis presented in Chapter 3 which demonstrates that binding sites for multiple transcription factor families occur within Sox:POU bound regions. The differential binding of other factors may play a key role in recruiting co-activators or repressors to specific loci. Another unanswered question is why Sox2:Oct4 and Sox2:Brn modules select different Sox:Octamer composite motifs. Constructs expressing chimeric POU factors containing differing domains from Oct4, Brn1, and Brn2 should be engineered and tested for the ability to cooperatively bind with Sox2 at known ESC or NPC specific targets. This may allow for the identification of domains or residues that mediate selection of certain composite Sox:Octamer motifs. Also, it would be interesting to more precisely define the relationship between Sox2, the POU factors and chromatin marks H3K4me1 and H3K27Ac at enhancers. Does the presence of these marks allow the cell-type-specific binding of Sox2 and POU factors in ESCs and NPCs? Or, more likely, do these factors recruit the enzymes that catalyze the addition of these modifications to histones? What are these enzymes? If a direct link between the establishment of H3K4me1 and H3K27Ac and the binding of Sox2 and POU factor partners could be found, the role of these proteins as pioneer factors would become more clear. Finally, this thesis suggests that combinatorial control of gene expression by transcription factor co-binding may be a general phenomenon in development. This may be one way that

the limited number of transcription factors encoded in the mammalian genome (less than 2,000) can control the expression of approximately 20,000 genes in the constant state of flux that is the process of development, in the 200 adult cell types, and in the face of countless external stimuli. Thus, only by defining these combinatorial interactions can the regulatory network of a cell be truly understood.

Appendix: The role of cell density and oxygen tension on factor-mediated reprogramming efficiency

Michael A. Lodato[1,2], Christopher J. Lengner[2,3], Maris Wernig[2,4], Ruth Foreman[1,2], Jacob Hanna[2,5], Rudolf Jaenisch[1,2]

[1]Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue Cambridge, MA 02139 USA; [2]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142 USA [3]Current Address Department of Animal Biology, School of Veterinary Medicine, 3800 Spruce Street, Philadelphia, PA 19104 [4]Current Address - Institute for Stem Cell Biology and Regenerative Medicine 265 Campus Drive, Stanford, CA 94305 [5]Current Address - Department of Molecular Genetics and The Helen and Martin Kimmel Institute for Stem Cell Research, Meyer Building, Weizmann Institute of Science, Rehovot, Isreal.

Synopsis

By the amazing process of factor-mediated reprogramming, a fully differentiated cell can be completely converted back into a pluripotent, embryonic-stem-cell-like state, termed induced pluripotency by the ectopic expression of four transcription factors: Oct4, Sox2, cMyc, and Klf4. Early reprogramming systems were inefficient and highly variable, stymieing efforts to study the mechanics of this process. To circumvent these issues, our lab developed "secondary" reprogramming systems, which controlled for heterogeneity introduced by the viral transduction of transgenes; secondary systems displayed increased efficiency and reproducibility. I used such a system to study the effects of two variables in the reprogramming process: cell density oxygen tension. I was able to define optimal conditions for each parameter, and show that increases or decreases in cell density from this optimal value decreased reprogramming efficiency. Taken together with other published work, these data suggest the balance between cell growth and cell senescence plays a key role in determining reprogramming efficiency.

Author Contributions

M.W. and R.J. conceived secondary reprogramming system. M.W. developed the dox-inducible vectors and secondary system. J.H. and R.F. generated the iPS cell line NGFP 2. M.A.L., and C.J.L., and designed the oxygen tension and density experiments. C.J.L derived NGFP2 MEFs. M.A.L performed secondary oxygen tension and reprogramming experiments. M.A.L wrote this chapter.

Introduction

Normal development is unidirectional; cell types with wide developmental potential, such as the pluripotent cells of the inner cell mass which are capable of differentiating into the cells of the embryo proper, progressively give rise to more specialized cell types with less developmental potential, such as tissue stem cells and non-dividing, differentiated cells. Studies of the frog *Xenopus* demonstrated that this directionality is experimentally reversible by transferring a differentiated, somatic nucleus into an enucleated oocyte, which upon activation was able to develop into a fully formed frog (Briggs and King, 1952; Gurdon, 1962). This technique, known as somatic cell nuclear transfer (SCNT) or cloning, proved that development and differentiation occur by reversible, non-genetic changes, and was later used to create cloned mammals, such as Dolly the sheep, even from terminally differentiated and non-mitotic cells (Wilmut, 1997; Wakayama et al., 2008; Hochedlinger et al., 2002; Eggan et al., 2004). Thus, development is experimentally reversible by factors contained in the oocyte.

