Transcriptional Regulation of Neuroectodermal Lineage Commitment in Embryonic Stem Cells

Yuan-Ping Huang

Submitted in partial fulfillment of the

requirements for the degree of

Doctorate of Philosophy

under the Executive Committee

of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSUTY

2014

Abstract

Transcriptional Regulation of Neuroectodermal Lineage Commitment in Embryonic Stem Cells

Yuan-Ping Huang

Lineage commitment of pluripotent cells is a critical step in the development of multicellular organisms and a prerequisite for efficient differentiation of stem cells into terminal cell types. During successful neuroectodermal lineage commitment, extracellular signals terminate the pluripotency program, activate neural transcriptional program, and suppress alternative mesendodermal fate. Retinoic acid (RA) has been identified as a potent inducer of neural differentiation in embryonic stem cells (ESCs), yet the transcriptional program initiated by RA is poorly understood. Expression profiling of differentiating ESCs revealed delayed response of the pluripotency marker Oct4 and neural marker Sox1 following RA treatment, suggesting that RA regulates the pluripotency program and neural transcriptional program indirectly through induction of additional transcription factors.

In this study, I identified a zinc finger factor Zfp703 as a downstream effector of RA-mediated neuroectodermal lineage commitment. Zfp703 expression in ESCs resulted in Oct4 repression, Sox1 induction, and neural differentiation. Moreover, Zfp703 strongly suppresses mesendodermal fate by repressing genes such as *Brachyury, Eomes,* and *Mixl1* even under conditions favoring mesendoderm specification. Zfp703 binds to and represses Lef1 promoter, raising the possibility that it might modulate Wnt signaling via regulating Lef1. Finally, Zfp703 is not required for RA-mediated Oct4 repression and Sox1 induction. However, it is necessary for efficient *Brachyury* repression by RA. Based on these data, I propose that Zfp703 is involved in the transcription regulation during neural progenitor specification. Through downregulating of

both mesendodernal fate and pluripotency, Zfp703 de-represses neural transcriptional program and indirectly promotes the default neuroectodermal lineage commitment.

Table of Contents

Tables & illustrations

Acknowledgements

This thesis could not have been possible without the guidance from my mentor, Dr. Hynek Wichterle, and members of my thesis committee: Dr. Lori Sussel, Dr. Boris Reizis, Dr. Alla Grishok, and Dr. Lorenz Studer. It also benefitted significantly from the exchange of ideas during discussion in the meetings with current and past members of the Wichterle and the Doetsch labs throughout my study. I deeply appreciate their input. Finally, I am grateful to my family and friends for their unconditional support along the way. And I would like to dedicate this thesis to them.

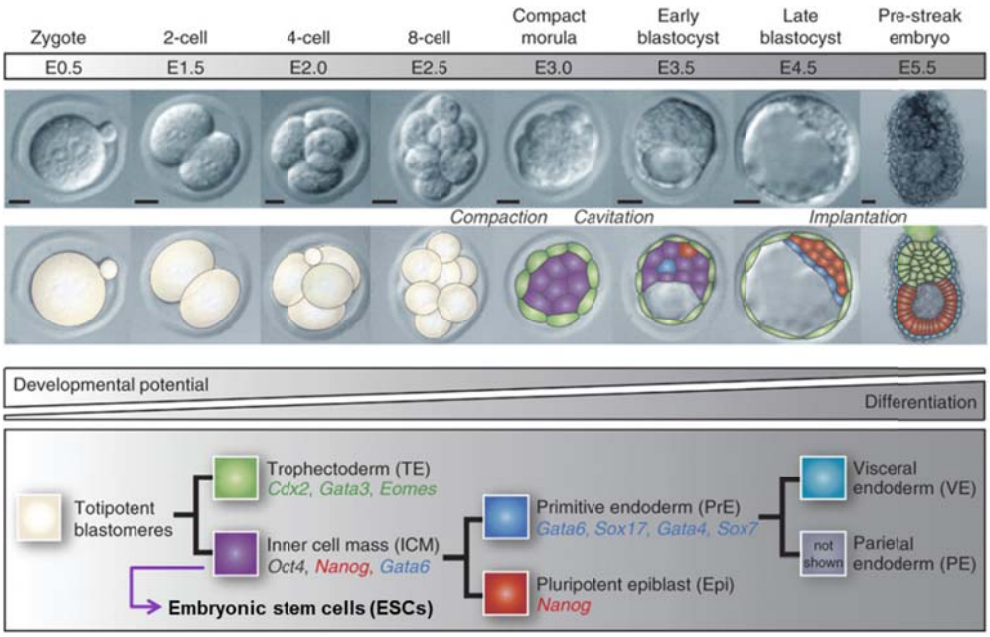
Chapter 1. Introduction

One of the most fascinating phenomena during the embryonic development of metazoans is the generation of multitudinous cell types from a single cell: the zygote. In mouse, the zygote undergoes waves of divisions after fertilization to generate an apparent homogeneous cell mass known as the morula (Figure 1.1-A). Subsequently, these cells segregate into two populations with distinct fates and development potential: the inner cell mass (ICM) which will contribute mostly to the embryo proper and the trophectoderm (TE) giving rise primarily to extraembryonic tissues supporting placenta development (Acloque et al., 2012). After TE has emerged, around E3.5 (at early blastocyst stage), ICM cells further split into two groups: the epiblast, a pluripotent embryonic layer and the primitive endoderm (PE) which contributes to extraembryonic tissues. At the onset of gastrulation from E6.5 (Figure 1.1-B), subsets of cells in the epiblast move to the primitive streak (PS) where they undergo epithelial to mesenchymal transition (EMT), fold inward to form mesendoderm- a transient population that quickly separates into the embryonic endoderm and mesoderm (Johnson, 2009; Thiery et al., 2009). At this point, the remaining epiblast cells will give rise to the embryonic neuroectoderm. So far, the basic embryonic body plan comprising three primary germ layers has been established, and cells within these three lineages are poised to develop into all mature cell types in the animal. Hence, the fate choice between neuroectoderm and mesendoderm by pluripotent epiblast cells at early gastrulation marks the first cell lineage commitment in the embryo proper. And the segregation between these two cellular identities is fundamental for the generation of diversified cell types down the developmental path.

Embryonic stem cells as a model system to study early development

To fully understand the regulation of germ layer development, it is necessary to characterize the molecular basis of lineage commitment upon gastrulation. However, such attempts are hindered by the relative inaccessibility of embryos in placental mammals (Chenoweth et al., 2010). Even if the technical difficulties to isolate embryos are overcome, it remains challenging to acquire sufficient amount of homogeneous cells for gene profiling and biochemical analyses. This road block was cleared by the derivation of mouse embryonic stem cells (ESCs, Figure 1.1-A), first achieved through explanting blastocysts or ICMs on a layer of mitotically inactivated fibroblasts (feeder cells) (Evans and Kaufman, 1981; Martin, 1981). ESCs self-renew and are capable of expansion in culture extensively. Moreover, they can be differentiated to form cells of three germ layers *in vitro* and contribute to all tissues (including germline) in chimeric mice generated by blastocyst injection (Bradley et al., 1984), a property defined as pluripotency. Importantly, the molecular events and timeline of ESC differentiation *in vitro* faithfully recapitulate embryonic development *in vivo* (Niwa, 2010). With these characteristics, ESCs represent a powerful model system. They could provide unlimited source of relatively pure cell populations at specific developmental stages, as long as proper differentiation protocols are adopted. Therefore, ESCs can greatly facilitate the molecular characterization at different time points in development, as well as the investigation of the transition between earlier and later cell fates.

A


Zygote
2-cell
4-cell
8-cell
Compact morula
Early blastocyst
Late blastocyst
Pre-streak embryo
E0.5
E1.5
E2.0
E2.5
E3.0
E3.5
E4.5
E5.5
Compaction
Cavitation
Implantation
Developmental potential
Differentiation
Totipotent blastomeres
Trophectoderm (TE)
Cdx2, Gata3, Eomes
Inner cell mass (ICM)
Oct4, Nanog, Gata6
Embryonic stem cells (ESCs)
Primitive endoderm (PrE)
Gata6, Sox17, Gata4, Sox7
Pluripotent epiblast (Epi)
Nanog
Visceral endoderm (VE)
not shown
Parietal endoderm (PE)

B

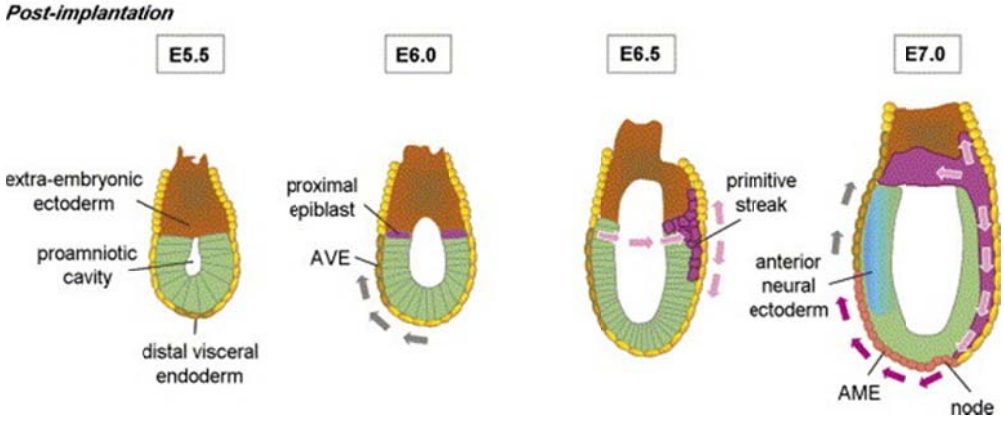

Post-implantation
E5.5
E6.0
E6.5
E7.0
extra-embryonic ectoderm
proamniotic cavity
distal visceral endoderm
proximal epiblast
AVE
primitive streak
anterior neural ectoderm
AME
node

Figure 1.1

Figure 1.1 Early development of mouse embryo

A. Pre-implantation development modified from (Niakan et al., 2013). After fertilization, cells in pre-implanting embryos first segregated into trophectoderm (TE) and inner call mass (ICM) at morula stage around E3.0. Embryonic stem cells (ESCs) are derived from ICM and expressing key pluripotency genes Oct4 and Nanog. Subsequently in the late blastocyst around E4.5, cells in ICM further split into primitive endoderm (PE) and epiblast. B. Early post-implantation development taken from (Marikawa, 2006). At the onset of gastrulation around E6.5, subsets of the cells in the epiblast move to the primitive streak (PS) and fold inward to form mesoderm and endoderm; whereas cells remaining in the epiblast will give rise to neuroectoderm.

Transcription network maintaining ESC pluripotency

During the transition from morula to blastocyst, TE and ICM are separated by the regional activity of the transcription factor Tead4 (Nishioka et al., 2009), resulting in TE cells expressing Cdx2 and the T-box transcription factor eomesodermin (Eomes), and ICM cells expressing the Pou-domain transcription factor Oct4. Oct4 is essential for ICM identity. In morulas lacking Oct4, inner cells differentiate along TE lineage, and they do not develop any embryonic rudiment although the mutant embryos can implant in the uterus (Nichols et al., 1998). Oct4 acts cooperatively with the SRY-box transcription factor Sox2 to regulate multiple target genes, including fibroblast growth factor 4 (FGF4) (Basilico et al., 1997) and key transcription factors such as Nanog (Boyer et al., 2005). Sox2-null mutants form normal blastocysts but fail development before gastrulation (Avilion et al., 2003). The last pluripotent stage before lineage commitment between mesendoderm and neuroectoderm, the epiblast, is marked by and dependent on the expression of Nanog. Nanog is initially expressed in a "salt and pepper" manner throughout the ICM (Plusa et al., 2008). However, it is restricted to the epiblast at late blastocyst stage. Embryos lacking Nanog fail to establish epiblast identity due to ICM degeneration (Silva et al., 2009). Based on these findings, it has been argued that Oct4, Sox2, and Nanog are critical for the establishment of pluripotent state *in vivo*. However, the molecular mechanisms underlying their actions could not be readily elucidated by merely studying the embryos. And this is where the advantages of ESCs as a model system have been effectively utilized.

Research on ESCs has revealed a transcription factor network that sustains pluripotent state in which Oct4, Sox2, and Nanog play essential roles along with other factors such as Klf4, Esrrb, Zfx, and Tbx3 (Galan-Caridad et al., 2007; Jaenisch and Young, 2008; Nichols and Smith,

2012). Among them, Oct4 might be the most crucial factor. It is uniformly expressed continuously throughout the epiblast stage and in ESC culture. Moreover, Oct4-null ESCs cannot be recovered by serial gene targeting, and conditional elimination of Oct4 in ESCs results in their conversion to trophoblast-like cells (Niwa et al., 2000). Chromatin immunoprecipitation-sequencing (ChIP-seq) experiments established that Sox2 is a key Oct4 partner (Nichols and Smith, 2012). Oct4/Sox2 binding elements have been identified in multiple pluripotency genes, including Nanog. (Catena et al., 2004; Kuroda et al., 2005; Okumura-Nakanishi et al., 2005). Genome-wide ChIP-seq analyses demonstrated that Oct4, Sox2, and Nanog co-occupy a significant proportion of their target genes (Boyer et al., 2005; Chambers and Tomlinson, 2009). These genes encode transcription factors regulating pluripotency and differentiation. Oct4, Sox2, and Nanog thus construct a feed forward loop (Figure 1.2) in which they act cooperatively to promote the expression of themselves and other genes supporting ESC identity, meanwhile repressing genes inducing differentiation (Boyer et al., 2005). Because of the self-sustained nature of this transcription network, pluripotency in epiblast as well as ESCs can be maintained indefinitely. In fact, this transcription network is sufficient to reprogram mature differentiated cells back to ESC-like state when reinitiated through overexpression of only four factors: Oct4, Sox2, Klf4, and c-Myc (Takahashi and Yamanaka, 2006). Therefore when pluripotent cells commit to differentiate towards specific lineages, this self-sustained transcription network must be terminated. (Smith, 2009).

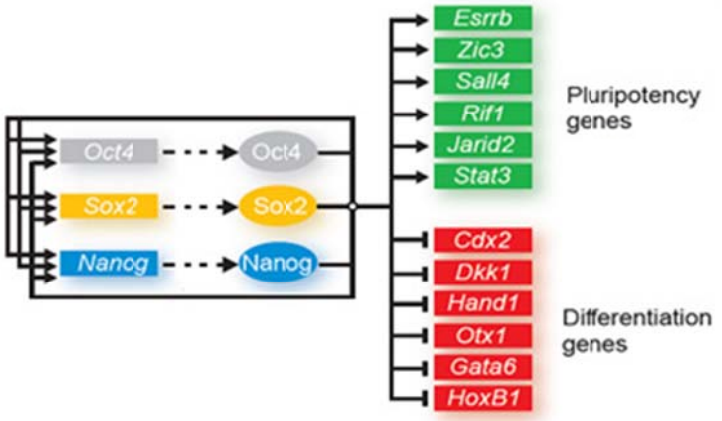

Oct4
Oct4
Sox2
Sox2
Nanog
Nanog
Esrrb
Zic3
Sall4
Rif1
Jarid2
Stat3
Pluripotency genes
Cdx2
Dkk1
Hand1
Otx1
Gata6
HoxB1
Differentiation genes

Figure 1.2

Figure 1.2 ESC transcriptional network maintaining pluripotency

The core transcription factors Oct4, Sox2, and Nanog cooperatively activate the expression of themselves and other pluripotency genes to form the ESC transcriptional network to maintain the ESC identity. At the same time, Oct4, Sox2, and Nanog also repress differentiation genes. (Taken from (Loh et al., 2011))

Lineage commitment of mesoderm and endoderm

The formation of primitive streak (PS) in the posterior region of the embryo is the critical initial step in the lineage commitment of epiblast. It marks the onset of gastrulation, which subsequently results in the generation of the primary germ layers and the establishment of the basic body plan in the embryo (Tam and Loebel, 2007). PS cells express markers Brachyury (T), Mixl1, Eomes, and Evx1 (Arnold et al., 2008; Dush and Martin, 1992; Rivera-Perez and Magnuson, 2005; Robb et al., 2000) , and its formation depends on the integration of Nodal, WNT, and Bmp signaling pathways (Rossant and Tam, 2004). BMP4 is secreted from the extraembryonic ectoderm (ExE) (Winnier et al., 1995) whereas Wnt3 is induced by the unprocessed form of nodal (Ben-Haim et al., 2006). Moreover, Wnt3 activates a feedback loop sustaining nodal expression in the epiblast, which in turn maintains Bmp4 expression in the ExE (Ben-Haim et al., 2006; Brennan et al., 2001).

Following its appearance, groups of cells move from epiblast through the PS where they undergo EMT and emerge as different mesoderm or endoderm tissues depending on their allocation of these cells along the anterior-posterior axis (Figure 1.3) (Johnson, 2009; Thiery et al., 2009). The formation of anterior and posterior PS derivatives is dependent on nodal and BMP activities, respectively (Tam and Loebel, 2007); whereas the specification of middle PS derivatives is subject to the interaction of TGF-β, FGF, and Wnt signaling (Marikawa, 2006).

Primary germ layer induction and patterning can be modeled by ESCs *in vitro*. Treatment of ESCs with Bmp4, Wnt3a, and/or Activin A in serum or serum-free cultures promotes differentiation towards endodermal and mesodermal cell types (D'Amour et al., 2005; Jackson et al., 2010). Gene expression profiling demonstrated that there is an initial down-regulation of pluripotency genes such as Oct4 and Rex1 along with increased Fgf5 expression, which marks

the epiblast stage (Hirst et al., 2006). Subsequent transient expression of Brachyury and Mixl1 suggests the passing of the differentiating cells through a PS-like state before the emergence of Flk1 and Sox17, whose expression denotes early mesoderm and endoderm development (Hirst et al., 2006; Ng et al., 2005). Thus, differentiation of ESCs recapitulates key processes of mesendoderm specification and facilitates its detailed molecular characterization (Pereira et al., 2012).

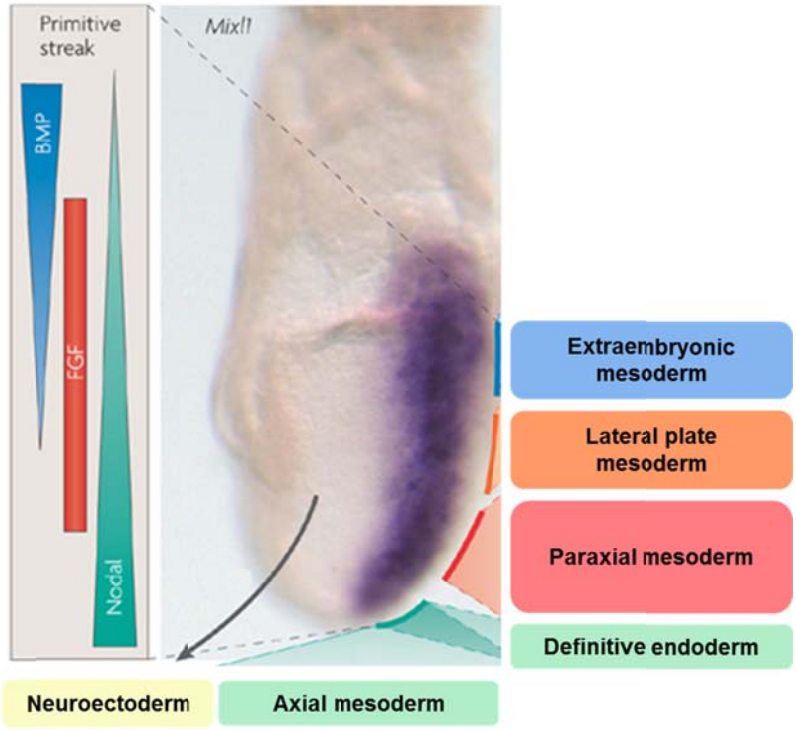

Primitive streak
BMP
FGF
Nodal
Mixl1
Extraembryonic mesoderm
Lateral plate mesoderm
Paraxial mesoderm
Definitive endoderm
Neuroectoderm
Axial mesoderm

Figure 1.3

Figure 1.3 Lineage commitment during early gastrulation

Primitive streak (PS, visualized here by *in situ* hybridization of PS/mesendodermal marker Mixl1) formation requires the Wnt, Bmp, and Nodal signaling. At early gastrulation, epiblast cells committed to mesendodermal fate move through the PS and are further specified by the interaction between Bmp, Fgf, and Nodal signaling: the anterior-most region of the PS (the node) will become axial mesoderm and definitive endoderm; the anterior-third segment of the PS will give rise to paraxial mesoderm; The middle-third area of the PS will generate lateral plate mesoderm; and the posterior-third part of the PS will develop into the extraembryonic mesoderm. Meanwhile, the cells remain in the epiblast will commit to the neuroectodermal lineage (modified from (Tam and Loebel, 2007)).

Neural induction

Epiblast cells that are not recruited to the PS will form the neuroectoderm (Figure 1.3) (Tam and Loebel, 2007). At the onset of gastrulation, these cells possess the ability to develop as either epidermis or neural tissues, however by the end of gastrulation the lineage commitment has completed (Ozair et al., 2013). In vertebrates, the generation of nervous system from neuroectodermal precursor cells is termed neural induction, during which the neuroectoderm forms on the dorsal side of the embryo in response to signals from the adjacent dorsal mesoderm called the "node" in chick and mouse (Stern, 2005). In chick, a population of epiblast cells begins to express Fgf3 and Fgf8 before the formation of the node (Streit et al., 2000; Wilson et al., 2000). With sustained Fgf signaling, cells in the node secret small diffusible molecules- such as noggin, chordin, and follistatin- which bind to extracellular Bmps or Wnts and prevent these ligands from activating their cognate receptors (Hemmati-Brivanlou et al., 1994; Sasai et al., 1994; Smith and Harland, 1992). As a result, ectodermal cells exposed to Bmps and Wnts become epidermis, while they become neural in the absence of those two signals (Moody et al., 2013). By the end of gastrulation, the neuroectoderm cells form the neural plate which subsequently folds into a tube before developing into the brain at its anterior and the spinal cord at its posterior end (Ozair et al., 2013).

Transcription regulation of neural lineage commitment

Once committed in response to extracellular signals, the nascent neuroectodermal precursors begin to express a host of transcription factors which maintain those precursors as neurogenic, expand the neuroectodermal precursor population in the neural plate and initiate the differentiation of committed neural progenitors (Moody et al., 2013). As neural development proceeds, the profile of transcription factors changes progressively, reflecting different stages of

neurogenesis. First, transcription factor Zic1 sensitizes the neuroectodermal precursors to Noggin (Kuo et al., 1998), whereas Churchill and Zeb2 prevent these cells from acquiring mesendodermal fate by limiting the mesodermal-inducing activities of Fgf and Activin/Nodal signals, respectively (Chng et al., 2010; Sheng et al., 2003). At this point, Zfp521 is intrinsically induced (Kamiya et al., 2011). Together with p300, Zfp521 further stabilizes neuroectodermal fate by activating transcription factors Geminin, Irx1, and SoxB1 family members (Sox1, Sox2, and Sox3), which antagonize Wnt and Bmp signaling (Cavodeassi et al., 2001; Kamiya et al., 2011; Kroll et al., 1998; Rogers et al., 2009). Subsequently, Zic2, Geminin and SoxB1 factors facilitate the expansion of neural plate by promoting the proliferation of committed neural progenitors while preventing them from further neural differentiation via repressing bHLH proneural genes (Brewster et al., 1998; Bylund et al., 2003; Seo et al., 2005). Finally, Zic and Irx genes act downstream of the SoxB1 genes to promote the onset of bHLH proneural gene expression, which results in the transition from committed neural progenitors to postmitotic neurons (Gomez-Skarmeta and Modolell, 2002; Houtmeyers et al., 2013).

The experimental observations above summarize the function of the transcription factors during neural lineage commitment from a gene-by-gene perspective. However, it is highly possible that these factors function in a gene regulatory network similar to that in the ESCs. An initial set of transcription factors establishes the identity of neuroectodermal precursors and sustains the expansion of neuroectoderm before being down-regulated. Then another set of factors is induced and converts the neuroectodermal cells to committed neural progenitors. Finally, these neural progenitor factors are down-regulated, and a third group of transcription factors is activated to initiate neural differentiation (Moody et al., 2013). Despite extensive studies in the embryos, the functional interactions among transcription factors controlling neural

specification are not very well understood, and their direct targets are generally unknown. Yet this information is crucial to elucidate the hierarchy of transcription factors within transcriptional cascades controlling neural lineage commitment.

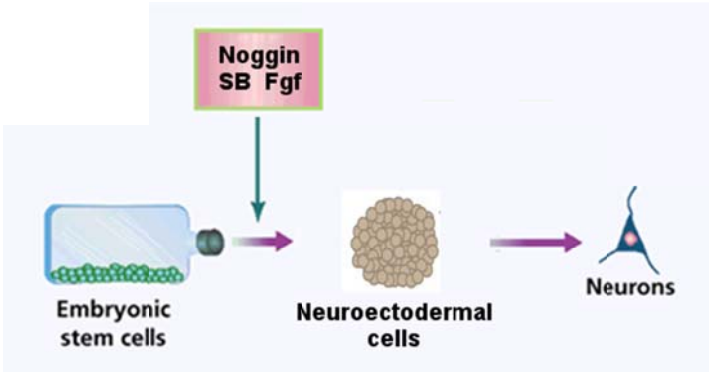
Neural induction in ESCs

There are two principle approaches for induction of neural identity in ESCs. The first method closely recapitulates processes characteristic for neural induction *in vivo* and relies primarily on the inhibition of mesendoderm-inducing SMAD signaling pathways (activated by Nodal and Bmp signals) and activity of endogenous Fgf signals (Figure 1.4-A) (Chambers et al., 2009; Tropepe et al., 2001). Fgf signaling is required for mouse ESCs to progress to a primed state of pluripotency (the primitive ectoderm stage) before becoming competent for neural induction (Kunath et al., 2007; Ying et al., 2003). Inhibition of Nodal pathway down-regulates Nanog and promotes expression of Zeb2 as well as Coup-TFII (Nr2f2), which directly represses Oct4 expression and promotes expression of other neural-specific genes (Chng et al., 2010; Rosa and Brivanlou, 2011). Bmp inhibition contribute to the neuroectodermal differentiation by suppressing the induction of non-neural germ layers, maintaining the expression of shared pluripotency and neural genes like Sox2, and promoting the expression of cell-intrinsic neural determinants, such as Zfp521 (Greber et al., 2011; Kamiya et al., 2011). Together, Fgf and Bmp/Nodal inhibitors consolidate neural lineage commitment through down-regulating the transcription network in ESCs, preventing the induction of mesendoderm determinants, and promoting the onset of neural transcription programs.

The second, historically older approach relies on treatment of ESCs cultured as suspension aggregates termed embroid bodies (EBs) with retinoic acid (RA) (Figure 1.4-B) (Bain and Gottlieb, 1994; Gottlieb and Huettner, 1999). After several days of continued culture, robust

neurogenesis is observed. The changes in morphology and gene expression in EBs bear striking resemblance to the *in vivo* neural development. After the addition of RA, the expression of pluripotency gene Oct4 is rapidly and efficiently extinguished; followed by the emergence of the neuroectodermal marker Sox1. This indicates the termination of the ESC transcription network and the initiation of neuroectodermal lineage commitment. Subsequently, EB cells start to express neural progenitor markers (e.g. Pax6) and stem cell marker Nestin. Finally, postmitotic neurons expressing Tuj1 and NF appear in the culture (Moody et al., 2013). While there is no evidence that RA would be required for neural induction *in vivo*, it has been reported to repress mesodermal marker Brachyury and Wnt-3a in developing mouse embryo (Iulianella et al., 1999). So, similar to the neuralizing method with FGF and SMAD inhibitors, RA-induced neural lineage commitment also involves the down-regulation of the ESC transcription network, the initiation of neural transcription programs, and the suppression of alternative fates. Given the parallel molecular and temporal events between RA-induced neurogenesis and embryonic development, it can serve as a surrogate system to investigate molecular mechanisms of neural induction. Important advantage of the RA driven approach is that the process is anchored by a defined inductive signal that provides convenient entry point for mapping the transcriptional cascade leading to neural specification.

A


Noggin
SB Fgf
Embryonic
stem cells
Neuroectodermal
cells
Neurons

B

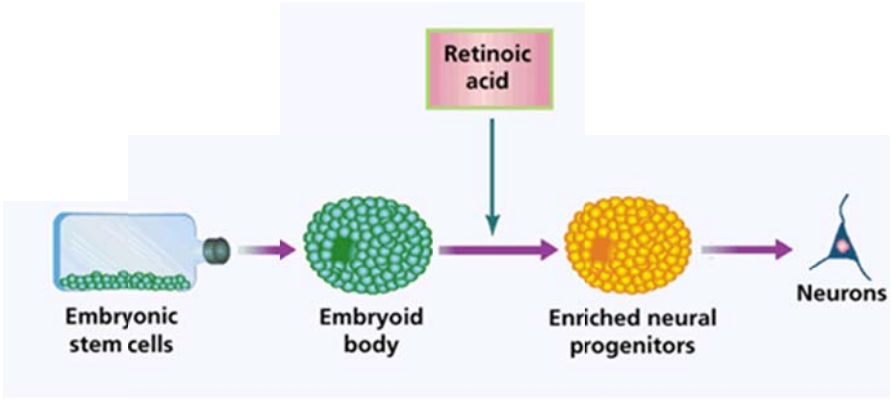

Retinoic
acid
Embryonic
stem cells
Embryoid
body
Enriched neural
progenitors
Neurons

Figure 1.4

Figure 1.4 Neural induction in ESCs

There are two approaches to induce neural differentiation in ESCs. A. The combination of Fgf and inhibitors of both Bmp (Noggin) and Nodal signaling (SB431542, SB) can drive ESCs to become committed neuroectodermal cells and subsequently postmitotic neurons through progressive stages closely reflecting the *in vivo* process (modified from(Gottlieb, 2002; Wada et al., 2009)). B. Alternatively, the ESCs can be cultured as aggregates known as embryoic bodies (EBs) and treated with retinoic acid (RA) to induce neural fate (modified from (Gottlieb, 2002)).

Retinoic acid signaling

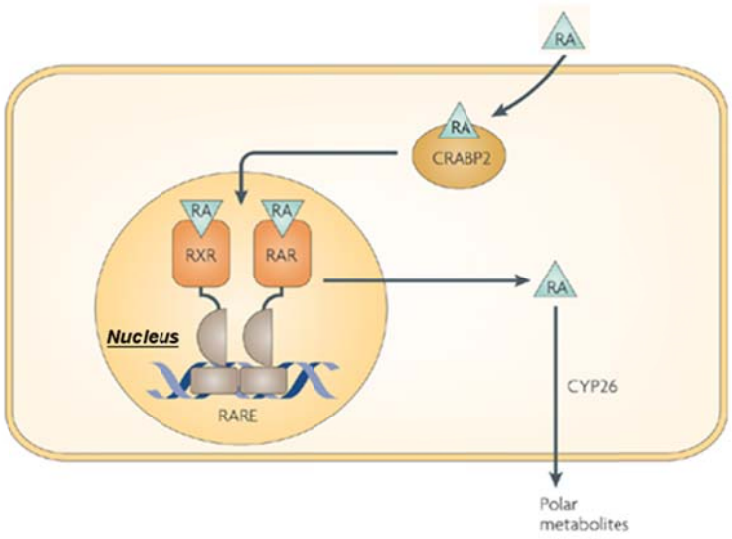
RA is a metabolic product of liposoluble vitamin A (retinol). Most animals cannot synthesize vitamin A so they need to get it from the diet and store as retinoids in the liver (Blomhoff and Blomhoff, 2006). Retinoids are transported as retinol, which is released into the bloodstream and bound to retinol-binding protein 4, plasma (RBP4) before taken up by target cells via membrane receptor for RBP4, STRA6 (Kawaguchi et al., 2007). Once in the cytoplasm, retinol binds to retinol-binding protein1, cellular (RBP1) and is converted to *all-trans* RA by retinol dehydrogenase and retinaldehyde dehydrogenases (RALDHs) (Sandell et al., 2007). The newly-synthesized RA is then bound by cellular retinoic acid binding proteins 1 and 2 (CRABP1 and CRABP2) in the cytoplasm. Subsequently, RA can act in a paracrine manner where it is released from the secreting cells and taken up by receiving cells, or it can act in an autocrine fashion. Either way, RA enters the nucleus with the assistance of CRABP2 and binds to heterodimers of ligand-inducible transcription factors comprising the RA receptors (RARα, RARβ, and RARγ) and retinoic X receptors (RXRα, RXRβ, and RXRγ) (Budhu and Noy, 2002). RA-bound RAR/RXR complexes regulate gene expression through binding to DNA sequences within the promoter of target genes called retinoic acid response elements (RAREs) (Bastien and Rochette-Egly, 2004). After RA has activated RARs, it exits the nucleus and is catabolized by the CYP26 class of P450 enzymes in the cytoplasm (Figure 1.5-A) (Ross and Zolfaghari, 2011).

It has been proposed that in the absence of RA, the apo-receptor heterodimer RAR/RXR binds to the RAREs and recruits corepressors and histone deacetylase complexes (HDACs) to maintain target gene repression (Hart, 2002). In the presence of ligand, a conformational change leads to the replacement of corepressors by co-activator complexes (Figure 1.5-B). This induces chromatin remodeling and facilitates the assembly of the transcription pre-initiation complex,

therefore activating target gene expression (Wei, 2003). A recent study from our lab using RA-mediated ESC differentiation and whole genome chromatin immunoprecipitation sequencing (ChIP-seq) suggests that RA also activates target genes via inducing *de novo* RAR/RXR binding to the RAREs that are not occupied by the apo-receptor dimers (Mahony et al., 2011).

Based on the mechanism of RA-mediated gene expression, there are two criteria for identifying direct target genes of RA signal. First, because RA-bound RAR/RXR induces gene expression independent of *de novo* protein synthesis, direct target genes should respond very quickly to RA signal and exhibit upregulated expression within short timeframe. Second, direct target genes have RAREs bound by RAR/RXR in their regulatory elements and should show enrichment of receptor binding in ChIP-seq experiments.

A


RA
RA
CRABP2
RA
RA
RXR
RAR
Nucleus
RARE
RA
CYP26
Polar
metabolites

B

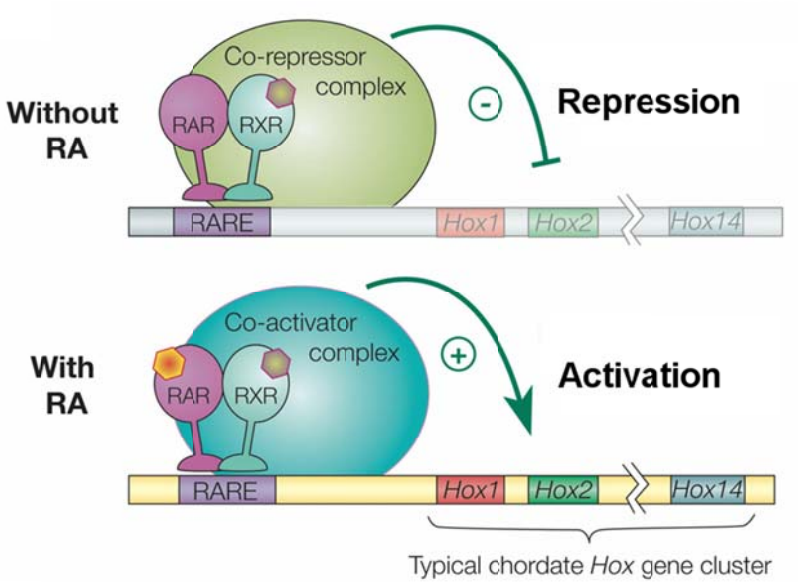

Without
RA
Co-repressor
complex
RAR
RXR
Repression
RARE
Hox1
Hox2
Hox14
With
RA
Co-activator
complex
RAR
RXR
Activation
RARE
Hox1
Hox2
Hox14
Typical chordate Hox gene cluster

Figure 1.5

Figure 1.5 RA signaling

A. After the entry of RA to the cytoplasm, cellular retinoic-acid-binding protein 2 (CRABP2) binds to RA and facilitates the access into the nucleus. In the nucleus, RA binds to RA receptors (RARs) and retinoid X receptors (RXRs), which themselves heterodimerize and bind to a sequence of DNA that is known as the retinoic acid-response element (RARE). This binding activates the transcription of target genes. At the end of the signaling event, RA is catabolized in the cytoplasm by the CYP26 class of P450 enzymes (taken from (Maden, 2007)). B. Without RA, RARE-bound apo-receptor heterodimers associate with corepressor complex and repress gene expression. Upon binding of RA, coactivator complex replaces corepressors and therefore activate target gene expression. (modified from (Marletaz et al., 2006)).