While it was possible that hundreds of the factors present in the unfertilized egg where responsible for reprogramming the somatic nucleus, it was also conceivable that only a few of these factors were sufficient for reprogramming. In 2006, Takahashi and Yamanka showed that indeed the ectopic expression by viral transduction of only four transcription factors, Oct4, Sox2, c-Myc, and Klf4, was sufficient to reprogram somatic cells (in this case, mouse embryonic fibroblasts [MEFs]) back into pluripotent, embryonic stem cell-like cells, termed induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka, 2006). Factor-mediated reprogramming was sufficient to generate iPSCs from terminally differentiated and non-mitotic cells, and could yield iPSCs which pass the most stringent tests for pluripotency, germline contribution and tetraploid complementation (Wernig et al., 2007; Takahashi and Yamanaka, 2007; Hanna et al., 2008; Zhao et al., 2008; Kim et al., 2011; Carey et al., 2011). Thus, four factors, Oct4, Sox2, c-Myc, and Klf4 are sufficient to reprogram differentiated cells into a pluripotent state.

To gain insight into the mechanism by which factor-mediated reprogramming occurred, technical hurdles had to be overcome. One major challenge was the heterogeneity of viral integrations in populations of reprogramming cells, which lead to cell-to-cell variation of factor expression. To circumvent this issue, "secondary" systems were designed to create populations of cells with homogenous factor integrations (Wernig et al., 2008) (Figure 1). In these systems, somatic cells which already expressed the reverse tetracyclin transactivator, rtTA, and puromycin resistance genes were transduced with tetracycline inducible viral vectors expressing the four factors and induced with the tetracyclin analog, doxycycline (dox). Then, once dox-independent iPSC lines were established, the resulting cells were injected into blastocysts to create chimeras. Puromycin-selected MEFs harvested from these mice were then homogenous for viral integrations known to be permissive for reprogramming, and upon exposure to dox yielded "secondary" iPS cells. This system also overcame another obstacle to mechanistic studies in reprogramming, that of low efficiency. While primary iPSC generation had an efficiency of 1/1,000, secondary systems reached efficiencies of 2-4% depending on the line (Wernig et al., 2008). It was using this system that we performed analyses on the effect of two parameters on the efficiency reprogramming process, starting-cell density and oxygen tension.