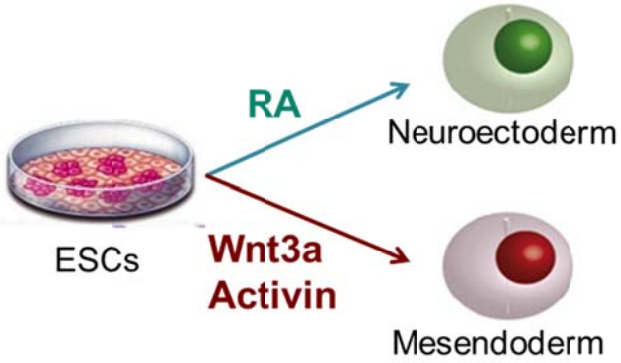
RA-regulated genes in ESC differentiation

RA treatment of ESCs results in both increase and decrease in the expression of a set of genes encoding transcription factors, RA metabolism and transport proteins, extracellular matrix components, protooncogenes, growth factors and their receptors, cytoskeletal proteins, proteins involved in cell metabolism, cell surface antigens, apoptosis-related factors, cell-cycle regulators, and signaling mediators (Balmer and Blomhoff, 2002; Soprano et al., 2007). The analysis of temporal changes in gene expression subdivides the regulated genes into early and late responding categories. A subset of genes is induced rapidly (within 8-16 hours) upon RA treatment in the presence of cycloheximide. These genes are primary response genes and based on ChIP-seq experiments can be further divided into 3 classes (Mahony et al., 2011). The genes in the first class possess RAREs in the proximity (within 20 Kbp) of their transcription start sites bound by the apo-receptor heterodimer RAR/RXR. Cyp26a1, Hoxa1, Cdx1, Meis2, Stra8, Hoxa2 and Rarb belong to this class. The second class of genes, such as Hoxb5, Hoxb1, Hoxa3, Hoxa5, Tshz1, etc., is activated by *de novo* RAR/RXR binding to the RAREs in response to RA treatment. Genes of the third class, including Zfp703, Hoxc4, Kchn1, Zfp503, and Fbp1, do not have proximal RAREs and therefore might be activated indirectly by secondary transcription factors or by distal RAR/RXR bound enhancers. In addition to the primary response genes, a much larger group genes show altered expression at later time points (1 or more days) followed RA exposure, and these changes in expression are dependent on new protein synthesis. These genes, such as Rex-1, Pbx1, Mash-1, NeuroD and N-cadherin, are secondary respondents and therefore regulated indirectly by RA. It is noteworthy that many of the genes in this group are implicated in specific differentiation pathways. Lastly, 5-6 days after RA stimulation, genes marking terminal differentiation are expressed (Soprano et al., 2007).

Mechanism of RA-induced neural lineage commitment

In the developing neural tube, RA signaling initiates neural differentiation (Diez del Corral and Storey, 2004). It is also involved in specifying caudal hindbrain and rostral cervical spinal cord identity (Liu et al., 2001; Niederreither et al., 2000); as well as patterning and differentiation of spinal motor neurons and interneurons (Novitch et al., 2003; Pierani et al., 1999; Sockanathan and Jessell, 1998). As previously mentioned, RA treatment is one of the most common approaches to direct *in vitro* neural differentiation of ESCs that recapitulates multiple important features of neuroectodermal lineage commitment *in vivo*: the termination of the ESC transcription network, the initiation of neuroectodermal transcription programs, and the suppression of mesendodermal differentiation. These phenomena require RA signal to downregulate key transcription factors involved in maintaining pluripotency and specifying mesendodermal fate; meanwhile upregulate transcription factors promoting neuroectodermal formation (Figure 1.6). However, the mechanism through which RA regulates these crucial factors remains poorly understood. Mapping and understanding the operation of transcription regulatory network downstream of RA signaling would provide important insights in the earliest cell fate choice in the embryo proper during development and enable us to devise better methods for cell differentiation to harness the potential of ESCs in regenerative medicine.

A


RA
Neuroectoderm
ESCs
Wnt3a
Activin
Mesendoderm

B

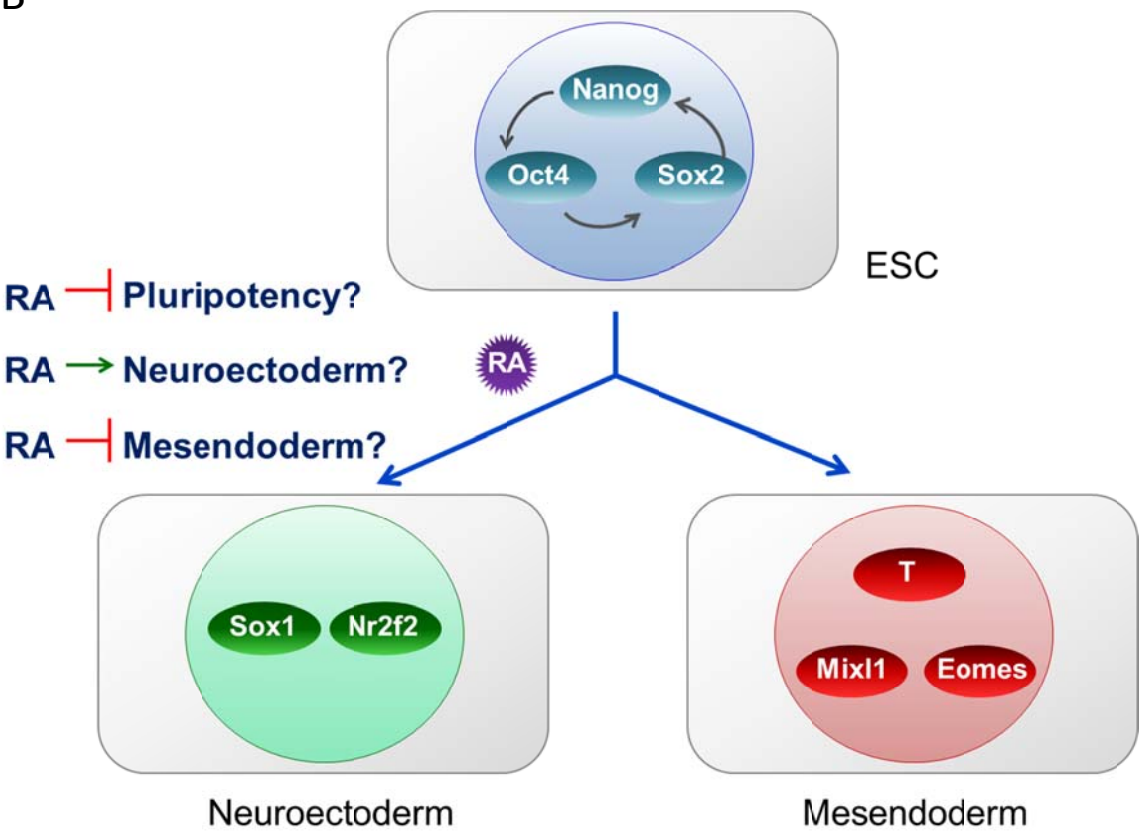

Nanog
Oct4
Sox2
ESC
RA ⊣ Pluripotency?
RA → Neuroectoderm?
RA ⊣ Mesendoderm?
RA
Sox1
Nr2f2
T
Mixl1
Eomes
Neuroectoderm
Mesendoderm

Figure 1.6

Figure 1.6 RA promotes neuroectodermal differentiation of ESCs

A. ESCs can differentiate in vitro along neuroectodermal or mesendodermal lineages in response to RA and Wnt3A/Activin, respectively. B. For RA signal to promote neural differentiation, it has to repress pluripotency (sustained by Oct4, Sox2 and Nanog), suppress mesendodermal fate (marked by T/Brachyury, Mixl1 and Eomes), and promote neuroectodermal specification (induced by Sox1 and Nr2f2). However, molecular mechanism underpinning RA-mediated cell fate choice is not well characterized.

RA represses the ESC transcription network

The first step in ESC differentiation is the termination of the transcription network sustaining the pluripotent state. Indeed, RA treatment of EBs causes a rapid drop in the expression of key transcription factors that maintain ESC identity such as Oct4 and Nanog at both mRNA and protein levels, indicating that these factors are subjected to transcription repression downstream of RA signaling (Shahhoseini et al., 2010). Since the output of RA-bound RAR/RXR dimers on primary response genes is almost exclusively transcriptional activation (Wei, 2003), the repression of pluripotent genes is likely to be an indirect effect of RA, mediated by the transcriptional repressors downstream of RA treatment. Consistent with this notion, time-course expression profiling of differentiating EBs indicates that there is little change in the expression of Oct4 mRNA during the first 8 hours of RA exposure, and the most profound mRNA decrease happens between 8-24 hours after RA addition (Mahony et al., 2011). Furthermore, genome wide ChIP-seq analysis does not identify RAR binding sites in the vicinity of Oct4 promoter region (Mahony et al., 2011). These lines of evidence strongly suggest that Oct4 is regulated indirectly by RA. In order to better characterize the mechanism of RA-mediated repression of the ESC transcription network, it is crucial to identify which of the direct response genes are involved in the repression of pluripotency genes.

The initial attempts to elucidate the mechanism of Oct4 repression by RA in embryonic carcinoma cells (ECs) and ESCs identified several hormone response elements (HREs) in Oct4 enhancer regions required for its repression (Pikarsky et al., 1994). Subsequent investigation revealed that these HREs, despite having features of RAREs, are not directly bound by RAR/RXR. Instead, gel bandshift assay suggested that nuclear orphan receptors Coup-TFs bind to these elements and mediate the repression (Schoorlemmer et al., 1994). COUP-TFs such as

COUP-TF1, Nr2F2 (COUP-TF2), ARP-1 and EAR-2 recognize these HREs and repress Oct4 promoter when overexpressed in ECs and ESCs (Rosa and Brivanlou, 2011; Schoorlemmer et al., 1994). Given the induced expression of Coup-TF1 and ARP-1 during RA-induced ESC differentiation, these results indicate that COUP-TFs might operate as downstream effectors of RA-mediated exit of pluripotency.

Another transcription factor that has been implicated in the repression of Oct4 by RA is the primary response gene Hoxa1. It belongs to the homeobox (Hox) family of proteins that regulates embryonic patterning and organogenesis (Mallo et al., 2010). With the presence of an RARE in its 3' enhancer, Hoxa1 is a direct target of RA signal (Langston et al., 1997). Targeted inactivation of Hoxa1 in mice leads to developmental defects, including hindbrain deficiencies and abnormal skull ossification, which ultimately lead to neonatal death (Chisaka et al., 1992; Mark et al., 1993). Interestingly, RA treatment of Hoxa1 null ESCs results in higher mRNA level of Oct4 when comparing to heterozygous or wild-type ESCs. Upon LIF removal, Hoxa1 null ESCs shows reduced Fgf5 expression and elevated Gata4, Sox17 and Dab2 levels (Martinez-Ceballos et al., 2005), suggesting that instead of differentiating along the neuroectodermal lineage as wild-type ESCs, Hoxa1 mutant cells are prone to follow an endodermal pathway and failed to differentiate into neurons (Martinez-Ceballos and Gudas, 2008). These cells express several-fold lower levels of many neural differentiation markers such as nestin, β-tubulin III, and MAP2 than wild-type cells. Conversely, higher levels of endodermal markers (i.e. Sox17 and Col4a1) are observed. Importantly, re-introduction of exogenous Hoxa1 to the null cells restores their ability to produce neurons. Hence these observations indicate that Hoxa1 might be a critical downstream mediator of RA-induced neural lineage commitment.

Lastly, the orphan nuclear receptor GCNF has also been proposed to be required for the repression of pluripotency genes in RA-induced ESC differentiation (Gu et al., 2005). GCNF is a transcriptional repressor that down-regulates target genes by binding to responsive elements in the promoter region (Cooney et al., 1998). During mouse development, the expression pattern of GCNF is inversely correlated with that of Oct4; and Oct4 expression is not efficiently turned off in somatic cells at gastrulation in GCNF null mutants (Fuhrmann et al., 2001). The mutant embryos subsequently die around E10.5 (Chung et al., 2001). In wild-type ESCs, GCNF protein is induced by RA treatment, peaks between 24-36 hours, and drops to undetectable levels after 3 days. While Oct4 protein is dramatically decreased in wild-type cells between 24-72 hours after RA induction, increased level of Oct4 and other pluripotency genes such as, Nanog, Sox2, FGF4, and Stella is found in GCNF mutant ESCs (Gu et al., 2005). GCNF binds to response elements located in the promoter of Oct4, as well as the promoter and 3' untranslated region of Nanog (Fuhrmann et al., 2001; Gu et al., 2005). When co-expressed with luciferase constructs containing these elements, GCNF is capable of repressing the luciferase activity (Gu et al., 2005). Together, these results suggest that GCNF mediates the repression of ESC transcription network downstream of RA signal by directly repressing Oct4 and Nanog. Recently, GCNF has been shown to be important for the transition from primitive to definitive neural stem cells in the early neural stem cell lineage through the repression of Oct4 (Akamatsu et al., 2009), indicating that GCNF might be involved in multiple aspects of RA-induced neural lineage commitment.

Transcriptional network in RA-induced neuroectodermal lineage commitment

So far, COUP-TFs, Hoxa1, and GCNF have all been identified as potential downstream effectors of RA-induced repression of pluripotency. And both Hoxa1 and GCNF seem to be required for the effective down-regulation of Oct4 and other key ESC transcription factors at the

onset of differentiation. However, loss of either Hoxa1 or GCNF does not completely abolish the repression of Oct4 by RA, suggesting that Hoxa1 and GCNF might be able to compensate each other or be compensated by additional factor. Like the key transcription factors that act cooperatively in a regulatory network to maintain ESC identity and specify neural fate, it is possible that RA induces neuroectodermal lineage commitment through initiating a transcriptional network consisting of groups of primary and secondary respondents to effectively terminate the pluripotency network and promote neural differentiation. Moreover, little is currently known about mechanisms mediating suppression of the mesendodermal fate in RA-mediated neural induction. Therefore, it will be important to map the transcriptional cascade downstream of RA signaling in greater detail to delineate the key steps and regulatory interactions leading to efficient acquisition of neural identity.

NET family transcription factors

The transcription factors identified as putative downstream effectors of RA-induced neural lineage commitment in this study are zinc finger proteins Zfp703 and Zfp503. They belong to the NET family (Noc, Nlz, Elbow, and Tlp-1) of C_2H_2 zinc finger proteins important for diverse developmental processes across multiple organisms (Nakamura et al., 2004). In drosophila, *noc* mutants display defect in the light sensitive organs (the ocelli) (Woodruff and Ashburner, 1979a, b) and embryonic superaesophageal ganglion (Cheah et al., 1994); whereas *elbow* mutant flies have small bent wings (Davis et al., 1997). Homozygous *elbow* and *noc* mutants also exhibit defects in tracheal development, specifically stalled and aberrant migration of the dorsal branch and lateral trunks of the trachea (Dorfman et al., 2002). Meanwhile, mutations in *tlp-1* affect specification of asymmetric cell fate and cell fusion, causing abnormal tail morphogenesis in C. elegans (Zhao et al., 2002). Lastly, in zebrafish, *nlz1* (also known as Zfp703) and *nlz2* (also

known as Zfp503) genes are required for proper patterning and development of the hindbrain (Hoyle et al., 2004; Runko and Sagerstrom, 2003).

The NET family proteins are related to the Sp/Buttonhead family of transcription factors with which they share three conserved domains: an N-terminal Sp motif that might regulate protein degradation (Su et al., 1999) or transcriptional activity (Murata et al., 1994), a Buttonhead (Btd) box which appears to be required for transcriptional activation in some cases (Athanikar et al., 1997), and a C_2H_2 zinc finger (Suske, 1999) (Figure 1.6). While proteins in the Sp family bind GC-rich DNA sequences with three tandem zinc fingers (Kadonaga et al., 1987), it is not clear whether NET proteins can bind DNA via its single zinc finger. First, usually 2-4 C_2H_2 zinc fingers are required for efficient DNA binding (Iuchi, 2001). Single zinc fingers in GATA and GAGA family proteins require adjacent basic domains, which NET proteins lack, to interact with DNA (Omichinski et al., 1997; Pedone et al., 1996; Pedone et al., 1997). Moreover, although Nlz proteins can form homo- or heterodimers, thereby bringing multiple zinc fingers together, such complexes are not required for Nlz function (Runko and Sagerstrom, 2004). Finally, among four conserved residues involved in contacting DNA within DNA-binding C_2H_2 zinc fingers, only two residues are conserved in the C_2H_2 zinc finger of the NET proteins (Nakamura et al., 2004; Wolfe et al., 2000). Therefore these observations indicate that the C_2H_2 zinc finger of the NET proteins does not bind DNA *per se* but might mediate protein-protein interactions.

While it is not known whether NET proteins interact with DNA, it is likely that they control transcription as transcriptional co-repressors or repressors. First, Elbow, Tlp-1, Nlz1, and Nlz2 are located in the nucleus (Dorfman et al., 2002; Runko and Sagerstrom, 2003, 2004; Zhao et al., 2002), and Nlz1 must be nuclear to be fully functional (Runko and Sagerstrom, 2004). Second,

ectopic Elbow represses the expression of tracheal genes in Drosophila, and Nlz1 or Nlz2 overexpression leads to loss of gene expression in the rostral hindbrain of zebrafish (Hoyle et al., 2004; Runko and Sagerstrom, 2003, 2004). Conversely, *elbow* and *noc* mutant flies exhibit an expansion in the expression of tracheal branch-specific genes (Dorfman et al., 2002), and the expression of a dominant-negative form of Nlz1 results in the expansion of rhombomere 5-specific gene expression (Runko and Sagerstrom, 2003). Furthermore, Gal4-Nlz1 and Gal4-Nlz2 fusion proteins repress transcription from Gal4-responsive promoters in both cell lines and zebrafish embryos (Runko and Sagerstrom, 2003). Lastly, Elbow, Nlz1 and Nlz2 bind to the corepressor Groucho through the repressor interaction domain containing a serine/threonine-rich region, and Nlz1 and Nlz2 also interact with histone deacetylases HDAC1 and HDAC2 (Dorfman et al., 2002; Runko and Sagerstrom, 2003, 2004). Together, this suggests that NET family proteins act primarily as transcriptional repressors or co-repressors.

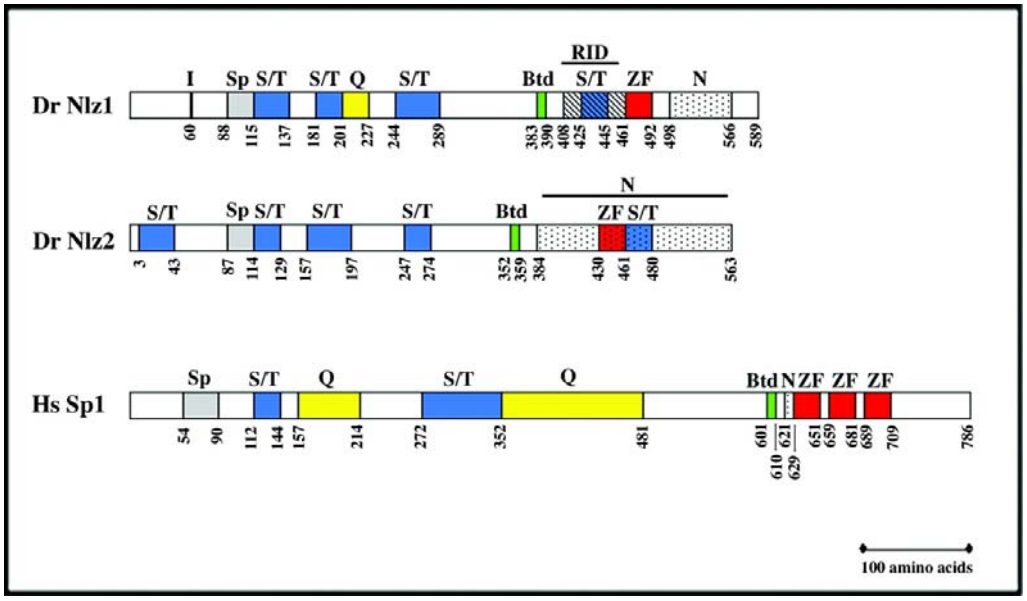

Dr Nlz1
I
Sp
S/T
S/T
Q
S/T
Btd
RID
S/T
ZF
N
60
88
115
137
181
201
227
244
289
383
390
408
425
445
461
492
498
566
589
Dr Nlz2
S/T
Sp
S/T
S/T
S/T
Btd
N
ZF
S/T
3
43
87
114
129
157
197
247
274
352
359
384
430
461
480
563
Hs Sp1
Sp
S/T
Q
S/T
Q
Btd
N
ZF
ZF
ZF
54
90
112
144
157
214
272
352
481
601
610
621
629
651
659
681
689
709
786
100 amino acids

Figure 1.7

Figure 1.7 Net family proteins share conserved domains

Domains of the zebrafish Nlz1, Nlz2 and the human Sp1 proteins are shown. The proteins are shown as **white boxes,** except for the internal start site (**I; black**), the Sp motif (**Sp; gray**), serine/threonine-rich regions (**S/T; blue**), the glutamine-rich region (**Q; yellow**), the Btd box (**green**), the C_2H_2 zinc finger (**ZF; red**), the repressor interaction domain (**RID; hatched**), and a region required for nuclear localization (**N; stippled**). **Numbers** indicate amino acid positions. **Dr, Danio rerio; Hs, Homo sapiens (taken from (Runko and Sagerstrom, 2004)).**

Zfp703 as oncogene in tumor progression

Zfp703 (ZNF703 in human) is the mammalian homolog of Nlz1. Human ZNF703 is located in the proximal chromosome arm 8p (8p12) within a region that is commonly amplified in estrogen-positive (ER^+) breast cancers (Chin et al., 2006). Expression profile of primary breast tumors revealed that ZNF703 is amplified predominately in the Luminal B subtype, with increased gene and protein expression associated with poor clinical outcomes (Chin et al., 2007; Reynisdottir et al., 2013; Sircoulomb et al., 2011). ZNF703 is capable of transforming NIH 3T3 fibroblasts and regulate cell proliferation in human breast epithelial cells (Holland et al., 2011). Interestingly, ZNF703 binds to the promoter of TGFBR2 together with HDAC1 and mediates repression of the receptor therefore modulating the effect of TGF-β signaling on the proliferation of breast tumor cell line MCF7 (Holland et al., 2011). The data above suggests that ZNF703 acts as a novel oncogene in Luminal B breast cancer. Recently, ZNF703 has also been shown to act as an oncogene to promote gastric cancer progression (Yang et al., 2014).

Mouse Zfp703 (also known as Zeppo1) shares 96% amino acids with human ZNF703, and it is expressed in the mammary epithelium (Slorach et al., 2011). Similar to the human protein, Zfp703 overexpression in a mouse breast cancer model increases lung metastasis; while Zfp703 knockdown has the opposite effect (Slorach et al., 2011). Zfp703 reduces cell-cell adhesion and stimulates cell migration and proliferation (Slorach et al., 2011). It regulates transcription, interacts with Groucho, and represses E-cadherin as well as Wnt and TGF-β reporters (Slorach et al., 2011). Together, this indicates that Zfp703 is a key regulator of breast cancer progression.

Despite the emerging evidence supporting the oncogenic characteristics of Zfp703, little is known about its function during normal mammalian development. Nevertheless, it is highly possible that Zfp703 participates in embryogenesis, particularly in the development of the

nervous system. First, Zfp703 homolog in Drosophila, *nocA*, has been involved in the development of embryonic brain (the superaesophageal ganglion) (Cheah et al., 1994). Second, the homolog of Zfp703 in zebrafish (Nlz1) plays an important role in the specification of rhombomere identity in developing hindbrain (Hoyle et al., 2004). Finally, its closely related paralog in the same family, Zfp503 (Nolz1) is implicated in the specification of motor neurons (Ji et al., 2009) and striatal projection neurons in mice (Ko et al., 2013; Urban et al., 2010).

Dissertation outline

Multiple studies have examined RA-mediated neural differentiation and proposed several transcription factors, including COUP-TFs, Hoxa1, and GCNF, as repressors involved in Oct4 repression. However, my analysis of RA targets indicated that besides Hoxa1, COUP-TFs and GCNF are not primary targets of RA signaling, raising the questions how they are regulated and whether additional factors contribute to the effects of RA. Moreover, it is not clear how RA signaling initiates neural transcription program and suppress non-neural (mesendoderm) differentiation, two premises for consolidating neural identity. To revisit these questions and build more accurate transcriptional map reflecting the cascade of events triggered by RA signaling, I performed a screen that identified Zfp703 as a new factor downstream of RA that exhibited the strongest effect on the repression of Oct4. Next I identified putative target genes of Zfp703, examined whether Zfp703 controls neural lineage specification, and probed its ability to repress mesendodermal genes. Based on the results I propose a revised model of the hierarchy of regulatory events downstream of RA signaling during neuroectodermal lineage commitment.

Chapter 2. Screen for primary response genes of RA signaling that repress pluripotency factor Oct4

Introduction

RA treatment is the most commonly adopted approach to differentiate ESCs along the neuroectodermal lineage (Bain and Gottlieb, 1994; Gottlieb and Huettner, 1999). However, molecular mechanisms underlying RA-mediated neural induction remain elusive. Since the ESC state (i.e. self-renewal and pluripotency) is sustained by the transcriptional network comprising the Oct4, Sox2, and Nanog feed forward loop, neural induction has to be accompanied by the termination of this transcription program (Jaenisch and Young, 2008; Nichols and Smith, 2012). Originally, it was proposed that RA extinguishes pluripotency by repressing Oct4 via the direct interaction of RAR/RXR with the HREs in its promoter (Pikarsky et al., 1994). But this notion was ruled out based on the following observations: (1) Oct4 promoter HREs are not bound by RAR/RXR (Schoorlemmer et al., 1994); (2) RA-bound RAR/RXRs almost always act as transcriptional activators rather than repressors (Wei, 2003); and (3) the delayed change in expression of Oct4 suggests that it is not a direct target of RA (Mahony et al., 2011). Therefore, the most parsimonious hypothesis is that RA signal terminates the ESC transcriptional network indirectly through the induction transcriptional repressors of Oct4 and other pluripotency factors.

Previous attempts to characterize Oct4 repressors downstream of RA have identified several genes participating in the process. COUP-TFs such as COUP-TF1, ARP-1 and EAR-2 recognize Oct4 promoter HREs and function as repressors in Oct4 promoter reporter assays (Schoorlemmer et al., 1994). Moreover, both Hoxa1 and GCNF have been shown to be required for effective Oct4 repression and neural differentiation following RA treatment (Akamatsu et al., 2009; Gu et al., 2005; Martinez-Ceballos et al., 2005; Martinez-Ceballos and Gudas, 2008). However, loss of either Hoxa1 or GCNF does not completely abolish the repression of Oct4 by RA, suggesting

that they might act redundantly, and additional factors likely contribute to RA-mediated termination of the ESC state. Importantly, the exit of pluripotency is just the first step of neural lineage commitment. In order to consolidate the neuronal fate, RA must also be able to induce the neural transcription program and suppress the alternative mesendodermal fate. So far, little is known about the effectors and mechanisms in these processes. Therefore to better understand the transcriptional network underpinning RA-mediated neuroectodermal lineage commitment, it is critical to identify additional RA target genes involved in the repression of pluripotecy, the induction of neuronal fate, and the suppression of mesendodermal differentiation.

In this chapter, I performed a small-scale screen of RA primary response genes to identify putative downstream effectors of RA-mediated neuroectodermal lineage commitment. I first examined the dynamics in expression levels of Oct4 and Sox1 following RA treatment. Next I selected a set of seven candidate primary response genes involved in the regulation of transcription, cloned their cDNAs and generated inducible ESC lines to test the effects of their expression in the absence of RA signaling on Oct4 expression levels.

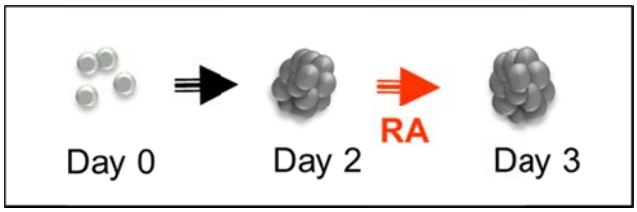
Results

Delayed response in Oct4 and Sox1 expression upon RA treatment

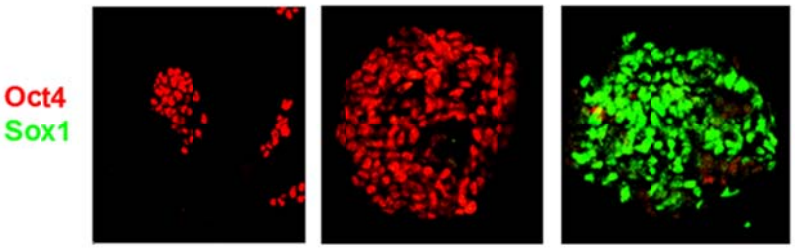
To identify additional players in RA-mediated neuroectodermal lineage commitment, I adopted the *in vitro* differentiation protocol of mouse ESCs (Wichterle et al., 2002). In this protocol EBs are treated with 1μM RA two days (Day2, Figure 2.1-A) after the start of differentiation (as opposed to four days in the original protocol developed by Bain et al. (Bain and Gottlieb, 1994)), at a point when cells still express uniformly high levels of the pluripotency marker Oct4 (Figure 2.1-B). One day later (Day3) Oct4 was significantly downregulated and simultaneously early neural marker Sox1 was up-regulated in EBs; while EBs not retreated with RA maintained high level of Oct4 expression and exhibited minimal induction of Sox1 (Figure 2.4-A, Ctrl). Importantly, Sox1 and Oct4 were expressed in mutually exclusive pattern upon RA treatment, indicating that in one day RA has efficiently directed many cells to exit the ESC state and commit to neural lineage. To examine the dynamics of Oct4 downregulation and Sox1 upregulation, I analyzed mRNA levels by collecting samples at 0, 8, and 24 hours following the addition of RA and performed qRT-PCR (Figure 2.1-C). There was no significant difference in Oct4 mRNA level between RA treated and untreated cells (p value: 0.15) during the first 8 hours. However, 16 hours later Oct4 mRNA level in RA-treated cells was significantly lower than that in control cells (7-fold; p value: 0.0004), suggesting that the most profound decrease of Oct4 mRNA occurred between 8-24 hours after RA was added. Interestingly, Sox1 was induced in a similar pattern, with small difference in mRNA level in RA-treated cells compared to untreated cells (1.49 fold; p value 0.02) 8 hours following RA treatment and more significant increase in RA-treated cells at 24 hours following RA treatment (7 folds more than untreated cells; p value: 0.004). This kinetics was markedly different from the regulation of primary RA target genes such

as Hoxa1 that exhibited peak of induction already at 8 hours after RA treatment (Figure 2.1-C). The delayed response in Oct4 and Sox1 mRNA expression upon RA treatment suggests that Oct4 and Sox1 are not direct targets of RA signaling. Given the central role of Oct4 and Sox1 in the transcriptional program in ESCs and neural progenitors, this result indicates that RA promotes the transition from ESCs to neural progenitors through induction of primary response genes that directly or indirectly repress Oct4 and induce Sox1.

A



B



C

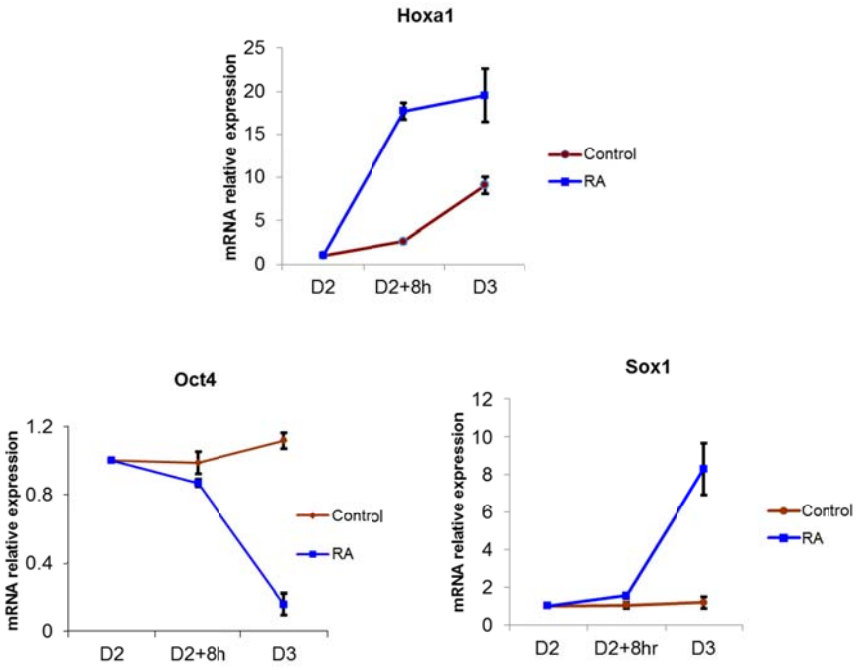


Figure 2.1

Figure 2.1 RA treatment promotes commitment of neuroectodermal fate

Mouse ESCs were differentiated as embroid bodies (EBs) for 2 days before treatment of RA (1μM). The expression of Oct4 and Sox1 proteins were analyzed on Day0, Day2, and Day3 of differentiation. And the mRNA of Hoxa1, Oct4 and Sox1 was analyzed by qRT-PCR at 0, 8, and 24 hours following the addition of RA. A. Schematic of ESC differentiation. B. Cross sections of ESCs on Day0 as well as EBs on Day2 and Day3. Oct4 protein was expressed uniformly from Day0 to Day2. 24 hours after RA treatment, Oct4 protein was quickly extinguished, and a concomitant emergence of Sox1-positive cells can be observed in EBs, suggesting that these cells have committed to the neuroectodermal lineage in response to RA signal. C. Time-course qRT-PCR of Hoxa1, Oct4 and Sox1through the differentiation. Whereas a typical RA target gene such as Hoxa1 was induced within 8 hours of RA treatment, the expression level of both Oct4 and Sox1 showed little change through the same period. Significant downregulation of Oct4 and upregulation of Sox1 occurred between Day2+8h and Day3, suggesting that Oct4 and Sox1 are not direct targets of RA (3 independent experiments; mRNA relative expression levels were shown as average ± standard deviation).

Candidate RA target genes regulating transcription

Since the first premise of lineage commitment regardless of terminal fates is the exit of the ESC state. I began my project by identifying putative primary response genes that act as downstream effectors of RA to repress Oct4. I analyzed Affymetrix expression data of cDNA libraries obtained from differentiating ESCs on Day0, Day2 prior to RA treatment, Day2 8h after the addition of RA, and Day3 (Mahony et al., 2011). First I checked the expression patterns of the reported repressors of Oct4: GCNF, COUP-TFs (Nr2f1 and Nr2F2), and Hoxa1 (Figure 2.2). Interestingly, the expression of COUP-TFs peaked on Day3 of differentiation when Oct4 repression has already been completed. Meanwhile, GCNF was already expressed in ESCs and increased only modestly (< 2 fold) following RA treatment, indicating that it unlikely initiates Oct4 repression in response to RA. Hoxa1 was barely detectable from Day0 to Day2 and was quickly induced after RA addiction. Its expression reached the highest level after 8 hours of RA treatment. This pattern of repression best matched the profile of putative Oct4 repressors.

To identify additional candidate transcription factors with similar expression profile, I filtered genes with Log2 change expression greater than 2.5 for ones annotated by Gene Ontology as regulators of transcription (Table 2.1). Besides promoting neural differentiation, RA is also involved in the specification of rostro-caudal identity (Liu et al., 2001). Indeed, 14 of the 20 transcription regulators induced by RA are Hox factors or co-factors previously shown to be regulated by RA signaling (Mahony et al., 2011).