Figure 1

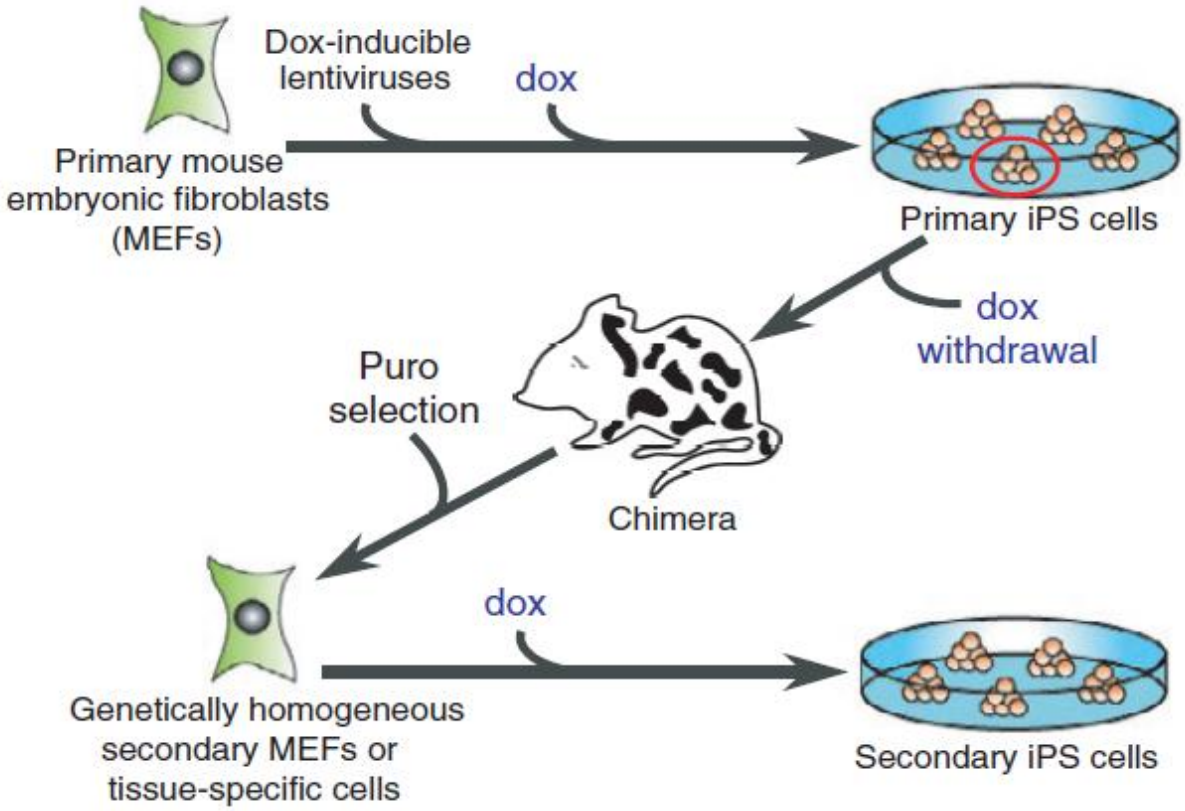


Figure 1: Schematic of the secondary system for obtaining genetically homogenous populations of reprogramming somatic cells (Wernig et al., 2008).

Results

Reprogramming efficiency is dependent on cell density

In order to characterize this new "secondary" system, we sought to define the optimal parameters in which reprogramming could occur. The high oxygen tension experienced by cells brought into culture relative to the level of oxygen they experience *in vivo* is known to have a negative effect on cell viability and induce cellular senescence, particularly in MEFs (Parrinello et al., 2003). To test whether oxygen tension had an effect on reprogramming efficiency, we cultured secondary MEFs harboring a GFP reporter knocked into the Nanog locus (Nanog-GFP; cell line NGFP2) in 20% oxygen and 5% oxygen

with dox for three weeks. We then withdrew dox for one week to ensure we only scored transgene independent lines, and counted the number of GFP-positive colonies under each condition. Indeed, reprogramming efficiency was higher in cultures kept under low oxygen, indicating that cellular senescence or another process induced by high oxygen may be detrimental to reprogramming efficiency.

Figure 2

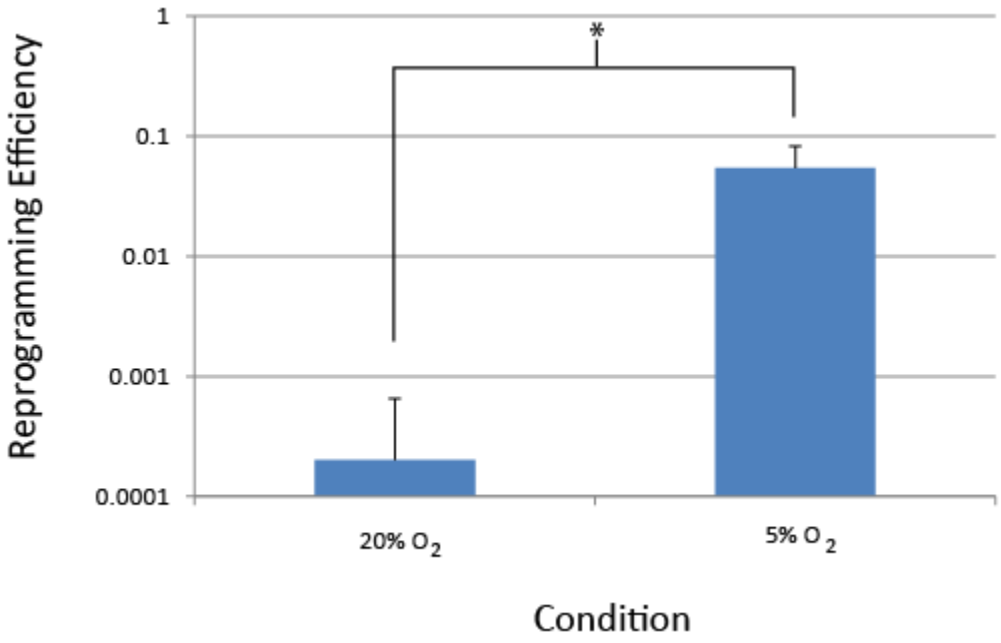


Figure 2: Effect of oxygen tension on reprogramming efficiency. Equal numbers of Nanog-GFP MEFs were plated and cultured at either 5% or 20% oxygen in four independent wells with dox for 3 weeks and without dox for 1 week, and colonies were scored for GFP visually. * indicates p-value < .05, Student's two tailed, paired T test