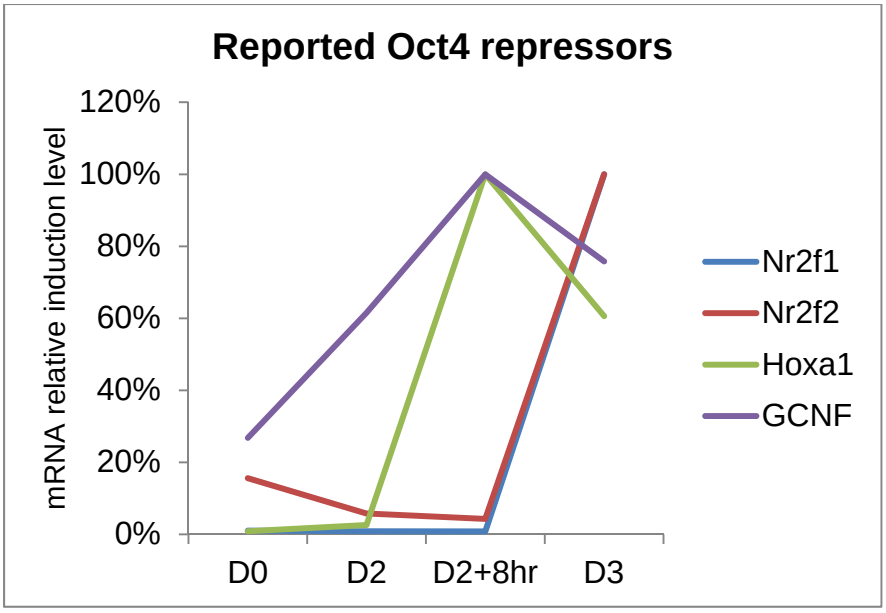

Reported Oct4 repressors
mRNA relative induction level
120%
100%
80%
60%
40%
20%
0%
D0
D2
D2+8hr
D3
Nr2f1
Nr2f2
Hoxa1
GCNF

Figure 2.2

Figure 2.2 Dynamics in the expression of reported effectors downstream of RA-mediated Oct4 repression

Mouse ESCs were differentiated as embroid bodies (EBs) for 2 days before treatment of RA (1μM). Total mRNA samples were collected at different time points from Day0 to Day3 and analyzed by microarray profiling (Mahony et al., 2011). Relative expression levels of reported effectors responsible for RA-mediated Oct4 repression, including COUP-TFs (Nr2f1 and Nr2f2), Hoxa1, and GCNF, were obtained on Day0, Day2, Day2+8hrs, and Day3, then normalized to their the maximize expression level among these time points. The induction pattern of Hoxa1 confirmed that it is a typical primary response gene and direct target of RA, with rapid upregulation during the first 8 hours of RA treatment followed by downregulation as the cells proceeded to Day3. On the other hand, both Coup-TFs were not induced between Day2 and Day2+8hr, suggesting their upregulation might be a secondary effect of RA. While the expression of GCNF peaked at 8 hours after RA treatment, its induction began before the addition of RA, suggesting that RA signal might not have major effect on GCNF expression (Microarray profiling and analysis done by E. Mazzoni and S. Mahony (Mahony et al., 2011)).

Table 2.1 Genes annotated by Gene Ontology Biological Process as regulators of transcription with Log2 expression change >2.5 at 8 hours after RA treatment.

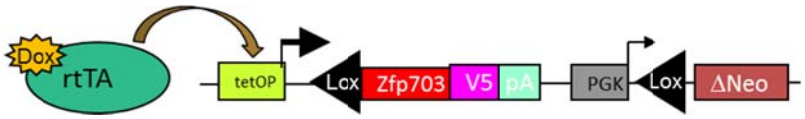
Shaded are Hox factors and co-factors. Highlighted are genes cloned and tested for Oct4 repression. (Microarray profiling and analysis done by E. Mazzoni and S. Mahony (Mahony et al., 2011))

Probe	Foldchange [Log2]	P-value	Day2-RA RAR binding site within 20Kbp	Day2+RA RAR binding site within 20Kbp of TSS
Zfp703	7.98	0.000001		
Hoxb5	7.74	0.000001		✓
Hoxb1	7.57	0.000007		✓
Hoxa1	5.56	0.000002	✓	✓
Hoxb4	5.49	0.000036		✓
Cdx1	5.44	0.000000	✓	✓
Hoxa3	5.17	0.001039		✓
Hoxa5	4.98	0.000001		✓
Hoxb6	4.70	0.001732		✓
Hoxa4	4.53	0.000059		✓
Meis2	4.11	0.000041	✓	✓
Hoxb2	4.07	0.000069		✓
Hoxc4	3.87	0.000001		
Hoxb3	3.86	0.000126		✓
Hoxa2	3.54	0.000321	✓	✓
Tshz1	3.54	0.004358		✓
Rarb	3.52	0.000112	✓	✓
Nrip1	2.98	0.000072		
Zfp503	2.97	0.009475		
Hoxa10	2.91	0.000112		

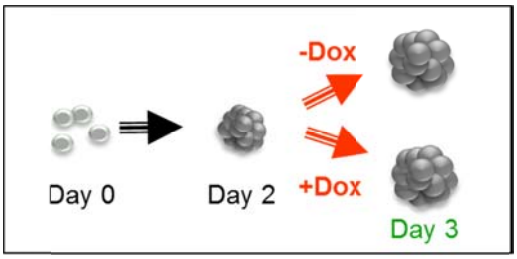
Zfp703 is sufficient to repress Oct4 on the transcription level

To test the ability of identified candidate transcriptional regulators to repress Oct4 expression, I decided to focus on factors other than Hox genes and their co-factors with the exception of Hoxa1, which has been previously implicated in repression of Oct4. I cloned *Hoxa1* and five of the six non-Hox genes (all except Nrip1), V5 epitope-tagged at C-terminus, and recombined their cDNAs downstream of the TetO promoter to generate six doxycycline (Dox) inducible ESC lines (Figure 2.3-A). Treatment of the cell lines with Dox led to robust expression of V5 tagged factors 24 hours later (Figure 2.3-C). To test the ability of these factors to modulate Oct4 expression, I differentiated the 6 lines as EBs. On Day 2 I split the EBs into four dishes – one kept as a control, one treated with RA, one treated with Dox and one treated with both Dox and RA. Cells were harvested 24 hours later on Day 3, and Oct4 mRNA was quantified by qPCR (Figure 2.3-B). As shown in Figure 2.3-D, Cdx1 and the previously-identified Oct4 repressor Hoxa1 reduced Oct4 mRNA only moderately by ~30% (p values: 0.06 and 0.06, respectively). Importantly I identified two new factors - Zfp703 and Zfp503 that were sufficient to repress Oct4 mRNA by ~80% (p values: 0.01 and 0.03, respectively). The effect of Zfp703 and 503 expression was not significantly different from the effect of RA treatment (p value: 0.33 for Zfp703; 0.13 for Zfp503). This result suggests that Zfp703 and Zfp503 might act as primary downstream effectors of RA-mediated termination of ESC transcription network.

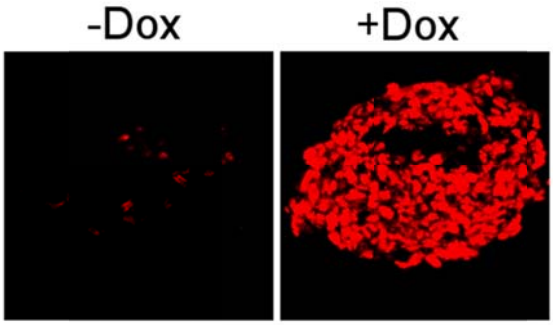
A



B



C



D

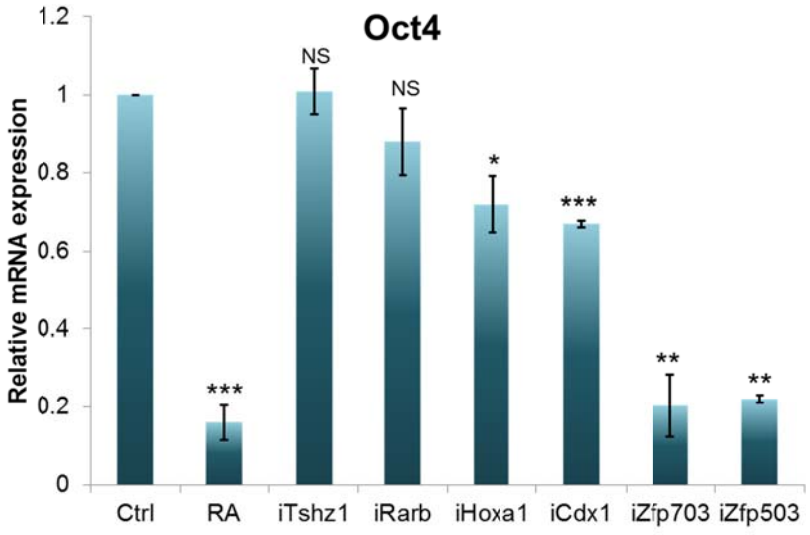


Figure 2.3

Figure 2.3 Using inducible ESC lines to screen for additional effectors of RA

A panel of candidate transcription factors induced by RA within 8 hours was cloned into a TetO inducible vector (Mazzoni et al., 2011), which then was used to construct inducible ESC lines for each factor. Subsequently, these inducible ESCs were differentiated as EBs for 2 days before treated with RA (1μM) or doxycycline (1.5μg/ml). The expression of Oct4 mRNA was analyzed by qRT-PCR on Day3. A. Schematic of the inducible vector. B. Schematic of ESC differentiation. C. Expression of candidate TFs in EBs revealed by V5-tag immunostaining 24 hours following doxycycline treatment, confirming the robust induction of the inducible ESC system. D. Oct4 mRNA expression in non-treated and doxycycline- or RA-treated inducible EBs on Day3. While expression of Tshz1 and Rarb had no effect on Oct4, Hoxa1 and Cdx1 only led to moderate decrease in Oct4 mRNA. Importantly, Zfp703 and Zfp503 induction resulted in significant downregulation of Oct4. (3 independent experiments; mRNA relative expression levels were shown as percentage values normalized to –Dox control value ± standard deviation; *$p<0.05$; **$p<0.01$; ***$p<0.001$).

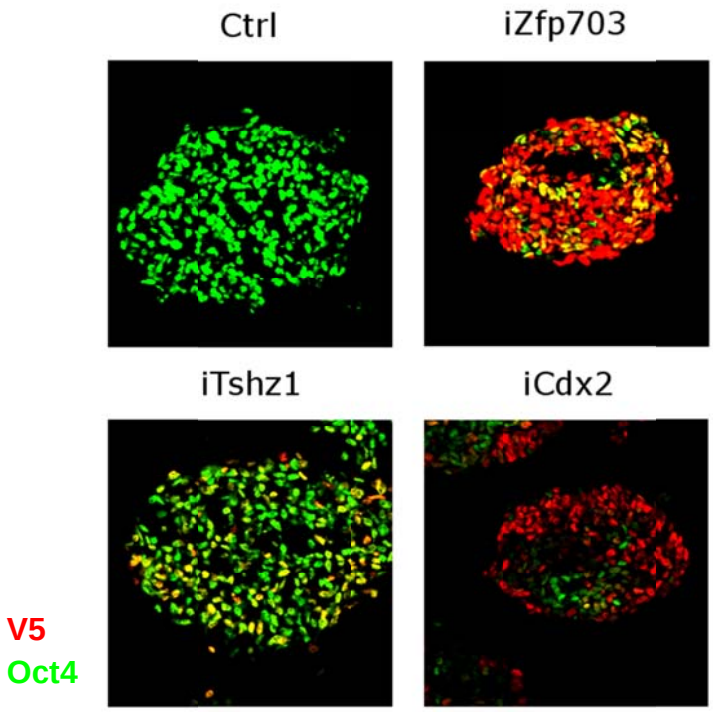
Zfp703 is sufficient to repress Oct4 protein

To examine effects of inducible expression of transcriptional regulators on Oct4 protein level and to study the effect with cellular resolution, I performed immunostaining of Day3 EBs with antibodies against Oct4 and anti-V5 antibodies to visualize the expression of the epitope-tagged induced factors. As shown in Figure 2.4-A, the addition of doxycycline resulted in various degrees of induction in different ESC lines. While the V5 appeared patchy in iTshz1 EBs, the majority of iZfp703 and iCdx2 cells expressed the induced factor following doxycycline treatment. Zfp703 expression resulted in significant decrease in the number of cells expressing Oct4 protein, so did the induction of Cdx2, a well-established Oct4 repressor (Wang et al., 2010). Meanwhile, Tshz1 expression led to similar protein level of Oct4 as untreated control. To further analyze the correlation between the candidate factors and Oct4, I quantified the proportion of cells co-expressing the candidate factor and Oct4 protein over the total number of cells expressing the candidate factor. As shown in Figure 2.4-B, more than 90% of the cells expressing Tshz1 were Oct4-positive, whereas only 20% of Zfp703-positive cells and less than 10% of Cdx2-positive cells expressed Oct4, suggesting that similar to Cdx2, Zfp703 is sufficient to repress Oct4 protein when expressed in EBs. The fact that the majority Zfp703-expressing cells did not express Oct4 suggested that the repression of Oct4 by Zfp703 is likely cell-autonomous. However, the level of Oct4 protein did decrease in some cells in iZfp703 EBs that were V5- negative, arguing the possibility of non-cell-autonomous effects of Zfp703 on Oct4.

Although Zfp503 expression was also sufficient to repress Oct4 mRNA on Day3 (Figure 2.3-C), closer observation by immunostaining revealed that the EBs looked unhealthy with increased cell death (Figure 2.5) thus making the results with iZfp503 difficult to interpret.

Together with the fact that Zfp703 was induced much higher than Zfp503 (Table 2.1), I hence focused the remaining study on Zfp703.

A


Ctrl
iZfp703
iTshz1
iCdx2
V5
Oct4

B

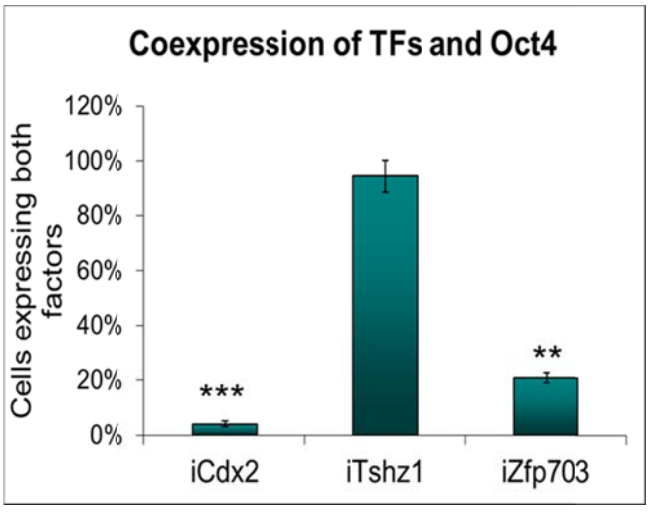

Coexpression of TFs and Oct4
Cells expressing both factors
120%
100%
80%
60%
40%
20%
0%
***
**
iCdx2
iTshz1
iZfp703

Figure 2.4

Figure 2.4 Zfp703 represses Oct4 protein

Inducible ESC lines expressing Zfp703, Tshz1, and Cdx2 were differentiated as EBs and treated with doxycycline on Day2. Then the protein expression of Oct4 and the induced factors was analyzed by immunostaining 24 hours later on Day3. A. Cross sections of Day3 EBs which were either untreated or treated with doxycycline therefore expressing Zfp703, Tshz1, or Cdx2, respectively. Expression of Zfp703 and Cdx2 led to marked decrease in Oct4-positive cells in EBs, whereas extensive Oct4 protein expression sustained in untreated EBs and EBs expressing Tshz1. B. quantification of cells co-expressing Oct4 and the induced factors. About 20% of Zfp703-expressing cells and less than 5% of Cdx2-expressing cells were Oct4-positive. While more than 90% of Tshz1-expressing cells also expressed Oct4. These results suggested that both Zfp703 and Cdx2 can effectively repress Oct4 on the protein level (3 independent experiments; numbers of cells expressing both induced TFs and Oct4 relative to all TF-expressing cells were shown as average percentage ± standard deviation; **$p<0.01$; ***$p<0.001$).

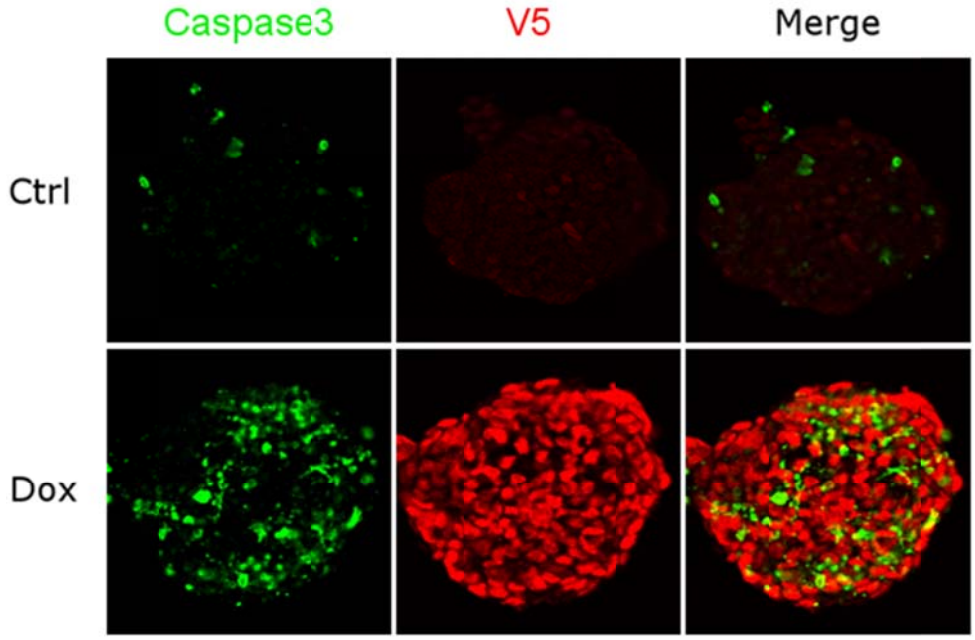


Figure 2.5

Figure 2.5 Zfp503 induction leads to cell death

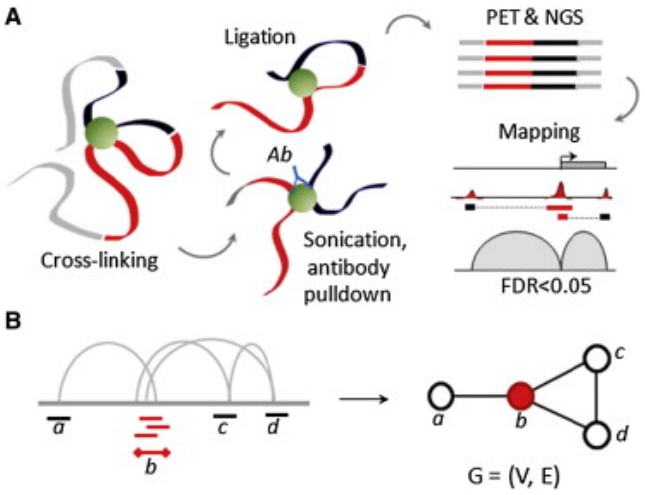
Inducible ESCs expressing Zfp503 were differentiated as EBs and treated with doxycycline on Day2. Then the expression of V5-tagged Zfp503 and cleaved active Caspase 3 was analyzed by immunostaining 24 hours later on Day3. Cross sections of untreated or doxycycline-treated EBs were shown. Zfp503 expression resulted in increased level of cleaved Caspase 3 in EBs, suggesting that Zfp503 induces cell death.

Zfp703 is directly regulated by RA

While Zfp703 is the top gene induced at 8 hours following RA treatment, ChIP-seq of RAR did not reveal binding sites within 20Kbp region of its start site (Table 2.1). Therefore it remained unclear whether Zfp703 is actually a direct target of RA signaling. Based on the fast induction of Zfp703, I speculated whether Zfp703 might be regulated by distal chromatin interactions beyond the 20 Kbp window we used in the initial analysis (Mahony et al., 2011). To test this hypothesis, I took advantage of recently obtained data in the lab that mapped long-distance chromatin interactions by paired-end tag sequencing (ChIA-PET) using RNA polymerase II (PolII) as an anchor to identify long-range interactions between transcription start sites (TSS) and distal elements (Figure 2.6-A & B). Consistent with the low level expression of Zfp703 on Day0, I detected few interactions between Zfp703 TSS and distal genomic regions. In contrast, analysis on Day 3 of differentiation revealed strong interactions between Zfp703 TSS and two sites located ~140Kbp upstream (Figure 2.6-C). Importantly, when we aligned ChIP-seq data mapping binding sites of RARs, these two distal sites were enriched for RAR binding 8 hours after RA treatment. Together, these data suggested that RARs are recruited to two distal elements following RA treatment that loop and engage Zfp703 TSS to regulate its expression.

To verify whether these 2 elements function as distal enhancers, I aligned the ChIA-PET results on Day3 with ChIP-seq data mapping enhancer signatures in neural progenitors (Figure 2.7-A). Interestingly, both elements are associated with the active enhancer marker H3K27ac, the co-factor p300, and the pan-enhancer marker H3K4me1. This information suggests that these 2 elements might be *bona fide* enhancers. Next, I performed conservation analysis for these elements and found significant enrichment of conserved sequences in the first RAR binding site among mammals in comparison to adjacent sites (Figure 2.8-A). In the second RAR site I also

identified multiple highly conserved regions (Figure 2.8-B). Therefore these two sites seemed to be conserved in mammals. Based on the facts that (1) these two sites interact with Afp703 TSS during active transcription; (2) they are bound by RAR following RA treatment; (3) they are colocalized with active enhancer signatures; and (4) they are evolutionarily conserved in mammals, Zfp703 is likely a direct target of RA regulated by these two RAR-bound distal enhancers.


A
Ligation
PET & NGS
Mapping
Ab
Cross-linking
Sonication,
antibody
pulldown
FDR<0.05
B
a
b
c
d
G = (V, E)

C

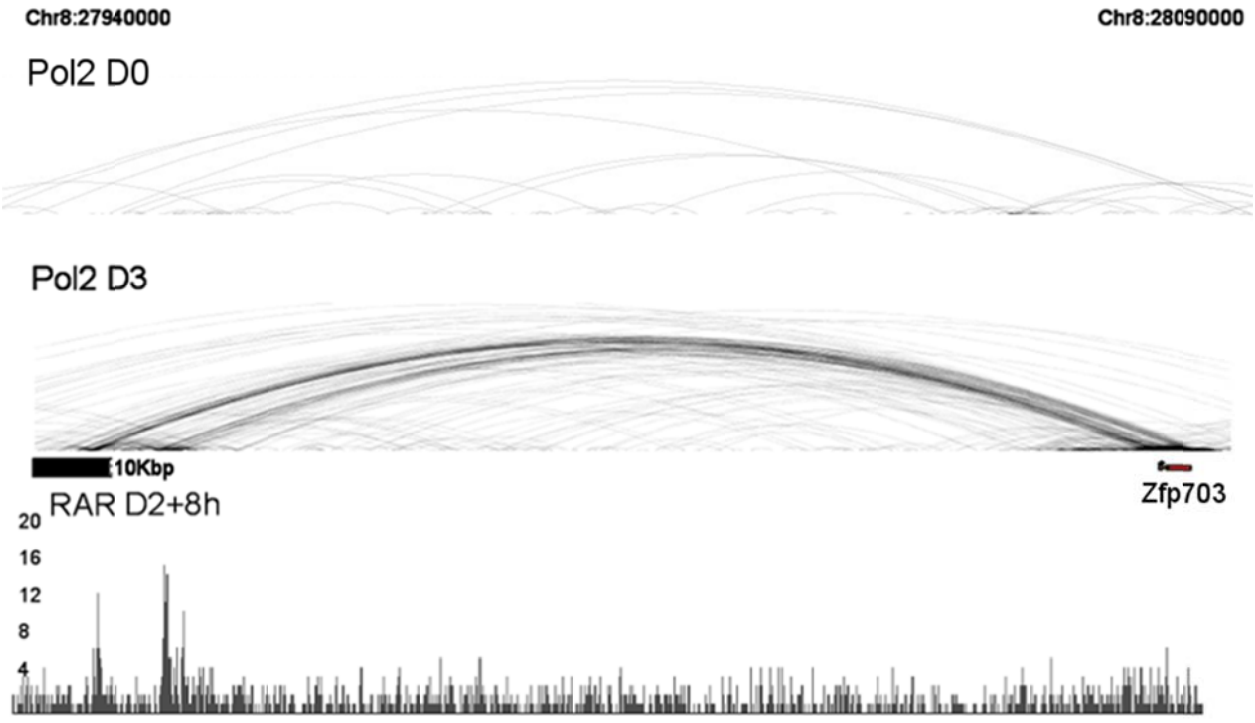

Chr8:27940000
Chr8:28090000
Pol2 D0
Pol2 D3
10Kbp
Zfp703
RAR D2+8h
20
16
12
8
4

Figure 2.6

Figure 2.6 RA might activate Zfp703 through long-range chromatin interactions

A. and B. Schematic of ChIA-PET. Formaldehyde crosslinked chromatin was sonicated and chromatin complexes bound with RNA polymerase II (green) were pulled down using PolII antibody (blue). Specific linkers were added to the open ends and subsequently ligated in the diluted conditions. After the chromatin complexes were reverse-crosslinked, DNA material was subjected to PET extraction and next-generation sequencing statistically significant interactions were called at FDR ≤ 0.05 using the ChIA-PET tool (Li et al., 2010) (taken from (Sandhu et al., 2012)). C. Alignment between PolII ChIA-PET and RAR ChIP-seq around Zfp703 locus during differentiation. While few interactions showed up in this region on Day0, the transcription start site (TSS) of Zfp703 exhibited strong interactions with 2 distal elements (approximately 140Kb upstream) on Day3 after 24hours of RA treatment. Interestingly, these elements were enriched for RAR binding after 8 hours of RA treatment on Day2, suggesting that RA might activate Zfp703 expression through long-range chromatin interactions. (ChIA-PET and analysis done by E. Mazzoni in collaboration with the C. L. Wei group)

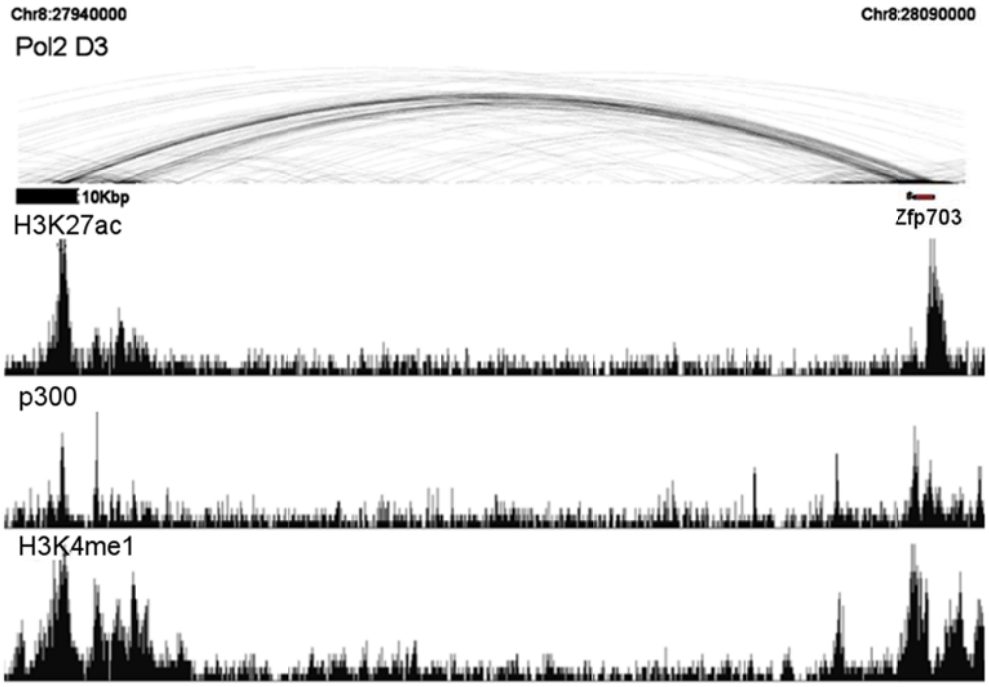
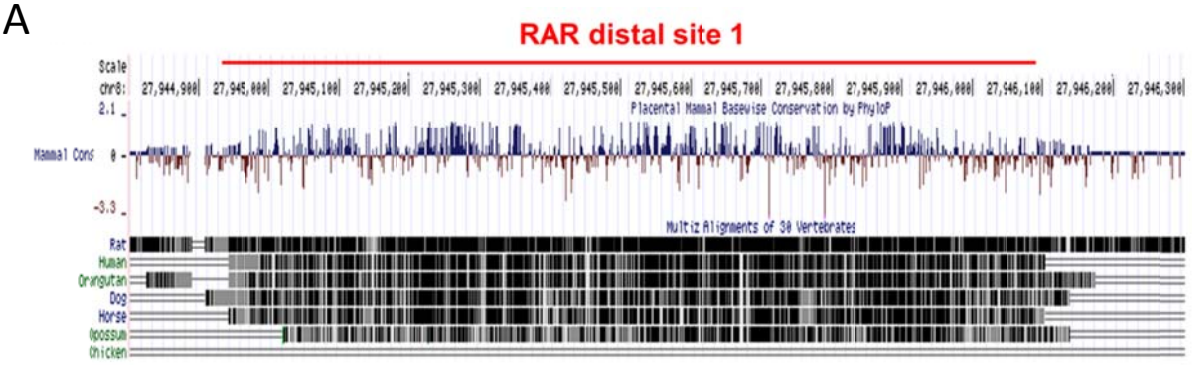

Chr8:27940000
Chr8:28090000
Pol2 D3
10Kbp
H3K27ac
Zfp703
p300
H3K4me1

Figure 2.7

Figure 2.7 Distal RAR binding sites colocalize with enhancer signatures

Alignment of PolII ChIA-PET on Day3 EBs treated with RA with ChIP-seq data for enhancer signatures H3K27ac, p300, and H3K4me1 in neural progenitors around Zfp703 locus. The colocalization of all three enhancer signatures with the two distal elements interacting with Zfp703 and RAR (Figure 2.6-C) suggests that these two elements function as enhancers through which RA regulates the expression of Zfp703. (ChIA-PET and analysis done by E. Mazzoni in collaboration with the C. L. Wei group; ChIP-seq and analysis of enhancer signatures in neural progenitors done by M. Closser in collaboration with the D. Gifford group)

A


RAR distal site 1
Scale
chr8:
27,944,900
27,945,000
27,945,100
27,945,200
27,945,300
27,945,400
27,945,500
27,945,600
27,945,700
27,945,800
27,945,900
27,946,000
27,946,100
27,946,200
27,946,300
2.1
Mammal Cons
0
-3.3
Placental Mammal Basewise Conservation by PhyloP
Multiz Alignments of 30 Vertebrates
Rat
Human
Orangutan
Dog
Horse
Opossum
Chicken

B

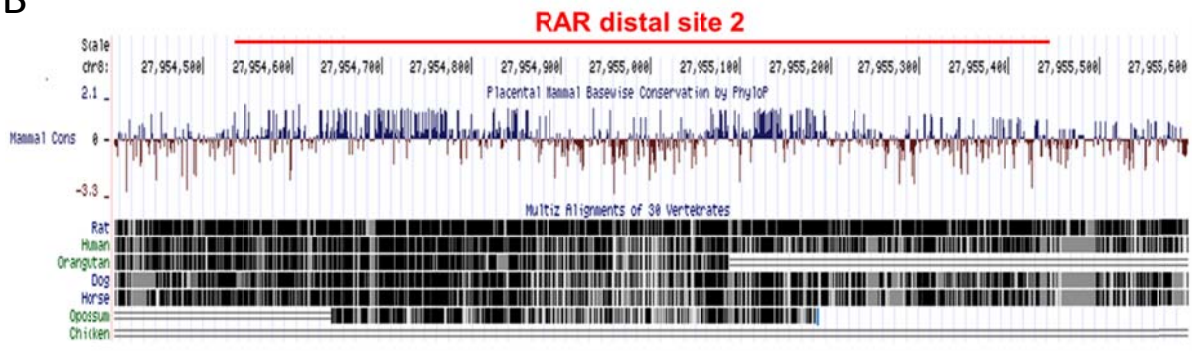

RAR distal site 2
Scale
chr8:
27,954,500
27,954,600
27,954,700
27,954,800
27,954,900
27,955,000
27,955,100
27,955,200
27,955,300
27,955,400
27,955,500
27,955,600
2.1
Mammal Cons
0
-3.3
Placental Mammal Basewise Conservation by PhyloP
Multiz Alignments of 30 Vertebrates
Rat
Human
Orangutan
Dog
Horse
Opossum
Chicken

Figure 2.8

Figure 2.8 Distal RAR binding sites are conserved in mammals

Multiple alignment and conservation analysis around 1Kbp region flanking the two distal RAR binding sites upstream of Zfp703 TSS. Shown in the middle of each panel are scores of conservation among mammals using the PhastCons package (Siepel et al., 2005). Pair-wise alignment between mouse and 7 organisms can be found at the bottom of each panel. The analysis was performed using UCSC genome browser. Both sites are conserved across mammals. While there is significant enrichment of conserved sequences proximal to RAR site1 (A) compared to the adjacent regions; patches of highly conserved sections can be observed in the vicinity of RAR site 2 (i.e., Chr8:27954650-27954750 and Chr8: 27955050-27955200 in B).

Discussion

In this chapter, I established that RA treatment in 2 day old EBs represses Oct4 and promotes the exit of pluripotency. Importantly, I determined that Oct4 (as well as Sox1) is likely not a direct target of RA signaling. I examined expression data and identified 6 candidate transcription factors that might mediate effects of RA. Finally, I generated inducible ESC lines to test the function of these factors and discovered that both Zfp703 and Zfp503 are sufficient to effectively repress Oct4 in the absence of RA. Therefore, I hypothesize that Zfp703 and Zfp503 are novel downstream factors of RA signaling that contribute to the termination of the ESC transcription network in neural lineage commitment.

RA indirectly regulates the ESC and neural transcriptional network

In our system, as previously reported, Oct4 is not a direct target of RA signaling. Its mRNA expression changed minimally in the first 8 hour of RA treatment but dramatically decreased during the following 16 hours. The facts that RA-bound RAR/RXRs act as transcriptional activators and that no RAR binding sites were identified in the proximity of Oct4 promoter further support this notion (Mahony et al., 2011). So RA must induce some other transcription factors to repress Oct4 and terminate the ESC state. Interestingly, although Sox1 is induced by RA, and theoretically this could be achieved by the binding of RA-bound RAR/RXR to its promoter, the delayed induction of Sox1 is inconsistent with direct regulation by RA. Similar to Oct4, its mRNA expression level did not change much during the first 8 hours of RA treatment but was significantly upregulated between 8-24 hours after RA addition (Figure 2.1-C). There is no RAR binding site around Sox1 promoter either (Mahony et al., 2011). Taken together, these results suggest that RA regulates both ESC and neural transcriptional networks indirectly

through induction of additional transcription factors that can repress Oct4 and promote Sox1 expression during neuroectodermal lineage commitment.

There can be two reasons why RA signaling regulates the ESC and neural transcriptional programs indirectly through an additional layer of transcription factors rather than directly repressing Oct4 and inducing Sox1 by RAR binding. First, because the factors mediating neuroectodermal lineage commitment downstream of RA might also be regulated by other signaling pathways, they can serve as hubs for the integration of various signal inputs to ensure coordinated expression of genes involved in the commitment of ESCs towards neural fate. Second, since one transcription factor can regulate multiple genes, this mechanism provides a convenient way for RA signaling to simultaneously control large number of genes without RAREs. It could also enable RA to effectively modulate different transcription networks by directly regulating only a small number of direct targets.

Additional transcription factors are involved in Oct4 repression by RA

Although multiple transcription factors, including COUP-TFs, Hoxa1, and GCNF have been shown to participate in RA-mediated Oct4 repression (Gu et al., 2005; Martinez-Ceballos et al., 2005; Schoorlemmer et al., 1994), only Hoxa1 appears to be a direct target of RA signaling. However, expression of Hoxa1 in inducible ESCs only led to moderate Oct4 repression, significantly weaker that the effects of RA (Figure 2.3-D), indicating that additional transcription factors downstream of RA might be involved in Oct4 repression.

To identify the putative Oct4 repressors, I analyzed expression data and found 6 candidate transcription factors that were significantly induced within 8 hours of RA treatment. In the screen using the inducible ESC lines established with these factors, I discovered that both

Zfp703 and Zfp503 can significantly repress Oct4 on mRNA and protein level when expressed without RA, suggesting that they might also be downstream effectors of RA-mediated exit of pluripotency. Zfp703 and Zfp503 belong to the NET family transcription factors whose members have been implicated in various events during embryonic development and cancer (Pereira-Castro et al., 2013). While Zfp703 and Zfp503 are required for proper patterning of hindbrain in zebrafish, little is known about their role during early neurogenesis. Importantly, Zfp703 and Zfp503 have been shown to interact with corepressors Groucho and HDACs and act as transcriptional repressor (Runko and Sagerstrom, 2003, 2004). Therefore, the findings in this chapter indicate that Zfp703 and Zfp503 might be induced by RA signaling and initiate the repression of Oct4 and the exit of pluripotency during the first stage of RA-induced neuroectodermal lineage commitment (Figure 2.9).