We next sought to define the optimal cell density under which reprogramming could occur. Mammalian cells in culture are known to respond to cell density by sensing paracrine signals from neighboring cells (Pedroja et al., 2009; Hanahan and Weinberg, 2010), so we tested whether we could define an optimal density under which reprogramming could occur. Nanog-GFP secondary MEFs were plated at varying cell densities, and reprogramming efficiency was calculated by scoring GFP+, dox-

independent colonies, as above. We found that at both very low and very high density, reprogramming efficiency was reduced (Figure 3). We observed that at low densities, plated cells seemed to be non-proliferative and senescent (data not shown), indicating that paracrine factors secreted by neighboring MEFs may support cell growth, and the lack of these signals induces mitotic exit, senescence, and decrease reprogramming efficiency. At high densities, it is likely that contact inhibition, a well characterized regulator cell growth (Abercrombie, 1979; Küppers et al., 2010), induced mitotic arrest and subsequently decreased reprogramming efficiency. Thus, an optimal density for reprogramming exists in which cells maintain paracrine signaling and are not contact inhibited.

Figure 3

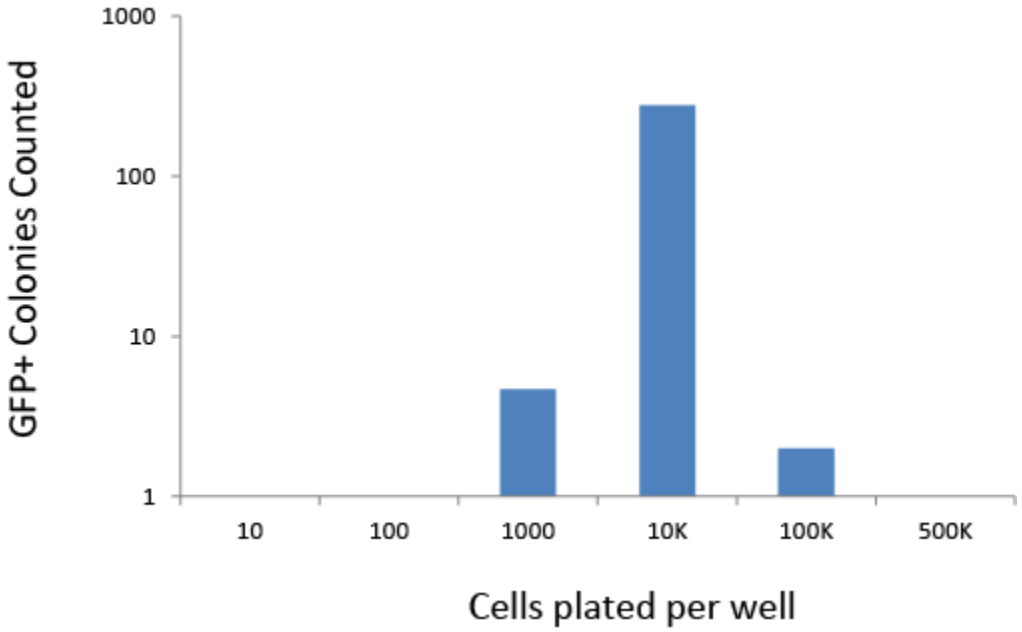


Figure 3: Effect of cell density on reprogramming efficiency. Varying numbers of secondary Nanog-GFP MEFs were plated in wells of a 12-well dish (surface area, 400 mm^2) in triplicate, cultured with dox for 3 weeks and without dox for 1 week, and colonies were scored for GFP visually. Bars indicate number of colonies obtained.