Transcriptional network terminating the ESC state

In my screen of 6 early-response transcription factors downstream of RA (Figure 2.3-D), expression of Zfp703 and Zfp503 resulted in significant down-regulation of Oct4 mRNA in the absence of RA. Whereas Hoxa1 induction only led to moderate Oct4 repression during the same period, despite that it has been reported to be required for efficient RA-mediated down-regulation of Oct4 and neural differentiation (Martinez-Ceballos et al., 2005; Martinez-Ceballos and Gudas, 2008). Furthermore, COUP-TFs have been shown to bind to the HREs in Oct4 promoter and repress transcription upon binding (Schoorlemmer et al., 1994). And Nr2f2 is required for neuronal specification of human ESCs through directly repressing Oct4 and turning on genes driving neuronal fate (Rosa and Brivanlou, 2011). However, they are upregulated after the initiation of Oct4 repression, similar to Sox1 in our system (Figure 2.2). Finally, despite that GCNF has been demonstrated as necessary for RA-induced exit of pluripotency and neural

differentiation (Akamatsu et al., 2009; Gu et al., 2005), its mRNA is expressed at high levels already in ESCs and is only mildly induced following RA treatment (Figure 2.2). These observations suggest that RA does not regulate Oct4 and thus the ESC state through only one downstream transcription factor. Rather, it might rely on a group of factors that function redundantly and interdependently to achieve fast and precise repression of the genes maintaining the ESC identity. Consistent with this notion, Oct4 is still repressed by RA in ESCs lacking either Hoxa1 or GCNF (Gu et al., 2005; Martinez-Ceballos et al., 2005). Based on this information, I hypothesize that Zfp703 and Zfp503 are the key initial repressors of Oct4 assisted by Hoxa1 and Cdx1. Secondary repressors induced 24 hours after RA treatment (COUP-TFs) likely consolidate the repression of Oct4 and the termination of pluripotency. To test this hypothesis and to map the transcriptional cascade downstream of Zfp703, I set out to characterize the effects of Zfp703 overexpression at genome-wide level.

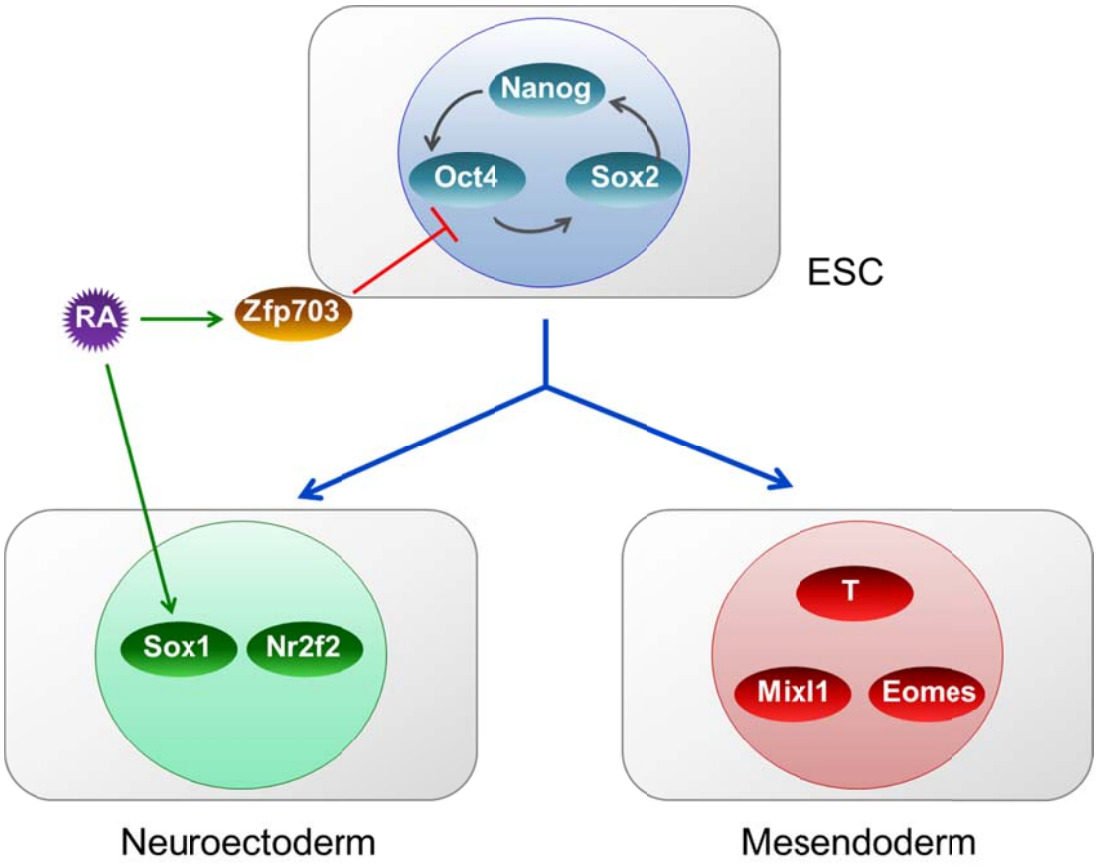

Nanog
Oct4
Sox2
ESC
RA
Zfp703
Sox1
Nr2f2
T
Mixl1
Eomes
Neuroectoderm
Mesendoderm

Figure 2.9

Figure 2.9 RA might initiate neuroectodermal lineage commitment by inducing Zfp703

RA promotes neuroectodermal differentiation of ESCs by indirectly repressing pluripotency gene such as Oct4 and upregulating key neuroectodermal regulators like Sox1. Zfp703, a transcription factor and a putative direct target of RA, is likely the major effector mediating the onset of pluripotency termination downstream of RA.

Chapter 3. Global effects of Zfp703 on gene expression profile and neurogenesis

Introduction

Members of the NET family proteins have been shown to regulate gene expression during embryonic development in various organisms including Drosophila, zebrafish, and mice (Pereira-Castro et al., 2013) as well as in cancer progression (Holland et al., 2011; Slorach et al., 2011). Hence, NET family proteins seem to regulate different sets of target genes in a context-dependent manner. I identified Zfp703 as an RA-induced transcriptional regulator that effectively represses Oct4 expression. However, RA signaling regulates additional genes and pathways during ESC neuroectodermal induction – namely it effectively induces expression of neuroectodermal genes and represses mesendodermal lineage. In the next two chapters I will examine whether Zfp703 recapitulates any of these additional activities attributed to the fate choice of neuroectoderm over mesendoderm by RA signaling. I will test whether Zfp703: (1) regulates expression of other genes implicated in Oct4 repression (COUP-TFs, GCNF, etc.); (2) represses additional pluripotency genes besides Oct4; (3) activates expression of neuroectodermal specific genes; and (4) represses mesendodermal lineage-specific genes. To address these questions I performed genome-wide expression profiling and gene ontology analysis in control and Zfp703-expressing cells, confirmed key downstream targets of Zfp703, identified Zfp703 binding sites by ChIP-seq analysis, and derived Zfp703 conditional knockout cells to determine requirements for Zfp703 in RA-mediated neuroectodermal differentiation of ESCs.

Results

Expression profiling of EBs expressing Zfp703

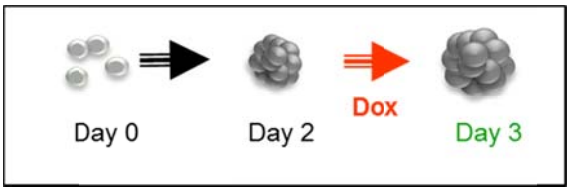
To characterize the effects of Zfp703 expression on the global pattern of gene expression, I differentiated the inducible Zfp703 ESC line as EBs for 2 days and treated them with or without doxycycline for 24 hours. On Day3 I collected EBs, prepared cDNA libraries, and analyzed gene expression by GeneChip 430A 2.0 Array from Affymetrix which comprises more than 39,000 probesets representing approximately 14,000 genes (Figure 3.1-A). The result from 4 independent replicas indicated that compared to control, expression of Zfp703 in EBs led to > 2-fold downregulation of 109 genes and upregulation of 173 genes with p value less than 0.05 (Figure 3.1-B). The 50 most upregulated or downregulated genes are listed in Table 3.1.

Gene ontology (GO) analysis of putative Zfp703 target genes

To get a general idea of the functions involving these 282 Zfp703-regulated genes, I conducted GO analysis using Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 from National Institute of Allergy and Infectious Diseases (NIAID) (Table 3.2-A) and found that the top-ranked GO categories according to p values are developmental process (1.20E-18), multicellular organismal process (1.50E-08), biological regulation (2.40E-05), and cellular process (1.40E-04). Because developmental process ranked as the first GO category according to the p value, and I am most interested in understanding the role of Zfp703 during early development, I performed a second round of GO analysis on the 99 genes in the category of developmental process (Table 3.2-B). Interestingly, the top-ranked categories according to p values within developmental process comprised mesoderm development (3.00E-06), ectoderm development (1.00E-05), mRNA transcription regulation (1.50E-05), cell communication (6.50E-05), and neurogenesis (1.10E-04). Therefore the GO analysis of putative Zfp703 target

genes suggested that in addition to Oct4 repression, Zfp703 might participate in other relevant early embryonic developmental processes.

A



B

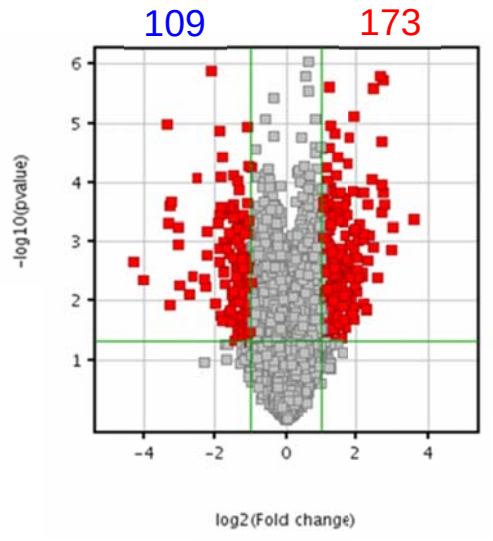


Figure 3.1

Figure 3.1 Gene expression profiling of Zfp703-expressing EBs

Inducible Zfp703 ESCs were differentiated as EBs for 2 days before treated with or without doxycycline. Then the samples were collected on Day3, and total cDNA was prepared and subjected to microarray analyses. A. Schematic of ESC differentiation. B. Volcano plot showing genes whose expression level significantly changed in doxycycline-treated EBs compared to untreated ones. Zfp703 expression led to downregulation of 109 genes and upregulation of 173 genes (4 independent experiments with cutoffs as fold change: 2 and p-value: 0.05).

Table 3.1 Top 50 genes affected by Zfp703 expression ranked by fold-change

Rank	Gene	Up-fold	Rank	Gene	Down-fold
1	Acta2	7.825689	1	Nkx1-2	17.71596
2	Csn3	7.50006	2	T	16.30743
3	Igfbp5	7.19221	3	Mixl1	14.75618
4	Synpo2l	6.947807	4	Hoxb1	13.35986
5	Col1a2	6.85199	5	Wnt8a	10.89776
6	Igfbp5	6.77359	6	Eomes	10.76884
7	Col1a2	6.622636	7	Nkx1-2	10.18989
8	Ets1	6.462058	8	Eomes	9.931783
9	Nr2f2	6.4234	9	Fst	9.494982
10	Tal2	5.9036	10	Apln	9.210899
11	Thbs1	5.897475	11	Fgf8	8.567696
12	Nrk	5.766512	12	Fst	8.496554
13	Sox1	5.763515	13	Aplnr	7.787348
14	Cryab	5.309863	14	Cdx1	6.137744
15	Gpm6a	5.259384	15	Wnt3	5.995048
16	Tgfbi	4.852696	16	Rbp1	5.161519
17	Anxa1	4.759348	17	Hoxa1	4.971188
18	Gpm6a	4.684581	18	Cnpy1	4.952869
19	Ets1	4.670638	19	Defa-rs2	4.438122
20	Phlda1	4.463592	20	Cnpy1	4.330779
21	Fosl2	4.456245	21	Dkk1	4.321925
22	Cadm3	4.428384	22	Nefl	4.311172
23	Tgfbi	4.383843	23	Cdx2	4.166731
24	Igf2	4.370416	24	Cxx1c	4.158531
25	Cotl1	4.272188	25	Pim2	4.121471
26	Tgfbi	4.207836	26	Cabp7	4.033081
27	Synpo2l	4.02347	27	Ntn1	3.823372
28	Gab2	3.977862	28	Ror1	3.803289
29	Hdac9	3.949652	29	Alox15	3.787318
30	Clic5	3.831111	30	Cnpy1	3.77149
31	Hdac9	3.813861	31	Gbx2	3.746454
32	Armcx3	3.789828	32	Tdgf1	3.733802
33	Krt18	3.788724	33	AU020094	3.727307
34	Gadd45b	3.686982	34	Enpp2	3.628458
35	Prnd	3.634711	35	Lhpp	3.622605
36	Cyr61	3.572909	36	Nkx1-2	3.541297
37	Thbs1	3.564796	37	Sp5	3.528271
38	Pax3	3.472756	38	Prr18	3.523384
39	Lrp2	3.46073	39	Wnt5b	3.447873
40	Camk2n1	3.432875	40	Nefl	3.445313
41	Gbp2	3.431339	41	Btla	3.437892
42	Synpo2l	3.425609	42	Evx1	3.39707
43	Nrk	3.421276	43	Cer1	3.263713
44	Hrc	3.413748	44	Hoxb2	3.245185
45	Acta1	3.411877	45	Zfp428	3.239637
46	Dsp	3.381145	46	Prickle1	3.168788
47	Gbp1	3.342031	47	Lef1	3.152981
48	Ppm1h	3.327644	48	Nme5	3.140101
49	Krt8	3.308301	49	1700097N02Rik	3.083912
50	Zfp516	3.267424	50	Fgf5	3.048187

Table 3.2 Gene ontology (GO) analyses of putative Zfp703 target genes

A. 282 putative Zfp703 target genes were subjected to GO analysis, and top-ranked functional categories were shown with p values. B. A second round of GO analysis for the 99 genes belonging to "Developmental process" from the first round GO analysis.

A

GO Term	Count	%	P-Value
Developmental process	99	34.6	1.20E-18
Multicellular organismal process	106	37.1	1.50E-08
Biological regulation	136	47.6	2.40E-05
Locomotion	17	5.9	1.40E-04
Cellular component organization	47	16.4	7.50E-04
Death	18	6.3	2.50E-03
Cellular process	156	54.5	2.30E-02
Biological adhesion	15	5.2	4.90E-02
Growth	7	2.4	7.40E-02
Immune system process	18	6.3	9.60E-02

B

GO Term	Count	%	P-Value
Mesoderm development	24	8.4	3.00E-06
Ectoderm development	26	9.1	1.00E-05
mRNA transcription regulation	41	14.3	1.50E-05
Cell communication	34	11.9	6.50E-05
Neurogenesis	22	7.7	1.10E-04
Cell structure and motility	30	10.5	3.40E-04

Lineage specific markers regulated by Zfp703

Gene-by-gene examination of the most important lineage-specific markers provides important insight into the programs regulated by Zfp703. First I extended analysis of pluripotency markers upon Zfp703 induction. In addition to Oct4, I detected downregulation of *Fgf4* (2.5 fold) and *Lef1* (3.1 fold) (Figure 3.2 and Table 3.1), two genes involved in the maintenance of the ESC state, and 6.4-fold upregulation of *Nr2f2* (*COUP-TF2*), a gene implicated in *Oct4* repression. Interestingly, Zfp703 had only negligible effect on *Nanog* expression (1.1-fold downregulation as compared to nearly 20-fold downregulation by RA treatment), indicating that Zfp703 targets only a subset of pluripotency genes, and that successful repression of the entire pluripotency program might depend on additional downstream regulators such as Nr2f2.

Next I examined effects of Zfp703 on markers of neuroectodermal lineage. Zfp703 induced early markers of neural differentiation *Sox1* (5.6 fold) and *Zeb2* (3.1 fold) (Figure 3.2 and Table 3.1). However, many additional neural lineage genes (*Pax6*, *Neurog2*, *Ascl1*, or *Sox21*) were not significantly induced 24 hours following Dox treatment, in contrast to their induction following RA application. Meanwhile, Zfp703 increased expression of several epidermal genes, including *Col1a2* (6.8 fold), *Krt18* (3.8 fold), *Krt8* (3.3 fold), and *Col1a1* (2.9 fold), raising the possibility that Zfp703 might be initially biasing neuroectodermal differentiation of ESCs towards epidermal phenotype.

Finally, I examined the expression of mesendodermal markers to determine whether Zfp703 promotes differentiation of ESCs along all major germ layers. Expression profiling revealed that mesodermal (*Mixl1, Lhx1, Gata6, Sox17,* and *Gsc*) and endodermal (*Wnt3, Eomes,* and *Lef1*) as well as mesendoderm lineage marker *Brachyury* (*T*) were among the most strongly

downregulated genes upon Zfp703 expression (Figure 3.2 and Table 3.1). Together, these data suggest that Zfp703 might be involved in key decision steps in RA-mediated differentiation of ESCs, acting as a suppressor of pluripotency as well as mesendodermal lineage and an activator of neuroectodermal fate.

up

Epidermal	Fold
Col1a2	6.67
Krt18	4.00
Krt8	4.13
Col1a1	2.99

Neurogenesis	Fold
Sox1	2.25
Nr2f2	6.46
Zeb2	2.37

down

Endoderm	Fold
Mixl1	0.05
Lhx1	0.24
Gata6	0.44
Sox17	0.35
Gsc	0.47

Mesoderm	Fold
Wnt3	0.19
Eomes	0.10
Lef1	0.15

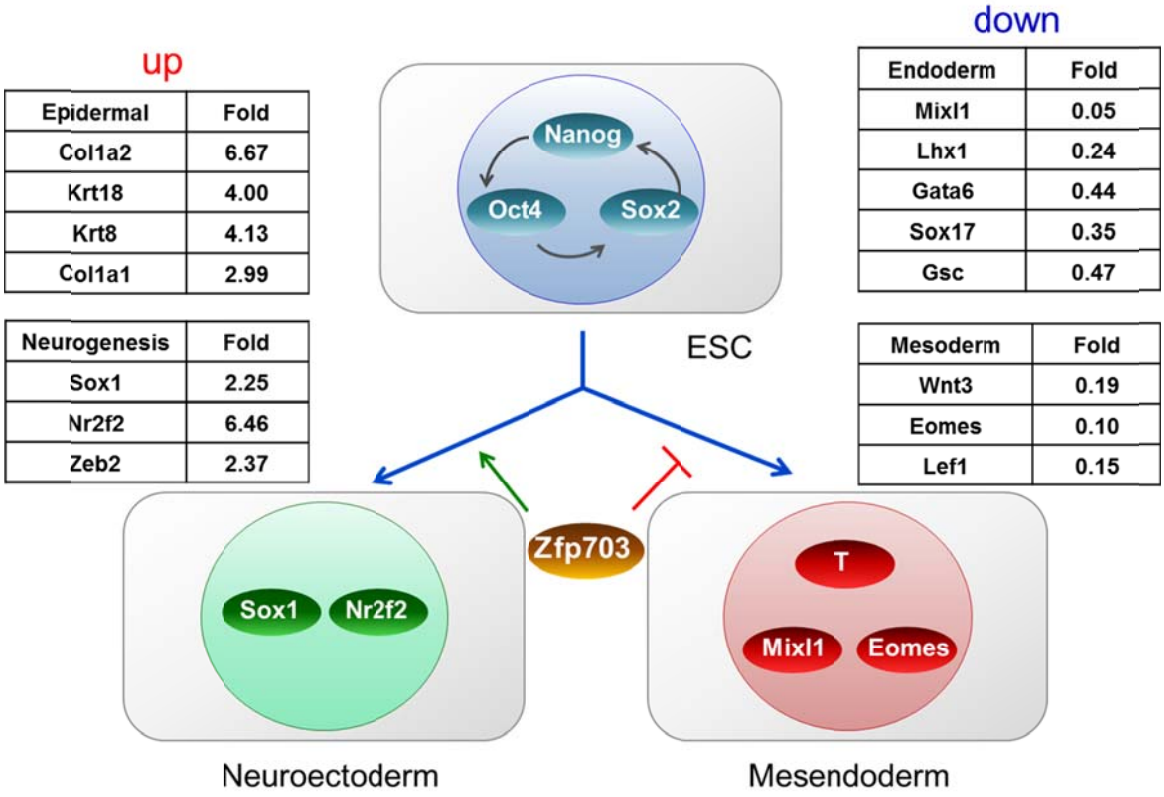


Figure 3.2

Figure 3.2 qRT-PCR verification of putative Zfp703 target genes

Inducible Zfp703 ESCs were differentiated as EBs for 2 days before treated with or without doxycycline. 24 hours later the samples were collected and qRT-PCR was performed. Relative expression levels in Zfp703-expressing (doxycycline-treated) EBs were normalized to untreated ones. Consistent with the microarray data, Zfp703 expression resulted in downregulation of genes implicated in maintaining pluripotency (*Lef1* and *Oct4*) as well as differentiation of mesoderm (*Mixl1, Lhx1, Gata6, Sox17,* and *Gsc*) and endoderm (*Wnt3, Eomes,* and *Lef1*). Meanwhile, Zfp703 expression also led to upregulation in epidermis markers (*Col1a2, Krt18, Krt8,* and *Col1a1*) as well as genes involved in neurogenesis (*Sox1, Nr2f2,* and *Zeb1*) (3 independent experiments; mRNA relative expression levels were normalized to non-treated control EBs and shown as average).

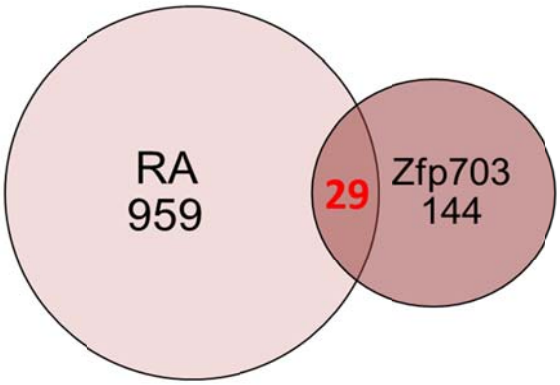
Zfp703 regulates a subset of important genes involved in RA-mediated neuroectodermal lineage commitment

Zfp703 is putatively a direct target of RA. And expression profiling as well as qPCR analyses suggest that Zfp703 is likely implicated in all major events of RA-mediated neuroectodermal lineage commitment, raising the question to which extent can Zfp703 recapitulate the effects of RA on gene expression. To address this issue, I compared the genes whose expression level is significantly changed (fold change ≥ 2 and p-value ≤ 0.05) upon RA treatment and Zfp703 expression through analyzing the microarray profiling data on Day3 EBs that has been treated with RA or induced to express Zfp703 from Day2. As shown in Figure 3.3-A, RA treatment resulted in 988 genes significantly upregulated on Day3; while Zfp703 expression led to 173 upregulated genes at the same point. 29 genes are upregulated by both RA and Zfp703 (16% of total genes upregulated by Zfp703 expression). On the other hand, RA treatment caused 963 genes significantly downregulated on Day3; whereas Zfp703 expression resulted in 109 downregulated genes during that period. 47 genes are downregulated by both RA and Zfp703 (43% of total genes downregulated by Zfp703 expression). Therefore, Zfp703 seems to recapitulate the effect of RA on only a small subset of RA-regulated genes. Because 43% of genes downregulated by Zfp703 overlap with genes downregulated by RA, the major function of Zfp703 might be a transcription repressor downstream of RA.

The genes regulated by both RA and Zfp703 are listed in Table 3.3. Among them are key transcription factors in ESC lineage commitment. Critical neuroectodermal fate determinant *Sox1* and *Nr2f2* are upregulated by both RA and Zfp703; while genes specifying mesendoderm such as *Brachyury (T)*, *Eomes*, *Wnt3*, and *Lef1* are repressed by both RA and Zfp703. Interestingly, epiblast marker Fgf5 is also downregulated, suggesting that both RA and Zfp703

promote the progression of ESCs through the epiblast stage towards lineage commitment. In summary, although Zfp703 causes similar changes in expression of only a small number of RA-regulated genes (Figure 3.3), it is sufficient to recapitulate the effect of RA on key regulators of ESC lineage commitment, specifically genes involved in the fate choice between neuroectoderm and mesendoderm. These data further support the notion that Zfp703 might be a suppressor of pluripotency and mesendodermal lineage and an activator of neuroectodermal fate downstream of RA.

A



B

Figure 3.3

Figure 3.3 Overlap between genes regulated by Zfp703 and RA on Day3

Microarray profiling data from Zfp703-expressing and RA-treated Day3 EBs were compared, and Van diagrams for genes upregulated (A) as well as downregulated (B) under each condition are shown (Cutoff: fold change ≥ 2 and p-value ≤ 0.05). A. RA treatment led to significant upregulation of 988 genes on Day3; whereas Zfp703 expression resulted in significant upregulation of 173 genes. 29 genes are significantly upregulated in both RA treatment and Zfp703 expression. B. RA treatment caused significant downregulation of 963 genes; while Zfp703 expression led to significant downregulation of 109 genes. There are 47 genes significantly downregulated in both RA treatment and Zfp703 expression (Microarray profiling and analysis on RA-treated Day3 EBs done by E. Mazzoni and S. Mahony (Mahony et al., 2011)).

Table 3.3 Genes regulated by both RA and Zfp703 expression

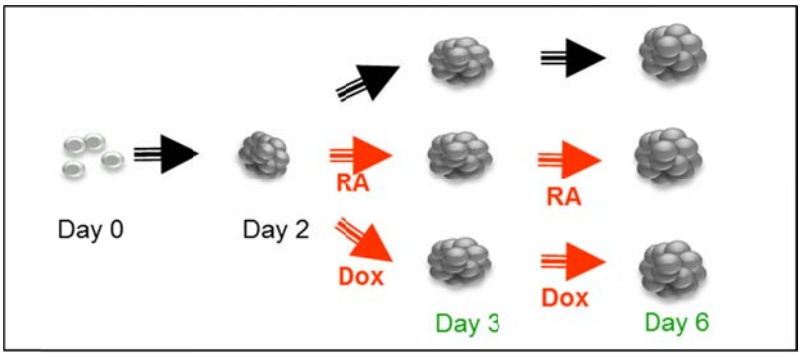
Microarray profiling data of RA-treated and Zfp703-expressing Day3 EBs were compared and genes whose expression is affected by both RA and Zfp703 are listed. (Cutoff: fold change ≥ 2 and p-value ≤ 0.05)

Up	Down	Down
Tgfbi	Sgk1	Prickle1
Nr2f2	Pla2g1b	1700097N02Rik
Fzd2	Pim2	Eras
Fbn2	Slc7a7	Ror1
Gap43	Tnfsf11	Cnpy1
Dsc2	T	Fgf5
Tgm2	Fst	Grik3
Gbp7	Spry4	Vrtn
Dsp	Nkx1-2	Slc7a7
Cotl1	Foxd3	Lef1
Tgm2	Wnt8a	AU020094
Rhou	Epha1	Cnpy1
Col1a2	Etv4	Btla
Crabp2	Timp4	Cnpy1
Csf1	Zic3	
Gab2	Eomes	
8430427H17Rik	Nefl	
Sema3c	Celf4	
Fndc3b	Ggct	
Frk	1700097N02Rik	
Trp53i11	Spry2	
Zfp516	Pla2g1b	
Sox1	Enpp2	
Cpm	Fbp1	
Cxxc4	Pllp	
	Wnt3	
	Rerg	
	L1td1	
	Tdgf1	

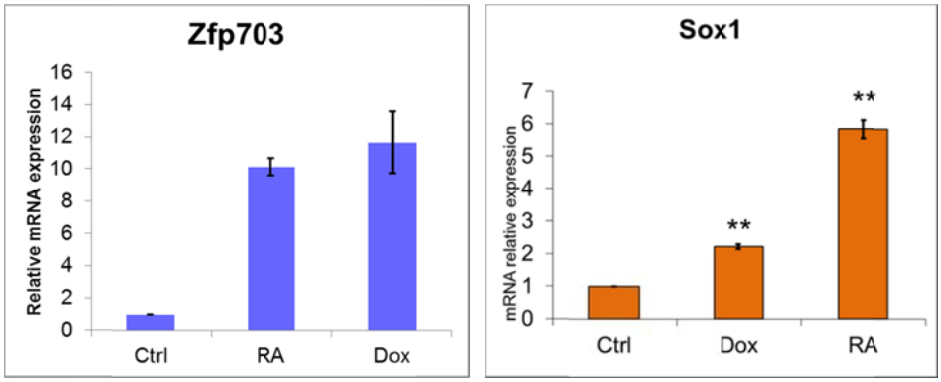
Zfp703 induces neural progenitor marker Sox1

While early neural lineage marker is among the top Zfp703-induced genes, presence of multiple markers of epidermal lineage prompted me to more carefully examine whether Zfp703 leads primarily to the specification of epidermis or neural tissue (Figure 3.2). To confirm the gene expression analysis and to examine expression of *Sox1* and epidermal markers with cellular resolution I turned to qPCR and to immunocytochemical analyses of Day3 EBs expressing Zfp703 (Figure 3.4-A). While RA treatment resulted in almost 6-fold increase in *Sox1* mRNA level, Zfp703 induction by doxycycline led to a more moderate but still significant ~2-fold upregulation in *Sox1* mRNA (Figure 3.4-B), compared to non-treated control EBs (p value: 0.02). Importantly, the expression level of Zfp703 is similar in RA-treated and doxycycline treated EBs (Figure 3.4-B), ruling out the possibility that the observed effects of Zfp703 are due to abnormally high expression level resulting from the inducible system. At this point, changes in mRNA level were accompanied with comparable increase in the number of cells expressing Sox1 protein (Figure 3.4-C). Together, these data suggested that Zfp703 expression is sufficient to partially recapitulate the effect of RA in converting pluripotent cells to Sox1-positive neural progenitors.

A



B



C

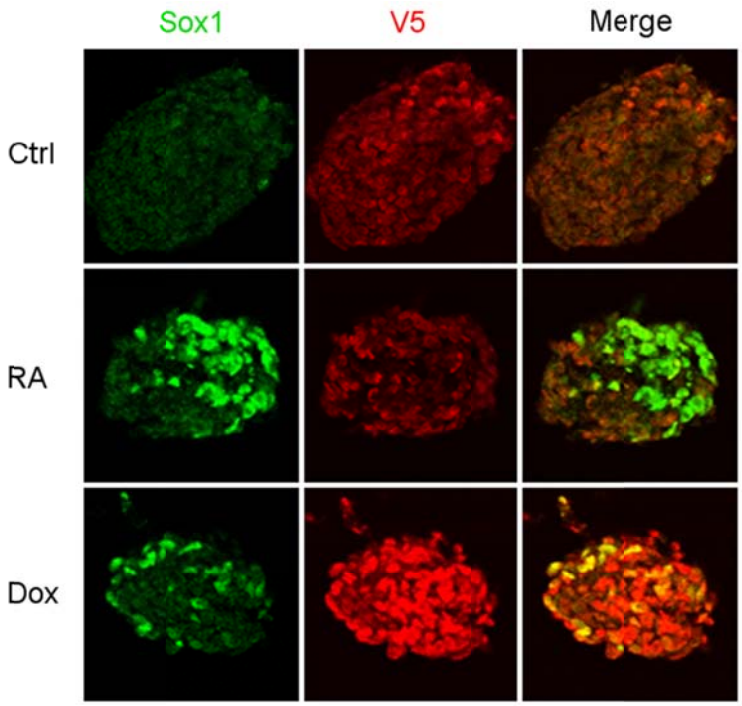


Figure 3.4

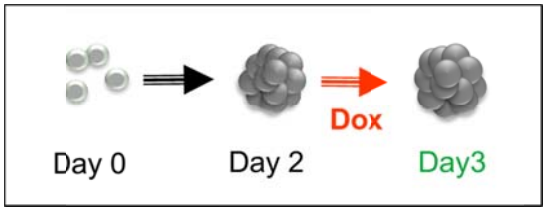
Figure 3.4 Zfp703 promotes the commitment of neural progenitors

Inducible Zfp703 ESCs were differentiated as EBs for 2 days before treated with RA or doxycycline. 24 hours later the EBs were collected for qRT-PCR and immunocytochemistry analyses. A. Schematic of experiments on the effect of Zef703 on neurogenesis. After induction by RA or doxycycline on Day2, EBs were analyzed on Day3 for neural progenitor marker and Day6 for postmitotic neuronal markers. B. The expression of *Zfp703* and *Sox1* in Day3 EBs. While both RA treatment and Zfp703 induction by doxycycline resulted in about 10-fold increase in Zfp703 mRNA level compared to non-treated EBs, RA treatment resulted in 6-fold increase in Sox1 mRNA, and Zfp703 expression (Dox) led to 2-fold upregulation of Sox1 mRNA (3 independent experiments; mRNA relative expression levels were normalized to non-treated control EBs and shown as average ± standard deviation; **$p<0.01$). C. Sections of Day3 EBs stained for V5 and Sox1. Both RA treatment and the expression of V5-tagged Zfp703 (doxycycline treatment) resulted in more Sox1-positive neural progenitors compared to non-treated control EBs.

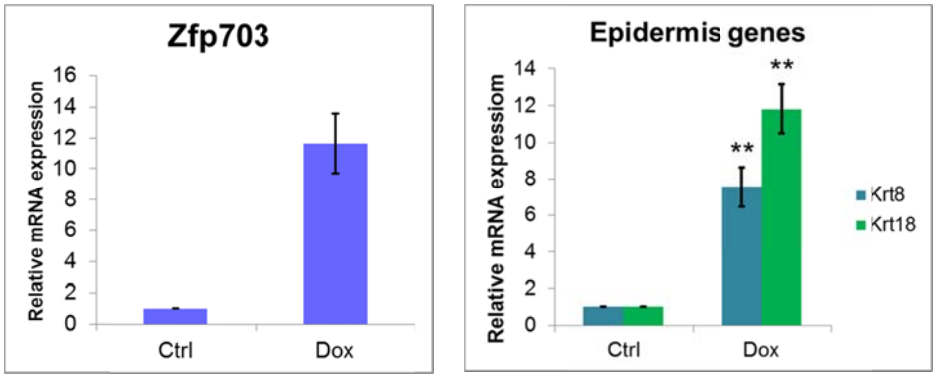
Zfp703 transiently induces epidermal genes

In addition to Sox1, Zfp703 expression also led to upregulation of epidermal genes (Figure 3.2). Indeed, I observed upregulation of keratins Krt8 and Krt18 on Day3 by qPCR (Figure 3.5-B) and immunostaining (Figure 3.5-C). Importantly, keratins and Sox1 were expressed in non-overlapping populations of cells (keratin-positive cells close to the periphery of EBs surrounding more centrally-located Sox1 cells), indicating that the two lineages are effectively resolved in Zfp703-expressing cells (Figure 3.6-A). Furthermore, by Day 4 of differentiation Sox1-positive cells appeared to accumulate relative to keratin-positive cells (Figure 3.6-A) and by Day 6 the EBs are composed primarily of Sox1-positive neural cells (Figure 3.6-B). These data indicate that while initially both epidermal and neural lineages are specified in response to Zfp703 induction, neural cells become the dominant population in Day 4-6 EBs. The delayed emergence of Sox1 cells in Zfp703-expressing EBs compared to EBs treated with RA (Figure 3.4-B) suggests that the neural induction is likely to be an indirect consequence of Zfp703 induction.

A



B



C

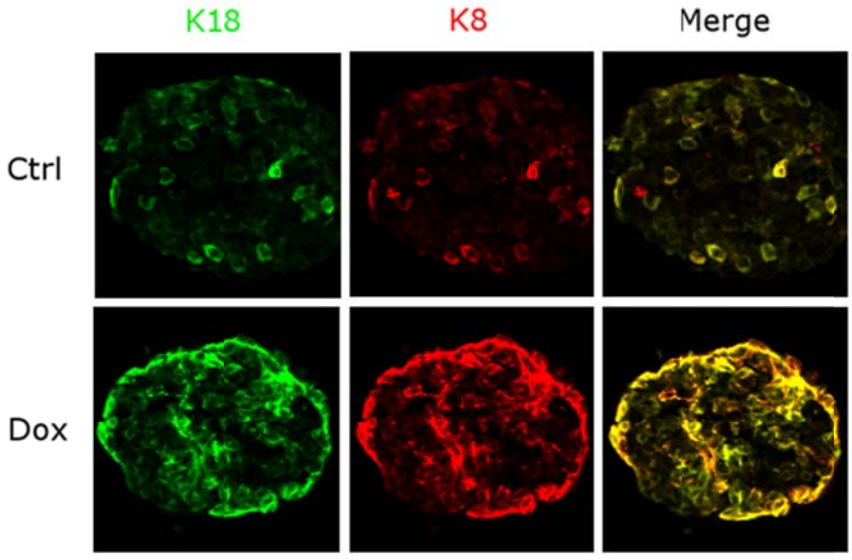


Figure 3.5

Figure 3.5 Zfp703 induces epidermal genes

Inducible Zfp703 ESCs were differentiated as EBs for 2 days before treated with or without doxycycline. On Day3 the EBs were collected for qRT-PCR and immunocytochemistry analyses. A. Schematic of ESC differentiation. B. Zfp703 expression (Dox) resulted in ~7-fold upregulation in Krt8 mRNA and ~11-fold increase in Krt18 mRNA compared to non-treated control (Ctrl) on Day3, consistent with the microarray data (Figure 3.2) (3 independent experiments; mRNA relative expression levels were normalized to non-treated control EBs and shown as average ± standard deviation; **$p<0.01$). C. Sections of Day3 EBs showed that there were more cells positive for Krt8 and Krt18 proteins when treated with doxycycline (Dox) compared with non-treated control (Ctrl), suggesting that Zfp703 expression induced Krt8 and Krt18 proteins on Day3.

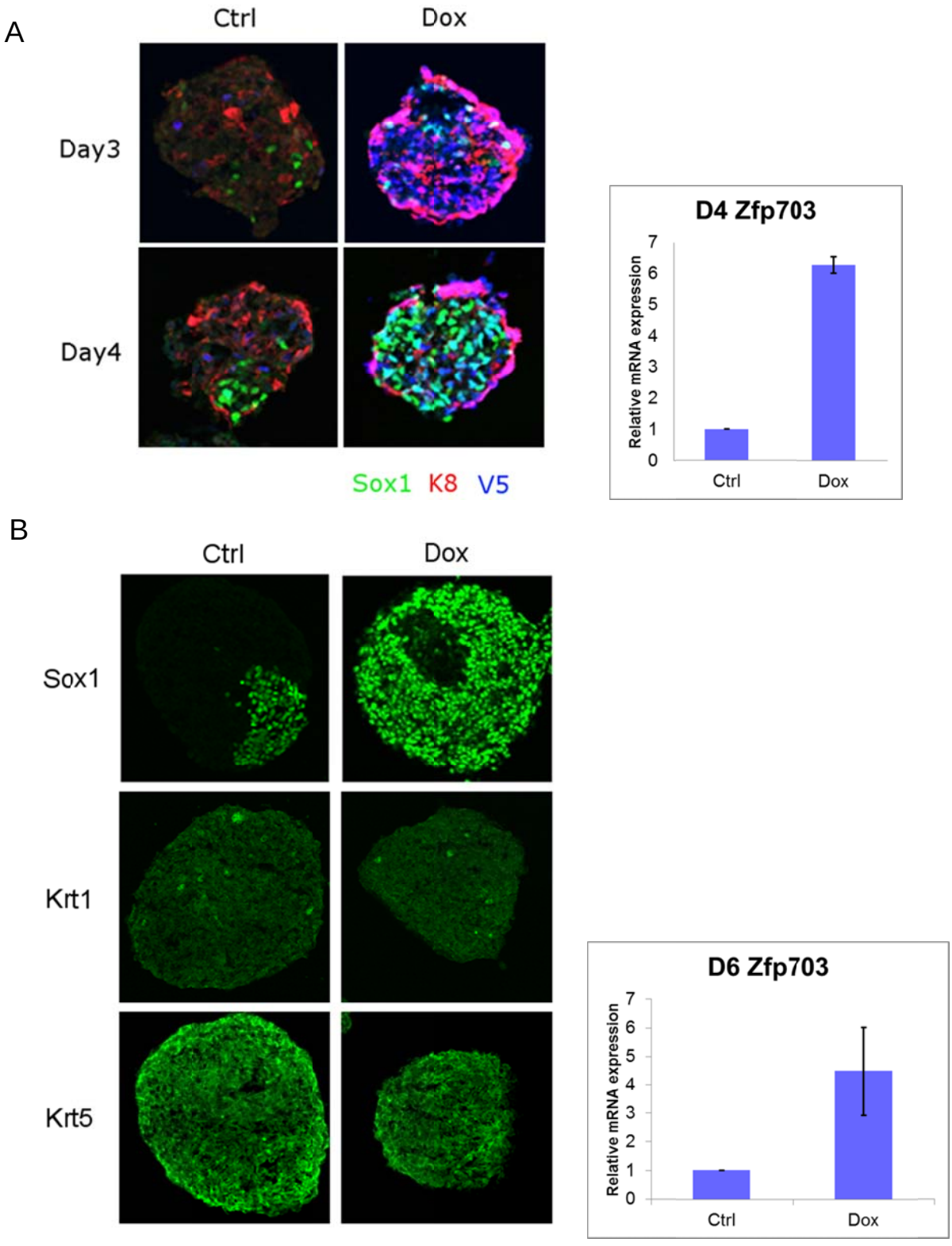

A
Ctrl
Dox
Day3
Day4
Sox1 K8 V5
D4 Zfp703
Relative mRNA expression
0
1
2
3
4
5
6
7
Ctrl
Dox
B
Ctrl
Dox
Sox1
Krt1
Krt5
D6 Zfp703
Relative mRNA expression
0
1
2
3
4
5
6
7
Ctrl
Dox

Figure 3.6

Figure 3.6 Zfp703 mainly promotes neural progenitor fate rather than epidermis

Inducible Zfp703 ESCs were differentiated as EBs for 2 days before treated with or without doxycycline. The EBs were then collected on Day3, Day4, and Day6 for immunocytochemistry analysis using antibodies against neural progenitor marker Sox1, epidermis gene Krt8 (K8), V5 (tagged Zfp703), differentiated epidermis marker Krt1 and epidermal progenitor marker Krt5. Meanwhile, Zfp703 expression was examined by qRT-PCR on Day4 and Day6 (3 independent experiments; mRNA relative expression levels were normalized to non-treated control EBs and shown as average ± standard deviation). A. While there were significantly more K8-positive cells and Sox1-positive cells in doxycycline-treated EBs compared to non-treated (Ctrl) EBs on Day3, the difference in the number of K8-positive cells was less profound on Day4. Importantly, there was a dramatic increase in Sox1-positive cells in doxycycline-treated EBs on Day4, and no cells co-expressed K8 and Sox1. Moreover, there appeared to be a decrease in cells expressing K8 in doxycycline-treated EBs, suggesting the induction of epidermis genes by Zfp703 expression might be transient. B. Sections of Day6 EBs stained for Sox1, Krt1, and Krt5. At this stage, Zfp703-expressing EBs were predominately populated by Sox1-positive neural progenitors compared to non-treated control while staining of Krt1 and Krt5 showed no difference. Together, these data indicated that Zfp703 expression mostly led to neural progenitor specification rather than epidermis fate.

Zfp703 promotes neurogenesis

Detection of a significant number of Sox1 positive cells in Day 6 EBs (Figure 3.6-B) was surprising, since by this time most cells in RA-treated EBs progress from Sox1-positive neural progenitors to NeuN- and Tuj1-positive postmitotic neuronal stage. To address whether Sox1-positive cells in Zfp703-expressing EBs are *bona fide* neural progenitors and committed to differentiate down the path of neural lineage, I stained Day 6 EBs for postmitotic neuronal marker NeuN (Figure 3.7-A). Importantly, I detected NeuN-positive cells in EBs expressing Zfp703 and their number appeared larger compared to the number of NeuN-positive cells in control EBs. However, since the distribution of NeuN-positive cells was heterogeneous, I decided to perform quantification on dissociated and replated cells. I dissociated EBs on Day6 and cultured them as single cells for 2 additional days before staining them for postmitotic neuronal markers NeuN and Tuj1. As shown in Figure 3.7-C, the induced Tuj1-positive cells exhibited typical neuronal morphology with several neurites extending from cell bodies. Quantification of NeuN-positive cells revealed that while there were about 20% NeuN-positive cells in non-treated EBs, 80% and 50% of the cells are NeuN-positive in EBs treated with RA and doxycycline, respectively (Figure 3.7-B). Therefore, Zfp703 is capable of promoting neurogenesis in the absence of exogenous retinoids.

Only 50% of the cells in Zfp703-expressing EBs exhibited postmitotic neuronal markers NeuN and Tuj1 (Figure 3.7-B and 3.7-C), raising the question about the identity of the remaining cells. To address this, I stained Day6 iZfp703 EBs with a panel of neural, epidermal, and oligodendroglial markers including Pax6, Nestin, Krt1, Krt5, and Olig2. None of these markers exhibited significant difference between non-treated, RA-treated, and doxycycline-treated EBs (data not shown), indicating that Sox1-positive cells present in the Zfp703 expressing EBs might

be arrested in a naïve neural progenitor stage. Because the majority of the cells in Zfp703-expressing EBs were positive either for postmitotic neuronal markers NeuN and Tuj1 or neural progenitor marker Sox1, I concluded that the primary effect of Zfp703 is to specify neural identity on top of repressing Oct4.

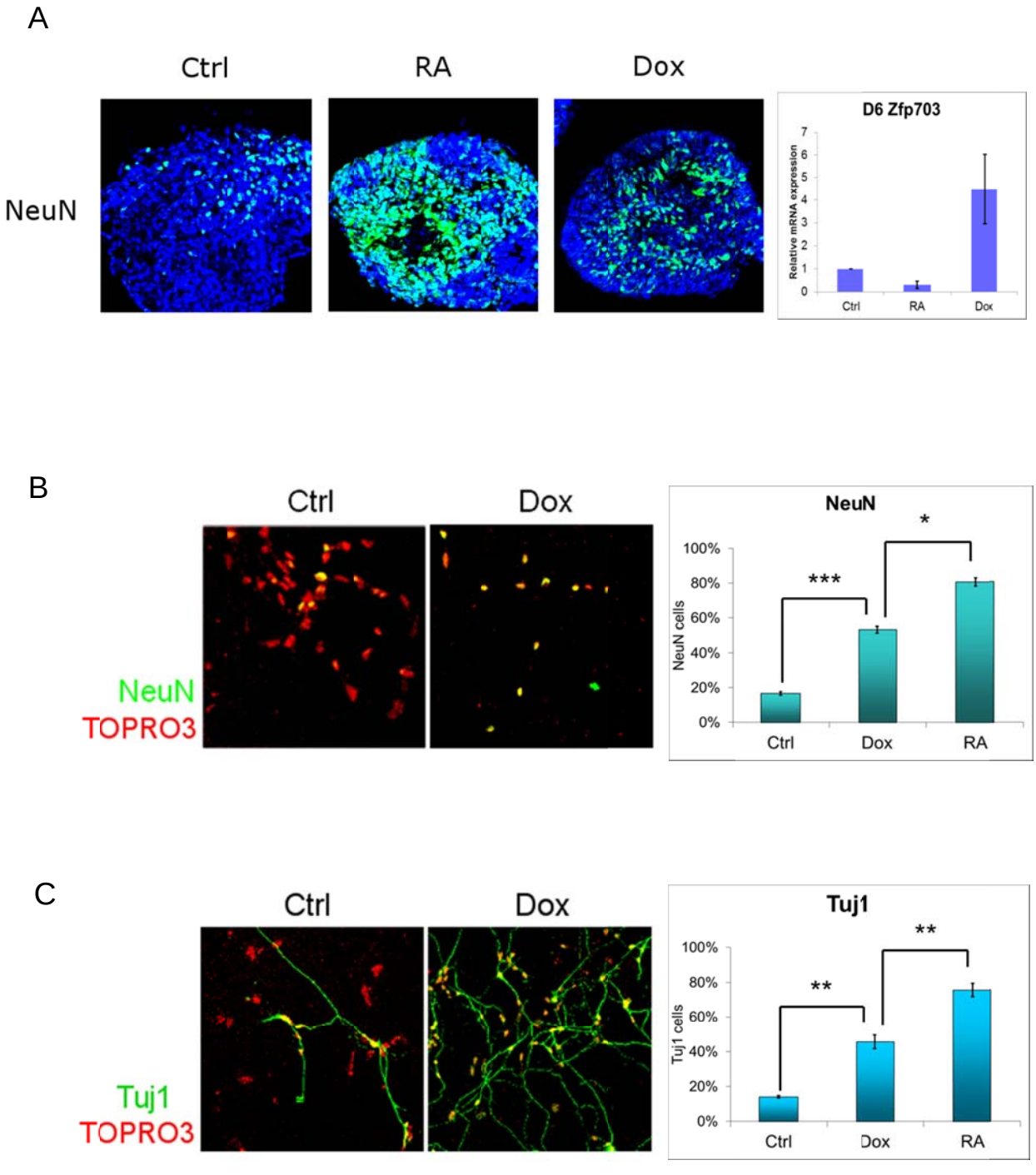

A
Ctrl
RA
Dox
NeuN
D6 Zfp703
Relative mRNA expression
Ctrl
RA
Dox
B
Ctrl
Dox
NeuN
TOPRO3
NeuN
NeuN cells
100%
80%
60%
40%
20%
0%
***
*
Ctrl
Dox
RA
C
Ctrl
Dox
Tuj1
TOPRO3
Tuj1
Tuj1 cells
100%
80%
60%
40%
20%
0%
**
**
Ctrl
Dox
RA

Figure 3.7

Figure 3.7 Zfp703 promotes neurogenesis

Inducible Zfp703 ESCs were differentiated as EBs for 2 days before treated with RA from Day2 to Day5 or doxycycline from Day2 to Day6. On Day6 the EBs were either collected for immunostaining or dissociated and subjected to culture for 2 more days. Moreover, Zfp703 expression was examined by qRT-PCR on Day6 EBs (3 independent experiments; mRNA relative expression levels were normalized to non-treated control EBs and shown as average ± standard deviation). A. Sections of Day6 EBs stained for postmitotic neuronal marker NeuN. While significantly more NeuN-positive cells can be observed in EBs treated with RA or doxycycline, the distribution of NeuN-positive cells was not homogeneous cross EBs subjected to identical treatment. B. Image and quantification of NeuN-positive cells in dissociated cultures on Day8. While ~50% of the cells expressed postmototic neuronal marker NeuN after doxycycline treatment, less than 20% of the cells were NeuN-positive in non-treated group (Ctrl) C. Image and quantification of Tuj1-positive cells in dissociated cultures on Day8 showed similar results. (3 independent experiments; proportion of cells expressing markers relative to total cell number were expression shown as average percentage ± standard deviation; *$p<0.05$, **$p<0.01$, ***$p<0.001$).

Discussion

Effects of Zfp703 on gene expression

In order to identify putative Zfp703 target genes and shed light on its function during neuroectodermal lineage commitment, I conducted gene expression profiling and GO analysis of EBs expressing Zfp703. In addition to the repression of Oct4 which has already been demonstrated through the experiments in chapter 2, Zfp703 overexpression also resulted in strong repression of genes driving mesoderm and endoderm differentiation (i.e. *Brachyury*, *Mixl1*, and *Eomes*) with concomitant induction of genes participating in neurogenesis and epidermis formation. This suggests that Zfp703 might have a broader role in RA-mediated neuroectodermal lineage commitment by promoting neuroectodermal lineage at the expense of mesendodermal differentiation rather than merely repressing Oct4.

The Net family proteins have been reported to function as transcriptional repressors (Nakamura et al., 2004). However, Zfp703 expression in EBs led to more genes significantly upregulated than downregulated (173 vs. 109). This could be explained by two scenarios. First, Zfp703 might repress a repressor of the induced genes. Second, despite that the NET proteins have been reported to interact with corepressors Groucho and HDACs (Runko and Sagerstrom, 2003, 2004), it remains plausible that they can act as context-dependent activators and repressors. Notably, it is thought that NET family proteins cannot bind DNA (Nakamura et al., 2004). Therefore the sequence-specificity and the outcome of transcription regulation for Zfp703 could be determined by the cofactors it interacts with.

Zfp703 modulates a subset of key lineage determinants regulated by RA

Zfp703 is potentially a direct target of RA (Figure 2.6 and 2.7). And its expression in EBs without additional signaling molecules resulted in decreased expression of pluripotency genes and genes specifying mesendodermal lineage as well as increased expression of neuroectodermal genes – the exact changes of gene expression during neuroectodermal lineage commitment exerted by RA. This argues that Zfp703 might be an important downstream effector of RA that is involved in the major transcriptional regulatory events during neural differentiation. It also flags the question to what extent Zfp703 can recapitulate the effect of RA on gene expression. To address this issue, I compared microarray profiling data on Day3 EBs treated with RA or induced to express Zfp703. RA treatment resulted in much more genes significantly upregulated and downregulated than Zfp703 expression (about 10 times more genes in each category, Figure 3.3). This is not surprising because of the pleiotropic effects of RA on gene expression. And Zfp703 might be more specific on gene regulation during ESC differentiation.

When I look at the genes with similar expression pattern in both RA-treated and Zfp703-expressing EBs, I found that Zfp703 only affects a small subset of RA-regulated genes (29 out of 988 genes upregulated and 47 out of 963 genes downregulated by RA, Figure 3.3). Because RA treatment induces a hoard of transcription factors, and Zfp703 is only one of them (Table 2.1), it is reasonable that Zfp703 expression alone cannot mimic the full effects of RA on gene expression. Nevertheless, 43% of the genes downregulated by Zfp703 (47 out of 107, Figure 3.3) are also downregulated by RA; while only 16% of the genes upregulated by Zfp703 (29 out of 173, Figure 3.3) are also upregulated by RA. It appears that Zfp703 is more important for transcriptional repression than activation during neuroectodermal specification, consistent with its proposed role as a transcriptional repressor (Nakamura et al., 2004). Finally, despite the small

overlap of genes regulated by both RA and Zfp703, this group encompasses key players in lineage commitment: the neuroectodermal fate determinants *Sox1* and *Nr2f2*, genes specifying mesendoderm *Brachyury (T)*, *Eomes*, *Wnt3*, and *Lef1*, as well as the epiblast marker Fgf5 (Table 3.3). In conclusion, Zfp703 might be an important downstream effector of RA which exerts transcriptional regulation on critical genes during the fate choice of differentiating ESCs between neuroectoderm and mesendoderm.

Zfp703 and transcriptional network of neural induction

The data in this chapter suggested that Zfp703 might have a broader role in RA-induced neuroectodermal lineage commitment beyond Oct4 repression. Specifically, Zfp703 expression led to the emergence of neural progenitors and subsequently, postmitotic neurons in EBs without RA treatment. Thus Zfp703 is sufficient to induce neural fate. In addition to Sox1, Zfp703 expression also resulted in upregulation of Nr2f2 and Zeb2 (Figure 3.2); both are critical in neural differentiation (Ozair et al., 2013). Interestingly, Nr2f2 has been shown to be required for Oct4 repression and neural lineage commitment in human ESCs (Rosa and Brivanlou, 2011). Hence it is possible that Zfp703 specifies neural fate by initiating a transcriptional network encompassing Sox1, Nr2f2, and Zeb2 which both consolidates the repression of the pluripotency program and initiates neural differentiation program.

Zfp703 mainly promotes neural rather than epidermis differentiation

While I observed transient upregulation of epidermal genes including keratins and collagens in Zfp703-expressing EBs (Figure 3.2), keratins were not induced following RA treatment (although Col1a2 was upregulated by both Zfp703 and RA, Table 3.3), indicating that Zfp703 initially promotes neuroectodermal identity that is later refined to produce mostly neural cells. Lack of keratin induction in RA-treated EBs suggests the existence of Zfp703-independent

repressors of early epidermal fate that are induced in response to RA treatment. Despite initial elevation of Krt8 and Krt18 mRNA and protein in Day3 EBs expressing Zfp703 (Figures 3.5 and 3.6-A), the Krt8- or Krt18-positive cells tend to localize mainly at the periphery of EBs (Figure 3.6-A), and their numbers decrease as differentiation progresses. Already by Day4, Sox1-positive neural progenitors greatly outnumbered cells expressing Krt8 (Figure 3.6-A). Importantly, there were no cells co-expressing Sox1 and Krt8. The exact identity of Krt8- and Krt18-positive cells has not been determined. While Krt8 and Krt18 are highly expressed in Merkel cells mediating the sense of touch (Van Keymeulen et al., 2009), the definitive marker of Merkel cells, Atoh1 (Morrison et al., 2009) was not induced throughout the differentiation (data not shown). Lastly, at the end of differentiation (Day6), preliminary immunostaining of epidermis progenitor maker Krt5 and differentiated epidermis marker Krt1 showed no difference between untreated EBs and EBs expressing Zfp703 (Figure 3.6-B), indicating that the epidermal fate is not a major output of Zfp703 induction.

Gap in efficiency between Zfp703- and RA-induced neurogenesis

Zfp703 is sufficient to induce neural fate. However, when compared to RA, Zfp703 expression alone in EBs led to markedly lesser extent of upregulation in Sox1 mRNA and fewer neural progenitors on Day3 (Figure 3.4-B and 3.4-C) as well as fewer postmitotic neurons on Day6 (Figure 3.7). The difference in the efficiency of specifying neural fate between Zfp703 and RA can be explained in two scenarios. First, Zfp703 might not be the only effector downstream of RA responsible for promoting neurogenesis. There are additional factors functioning in a redundant manner. This is supported by the fact that Zfp703 can only recapitulate the effects of RA among a small group of genes (Figure 3.3). Second, although Zfp703 expression only led to moderate increase of neural progenitors in Day3 EBs, significantly more Sox1-positive cells are

present in Day4 and Day6 EBs expressing Zfp703 (Figure 3.6). It remains to be determined whether Sox1- positive cells present in Day6 EBs are frozen in an immature naïve state that they cannot escape, or whether they simply lack a patterning signal that could initiate a delayed program on neural differentiation. Nevertheless, the delayed emergence of Sox1 cells in Zfp703 expressing EBs suggests that the effect on neurogenesis of Zfp703 is likely to be indirect. And the accumulation of Sox1 cells indicates that there are Zfp703-independent mechanisms to promote the maturation of neural progenitors downstream of RA.

Zfp703 partially recapitulates important effects of RA on lineage commitment

In conclusion, Zfp703 expression promotes neurogenesis with the initial induction of both neural and epidermal fates. Because I already demonstrated that Zfp703 is sufficient to repress key pluripotency gene Oct4, Zfp703 seems to be involved in both the termination of the ESC transcriptional program and the initiation of neuroectodermal transcriptional networks, two critical events during RA-mediated neuroectodermal lineage commitment (Figure 3.8). Based on microarray profiling, Zfp703 expression also strongly represses mesendodermal fate (Figure 3.2). In the next chapter I will verify the effect of Zfp703 on medendoderm specification and examine the transcriptional cascade of Zfp703-mediated fate choice during ESC differentiation.

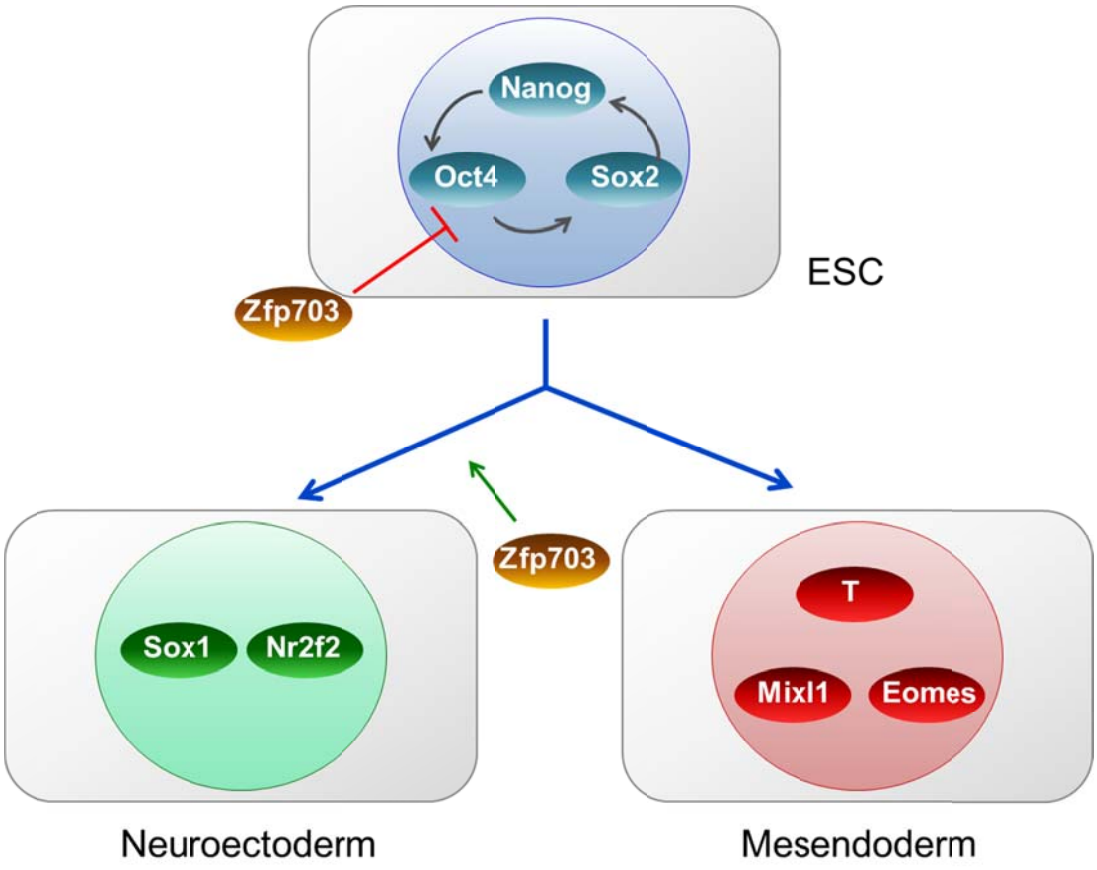

Nanog
Oct4
Sox2
ESC
Zfp703
Zfp703
Sox1
Nr2f2
T
Mixl1
Eomes
Neuroectoderm
Mesendoderm

Figure 3.8

Figure 3.8 Zfp703 suppresses pluripotency and promotes neurogenesis

Zfp703 expression alone is sufficient to repress key pluripotency gene Oct4 and promote the exit of the ESC state, rendering the cells poised to make the fate choice between neuroectoderm and mesendoderm. Moreover, Zfp703 expression is also sufficient to induce neuroectodermal determinants Sox1 and Nr2f2, pushing the cells to differentiate along the neuroectodermal lineage and eventually leading to enhanced neurogenesis.

Chapter 4. Zfp703 suppresses mesendodermal fate

Introduction

Proper specification of neuroectodermal lineage has to be accompanied by the suppression of mesendodermal fate (Thomson et al., 2011). Indeed, exogenous RA has been reported to repress mesodermal marker Brachyury and Wnt-3a in developing mouse embryo (Iulianella et al., 1999). However it remains unknown whether RA primarily activates the neural transcriptional program that represses mesendoderm differentiation or whether it enables execution of the default neural fate by primarily repressing mesendodermal transcriptional program. Identification of Zfp703 as an important target of RA signaling provided an entry point to start dissecting the hierarchy of regulatory network underlying RA-mediated neural specification.

To examine the role of Zfp703 in the fate choice between neuroectoderm and mesendoderm in differentiating ESCs, in this chapter I set out to address 3 questions: (1) Is Zfp703 sufficient to counteract activity of mesendoderm-inducing signals? (2) Does expression of Zfp703 in committed mesendodermal cells lead primarily to induction of neural markers or to repression of mesendodermal markers? (3) Does Zfp703 bind in the proximity of neural or mesendodermal lineage genes? (4) Is Zfp703 required for neural induction and/or mesendoderm repression? For the first two questions, I adopted ESC differentiation conditions promoting specification of mesendodermal lineage and tested the effect of RA treatment and Zfp703 overexpression on the expression of mesendodermal markers. For the last two questions, I performed ChIP-seq experiments on ZFp703-expressing EBs to identify its genome-wide binding sites as well as luciferase assay to examine the effect of Zfp703 binding on the expression of adjacent genes. Finally, I derived Zfp703 null ESCs and conducted preliminary loss-of-function analyses of Zfp703.

Results

Zfp703 suppresses the acquisition of mesendodermal fate

The microarray profiling data from Zfp703-expressing Day3 EBs indicates that Zfp703 strongly represses mesendodermal determinants such as *Brachyury*, *Eomes*, and *Mixl1* (Figure 3.2). To test whether Zfp703 expression is sufficient to repress mesendodermal fate, I adopted a mesendodermal differentiation protocol in which inducible Zfp703 ESCs were kept in serum-free media as monolayer for 2 days followed by treatment with GSK3β-inhibitor (GSK3i) and ActivinA. As shown in Figure 4.1-B, treatment of these two factors led to robust induction of Brachyury-expressing cells compared to control, confirming the effective mesendoderm differentiation. After verifying the protocol, I first tested if Zfp703 expression is sufficient to prevent the induction of mesendodermal fate by adding either RA or doxycycline to the culture on Day1 (Figure 4.2-A). As shown in Figure 4.2-B, both RA treatment and Zfp703 expression prior to the addition of GSK3i and ActivinA strongly interfered the induction of *Brachyury, Eomes,* and *Mixl1* mRNA. Importantly, pretreatment of cells with RA or doxycycline also led to the upregulation of *Sox1* mRNA in cells despite the subsequent exposure to mesendoderm-inducing agents (Figure 4.2-C). Consistent with these observations, on Day3 around 80% of the cells receiving GSK3i and ActivinA expressed Brachyury protein; whereas very few cells pretreated with RA or doxycycline were positive for Brachyury. Conversely, while plenty of Sox1-positive cells existed in cultures with RA or doxycycline, they were rare among control cells receiving GSK3i and ActivinA (Figure 4.3). Together these results indicate that Zfp703 is sufficient to efficiently prevent the induction of mesendodermal identity and induce neuroectodermal differentiation when expressed prior to the exposure of cells to mesendoderm-specifying signal.

A


Day 0
Day 2
GSK3i
ActvinA
Day 3

B

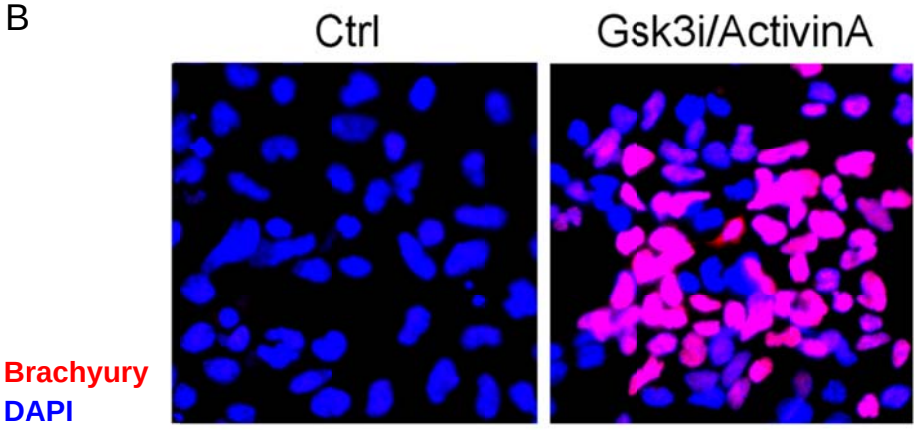

Ctrl
Gsk3i/ActivinA
Brachyury
DAPI

Figure 4.1

Figure 4.1 Gsk3i and Activin A drives mesendodermal differentiation

ESCs were dissociated and plated as monolayer in serum-free medium on Day0 and kept in serum-free condition from Day0 to Day2. On Day2 Gsk3i and ActivinA were added. Subsequently the cells were fixed and stained on Day3. A. Schematic of ESC differentiation. B. immunostaining of Day3 cultures for mesendodermal marker Brachyury and nuclear marker DAPI. Treatment of Gsk3i and ActivinA resulted in robust emergence of Brachyury-positive cells.

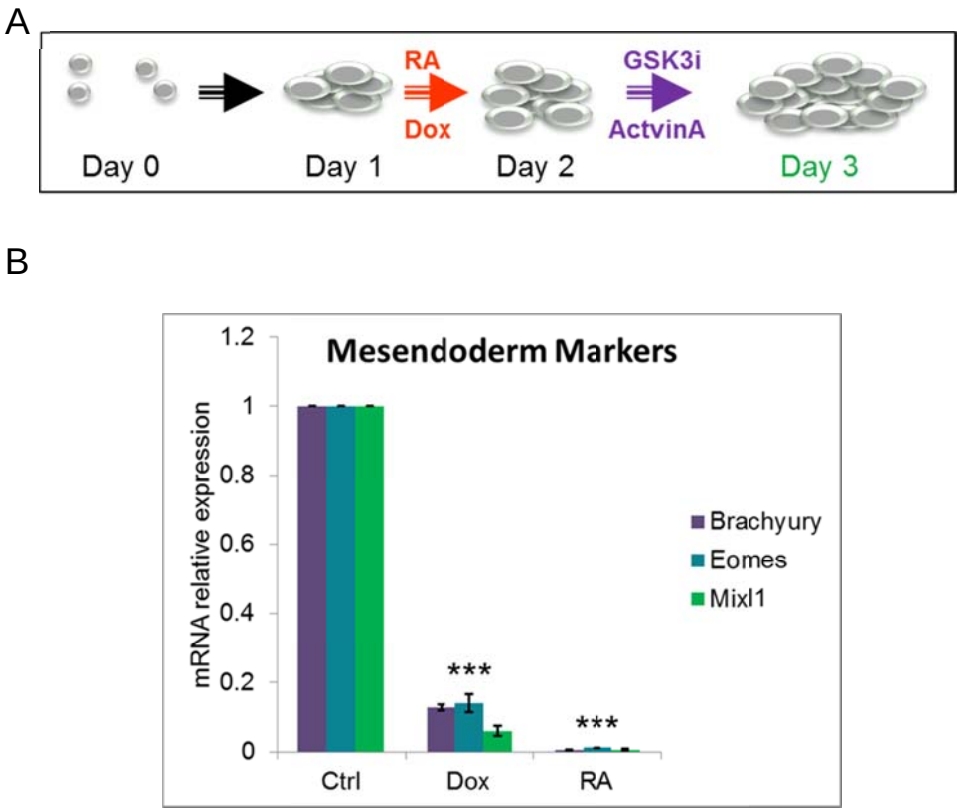

A
RA
Dox
GSK3i
ActvinA
Day 0
Day 1
Day 2
Day 3
B
Mesendoderm Markers
mRNA relative expression
1.2
1
0.8
0.6
0.4
0.2
0
***
***
Brachyury
Eomes
Mixl1
Ctrl
Dox
RA
C
Sox1
mRNA relative epression
100
10
1
*
**
Ctrl
Dox
RA
D
Zfp703
Relative mRNA expression
14
12
10
8
6
4
2
0
Ctrl
Dox
RA

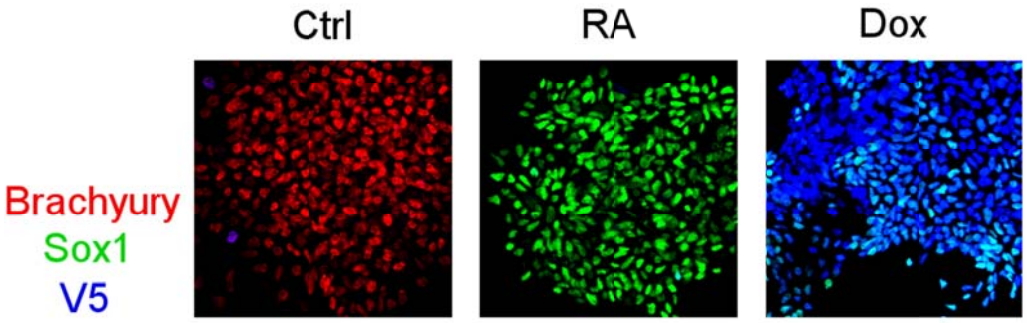
Figure 4.2

Figure 4.2 Zfp703 prevent the response to mesendoderm-inducing signals

Inducible Zfp703 ESCs were plated as monolayer on Day0. RA or doxycycline was added on Day1. Then the cells were treated with ActivinA and GSK3i on Day2 for 24 hours. The cells were collected on Day3 for qRT-PCR analysis. A. Schematic of ESC differentiation. B. Expression of mesendoderm markers *Brachyury, Eomes*, and *Mixl1* on Day3. Pretreatment of doxycycline or RA on Day1 led to significantly lower mRNA levels of these mesendoderm markers, suggesting that both Zfp703 and RA prevented the cells from responding to mesendoderm induction mediated by ActivinA and GSK3i. C. Expression of neural progenitor marker Sox1. D. Expression of Zfp703. Pretreatment of doxycycline or RA on Day1 led to significantly higher mRNA level of Sox1 on Day3, suggesting that both Zfp703 and RA promoted neural progenitor specification even that the cells were exposed to ActivinA and GSK3i from Day2 to Day3 (3 independent experiments; mRNA relative expression levels were normalized to ESCs treated with only ActivinA and GSK3i and shown as average ± standard deviation; *$p<0.05$; **$p<0.01$; ***$p<0.001$).