Discussion

This appendix summarized experiments attempting to characterize the optimal conditions for reprogramming using a "secondary" reprogramming system. We found that both oxygen tension and starting cell density influence the efficiency of iPSC generation as measured by activation of a Nanog-GFP reporter. Specifically, high oxygen, and low or high cell density inhibited reprogramming in this system. A commonality amongst these deleterious conditions is that they all are known to inhibit cellular proliferation. The hyperoxic environment in normal cell culture relative to an *in vivo* setting is known to induce of cellular senescence, characterized by an irreversible exit from the cell cycle, in MEFs (Parinello et al., 2003; Lanigan et al., 2011). Dense culture of cells are known to undergo contact inhibition, which has long been known to negatively regulate cellular growth in culture (Levine et al., 1965), possibly due to the important role density sensing plays in normal development and cancer (Hanahan and Weinberg, 2010). On the other extreme of cell density, sparse cultures of MEFs also failed to reprogram. MEFs are known to grow poorly at low density, probably due to a loss of paracrine signaling between neighboring MEFs, such as those mediated by TGF-β and extracellular matrix components (Smith and Braunschweiger, 1979; Connover, 1989; Postlethwaite et al., 1992; Pedroja et al., 2009). Thus, three separate conditions in which cell proliferation was inhibited resulted in decreased reprogramming efficiency.

Interestingly, work from our lab and others later showed that p53 knockout MEFs reprogrammed with higher efficiency than controls (Hong et al., 2009; Li et al., 2009; Kawamura et al., 2009; Utikal et al., 2009; Marión et al., 2009; Hanna et al., 2009), and that this increased efficiency can be fully explained by the increase in cell cycle of p53-null MEFs relative to wild-type MEFs (Hanna et al., 2009). Thus, my data and work form others suggest that proliferation is a key component of factor-mediated reprogramming. Factor mediated reprogramming is known to be associated with a near-complete erasure of the somatic epigenome and a rewriting of a pluripotent, ESC-like epigenetic state (Guenther et al., 2010). This includes remodeling of histone modifications and demethylation of modified bases in DNA

itself. One possible explanation for the link between cell proliferation and reprogramming is the idea that a certain number of cell divisions must occur to passively dilute out the somatic epigenetic marks. Thus, going forward it would be interesting to test whether active erasure of chromatin and DNA modifications is necessary for reprogramming by loss-of-function analyses of the enzymes thought to catalyze the removal of these marks, or if by ectopic expression of such enzymes it is possible to accelerate reprogramming in a cell-proliferation independent way. Thus, cell proliferation seems to be important for factor-mediated reprogramming, but the mechanistic relationship between these two processes remains unclear.

Methods

Primary iPSCs were generated and injected in host blastocysts as described (Wernig et al., 2008). For secondary MEF isolation, chimeric embryos were isolated at embryonic day 13.5 and the head and the internal organs were subsequently removed. The remaining somatic tissue was physically dissociated and then incubated in 0.5% trypsin in HEPES-EDTA at 37°C for 20 minutes, after which cells were resuspended in MEF media (DMEM supplemented with 10% Fetal Bovine Serum (Hyclone), 0.001 % β-mercapto-ethanol (Sigma, M7522)., 100 μM nonessential amino acids (Invitrogen, 11140-050), 2 mM L-glutamine (Invitrogen, 25030-081), 100 U/mL penicillin, 100 μg/mL streptomycin (Invitrogen, 15140-122) containing puromycin (2ug/mL) and expanded in low oxygen (5%) for two passages prior to freezing for long term storage. Secondary MEFs used for the described experiments were thawed into low oxygen and experiments plated 1-2 passages after thawing into 12-well plates with 2 μg/ml doxycycline in ES medium (Chapter 2) for 3 weeks, after which doxycycline was withdrawn for 1 week and GFP+ iPSC colonies were scored

References

Abercrombie, M. (1979). "Contact inhibition and malignancy." Nature **281**(5729): 259.

Briggs, R. and T. J. King (1952). "Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs." Proceedings of the National Academy of Sciences **38**(5): 455-463.