.

A


Ctrl
RA
Dox
Brachyury
Sox1
V5

B

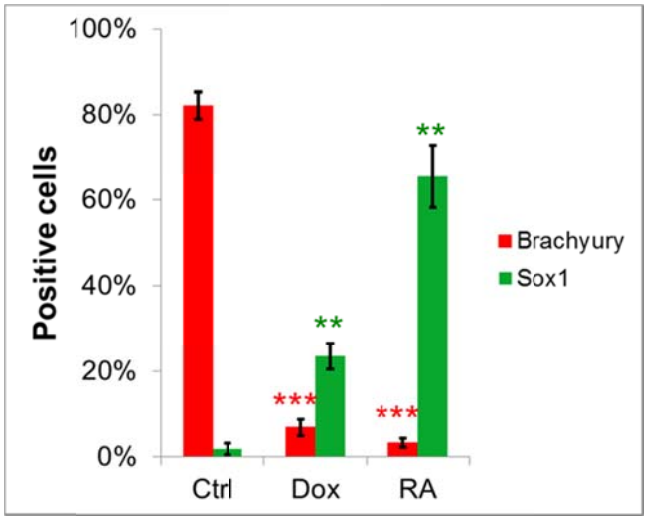

100%
80%
60%
40%
20%
0%
Positive cells
Ctrl
Dox
RA
Brachyury
Sox1
**
**
***
***

Figure 4.3

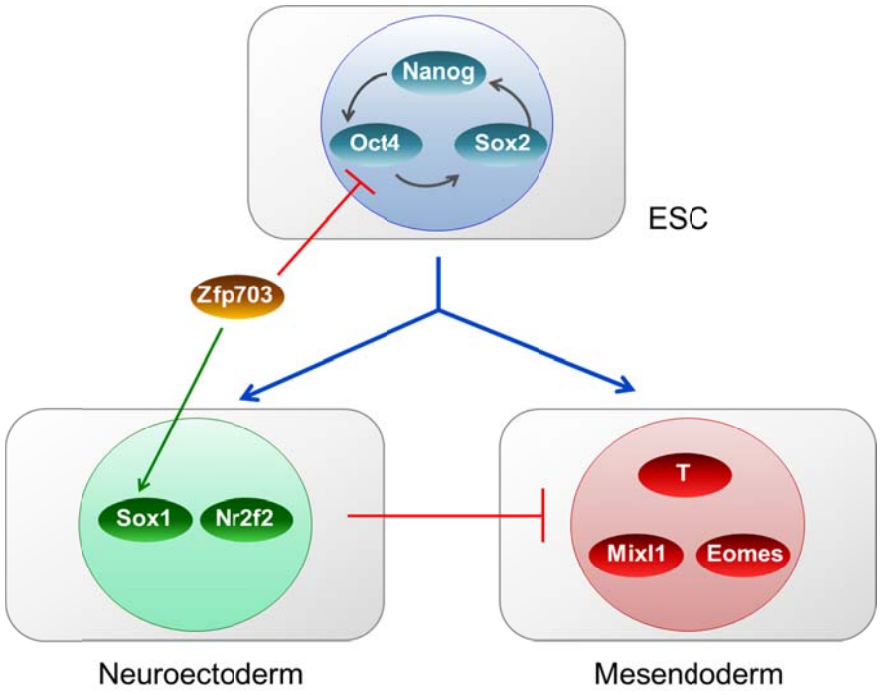
Figure 4.3 Zfp703 and RA favor neuroectoderm over mesendoderm fate

Inducible Zfp703 ESCs were plated as monolayer on Day0. RA or doxycycline was added on Day1. Then the cells were treated with ActivinA and GSK3i on Day2 for 24 hours. The cells were fixed on Day3 for immunocytochemistry analysis. A. Expression of the neuroectodermal marker Sox1, mesendodermal marker Brachyury, and Zfp703 (tracked by V5 staining). While ActivinA and GSK3i induced extensive Brachyury expression in control cells on Day3, pretreatment of doxycycline or RA resulted in very few Brachyury-expressing cells. Meanwhile, doxycycline and RA treatment led to the emergence of Sox1-positive cells on Day3. B. Quantification of cells expressing Brachyury or Sox1 protein on Day3 (3 independent experiments; proportion of cells expressing Sox1 or Brachyury relative to total cell numbers were show as average percentage ± standard deviation; **$p<0.01$; ***$p<0.001$).

Zfp703 represses mesendodermal marker Brachyury in committed cells

So far, I have demonstrated that Zfp703 expression is sufficient to induce neuroectodermal commitment (Figure 3.3 and 3.6) and suppress mesendodermal fate (Figure 4.2 and 4.3). However, the transcriptional hierarchy underpinning these phenomena has not been elucidated. Specifically, there are two putative models to explain the mechanism of Zfp703 in ESC fate choice (Figure 4.4). Zfp703 might actively induce important components in the neuroectodermal lineage such as Sox1 and Nr2f2, which in turn repress mesendodermal determinants Brachyury, Eomes and Mixl1 (Figure 4.4-A). Alternatively, Zfp703 might primarily repress mesendodermal determinants therefore removing the inhibitory effects of these factors on neuroectodermal genes and de-repressing the neuroectodermal fate (Figure 4.4-B). To distinguish between these two models, I examined whether Zfp703 can repress mesendodermal determinants without inducing neuroectodermal markers. To do this, I treated cells with GSK3i and ActivinA for 24 hours before the addition of RA or doxycycline (Figure 4.5-A). 12 hours later, there was significant downregulation of *Brachyury* mRNA in cells receiving RA or doxycycline (Figure 4.5-B). Immunostaining of Brachyury protein showed similar results (Figure 4.5-C). Interestingly, under this condition neither RA nor Zfp703 could induce *Sox1* (Figure 4.5-C). Taken together, these experiments demonstrate that in nascent mesendodermal cells Zfp703 is sufficient to repress key mesendodermal marker Brachyury, however it is not sufficient to induce expression of early neuroectodermal markers. Because Zfp703 represses Brachyury without inducing Sox1, the second model (Figure 4.4-B) seems to be more accurate that Zfp703 functions primarily as a repressor of mesendodermal fate which consequently leads to the de-repression of neuroectodermal potential in differentiating ESCs grown under basal conditions.

A


Nanog
Oct4
Sox2
ESC
Zfp703
Sox1
Nr2f2
T
Mixl1
Eomes
Neuroectoderm
Mesendoderm

B

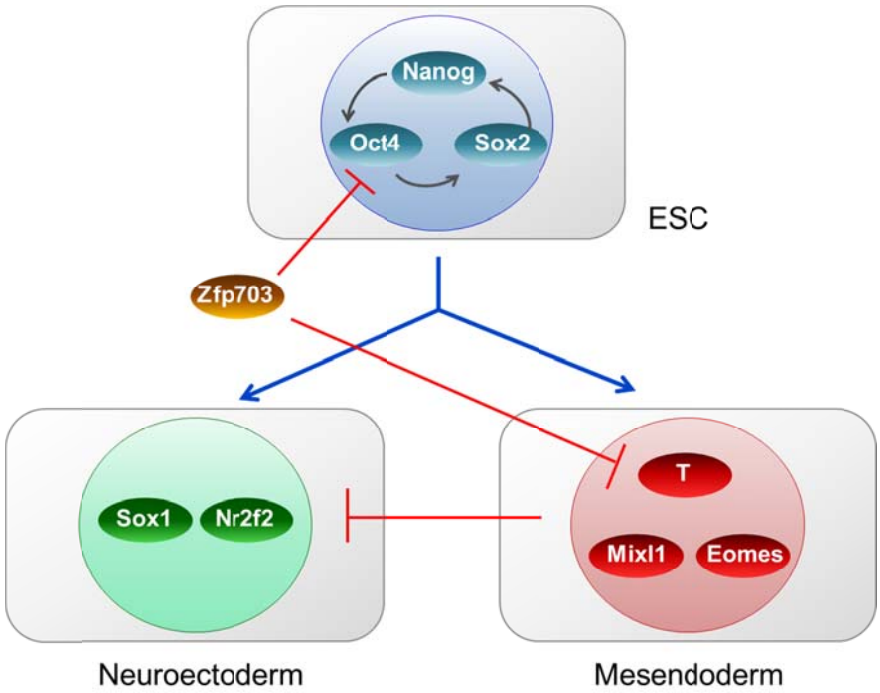

Nanog
Oct4
Sox2
ESC
Zfp703
Sox1
Nr2f2
T
Mixl1
Eomes
Neuroectoderm
Mesendoderm

Figure 4.4

Figure 4.4 Putative models of transcriptional cascade downstream of Zfp703

There are two possible scenarios of transcriptional hierarchy in Zfp703-mediated fate choice of neuroectoderm over mesendoderm. A. Zfp703 actively induces key transcriptional factors in neuroectodermal lineage such as *Sox1* and *Nr2f2*, which in turn repress mesendodermal determinants *Brachyury*, *Eomes*, and *Mixl1*. B. Zfp703 primarily represses mesendodermal determinants and thus removing their inhibition of neuroectodermal factors and de-repressing neuroectodermal fate.

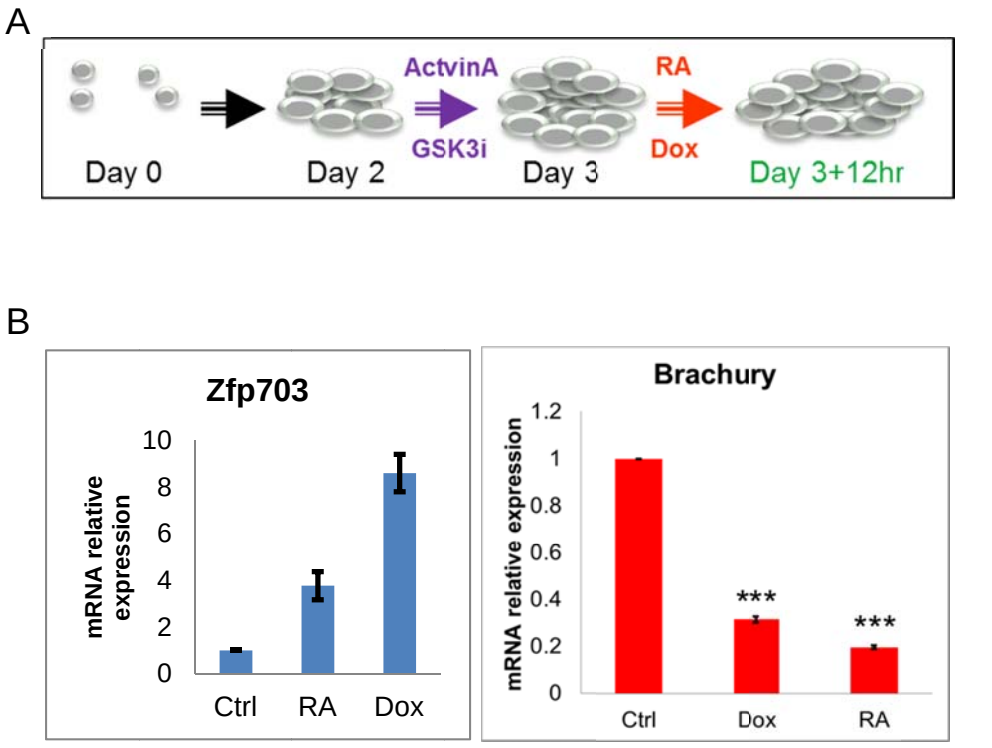

A
ActvinA
GSK3i
RA
Dox
Day 0
Day 2
Day 3
Day 3+12hr
B
Zfp703
mRNA relative expression
10
8
6
4
2
0
Ctrl
RA
Dox
Brachury
mRNA relative expression
1.2
1
0.8
0.6
0.4
0.2
0
***
***
Ctrl
Dox
RA
C
Ctrl
RA
Dox
Sox1
Brachyury
DAPI

Figure 4.5

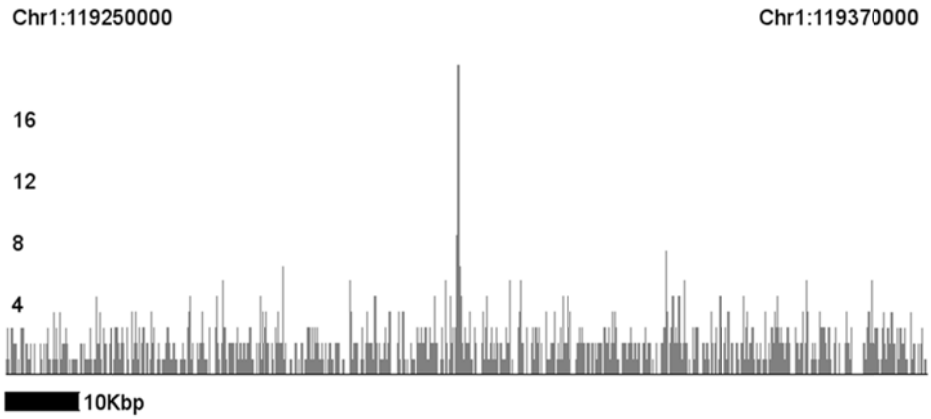
Figure 4.5 Zfp703 and RA represses Brachyury in committed mesendodermal cells

Inducible Zfp703 ESCs were plated as monolayer on Day0 and treated with ActivinA and GSK3i on Day2. RA or doxycycline was added on Day3, and cells were collected 12 hours later for qRT-PCR and immunocytochemistry analysis. A. Schematic of ESC differentiation. B. Expression of Zfp703 and *Brachyury* mRNAs at Day3+12hrs. Treating cells with doxycycline or RA after they were committed to mesendoderm fate through exposure of ActivinA and GSK3i still resulted in repression of *Brachyury* mRNA (3 independent experiments; mRNA relative expression levels were normalized to ESCs treated with only ActivinA and GSK3i and shown as average ± standard deviation; ***$p<0.001$). C. Expression of Brachyury and Sox1 proteins at Day3+12hrs. Brachyury was widely expressed in control cells, whereas treatment of doxycycline or RA greatly reduced the number of Brachyury-positive cells in culture. Meanwhile, the Sox1 staining was negative in all three conditions.

Zfp703 ChIP-seq yielded few binding sites

Now I have established and tested a model of transcriptional cascade downstream of Zfp703 in the fate choice of ESCs: Zfp703 primarily represses mesendodermal determinants and therefore de-represses neuroectodermal factors to steer the differentiation towards neuroectoderm and subsequently neuronal lineage. There are two outstanding questions remaining: (1) Whether Zfp703 directly represses pluripotency and mesendodermal genes; (2) whether Zfp703 is necessary for RA-mediated neuroectodermal lineage commitment. To address the first question I designed experiments to map genome-wide binding sites of Zfp703 by chromatin immunoprecipitation followed by deep sequencing (ChIP-seq). I differentiated inducible Zfp703 ESCs as EBs and treated them with doxycycline on Day2 for 24 hours before fixing and lysing them on Day3. Subsequently, the cell lysate was fractioned with sonication before subjected to immunoprecipitation with control IgG or V5 antibody. The protein-chromatin complexes pulled down were then sequenced and resulting data analyzed by our collaborators in David Gifford's computational lab at MIT. Among three independent ChIP-seq experiments, we identified only 330 high-confidence binding sites ($p<0.0002$, one such site is shown in Figure 4.6-A), much fewer than in other ChIP-seq experiments for developmentally regulated transcription factors that typically yield 5,000-30,000 binding sites. The low number of Zfp703 peaks might either reflect high degree of selectivity in its recruitment to DNA or might reflect less efficient immunoprecipitation because of indirect binding of Zfp703 to DNA. Motif analysis indicates that 163 out of 330 sites containing the sequence ACAA(A/T)(G/A)G. A search for conserved motifs suggests that this sequence is bound by Sox proteins of the HMG group (Figure 4.6-B). Among the top 140 sites, 7 are intragenic (2 with motif), and 133 are intergenic. Only 7 sites appeared in the 20Kbp vicinity of top 100 Zfp703-regulated genes (Table 3.1).

A


Chr1:119250000
Chr1:119370000
16
12
8
4
10Kbp

B


MN D3 iZfp703 3, #0,

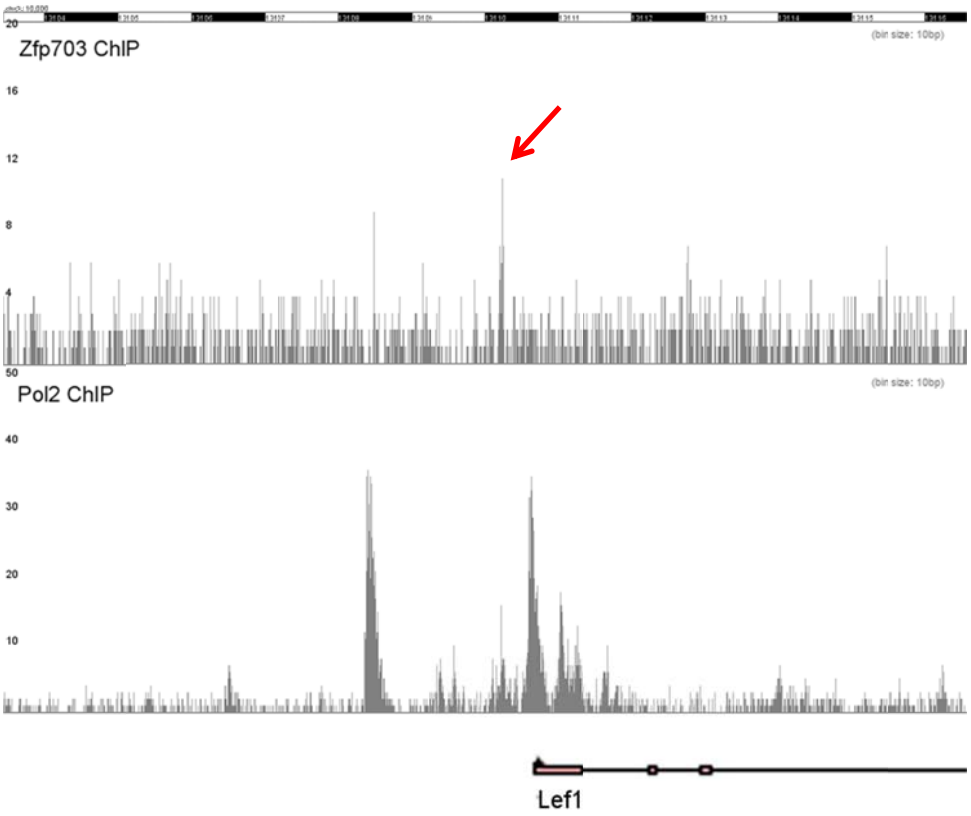
Figure 4.6 Zfp703 ChIP-Seq

Inducible Zfp703 ESCs were differentiated as EBs and treated with or without doxycycline on Day2. On Day3 the cells were fixed and subjected to ChIP with control IgG or V5 antibody. The precipitated DNA fragments then underwent deep sequencing and computational analysis conducted by David Gifford's lab at MIT. 330 high-confidence binding sites were identified with p value <0.0002. A. One representative high-confidence Zf703 binding site on chromosome 1. B. 163 out of 330 high-confidence sites possess the HMG protein (Sox2, Sox3, and Sox6) binding motif ACAA(A/T)(G/A)G.

Zfp703 represses Lef1

Because the majority of high-confidence binding sites were found far from transcription start sites of genes regulated following Zfp703 expression, it was difficult to correlate ChIP-Seq and microarray results. It also suggests that Zfp703 might regulate these genes came indirectly through additional mechanisms. To identify such mechanisms, we adopted less stringent criteria for peak-calling and detected one binding site in the promoter region of Lef1 (7Kb upstream from transcription start site, TSS; p value: 0.003), a critical transcription factor mediating the activation of genes in response to Wnt signaling (Figure 4.7-A). Since the fold enrichment for this binding site was 3.6 IP/WCE (whole cell extract), and the q-value is 0.089 after multiple hypothesis testing correction, it was not included in the 330 high-confidence binding sites. To validate whether this is a real binding site, I performed ChIP-PCR and confirmed the interaction (Figure 4.7-B). Because Lef1 also came up as a gene downregulated by Zfp703 expression in the array data (Figure 3.2), I then examined whether Zfp703 represses Lef1 through binding to its promoter. I cloned the 1Kb region of Lef1 promoter encompassing the Zfp703 binging site into a luciferase reporter vector (Figure 4.8-A) and co-expressed it with or without Zfp703 in 293T cells. Zfp703 expression resulted in ~2-fold reduction in luciferase activity compared to control (Figure 4.8-B). Therefore, Zfp703 likely directly binds to the promoter of Lef1 and seems to repress its expression. Given the close relationship between Lef1 and Wnt signaling (Mao and Byers, 2011), one plausible mechanism of gene regulation exerted by Zfp703 might be through the attenuation of Wnt pathway. This would argue that Zfp703 regulates pluripotency and mesendodermal genes indirectly.

A


Zfp703 ChIP
(bin size: 10bp)
16
12
8
4
50
Pol2 ChIP
(bin size: 10bp)
40
30
20
10
Lef1

B

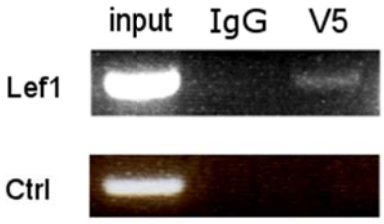

input
IgG
V5
Lef1
Ctrl

Figure 4.7

Figure 4.7 Zfp703 binds to the promoter of Lef1

Inducible Zfp703 ESCs were differentiated as EBs for 2 days before treated with doxycycline. 24 hours later the EBs were dissociated and fixed before lysed and sonicated. The lysate was then subjected to ChIP using either control IgG or V5 antibody. A. Image of ChIP-seq for Zfp703 and Pol2 around Lef1 locus. There was a peak at ~7 Kbp upstream of the transcription start site of Lef1 (arrow), suggesting Zfp703 binds to its promoter. B. ChIP-PCR using primer pairs flanking Zfp703 binding site in Lef1 promoter (Lef1) and an unbound control region (Ctrl). While product was generated from both lysate (input) and ChIP using V5 antibody with Lef1 primers, ChIP with control IgG and PCR with Lef1 primers yielded no band, neither did PCR using Ctrl primers, confirming that Zfp703 interacts with Lef1 promoter.

A


Zfp peak
Lef1
1Kbp
TSS
Zfp peak
Luciferase

B

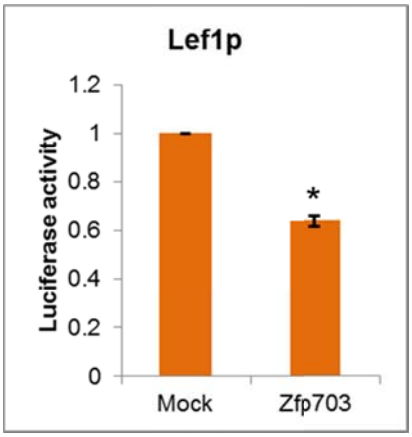

Lef1p
Luciferase activity
1.2
1
0.8
0.6
0.4
0.2
0
*
Mock
Zfp703

Figure 4.8

Figure 4.8 Zfp703 represses the activity of Lef1 promoter

HEK293T cells were transfected with plasmid containing luciferase under the control of Lef1 promoter together with mock or Zfp703 expression vectors. 24 hours later the cells were lysed, and the lysate was subjected to luciferase assay. A. The schematic of the luciferase construct. The 1Kbp sequence encompassing the Zfp703 binding site called in ChIP-seq analysis was cloned upstream of luciferase reporter gene. B. Expression of Zfp703 led to ~40% decrease in luciferase activity compared to control, suggesting that Zfp703 is able to repress Lef1 promoter (3 independent experiments; luciferase activities in Zfp703-expressing cells were normalized to that of mock transfected cells and shown as average percentage ± standard deviation; *$p<0.05$).

Conditional targeting of Zfp703

To address the second outstanding question whether Zfp703 is an indispensable downstream effector of RA-mediated neuroectodermal lineage commitment, I set to generate a conditional Zfp703 allele. The targeting vector was generated by recombineering using bacterial artificial chromosomes (BACs) (Malureanu, 2011). In the vector, the second exon of Zfp703, which encodes the bulk of the protein and major functional domains, was flanked by loxP sites (Figure 4.9). I screened 200 targeted ESC clones and generated chimera mice with the help of the transgenic facility. Afterwards, I mated the chimera with flippase-expressing mice to get rid of the Neomycin-resistant cassette and obtained 4 F1 pups carrying one floxed allele (2 still retained Neo cassettes) and one wild-type allele (heterozygous). During subsequent 2 rounds of the mating between the 2 pairs of heterozygous F1 animals, I obtained 24 F2 pups, of which 17 were heterozygous, and 7 were wild-type. The mating between 4 pairs of heterozygous F2 animals (half retained Neo cassettes) yielded 15 heterozygous and 10 wild-type. Together, these matings generated 32 heterozygous, 17 wild-type, and none homozygous F2/F3 pups (Table 4.1). The lack of homozygous progenies and the 1.9-to-1 ratio between heterozygous and wild-type mice among all F2 and F3 pups suggested surprising lethality of animals homozygous for the floxed Zfp703 allele. To circumvent the embryonic lethality of homozygous mice I decided to derive ESC lines from mating between heterozygous mice (Table 4.2). Out of 4 pregnant females I derived 20 ESC lines, of which 13 were heterozygous, 7 were wild-type, and none was homozygous. This suggested that the floxed allele is homozygous lethal during preimplantation stage. In this approach we had nearly 100% success in ESC derivation, therefore the lethality likely happens before E3.0. Hence, this strategy could not be used for Zfp703 loss-of-function analysis.

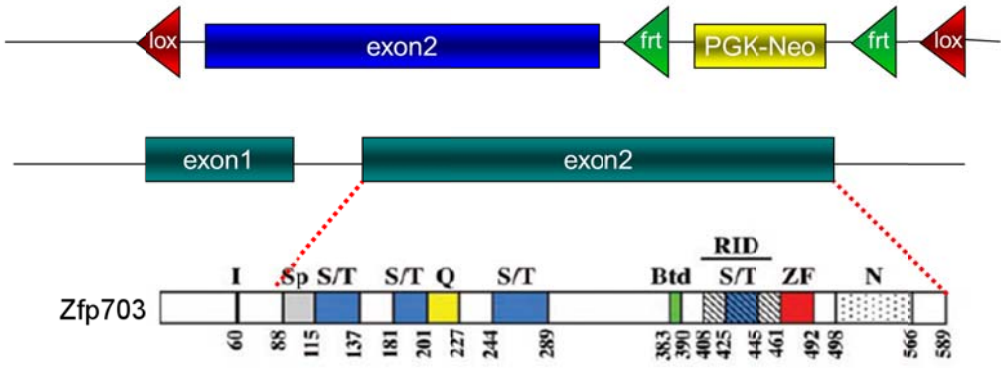

lox
exon2
frt
PGK-Neo
frt
lox
exon1
exon2
RID
I
Sp
S/T
S/T
Q
S/T
Btd
S/T
ZF
N
Zfp703
60
88
115
137
181
201
227
244
289
383
390
408
425
445
461
492
498
566
589

Figure 4.9

Figure 4.9 Conditional targeting strategy of Zfp703

Top: The targeting construct was designed that the exon2 of Zfp703 together with a Neomycin-resistant selection cassette were flanked by loxP sites. Bottom: Exon2 of Zfp703 encodes the bulk of the protein and major functional domains Sp: the SP motif, S/T: serine/threonine-rich regions, Q: the glutamine-rich region, Btd: the Buttonhead box, RID: the repressor interaction domain, ZF: the C_2H_2 zinc finger, N: a region required for nuclear localization (modified from (Runko and Sagerstrom, 2004)).

Table 4.1 Mating schemes of Zfp703 conditional targeting mice

After the generation of chimeras ESCs heterozygous for the floxed conditional allele, the chimeras were mated to flippase-expressing mice (FlpO) to remove the Neo-selection cassette. The resulting mice (F1) heterozygous for the floxed allele with (Zfp703f, Neo/+) or without Neo cassette (Zfp703 f/+) were mated to generate F2, and the heterozygous animals were mated again to produce F3. After 3 rounds of mating, 32 out of 49 total F2 and F3 progenies were heterozygous whereas 17 out of 49 were wild-type. The 1.9-to-1 ratio between these genotypes suggested that the floxed conditional allele caused homozygous lethality.

F1: Chimera ♂ x FlpO ♀
4/22 Het.; 18/22 Wt.

F2: Zfp703 f/+ (♂x♀)	Zfp703 f,Neo/+ (♂x♀)
11/11 Het.	6/13 het.; 7/13 Wt.
F3: Zfp703 f/+ (♂x♀)	Zfp703 f,Neo/+ (♂x♀)
9/16 Het.; 7/16 Wt.	6/9 Het.; 3/9 Wt.

Total F2 & F3: 32/49 Het.; 17/49 Wt. → homozygous lethal

Table 4.2 ESC derivation for the floxed conditional allele

Matings were set up between F2 and F3 animals heterozygous for the floxed allele with (Zfp703f, Neo/+) or without Neo-cassette (Zfp703 f/+), and ESC lines were derived from plugged females at E3.5. Among the 4 matings, 13 out of 20 ESC lines derived were heterozygous; while the remaining 7 were wild-type. The 1.9-to-1 ratio between these genotypes suggested that the floxed conditional allele caused homozygous lethality at preimplantation stage.

1st attempt: Zfp703 f,Neo/+ (♂x♀) 2 pairs
4 Het. lines; 4 Wt. lines

2nd attempt: Zfp703 f/+ (♂x♀) 2 pairs
9 Het. lines; 3 Wt. lines

Total: 13 Het. lines; 7 Wt. lines from 4 crosses

Zfp703 knockout first allele

As an alternative strategy I established collaboration with Dr. Brian P. Brooks in NIH who generously provided heterozygous mice carrying Zfp703 knockout first allele (Skarnes et al., 2011). The targeted allele was generated by recombineering through inserting a β-gal reporter and a Neomycin-resistant cassette between exon 1 and exon 2 of Zfp703 to disrupt the production of full-length functional protein (Figure 4.10). Based on the initial mating between 2 pairs of heterozygous mice, embryos homozygous for Zfp703 knockout first allele seemed to die during late gestation period between E12.5 and P1. With these mice, I was able to derive one Zfp703 mutant ESC line.

Zfp703 is required for efficient suppression of mesendoderm fate by RA

To test whether Zfp703 is required for RA-mediated suppression of mesendodermal fate, I cultured wild-type and Zfp703 mutant ESCs for 2 days before treating them with Gsk3i and ActivinA. Subsequently on Day3 I added RA to the culture before collecting mRNA samples 12 hours later for qRT-PCR (Figure 4.11-A). As shown in Figure 4.11-B, Zfp703 was readily induced by RA in wild-type cells; whereas in mutant cells it is barely detectable. While RA treatment led to ~2-fold repression of mesendodermal marker *Brachyury* in wild-type cells, the extent of repression was significantly attenuated in Zfp703 mutant cells (p value: 0.01; Figure 4.9-C), suggesting that Zfp703 might be required for optimal suppression of mesendoderm fate during RA-mediated neuroectodermal lineage commitment.

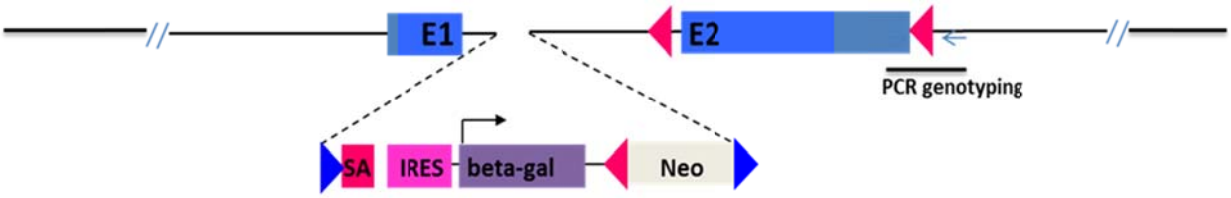
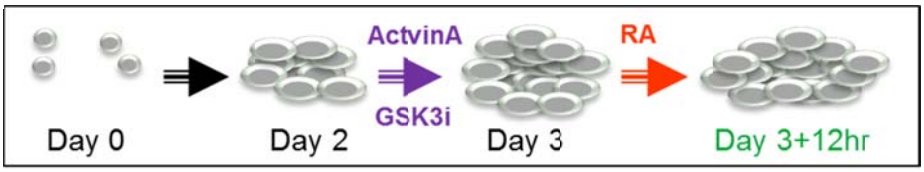


Figure 4.10

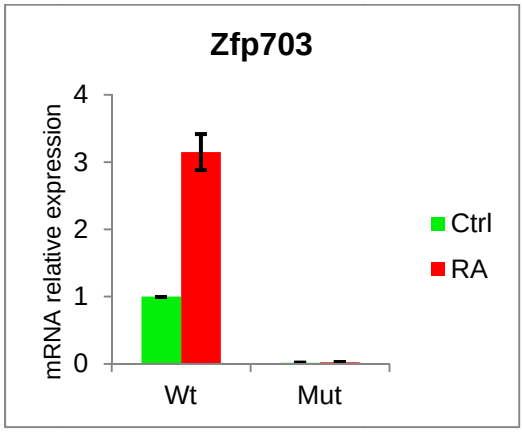
Figure 4.10 Zfp703 knockout first allele

A cassette comprising splicing acceptor (SA), IRES, beta-gal reporter, and Neomycin-resistant gene flanked by frt sites (blue triangles) was inserted between exon1 and exon2 of Zfp703, resulting in the prevention of the production of full-length Zfp703 protein therefore generating a null allele. Red triangles: loxP sites.

A



B



C

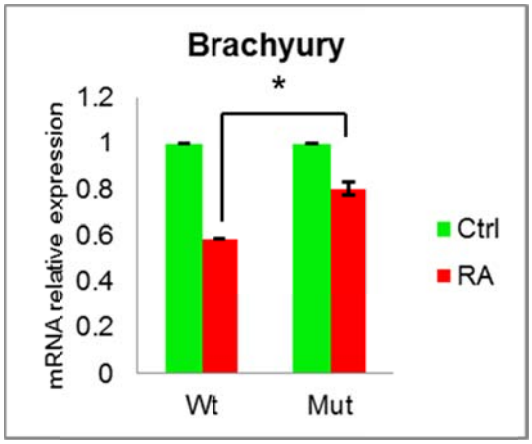


Figure 4.11

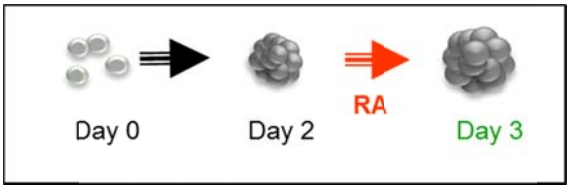
Figure 4.11 Loss of Zfp703 attenuated RA-mediated suppression of mesendoderm fate

Wild-type or Zfp703 mutant ESCs were plated as monolayer on Day0 and treated with ActivinA and GSK3i on Day2. RA was then added on Day3, and cells were collected 12 hours later for qRT-PCR. A. Schematic of ESC differentiation. B. Expression of Zfp703. While Zfp703 mRNA was induced by RA in wild-type cells, it is barely detectable in mutant cells (3 independent experiments; mRNA relative expression levels were normalized to untreated control wild-type EBs and shown as average ± standard deviation). C. Expression of mesendodermal marker *Brachyury*. RA treatment resulted in ~50% decrease in *Brachyury* mRNA compared to non-treated (Ctrl) among wild-type cells, whereas in mutant cells RA treatment led to less than 20% decrease in *Brachyury* mRNA (3 independent experiments; mRNA relative expression levels were normalized to untreated control EBs and shown as average percentage ± standard deviation; *$p<0.05$).

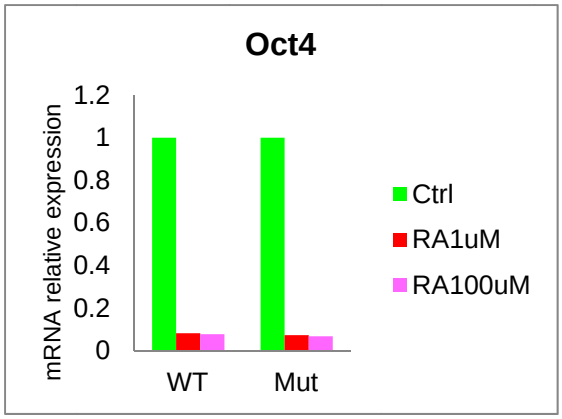
Zfp703 is not required for Oct4 repression or Sox1 induction by RA

Next I examined if Zfp703 is required for the other two events during RA-mediated neuroectodermal lineage commitment: the exit of pluripotency and the initiation of neural transcription program. I differentiated wild-type and Zfp703 mutant ESCs as EBs and treated them with RA on Day2 before performing qRT-PCR on Day3 mRNA samples (Figure 4.12-A). As shown in Figure 4.12-B and 4.12-C, no difference in RA-mediated Oct4 repression or Sox1 induction was observed between wild-type and mutant EBs, suggesting that the loss of Zfp703 might be compensated by other factors for the termination of the ESC transcriptional network and the specification of neural progenitors during neuroectodermal lineage commitment.

A



B



C

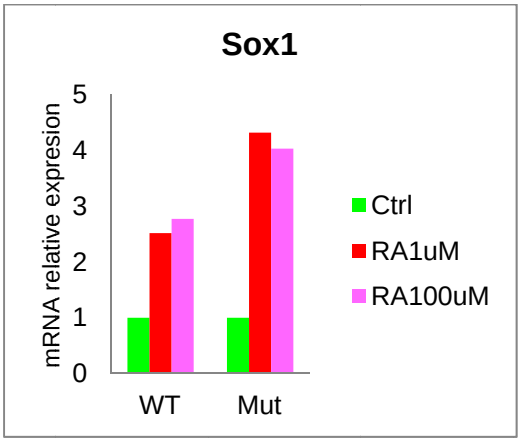


Figure 4.12

Figure 4.12 Zfp703 is not required for RA-mediated Oct4repression and Sox1 induction

Wild-type or Zfp703 mutant ESCs were differentiated as EBs for 2 days before treated with 1μM or 100nM RA. On Day3 the EBs were collected for qRT-PCR. A. Schematic of ESC differentiation. B. Expression of Oct4 after 24 hours of RA treatment. Oct4 mRNA was effectively downregulated in both wild-type and Zfp703 mutant cells by RA compared to non-treated cells (Ctrl). C. Expression of Sox1mRNA on Day3. In both wild-type and Zfp703 mutant cells RA treatment led to upregulation of Sox1 compared to non-treated cells (Ctrl) (1 experiment).

Discussion

Zfp703 and ESC fate choice during differentiation

Although Zfp703 expression in EBs caused significant change among only a small group of genes (~200, Figure 3.1), these genes are implicated in all three processes critical for neuroectodermal lineage commitment: the exit from pluripotency, the initiation of neural transcriptional program, and the suppression of non-neural fate. In previous chapters I demonstrated that Zfp703 expression is sufficient to repress Oct4 and induce neural specification. The results from experiments in this chapter suggested that Zfp703 acts also as a potent repressor of key transcription factors specifying mesendoderm. It represses *Brachyury*, *Mixl1*, and *Eomes* even under the condition when ESCs were induced to differentiate along mesendodermal lineage (Figures 4.2, 4.3 and 4.5). These results suggest that Zfp703 plays central role in the fate choice between neuroectoderm and mesendoderm of differentiating ESCs, after it first represses pluripotency genes Oct4 thereby terminating the ESC state. Because Zfp703 can repress mesendodermal determinants such as *Brachyury* without simultaneous induction of neuroectodermal factors such as Sox1, I concluded the primary function of Zfp703 is the suppression of mesendodermal fate (Figure 4.4-B). As this suppression leads to de-repression of neuroectodermal factors, Zfp703 indirectly drives the cells towards neuroectodermal lineage. There are two implications of this conclusion. First, it helps to explain the gap in neurogenesis capacity between Zfp703 expression and RA treatment, since RA could directly induce neural factors independent of Zfp703. It also suggests that Zfp703 is more important to function as a repressor of mesendodermal fate rather than an inducer of neural differentiation as a downstream effector of RA. Consistent with this notion, preliminary results from Zfp703 mutant ESCs indicate that Zfp703 is required for optimized suppression of mesendodermal determinants by

RA; while it is not required for RA-mediated Sox1 induction (Figure 4.12). Second, the de-repression nature of Zfp703-mediated neuroectodermal commitment is reminiscent of the "default" neural induction. Although Zfp703 was identified as an RA-induced factor in this study, it would be interesting to determine whether Zfp703 is also involved in "default" neural differentiation.

Zfp703 and Wnt signaling

To reveal the genome-wide binding sites and direct target genes of Zfp703, I performed ChIP-seq experiments on EBs expressing Zfp703. However, I obtained only 330 significant binding events. Specifically, very few of them were located proximal to the top 100 genes regulated by Zfp703. This could be explained by three scenarios. First, the quality of immunoprecipitation might be poor due to the indirect recruitment of Zfp703 to DNA therefore resulting in significant false-negative rate. Second, the regulation of these genes mediated through long-range distal enhancers bound by Zfp703 would not be unveiled using current analytic algorithm. Third, these genes could be secondary targets of signaling pathways regulated by Zfp703 and thus interacting with secondary transcriptional factors. The latter scenario is supported by my finding that Zfp703 binds to Lef1 promoter and represses its activity in luciferase assay (Figures 4.7 and 4.8). Lef1, a critical transcription factor mediating the activation of genes in response to Wnt signaling (Mao and Byers, 2011), was also downregulated by Zfp703 on mRNA level during differentiation (Figure 3.2). Moreover, Zfp703 has been reported to suppress the activation of the Wnt reporter (Slorach et al., 2011). Importantly, Wnt signaling is required for not only mesendodermal differentiation but also the maintenance of the ESC state (Lindsley et al., 2006; Nusse et al., 2008). Hence, one of the putative mechanisms for Zfp703 to mediate both the exit from pluripotency and the repression of mesendodermal fate

might be through attenuation of Wnt signaling pathway. Zfp703 has been shown to repress TGF-β signaling in human and mouse cells (Holland et al., 2011; Slorach et al., 2011), suggesting that Zfp703 might exert its function primarily through regulation of signaling pathways.

Phenotypes of Zfp703 mutants

Zfp703 expression is sufficient to bias differentiation of ESCs towards neuroectoderm. To address whether Zfp703 is required during normal RA-mediated neuroectodermal lineage commitment, I derived and differentiated Zfp703 mutant ESCs. The preliminary results indicated that Zfp703 is not required for RA-mediated Oct4 repression or Sox1 induction, raising the possibility that additional factors are also implicated in RA-mediated exit of pluripotency and specification of neural progenitors therefore compensating the loss of Zfp703 in the mutants. In fact, Zfp503, a close homolog of Zfp703 directly regulated by RA (Table 2.1) (Chang et al., 2004), is also sufficient to repress Oct4 when overexpressed (Figure 2.3-D) although further analyses of Zfp503 was complicated by the fact that it induces cell death (Figure 2.5). Because Zfp703 and Zfp503 can act as homo- and heterodimers (Runko and Sagerstrom, 2004), they might function cooperatively to regulate the ESC transcriptional network and the neural transcription program. Hence the more profound phenotypes of neuroectodermal lineage commitment *in vivo* might only be revealed when both Zfp703 and Zfp503 are targeted. Nevertheless, Zfp703 is required for full repression of *Brachyury* in RA-treated early mesendodermal cells (Figure 4.11).

Homozygous lethality of Zfp703 conditional allele

The death of Zfp703 mutant embryos homozygous for the knockout first allele between E12.5 to around birth suggested that Zfp703 is an essential gene critical for the normal development during late gestation stage (Figure 4.10). However, it cannot explain the

preimplantation lethality of embryos homozygous for the conditional allele I generated. The most likely explanation is that the insertion of loxP sites disrupted a yet unknown essential regulatory element (i.e. enhancers or non-coding RNAs), causing the observed lethality. In fact, the region proximal to the inserted loxP site upstream of Zfp703 exon2 is highly conserved among mammals (Figure 4.13) and might be important for preimplantation development.


```
Mouse       CGCCATCACCAGAAACCCGTCCCAAACGAAGTTCTCCCAAACTAAAGGCTTAAGTTT-TT-CCCTGGCCCAAAA-AGATGGGAAAATTATCTGGAAA---G
Rat         CGCCATCACCAGAAACCCGTCCCAAACGAAGTTCTCCCAAACTAAAGGCTTAAGTTT-TT-CCCTGGGCCAAAA-
Rabbit      CGCCATCGCCAGAAACCCGTCCCAAACGAATTTCCTCCAAACTAGAGAAATAAGTTT-CT-CCCCGGGCCTGGA-AGAAGGAGAAATTGGAGCGAA-CGAG
Marmoset    CGCTGTCGCCAGAAACCCGTCCCAAACGAAGTTCCCCCAAACTAACAGAATAAGTTT-TT-TCCTGCGCCTAGA-GGAGGGGGACATTAGAGAGAA----G
Gorilla     CGCCATCGCCAGAAACCCGTCCCAAGCGAAGTTCCCCCAAACTAAAGGAATAAGTTT-TT-TCCTGGGCCTGGA-GGAGGGGGACATTAGAGGGAA----G
Human       CGCCATCGCCAGAAACCCGTCCCAAGCGAAGTTCCCCCAAACTAAAGGAATAAGTTT-TT-TCCTGGGCCTGGA-GGAGGGGGACATTAGAGGGAA----G
Chimpanzee  CGCCATCGCCAGAAACCCGTCCCAAGCGAAGTTCCCCCAAACTAAAGGAATAAGTTT-TT-TCCTGGGCCTGGA-GGAGGGGGACATTAGAGGGAA----G
Orangutan   CGCCATCGCCAGAAACCCGTCCCAAGCGAAGTTCCCCCAAACTAAAGGAATAAGTTT-TT-TCCTGGGCCTGGA-GAAGGGGGACATTAGAGGGAA----G
Macaque     CGCCATTGCCAGAAACCCGTCCCAAGCGAAGTTGCCCCAAACTAAAGGAATAAGTTTTTT-TCCTGGGCCTGGA-AGAGGGGGACATTAGAGGGAA----G
Cow         CGTCACTGCCAGAAACCCGTCCTAAGCGAAGTTCC-CCAAACTGAAGGCATAAGTTT-TTAACCCGGGCCTAGA--GAGGACGAGATGAGAGAGTA----G
Pig         CGCCACCGCCTGAAACCCGTCCTAAGCGAAGTTAC-CCAAACTGAAGCAATAAGTTT-TT-CCCTGGGCATAGAC-TAGGGGGAGATTCGAGGGCA----G
Dog         CGCCG-CGCCAGAAACCCGTCCCA-GCGAAGTNNN-NNNNNN----------
Horse       CGCCACAGCCAGAAACCCGTCCCAAGCGAAGTTCC-CCAAACTGAAGGAGTAAGTTT-TT-CCCTGGACATAGG--GAAG----------
```

Figure 4.13

Figure 4.13 The region around one loxP site is highly conserved among mammals

Alignment of genomic sequences at the position of the first loxP site upstream of exon2 on the conditional targeting allele. The sequence disrupted by the loxP sites (red letters) was highly conserved across 13 mammalian species.

Chapter 5. Conclusion and future directions

Conclusion

I began this study by identifying downstream effectors of RA-mediated neuroectodermal lineage commitment based on the observation that critical transcription factors involved in three discrete cellular statuses during ESC lineage commitment: the pluripotent ESC state (Oct4, Nanog, and Sox2), the committed neuroectodermal (Sox1 and Nr2f2) precursors and mesendodermal progenitors (Brachyury, Eomes, and Mixl1) are not direct targets of RA. Through a screening with transcription factors induced rapidly by RA using an inducible ESC system, I found that Zfp703, a member of NET family zinc finger proteins, is sufficient to repress Oct4 when expressed alone. This suggests that Zfp703 might be a downstream effector of RA which represses pluripotency genes and thus leads to the exit of the ESC state, the perquisite of lineage specification during differentiation. Subsequent microarray profiling and qPCR analyses indicated that Zfp703 not only affects plutipotency genes but also upregulates genes specifying neuroectodermal fate and downregulate mesendodermal markers. Despite Zfp703 expression induces both neural progenitor marker Sox1 and some epidermal genes like keratins early in the differentiation, Sox1-expressing neural progenitors and NeuN/Tuj1-positive postmitotic neurons dominated the later culture. Hence Zfp703 promotes neuronal identity. On the other hand, Zfp703 suppresses mesendodermal differentiation even under conditions favoring mesendodermal fate. Because Zfp703 strongly represses mesendodermal determinants such as *Brachyury* without concomitant induction of neuroectodermal genes such as Sox1, I concluded that Zfp703 primarily represses mesendodermal fate and by doing so de-represses neuroectodermal genes therefore promoting neurogenesis. Therefore, Zfp703 seems to be

implicated in the lineage choice between neuroectoderm and mesendoderm, the first crucial fate decision toward making diversified cell types during embryogenesis.

In order to illuminate how Zfp703 regulates gene expression, I conducted ChIP-seq experiments to characterize the genome-wide binding sites of Zfp703. The results revealed very few high-confidence binding sites in the genome that are far from genes highly-regulated by Zfp703. This might due to the fact that Zfp703 does not bind DNA directly (Nakamura et al., 2004). Nevertheless, Lef1 was one of the few Zfp703-regulated genes with Zfp703 enriched site in the proximity of its transcription start site (~7Kb upstream). I confirmed that Zfp703 binds to the promoter of Lef1 and represses its activity. Based on these results, I propose that Zfp703 might modulate Wnt signaling during ESC differentiation through repressing Lef1. Given the implication of Wnt signaling and Lef1 in both the maintenance of pluripotency and the specification of mesendodermal fate (Lindsley et al., 2006; Nusse et al., 2008), this could be a plausible mechanism of Zfp703 in promoting neuroectodermal lineage commitment. On the other hand, I cannot rule out the possibility that Zfp703 directly represses key mesendoderm determinants (such as *Brachyury, Eomes,* and *Mixl1*) and Oct4. While I failed to detect Zfp703 binding sites in the proximity of these genes, the low signals of Zfp703 ChIP-seq suggests that I might miss many weaker sites, the genes are regulated by distal enhancers, or that secondary repressors might be involved in the repression of these genes.

Lastly, I performed preliminary Zfp703 loss-of-function analyses using ESCs carrying the knockout first allele. While RA-mediated repression of *Brachyury* under mesendoderm-inducing condition was attenuated in Zfp703 mutant ESCs, the repression of Oct4 and the upregulation of Sox1 by RA are unaffected in Zfp703 mutant EBs. Interestingly, Zfp503, a member in the NET protein family closely related to Zfp703 and directly regulated by RA signal (Chang et al., 2004),

was also sufficient to repress Oct4 when expressed in EBs. This information suggests that Zfp703 might be required for optimized suppression of mesendodermal fate in RA-mediated neuroectodermal lineage commitment, and other factors could also contribute to the termination of the ESC transcriptional network and the initiation of neural differentiation program downstream of RA. Consistent with this notion, neurogenesis resulting from Zfp703 expression is not as robust as RA treatment.

In conclusion, through this study I defined one of the putative important transcriptional cascades downstream of RA in neuroectodermal lineage commitment (Figure 5.1). Specifically, RA directly activates Zfp703, which represses pluripotency gene *Oct4* therefore promotes the exit of the ESC state. As cells are now poised to commit to two alternative fates: neuroectoderm and mesendoderm, Zfp703 further represses mesendodermal determinants *Brachyury*, *Eomes*, and *Mixl1* thus removing their repression of neuroectodermal genes such as *Sox1* and *Nr2f2*. This de-repression of key transcription factors eventually leads to the emergence of neuronal identity. There are several caveats about this model. First, it might not be the only mechanism of RA-mediated neuroectodermal lineage commitment, since Zfp703 cannot fully recapitulate the effects of RA. Second, I cannot conclude whether Zfp703 repress *Oct4* and *Brachyury* directly through binding to their regulatory elements in the genome or indirectly through secondary repressors due to the ChIP-seq results. Third, because the relative ratio of Oct4 to Sox2 is an important factor affecting the choice between neuroectodermal vs. mesendodermal lineages (Thomson et al., 2011), downregulation of Oct4 by Zfp703 would result in high Sox2/Oct4 ratio, which might also contribute to Zfp703-mediated neurogenesis. Interestingly, although RA is not involved in neural induction *in vivo*, this model bears striking resemblance to the "default" neural induction, the prevailing neutralizing mechanism in mammalian embryogenesis. RA-

mediated and “default” neural induction might converge upon Zfp703. And it would be important to determine the role of Zfp703 in the “default” neural induction.

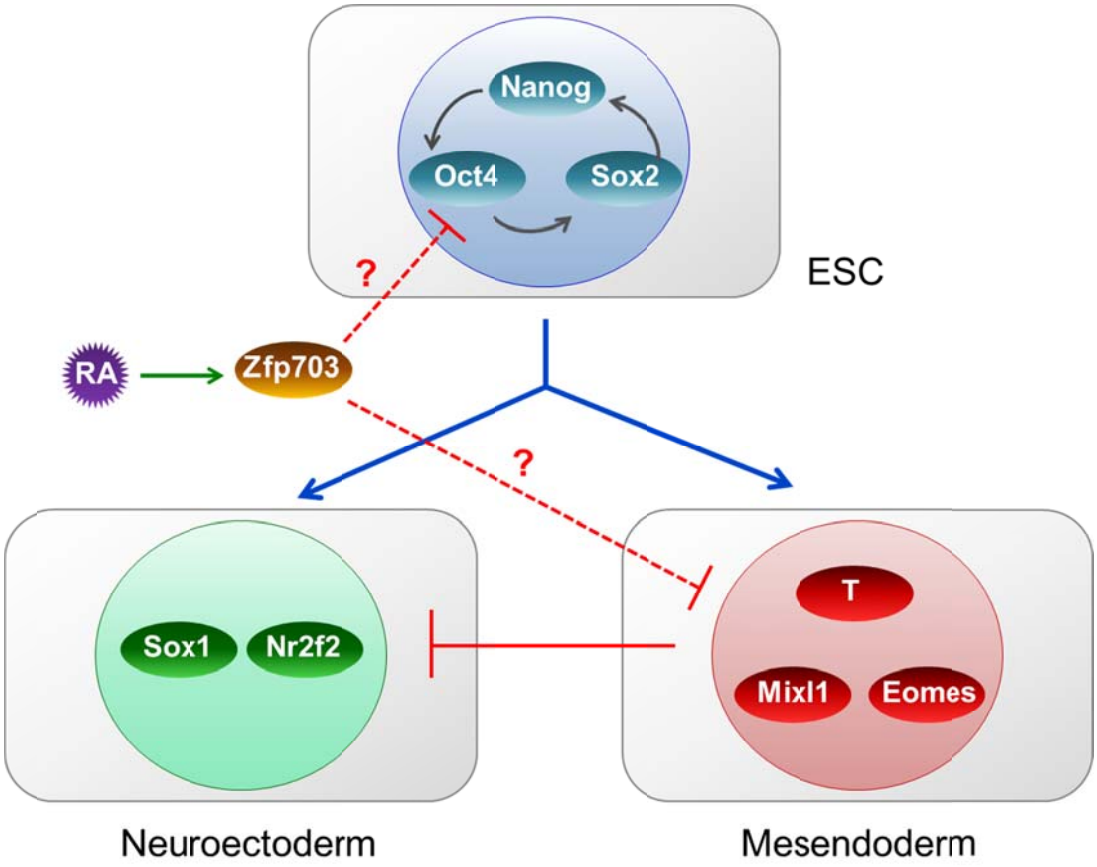

Nanog
Oct4
Sox2
ESC
?
RA
Zfp703
?
Sox1
Nr2f2
T
Mixl1
Eomes
Neuroectoderm
Mesendoderm

Figure 5.1

Figure 5.1 Model of RA-mediated neuroectodermal lineage commitment

RA directly induces Zfp703, which represses pluripotency gene *Oct4* and mesendodermal determinants *Brachyury*, *Eomes*, and *Mixl1*. By doing so Zfp703 removes the repression on neuroectodermal genes such as *Sox1* and *Nr2f2*. Subsequently, the de-repression of these transcription factors leads to the emergence of neuroectodermal fate.

Discussion

Cell state transition during lineage commitment

There are three distinct cell states involved in the commitment of ESCs to neuroectodermal or mesendodermal lineages, a process faithfully recapitulating the first step toward making diverse cell types in the embryo at the onset of gastrulation in developing epiblast. These cell states are the pluripotent ESC state maintained by the ESC transcriptional program composed of the core transcription factors Oct4, Sox2 and Nanog; the neuroectodermal progenitors specified by transcription factors Sox1 and Nr2f2; as well as the mesendodermal precursors established by transcription factors Brachyury, Eomes, and Mixl1. The transition between cell states during lineage commitment is driven by extracellular signals. For example, RA and FGF can promote neuroectodermal commitment, whereas Wnt and Activin induce mesendodermal identity (Thomson et al., 2011). These signaling pathways affect cell states through regulation of gene expression which leads to the termination or activation of specific transcriptional programs that subsequently consolidate cell fate. To ensure coordinated regulation of these transcriptional programs and to prevent cells from acquiring mixed identities, two general themes can be observed in the mechanisms of these signaling pathways in lineage commitment. First, there are multiple downstream effectors regulating the same set of genes in transcriptional programs specifying cell states. This redundancy will ensure robust and thorough regulation of key transcription factors. At the same time, it would also serve as fail-safe cushion that if one effector fails, the cell state transition can still proceed. Second, there are downstream effectors able to affect multiple transcriptional programs involved in lineage commitment. This is a convenient way for the signaling pathways to exert integrated effects during cell state transition with a few effectors as the hubs of transcriptional regulation.

Both themes are well-illustrated in RA-mediated neuroectodermal lineage commitment. Successful neuroectoderm specification requires the transition of the cells from pluripotent state to committed neuroectodermal identity. On the level of transcriptional regulation, RA must terminate the ESC transcriptional program, activate neuroectodermal factors and repress mesendodermal determinants to facilitate this transition. In the termination of ESC transcriptional program, Hoxa1, Coup-TFs, and GCNF have all been proposed to repress key pluripotency gene Oct4 and/or Nanog downstream of RA (Gu et al., 2005; Martinez-Ceballos et al., 2005; Rosa and Brivanlou, 2011; Schoorlemmer et al., 1994). I also identify two potent Oct4 repressors: Zfp703 and Zfp503 in this study. Therefore RA might terminate the ESC transcriptional program through multiple downstream effectors to achieve robust and complete exit of the pluripotent state. It is also possible that RA uses different combination of effectors at different time points throughout neuroectodermal commitment to ensure sustained repression of the ESC transcriptional program. Consistent with this notion, these effectors are induced at various time points after RA treatment (Figure 2.2). Among them, Zfp703 seems to play critical role in RA-mediated Oct4 repression and the subsequent exit of pluripotent state. First, it is rapidly and highly induced in response to RA signal (within 8 hours, Table 2.1). Second, it represses Oct4 to the extent comparable to RA (Figure 2.3-D). Finally, it upregulates Nr2f2 (Coup-TFII, Figure 3.2), another downstream effector involved in Oct4 repression. Therefore Zfp703 might be the effector that initiates the termination of the ESC transcriptional program by repressing Oct4 in response to RA. Meanwhile, it also facilitates sustained repression of pluripotency genes via inducing Nr2f2 thus constituting a feed forward loop leading to efficient exit of the ESC state. It should be noted, however, that Zfp703 is not the only effector in the termination of the ESC transcriptional program, as it proceeds normally in Zfp703 mutant ESCs

treated with RA (Figure 4.12-B), possibly due to the effects of Zfp503. This is consistent with the redundant nature of the regulation.

Regarding the second theme, Zfp703 also seems to be an ideal hub to integrate the regulation of multiple transcriptional programs during neuroectodermal lineage commitment. In addition to repressing Oct4, Zfp703 strongly represses mesendodermal determinants Brachyury, Eomes, and Mixl1. By repressing both Oct4 and mesendodermal fate, Zfp703 de-represses neuroectodermal factors Sox1 and Nr2f2 thus promoting neuroectodermal identity. Hence, RA is capable of coordinately regulating the ESC, neuroectoderm, and mesendoderm transcriptional programs through Zfp703 to facilitate the transition from ESC to neuroectodermal progenitors. This puts Zfp703 at the center of transcriptional regulation downstream of RA (Figure 5.1).

How can Zfp703 regulate such diverse genes given the fact that it might not bind DNA by itself (Nakamura et al., 2004)? One possible mechanism is that it modulates the signaling pathways controlling those genes. Lef1 is an important downstream effector of Wnt signaling which is involved in the maintenance of pluripotency and mesendoderm specification (Lindsley et al., 2006; Nusse et al., 2008). As Zfp703 binds to the promoter of Lef1 and represses its activity (Figure 4.7 and 4.8), it might regulate both the ESC and mesendoderm transcriptional program through modulating Wnt signaling.

A second putative mechanism is that Zfp703 could act as a cell type specific cofactor of transcription factors that modulates their activity and in turn regulates diverse target genes in a context dependent manner. One group of potential Zfp703 partners includes Sox proteins because their binding motif is enriched in Zfp703 biding sites in ChIP-seq experiments (Figure 4.6-B). Among them, Sox2 is a particularly intriguing candidate. It is expressed in ESCs and is a

member of the ESC transcriptional program maintaining pluripotency together with Oct4 and Nanog. Unlike the latter two factors whose expression is rapidly exhausted following the onset of neural differentiation, Sox2 stays present through neuroectodermal and neural progenitor stages (Collignon et al., 1996). In fact, it is required for the maintenance of neural progenitors (Wegner, 2011). ChIP-seq experiments revealed that Sox2 binds to the vicinity of both ESC-specific genes and a large number of neural genes (Bergsland et al., 2011). Therefore, while activating ESC genes to sustain pluripotency, Sox2 also establishes transcriptional competence for neural development in ESCs, as evidenced by the association with bivalent histone marks (H3K4me3 and H3K27me3) at the Sox2 binding sites in the proximity of neural genes. Subsequently, Sox3 is recruited to these sites to activate the neural differentiation program (Bergsland et al., 2011). Differential regulation of gene expression by Sox2 can be explained by the differential expression of a cell type specific cofactor. In pluripotent cells ESC-specific genes are co-bound by Sox2, Oct4, and Nanog thus providing a convenient way to sort out genes to be activated. When neural differentiation proceeds, induced Zfp703 might bind to Sox2 and repress its activity at Sox2-bound ESC genes. Induction of Sox1 and Sox3 then might result in activation of neural genes to further consolidate neural identity later in differentiation. To comprehensively assess the function of Zfp703 in the cell state transition during neuroectodermal lineage commitment, it would be important to determine whether Zfp703 partner with Sox2 or other proteins to regulate gene expression and to determine why Zfp703 is recruited only to a small subset of Sox2 bound sites in ESCs.

Zfp703 and neural induction

In mammalian embryonic development, neuroectoderm specification and the subsequent induction of neural identity follow the “default” mechanism. At the onset of gastrulation a subset

of the apparently homogeneous pluripotent epiblast cells form primitive streak (PS) in response to the combination of Nodal, Wnt, and BMP signaling (Rossant and Tam, 2004). Then groups of epiblast cells move through the PS and emerge as mesendoderm under the control of the same signals (Marikawa, 2006; Tam and Loebel, 2007). Epiblast cells that do not receive Nodal, Wnt, and BMP signaling will not be recruited to the PS and become neuroectodermal progenitors. Therefore, the "default" fate of epiblast cells is neuroectodermal commitment unless they receive these signals. Among neuroectodermal progenitors, those are exposed to Bmps and Wnts will become epidermis while cells do not receive these signals acquire neural identity. Again, the "default" fate of neuroectodermal progenitors is neural lineage unless they are exposed to Bmps and Wnts. This "default" model of neural induction suggests that the inhibition of these signaling pathways can effectively drive pluripotent cells towards neural lineage, which can be achieved *in vivo* through the secretion of Bmp/Wnt antagonists by the cells in the node or *in vitro* via Nodal and Bmp inhibition on ESCs (Chambers et al., 2009; Hemmati-Brivanlou et al., 1994; Sasai et al., 1994; Smith and Harland, 1992).

Although treatment of ESCs with RA can lead to efficient neurogenesis (Bain and Gottlieb, 1994; Gottlieb and Huettner, 1999), there is no evidence that RA is required for neural induction *in vivo*. Here, Zfp703 might provide a link between RA and the "default" model of neural induction. Zfp703 is a downstream effector of RA and sufficient to promote neuroectodermal identity when expressed in the absence of the signal (Figure 3.4). Importantly, Zfp703 likely achieves this through repressing pluripotency and mesendodermal fate therefore de-repressing the neuroectoderm specification (Figure 5.1). The de-repression mode of Zfp703-mediated neuroectodermal commitment bears striking resemblance to the "default" model of neural induction. Interestingly, one plausible mechanism underpinning the role Zfp703 in the fate

choice of neuroectoderm over mesendoderm is through modulating Wnt signaling because Zfp703 binds to and represses Lef-1, a critical Wnt signal transducer (Figure 3.2, Figure 4.7 & 4.8). As Wnt inhibition is part of the "default" neural induction, Zfp703 might drive neurogenesis following the similar route. It will be interesting to investigate whether Zfp703 is induced in an RA-independent manner to assist during the "default" neural induction.

Despite the apparent similarities between Zfp703-mediated neurogenesis and the "default" model, disparities exist. Specifically, there is an initial emergence of cells with epidermal markers in EBs expressing Zfp703 (Figure 3.5) which should not occur if the cells follow the "default" mechanism strictly. As population of cells expressing epidermal genes decreases and neural progenitors dominate the later culture (Figure 3.6), this suggests that Zfp703 alone cannot fully recapitulate the "default" neurogenesis. Other mechanisms might contribute to effective repression of epidermal identity during neural induction. Zfp521, another zinc-finger nuclear protein, might be involved in this process. It has been shown to be essential and sufficient to drive "default" neural differentiation of mouse ESCs (Kamiya et al., 2011). Zfp521 is intrinsically induced in differentiating ESCs and cooperates with the coactivator p300 to activate key neuroectoderm-specific genes Sox1, Sox3 and Pax6 which in turn consolidate neural identity. Interestingly, Zfp703 induction following RA treatment precedes the induction of Zfp521, indicating that it might act downstream in the RA triggered neurogenic transcriptional cascade. The interplay between Zfp703, Zfp521 and other intrinsic factors during neural differentiation should be further investigated to determine the implication of Zfp703 in the "default" neural induction. On the other hand, while Zfp703 is highly induced by RA, its expression is upregulated in untreated EB cultures albeit to a lesser extent (data not shown). The mechanisms

of RA-independent Zfp703 induction following the "default" neural induction remain to be investigated.

It should be noted that Zfp703 is not the only downstream effector mediating neurogenesis initiated by RA. First, the capacity of Zfp703 to induce Sox1-positive neural progenitors and subsequently NeuN/Tuj1-positive postmitotic neurons is lagging that of RA, indicating that additional downstream effectors of RA contribute to its neuralizing effects. Moreover, Sox1 induction by RA treatment is unaffected in Zfp703 mutant ESCs (Figure 4.12-C). And Hoxa1, GCNF, and Nr2f2 have all been proposed to be required for optimized neural differentiation in response to RA (Akamatsu et al., 2009; Martinez-Ceballos and Gudas, 2008; Rosa and Brivanlou, 2011). Hence, Hoxa1, GCNF, and Nr2f2 might be implicated in Zfp703-independent mechanisms downstream of RA contributing to the robust neurogenesis. Interestingly, Zfp703 expression leads to the upregulation of Nr2f2, suggesting the potential crosstalk and integration of transcriptional networks driving neurogenesis. Lastly, the accumulation of Sox1-positive neural progenitors in later cultures of Zfp703-expressing cells when the majority of RA-induced neural progenitors have already become postmitotic (Figure 3.6-B) indicates that additional Zfp703-independent factors promote the terminal differentiation of neural progenitors. In conclusion, while part of the neuralizing effects of RA can be attributed to the mechanism similar to the "default" model that is putatively mediated by Zfp703, RA-mediated neural induction still retains its unique features that are quite distinct from the prevailing *in vivo* model.

Gene regulation through long-range chromatin interactions

Traditionally, the standard way of assigning putative target genes to transcription factors has been based on the proximity (typically within 10-20Kbp upstream and downstream) of transcription factor binding site to a transcription start site (TSS). This approach has also been

adopted to identify genes that are direct targets of RA signaling during ESC differentiation through monitoring RAR binding (Mahony et al., 2011). However, applying proximal binding as the only criteria for direct targets might not provide the "whole picture" of regulation exerted by one transcription factor. For example, Zfp703 is rapidly induced following RA treatment (within 8 hours, Table 2.1). The fast response, similar to that of the canonical RA target genes such as Hoxa1 (Figure 2.1-C), suggests direct regulation by RA signal, but there is no enrichment of RAR binding within the 20Kbp window of Zfp703 TSS (Mahony et al., 2011). Recently, long-range chromatin interactions via enhancers have been demonstrated as an important mechanism of gene regulation (Harmston and Lenhard, 2013). The development of techniques such as ChIA-PET provides the access to genome-wide, long-range chromatin interactions (Li et al., 2010). When combined with ChIP-seq for enhancer signatures and conservation analysis, these techniques can offer a more comprehensive scope of direct targets regulated by specific transcription factors. Indeed, ChIA-PET data indicate that there are two distal RAR binding sites interacting with Zfp703 TSS following RA treatment (Figure 2.6). These sites colocalize with enhancer signatures H3K27ac, p300, and H3K4me1 (Figure 2.7), and they are conserved in mammals (Figure 2.8). Together, these observations strongly suggest that Zfp703 is directly regulated by RA through RAR binding to distal enhancers. Although mutagenesis of these elements is still necessary to verify that they are required for RA-mediated Zfp703 induction, this finding provides the first evidence for an involvement of distal regulatory elements engaged by RA signaling. It would be interesting to identify other RA targets regulated in the same manner. In contrast, Chip-seq experiments revealed very few interactions between genomic regions bound by Zfp703 and transcription start sites supporting the notion that Zfp703 might

function as a transcriptional repressor responsible for disruption, rather than establishment of long distance chromatin interactions.

Future directions

Characterization of Zfp703-interacting proteins

Given that Zfp703 expression leads to both activation and repression of genes, and the fact that Zfp703 might not bind DNA by itself (Nakamura et al., 2004), it will be critical to identify Zfp703-interacting proteins throughout RA-mediated neuroectodermal lineage commitment by proteomics and co-immunoprecipitation approaches. The results will not only address whether Zfp703 functions as both repressors and activators in a context-dependent manner, but also explain how Zfp703 achieve its sequence specificity when regulating its target genes. Finally, it will also clarify whether Zfp703 and Zfp503 act as a dimer in promoting neural fate.

Further investigating the effect of Zfp703 on Wnt signaling

Despite that Zfp703 has been reported to suppress the activation of the Wnt reporter in human breast cancer cell lines (Slorach et al., 2011), and it represses Lef1 likely through binging to its promoter, it remains to be confirmed whether the identified binding site is required for Zfp703-mediated regulation of Lef1. Luciferase assay using Lef1 promoter in which this site is mutated would help to clarify this issue. Moreover, it is not clear how Zfp703 affect Wnt signaling during neuroectodermal lineage commitment. Experiments using Wnt agonists /antagonists combined with Zfp703 expression or knockout and carefully monitoring the activity of Wnt signaling during neuroectodermal lineage commitment will be necessary to determine whether modulation of Wnt signaling is the major mechanism underlying the function of Zfp703.

Zfp703 ChIP-seq and enhancers

While Zfp703 is able to bind to the promoter of Lef1, a significant portion of binding sites identified in ChIP-seq experiments actually overlapped with enhancer signatures such as

H3K4me1, H3K27Ac, and p300 (Smith and Shilatifard, 2014). It would be interesting to verify the binding by ChIP-PCR, examine the effect of Zfp703 on these enhancers by luciferase assay, and in the long term knock out these enhancers to compare the phenotype with Zfp703 mutant in ESCs. It would also be important to identify transcription factors co-occupying these sites with Zfp703 through motif analysis and validate them by proteomics experiments. Interestingly, the motif present in 163 out of 330 high-confidence Zfp703 binding sites is potentially bound by Sox2, Sox3, and Sox6. Given the dual role of Sox2 in both the maintenance of pluripotency and the promotion of neuronal fate, it could be an attractive Zfp703 co-factor. It will be important to test the effect of Zfp703 on enhancers in the presence or absence of Sox2 by luciferase assay. Together, these experiments will offer a more comprehensive view of Zfp703-mediated gene regulation.