Carey, B. W., S. Markoulaki, et al. (2011). "Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells." Cell Stem Cell **9**(6):

588-598.
Conover, A., R. Hintz, et al. (1989). "Direct evidence that the insulin receptor mediates a mitogenic response in cultured human fibroblasts." Horm Metab Res **21**(2): 59-63.
Eggan, K., K. Baldwin, et al. (2004). "Mice cloned from olfactory sensory neurons." Nature **428**(6978): 44-49.
Guenther, M. G., G. M. Frampton, et al. (2010). "Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells." Cell Stem Cell **7**(2): 249-257.
Gurdon, J. B. (1962). "The transplantation of nuclei between two species of Xenopus." Developmental Biology **5**: 68-83.
Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.
Hanna, J., S. Markoulaki, et al. (2008). "Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency." Cell **133**(2): 250-264.
Hanna, J., K. Saha, et al. (2009). "Direct cell reprogramming is a stochastic process amenable to acceleration." Nature **462**(7273): 595-601.
Hochedlinger, K. and R. Jaenisch (2002). "Monoclonal mice generated by nuclear transfer from mature B and T donor cells." Nature **415**(6875): 1035-1038.
Hong, H., K. Takahashi, et al. (2009). "Suppression of induced pluripotent stem cell generation by the p53/p21 pathway." Nature **460**(7259): 1132-1135.
Kawamura, T., J. Suzuki, et al. (2009). "Linking the p53 tumour suppressor pathway to somatic cell reprogramming." Nature **460**(7259): 1140-1144.
Kim, J., C. J. Lengner, et al. (2011). "Reprogramming of Postnatal Neurons into Induced Pluripotent Stem Cells by Defined Factors." Stem Cells **29**(6): 992-1000.
Küppers, M., C. Ittrich, et al. (2010). "The transcriptional programme of contact-inhibition." Journal of Cellular Biochemistry **110**(5): 1234-1243.
Lanigan, F., J. Geraghty, et al. (2011). "Transcriptional regulation of cellular senescence." Oncogene **30**(26): 2901-2911.
Levine, E. and Y. Becker "BOONE, AND H. EAGLE. 1965. Contact inhibition, macromolecular synthesis and polyribosomes in cultured human diploid fibroblasts." Proceedings of the National Academy of Sciences **53**: 350-356.
Levine, E. M., Y. Becker, et al. (1965). "Contact inhibition, macromolecular synthesis, and polyribosomes in cultured human diploid fibroblasts." Proceedings of the National Academy of Sciences **53**(2): 350.
Li, H., M. Collado, et al. (2009). "The Ink4/Arf locus is a barrier for iPS cell reprogramming." Nature **460**(7259): 1136-1139.
Marión, R. M., K. Strati, et al. (2009). "A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity." Nature **460**(7259): 1149-1153.
Parrinello, S., E. Samper, et al. (2003). "Oxygen sensitivity severely limits the replicative lifespan of murine fibroblasts." Nature Cell Biology **5**(8): 741-747.
Pedroja, B. S., L. E. Kang, et al. (2009). "Plasminogen activator inhibitor-1 regulates integrin αvβ3 expression and autocrine transforming growth factor β signaling." Journal of Biological Chemistry **284**(31): 20708-20717.
Postlethwaite, A., M. Holness, et al. (1992). "Human fibroblasts synthesize elevated levels of extracellular matrix proteins in response to interleukin 4." Journal of Clinical Investigation **90**(4): 1479.
Smith, J. R. and K. I. Braunschweiger (1979). "Growth of human embryonic fibroblasts at clonal density: concordance with results from mass cultures." Journal of Cellular Physiology **98**(3): 597-601.
Takahashi, K., K. Tanabe, et al. (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-872.
Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.

Utikal, J., J. M. Polo, et al. (2009). "Immortalization eliminates a roadblock during cellular reprogramming into iPS cells." Nature **460**(7259): 1145-1148.
Wakayama, S., J. M. Cummins, et al. (2008). "Nuclear reprogramming to produce cloned mice and embryonic stem cells from somatic cells." Reprod Biomed Online **16**(4): 545-552.
Wernig, M., C. J. Lengner, et al. (2008). "A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types." Nature Biotechnology **26**(8): 916-924.
Wernig, M., A. Meissner, et al. (2007). "In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state." Nature **448**(7151): 318-324.
Wilmut, I., A. E. Schnieke, et al. (1997). "Viable offspring derived from fetal and adult mammalian cells." Nature **385**(6619): 810-813.
Zhao, X., W. Li, et al. (2009). "iPS cells produce viable mice through tetraploid complementation." Nature **461**(7260): 86-90.