Zfp703 and Zfp503 double-knockout

Because both Zfp703 and Zfp503 are able to repress Oct4 and induce Sox1 when overexpressed, and Zfp703 mutant ESCs behave normally in RA-induced Oct4 downregulation and Sox1upregulation, it is likely that Zfp503 is compensating the function of Zfp703, and both of them are required for RA-mediated neuroectodermal lineage commitment. To test this, it would be necessary to generate Zfp703 and Zfp503 double-knockout ESCs and mice to characterize the phenotypes *in vitro* as well as *in vivo*.

Material and methods

Cell culture and neural differentiation

ESCs were differentiated as previously described (Wichterle et al., 2002). Briefly, ESCs were trypsinized and seeded at 5×10^5 cells/ml in ANDFK medium (Advanced DMEM/F12:Neurobasal (1:1) medium, 10% knockout-SR, Pen/Strep, 2 mM L-glutamine, and 0.1 mM 2-mercaptoethanol) to initiate formation of embroid bodies (Day 0). Medium was exchanged on day 2 of differentiation. Patterning of embroid bodies was induced by supplementing media on day 2 with 1 μM RA (Sigma). For ChIP experiments, the same conditions were used but scaled to seed 1×10^7 cells on Day 0.

Generation of inducible lines

The p2Lox-V5 plasmid was generated by replacing GFP with the L1-L2 Gateway cassette from pDEST-40 (Invitrogen) in the p2Lox plasmid. The cassette contains a V5-His double epitope tag in frame downstream of the L2 recombination site. Open reading frames of genes are cloned by polymerase chain reaction (PCR). To minimize the introduction of mutations during PCR amplification, Phusion polymerase was used (New England Biolabs). Open reading frames were directionally inserted into pENTR/D-TOPO vector (Invitrogen) following manufacturer instructions. The 5' primer always contains the addition of the CACC sequence to ensure directional integration. Then the pENTR plasmid with NO STOP codon is recombined with the p2Lox-V5 to construct a V5-His C-terminal fusion protein in LR recombination. Inducible lines were generated by treating the recipient ESCs for 16 hours with doxycycline to induce Cre followed by electroporation of p2Lox-V5 plasmids harboring the desired construct. After G418 selection, on average three resistant clones were picked, characterized and expanded.

Mesendoderm differentiation

ESCs were plated on gelatine-coated dishes and cultured with ES medium (EmbryoMax D-MEM (Fisher) supplemented with 10% ES-FBS (Invitrogen), L-Glutamine (Gibco), 0.1 mM β-mercaptoethanol and 100 U/ml LIF) with initial density of 1.5×10^4 cells/cm^2 (Day -1). The next day (Day0) the medium was exchanged to D0 differentiation media (Advanced DMEM (Invitrogen) + 1% N-2 (1 vial, Invitrogen) + 2% B-27 (1 vial, Invitrogen) + 1% Glutamax (Invitrogen)). 48 hours later on Day2, the medium was refreshed with D0 differentiation media supplemented with 50 ng/mL ActivinA (Peprotech E Coli produced) and 5 nM GSK3 inhibitor XV (Calbiochem). Cells were harvest on Day3 for qPCR and immunocytochemistry analysis.

Immunocytochemistry

Antibodies used in this study include: Goat anti-Oct4 (Santa Cruz, sc-8628), Rabbit anti-Sox1 (Cell Signaing, #4194), Mouse anti-V5 (Invitrogen, R960-25), Rabbit anti-V5 (Abcam, ab9116), Goat anti-Brachyury (Santa Cruz, sc-17745), Rabbit anti-Krt1, Rabbit anti-Krt5, Rat anti-Krt8, Mouse anti-Krt18, Mouse anti-NeuN and mouse anti-Tuj1. Alexa488-, FITC-, Cy3- and Cy5-conjugated secondary antibodies were obtained from either Invitrogen or Jackson Immunoresearch.

Expression analysis

Total RNA was extracted from ES cells or embryoid bodies using Qiagen RNAeasy kit (Qiagen). For quantitative PCR analysis, cDNA was synthesized using SuperScript III (Invitrogen) and amplified using SYBR green brilliant PCR amplification kit (Stratagene) and Mx3000 thermocycler (Stratagene). For GeneChip expression analysis, RNA was amplified using Ovation amplification and labeling kit (NuGen) and hybridized to Affymetrix Mouse Genome 430 2.0

microarrays. Expression microarray experiments were performed in biological quadruplicate for each analyzed time point. Arrays were scanned using the GeneChip Scanner 3000. Data analysis was carried out using GeneSpring GX (Agilent Technologies). Differentially expressed genes after 24 hours of doxycycline treatment were defined by ranking all probesets by the moderated t-statistic-derived P-value and setting thresholds of $P < 0.05$ and a fold-change of at least 2.

ChIP-seq protocols

ChIP protocols were adapted from previously published method (Guenther et al., 2008). Briefly, approximately 6×10^7 cells were cross-linked using formaldehyde and snap-frozen in liquid nitrogen. Cells were thawed on ice, resuspended in 5 ml lysis buffer 1 (50 mM Hepes-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) and mixed on a rotating platform at 4°C for 5 minutes. Samples were spun down for 3 minutes at 3,000 rpm, resuspended in 5 ml lysis buffer 2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA), and mixed on a rotating platform for 5 minutes at room temperature. Samples were spun down once more, resuspended in lysis buffer 3 (10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine) and sonicated using a Misonix 3000 model sonicator to sheer cross-linked DNA to an average fragment size of approximately 500 bp. Triton X-100 was added to the lysate after sonication to final concentrations of 1% and the lysate spun down to pellet cell debris. The resulting whole-cell extract supernatant was incubated on a rotating mixer overnight at 4°C with 100 µl of Dynal Protein G magnetic beads that had been preincubated for 24 hours with 10 µg of the appropriate antibody in a phosphate-buffered saline/bovine serum albumin solution. Rabbit anti-V5 (Abcam, ab9116) was used for ChIP experiments. After approximately 16 hours of bead-lysate incubation,

beads were collected with a Dynal magnet. ChIP samples were washed with the following regimen, mixing on a rotating mixer at 4°C for 5 minutes per buffer: low-salt buffer (20 mM Tris at pH 8.1, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), high-salt buffer (20 mM Tris at pH 8.1, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), LiCl buffer (10 mM Tris at pH 8.1, 250 mM LiCl, 1 mM EDTA, 1% deoxycholate, 1% NP-40), and TE containing 50 mM NaCl. After the final bead wash, samples were spun down to collect and discard excess wash solution, and bound antibody-protein-DNA fragment complexes were eluted from the beads by incubation in elution buffer at 65°C with occasional vortexing. Cross-links were reversed by overnight incubation at 65°C. Samples were digested with RNase A and Proteinase K to remove proteins and contaminating nucleic acids, and the DNA fragments precipitated with cold ethanol.

Luciferase Reporter Assay

Genomic DNA fragments of ~1Kb encompassing putative Zfp703 binding sites were cloned into pGL4.3 vector (Promega). HEK293 cells were plated at 1×10^5 per well (24-well plate), expanded for 16 hr and cotransfected with a mixture of 500ng pDEST41-Zfp703, 450ng pGL4.3 construct, and 50ng pGL4.74[*hRluc/TK*] plasmids using 1 μl Lipofectamine 2000 (Invitrogen). Cells were lysed 24 hr later and processed for luciferase assay using Dual-Luciferase Reporter Assay System (Promega). Luciferase activity was measured by 20/20n luminometer (Turner Biosystems).

Bibliography

Acloque, H., Ocana, O.H., and Nieto, M.A. (2012). Mutual exclusion of transcription factors and cell behaviour in the definition of vertebrate embryonic territories. Current opinion in genetics & development *22*, 308-314.

Akamatsu, W., DeVeale, B., Okano, H., Cooney, A.J., and van der Kooy, D. (2009). Suppression of Oct4 by germ cell nuclear factor restricts pluripotency and promotes neural stem cell development in the early neural lineage. The Journal of neuroscience : the official journal of the Society for Neuroscience *29*, 2113-2124.

Arnold, S.J., Hofmann, U.K., Bikoff, E.K., and Robertson, E.J. (2008). Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. Development (Cambridge, England) *135*, 501-511.

Athanikar, J.N., Sanchez, H.B., and Osborne, T.F. (1997). Promoter selective transcriptional synergy mediated by sterol regulatory element binding protein and Sp1: a critical role for the Btd domain of Sp1. Molecular and cellular biology *17*, 5193-5200.

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

Bain, G., and Gottlieb, D.I. (1994). Expression of retinoid X receptors in P19 embryonal carcinoma cells and embryonic stem cells. Biochemical and biophysical research communications *200*, 1252-1256.

Balmer, J.E., and Blomhoff, R. (2002). Gene expression regulation by retinoic acid. Journal of lipid research *43*, 1773-1808.

Basilico, C., Ambrosetti, D., Fraidenraich, D., and Dailey, L. (1997). Regulatory mechanisms governing FGF-4 gene expression during mouse development. Journal of cellular physiology *173*, 227-232.

Bastien, J., and Rochette-Egly, C. (2004). Nuclear retinoid receptors and the transcription of retinoid-target genes. Gene *328*, 1-16.

Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E.J., and Constam, D.B. (2006). The nodal precursor acting via activin receptors induces mesoderm by maintaining a source of its convertases and BMP4. Developmental cell *11*, 313-323.

Bergsland, M., Ramskold, D., Zaouter, C., Klum, S., Sandberg, R., and Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. Genes & development *25*, 2453-2464.

Blomhoff, R., and Blomhoff, H.K. (2006). Overview of retinoid metabolism and function. Journal of neurobiology *66*, 606-630.

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

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255-256.

Brennan, J., Lu, C.C., Norris, D.P., Rodriguez, T.A., Beddington, R.S., and Robertson, E.J. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. Nature *411*, 965-969.

Brewster, R., Lee, J., and Ruiz i Altaba, A. (1998). Gli/Zic factors pattern the neural plate by defining domains of cell differentiation. Nature *393*, 579-583.

Budhu, A.S., and Noy, N. (2002). Direct channeling of retinoic acid between cellular retinoic acid-binding protein II and retinoic acid receptor sensitizes mammary carcinoma cells to retinoic acid-induced growth arrest. Molecular and cellular biology *22*, 2632-2641.

Bylund, M., Andersson, E., Novitch, B.G., and Muhr, J. (2003). Vertebrate neurogenesis is counteracted by Sox1-3 activity. Nature neuroscience *6*, 1162-1168.

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

Cavodeassi, F., Modolell, J., and Gomez-Skarmeta, J.L. (2001). The Iroquois family of genes: from body building to neural patterning. Development (Cambridge, England) *128*, 2847-2855.

Chambers, I., and Tomlinson, S.R. (2009). The transcriptional foundation of pluripotency. Development (Cambridge, England) *136*, 2311-2322.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nature biotechnology *27*, 275-280.

Chang, C.W., Tsai, C.W., Wang, H.F., Tsai, H.C., Chen, H.Y., Tsai, T.F., Takahashi, H., Li, H.Y., Fann, M.J., Yang, C.W., *et al.* (2004). Identification of a developmentally regulated striatum-enriched zinc-finger gene, Nolz-1, in the mammalian brain. Proceedings of the National Academy of Sciences of the United States of America *101*, 2613-2618.

Cheah, P.Y., Meng, Y.B., Yang, X., Kimbrell, D., Ashburner, M., and Chia, W. (1994). The Drosophila l(2)35Ba/nocA gene encodes a putative Zn finger protein involved in the development of the embryonic brain and the adult ocellar structures. Molecular and cellular biology *14*, 1487-1499.

Chenoweth, J.G., McKay, R.D., and Tesar, P.J. (2010). Epiblast stem cells contribute new insight into pluripotency and gastrulation. Development, growth & differentiation *52*, 293-301.

Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T., *et al.* (2006). Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer cell *10*, 529-541.

Chin, S.F., Teschendorff, A.E., Marioni, J.C., Wang, Y., Barbosa-Morais, N.L., Thorne, N.P., Costa, J.L., Pinder, S.E., van de Wiel, M.A., Green, A.R., *et al.* (2007). High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome biology *8*, R215.

Chisaka, O., Musci, T.S., and Capecchi, M.R. (1992). Developmental defects of the ear, cranial nerves and hindbrain resulting from targeted disruption of the mouse homeobox gene Hox-1.6. Nature *355*, 516-520.

Chng, Z., Teo, A., Pedersen, R.A., and Vallier, L. (2010). SIP1 mediates cell-fate decisions between neuroectoderm and mesendoderm in human pluripotent stem cells. Cell stem cell *6*, 59-70.

Chung, A.C., Katz, D., Pereira, F.A., Jackson, K.J., DeMayo, F.J., Cooney, A.J., and O'Malley, B.W. (2001). Loss of orphan receptor germ cell nuclear factor function results in ectopic development of the tail bud and a novel posterior truncation. Molecular and cellular biology *21*, 663-677.

Collignon, J., Sockanathan, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P.N., and Lovell-Badge, R. (1996). A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. Development (Cambridge, England) *122*, 509-520.

Cooney, A.J., Hummelke, G.C., Herman, T., Chen, F., and Jackson, K.J. (1998). Germ cell nuclear factor is a response element-specific repressor of transcription. Biochemical and biophysical research communications *245*, 94-100.

D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nature biotechnology *23*, 1534-1541.

Davis, T., Ashburner, M., Johnson, G., Gubb, D., and Roote, J. (1997). Genetic and phenotypic analysis of the genes of the elbow-no-ocelli region of chromosome 2L of Drosophila melanogaster. Hereditas *126*, 67-75.

Diez del Corral, R., and Storey, K.G. (2004). Opposing FGF and retinoid pathways: a signalling switch that controls differentiation and patterning onset in the extending vertebrate body axis. BioEssays : news and reviews in molecular, cellular and developmental biology *26*, 857-869.

Dorfman, R., Glazer, L., Weihe, U., Wernet, M.F., and Shilo, B.Z. (2002). Elbow and Noc define a family of zinc finger proteins controlling morphogenesis of specific tracheal branches. Development (Cambridge, England) *129*, 3585-3596.

Dush, M.K., and Martin, G.R. (1992). Analysis of mouse Evx genes: Evx-1 displays graded expression in the primitive streak. Developmental biology *151*, 273-287.

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

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

Galan-Caridad, J.M., Harel, S., Arenzana, T.L., Hou, Z.E., Doetsch, F.K., Mirny, L.A., and Reizis, B. (2007). Zfx controls the self-renewal of embryonic and hematopoietic stem cells. Cell *129*, 345-357.

Gomez-Skarmeta, J.L., and Modolell, J. (2002). Iroquois genes: genomic organization and function in vertebrate neural development. Current opinion in genetics & development *12*, 403-408.

Gottlieb, D. (2002). Signposts on the neural path. Nature biotechnology *20*, 1208-1210.

Gottlieb, D.I., and Huettner, J.E. (1999). An in vitro pathway from embryonic stem cells to neurons and glia. Cells, tissues, organs *165*, 165-172.

Greber, B., Coulon, P., Zhang, M., Moritz, S., Frank, S., Muller-Molina, A.J., Arauzo-Bravo, M.J., Han, D.W., Pape, H.C., and Scholer, H.R. (2011). FGF signalling inhibits neural induction in human embryonic stem cells. The EMBO journal *30*, 4874-4884.

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

Guenther, M.G., Lawton, L.N., Rozovskaia, T., Frampton, G.M., Levine, S.S., Volkert, T.L., Croce, C.M., Nakamura, T., Canaani, E., and Young, R.A. (2008). Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. Genes & development *22*, 3403-3408.

Harmston, N., and Lenhard, B. (2013). Chromatin and epigenetic features of long-range gene regulation. Nucleic acids research *41*, 7185-7199.

Hart, S.M. (2002). Modulation of nuclear receptor dependent transcription. Biological research *35*, 295-303.

Hemmati-Brivanlou, A., Kelly, O.G., and Melton, D.A. (1994). Follistatin, an antagonist of activin, is expressed in the Spemann organizer and displays direct neuralizing activity. Cell *77*, 283-295.

Hirst, C.E., Ng, E.S., Azzola, L., Voss, A.K., Thomas, T., Stanley, E.G., and Elefanty, A.G. (2006). Transcriptional profiling of mouse and human ES cells identifies SLAIN1, a novel stem cell gene. Developmental biology *293*, 90-103.

Holland, D.G., Burleigh, A., Git, A., Goldgraben, M.A., Perez-Mancera, P.A., Chin, S.F., Hurtado, A., Bruna, A., Ali, H.R., Greenwood, W., *et al.* (2011). ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. EMBO molecular medicine *3*, 167-180.

Houtmeyers, R., Souopgui, J., Tejpar, S., and Arkell, R. (2013). The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. Cellular and molecular life sciences : CMLS *70*, 3791-3811.

Hoyle, J., Tang, Y.P., Wiellette, E.L., Wardle, F.C., and Sive, H. (2004). nlz gene family is required for hindbrain patterning in the zebrafish. Developmental dynamics : an official publication of the American Association of Anatomists *229*, 835-846.

Iuchi, S. (2001). Three classes of C2H2 zinc finger proteins. Cellular and molecular life sciences : CMLS *58*, 625-635.

Iulianella, A., Beckett, B., Petkovich, M., and Lohnes, D. (1999). A molecular basis for retinoic acid-induced axial truncation. Developmental biology *205*, 33-48.

Jackson, S.A., Schiesser, J., Stanley, E.G., and Elefanty, A.G. (2010). Differentiating embryonic stem cells pass through 'temporal windows' that mark responsiveness to exogenous and paracrine mesendoderm inducing signals. PloS one *5*, e10706.

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

Ji, S.J., Periz, G., and Sockanathan, S. (2009). Nolz1 is induced by retinoid signals and controls motoneuron subtype identity through distinct repressor activities. Development (Cambridge, England) *136*, 231-240.

Johnson, M.H. (2009). From mouse egg to mouse embryo: polarities, axes, and tissues. Annual review of cell and developmental biology *25*, 483-512.

Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tjian, R. (1987). Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell *51*, 1079-1090.

Kamiya, D., Banno, S., Sasai, N., Ohgushi, M., Inomata, H., Watanabe, K., Kawada, M., Yakura, R., Kiyonari, H., Nakao, K., *et al.* (2011). Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. Nature *470*, 503-509.

Kawaguchi, R., Yu, J., Honda, J., Hu, J., Whitelegge, J., Ping, P., Wiita, P., Bok, D., and Sun, H. (2007). A membrane receptor for retinol binding protein mediates cellular uptake of vitamin A. Science (New York, NY) *315*, 820-825.

Ko, H.A., Chen, S.Y., Chen, H.Y., Hao, H.J., and Liu, F.C. (2013). Cell type-selective expression of the zinc finger-containing gene Nolz-1/Zfp503 in the developing mouse striatum. Neuroscience letters *548*, 44-49.

Kroll, K.L., Salic, A.N., Evans, L.M., and Kirschner, M.W. (1998). Geminin, a neuralizing molecule that demarcates the future neural plate at the onset of gastrulation. Development (Cambridge, England) *125*, 3247-3258.

Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development (Cambridge, England) *134*, 2895-2902.

Kuo, J.S., Patel, M., Gamse, J., Merzdorf, C., Liu, X., Apekin, V., and Sive, H. (1998). Opl: a zinc finger protein that regulates neural determination and patterning in Xenopus. Development (Cambridge, England) *125*, 2867-2882.

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

Langston, A.W., Thompson, J.R., and Gudas, L.J. (1997). Retinoic acid-responsive enhancers located 3' of the Hox A and Hox B homeobox gene clusters. Functional analysis. The Journal of biological chemistry *272*, 2167-2175.

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

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

Liu, J.P., Laufer, E., and Jessell, T.M. (2001). Assigning the positional identity of spinal motor neurons: rostrocaudal patterning of Hox-c expression by FGFs, Gdf11, and retinoids. Neuron *32*, 997-1012.

Loh, Y.H., Yang, L., Yang, J.C., Li, H., Collins, J.J., and Daley, G.Q. (2011). Genomic approaches to deconstruct pluripotency. Annual review of genomics and human genetics *12*, 165-185.

Maden, M. (2007). Retinoic acid in the development, regeneration and maintenance of the nervous system. Nature reviews Neuroscience *8*, 755-765.

Mahony, S., Mazzoni, E.O., McCuine, S., Young, R.A., Wichterle, H., and Gifford, D.K. (2011). Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. Genome biology *12*, R2.

Mallo, M., Wellik, D.M., and Deschamps, J. (2010). Hox genes and regional patterning of the vertebrate body plan. Developmental biology *344*, 7-15.

Malureanu, L.A. (2011). Targeting vector construction through recombineering. Methods in molecular biology (Clifton, NJ) *693*, 181-203.

Mao, C.D., and Byers, S.W. (2011). Cell-context dependent TCF/LEF expression and function: alternative tales of repression, de-repression and activation potentials. Critical reviews in eukaryotic gene expression *21*, 207-236.

Marikawa, Y. (2006). Wnt/beta-catenin signaling and body plan formation in mouse embryos. Seminars in cell & developmental biology *17*, 175-184.

Mark, M., Lufkin, T., Vonesch, J.L., Ruberte, E., Olivo, J.C., Dolle, P., Gorry, P., Lumsden, A., and Chambon, P. (1993). Two rhombomeres are altered in Hoxa-1 mutant mice. Development (Cambridge, England) *119*, 319-338.

Marletaz, F., Holland, L.Z., Laudet, V., and Schubert, M. (2006). Retinoic acid signaling and the evolution of chordates. International journal of biological sciences *2*, 38-47.

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 of the United States of America *78*, 7634-7638.

Martinez-Ceballos, E., Chambon, P., and Gudas, L.J. (2005). Differences in gene expression between wild type and Hoxa1 knockout embryonic stem cells after retinoic acid treatment or leukemia inhibitory factor (LIF) removal. The Journal of biological chemistry *280*, 16484-16498.

Martinez-Ceballos, E., and Gudas, L.J. (2008). Hoxa1 is required for the retinoic acid-induced differentiation of embryonic stem cells into neurons. Journal of neuroscience research *86*, 2809-2819.

Mazzoni, E.O., Mahony, S., Iacovino, M., Morrison, C.A., Mountoufaris, G., Closser, M., Whyte, W.A., Young, R.A., Kyba, M., Gifford, D.K., *et al.* (2011). Embryonic stem cell-based mapping of developmental transcriptional programs. Nature methods *8*, 1056-1058.

Moody, S.A., Klein, S.L., Karpinski, B.A., Maynard, T.M., and Lamantia, A.S. (2013). On becoming neural: what the embryo can tell us about differentiating neural stem cells. American journal of stem cells *2*, 74-94.

Morrison, K.M., Miesegaes, G.R., Lumpkin, E.A., and Maricich, S.M. (2009). Mammalian Merkel cells are descended from the epidermal lineage. Developmental biology *336*, 76-83.

Murata, Y., Kim, H.G., Rogers, K.T., Udvadia, A.J., and Horowitz, J.M. (1994). Negative regulation of Sp1 trans-activation is correlated with the binding of cellular proteins to the amino terminus of the Sp1 trans-activation domain. The Journal of biological chemistry *269*, 20674-20681.

Nakamura, M., Runko, A.P., and Sagerstrom, C.G. (2004). A novel subfamily of zinc finger genes involved in embryonic development. Journal of cellular biochemistry *93*, 887-895.

Ng, E.S., Azzola, L., Sourris, K., Robb, L., Stanley, E.G., and Elefanty, A.G. (2005). The primitive streak gene Mixl1 is required for efficient haematopoiesis and BMP4-induced ventral mesoderm patterning in differentiating ES cells. Development (Cambridge, England) *132*, 873-884.

Niakan, K.K., Schrode, N., Cho, L.T., and Hadjantonakis, A.K. (2013). Derivation of extraembryonic endoderm stem (XEN) cells from mouse embryos and embryonic stem cells. Nature protocols *8*, 1028-1041.

Nichols, J., and Smith, A. (2012). Pluripotency in the embryo and in culture. Cold Spring Harbor perspectives in biology *4*, a008128.

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

Niederreither, K., Vermot, J., Schuhbaur, B., Chambon, P., and Dolle, P. (2000). Retinoic acid synthesis and hindbrain patterning in the mouse embryo. Development (Cambridge, England) *127*, 75-85.

Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R.O., Ogonuki, N., *et al.* (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. Developmental cell *16*, 398-410.

Niwa, H. (2010). Mouse ES cell culture system as a model of development. Development, growth & differentiation *52*, 275-283.

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

Novitch, B.G., Wichterle, H., Jessell, T.M., and Sockanathan, S. (2003). A requirement for retinoic acid-mediated transcriptional activation in ventral neural patterning and motor neuron specification. Neuron *40*, 81-95.

Nusse, R., Fuerer, C., Ching, W., Harnish, K., Logan, C., Zeng, A., ten Berge, D., and Kalani, Y. (2008). Wnt signaling and stem cell control. Cold Spring Harbor symposia on quantitative biology *73*, 59-66.

Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. (2005). Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. The Journal of biological chemistry *280*, 5307-5317.

Omichinski, J.G., Pedone, P.V., Felsenfeld, G., Gronenborn, A.M., and Clore, G.M. (1997). The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. Nature structural biology *4*, 122-132.

Ozair, M.Z., Kintner, C., and Brivanlou, A.H. (2013). Neural induction and early patterning in vertebrates. Wiley interdisciplinary reviews Developmental biology *2*, 479-498.

Pedone, P.V., Ghirlando, R., Clore, G.M., Gronenborn, A.M., Felsenfeld, G., and Omichinski, J.G. (1996). The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding. Proceedings of the National Academy of Sciences of the United States of America *93*, 2822-2826.

Pedone, P.V., Omichinski, J.G., Nony, P., Trainor, C., Gronenborn, A.M., Clore, G.M., and Felsenfeld, G. (1997). The N-terminal fingers of chicken GATA-2 and GATA-3 are independent sequence-specific DNA binding domains. The EMBO journal *16*, 2874-2882.

Pereira-Castro, I., Costa, A.M., Oliveira, M.J., Barbosa, I., Rocha, A.S., Azevedo, L., and da Costa, L.T. (2013). Characterization of human NLZ1/ZNF703 identifies conserved domains essential for proper subcellular localization and transcriptional repression. Journal of cellular biochemistry *114*, 120-133.

Pereira, L.A., Wong, M.S., Mei Lim, S., Stanley, E.G., and Elefanty, A.G. (2012). The Mix family of homeobox genes--key regulators of mesendoderm formation during vertebrate development. Developmental biology *367*, 163-177.

Pierani, A., Brenner-Morton, S., Chiang, C., and Jessell, T.M. (1999). A sonic hedgehog-independent, retinoid-activated pathway of neurogenesis in the ventral spinal cord. Cell *97*, 903-915.

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

Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., and Hadjantonakis, A.K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development (Cambridge, England) *135*, 3081-3091.

Reynisdottir, I., Arason, A., Einarsdottir, B.O., Gunnarsson, H., Staaf, J., Vallon-Christersson, J., Jonsson, G., Ringner, M., Agnarsson, B.A., Olafsdottir, K., *et al.* (2013). High expression of ZNF703 independent of amplification indicates worse prognosis in patients with luminal B breast cancer. Cancer medicine *2*, 437-446.

Rivera-Perez, J.A., and Magnuson, T. (2005). Primitive streak formation in mice is preceded by localized activation of Brachyury and Wnt3. Developmental biology *288*, 363-371.

Robb, L., Hartley, L., Begley, C.G., Brodnicki, T.C., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., and Elefanty, A.G. (2000). Cloning, expression analysis, and chromosomal localization of murine and human homologues of a Xenopus mix gene. Developmental dynamics : an official publication of the American Association of Anatomists *219*, 497-504.

Rogers, C.D., Harafuji, N., Archer, T., Cunningham, D.D., and Casey, E.S. (2009). Xenopus Sox3 activates sox2 and geminin and indirectly represses Xvent2 expression to induce neural progenitor formation at the expense of non-neural ectodermal derivatives. Mechanisms of development *126*, 42-55.

Rosa, A., and Brivanlou, A.H. (2011). A regulatory circuitry comprised of miR-302 and the transcription factors OCT4 and NR2F2 regulates human embryonic stem cell differentiation. The EMBO journal *30*, 237-248.

Ross, A.C., and Zolfaghari, R. (2011). Cytochrome P450s in the regulation of cellular retinoic acid metabolism. Annual review of nutrition *31*, 65-87.

Rossant, J., and Tam, P.P. (2004). Emerging asymmetry and embryonic patterning in early mouse development. Developmental cell *7*, 155-164.

Runko, A.P., and Sagerstrom, C.G. (2003). Nlz belongs to a family of zinc-finger-containing repressors and controls segmental gene expression in the zebrafish hindbrain. Developmental biology *262*, 254-267.

Runko, A.P., and Sagerstrom, C.G. (2004). Isolation of nlz2 and characterization of essential domains in Nlz family proteins. The Journal of biological chemistry *279*, 11917-11925.

Sandell, L.L., Sanderson, B.W., Moiseyev, G., Johnson, T., Mushegian, A., Young, K., Rey, J.P., Ma, J.X., Staehling-Hampton, K., and Trainor, P.A. (2007). RDH10 is essential for synthesis of embryonic retinoic acid and is required for limb, craniofacial, and organ development. Genes & development *21*, 1113-1124.

Sandhu, K.S., Li, G., Poh, H.M., Quek, Y.L., Sia, Y.Y., Peh, S.Q., Mulawadi, F.H., Lim, J., Sikic, M., Menghi, F., *et al.* (2012). Large-scale functional organization of long-range chromatin interaction networks. Cell reports *2*, 1207-1219.

Sasai, Y., Lu, B., Steinbeisser, H., Geissert, D., Gont, L.K., and De Robertis, E.M. (1994). Xenopus chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes. Cell *79*, 779-790.

Schoorlemmer, J., van Puijenbroek, A., van Den Eijnden, M., Jonk, L., Pals, C., and Kruijer, W. (1994). Characterization of a negative retinoic acid response element in the murine Oct4 promoter. Molecular and cellular biology *14*, 1122-1136.

Seo, S., Herr, A., Lim, J.W., Richardson, G.A., Richardson, H., and Kroll, K.L. (2005). Geminin regulates neuronal differentiation by antagonizing Brg1 activity. Genes & development *19*, 1723-1734.

Shahhoseini, M., Taei, A., Mehrjardi, N.Z., Salekdeh, G.H., and Baharvand, H. (2010). Epigenetic analysis of human embryonic carcinoma cells during retinoic acid-induced neural differentiation. Biochemistry and cell biology = Biochimie et biologie cellulaire *88*, 527-538.

Sheng, G., dos Reis, M., and Stern, C.D. (2003). Churchill, a zinc finger transcriptional activator, regulates the transition between gastrulation and neurulation. Cell *115*, 603-613.

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

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722-737.

Sircoulomb, F., Nicolas, N., Ferrari, A., Finetti, P., Bekhouche, I., Rousselet, E., Lonigro, A., Adelaide, J., Baudelet, E., Esteyries, S., *et al.* (2011). ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. EMBO molecular medicine *3*, 153-166.

Skarnes, W.C., Rosen, B., West, A.P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A.O., Thomas, M., Harrow, J., Cox, T., *et al.* (2011). A conditional knockout resource for the genome-wide study of mouse gene function. Nature *474*, 337-342.

Slorach, E.M., Chou, J., and Werb, Z. (2011). Zeppo1 is a novel metastasis promoter that represses E-cadherin expression and regulates p120-catenin isoform expression and localization. Genes & development *25*, 471-484.

Smith, A. (2009). Design principles of pluripotency. EMBO molecular medicine *1*, 251-254.

Smith, E., and Shilatifard, A. (2014). Enhancer biology and enhanceropathies. Nature structural & molecular biology *21*, 210-219.

Smith, W.C., and Harland, R.M. (1992). Expression cloning of noggin, a new dorsalizing factor localized to the Spemann organizer in Xenopus embryos. Cell *70*, 829-840.

Sockanathan, S., and Jessell, T.M. (1998). Motor neuron-derived retinoid signaling specifies the subtype identity of spinal motor neurons. Cell *94*, 503-514.

Soprano, D.R., Teets, B.W., and Soprano, K.J. (2007). Role of retinoic acid in the differentiation of embryonal carcinoma and embryonic stem cells. Vitamins and hormones *75*, 69-95.

Stern, C.D. (2005). Neural induction: old problem, new findings, yet more questions. Development (Cambridge, England) *132*, 2007-2021.

Streit, A., Berliner, A.J., Papanayotou, C., Sirulnik, A., and Stern, C.D. (2000). Initiation of neural induction by FGF signalling before gastrulation. Nature *406*, 74-78.

Su, K., Roos, M.D., Yang, X., Han, I., Paterson, A.J., and Kudlow, J.E. (1999). An N-terminal region of Sp1 targets its proteasome-dependent degradation in vitro. The Journal of biological chemistry *274*, 15194-15202.

Suske, G. (1999). The Sp-family of transcription factors. Gene *238*, 291-300.

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

Tam, P.P., and Loebel, D.A. (2007). Gene function in mouse embryogenesis: get set for gastrulation. Nature reviews Genetics *8*, 368-381.

Thiery, J.P., Acloque, H., Huang, R.Y., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871-890.

Thomson, M., Liu, S.J., Zou, L.N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell *145*, 875-889.

Tropepe, V., Hitoshi, S., Sirard, C., Mak, T.W., Rossant, J., and van der Kooy, D. (2001). Direct neural fate specification from embryonic stem cells: a primitive mammalian neural stem cell stage acquired through a default mechanism. Neuron *30*, 65-78.

Urban, N., Martin-Ibanez, R., Herranz, C., Esgleas, M., Crespo, E., Pardo, M., Crespo-Enriquez, I., Mendez-Gomez, H.R., Waclaw, R., Chatzi, C.*, et al.* (2010). Nolz1 promotes striatal neurogenesis through the regulation of retinoic acid signaling. Neural development *5*, 21.

Van Keymeulen, A., Mascre, G., Youseff, K.K., Harel, I., Michaux, C., De Geest, N., Szpalski, C., Achouri, Y., Bloch, W., Hassan, B.A.*, et al.* (2009). Epidermal progenitors give rise to Merkel cells during embryonic development and adult homeostasis. The Journal of cell biology *187*, 91-100.

Wada, T., Honda, M., Minami, I., Tooi, N., Amagai, Y., Nakatsuji, N., and Aiba, K. (2009). Highly efficient differentiation and enrichment of spinal motor neurons derived from human and monkey embryonic stem cells. PloS one *4*, e6722.

Wang, K., Sengupta, S., Magnani, L., Wilson, C.A., Henry, R.W., and Knott, J.G. (2010). Brg1 is required for Cdx2-mediated repression of Oct4 expression in mouse blastocysts. PloS one *5*, e10622.

Wegner, M. (2011). SOX after SOX: SOXession regulates neurogenesis. Genes & development *25*, 2423-2428.

Wei, L.N. (2003). Retinoid receptors and their coregulators. Annual review of pharmacology and toxicology *43*, 47-72.

Wichterle, H., Lieberam, I., Porter, J.A., and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. Cell *110*, 385-397.

Wilson, S.I., Graziano, E., Harland, R., Jessell, T.M., and Edlund, T. (2000). An early requirement for FGF signalling in the acquisition of neural cell fate in the chick embryo. Current biology : CB *10*, 421-429.

Winnier, G., Blessing, M., Labosky, P.A., and Hogan, B.L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. Genes & development *9*, 2105-2116.

Wolfe, S.A., Nekludova, L., and Pabo, C.O. (2000). DNA recognition by Cys2His2 zinc finger proteins. Annual review of biophysics and biomolecular structure *29*, 183-212.

Woodruff, R.C., and Ashburner, M. (1979a). The genetics of a small autosomal region of Drosophila melanogaster containing the structural gene for alcohol dehydrogenase. I. Characterization of deficiencies and mapping of ADH and visible mutations. Genetics *92*, 117-132.

Woodruff, R.C., and Ashburner, M. (1979b). The genetics of a small autosomal region of Drosophila melanogaster containing the structural gene for alcohol dehydrogenase. II. Lethal mutations in the region. Genetics *92*, 133-149.

Yang, G., Ma, F., Zhong, M., Fang, L., Peng, Y., Xin, X., Zhong, J., Yuan, F., Gu, H., Zhu, W., *et al.* (2014). ZNF703 acts as an oncogene that promotes progression in gastric cancer. Oncology reports *31*, 1877-1882.

Ying, Q.L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nature biotechnology *21*, 183-186.

Zhao, X., Yang, Y., Fitch, D.H., and Herman, M.A. (2002). TLP-1 is an asymmetric cell fate determinant that responds to Wnt signals and controls male tail tip morphogenesis in C. elegans. Development (Cambridge, England) *129*, 1497-1508.